

Forschungsdatenmanagement  
am  
Institut für interdisziplinäre Konflikt- und Gewaltforschung (IKG)

Kurt Salentin

Vortrag am 5.11.2010

im Kolloquium Wissensinfrastruktur  
der Universität Bielefeld

## Zweck des Vortrags

- Das Institut
- Welche Daten?
- Technische Aspekte
- Strukturelle und organisatorische Fragen der Datenproduktion, -dokumentation, -weitergabe
- Interessen der Akteursgruppen

## Was macht das Institut für interdisziplinäre Konflikt- und Gewaltforschung (IKG)?

Es untersucht

- Rechtsextremismus (bevölkerungsweite Einstellungen, extremistische Gruppen)
- ethnisch-kulturelle Konflikte (religiöse Konflikte, Minderheitenkonflikte, Diskriminierung, ablehnende Einstellungen)
- Gesellschaftliche Integration
- Gewalt (Terrorismus, Krieg, interpersonal)

## Warum ist das IKG interdisziplinär?

- Sozialwissenschaften
- Erziehungswissenschaft
- Psychologie
- Sozialpsychologie
- Geschichtswissenschaft
- Rechtswissenschaft
- Religionswissenschaft
- Politikwissenschaft

## Welche Projekte betreibt das IKG? (Auswahl)

- Kontrolle der Gewalt
- Gruppenkonflikte in kleinen Stadtgesellschaften mit hohem Aussiedleranteil
- Öffentliche Gewalt im Stadtquartier
- Konfliktregulierung bei Straftaten
- Entsolidarisierung und ihre sozialen und politischen Folgen
- Stadtgesellschaften im Umgang mit Rechtsextremismus und Fremdenfeindlichkeit
- Konflikte um religiöse Symbole: Moscheebau und Muezzinruf in deutschen Städten
- Die Reaktion der Polizei auf ethnisch-kulturelle Konflikte
- Gruppenbezogene Menschenfeindlichkeit
- Zuwanderung, Figuration, Konflikt
- Sexuelle Belästigung und Soziosexualität
- Cyberbullying von Schülerinnen und Schülern

## Welche Rolle haben Daten in der Konflikt- und Gewaltforschung?

- Im IKG überwiegt erfahrungswissenschaftliche (empirische) Forschung.
- Die Methoden der Konflikt- und Gewaltforschung ähneln zusehends denen der Naturwissenschaften, unterscheiden sich aber auch.
- Erkenntnisse werden durch kontrollierte, nachvollziehbare Verfahren gewonnen.
- Daten dienen als einzig legitime Repräsentation der Welt (vorherrschende Position, kein Konsens).
- Daten sind zentraler Bestandteil wissenschaftlicher Verfahren in der Konflikt- und Gewaltforschung.
- Erweiterter Datenbegriff: Nicht nur Zahlen, Bits und Bytes, auch qualitative Daten

## Was bedeutet Methodenpluralismus?

### Auswahl

- qualitative Methoden
  - Interviews
  - teilnehmende Beobachtung
  - Medienanalyse
  - Dokumentenanalyse
- quantitative Methoden
  - Surveys (regional, bundesweit, international)
  - experimentelle Verfahren
  - Sekundäranalysen

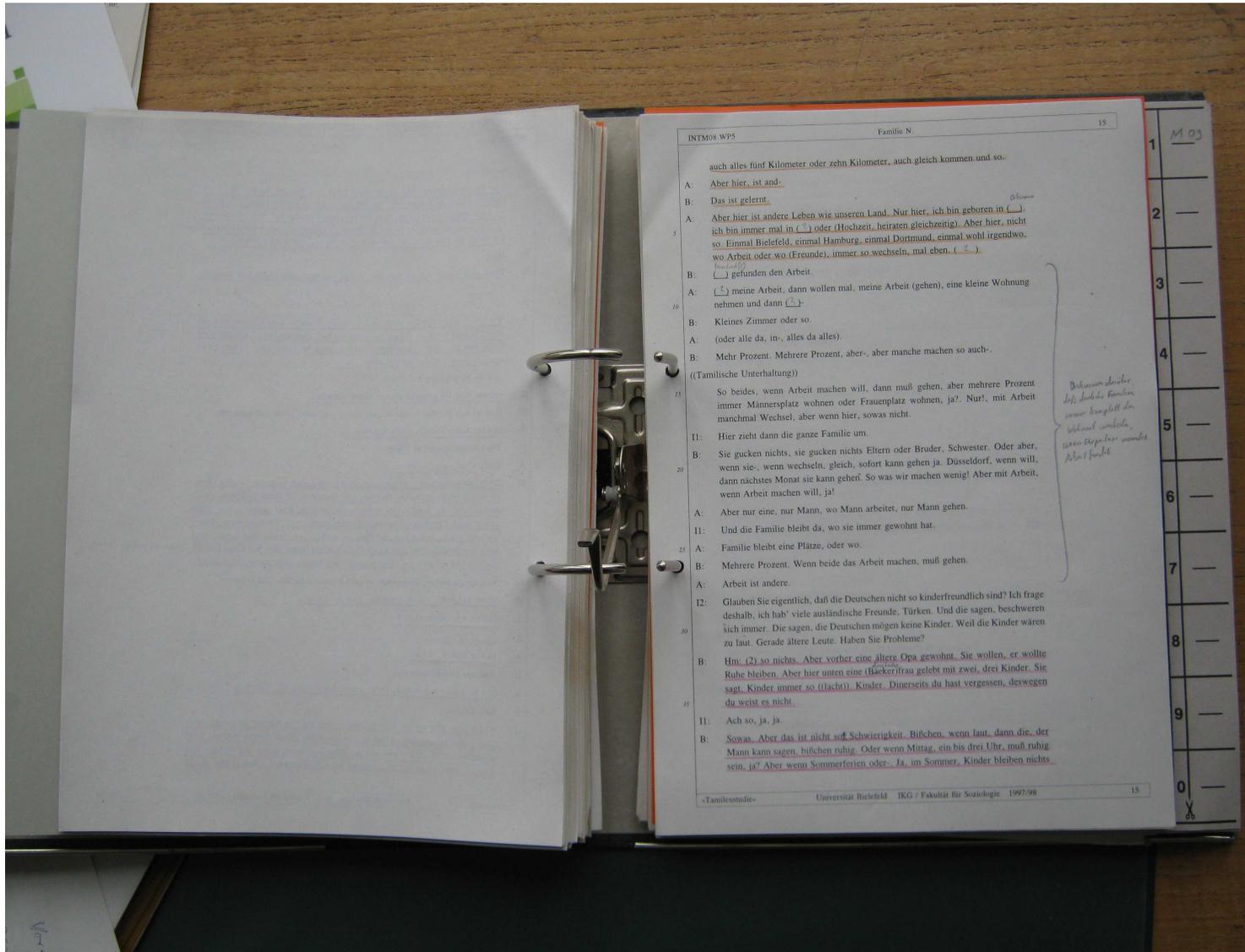
## Wie geschieht die Datengewinnung? Beispiele



# Qualitative Daten: ein herkömmliches Interviewtranskript

INTM08.WP5	Familie N.	15
	auch alles fünf Kilometer oder zehn Kilometer, auch gleich kommen und so.	
A:	<u>Aber hier, ist and-</u>	
B:	<u>Das ist gelernt.</u>	
A:	<u>Aber hier ist andere Leben wie unseren Land. Nur hier, ich bin geboren in ( ),</u> <u>ich bin immer mal in ( ) oder (Hochzeit, heiraten gleichzeitig). Aber hier, nicht</u> <u>so. Einmal Bielefeld, einmal Hamburg, einmal Dortmund, einmal wohl irgendwo,</u> <u>wo Arbeit oder wo (Freunde), immer so wechseln, mal eben. ( )</u>	<i>als immer</i>
B:	<u>( ) gefunden den Arbeit.</u>	
A:	<u>( ) meine Arbeit, dann wollen mal, meine Arbeit (gehen), eine kleine Wohnung</u> <u>nehmen und dann ( )-</u>	
B:	<u>Kleines Zimmer oder so.</u>	
A:	<u>(oder alle da, in-, alles da alles).</u>	
B:	<u>Mehr Prozent. Mehrere Prozent, aber-, aber manche machen so auch-</u>	
	<u>((Tamilische Unterhaltung))</u>	
	<u>So beides, wenn Arbeit machen will, dann muß gehen, aber mehrere Prozent</u> <u>immer Männersplatz wohnen oder Frauenplatz wohnen, ja?. Nur!, mit Arbeit</u> <u>manchmal Wechsel, aber wenn hier, sowas nicht.</u>	
I1:	<u>Hier zieht dann die ganze Familie um.</u>	
B:	<u>Sie gucken nichts, sie gucken nichts Eltern oder Bruder, Schwester. Oder aber,</u> <u>wenn sie-, wenn wechseln, gleich, sofort kann gehen ja. Düsseldorf, wenn will,</u> <u>dann nächstes Monat sie kann gehen. So was wir machen wenig! Aber mit Arbeit,</u> <u>wenn Arbeit machen will, ja!</u>	<i>Diskussion darüber, daß deutsche Familien wenn umgezogen Wohnort wechseln, wenn Ehepartner wandert Aber nicht</i>
A:	<u>Aber nur eine, nur Mann, wo Mann arbeitet, nur Mann gehen.</u>	
I1:	<u>Und die Familie bleibt da, wo sie immer gewohnt hat.</u>	
A:	<u>Familie bleibt eine Plätze, oder wo.</u>	
B:	<u>Mehrere Prozent. Wenn beide das Arbeit machen, muß gehen.</u>	
A:	<u>Arbeit ist andere.</u>	
I2:	<u>Glauben Sie eigentlich, daß die Deutschen nicht so kinderfreundlich sind? Ich frage</u> <u>deshalb, ich hab' viele ausländische Freunde, Türken. Und die sagen, beschweren</u> <u>sich immer. Die sagen, die Deutschen mögen keine Kinder. Weil die Kinder wären</u> <u>zu laut. Gerade ältere Leute. Haben Sie Probleme?</u>	
B:	<u>Hm. ( ) so nichts. Aber vorher eine ältere Opa gewohnt. Sie wollen, er wollte</u> <u>Ruhe bleiben. Aber hier unten eine (Bäcker)frau gelebt mit zwei, drei Kinder. Sie</u> <u>sagt, Kinder immer so ((lacht)). Kinder. Dinerseits du hast vergessen, deswegen</u> <u>du weist es nicht.</u>	
I1:	<u>Ach so, ja, ja.</u>	
B:	<u>Sowas. Aber das ist nicht so Schwierigkeit. Bißchen, wenn laut, dann die, der</u> <u>Mann kann sagen, bißchen ruhig. Oder wenn Mittag, ein bis drei Uhr, muß ruhig</u> <u>sein, ja? Aber wenn Sommerferien oder-. Ja, im Sommer, Kinder bleiben nichts</u>	
»Tamilienstudie«	Universität Bielefeld IKG / Fakultät für Soziologie 1997/98	15

# Ein File früher



# Ein File heute: Qualitative Daten in MAXQDA

The screenshot displays the MAXQDA software interface with the following components:

- Project Information:** P:\Unterlagen\Inh. analyse\alte Versionen von MAXQDA\2.6.9 Muslima in Deutschland.m2k - MAXQDA
- Project Menu:** Projekt, Bearbeiten, Codes, Memos, Variablen, Textsuche, Analyse, Fenster, ?
- Liste der Texte (List of Texts):** A tree view showing the file structure. The selected file is "Ehrenmord" (208 lines).
- Text Editor:** Displays the text "Vogelfreie Frauen" with line numbers 1-5. The text discusses the murder of Hatun Sürücü and its political implications. The word "Ehrenmord" is highlighted in blue.
- Liste der Codes (List of Codes):** A tree view showing the code system. The selected code is "Ehrenmord" (601 lines).
- Liste der Codings (List of Codings):** A tree view showing the codings for the selected code. The selected coding is "Ehrenmord" (100 lines).
- Bottom Panel:** Shows the current coding assignment for the selected text segment, including the code "Ehrenmord" and its sub-coding "Vogelfreie Frauen".

## Ein standardisierter Fragebogen (Migrantenbefragung)

77 உங்கள் வீட்டு வருமானத்தை சொல்லக்கூடியதாகவிருந்தால் அது எங்களுக்கு உதவியாகவிருக்கும். கருதப்படும் வருமானமானது ஒரு மாதத்தில் உங்கள் வீட்டிலுள்ள எல்லாவுறுப்பினர்களதும் வேலைவருமானம் பக்கவேலைவருமானம் வாடகை வருமானம் அதுபோல் ஓய்வு ஊதியம் பிள்ளைகளுக்கானபணம் மற்றும் அரசாங்கத்திடமிருந்து கிடைக்கும் பணங்கள் எல்லா வற்றினதும் நெற்றோ தொகை அதாவது வரிகளையும் சமூகக் காப்புறுதிகளையும் கழித்தபின் கிடைக்கும் தொகை உங்கள் வீட்டிற்குப் பொருத்தமான வருமானத்திற்கு புள்ளடி போடவும்.

400 மார்க்குக்கு குறைவு  Unter 400 DM  
 800-1200 DM  800-1200 DM  
 1600-2000 DM  1600-2000 DM  
 2500-3000 DM  2500-3000 DM  
 3500-4000 DM  3500-4000 DM  
 4500-5000 DM  4500-5000 DM  
 6000-8000 DM  6000-8000 DM

77 Es würde uns helfen, wenn Sie die Einkommensgruppe nennen könnten, zu der Ihr Haushalt gehört. Gemeint ist das Einkommen in einem Monat, für alle Haushaltsmitglieder zusammengerechnet, die Summe aus Arbeitseinkommen, Nebentätigkeiten, Vermietung etc. sowie Rente, Kindergeld und sonstige Zahlungen des Staates, netto, d.h. nach Abzug von Steuern und Sozialbeiträgen.

Bitte kreuzen Sie die Kategorie an, zu der Ihr Haushalt gehört.

400-800 DM  400-800 DM  
 1200-1600 DM  1200-1600 DM  
 2000-2500 DM  2000-2500 DM  
 3000-3500 DM  3000-3500 DM  
 4000-4500 DM  4000-4500 DM  
 5000-6000 DM  5000-6000 DM  
 8000 மார்க்குக்கு கூட  über 8000 DM

78 உங்களது தற்போதைய சொந்தப் பொருளாதார நிலையை நீங்கள் எப்படிக் கணிக்கிறீர்கள்?

78 Wie beurteilen Sie Ihre eigene wirtschaftliche Lage heute?

மிக நன்று  
Sehr gut

நன்று  
Gut

பாரதி நல்லது  
பாரதி கூடாது  
Teils gut,  
teils schlecht

கூடாது  
Schlecht

மிகக் கூடாது  
Sehr schlecht

# Quantitative Daten in SPSS:

Mega 02 bis 10.sav [DatenSet1] - SPSS Statistics Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Diagramme Extras Add-Ons Fenster Hilfe

1 : qcp\_ser 9,0 Sichtbar: 1732 von 1732 Variablen

	qcp_ser	datum	anzfzth	anzfztm	bula	regb	bik	bikneu	nrpers
1	9	20603	20	33	9	1	0	0	1
2	12	20521	18	28	12	0	7	6	1
3	15	20521	19	18	5	3	0	0	2
4	18	20521	18	29	14	2	0	0	2
5	22	20521	18	30	15	2	7	6	1
6	25	20612	17	1	3	3	8	3	2
7	27	20521	18	32	7	3	1	3	4
8	28	20606	17	2	10	0	6	3	3
9	35	20618	19	22	14	3	7	1	2
10	37	20521	18	35	14	1	3	6	2
11	39	20521	19	27	6	4	0	0	2
12	44	20521	18	36	16	0	6	6	2
13	47	20521	18	38	18	1	0	0	3
14	48	20624	20	5	9	4	7	3	3
15	51	20524	18	30	14	2	7	6	2
16	63	20528	19	37	5	1	0	0	1
17	65	20521	18	45	15	1	6	6	3
18	69	20604	18	32	6	5	3	7	2
19	83	20611	17	36	7	1	9	7	2
20	86	20607	17	36	8	1	6	6	2
21	102	20624	17	9	15	1	2	3	2
22	108	20521	19	5	5	3	0	0	2

Datenansicht Variablenansicht

SPSS Statistics Prozessor ist bereit

Ein typisches Produkt des IKG

---

Deutsche Zustände

---

Folge 9

---

Herausgegeben von

---

Wilhelm Heitmeyer

---

edition suhrkamp

---

SV

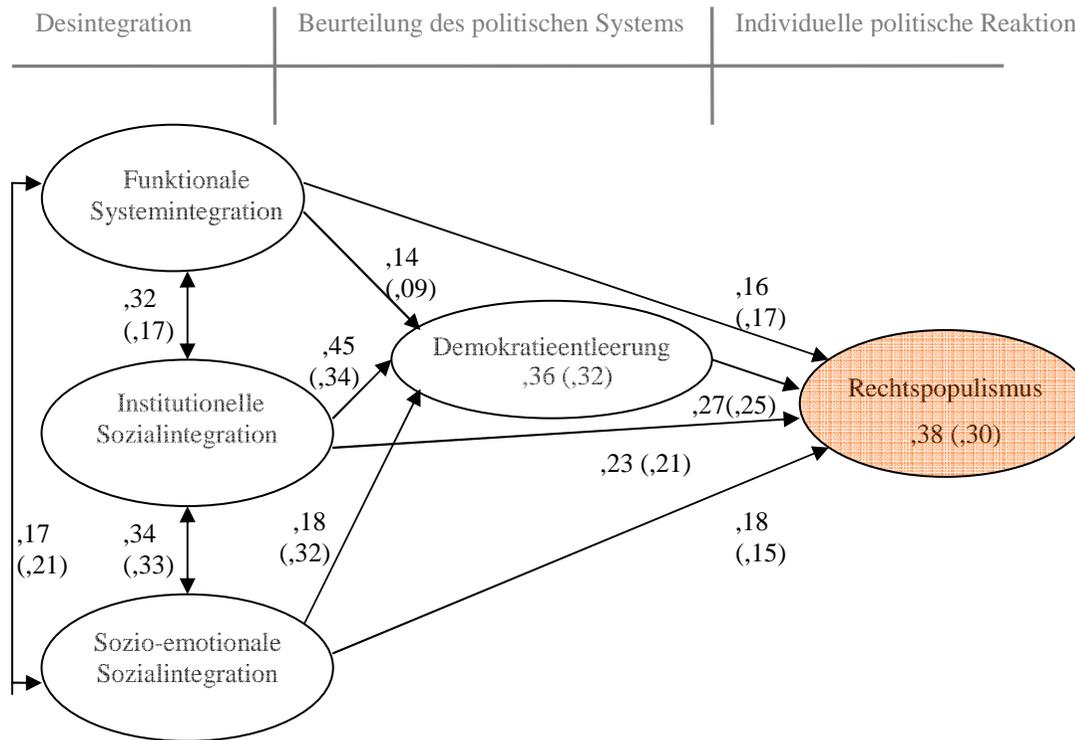
## Beispielhafte Ergebnisse

Positionen zur Vereinigung im GMF Survey 2008, Angaben in Prozent

	West	Ost
	trifft eher/voll und ganz zu	
Ost- und Westdeutschland sind zusammengewachsen.**	59,8	41,3
	sehr/eher gut	
Wie gut oder schlecht finden Sie rückblickend den politischen, ökonomischen und sozialen Zustand Deutschlands vor der Wende?	84,6	41,8

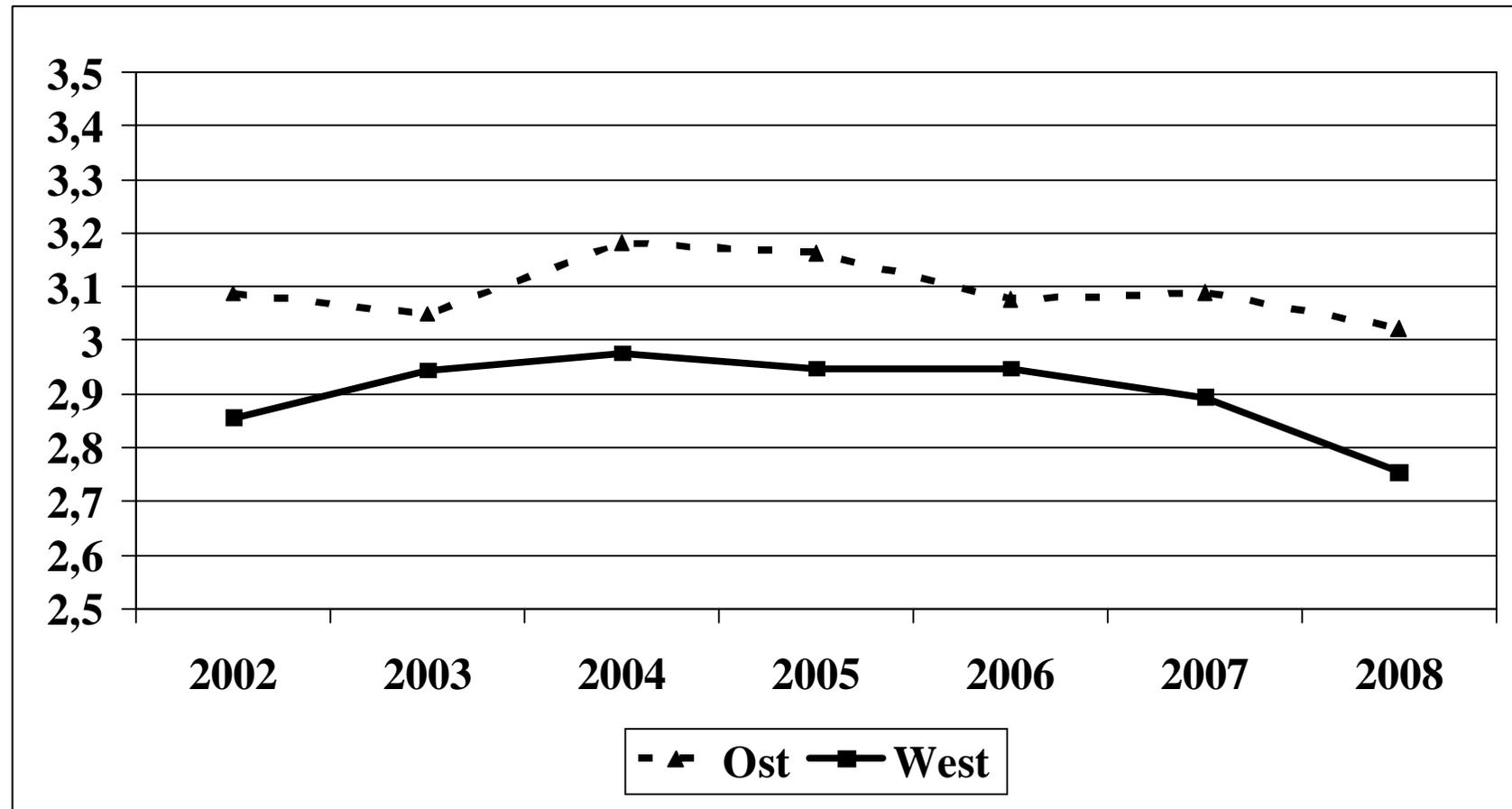
Quelle: Wilhelm Heitmeyer: Leben wir immer noch in zwei Gesellschaften? Deutsche Zustände 2008.

# Strukturgleichungsmodell



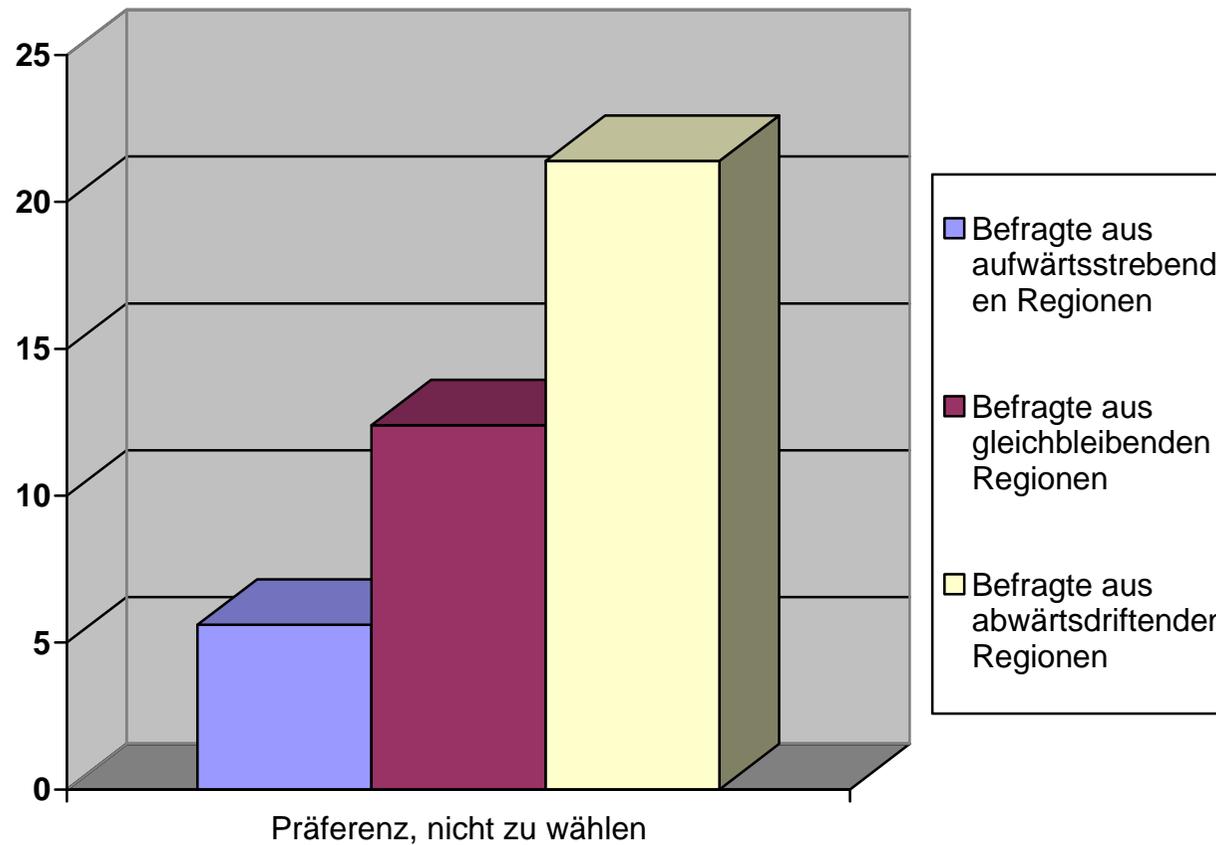
Quelle: Wilhelm Heitmeyer: Leben wir immer noch in zwei Gesellschaften? Deutsche Zustände 2008

## Rechtspopulismus in Ost und Westdeutschland 2002-2008



Quelle: Anna Klein / Beate Küpper / Andreas Zick: Erniedrigte und Feindselige - Rechtspopulismus im vereinigten Deutschland als Ergebnis von gefühlter Benachteiligung und Demokratiekritik. Deutsche Zustände 2008.

## Die Wahlintention in drei Regionentypen, GMF-Survey 2005



Quelle: Reinecke/Hüpping: in Deutsche Zustände 2008.

## Wer sind die Datenproduzenten im IKG?

26 Wissenschaftler (teilzeit/vollzeit)

davon

2 Professoren (Lehrstühle in der Fak. für Erziehungswissenschaft)

2 unbefristete Mitarbeiter

22 befristete Mitarbeiter (Drittmittelstellen, DFG, VolkswagenStiftung etc.)

9 Graduierte im DFG-Graduiertenkolleg 884: "Gruppenbezogene Menschenfeindlichkeit"

sowie

19 studentische Hilfskräfte

## Welche Infrastruktur hat das IKG? Verwaltung und Dienstleistungen

- 1 Wiss. Mitarbeiter mit Geschäftsführungsaufgaben (Projektentwicklung, Finanzplanung, universitäre Gremien, FoDa, Raumverwaltung etc.)
- 1 Wiss. Mitarbeiter auf Stabsstelle (Projektentwicklung, Research Consultant, interne Begutachtung, Praktikumsbeauftragter, Zeitschriftenredakteur, IT-Beauftragter, Datensicherheitsbeauftragter, Gelegenheitsprogrammierer, "Faktotum")
- 1 Sachbearbeiterin (Personal, Drittmittelverwaltung)
- 1/2 Sekretariatsstelle
- 2 SHK à 10 Wst. für EDV
- 0 Programmierer
- 0 Datenbankadministratoren

## Wie ist die Forschung im IKG organisiert?

- Projekte sind klein, typisch: 1,5 Stellen, 2-3 Jahre Laufzeit
- Projekte sind autonom
  - Drittmittelfinanzierung unabhängig vom Institut
  - starke Rolle des Projektleiters als Mittelempfänger
  - spezialisierte Kompetenz, disziplinärer Zuschnitt
  - Heterogene thematische und methodische Zugänge
- Keine Weisungsbindung in inhaltlichen und technischen Fragen
- IKG ist Container für Forschungsprojekte
- Geringe Arbeitsteilung in Erhebung, Analyse, Datenmanagement
- Begrenzte Synergien in Theorie, Methoden und Datenmanagement:  
Jedes Projekt ist eine Maßanfertigung.

## Mitarbeiter

- sind thematisch und methodisch qualifiziert (wissenschaftliche Verfahren, z. B: Statistik)
- stehen unter Qualifikations- und Publikationsstress
- besitzen Defizite in Softwarebeherrschung und teils in Analysetechniken
- bewerten Zeit- und Humankapitalinvestitionen vor dem Hintergrund kurzer Produktionszyklen und Amortisationsperspektiven
- besitzen geringe Technikaffinität

Diese Strukturen sind typisch für Geistes- und Sozialwissenschaften.

## Zwischenbilanz

- emsige Datenproduktion
- heterogenes, teils schwer standardisierbares Datenmaterial
- schwache Infrastruktur
- geringe Arbeitsteilung und Spezialisierung im Daten-Lebenszyklus

# Datenbestand des IKG:

## Statistik der Dateitypen im Active Directory

(nur Gruppenlaufwerke, 2.11.2010)

	Dateien		Speicher	
	N	%	MB	%
<b>Texte:</b>				
doc	22.951	32.05	4.817	5.38
pdf	7.853	10.97	12.245	13.69
ppt	874	1.22	1.670	1.87
txt	1.997	2.79	3.502	3.91
xls	1.775	2.48	449	0.50

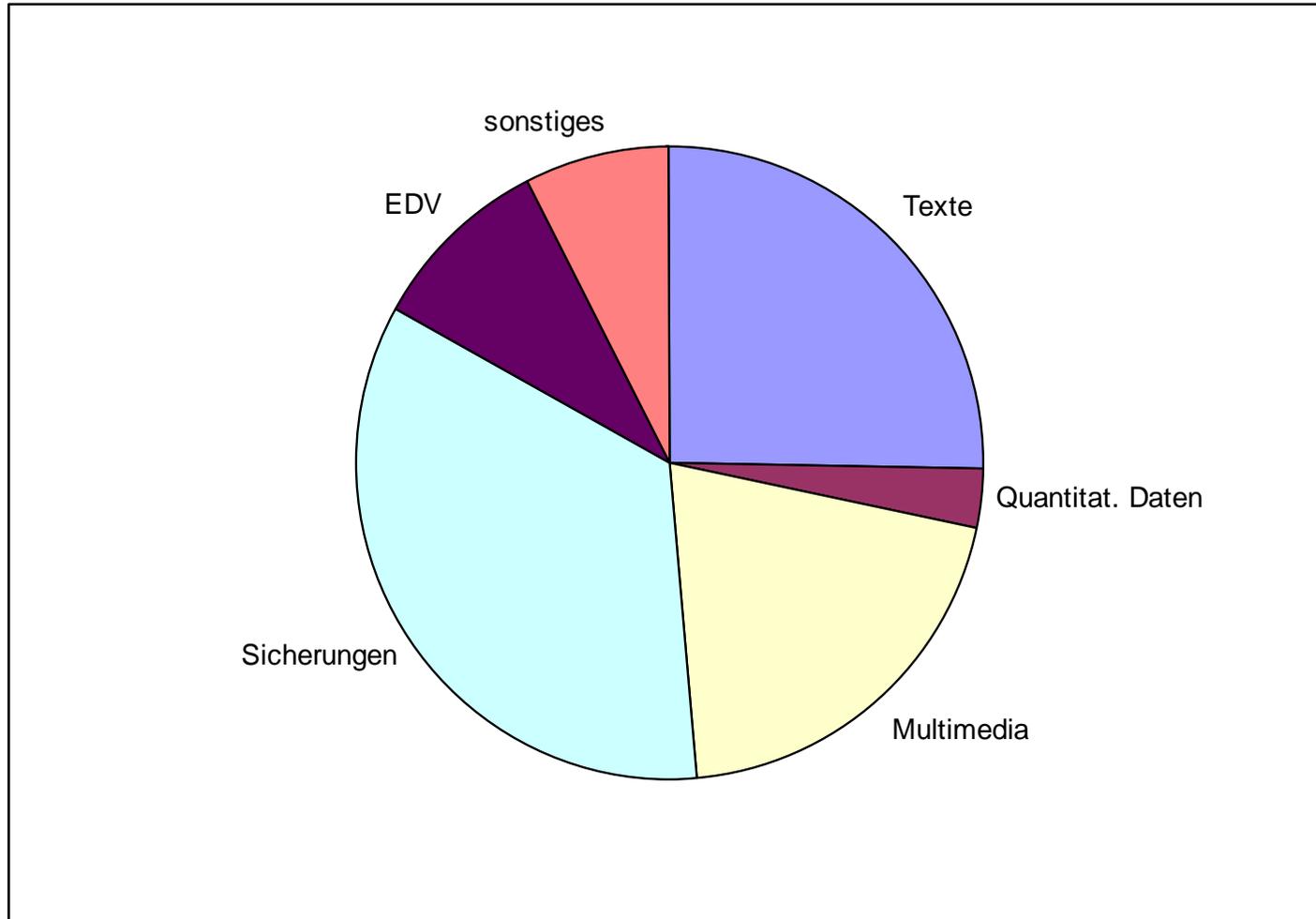
<b>Statistik:</b>				
sav	1.645	2.30	2.597	2.90
sps	983	1.37	24	0.03
rec	751	1.05	128	0.14

<b>Multimedia:</b>				
jpg	4.775	6.67	4.955	5.54
gif	1.585	2.21	17	0.02
mp3	115	0.16	3.969	4.44
wav	42	0.06	1.148	1.28
m4a	34	0.05	2.424	2.71
wma	19	0.03	476	0.53
avi	6	0.01	1.106	1.24
vob	5	0.01	4.046	4.52

<b>Sonstiges</b>				
7z	13	0.02	24.253	27.11
zip	450	0.63	3.520	3.93
[---]	1.229	1.72	3.075	3.44

<b>Software:</b>				
dll	1.114	1.56	196	0.22
au3	2.123	2.96	9	0.01
exe	588	0.82	3.295	3.68
iso	7	0.01	3.672	4.10
nrg	7	0.01	1.181	1.32
Summe	71.608	100.00	89.458	100.00

## Datenmenge nach Verwendungsfeldern



## Lebenszyklus von Forschungsdaten:

Fragestellung

Erhebung

Datenedition

*Primärnutzung* | *Sekundärnutzung*

interne Dokumentation | publizierbare Dokumentation

Primäranalyse | Datenweitergabe

Publikation | Sekundäranalyse

Publikation

# Beispiel eines dokumentierten Datensatzes

Datenbestandskatalog  
Studienbeschreibung 4391



<b>ZA4391: Gruppenbezogene Menschenfeindlichkeit (GMF-Survey 2002)</b>	
Studiennummer	ZA4391
Titel	Gruppenbezogene Menschenfeindlichkeit (GMF-Survey 2002)
Aktuelle Version zeige <a href="#">Versionshistorie &amp; Errata</a>	<ul style="list-style-type: none"><li>1.0.0, 13.04.2010, <b>doi:10.4232/1.4391</b> Versionsnummer automatisch erzeugt (Einführung einer einheitlichen Versionierung)</li></ul>
Erhebungszeitraum	Mai 2002 bis Juni 2002
Primärforscher/Wissenschaftlicher Beirat Institution	<ul style="list-style-type: none"><li>W. Heitmeyer Institut für interdisziplinäre Konflikt- und Gewaltforschung der Universität <b>Bielefeld</b></li></ul>
Datenerhebung	Infratest Sozialforschung, München
Inhalt	Rassismus, Fremdenfeindlichkeit, Antisemitismus, Islamphobie, Etabliertenvorrechte, Sexismus, Heterophobie.  Themen: Einschätzung der wirtschaftlichen Lage in der Bundesrepublik Deutschland sowie der eigenen wirtschaftlichen Lage und des eigenen Lebensstandards; Einschätzung der eigenen Zukunftssicherung; Beurteilung der wirtschaftlichen Lage in der Wohngegend im Vergleich mit der wirtschaftlichen Lage in der BRD;

**Datendokumentation und Datenweitergabe:**

**Welche konfligierenden Interessen bestehen?**

**Systemische Betrachtung (Finanziers, Öffentlichkeit, Wissenschaftssystem):**

- Forschungsdaten sind steuerfinanziertes Kollektivgut
- Informationsfreiheit auch für Forschungsdaten
- wissenschaftlicher Erkenntnisfortschritt nur durch fundierte Kritik (Merton)
- ergo: Sekundäranalysen dringend notwendig

**Dagegen: Egoismen der Akteure ...**

## Instituts- und Projektleitung:

- Renommee durch "tributpflichtige" Sekundärnutzer
- ohne eigene Mehrbelastung

## Mitarbeiter:

- Profilierung eher durch Primäranalysen
- Vorbereitung der Sekundärnutzung: unrentabler Mehraufwand
- Thematische Monopolstellung nicht unerwünscht
- Furcht vor Kritik
- Verweis auf bestehende Repositorien:
  - Datenarchiv für Sozialwissenschaften (Zentralarchiv)
  - ZPID

## Potentielle Nutzer

- Eigene Erhebung erbringt mehr Prestige als Sekundäranalyse
- Probleme der Konstruktvalidität bei Sekundäranalysen
- Veraltete Daten

## Warum ist für den Datenproduzenten die Dokumentation aufwendig?

- Grundgesamtheit und Stichprobe
- Befragungsform
- Antwortbereitschaft, Rücklauf, Repräsentativität, Bias
- Erhebungsinstrument
- Skalen (moralisches Urheberrecht)
- generierte Variablen (Indizes etc.)
- Validität und Reliabilität
- Gewichtung
- u. U. Anonymisierung



## Wie werden Datenbestände des IKG aufbewahrt?

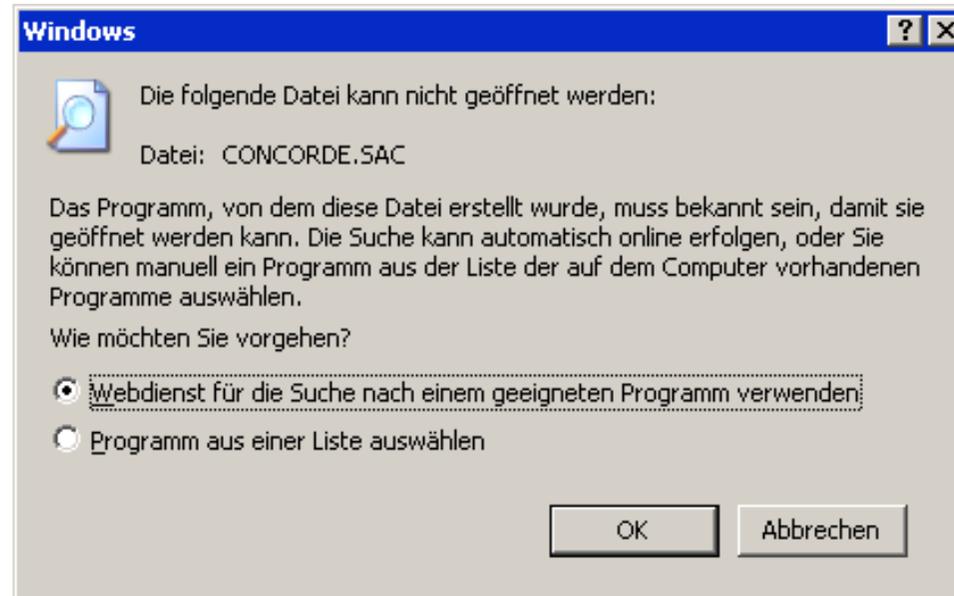
- Archivkeller (Papiere, Fragebögen): 10 Jahre
- Speicherung im Active Directory: befristet
- Sicherung auf optischen Medien
- Externe Festplatten

## Problem Medienhaltbarkeit



- physikalisch-chemische Haltbarkeit
- Geräte
- Schnittstellen

# Problem Dateiformate



## Dateiformate

- qualitative und quantitative Analyse:  
überwiegend herstellerspezifische, proprietäre Formate
- stets Konvertierungsaufwand (SPSS, Stata, SAS)
- massive Konvertierungsprobleme bei qualitativen Formaten (MAXQDA, AtlasTI)
- Kompatibilitätsprobleme zwischen Programmversionen eines Herstellers

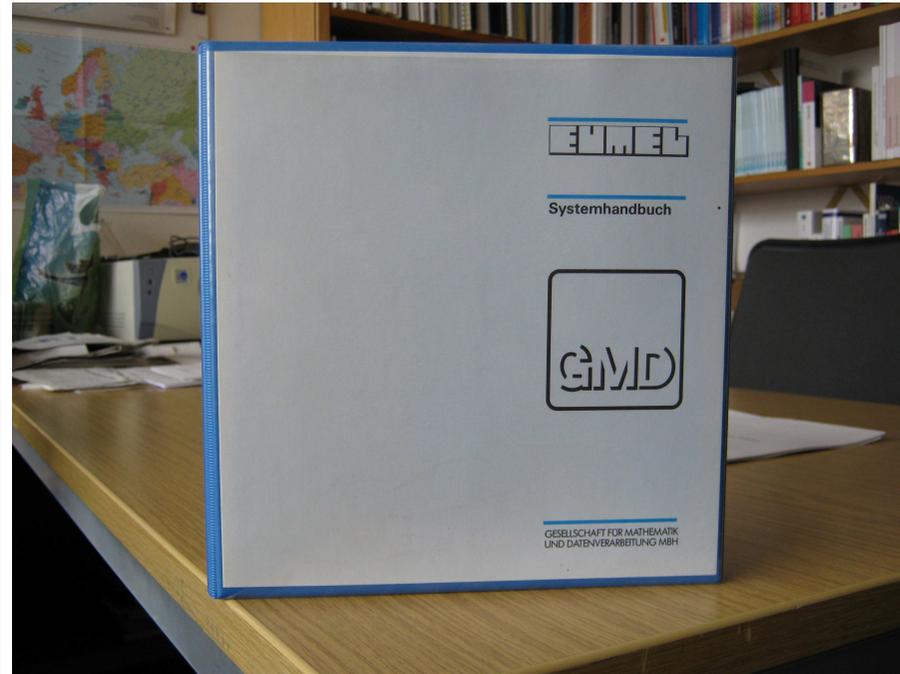
## Langfristige Archivierung ist ein ungelöstes Problem.

- unlesbare Medien
- überholte Geräte und Anschlußtechnik
- obsolete Dateiformate
- Papier ist das haltbarste Medium

# Hoffnungslose Fälle 1



## Hoffnungslose Fälle 2



**Ihre Fragen bitte!**