

gesis

Leibniz Institute  
for the Social Sciences

# Konzept für die Definition Signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten

Natascha Schumann, Astrid Recker,  
Stefan Müller, Reiner Mauer  
GESIS Datenarchiv für  
Sozialwissenschaften

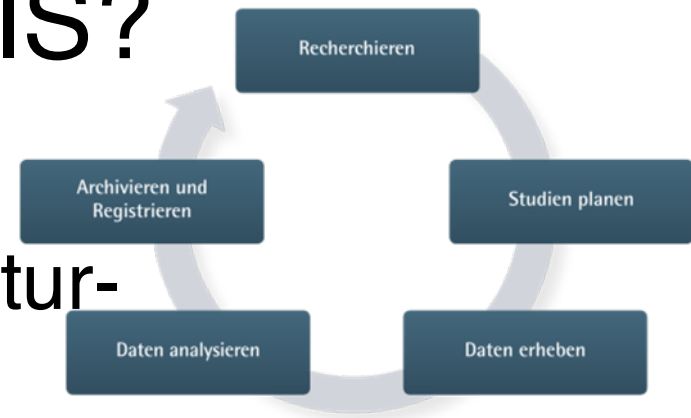
# Gliederung

- GESIS Datenarchiv
  - ▶ Aufgaben und Workflow
- Signifikante Eigenschaften
  - ▶ Konzepte
- Signifikante Eigenschaften sozialwissenschaftlicher Forschungsdaten

# Forschungsdaten in den Sozialwissenschaften

- Sozialwissenschaften: Verschiedene Fächer und „Bindestrichsoziologien“
- Vielzahl unterschiedlicher Methoden zur Datengewinnung
- Studien sind oftmals ....
  - ▶ heterogen und bestehen aus mehreren Objekten, die in ihrer Zusammensetzung variieren (ein oder mehrere Datensätze, Messinstrumente, Metadaten, begleitende Materialien ... )
  - ▶ nicht statisch: Daten und Metadaten werden im Archiv (kontinuierlich) verändert bzw. erzeugt (korrigiert, um neue Datenpunkte erweitert, integriert, aufgewertet)

# Was ist GESIS?



- Forschungs- und Infrastrukturanrichtung für die Sozialwissenschaften, Schwerpunkt empirische Sozialforschung

- ▶ **Forschungsdaten** (Erhebung, Archivierung, Registrierung, Analyse)
- ▶ **Integrierte Fachinformationen** (Daten, Literatur, Projekte)
- ▶ **Methodenberatung und –entwicklung**
- ▶ **Wissensvermittlung** (Summer School, Workshops, Seminare etc.)

# GESIS Datenarchiv

- Bereitstellung quantitativer Forschungsdaten, vorwiegend aus nationalen und internationalen vergleichenden Umfragen
- Forschungsdaten werden gemäß klar definierter Anforderungen ausgewählt, aufbereitet, dokumentiert, langfristig gesichert und zugänglich gemacht

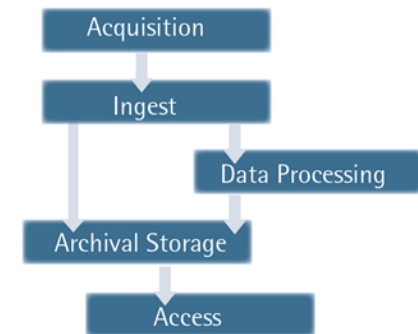


# GESIS Datenarchiv

- Langzeitarchivierung und Sicherung der Daten für die Nachnutzung
- Dienstleister für Primärforscher\*innen bei der Sicherung, Dokumentation, Aufwertung und Bereitstellung ihrer Daten
- Besonderes Augenmerk liegt auf der Einhaltung von datenschutzrechtlichen Bestimmungen

# Workflow Datenarchivierung

- **Akquise** (aktiv, passiv)
- Archivierungsvertrag
- **Aufnahme ins Archiv** (Ingest)
- **Datenaufbereitung und –dokumentation**  
(Standard für alle, Added-value für ausgewählte Studien)
- **Langzeitarchivierung**
- **Datenservice (Access)**: Beratung, Datenzugang  
(Download, Online-Analyse, Secure Data Center)
- **Datenregistrierung (dalra)**: Vergabe von DOIs im  
DataCite Verbund



# Signifikante Eigenschaften

*„Da [...] im Laufe der Zeit nicht alle Eigenschaften erhalten werden können, muss eine Auswahl getroffen werden. Diese Teilmenge aller Eigenschaften sind die signifikanten Eigenschaften. [...] Die signifikanten Eigenschaften sollen durch den gesamten Archivierungsprozess hindurch erhalten werden. An ihnen kann bis zu einem bestimmten Grad die Authentizität der aus migrierten oder emulierten Repräsentationen hervorgegangenen Performances festgestellt werden.“*



Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung, Version 2.0



# Significant Properties

*„The characteristics of digital objects that must be preserved over time in order to ensure the continued **accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record.**“*

*“Properties are considered to exist in one of five categories: **content, context, appearance, structure and behaviour**”*

Grace, S., Knight, G., & Montague, L. (2009). InSPECT: Final Report. pp. 4-5.

# Inhalt und Kontext

## Konzept für die Definition Signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten

Natascha Schumann, Reiner Mauer, Stefan Müller, Astrid Recker  
 GESIS Datenarchiv für Sozialwissenschaften

Das GESIS Datenarchiv stellt quantitative Forschungsdaten, vorwiegend aus nationalen und internationalen vergleichenden Umfragen, für die Nachnutzung bereit. Forschungsdaten werden gemäß klar definierten Anforderungen ausgewählt, aufbereitet, dokumentiert, langfristig gesichert und zugänglich gemacht. Ein besonderes Augenmerk bei der Archivierung und Bereitstellung von Forschungsdaten aus den Sozialwissenschaften liegt auf der Einhaltung von datenschutzrechtlichen Bestimmungen



DEUTSCHE INITIATIVE FÜR NETZWERKINFORMATION E.V.

DINI / Veranstaltungen / Workshops / Langzeitarchivierung von Forschungsdaten / Programm

### Programm

Programm – 3. Workshop der DINI/nestor AG Forschungsdaten

Langzeitarchivierung von Forschungsdaten – 19.06.2015  
 Konrad-Zuse-Zentrum für Informationstechnik, Takustr. 7, 14195 Berlin  
 10:30 – 15:30 Uhr

09:30-10:30	Ankommen und Registrierung
10:30	Begrüßung und Einführung durch die DINI/nestor AG Forschungsdaten und Begrüßung durch das Konrad-Zuse-Zentrum für Informationstechnik (Janna Neumann (TIB), Tim Hasler (ZIB))
<b>10:45-12:00</b>	<b>Block 1 (Moderation: J. Neumann)</b>
10:45-11:10	Aspekte der Langzeitarchivierung in Policies von Forschungsdatenrepositorien (Jessika Rücknagel, HU Berlin)
11:10-11:35	Vertrauenswürdige digitales Langzeitarchiv – Zertifizierung an der TIB (Franziska Schwab, Technische Informationsbibliothek)
11:35-12:00	Konzept für die Definition signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten (Natascha Schumann, Reiner Mauer, Stefan Müller, Astrid Recker, GESIS – Leibniz Institut für Sozialwissenschaften)
12:00-13:00	Mittagspause
<b>13:00-13:50</b>	<b>Block 2 (Moderation: J. Ludwig)</b>
13:00-13:25	NaLiDa: Nachhaltigkeit linguistischer Daten (Claus Zinn, Erhard Hinrichs, Thorsten Trippel, Universität Tübingen)
13:25-13:50	Digitales Datenmanagement seit 1988 – Überführung eines digitalen Forschungsarchivs aus dem Bereich Systemökologie in das ETH Data Archive (Ana Sesartic, ETH Zürich)
13:50-14:10	Kaffeepause



### Übersicht

- Startseite
- Programm
- Anmeldung
- Anmeldeformular
- Teilnehmerliste
- Abstracts
- Veranstaltungsort
- Kontakt

# Erscheinung und Struktur

**DINI**  
DEUTSCHE INITIATIVE  
FÜR NETZWERKINFORMATION E.V.

DINI / Veranstaltungen / Workshops / Langzeitarchivierung von Forschungsdaten / Programm

**Programm**  
Programm - 3, Workshop der DINI/nestor AG Forschungsdaten

Langzeitarchivierung von Forschungsdaten - 19.06.2015  
Konrad-Zuse-Zentrum für Informationstechnik, Takust 7, 14195 Berlin  
10:30 - 16:30 Uhr

**Übersicht**

- Startseite
- Programm
- Anmeldung
- Anmeldeformular
- Teilnehmerliste
- Abstracts
- Veranstaltungsort
- Kontakt

09:30-10:30	Ankommen und Registrierung
10:30	Begrüßung und Einführung durch die DINI/nestor AG Forschungsdaten und Begrüßung durch das Konrad-Zuse-Zentrum für Informationstechnik (Janna Neumann (TZB), Tim Hasler (ZIB))
<b>10:45-12:00</b>	<b>Block 1 (Moderation: J. Neumann)</b>
10:45-11:30	Aspekte der Langzeitarchivierung in Repos von Forschungsdatenrepositorien (Jessika Röcknagel, HU Berlin)
11:10-11:35	Vertrauenswürdiges digitales Langzeitarchiv - Zertifizierung an der TZB (Franziska Schwab, Technische Informationsbibliothek)
11:35-12:00	Konzept für die Definition signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten (Natascha Schumann, Reiner Mauer, Stefan Müller, Astrid Recker, GESIS - Leibniz Institut für Sozialwissenschaften)
12:00-12:30	Mittagspause
<b>13:00-13:50</b>	<b>Block 2 (Moderation: J. Ludwig)</b>
13:00-13:25	NaLiDa: Nachhaltigkeit linguistischer Daten (Claus Zinn, Erhard Hinrichs, Thorsten Trippel, Universität Tübingen)
13:25-13:50	Digitales Datenmanagement seit 1988 - Überführung eines digitalen Forschungsarchivs aus dem Bereich Systemökologie in das ETH Data Archive (Ana Szarits, ETH Zürich)
13:50-14:10	Kaffeepause

**Folien** | **Gliederung**

**1**

gesis  
Leibniz-Institut für Sozialwissenschaften

Konzept für die Definition signifikanter Eigenschaften von sozialwissenschaftlichen Forschungsdaten

J. Neumann  
G. Ludwig

**2**

gesis

**Gliederung**

- GESIS-Blockworkshop
  - Aufgaben und Wertever
- Signifikante Eigenschaften
  - Beispiele
- Signifikante Eigenschaften für sozialwissenschaftliche Forschungsdaten

**3**

gesis

**Forschungsdaten in den Sozialwissenschaften**

- Sozialethik und die Rolle von Daten: Ethik und „Industrieunternehmen“
- Welche Anforderungen an Daten für die Programmierung
- Daten und die Ethik...
  - Anforderungen an Daten aus ethischen Gründen: die Rolle der Programmierung, soziale Ethik und andere Beispiele, Maximalismus in der Daten-Analyse (G. Ludwig ...)
  - Wie ist die Ethik in der Daten-Analyse zu verstehen?
  - Wie ist die Ethik in der Daten-Analyse zu verstehen? (G. Ludwig ...)

# Verhalten

The screenshot shows a Beamer presentation window. The main slide area displays the gesis logo in the top left, the title 'Verhalten' in the center, and footer text at the bottom right: 'DINI/nestor WS „Langzeitarchivierung von Forschungsdaten“, 19.06.2015 Berlin' and the page number '12'. A small logo for the Leibniz Association is visible in the bottom left of the slide area. The sidebar on the left shows a table of contents with slides 8 through 12. Slide 12, titled 'Verhalten', is highlighted with a yellow background.

# Signifikante Eigenschaften: Ansätze

- Bezogen auf Daten
  - ▶ Datenformate (Audio, Text, etc.), Bestimmung z.B. von Bildgröße und Auflösung
  - ▶ Sozialwissenschaften: hauptsächlich Statistikformate, tabellenbasierte Formate

*„Die Funktion der signifikanten Eigenschaften wird im Zuge des Erhaltungsprozesses deutlich. Sobald sich eine Performance aufgrund einer Migration oder eines Emulators ändert, müssen die signifikanten Eigenschaften der alten Performance mit jenen der neuen Performance verglichen werden. Beispielsweise sollte ein Audioobjekt, das im Dateiformat A 11 Minuten lang ist, auch im Dateiformat B noch 11 Minuten lang sein.“*

Leitfaden Digitale Bestandserhaltung, S. 22

# Signifikante Eigenschaften: Ansätze

- Bezogen auf Personen
  - ▶ Bestimmung der Designated Community
  - ▶ Bestimmung von Nutzungszielen
  - ▶ „Wie nutzt die Designated Community die Daten? Was tut sie mit den Daten?“
  - ▶ Gruppierung von Objekten mit derselben Zielgruppe und demselben Nutzungszweck (neben weiteren Gruppierungsmöglichkeiten, z.B. Objekttyp)
  - ▶ Sozialwissenschaften: In der Regel wollen Sozialwissenschaftler\*innen die Daten maschinell weiterverarbeiten





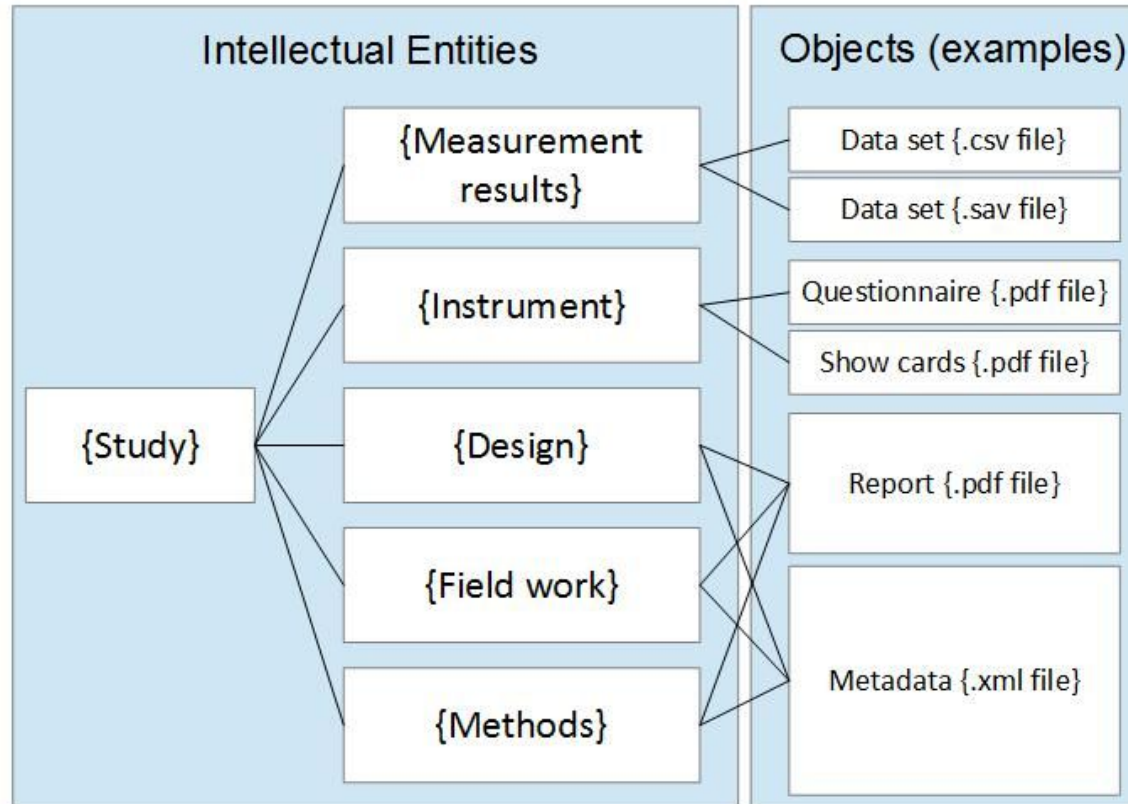
# Signifikante Eigenschaften: Ansätze

- Bezogen auf Prozesse
  - Ergebnisse einer Datenerhebung (Datensatz) allein sind nicht verständlich und nicht nachnutzbar
- Es braucht zum Verständnis auch Informationen über den Entstehungsprozess der Daten, z.B.:
  - ▶ Methodenbericht
  - ▶ Fragebogen
  - ▶ Codebuch



# Was bedeutet das?

- Daten (Studie) als „intellectual entity“, die auch prozessbezogene Informationen erhalten, um eine sinnvolle Nachnutzung möglich zu machen
- Bestimmung der verschiedenen Elemente der Erhebung und ihres Kontextes und ihrer Repräsentation



# Unser Vorgehen

- Eingehendere Beschäftigung mit vorhandenen Konzepten und Projektergebnissen
- Beschreibung AIPs mit zusätzlichen Metadaten
  - ▶ Im europäischen Kontext von CESSDA
  - ▶ Blick auf PREMIS 3.0
- Experteninterviews (Archivmitarbeiter\*innen, Sozialwissenschaftler\*innen)

Wir bleiben dran 😊



Kontakt:  
[natascha.schumann@gesis.org](mailto:natascha.schumann@gesis.org)