

# gesis

Leibniz Institute  
for the Social Sciences



## Secure Data Solutions for Social Media Data Analysis

IASSIST 2017, Lawrence, Kansas (USA)

*David Schiller (GESIS), Katharina Kinder-Kurlanda (GESIS)*

# GESIS

- Largest infrastructure institution for the Social Sciences in Germany
- Renders substantial, nationally and internationally relevant research-based infrastructure services
- Established 1986, restructured 2007
- Two locations: Cologne and Mannheim , more than 300 employees
- Five scientific departments:
  - ▶ Survey Design and Methodology
  - ▶ Monitoring Society and Social Change
  - ▶ Knowledge Technologies for the Social Sciences
  - ▶ Computational Social Science
  - ▶ Data Archive for the Social Sciences (DAS)

# Social Media Data



But also smaller social  
media provider, e.g.  
comments in newspapers.



# GESIS and Social Media Data I

- The “**Data Archive for the Social Sciences**” is providing data service for national and international comparative surveys from the fields of social and political science research. These surveys are archived and processed according to **internationally recognized standards** and made accessible to the scientifically interested public in a user-friendly manner.
- **User advice** is provided on finding relevant data and on data analysis, and researchers are introduced to complex techniques of data analysis in our training seminars.
- Substantive and methodological research is continuously carried out in the form of **secondary analyses** that demonstrate the analytical potential of the archived surveys.
- A Secure Data Center was established to allow **research with sensitive data**.

## GESIS and Social Media Data II

- In a nutshell: the data archive provides research infrastructures and services that assist in maintaining **good scientific practice**.
- Thereby GESIS has to adjust to new research needs and new data sources
- Research with Social Media Data is one of those new challenges
- According to that GESIS is working on an infrastructure to enable **good scientific practice** for research with social media data.

# Working with Social Media Data I

Such an infrastructure requires a number of specific modules:

- System for collection Social Media Data
- System for storing Social Media Data
- System for Social Media Data Documentation
- System to reference Social Media Data (DOI)
- System to Link Social Media Data to other sources
- Training for professional use of Social Media Data
  - ▶ Methods, reproducibility, reuse, security, legal issues

# Working with Social Media Data II

Goal: GESIS as “scientific privat cloud” for Social Media Analysis

- Researchers should be assisted
- Data collections should be replicable
- Good scientific practice should be ensured
- Data should be ready for reuse
- Privacy and ownership should be respected
- Social Media Data as Research Data should be “FAIR” (H2020):
  - ▶ Findable, Accessible, Interoperable, Re-useable

# Secure Access to Social Media Data

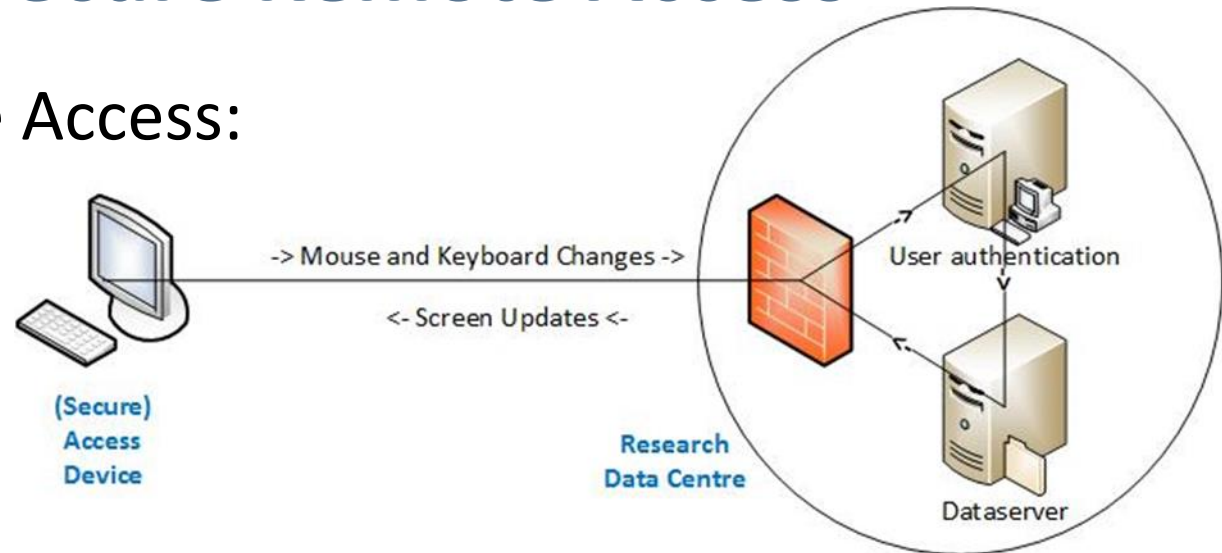
But why secure access?

- Depending on the Social Media Source, content could be sensitive (re-identification)
  - Sources may be hard or impossible to anonymize
  - Ownership issues may hinder un-secured access
  - Ownership issues may hinder un-secured reuse
- 
- Secure access may go hand in hand with convenient access

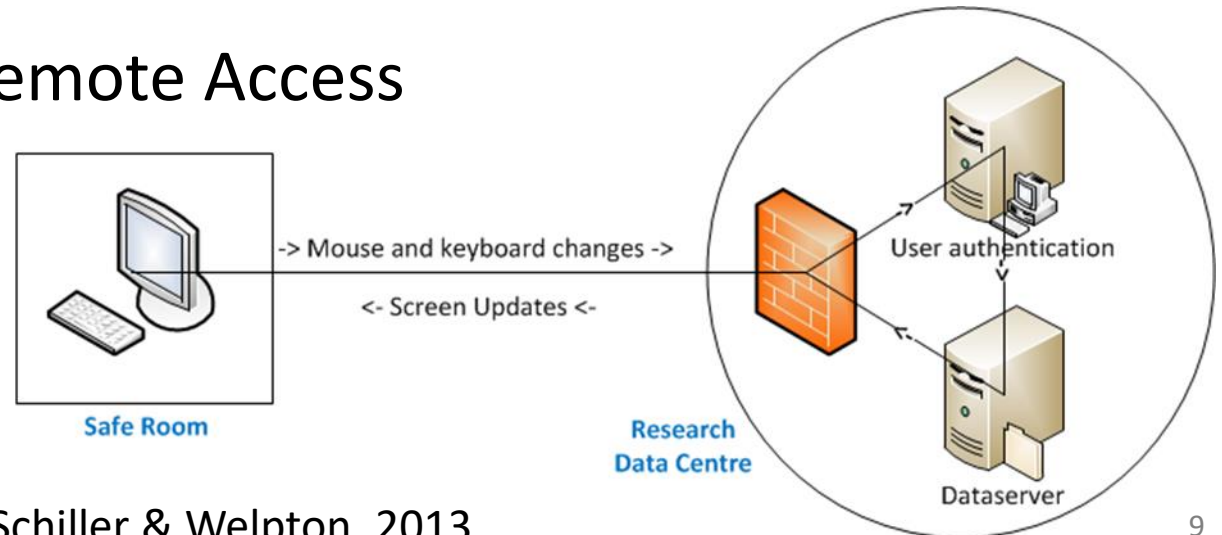


# Secure Remote Access

## Pure Remote Access:



## Safe Room Remote Access



# The (German) Law

**Microdata for scientific research: „provide [...] with individual data if [...] (de facto anonymised individual data),“ (BStatG §16 (6))**

- Interpretation: Remote Access is “providing” data (transmission). Closer look needed!
- It is data transmission but...
- It is not automatically research data transmission and...
- It is not automatically sensitive research data transmission
  
- Look at the research data:
  - ▶ surveys, geo-, administrative-, aggregated data etc.
- Always keep the „5 Saves“ in mind!

## Need for international access

- Solving the interpretation issue on a national (German level) is crucial but not enough
- Research and also Social Media Data research should not be limited to national borders
- Remote Access makes research from location far away easier
- Research that sometimes were not possible otherwise
- According to that: need for international agreements

## Summary

- Social Media Data research is increasing – that is good!
- But it has to be done in line with good scientific practice
- Infrastructures to support researchers need to be in place
- Legal issues have to be solved on a national and international level
- Re-use and Open Science should be ensured

gesis

Leibniz Institute  
for the Social Sciences



david.schiller@gesis.org, Katharina.kinder-kurlanda@gesis.org, [www.gesis.org](http://www.gesis.org)