

CSV2SIARD – Toolunterstützte Übernahme von Textdatei-basierten Datenbeständen der ehem. DDR

*17. Tagung des Arbeitskreises "Archivierung von
Unterlagen aus digitalen Systemen", März 2013, Dresden*



Ausgangslage „DDR-Altdateien“ im Bundesarchiv:

- Mit der Wiedervereinigung wurden auf zentraler Ebene der DDR zahlreiche Behörden und Rechenzentren geschlossen
- 1992, 1995 und 1997 erfolgten größere Datenübernahmen, ab 1998 erfolgten grundsätzl. archivischen Bewertung, umfangreiche Kassationen von Magnetbändern und Umkopierung
- Die zu archivierenden Daten wurden im Bundesarchiv grundsätzlich als Flat-Files vom **EBCDIC**-Format in das softwareunabhängigere Standard-Zeichencode **ASCII** konvertiert und gespeichert
- Um den Import in Anwendungsprogramme (Datenbankmanagementsysteme MS Access, Oracle, Informix oder Statistiksoftware wie SPSS, SAS oder TDA) zu erleichtern, wurden die Daten zusätzlich als „comma separated values“ (**csv**) konvertiert und gespeichert.

Problemstellung:

- Keine archivierbare Datenbank mehr (im engeren Sinne)
- Ständiges Umkopieren in den letzten Jahren auf neueste Speichertechnik (Bänder → CD → DAT → FP)
- Zwar weitgehend unabhängiges Datenformat, aber kein Archivierungsformat
- „Archivobjekte“ nicht im zusammengeführt (Datenfiles, Codelisten ...)
- „Nur“ Metadaten-Recherche über archivfachliche Anwendung BASYS
- Speicherung nicht OAIS - konform

Bestände:

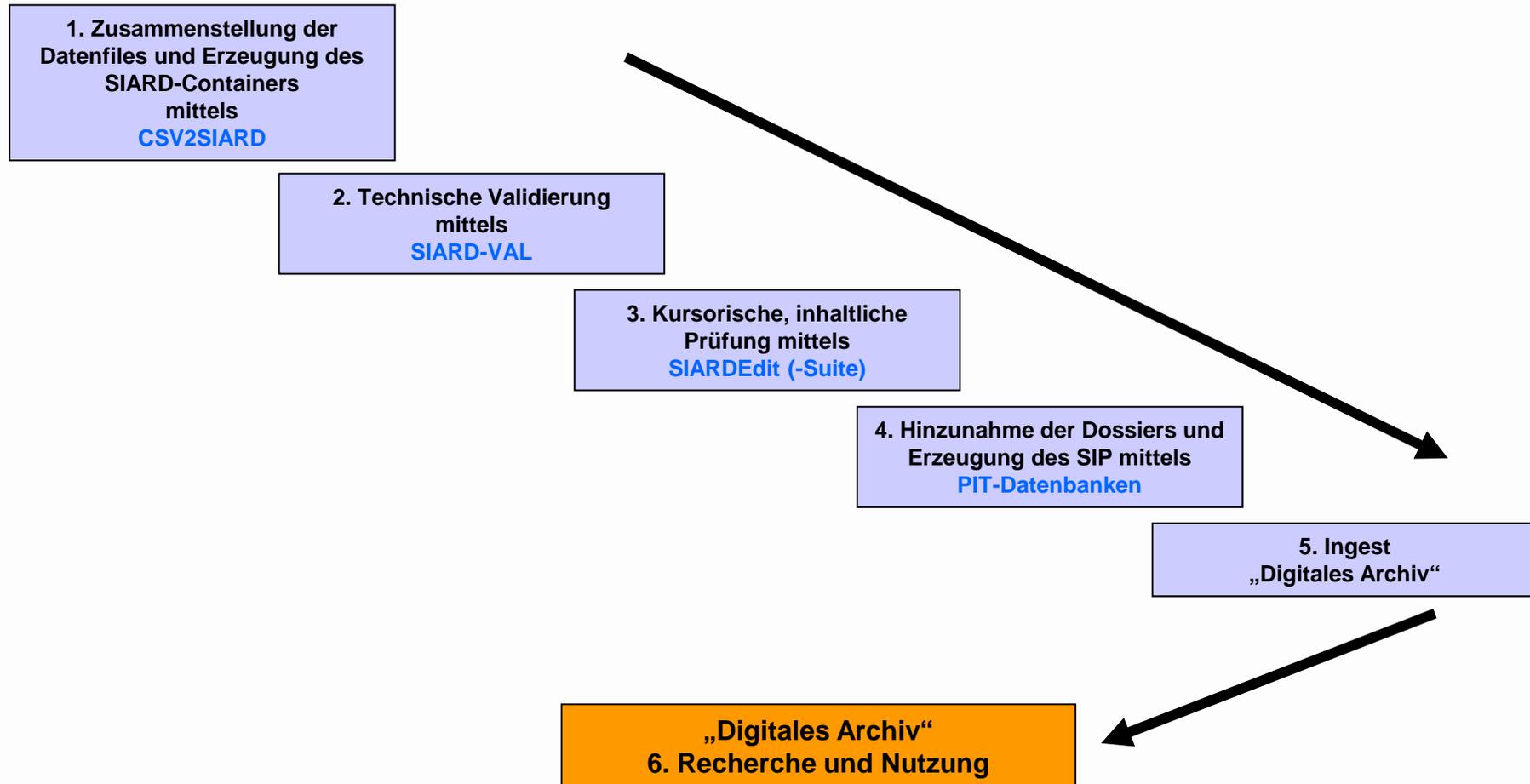
Ca. 182 DDR-Datenbestände, von

- „**Staatsrat der DDR – Bürgereingaben**“ (Sign. DA 1 MD 1) über
 - „**STUBEW – Studienbewerbungen, -zulassungen, -ablehnungen**“ (Sign. DR 306 MD 1) bis zu
 - „**NIQ – Zentraler Täterindex**“ (Sign. DO 1 MD 18)
- Testfall für csv2siard → „**NMA – Bahnbetriebsunfallstatistik**“ (Sign. DO 1 MD 9)



Ausgangslage aus: Ulf Rathje, BArch, für Magazin Historical Social Research, Vol. 28 — 2003 — No. 1/2, 57 – 72 und Bild-Sign.: Bundesarchiv, Bild_183-Z0612-027

Ablauf der Archivierung



1. Zusammenstellung der Datenfiles / Erzeugung SIARD-Container mittels CSV2SIARD

```
C:\ Eingabeaufforderung

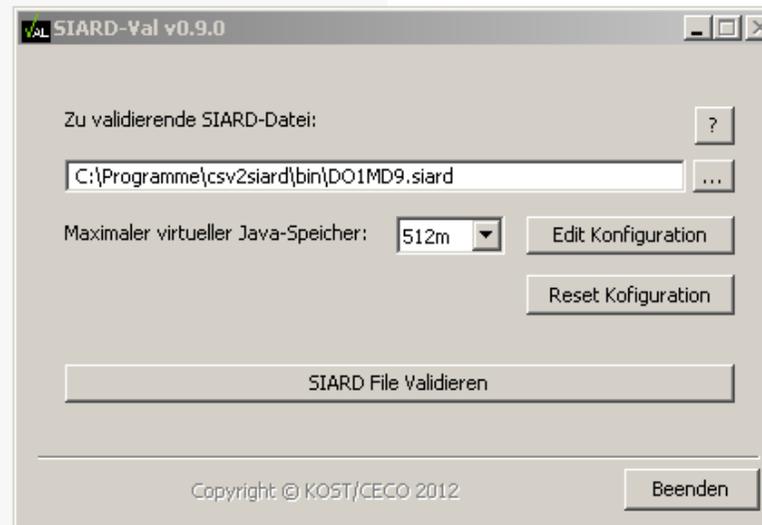
C:\Programme\csv2siard\bin>
C:\Programme\csv2siard\bin>
C:\Programme\csv2siard\bin>csv2siard.exe :NO_DB_MODEL c:\programme\csv2siard\csvdata D01MD9.siard

csv2siard v 1.8.2, Copyright (C) 2011 Martin Kaiser (KOST-CECO)
This program comes with ABSOLUTELY NO WARRANTY.
This is free software, and you are welcome to redistribute it under certain conditions;
see GPL-2.0_COPYING.txt for details.

Preferences:
[ERR] => 0
[DB_MODEL] => NO_DB_MODEL
[CSU_FOLDER] => c:/programme/csv2siard/csvdata
[SIARD_FILE] => C:/Programme/csv2siard/bin/D01MD9.siard
[DELIMITED] => !
[QUOTE] => "
```

```
addFolder: content/schema0/table7/
addFile: content/schema0/table8/table8.xml/ >crc32 >stuff
addFile: content/schema0/table8/table8.xsd/ >crc32 >stuff
addFolder: content/schema0/table8/
addFile: content/schema0/table9/table9.xml/ >crc32 >stuff
addFile: content/schema0/table9/table9.xsd/ >crc32 >stuff
addFolder: content/schema0/table9/
addFolder: content/schema0/
addFolder: content/
addFolder: header/
addFile: header/metadata.xsd/ >crc32 >stuff
addFile: header/metadata.xml/ >crc32 >stuff
SIARD file created: C:/Programme/csv2siard/bin/D01MD9.siard
Conversion completed
```

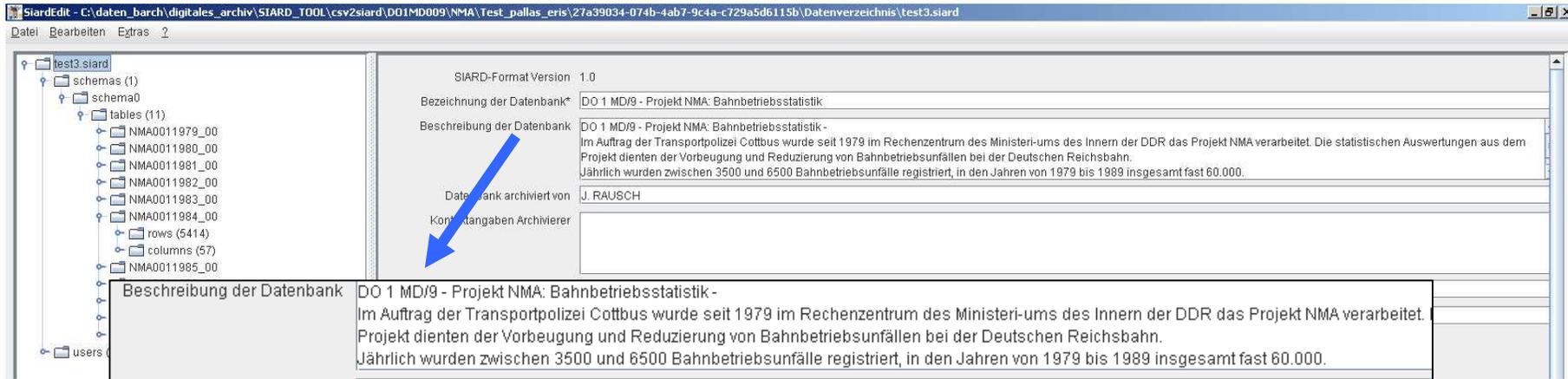
2. Technische Validierung mittels SIARD-VAL



The screenshot shows a Java command prompt window titled "C:\Programme\Java\jre7\bin\java.exe". The output text is as follows:

```
SIARD-Val v0.9.0, Copyright (C) 2012 Claire Roethlisberger (KOST-CECO),  
Christian Eugster, Olivier Debenath, Peter Schneider (Staatsarchiv Aargau)  
This program comes with ABSOLUTELY NO WARRANTY.  
This is free software, and you are welcome to redistribute it under  
certain conditions; see GPL-3.0_COPYING.txt for details.  
  
SIARD Validierung -> DO1MD9.siard  
  
[A] = Valid  
[B] = Valid  
[C] = Valid  
[D] = Valid
```

3. Inhaltliche Prüfung mittels SIARDEdit (-Suite)



SIARDEdit - C:\daten_barch\digitales_archiv\SIARD_TOOL\csv2siard\DO1MD009\NMA\Test_pallas_eri\27a39034-074b-4ab7-9c4a-c729a5d6115b\Datenverzeichnis\test3.siard

test3.siard

- schemas (1)
 - schema0
 - tables (11)
 - NMA0011979_00
 - NMA0011980_00
 - NMA0011981_00
 - NMA0011982_00
 - NMA0011983_00
 - NMA0011984_00
 - rows (5414)
 - columns (57)
 - NMA0011985_00

SIARDEdit - C:\daten_barch\digitales_archiv\SIARD_TOOL\csv2siard\DO1MD009\NMA\Test_pallas_eri\27a39034-074b-4ab7-9c4a-c729a5d6115b\Datenverzeichnis\test3.siard

SIARD-Format Version 1.0

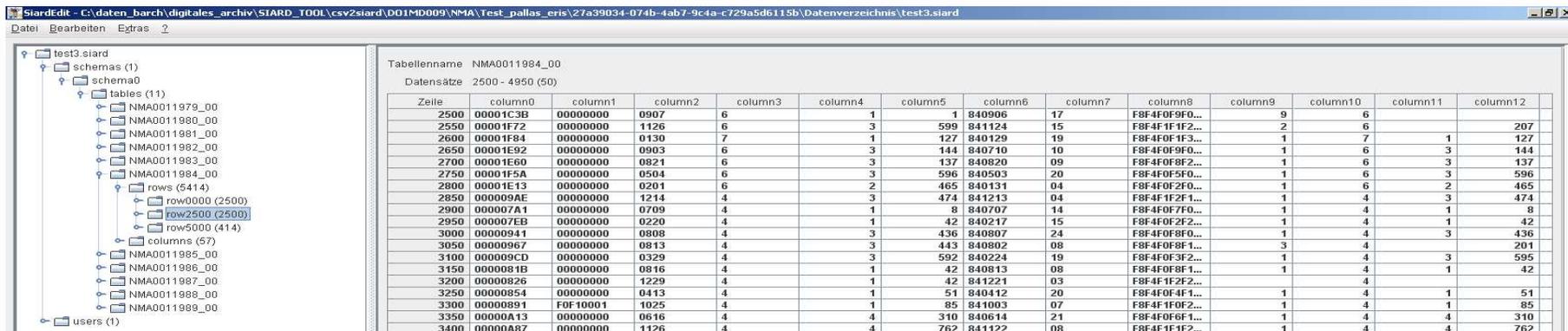
Bezeichnung der Datenbank* DO 1 MD/9 - Projekt NMA: Bahnbetriebsstatistik

Beschreibung der Datenbank DO 1 MD/9 - Projekt NMA: Bahnbetriebsstatistik -
Im Auftrag der Transportpolizei Cottbus wurde seit 1979 im Rechenzentrum des Ministeriums des Innern der DDR das Projekt NMA verarbeitet. Die statistischen Auswertungen aus dem Projekt dienen der Vorbeugung und Reduzierung von Bahnbetriebsunfällen bei der Deutschen Reichsbahn.
Jährlich wurden zwischen 3500 und 6500 Bahnbetriebsunfälle registriert, in den Jahren von 1979 bis 1989 insgesamt fast 60.000.

Datenbank archiviert von J. RAUSCH

Kontaktdaten Archivierer

Beschreibung der Datenbank DO 1 MD/9 - Projekt NMA: Bahnbetriebsstatistik -
Im Auftrag der Transportpolizei Cottbus wurde seit 1979 im Rechenzentrum des Ministeriums des Innern der DDR das Projekt NMA verarbeitet. Die statistischen Auswertungen aus dem Projekt dienen der Vorbeugung und Reduzierung von Bahnbetriebsunfällen bei der Deutschen Reichsbahn.
Jährlich wurden zwischen 3500 und 6500 Bahnbetriebsunfälle registriert, in den Jahren von 1979 bis 1989 insgesamt fast 60.000.



SIARDEdit - C:\daten_barch\digitales_archiv\SIARD_TOOL\csv2siard\DO1MD009\NMA\Test_pallas_eri\27a39034-074b-4ab7-9c4a-c729a5d6115b\Datenverzeichnis\test3.siard

test3.siard

- schemas (1)
 - schema0
 - tables (11)
 - NMA0011979_00
 - NMA0011980_00
 - NMA0011981_00
 - NMA0011982_00
 - NMA0011983_00
 - NMA0011984_00
 - rows (5414)
 - row0000 (2500)
 - row2500 (2500)
 - row5000 (414)
 - columns (57)
 - NMA0011985_00

Tabellenname NMA0011984_00

Datensätze 2500 - 4950 (50)

Zeile	column0	column1	column2	column3	column4	column5	column6	column7	column8	column9	column10	column11	column12
2500	00001C3B	00000000	0907	6	1	1	840906	17	F8F4F0F9F0...	9	6		
2550	00001F72	00000000	1126	6	3	599	841124	15	F8F4F1F1F2...	2	6		207
2600	00001F84	00000000	0130	7	1	127	840129	19	F8F4F0F1F3...	1	7	1	127
2650	00001E92	00000000	0903	6	3	144	840710	10	F8F4F0F9F0...	1	6	3	144
2700	00001E60	00000000	0821	6	3	137	840820	09	F8F4F0F8F2...	1	6	3	137
2750	00001F5A	00000000	0504	6	3	596	840503	20	F8F4F0F5F0...	1	6	3	596
2800	00001E13	00000000	0201	6	2	465	840131	04	F8F4F0F2F0...	1	6	2	465
2850	000009AE	00000000	1214	4	3	474	841213	04	F8F4F1F2F1...	1	4	3	474
2900	000007A1	00000000	0709	4	1	8	840707	14	F8F4F0F7F0...	1	4	1	8
2950	000007EB	00000000	0220	4	1	42	840217	15	F8F4F0F2F2...	1	4	1	42
3000	00000941	00000000	0808	4	3	436	840807	24	F8F4F0F9F0...	1	4	3	436
3050	00000967	00000000	0813	4	3	443	840802	08	F8F4F0F8F1...	3	4	3	201
3100	000009CD	00000000	0329	4	3	592	840224	19	F8F4F0F3F2...	1	4	3	595
3150	0000081B	00000000	0816	4	1	42	840813	08	F8F4F0F8F1...	1	4	1	42
3200	00000826	00000000	1229	4	1	42	841221	03	F8F4F1F2F2...	1	4		
3250	00000854	00000000	0413	4	1	51	840412	20	F8F4F0F4F1...	1	4	1	51
3300	00000891	F0F10001	1025	4	1	85	841003	07	F8F4F1F0F2...	1	4	1	85
3350	00000813	00000000	0616	4	4	310	840614	21	F8F4F0F6F1...	1	4	4	310
3400	00000A87	00000000	1126	4	4	762	841122	08	F8F4F1F1F2...	1	4	4	762

4. Erzeugung eines SIP mittels PIT-DB

PIT- Bundesarchiv (Pre Ingest Tool)

Datei Aktionen Exportieren Hilfe

Lieferung

- Datenbank AIP
 - AbschlussvermerkDO1MD009.pdf
 - Codeliste_DO1MD009.pdf
 - DO1MD009.pdf
 - DO1MD009dossier.pdf
 - DO1MD009txt.pdf
 - DO1MD9.siard

Lieferungsmetadaten

Daten: Transaction-Envelope

Daten zur abgebenden Behörde:

Transaktions-ID:

Datenbanktitel:

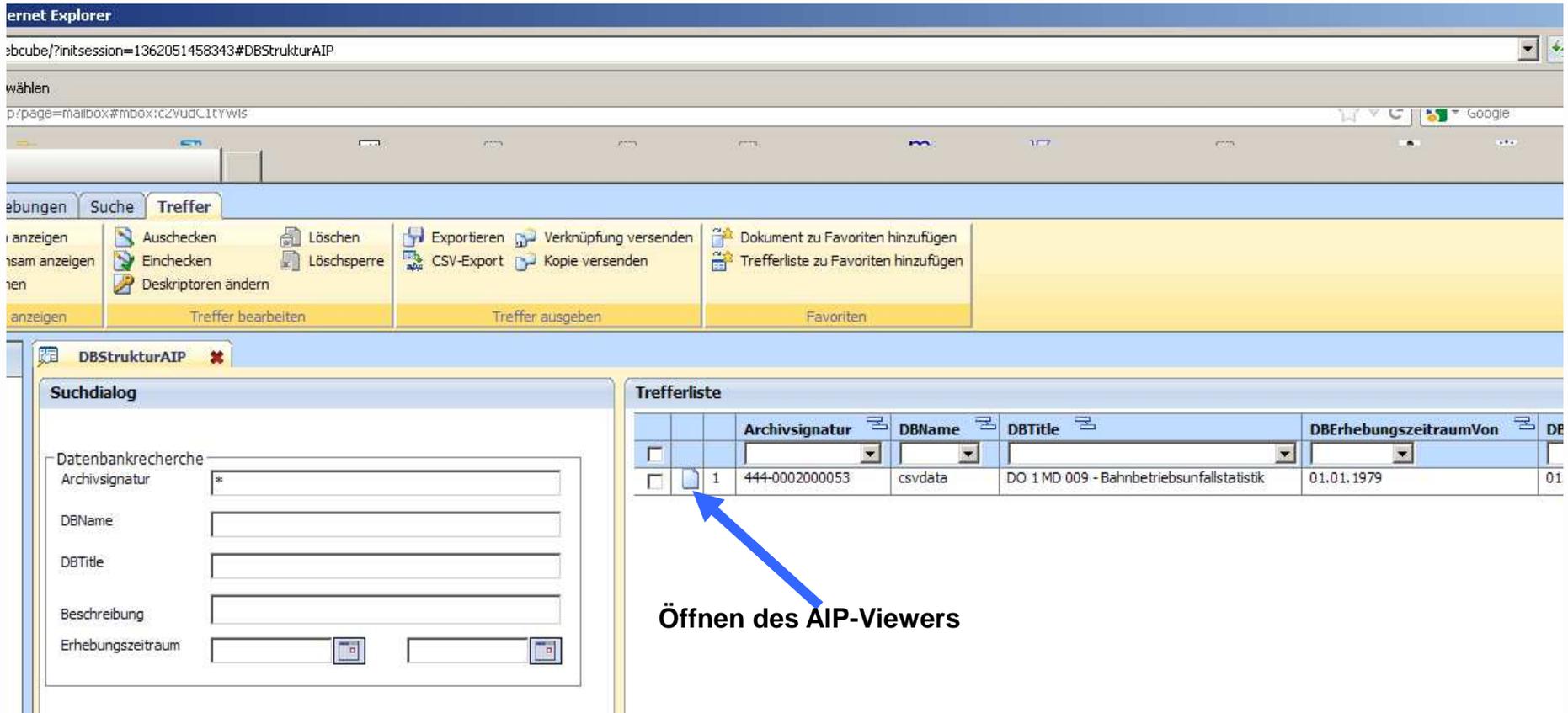
Erhebungszeitraum von:

Fristvorgabe personenbezogene Daten:

Erhebungszeitraum bis:

SIARD-Container **Dokumente / Dossiers etc.**

6.1 Recherche und Nutzung



The screenshot shows the KOST web interface. On the left, a 'Suchdialog' (Search dialog) is open, showing search criteria: 'Archivsignatur' with a wildcard '*', 'DBName', 'DBTitle', 'Beschreibung', and 'Erhebungszeitraum'. On the right, a 'Trefferliste' (Results list) table is displayed. A blue arrow points to the first row of the table, which contains the following data:

		Archivsignatur	DBName	DBTitle	DBErhebungszeitraumVon	DBErhebungszeitraumBis
<input type="checkbox"/>	1	444-0002000053	csvdata	DO 1 MD 009 - Bahnbetriebsunfallstatistik	01.01.1979	01.01.2000

Öffnen des AIP-Viewers

6.2. Recherche und Nutzung

Ministerium des Innern
Zentralstelle für Datenverarbeitung

Projekt NMA:
Bahnbetriebsunfallstatistik

DO 1 MD/9

Dossier

Marion Teichmann

Koblenz
März 2008

Bytes	
02-1d85e8ba5890	AbschlussvermerkDO 1MD0
81-631fb3600f4a	Codeliste_DO 1MD009.pdf
de-5cfbebcd8ddf	DO 1MD009.pdf
41596663	Export

DBImport AIPViewer - Windows Internet Explorer

XBarch Metadaten

Export started: /barch/ser/itaflow/data/SiardExport/export/DO1MD9.siard

Vielen Dank für Ihre Aufmerksamkeit !