



Auswahl und Bewertung von Forschungsdaten für die Archivierung und Nachnutzung

Reiner Mauer

GESIS – Leibniz-Institut für
Sozialwissenschaften

Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).



Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

Bewertung

- Dient der Auswahl von digitalen Objekten für die Archivierung
- Informiert über den Zustand der digitalen Objekte
- Kann Grundlage für Kuratierungsentscheidungen, die Zuweisung von Service-/Kuratierungslevels und Aufbewahrungsfristen sein

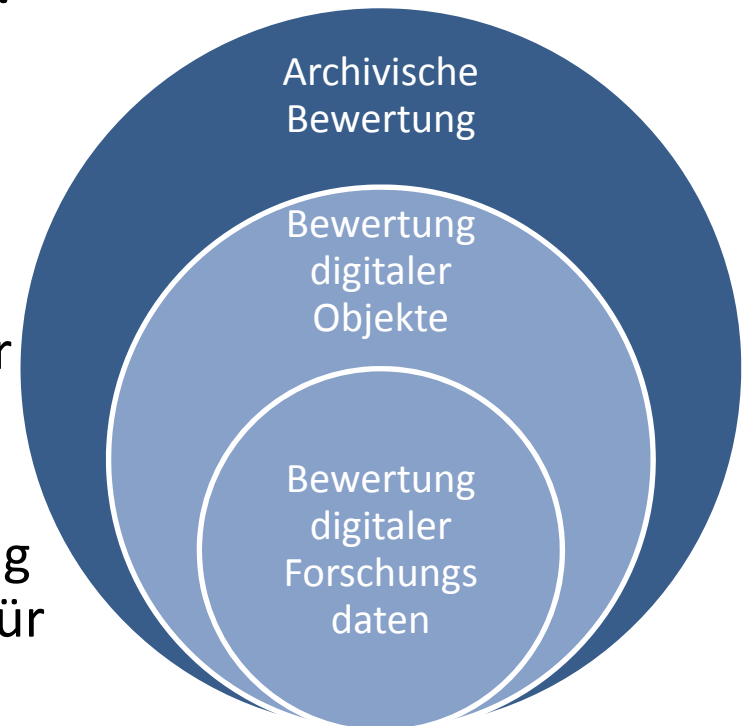
„Was soll für wen zu welchem Zweck für wie lange archiviert und in welchem Umfang kuratiert werden?“



Bild: CC-0

Sonderfall Forschungsdaten?

- Unterscheiden sich Daten in Auswahl und Bewertung von anderen digitalen Objekten?
 - Eher nein!
 - Tendenziell Archivierung von Daten weniger zur „reinen“ Überlieferungsbildung, sondern als Grundlage der Nutzung in zukünftiger Forschung
 - Beispiel statistische Daten: einzelne beispielhafte Fälle für die Archivierung auswählen vs komplette Datenbank für die spätere Nutzung bewahren

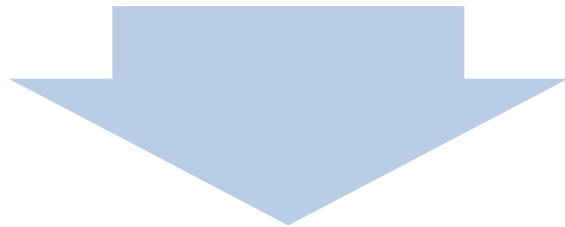


Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- **Auswahl – warum?**
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

Auswahl – warum?

- Alle Daten sind wertvoll
- Technologischer Fortschritt
 - immer leistungsfähigere Speicherlösungen (z. B. „die Cloud“)
 - Speicherplatz wird billiger



Müssen oder können nicht alle Daten aufgehoben bzw. archiviert werden?



Bild: CC-0

Auswahl - warum?

- Technische Einschränkungen
 - Speicherkapazität bei extrem großen Datenmengen nicht ausreichend
 - Beispiel CERN, Large Hadron Collider: ca. 99,9 % der Daten, die in den Detektoren anfallen, wird unmittelbar ‚aussortiert‘, 0,0025 % werden vollständig analysiert und am Ende werden nur 0,0005 % dauerhaft gespeichert
- Forschungsethische und rechtliche Einschränkungen
 - z.B. keine informierte Einwilligung, sensible Daten, Datenschutz

Auswahl - warum?

- Kosten
 - einfaches Abspeichern nicht ausreichend – Auffindbarkeit, Interpretierbarkeit und langfristige Zugänglichkeit müssen gewährleistet sein
 - Maßnahmen sind (je nach Datentyp) kostenintensiv
- Informationsüberflutung
 - Je mehr (nur gering erschlossene) Informationen verfügbar sind, desto schwieriger wird es, Relevantes aus dem „Rauschen“ herauszufiltern

Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- **Auswahl und Bewertung – wie?**
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

Auswahl und Bewertung – wie?

- Auswahlprozess sollte sich auszeichnen durch
 - Konsistenz der Bewertung (sowohl über Bewerter_innen als auch über die Zeit hinweg)
 - Nachvollziehbarkeit
 - Transparenz
- Erreichbar durch
 - definierten, dokumentierten Prozess
 - klar formulierte Auswahl-/Bewertungskriterien

Kriterien – Beispiele

Kategorie	Beispiel
Technisch	<ul style="list-style-type: none">• Dateiformat• Volumen (Größe der Datei)• technische Restriktionen (DRM)• Integrität/Authentizität
Inhaltlich-methodisch	<ul style="list-style-type: none">• Disziplin• Forschungsgegenstand• Methode• (methodische) Qualität• Verständlichkeit (Qualität der Dokumentation)• Vollständigkeit• Einzigartigkeit• wissenschaftliche Güte

Kriterien – Beispiele

Kategorie	Beispiel
Rechtlich	<ul style="list-style-type: none">• gesetzlicher Auftrag• Aufbewahrungspflichten und -fristen• Urheberrecht• Datenschutz
Ethisch	<ul style="list-style-type: none">• (informierte) Einwilligung von Teilnehmer_innen
Institutioneller Rahmen	<ul style="list-style-type: none">• Passung mit Sammelrichtlinie, institutioneller Mission• institutionelle Zugehörigkeit der Datengeber_in• Kosten der Archivierung• Bedürfnisse der Zielgruppen

Kriterientypen

- Formale vs inhaltliche Kriterien
 - Formal: Benutzbarkeit/Archivierbarkeit (Format, Größe, Virenfreiheit, Vollständigkeit); Rechtliche Situation (Urheberrecht, Datenschutz); Affiliation Primärforscher_in; sind Daten schon in einem anderen Archiv/ Repository
 - Inhaltlich: Untersuchungsthema, Methodik, wissenschaftlicher Wert, Qualität
- K.O./Ausschluss-Kriterien vs qualifizierende Kriterien
 - Ausschluss: rechtl. Restriktionen (Nutzungsrechte, Datenschutz); außerhalb des Sammelgebiets
 - Qualifizierende Kriterien: „Ausgangszustand“ der Daten; wissenschaftlicher Wert (thematisch, methodisch), Einzigartigkeit (Messung kann nicht wiederholt werden)

Kriterienbildung

- Operationalisierung
 - relativ einfach bei formalen Kriterien (Dateiformat, Volumen, Erhebungs-/Meßmethode, Urheberrechte...)
 - abstraktere Konstrukte (Qualität, Nachnutzungspotenzial, wissenschaftlicher Wert) müssen „messbar“ und Bewertung für Dritte nachvollziehbar gemacht werden

Kriterienbildung

- Wie bestimme ich die für meine Organisation relevanten Kriterien?
 - ⇒ Rahmenbedingungen und Kontext betrachten und Kriterien hieraus ableiten

Relevante Rahmenbedingungen

- Aufgabe/Zweck der Institution
 - z.B. Sammelrichtlinie, Leitbild, Satzung
- Zur Verfügung stehende Ressourcen
 - finanzielle, technische und personelle
- Zielgruppen
 - Bedürfnisse, Erwartungen
- Gesetzliche Vorgaben
 - z.B. Aufbewahrungsfristen für bestimmte Datentypen,
- Markt/Mitbewerber

Ableitung von Kriterien: Beispiel GESIS Datenarchiv

- Für wen sollen wir sammeln?

Auszug GESIS Leitbild

- Unsere Angebote richten sich primär an Forscher der empirischen Sozialforschung mit einem Schwerpunkt auf den Fachgebieten Soziologie und Politikwissenschaft sowie an solche der Sozialwissenschaften insgesamt.

Auszug GESIS Satzung (§2 Zweck)

- e) Schaffung eines benutzerfreundlichen und hochqualitativen Zugangs zu allen für die empirische Sozialforschung relevanten Informationen und Daten

Abzuleitendes Kriterium:

- Daten müssen für die Zielgruppe gut zu benutzen sein

Ableitung von Kriterien: Beispiel GESIS Datenarchiv

- Was sollen wir sammeln?

Auszug GESIS Satzung (§2 Zweck)

b) Beschreibung und Erklärung gesellschaftlicher Entwicklungen in nationaler, international vergleichender und historischer Perspektive einschließlich der Datenerhebung, statistischen Modellierung und Dauerbeobachtung,

c) Archivierung, Dokumentation und Langzeitsicherung sozialwissenschaftlicher Daten, einschließlich ihrer Erschließung sowie qualitativ hochwertigen Aufbereitung besonders relevanter Daten für Sekundäranalysen,

Abzuleitende Kriterien:

- Daten sind geeignet, Aussagen über genannte Themen zu treffen
- Daten, die einen Vergleich in zeitlicher, räumlicher bzw. kultureller Perspektive erlauben, sind von besonderer Bedeutung
- Relevanz der Daten (wissenschaftlicher Wert) für die Zielgruppe
- Eignung für Sekundäranalysen

Ableitung von Kriterien: Beispiel GESIS Datenarchiv

	Kriterien (Auszug)	Inhaltliche Passung	Eignung für Sekundäranalysen	wiss. Relevanz
Technisch	Dateiformat ist mit gängiger Software zu verarbeiten			
	Dateien überschreiten max. Volumen nicht			
	Dateien unterliegen keinen technischen Zugriffsbeschränkungen			
	Dateien sind virenfrei und nicht beschädigt			
Inhaltlich-methodisch	Die Untersuchung ist geeignet, sozialwissenschaftlich relevante Fragestellungen zu beantworten.			
	Die Daten sprechen voraussichtlich einen breiten sozialwissenschaftlichen Nutzerkreis an.			
	Die Untersuchung ist hinsichtlich des Fragenprogramms bzw. der erhobenen Daten innovativ oder einzigartig.			
	Die Daten sind umfassend dokumentiert			
	a. Es ist ein detaillierter Projekt- bzw. Methodenbericht vorhanden.			
	b. Der Fragebogen sowie alle weiteren Untersuchungsmaterialien sind dokumentiert.			
	Die Untersuchung ermöglicht Vergleiche über die Zeit oder über Länder hinweg.			
Ethisch-rechtlich	Es liegt eine Einverständniserklärung der Teilnehmer_innen vor			
	Die Urheberrechte der Daten sind geklärt (insbesondere Rechteinhaber/Rechtsnachfolger).			

Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- **Kriterien in der Anwendung: Auswahlprozess implementieren**
- Herausforderungen

Kriterien in der Anwendung: Auswahlprozess implementieren

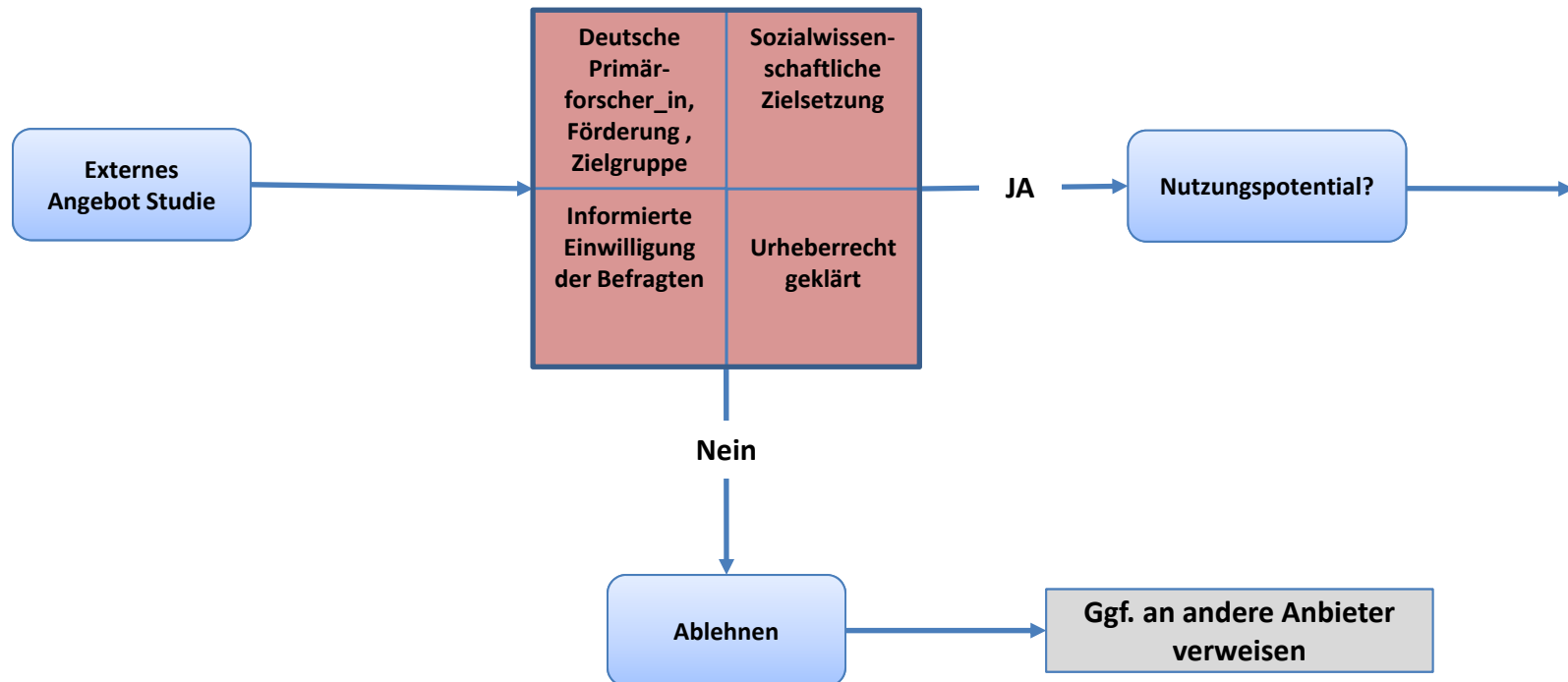
- Rollen und Verantwortlichkeiten müssen geklärt werden (Archiv, Datenproduzent_in, Zielgruppe/Community, Förderer, Politik)
- Konsistenz und Nachvollziehbarkeit gewährleisten
 - Verständliche Beschreibung der Kriterien,
 - Schulung von Mitarbeiter_innen,
 - Checkliste oder Leitfragenkatalog,
 - Vier-Augen-Prinzip oder Auswahlkonferenzen
- Transparenz gewährleisten
 - Dokumentation des Prozesses / der Entscheidung
- Effizienz gewährleisten
 - Bewertung muss auf möglichst hohem Abstraktionsniveau getroffen werden (insbesondere wenn große Volumina bewertet müssen; im Extremfall kostet Selektion und Bewertung ansonsten mehr als die Archivierung)

Kriterien in der Anwendung: Auswahlprozess implementieren

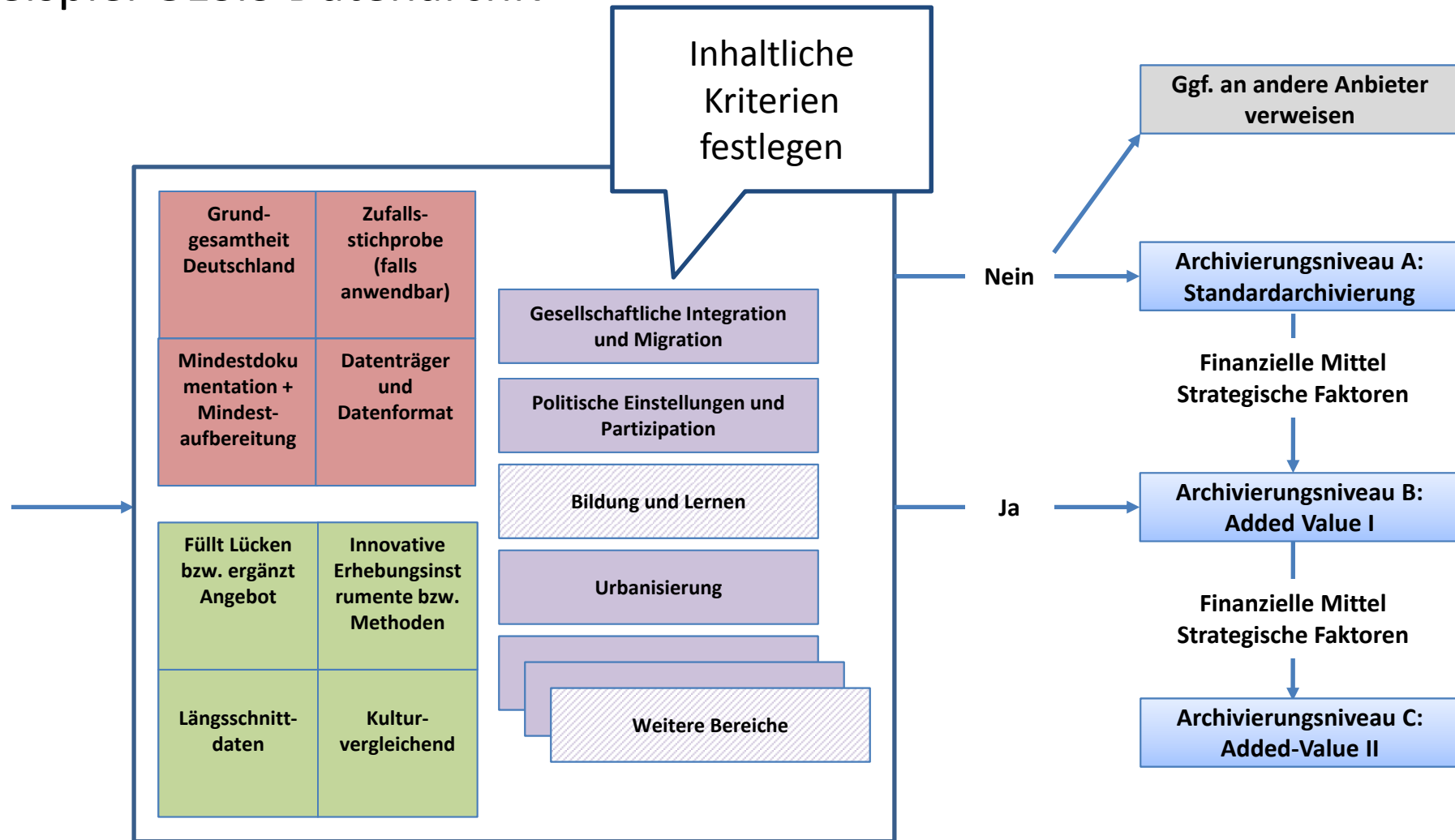
- Typischerweise: mehrstufiger Auswahlprozess
 - Passen angebotene Daten grundsätzlich ins Profil? Sind alle „Ausschlusskriterien“ positiv erfüllt?
 - Wenn dies der Fall ist, beginnt die Bewertung der Daten mit Blick auf Qualität etc.
 - Es folgt eine Entscheidung über Übernahme oder Kassation beziehungsweise bei Übernahme ggf. Entscheidung für ein Kuratierungslevel

Beispiel GESIS Datenarchiv

Aufzunehmende Studien müssen bestimmte Mindestanforderungen erfüllen



Beispiel GESIS Datenarchiv



Kriterien in der Anwendung: Auswahlprozess implementieren

- Wie lehnt man eingereichte Forschungsdaten ab?

Nicht in jedem Fall endgültig...

... aber immer mit Begründung!

- ⇒ Je nach Ablehnungsgrund kann Einreicher_in Nachbesserungen vornehmen (z.B. fehlende Dokumentation vervollständigen)
- ⇒ Bewertungs-/Entscheidungsprozess so dokumentieren, dass Datenproduzent_innen und andere verstehen können, warum Daten aufbewahrt bzw. zurückgewiesen wurden.

Gliederung

- Einleitung: Auswahl und Bewertung von Forschungsdaten
- Auswahl – warum?
- Auswahl und Bewertung – wie?
- Kriterien in der Anwendung: Auswahlprozess implementieren
- Herausforderungen

Herausforderungen (1)

- Beurteilung des Werts von Daten für zukünftige Forschung:
 - Welche Themen, welche Arten von Daten sind zukünftig relevant?
 - Können Daten, die aus Forschungsprojekten stammen, die bereits mehrfach bewertet wurden (Förderantrag, auf Daten basierende Publikationen, Evaluationen ...), “wertlos” sein?
- ⇒ Bewertung kann nicht nur von Archiven, Repositorien, Infrastrukturen vorgenommen werden. Alle Stakeholder (Fachgesellschaften, Forschungsförderer, Forschungspolitik und Datenprofis müssen in Auswahlprozess einbezogen werden
- ⇒ Es lässt sich nicht vermeiden, dass ‘Fehlentscheidungen’ im Auswahlprozess getroffen werden

Herausforderungen (2)

- Um Nachvollziehbarkeit und Transparenz von Wissenschaft zu unterstützen, sollen auch Forschungsdaten offengelegt werden
 - Zunehmend auch verpflichtende Auflagen von Forschungsförderern oder Zeitschriften
- ⇒ D.h. eigentlich müssen ***alle*** Daten aufbewahrt werden (zumindest für einen bestimmten Zeitraum)

Einführende Literatur

- Keitel, Ch. (2015). Auswahl und Bewertung digitaler Forschungsdaten aus archivwissenschaftlicher Perspektive. DINI/nestor-Workshop Forschungsdaten. Essen, 17.11.2015.
<http://www.forschungsdaten.org/index.php/Datei:04--keitel--Bewertung-Forschungsdaten.pdf>
- Tjalsma, H., & Rombouts, J. (2010). Selection of Research Data. Guidelines for Appraising and Selecting Research Data.
<https://www.surf.nl/en/knowledge-base/2010/research-report-selection-of-research-data.html>
- Whyte, A. & Wilson, A. (2010). How to Appraise and Select Research Data for Curation. DCC How-to Guides. Edinburgh: Digital Curation Centre.
<http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>
- Präsentationen des 4. DINI/nestor Workshops „Forschungsdaten auswählen und bewerten“. Essen, 17.11.2015.
<http://www.forschungsdaten.org/index.php/DINI-nestor-WS4>