

Die Entwirrung des Datenknäuels – Herangehensweise bei der Bewertung und Übernahme schwach strukturierter digitaler Unterlagen







# **Agenda**



- 1. Allgemein
- 2. Digitale Fotos der Pressestelle des Kreises Soest
- 3. Fazit
- 4. Quellen und Linksammlung





#### Dateisammlungen

"Menge von Einzeldateien, die von einem oder mehreren Bearbeitern zur Erledigung einer oder mehrerer Aufgaben über einen bestimmten Zeitraum erstellt und nach individuellen Ordnungskriterien zusammengestellt wurden.

Die Dateien liegen auf einer Ebene und/oder hierarchisch in einer Verzeichnisstruktur vor. Es können in einer Dateisammlung unterschiedlichste Dateiformate enthalten sein."

Annekathrin MIEGEL/Sigrid SCHIEBER/Christoph SCHMIDT, Vom richtigen Umgang mit kreativen digitalen Ablagen, in: Kai Naumann/Michael Puchta (Hrsg.), Kreative digitale Ablagen und die Archive. Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns, München 2017, S. 7-16, hier: S. 7.





- Dateisammlungen werden jedem Archivmitarbeitenden in unterschiedlichen Umfängen fortlaufend begegnen
  - Überreste der "digitalen Akte" vor Einführung der E-Akte
  - Dateien, die für eine Übernahme in die Papierakte nicht geeignet waren (z.B. Technische Zeichnungen, digitale Pressefotos, GIS-Daten)
  - Dateisammlungen von Privatpersonen, Vereinen, Verbänden, Parteien etc.
- Gefahr, dass vielfach potenziell archivwürdige Unterlagen keinen Weg in die Papierakte finden





- Umfängen, Ausmaßen, Gliederungstiefen und allgemein der Kreativität der erstellenden Personen sind dabei keine Grenzen gesetzt
- Bewertung und sinnvolle Strukturierung zu Übernahme häufig sehr zeitintensiv und komplex
- Herausforderungen:
  - Fehlende Metadaten
  - Fehlender Kontext
  - Mehrfach vorhandene Dateien
  - Beschädigte/nicht mehr lesbare Dateien
  - Exotische Dateiformate
  - Etc.





- Nachteile Dateisammlungen gegenüber DMS oder Papierakte:
  - Keine Revisionssicherheit
  - Keine Hinterlegung von Aufbewahrungsfristen möglich
  - Keine Möglichkeit zur strukturierten Bewertung
  - Kein geregeltes Aussonderungsverfahren
  - Fehlender Kontext und fehlende Metadaten





- Vorgehensweise folgt einem Drei-Phasen-Modell nach Kai Naumann und Marco Birn:
  - 1) Analyse
  - 2) (Nach-)Bewertung und Formierung der Archivpakete
  - 3) Erzeugung der Archivpakete für das digitale Langzeitarchiv und Verzeichnungssystem





#### Allgemein

- Digitale Fotos seit ca. Anfang der 2000er Jahre auf einem eigenen Laufwerk
- Umfang:
  - Ca. 174.000 Dateien in ca. 9.500 Ordnern
  - Ca. 1 TB an Daten
  - Struktur bis zu 12 Ebenen in die Tiefe
- Mehrstufiger Prozess zur Übernahme (Beginn: Mitte 2021)
- Protokollierung aller Prozessschritte
- Allgemeine Dokumentation des gesamten Projekts → Projekt mit Pilotcharakter für das digitale Langzeitarchiv des Kreisarchivs Soest





	7	→ Dieser PC → kreisarchiv (\\i200d01s52) (P:	) > Fotoarchiv > _Pre	ssetotos_neu
me	Anderungsd	^	1020	
Diverses	22.11.2022 0	Name	Änderungsdatum	Тур
DQ-Tool-Monitorkalibrierung	22.11.2022 0	alt	22.11.2022 07:21	Dateiordr
Filmaufnahmen2016	22.11.2022 0	Pressefotos 2017	22.11.2022 07:24	Dateiord
FotoCDRomDerianRaffi	22.11.2022 0	The state of the s		
Fotos Eissportzentrum	22.11.2022 0	Pressefotos 2018	22.11.2022 07:26	Dateiord
Fotos Film	22.11.2022 0	_Pressefotos 2019	22.11.2022 07:31	Dateiord
Fotos fuer Schulbuch	22.11.2022 0	Pressefotos 2020	22.11.2022 07:32	Dateiord
Fotos für Kreisbroschüre	22.11.2022 0	_Pressefotos 2021	22.11.2022 07:34	Dateiord
Fotos neues Kreisbuch 2007-2008	22.11.2022 0	_Pressefotos 2022	22.11.2022 07:36	Dateiord
Fotos neues Kreisbuch 2012-2013	22.11.2022 0	Sicherung SD Karte 64 GB	22.11.2022 07:38	Dateiordi
FotosESG	22.11.2022 0		22.11.2022 07:11	Dateiorde
FotosfuerInstagram	22.11.2022 0	- 20010118 Neujahrsempfang	22.11.2022 07:11	Dateiord
Fotosfür Heimatkalender	22,11,2022 0	20141203BesuchergruppeKolpingbildung	22.11.2022 07:11	Dateiordi
Fotosfür Publikationen	22.11.2022 0	20150128Einbürgerungsurkunde-Symbol	22.11.2022 07:11	Dateiord
Fotosfür Sonderseite Glocke 2005	22.11.2022 0	20171127AnkuendigungAusstellungKaun	22.11.2022 07:11	Dateiord
		20171212FörderbescheidRuff+Lübben	22.11.2022 07:11	Dateiord
		AHA	22.11.2022 07:11	Dateiordr

Seite 9 · 29.04.2024

Südwestfalen

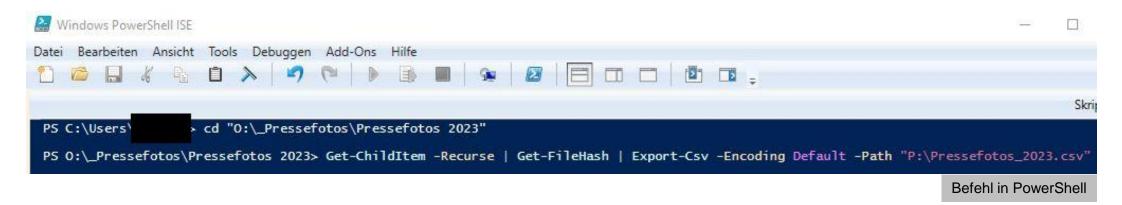


#### Bearbeitungsschritte vor der Bewertung

 1) Erzeugung Excel-Tabellen mit Ordnerstruktur des Laufwerks inkl. Dateipfad und Hashwerten → PowerShell oder TreeSize Professional







#### Erzeugte csv-Datei

1848 SHA256, "3E8DA276F2C64ECA80D80820937E6B962C53ACD010CA8090EA56D86A2A53F22E", "O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2699.JPG" 1849 SHA256,"12E3F7F7758D9B29197D59D56892D3784BB063C2BEB51D10D4342D8A3D9690C1","O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2700.JPG" 1850 SHA256, "E877E1897AEBF710ADE12E86D4A54DACA37B6E2722135AC0D4BCC45F106625C2", "O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2701.JPG" 1851 SHA256, "57E3115FB8265EA895DDC8196165D9248A0652B1AFEE6C7E17DEF35CF2AE656E", "O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2702.JPG" 1852 SHA256, "0BB73DD49EFBC26AB988C2C2E9429237DD858A1983938627220CD2CB8220D487", "O:\\_Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2703.JPG" 1853 SHA256, "E3AC5D667A9CB3F5E9812E326C3869F2A72C7FA50E0AD8A26DC67754DDBD91DE", "O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2704.JPG" 1854 SHA256,"1A7E3802E10D8998EA475EF1351984BED52E4D5FF026238817D5DF7135DF3559","O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2705.JPG" 1855 SHA256, "C8492C416549D7A670C7E7C31A7B6EF120760C153F7D25CDEBDF947D3DC8AC7A", "O:\ Pressefotos\Pressefotos 2023\20230703 Symbolfotos Führerscheine\DSC 2706.JPG" 1856 SHA256, "91D7CDF29D8E304C1E4E7FA562C4DFB778FF6FB585032E0CBD1DBD7F2825BDBE", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\20230704-2 1857 SHA256, "8EE372A10DBE75B90D53B80FD2CB323A11B8DF6FCF2AC0EBF911F1082AA13BBF", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\20230704-2 1858 SHA256, "DA0C89970E7F793C8B76E6B8487B67E31169BA10D0D3EE77EE1A9590C0B267C4", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2707.JF 1859 SHA256, "D483BD5203757B8793A0938152330D89A834004101E2735D725F8003521579B4", "O:\ Pressefotos\Pressefotos 2023\2023\704 Mobilfunkmesswoche Ergebnisse\DSC 2708.JPG" 1860 SHA256,"8CC4CF3499ABF9CF966575E6506CDD4F0BA2F094F22D838732267B1989BFBE9E","O:\ Pressefotos\Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2709.JP 1861 SHA256, "F433BAEF22154B0489C487D083FC521A2790D7F96D12CB2305D2515BDB383D9D", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2710.JP0 1862 SHA256, "B56CD870E93E21587BF32340543EDD9460710FEFABDE6A674FCBB146FB404A4E", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2711.JF 1863 SHA256,"163596521A7B6ACA8BFDFA996760E355FA44D5EE37877AD9F72AA25B438A63D4","O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2712.JF 1864 SHA256, "D43E5AECE97FE36CD2FBB67E5C8472B2D6F5853EEB2257B7FB6E25C73E8E376B", "O:\ Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Ergebnisse\DSC 2713.6

1865 SHA256 "56FEC24FC88D26F4AA0F10F118FF0882B4720AB7C16DD0F7BC64161F740DCFD6" "O:\ Pressefotos\Pressefotos\Pressefotos 2023\20230704 Mobilfunkmesswoche Fraebnisse\DSC 2714 | Südwestfalen



Südwestfalen

40,2	2 GB O:\_Pressefotos\Pressefotos 2023 ^	Diagramm 📰 Details 🧨 Dateitypen 👛 Benutzer 🧮 Dateialter 🕆	Top Dateien	Mistorie	
	3,4 GB 20230710 Naturdenkmäler, I	Name ^	Größe	Dateien	Pfadlänge
	2,9 GB 20230922 Bildungsregion Pc	202309 Angehörigen Pflegetag	1,7 MB	1	62
>	2,4 GB 20230901 Neue Auszubilder	20230102 Hochsitz Jagd Symbolfoto	16,9 MB	6	67
	1,0 GB 20230614 Firmenlauf	20230102 Windsack Symbolfoto	4,2 MB	3	62
> 📙	878,2 MB 20230830 Pflegestern	20230111 Treffen Bildungsrat der Bildungsregion	79,6 MB	12	81
	843,5 MB 20230620 Kreistag	20230112 Fenster öffnen Lüften Schimmel	1,5 MB	1	73
	817,1 MB 20230425 Dezernat 03	20230116-ss-Bildungsregion-Perspektivtagung	432,9 MB	48	77
	780,3 MB 20230427 Girls and Boys Day	20230119 Ausbildungsplätze	686,7 KB	2	60
	746,2 MB 20231115 PK Pflanzenschutz	20230120 Abendessen Firma Eaton INTERN	47,4 MB	6	72
	741,0 MB 20231106 FTZ	20230120 MINT Schulnetzwerk wfg	11,8 MB	5	65
	727,8 MB 20231011 Blitzerfotos	20230120 Symbolfotos Trans Transgender Divers jdw	119,0 MB	33	83
	688,0 MB 20230704 Mobilfunkmesswo	20230123 Förderlogos Familienfreundliche Unternehmen (von Gomes-wfg)	7,0 MB	7	102
	666,7 MB 20230314 Schuleingangsunt	20230123 Nebeneingang Auslaenderbehoerde Kreis Soest	43,2 MB	5	86
	654,5 MB 20230913 Aufsteller Klimaar	20230125 Ökoprofit Workshop	671,3 KB	1	61
	640,6 MB 20230703 Symbolfotos Führ	20230125 Rollcontainer	89,0 MB	17	56
	606.3 MB 20230426 Workshop Behind	20230126 FIS Werbung SM	4.2 MR	4	57



2 0	:\_Pressefotos\Pressefotos 2023\ (\\i200d01s52	\Fotoarchiv\_Pressefotos\Pressefotos 2023\) - Filter: Pfad ist ungleich	"P:\Besta	and F9 (Fo	toalben)" UNI
3 La	aufwerk: O:\ Größe: 2,9 TB Belegt: 900,3 GB	Frei: 2,0 TB			
4					
5 N	lame	Pfad	Größe	Dateien	Verzeichnisse
5 0	:\ Pressefotos\Pressefotos 2023\	O:\ Pressefotos\Pressefotos 2023\	40,1 GB	4927	269
7 20	0230710 Naturdenkmäler, Holger Meermann	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	3,4 GB	259	C
D	SC 1496.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	20,4 MB	1	(
D	SC 1497.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	20,0 MB	1	(
D	SC 1498.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	19,4 MB	1	
1 D	SC 1511.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	19,4 MB	1	(
2 D	SC 1499.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	18,9 MB	1	(
3 D	SC 1534.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	18,8 MB	1	C
1 D	SC 1510.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	18,8 MB	1	C
5 D	SC 1538.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	18,7 MB	1	C
5 D	SC 1529.JPG	O:\ Pressefotos\Pressefotos 2023\20230710 Naturdenkmäler, Holger	18,4 MB	1	C





#### Bearbeitungsschritte vor der Bewertung

- 1) Erzeugung Excel-Tabellen mit Ordnerstruktur des Laufwerks inkl. Dateipfad und Hashwerten → PowerShell oder TreeSize Professional
- 2) Abgleich der Hashwerte mit bereits im Archiv befindlichen Dateien aus vorherigen Übernahmen (optional)
- 3) Speicherung der Pressemitteilungen, die später als Bewertungsgrundlage herangezogen werden
- 4) Identifizierung von mehrfach vorhandenen Dateien und Auswahl nur eines Ordners/Speicherorts zur Bewertung
- 5) Kennzeichnung mit "K" bestimmter Dateigruppen in der Tabelle (z.B. Symbolfotos, Systemdateien, nicht archivfähige Dateiformate etc.)
- 6) Aufteilung in eine Tabelle pro Ordner auf dem Laufwerk zur Erleichterung der späteren Übernahme





Bewertungsent scheidung  A=Archivwürdig B=Bewertung K=Kassieren	Link  1. Gewünschte Zeile ansteuern/markieren 2. Strg + C 3. Windows Explorer öffnen 4. mit Strg + V aus der Zwischenablage in die Adresszeile des Explorers einfügen 5. Enter	Bemerkung •	Hash
K	O \_Pressefotos\Pressefotos 2023\20230102 Hochsitz Jagd Symbolfoto\20230102-PC290273-Hochsitz Jagd-mk JPG		ADDFC098B9FC8FDC5B5D841C0CD1D3DE81F32
×	On Presselotos Presselotos 2023/20230102 Hochsitz Jagd Symbolfoto PC290257 JPG		84AEBB4D97A25509D75A1F0C9030240D78863DI
K	O\_Pressefotos\Pressefotos 2023/20230102 Hochsitz Jagd Symbolfeto\PC290260 JPG		E910A447587AAA71EAE736BA4C14B2B372A195
K	O \_Pressefotos\Pressefotos 2023/20230102 Hochsitz Jagd Symbolfoto\PC290273 JPC		E0BF98FBCC63CFE0977881DA30086545D0931E
K	O \_Pressefotos\Pressefotos 2023\20230102 Windsack Symbolfoto\20221229_120332 jpg		5D07CAF697DD337B5F58D0DB86EAA18AE3378I
× ×	O \ Pressefotos\Pressefotos 2023\20230102 Windsack Symbolfoto\20221229_120340 windsack-mk.jpg		99E7FE4CAEA89E3A8FFD288114A5B9E14027F0
K	O:\_Pressefotos\Pressefotos 2023\20230102 Windsack Symbolfoto\20221229_120340.jpg		97F9FAEDA8B6C697357BB3EA314AD5B37616E.
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\20230111-1860-mk-Bildungsrat Ki	Pressemitteilung	4FD3093E6B89860AF49E7066AD1DD3124ADF36
	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\20231101-Bildungsregion-DSC_18		6F5835019EB0F736DCA01169EF343CECBFEA35
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\20231101-Bildungsregion-DSC_18	Pressemitteilung	563D10EB4C19DE5BDB675CA35A768F58A2DED
K	O \ Pressetotos\Pressetotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\debug log	Pressemitteilung	201CB1272D06AC96DF9AB8C400D84524DB18B6
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\DSC_1857.JPG	Pressemitteilung	D04AD86172C195AC35926BE61EB7B614534F025
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\DSC_1858.JPG	Pressemitteilung	9967E9EC68112B02E3EF142BA8A75E21612F9F8
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\DSC_1859.JPG	Pressemitteilung	96AA363A3F0FB07F58EA2109F5B279BBCB89F5
В	O:\_Pressefotos\Pressefotos 2023\20230111 Treffen Bildungsrat der Bildungsregion\DSC_1860.JPG	Pressemitteilung	A2A847FEBA9A4F83565DAA2BD4065E075BA10E
			Ansicht Bewertungstabelle





#### Bearbeitungsschritte zur Bewertung (ausstehend)

- 1) Bewertung Pressemeldungen und entsprechende Kennzeichnung der dazugehörigen Ordner/Dateien in der Tabelle
- 2) Einzelbewertung der mit "B" gekennzeichneten Ordner/Dateien
- 3) Einzelbewertung der Ordner/Dateien ohne Bezug zu Pressemitteilungen





# Bearbeitungsschritte zur Übernahme (ausstehend)

- 1) Erstellung einer Kopie der archivwürdigen Datei(en) in einem Verzeichnis des Archivs und Formierung zu Archivpaketen
- 2) Protokollierung des Kopiervorgangs anhand von Hashwerten zur Wahrung der Authentizität
- 3) Ingest DiPS.kommunal



#### 3. Fazit



- Projekt sehr zeitaufwendig und arbeitsintensiv, aber mit technischen Bordmitteln und fachlicher Einarbeitung machbar
- Erfahrungen aus dem Projekt sehr gut nachnutzbar, auch für andere Dateisammlungen
- Entwicklung einer Routine bzw. eines Workflows zur Bewertung und Übernahme von Dateisammlungen
- Nebeneffekt: Die Pressestelle hat ihre eigene Ordnerstruktur kritisch hinterfragt und verschlankt → Erleichterung zukünftiger Übernahmen



#### 4. Quellen und Linksammlung



#### Theorie

- Dr. Daniel Wilhelm: Umgang mit Dateisammlungen Status quo in der deutschsprachigen Fachliteratur und Handreichung zur Bearbeitung einer Dateisammlung
- Sonderveröffentlichungen der Staatlichen Archive Bayerns: Kreative digitale Ablagen und die Archive Ergebnisse eines Workshops des KLA-Ausschusses Digitale Archive am 22./23. November 2016 in der Generaldirektion der Staatlichen Archive Bayerns
- Landesarchiv NRW: Zum Umgang mit Dateisystemen Bewertung, Übernahme, Erschließung, Nutzung
- Musterworkflow zur Bearbeitung von Dateisammlungen für DIMAG-Anwender
- Karina Jaeger, Maria Kobold: Zwischen Datenwust und arbeitsökonomischer Bewertung Ein Werkstattbericht zum Umgang mit unstrukturierten Dateisammlungen am Beispiel des Bestandes der Odenwaldschule [In: Archivar 70/3 (2017), S. 307-311.]
- Michael Unger, Markus Schmalzl: Digitales Verwaltungshandeln nachvollziehbar archivieren oder: Was ist die (e)Akte? [In: Archivar 73/4 (2020), S. 371-378.]
- KOST-Projekt 20-039 Dateiablage
- Marco Birn, Kai Naumann: Bewertung schwach strukturierter Unterlagen. Berichte und Thesen aus Baden-Württemberg [In: Brandenburgische Archive: Berichte und Mitteilungen aus den Archiven des Landes Brandenburg 36/2019, S. 8-14.]
- Dr. Daniel Baumann: Ein Werkstattbericht: Die Übernahme unstrukturierter Dateisammlungen im Stadtarchiv München
- Bastian Gillner: Abfragen statt Anbieten Eine alternative Praxis im archivischen Umgang mit Dateisystemen [In: Archiv theorie & praxis 76/4 (2023), S. 317-321.]
- Gunnar Wendt, Sina Westphal: Eine Herausforderung der Übergangs Fileablagen als Quellen der digitalen Überlieferungsbildung. In: Transformation ins Digitale – 85. Deutscher Archivtag 2015 in Karlsruhe. S. 105-113.



#### 4. Quellen und Linksammlung



#### **Technik**

- Stephan Lenartz: Aufbereitung und Bewertung von Fileablagen mit Python am Beispiel einer digitalen Fotosammlung
- KOST-Wiki: Tools
- Community Owned digital Preservation Tool Registry (COPTR)
- nestor Toolbox Werkzeugübersicht
- nestor Toolbox Literaturhinweise
- Get-ChildItem Befehl für PowerShell zum Auslesen von Ordnerstrukturen.
  - <a href="https://learn.microsoft.com/en-us/powershell/module/microsoft.powershell.management/get-childitem?view=powershell-7.4">https://learn.microsoft.com/en-us/powershell/module/microsoft.powershell.management/get-childitem?view=powershell-7.4</a>
  - https://shellgeek.com/powershell-get-childitem/
- Übersicht gängiger Dateiformate



# 4. Quellen und Linksammlung



#### Kontakt

Kreisarchiv Soest

Kira Knappkötter

Niederbergheimer Straße 24

59494 Soest

Telefon: 02921 303895

Mail: kira.knappkoetter@kreis-soest.de

