

# Dienstleistung für eine internationale Community

-

## Das Datenarchiv für Sozialwissenschaften der GESIS

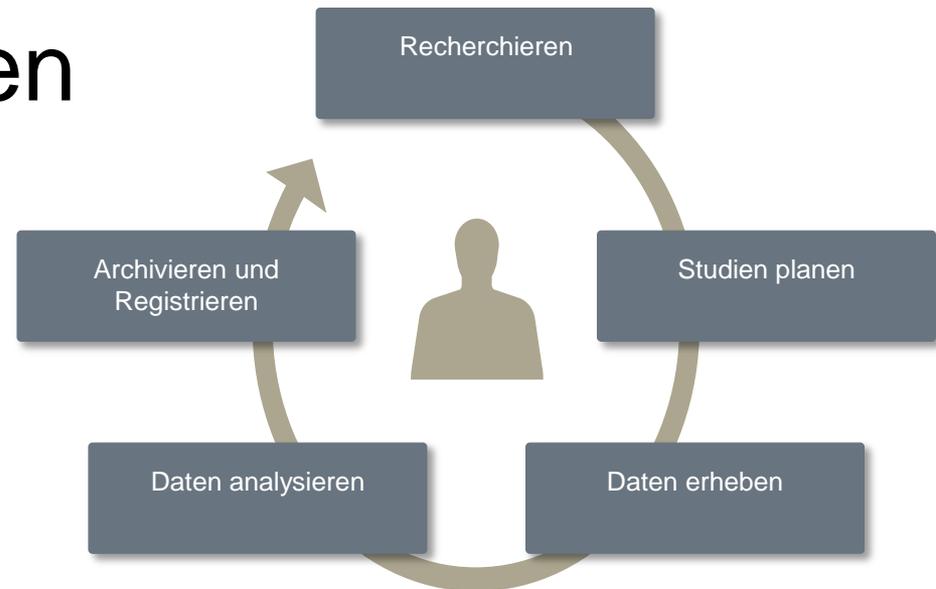
Uwe Jensen, GESIS – Leibniz-Institut für Sozialwissenschaften

*23.04 - Forschungsinfrastruktur für Geistes- und Sozialwissenschaften*

*101. Deutscher Bibliothekartag 2012, Hamburg*

## GESIS – Leibniz-Institut für Sozialwissenschaften

- Infrastruktureinrichtung für die Sozialwissenschaften
- 250+ Mitarbeiter/innen an 3 Standorten Mannheim, Köln, Berlin (seit 11.2011)
- „Forschungsdatenzzyklus“ dient als Leitbild zur Strukturierung und Verknüpfung der GESIS-Angebote



# Datenarchiv für Sozialwissenschaften (DAS)

- 1960: erstes sozialwissenschaftliches Datenarchiv in Europa (Zentralarchiv für Empirische Sozialforschung, Universität zu Köln)
  - 1986: Zentralarchiv wird Mitglied der neu gegründeten GESIS
  - 2008: Überführung der drei Teil-Institute der GESIS in ein Institut; seither ist das Archiv eine wissenschaftliche Abteilung der GESIS
- **Leitgedanke:** Langfristige Nutzung sozialwiss. Forschungsdaten ermöglichen („neue Fragen an alte Daten“) > Viele Motive:
- Forschungsökonomie** (Analysepotentiale für neue Forschung nutzen, Doppelung vermeiden);
  - komparative Forschung braucht** Zeitreihen (Analyse v. sozialem Wandel);
  - Gute wissenschaftliche Praxis** (Nachvollziehbarkeit von Forschungsergebnissen > Transparenz & Überprüfbarkeit sichern); **Daten für Methodenforschung, Lehre, neue Projekte ...**

# DAS Aufgaben

- Langzeitarchivierung von Forschungsdaten:  
Langzeitverfügbarkeit und Interpretierbarkeit sichern
- Sekundäranalysen, Replikationen, Raum- und  
Zeitvergleiche ermöglichen
- Zugang zu internationalen Forschungsdaten gewährleisten
- Dienstleister für Primärforscher bei der nachhaltigen  
Sicherung, Dokumentation, Zitation und Bereitstellung  
ihrer Daten

# DAS Datenbestand

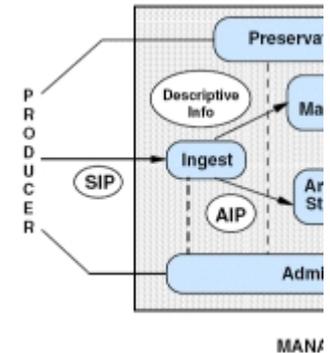
- > 6.100 Datensätze – Schwerpunkte:
  - Soziologie & Politikwissenschaft
  - Mikrodaten der Umfrageforschung, historische Zeitreihen / Aggregatdaten
- wichtige Datenschwerpunkte (Auszug)
  - interkulturell vergleichende Sozialforschung (ISSP, EB, EVS, CSES)
  - nationale Erhebungen (ALLBUS, GLES, Politbarometer, DeutschlandTrend)
- und umfangreiche Bestände von Studien zu
  - politischen Einstellungen u. Verhalten, Werten, Jugend, Gesundheit, Mediennutzung ...

# DAS Dienstleistungen

1. Akquisition & Archivierung
2. Datenaufbereitung u. -dokumentation
3. Datenzugang: Beratung & Datenbereitstellung
4. Langzeitarchivierung
5. Internat. Datenaustausch + F&E in Archivnetzwerken
6. Datenregistrierung (da|ra) gemeinsam mit ZBW
  - Infrastruktur für die dauerhafte Identifizierung, Sicherung, Lokalisierung und verlässliche Zitierbarkeit von Forschungsdaten dieser Disziplinen

# D1: Akquisition & Archivierung

- Vorbereitung der Archivierung mit Datenproduzenten
  - Ziel ist es möglichst früh im Lebenszyklus anzusetzen
  - **Akquisition & Beratung zum Datenmanagement:**  
Klärung: Umfang, Formate, Aufbereitungs- u. Dokumentationsziele, Nutzungsrechte
  - **Vertragliche Regelung der Archivierung** (Nutzungsrechte, Zugang ...)
- Aufnahme der Studie ins Archiv – Standardregeln & -prozesse
  - **Studienbeschreibung:** Administration (ID, PID, Versionierung) & Erschließung von Daten & Dokumentationen > Beschreibung: Inhalt, Methode, Datenstruktur, Literatur
  - **Technische und formale Kontrollen:** Formate, Lesbarkeit, Virenfreiheit etc. / Vollständigkeit, Nutzbarkeit, Konsistenz von Daten u. Dokumentation
  - **Sicherung** aller Originaldateien und Materialien
  - **Kontrollen zum Datenschutz** > ggf. korrigierende Aufbereitungen



## D2 Datenaufbereitung / -dokumentation

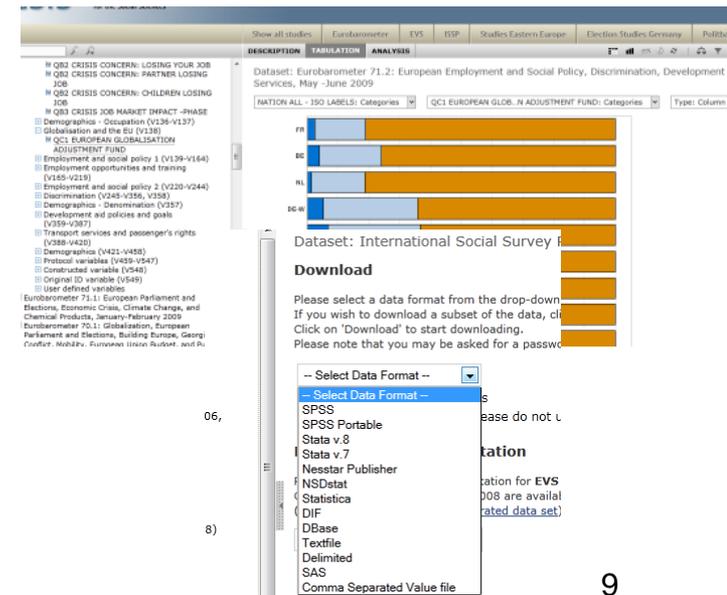
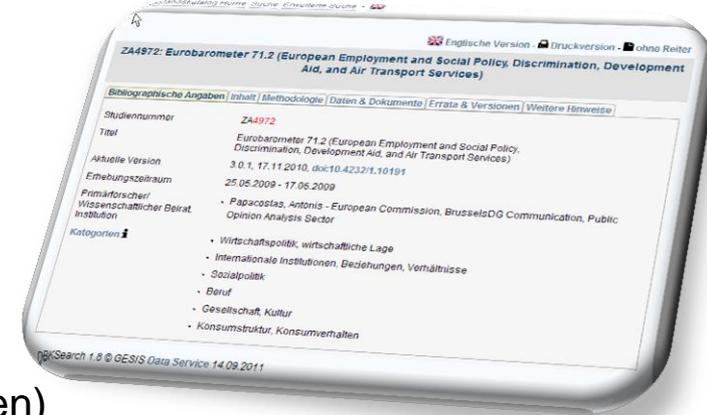
- Service in der Standardarchivierung:  
einfache Aufbereitung (Korrektur von Daten u. Metadaten) u. Bereitstellung
- Mehrwertdienste für ausgewählte komparative Studien
  - **Datenaufbereitung:**
    - Standardisierung von Datenstrukturen, Harmonisierung von Variablen,
    - Integration / Kumulation von Einzeldatensätzen (zeit- und / oder ländervergleichend),
    - Ergänzung mit Kontextdaten / Aggregatdaten
  - **Dokumentation:** Umfassende Produktion strukturierter Metadaten  
z.B. vollständige u. multilinguale Frage- und Antworttexte, Intervieweranweisung, Anmerkungen zur Datenqualität auf Variablenebene;
  - **Produkte:** Aufbereitete Daten + Dokumentationen  
Variablenreports (integrierte Dokumentation: Fragen, Daten, Datenaufbereitung), Methodenberichte - Originalfragebögen
  - **Service u. Forschung:** i.d.R. durch GESIS FDZs



FDZ ALLBUS
FDZ Amtliche Mikrodaten
FDZ Internationale Umfrageprogramme
FDZ Wahlen

## D3: Datenbereitstellung

- Datenzugang über Online-Portale und (individuelle) Bereitstellung auf Datenträgern oder per ftp
- (Meta-)Datenportale: **Datenbestandskatalog (DBK)** (Studienschließung & Download (Daten, Dokumentationen))  
**ZACAT, HISTAT** (Variablenrecherche, Online-Analyse, Download), **sowiport** (Literatur & Daten)
- DAS vermittelt Zugang zu Datenbeständen ausländischer Archive (ICPSR, CESSDA, IFDO)
- 30.000+ Datenweitergaben
- Weit überwiegend akademische Nutzung: Lehre und Forschung
- Je nach Angebot 30%-70% internationale Nutzer



# D4: Langzeitarchivierung



- **Archivspeicher:**

- Archivierung im engeren Sinne erfolgt durch die Überführung der zur Studie gehörenden Objekte (Originale + Archivprodukte) in zentralen Archivspeicher (STAR = Studienarchiv)
- Organisation: dateibasierte Verzeichnisstruktur (Regeln); Restriktiv. Zugangsrechte

- **Substanzerhaltung:**

- physischer Schutz vor Ort - räumlich getrennte und redundante Datenhaltung
- Diversität eingesetzter Speichertechnik - regelmäßige Medienmigration / Refreshment

- **Erhalt von Nutzbarkeit & Interpretierbarkeit:**

- hauptsächlich Migrationsstrategien (zur Überbrückung auch Emulation bzw. Virtualisierung)
- Bedrohung der digitalen Bestände durch technischen Fortschritt ist Konstante in der 50jährigen Archivarbeit

## 5. Datenaustausch & F&E in Archivnetzwerken

### Entwicklung informeller Netzwerke & Kooperationen

- Seit 1960iger Jahre Aufbau nationaler Datenarchive (ZA, ICPSR, ...)
- Forschungsorientierung: International vergleichende emp. SF und Sekundäranalyse von Daten (z.B. Ungleichheitsstrukturen, Werte, Sozialer Wandel ...)
- > Internationale Archivstrukturen: **CESSDA** (1976) > **IFDO** (1977)

### Ziele & Anforderungen transnationaler Archivkooperationen:

- Datenaustausch, Langzeitarchivierung, Datenbereitstellung
- Metadatenstandards: Erschließung<>Dokumentation<>Nutzung  
z. B. Study Description Standard (1986) > DDI 2 & DDI 3
- Aufbau kompatibler Regeln, Prozesse & Technologien  
> Betrieb gemeinsamer Dateninfrastrukturen

## CESSDA Council of European Social Science Data Archives

### 21 nationale Datenserviceeinrichtungen

- Service: CESSDA Datenportal & ELLST
- Entwicklung: CESSDA Expertenseminare
- Konferenzen & WS: z.B. mit DDI; IASSIST

### Entwicklung europ. Dateninfrastruktur:

- Themen: RDM>Data sharing>Infrastruktur
- Projekte: Forschung & Entwicklung (F&E)
  - > Vorbereitung CESSDA-ERIC
  - > Kooperation mit Stat. Ämtern
  - > Entwicklung SSH Dateninfrastrukturen



- Nutzer: > 30,000 Forschung & Lehre
- Zugang:> 25,000 Kollektionen
- Bereitstellung: > 70,000 Studien
- Akquisition: > 1,000 jährlich

## CESSDA Council of European Social Science Data Archives

Projekte: Entwicklung der europäischen Dateninfrastruktur (EU FP7)



### Data without Boundaries (DwB 2011-2015):

- **Ziel:** Verbesserung des europaweiten Zugangs zu Daten der amtlichen Statistik
- 28 Partner: CESSDA Datenarchive, Statistische Ämter und Universitäten
- **GESIS Beteiligung:** Datenquellen (WP5), Training (WP6), Standards (WP7), Datenzugang (WP9)

### Data Service Infrastructure for the Social Sciences & Humanities (DASISH 2012-2014)

- **Ziel:** Entwicklung von Synergien beim Aufbau von Infrastrukturen für den Datenservice
- Partner: CESSDA, CLARIN (Language Resources and Technology), DARIAH (Arts and Humanities), ESS (European Social Survey), SHARE (Survey of Health, Ageing and Retirement in Europe)
- **GESIS DAS Beteiligung:** Datenarchivierung, Datenzugang (PID, Metadaten), Datenschutz, Training

### CESSDA ERIC - European Research Infrastructure Consortium

- 2006: Aufnahme von CESSDA auf die ESFRI-Roadmap - 2008-2010: CESSDA-PPP (Vorbereitungsphase)
- Ab 2011: Etablierungsphase der neuen Organisationsform ERIC als Körperschaft europäischen Rechts
- Ziele: - Dauerhafte Koordinationsfunktionen für die sozialwissenschaftliche Datenlandschaft in Europa
- - Betrieb einer technischen Infrastruktur

## Akronyme und Verweise

### Nationaler und internationaler Datenservice

- [GESIS](#) > DAS Services: [Recherchieren](#) / [Daten analysieren](#) (Kollektionen) / [Archivieren & registrieren](#)
- [CESSDA](#) - Council of European Social Science Data Archives > [Datenportal](#)
- [IFDO](#) - International Federation of Data Organizations
- [ICPSR](#) - Interuniversity Consortium for Political and Social Research (USA)

### Europäische Dateninfrastrukturprojekte

- [CESSDA – PPP](#) (Preparatory Phase Project):
- [Data without Boundaries](#) (DwB):
- [Data Service Infrastructure for the Social Sciences & Humanities](#) (DASISH)
  - [CLARIN](#) (Common Language Resources and Technology Infrastructure),
  - [DARIAH](#) (Digital Research Infrastructure for the Arts and Humanities),
  - [ESS](#) (European Social Survey)
  - [SHARE](#) (Survey of Health, Ageing and Retirement in Europe)

### Weitere Quellen

- [DDI](#) - Data Documentation Initiative
- [IASSIST](#) - International Association of Social Science Information Service & Technology

### Literatur

- [Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme](#)
- [Handbuch Forschungsdatenmanagement](#)

Vielen Dank für Ihr Interesse  
und vielleicht auf ein Wiedersehen in  
Köln . . . . [Uwe.Jensen@gesis.org](mailto:Uwe.Jensen@gesis.org)

**IASSIST 39th Annual Conference - 2013**

**Köln, 28 – 31 Mai 2013**

**GESIS - Leibniz Institut für Sozialwissenschaften**

- IASSIST is an international organization of professionals working with information technology and data services to support research and teaching in the social sciences.
- Its 300 members are from a variety of workplaces, including data archives, statistical agencies, research centers, libraries, academic departments, government departments, and non-profit organizations.