

Auswahl und Bewertung digitaler Forschungsdaten aus archivwissenschaftlicher Perspektive

nestor-Workshop Forschungsdaten

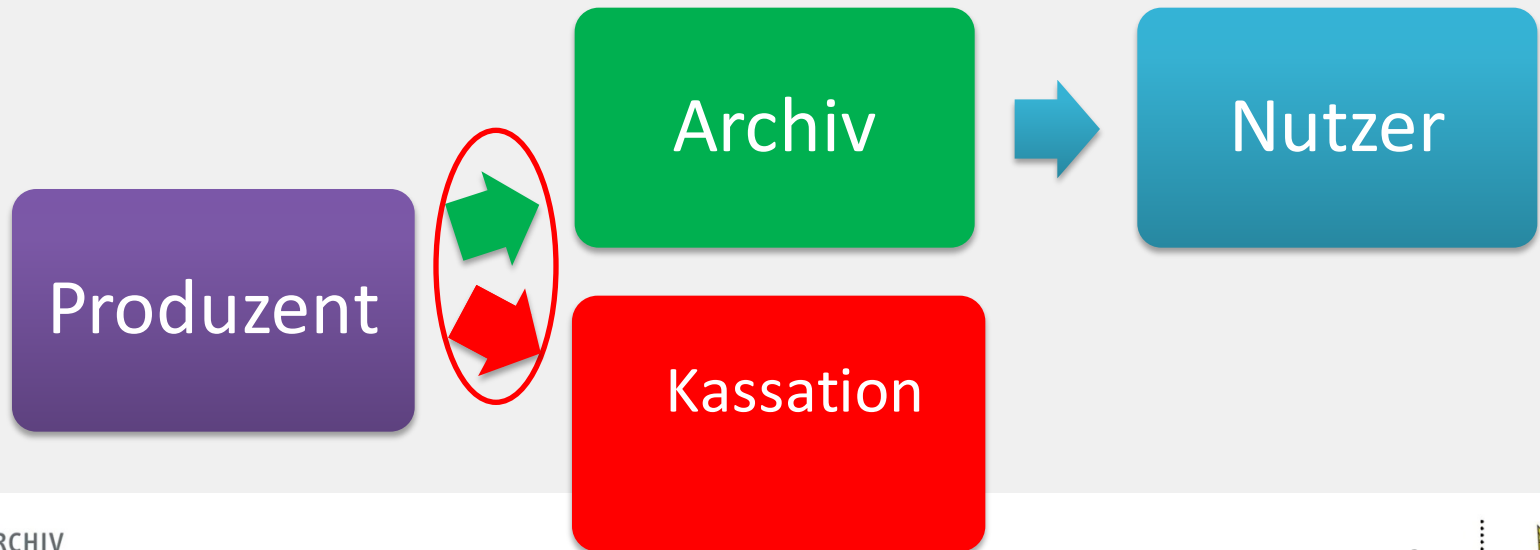
Essen, 17.11.2015

Prof. Dr. Christian Keitel



Basics

- **Bewertung** ist (wie alles archivische Handeln) stets auf den künftigen Nutzer ausgerichtet
- Unterlagen = alles, was zur Bewertung ansteht (Akten, Webseiten, Datenbanken...)
- i.d.R. werden 0 – 10% der angebotenen Unterlagen archiviert (Papier und digital).



Theodore Schellenberg (1903 – 1970)

- Primärwert: Wofür wurden die Unterlagen erstellt?
 - Bsp. 1: Volkszählung – Statistische Auswertung
 - Bsp. 2: Biotop – Naturschutz/Ermittlung Schutzbedürftigkeit
 - Bsp. 3: Anträge auf Asyl – Bearbeitung der Anträge
 - Bsp. 4: Webseite zum Hochwasserschutz – Information der Bevölkerung
- Sekundärwert: Wonach suchen die Nutzer im Archiv?
 - Bsp. 1: Volkszählung – Genealogie
 - Bsp. 2: Biotop – Historische Entwicklungen
 - Bsp. 3: Anträge auf Asyl – Sozialwissenschaftliche Analysen
 - Bsp. 4: Webseite zum Hochwasserschutz – Historische Analyse

Schellenberg über Sekundärwert

- Der Sekundärwert kann weiter untergliedert werden
 - Evidenzwert: Verweist auf Evidenz über den Produzenten und seine Arbeitsweise
 - Informationswert: Verweist auf einzelne Inhalte



Bewertung digitaler Unterlagen?

- Isn't it true that some of the 'graffiti' found at archaeological sites like Pompei have proven to be the most valuable and scientifically interesting artefacts? Without doubt, these examples are rather eye-catching and bold but make clear one point: acting as a guardian for future generations cannot be justified by reasons of dignity or decency; hushing up simply won't do. **We propose the supersession of appraisal, in favour of a process of random selection to complement the preservation of material in its entirety where completeness is essential.**

Robert Neumayer, Andreas Rauber, Why Appraisal is not 'Utterly' Useless and why it's not the Way to Go either. A Provocative Position Paper (PPP), 2007

Gründe für Bewertung

- Reduktion der Kosten
 - bei Papier: Platzbedarf (1815 bzw. 1918 ff. drohte massive Überfüllung)
 - bei Digitalem: Aufwände für Personal, Speicher und Strom
- Erleichterte Nutzung
 - bei Papier: Nutzer kann nur eine begrenzte Zahl an Dokumenten durchlesen
 - bei Digitalem: abhängig von technischen Suchmöglichkeiten
- Datenschutz bei personenbezogenen Angaben
 - Ausgleich zwischen Eingriff in die informationelle Selbstbestimmung und historischem Interesse der Gesellschaft
- Nicht-Bewertung ist passive Bewertung durch Zufall

Verfahren

- Unterlagen werden nicht isoliert, sondern im Abgleich mit den anderen Objekten/Informationsbereichen der abgebenden Stelle bewertet.
- Abgleich mit weiteren Stellen auf derselben Behördenebene und mit über- und nachgeordneten Stellen (= horizontale und vertikale Bewertung)
- Gespräche mit abgebender Stelle
 - Wichtige Hinweise auf relevante Unterlagen
 - Mitarbeiter markiert Akten während seiner alltäglichen Arbeit
 - Durch Gespräche bei Anbietung
 - Problematik
 - Mitarbeiter sehen manchmal Sekundärwert nicht
 - Sehr unterschiedliche Umsetzungsbereitschaft
- Zwei verschiedene Ausgangspunkte
 - Vorhandenes: Unterlagen/Aufgaben der Behörden
 - Wünschenswertes: Abgleich mit einem Dokumentationsprofil



Beispiel für Bewertungsmodell

Behörden

Archivieren
Bewerten
Vernichten

	Bewertung
IM:	A (Gesetze und Verwaltungsvorschriften) B (Beschwerden und Petitionen) V (Sperrerklärungen) B (übrige Unterlagen)
BPP:	V
BPA:	V
LPD:	V
	B (Dienstunfälle – einige wenige besondere Fälle)*
APD:	B (Dienstunfälle – einige wenige besondere Fälle)*

Vertikale und horizontale Bewertung der Unterlagen der Polizei in Baden-Württemberg



Beispiel für Dokumentationsprofil

1. Kategorisierung der lokaler Lebenswelt, z.B.

- Stadt und Raum
 - Verkehr
 - Ver- und Entsorgung
 - Grund und Boden
 - Bauwesen
 - ...

2. Bestimmung der Dokumentationsziele, z.B.

- Kategorie (Kommunal)Politik: Zu dokumentieren sind...
 - Gremien, Gruppen und Einzelpersonen, die in offiziellen und inoffiziellen Positionen politisch agieren
 - politische Ereignisse
 - Formen der politischen Beteiligung und der politischen Willensäußerung durch Mandatsträger/innen sowie Bürger/innen
 - Grundlagen (Werte, Vorstellungen, Normen)
 - ...

3. Suche nach Unterlagen



Statistische Unterlagen (1)

- Publikationen
 - Hohes Aggregationsniveau
 - Einführungen, Einbettung in Kontext
- Genesis etc. (Informationsdatenbanken im Internet)
 - Mittlere Aggregationsstufe
 - Manche Informationen sind statistikübergreifend verknüpfbar
- Mikrodaten/Rohdaten
 - nicht aggregiert
 - Kombinierte Abfrage über mehrere Merkmale der Erhebungseinheiten möglich

Statistische Unterlagen (2)

- Bewertung statistischer Unterlagen 2007,
https://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/ark/200805_ark_ag_statistikunterlagen_abschlussbericht.pdf
- Positive Bewertungskriterien für Mikrodaten:
 - Langfristige Relevanz
 - Gute Eignung für kleinräumige Auswertungen
 - Hohe Anzahl der Erhebungsmerkmale
 - Großzählungen
 - Gute Eignung zur Ergänzung der sonstigen archivischen Überlieferung
 - Hohe Zuverlässigkeit der Daten
 - Exemplarische Auswahl



Statistische Unterlagen (3)

- Negative Bewertungskriterien für Mikrodaten:
 - Publikation ausreichend
 - langfristig unwichtig
 - nichtstatistische Überlieferung ausreichend
 - Stichprobe zu klein
 - bezweckt nur Index-Erstellung
 - keine Erhebung (d.h. nur aus anderen Daten berechnet)
- Von den 520 Bundesstatistiken hat die Arbeitsgruppe 67 Statistiken aus 42 thematischen Einheiten positiv bewertet.

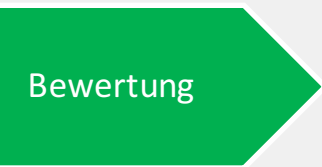
Modell zur Bewertung personen- bezogener Unterlagen (LABW)

- Perspektive: 1 Behörde mit Akten und Fachverfahren (Datenbanken)
- Vorgehen in zwei Schritten
 - Bestimmung der anzunehmenden Nutzungsziele (sind sie für die designated community relevant?)
 - Bezug auf die bestehenden Unterlagen (bestmögliche Abdeckung der relevanten Nutzungsziele)
- Nutzungsziele (Sekundärwert):
 - 1. Grundsicherung aus der Gesamtheit
 - 2. Statistisch auswertbare Teilmenge
 - 3. durchschnittliche (typische) Einzelfälle
 - 4. Herausragende Einzelfälle
 - 5. Evidenz/Dokumentation des Behördenhandelns
- Archivar 61 Heft 3 (2008), S. 275 – 278
http://www.archive.nrw.de/archivar/hefte/2008/ausgabe3/ARCHIVAR-03-2008_Internet.pdf
- 2011 Hinweise zur Anwendung des Bewertungsmodells beim Landesarchiv BW,
<http://www.landearchiv-bw.de/sixcms/media.php/120/52553/Hinweise.pdf>

Bewertung Akten

- Akten sind Container für Dokumente unterschiedlicher Form
- Auswahl ganzer Akten, teilweise über Datenbankabgleich
 - VIPs : Daten von Wikipedia + Landesbibliographie
 - Bedeutende Gerichtserfahren: Nachweis in JURIS
 - Zusätzlich: Ermittlung der „durchschnittliche“/zeittypischen Fälle

Bewertung Fachverfahren (1)



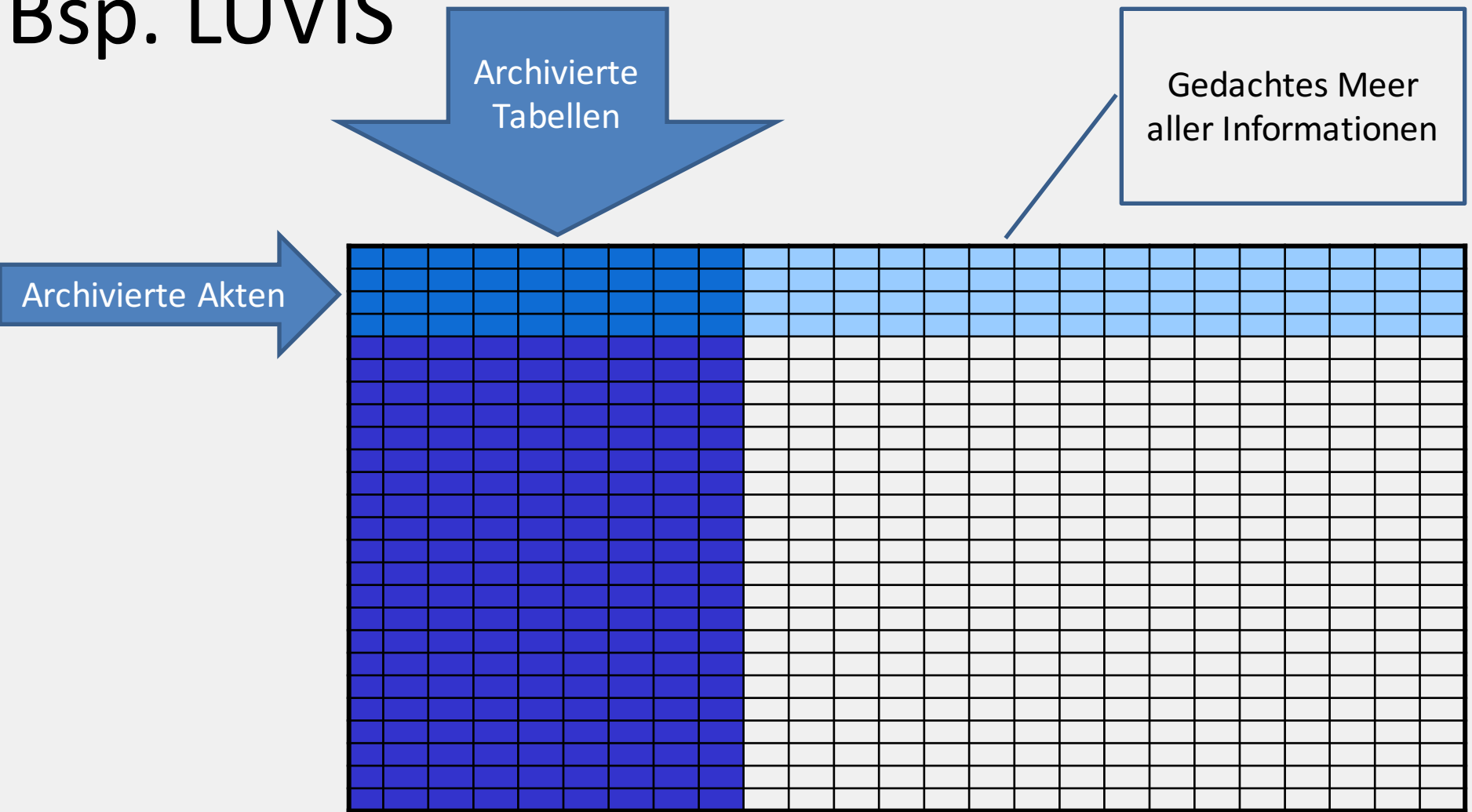
Die Bewertungsentscheidung wirkt sich auf alle Folgeprozesse aus.

Bewertung Fachverfahren (2)

- Fachverfahren basieren in der Regel auf Datenbanken und dienen der Erledigung von Aufgaben, die sich regelmäßig wiederholen.
- Einfache Verfahren (wenige Tabellen): Ggf. Komplettarchivierung
- Komplexe Verfahren: Auswahl von Tabellen
- Bsp. LÜVIS (Lebensmittelüberwachung in Baden-Württemberg)
 - > 300 Tabellen
 - 11 Abfragen als Tabellen archiviert: Tabelle entspricht einem Thema, keine zusätzliche Datensatzreduktion
 - Grundangaben zum Betriebe, Betriebsbesuche, Verstöße, Maßnahmen
 - zusätzlich Überlieferung einiger Papierakten (1 Akte mit allen Angaben zum Betrieb)



Bsp. LÜVIS



- Pro Spalte eine Angabe/Eigenschaft
- 1 „Datensatz“ enthält alle Angaben zu einem Betrieb

Umweltinformationssystem (1)

- Eine zentrale Datenbank in BaWü, die von n Fachverfahren der Kreise und Städte monatlich gespeist wird.
- Auswahl der vom Landesarchiv BW archivwürdig bewerteten Objektarten
 - Altlasten
 - Biotope
 - Wasserrechte
 - Amtliches digitales Wasserwirtschaftliches Gewässernetz
 - Fließgewässer
 - Stehende Gewässer
 - Einzugsgebiete
 - Wasserschutzgebiete
 - Quellenschutzgebiete
 - Grundwasseraufschlüsse
 - „Schutzgebiete nach Naturschutzrecht“
 - Kommunale Kläranlagen

Umweltinformationssystem (2): Archivwürdig und Verzichtbar

- **A** Wert für Forschungsarbeiten, für die zeitlich weit zurückliegende Daten erforderlich sind, in den Natur- und Geisteswissenschaften.
- **A** Wert für den Nachweis rechtsrelevanter Sachverhalte für Bürger, Staat und Institutionen.
- **A** Geobasisdaten im kleinsten erhältlichen Maßstab. Alle übrigen Geobasisdaten V.
- **A** Digitalisierte topographische Karten, Orthophotos und Schummerungskarten als Orientierungshilfe bei der Nutzung der Daten.
- **V** Objekte, die Teilmengen anderer Objekte darstellen – z.B. verschiedene Waldkategorien (Immissionsschutz, Wasserschutz etc.), sofern diese Flächen auch über die gesamte Landesfläche nach diesen Kategorien eingestuft sind.
- **V** Statistische Daten, sofern davon auszugehen ist, dass die Sachdaten eine hinreichende Basis für statistische Auswertungen bieten werden.
- **V** Wenn dieser Themenbereich in einem Bewertungsmodell der Archivverwaltung (Forst, Wasser, Umwelt, Straßenbau) insgesamt als „zu vernichten“ eingestuft wurde und keine weiteren Motive entgegenstehen.
- **V** Versuchs- und Probe-Objekte, da es sich nicht um aussagekräftige Endergebnisse handeln dürfte.
- **V** Objekte, die besondere Aufbereitungen anderer Objekte darstellen.
- **V** Objekte, die Angaben enthalten, die sich aus Kartendaten errechnen lassen (z. B. Flächenangaben über Gebiete, die als Polygone ausgewiesen sind)

Ausdünnung beim Produzenten: Bsp. Luftbilder

Arbeitsgemeinschaft der
Vermessungsverwaltungen
der Länder der
Bundesrepublik Deutschland
(AdV)
Leitfaden zur Sicherung von
Luftbilddatenbeständen

1. Jahr	2.	3.	4. Jahr	5.	6.	7.	8.	9.	10. Jahr	TB
R-G-B-NIR, PAN			R-G-B-NIR, PAN							
16 bit			16 bit							
TIFF			JPEG2000							
unkomprimiert			lossless							
14 TB	14	14	7	7	7	7	7	7	7	91

11. Jahr	12.	13.	14.	15.	16.	17.	18.	19.	20. Jahr	TB
R-G-B-NIR										
8 bit										
JPEG2000										
lossless										
3 TB	3	3	3	3	3	3	3	3	3	30

21. Jahr	22.	23.	24.	25.	26.	27.	28.	29.	30. Jahr	TB
—	—	—	R-G-B-NIR			—	—	—	—	
—	—	—	8 bit			—	—	—	—	
—	—	—	JPEG2000			—	—	—	—	
—	—	—	lossless			—	—	—	—	
—	—	—	3 TB	3 TB	3 TB	—	—	—	—	9

