



Managing the risks of the different access routes for sensitive data

Meet the Experts – GESIS online talks

Data Services, Data Archiving, and Research Data Management

Deborah Wiltshire, Secure Data Center , 14 December 2023

Today's Speaker



Deborah Wiltshire

- I'm originally a historical Demographer and Social Scientist, primarily teaching quantitative research methods with an interest in compiling historic trend data. For the last 10 years I've been working with sensitive data, and now head up the Secure Data Center where I specialise in data governance and statistical disclosure control. I regularly train and advise the research community in these areas, with my research focusing on ethical data governance as well as women's histories.
- Data governance, ethical data use, historical demography
- Contact: deborah.wiltshire@gesis.org

Data Services for the Social Sciences (DSS)

Data Services

Community Data
Collection

Metadata
& PIDs

RDM
& RDM Training

DP-R | EX

Risk
& Sensitive Data

Certification

KonsortSWD 
Konsortium für die
Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften



▼ Archiving

▼ Data Acquisitions and Access

▼ Metadata Standards and Interoperability



Data Services, Data Archiving, and Research Data Management

Certification	14.09.2023	Certification for data archives and research data centers <i>Jonas Recker, Kerstin Beck</i>
Data Services	28.09.2023	Data Services: An Overview <i>Oliver Watteler</i>
Community Data Collection	12.10.2023	Community Data Collection <i>Alexander Jedinger</i>
Metadata & PIDs	26.10.2023	Metadata and PIDs <i>Wolfgang Zenk-Möltgen, Jan Schwalbach & Kokila Jamwal</i>
DP-R EX	09.11.2023	An Introduction to Domain-Specific Data Infrastructures: DP-R EX <i>Alexander Jedinger, Marlene Hilgenstock & Pascal Kolkwitz-Anstötz</i>
RDM & RDM Training	23.11.2023	Research Data Management & RDM-Training <i>Anja Perry & Sebastian Netscher</i>
Risk & Sensitive Data	14.12.2023	Secure Data Center: Secure Access to Sensitive Data <i>Deborah Wiltshire & Jara Kampmann</i>

The roadmap for today's session

- What is sensitive data?
- The role of Trusted Research Environments (TREs) in enabling safe access to sensitive data
- Introducing the Secure Data Center & our work with secure data access
- The role of the 5 Safes Framework
- How it can be utilized in managing our move to new access routes



Thinking about sensitive 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

The role of Trusted Research Environments



Why are
TREs
important?

Making data available through a TRE means people can be **confident** that their **personal health data** is accessed **securely** and their **privacy protected**.

HDRUK
Health Data Research UK

A quick history to secure data access

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

The Secure Data Center at GESIS

- A developing Trusted Research Environment (TRE)
- Provide access to sensitive data from key German social surveys & digital behavioural data
- Also provide access to UK and French sensitive data
- Based in Cologne
- Primary access is on-site via our Safe Room
- On-site access also possible in Mannheim and the UK
- Development work underway for remote access systems



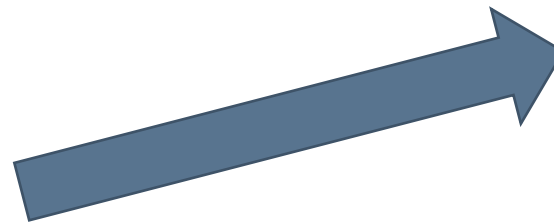
The basic model of access



**Secure Server in
location A**

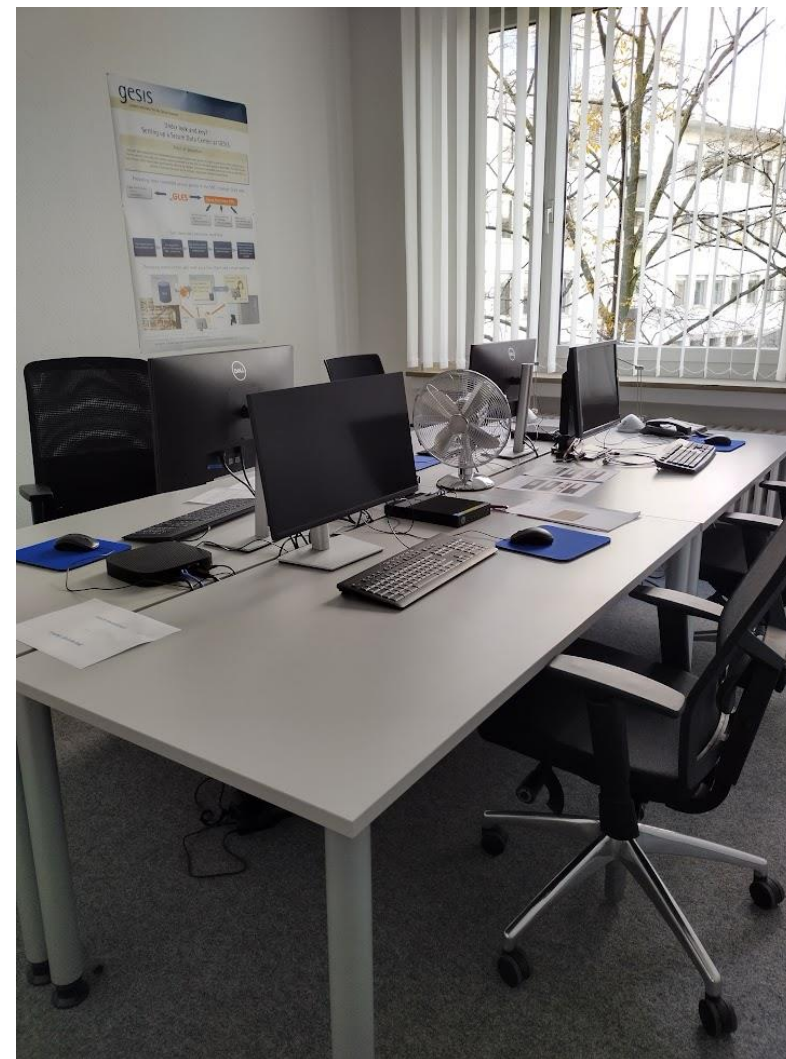


Person in Location B



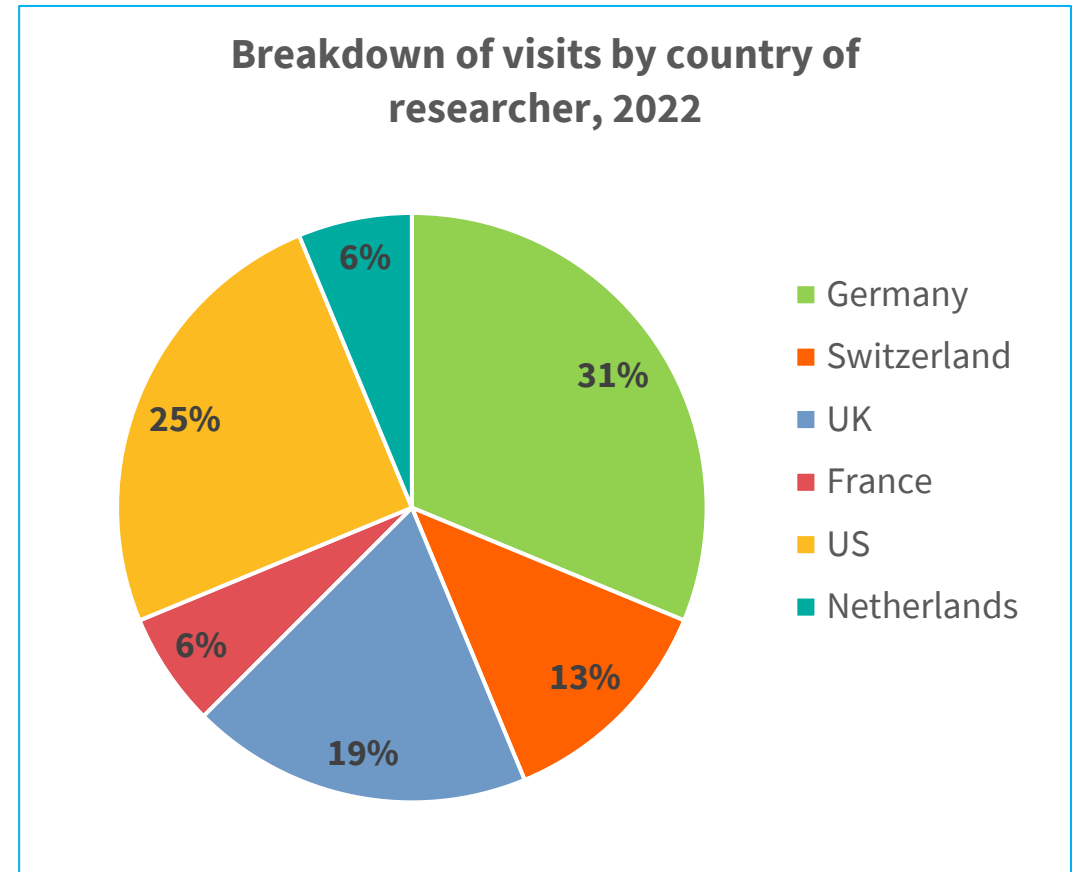
Safe Room Access

- 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



The downside to Safe Rooms

- Travel times & costs
- Restrictions on when you can work
- Limited availability
- Corona pandemic & other unforeseen events



Moving to remote access at the Secure Data Center

Option 1 - Remote Access

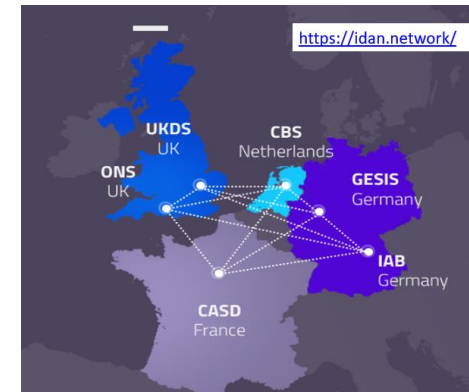
- Still based on Safe Room access
- Safe Room is at a partner organisation
- Access via bilateral agreements & secure technical connections
- Retains the physical controls of Safe Room access, but offers more flexibility



Location B
=
**on premises of partner
organisation**

Advantages of Remote Access

- Maintains physical controls of the Safe Room environment
- More capacity
- More likely to have a location closer to you
- “Easier” to implement



Use Study: SSHOC Remote Connection

"Our visit to the data centre went really smoothly. Instead of needing to travel to Germany, we were able to access excellent but restricted GESIS data from just down the road. The teams at both ends were really helpful throughout, and we are already thinking of new projects that allow us to make use of this fantastic data resource."

Option 2 - Remote Desktop Access

- Access is via secure encrypted internet connection from their own office
- Many advantages!
 - No need to travel
 - Can work when you want
 - Expands our capacity
- Some things to think about though...
 - More complex to implement
 - Greater capacity = more resources
 - Lose many of the physical controls of Safe Room access



Location B

=

Researchers own office

The 5 Safes Framework



Clearly specified project with valid statistical purpose

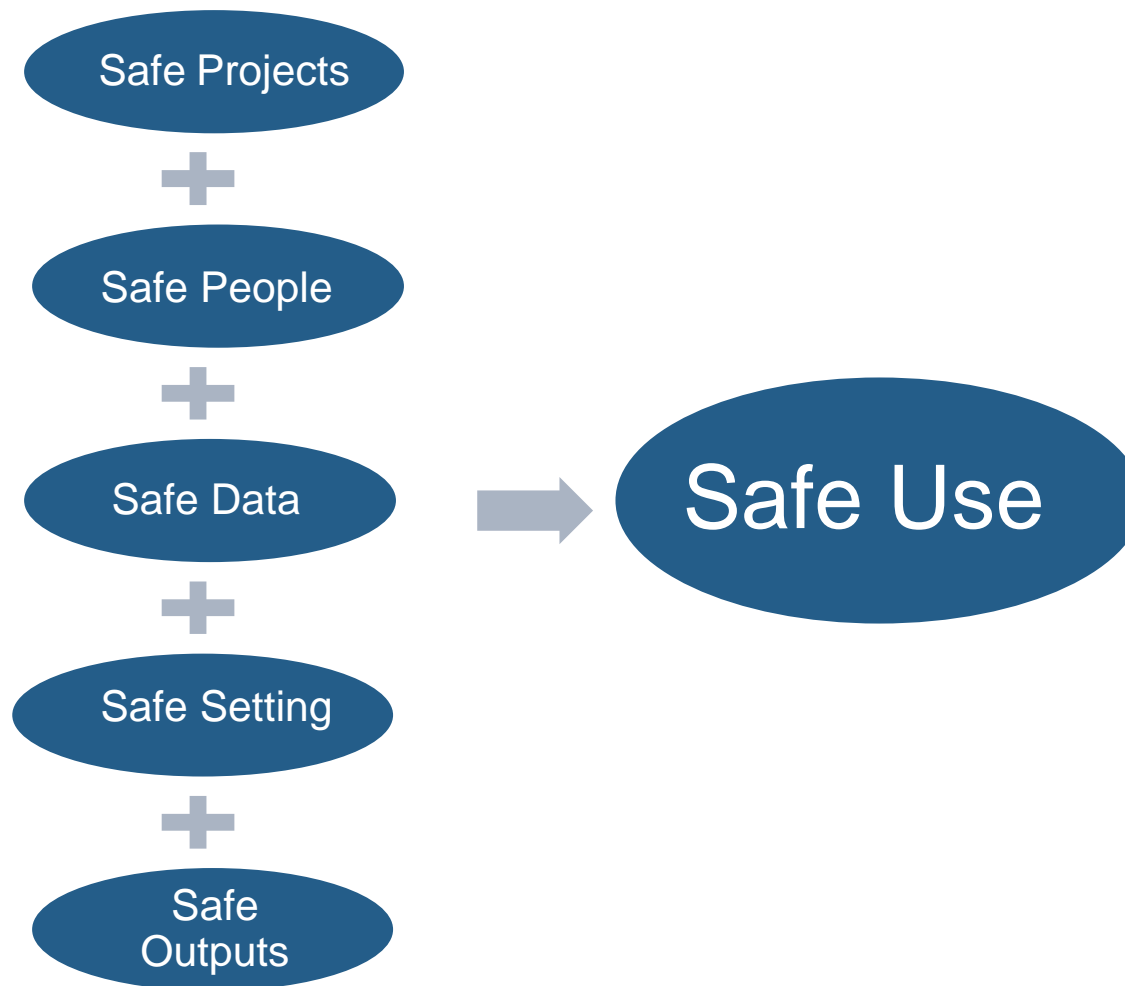
The right knowledge & experience to work with the data

→ 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

The 5 Safes Framework



Safe Setting - the technical set up

The working environment remains the same

But – we lose some controls of the Safe Room

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



We can add some extra technological controls in...

- 2-factor authentication
- Requirements for work station
 - Private office
 - Fixed IP address
- But we need to add in some non-technical controls like training, legal agreements

Non-technical controls – Safe Projects

- Agreements between data service & researcher
- Set out –
 - who can access what data,
 - for what purpose
 - & for how long
- Terms & conditions
- Is our current DUA sufficient?
- New licence compliance policy
- Institutional agreements?

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 –

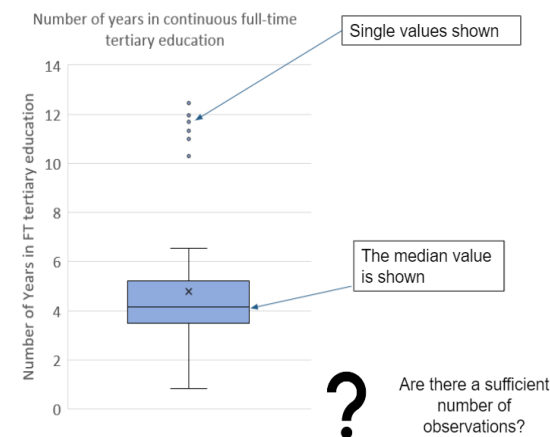
and

Last name	<input type="text"/>
First name	<input type="text"/>
E-mail	<input type="text"/>
Telephone number	<input type="text"/>
Institution	<input type="text"/>
Business address	<input type="text"/>
Position of data recipient ¹	<input type="text"/>

Non-technical controls – Safe People

- Researchers don't always have the knowledge necessary for working with secure data
- Researchers don't always read the instructions
- Specific training on statistical disclosure is often lacking in RDM training
- Training vital for remote working!
- 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

Box plots



- Shows how data are distributed
- Maxima, minima, and outliers may be an SDC risk
- This may also be true for the median.
- E.g, there are outliers who spent between 10 and 13 years in full time tertiary education
- Outliers stand out and may indicate unusual characteristics that aid identification

Accrediting Safe Use of Research Environments (ASSURED)

<i>Core Modules (Mandatory)</i>	<i>Service Specific Modules</i>	<i>Data-type Specific Modules</i>	<i>Role Specific Modules</i>
Introduction to Research Data Protection	How to use the Bundesbank Research Platform	Safe Handling of Genomics Data	Guidance for Output Checkers
The 5 Safes	How to use the GESIS Safe Haven	Principles for Protecting Qualitative Data	Guidance for Data Producers
Introduction to Output SDC	How to use GHGA Cloud	Performing Safe data Linkage	Guidance for Service Managers

Introduction to Safe Outputs

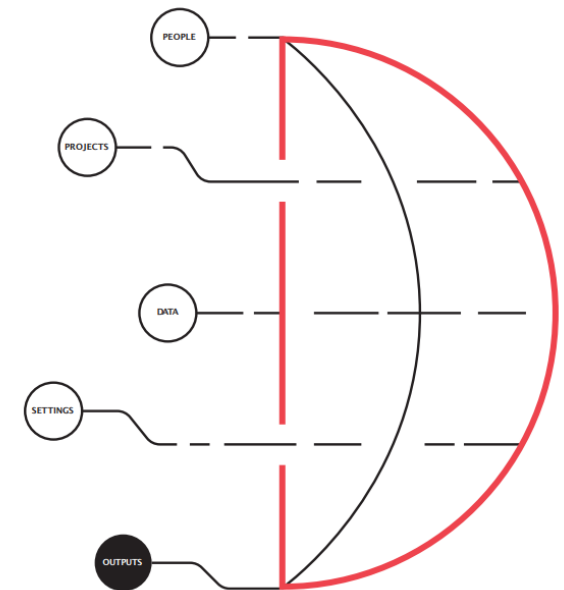
- The aim is to minimise the risk of an individual being identified, or assigning an attribute to someone, from a piece of analysis
- Residual risk in published results
- Statistical Disclosure Control (SDC) is a key method of doing so
- “The unprovability of safety”
- The aim is to demonstrate that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

Statistical Disclosure Control & statistical quality

- SDC is a set of sensitivity rules that are applied to outputs before release
- SDC is applied to research outputs before release or publication
- Generally a '4 eyes' approach is best practice
- Time-consuming!
- Currently working with a project in the UK to test semi-automated output checking
- Working with the SDAP team to update the SDC Handbook

Handbook on Statistical Disclosure Control for Outputs

Emily Griffiths (University of Manchester)
Carlotta Greci (The Health Foundation)
Yannis Kotrotsios (Cancer Research UK)
Simon Parker (Cancer Research UK)
James Scott (UK Data Archive, University of Essex)
Richard Welpton (The Health Foundation)
Arne Wolters (The Health Foundation)
Christine Woods (UK Data Archive, University of Essex)



What's next for the Secure Data Center?

- Expansion of secure data access
 - ▶ Easier and more flexible access
 - ▶ More data available
 - ▶ New data forms - DVD data
- Dissolving of boundaries
 - ▶ International boundaries
 - ▶ Disciplinary boundaries

Some final thoughts



Expert contact & GESIS consulting



Contact: you can reach the speaker/s via e-mail:
deborah.wiltshire@gesis.org

GESIS Consulting: GESIS offers individual consulting in a number of areas – including survey design & methodology, data archiving, digital behavioral data & computational social science – and across the research data cycle.

Please visit our website www.gesis.org for more [detailed information](#) on available services and terms.

Upcoming talks

- Please visit our meet-the-experts website:
 - <https://www.gesis.org/en/services/sharing-knowledge/consulting-and-guidelines/meet-the-experts>

Thank you for participating!