

KOST Koordinationsstelle für die dauerhafte Archivierung
elektronischer Unterlagen

Ein Gemeinschaftsunternehmen von Schweizer Archiven

csv  **siard**

csv2siard Version 1.7



csv2siard erlaubt die Konvertierung von CSV-Datei Sammlungen in eine SIARD-Datei .

Die Vorteile einer solchen Konvertierung:

1. einzelne CSV-Dateien aus der gleichen Datenbank werden in einer Datei zusammengefasst;
2. CSV-Dateien werden in ein standardisiertes Format gebracht und somit bezüglich Zeichensatz, Datentrennzeichen, Zeilenstruktur etc. vereinheitlicht.
3. Mit ***SiardEdit*** steht ein Viewer für SIARD-Dateien zur Verfügung.
4. zur Datenanalyse ist ein Export in eine relationale Datenbank problemlos möglich.

csv2siard ein Kommandozeilen Tool

csv2siard ist ein einfaches Kommandozeilen-Tool, das CSV-Dateien in Tabellen in einer SIARD-Datei umwandelt. Jede CSV-Datei wird zu einer Tabelle.

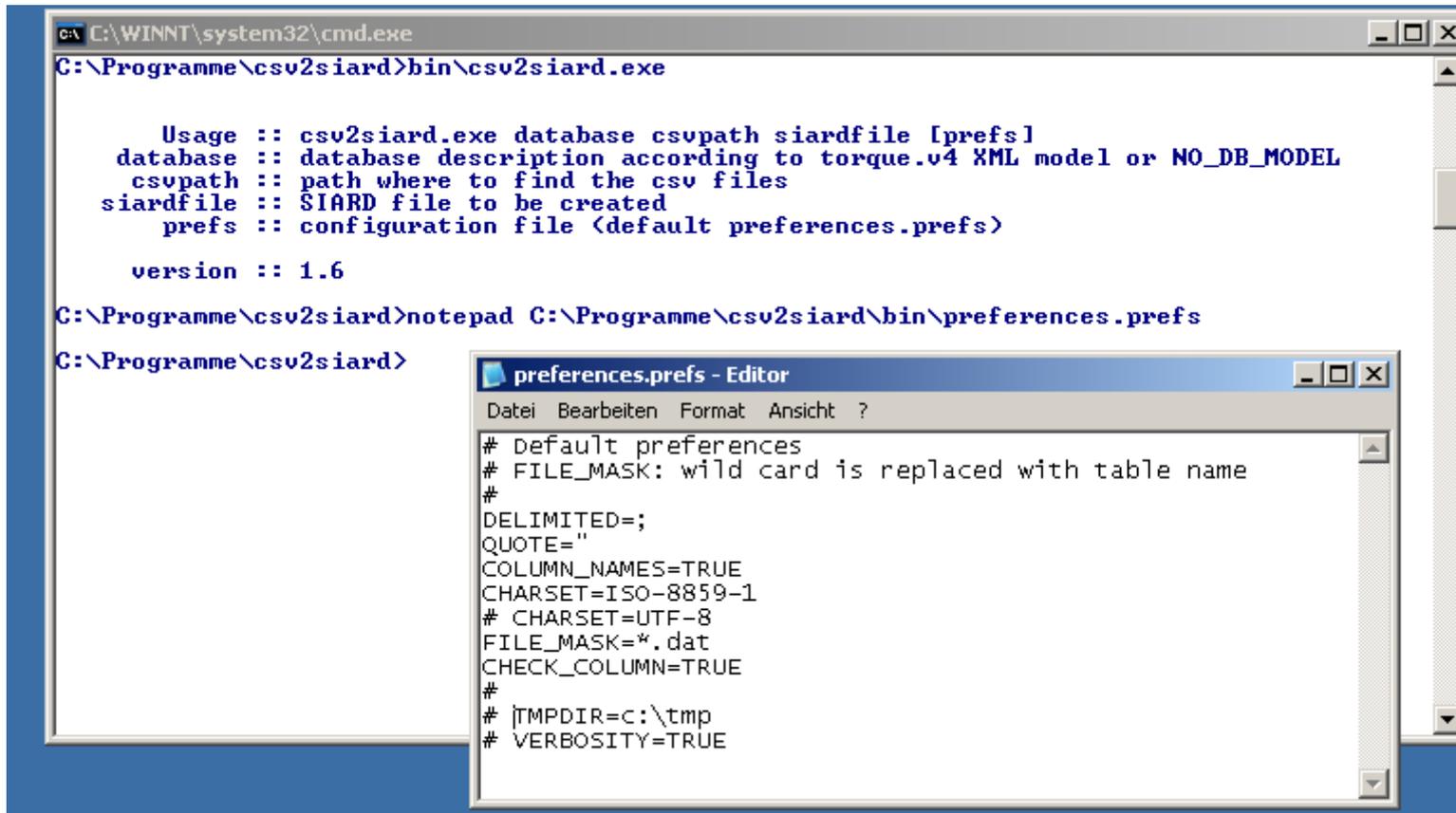
Da bei CSV-Dateien keine Strukturinformationen im eigentlichen Sinne zur Verfügung stehen, generiert das Tool eine einfache Tabellenbeschreibung mit Feldnamen und Feldattribut .

Diese Tabellenbeschreibung (XML Datenschema) kann wenn nötig in einem zweiten Schritt Manuell ergänzt, korrigiert und angepasst werden.

Beziehungen zwischen den Tabellen und Einschränkungen (*Constraints*) können in diesem Schema definiert werden.

csv2siard „hands on“

Wichtig sind vor allem sechs Einstellungen in der Konfigurationsdatei:



```
C:\WINNT\system32\cmd.exe
C:\Programme\csv2siard>bin\csv2siard.exe

Usage :: csv2siard.exe database csvpath siardfile [prefs]
database :: database description according to torque.v4 XML model or NO_DB_MODEL
csvpath  :: path where to find the csv files
siardfile :: SIARD file to be created
prefs   :: configuration file (default preferences.prefs)

version :: 1.6

C:\Programme\csv2siard>notepad C:\Programme\csv2siard\bin\preferences.prefs
C:\Programme\csv2siard>
```

```
preferences.prefs - Editor
Datei Bearbeiten Format Ansicht ?
# Default preferences
# FILE_MASK: wild card is replaced with table name
#
DELIMITED=;
QUOTE="
COLUMN_NAMES=TRUE
CHARSET=ISO-8859-1
# CHARSET=UTF-8
FILE_MASK=*.dat
CHECK_COLUMN=TRUE
#
# TMPDIR=c:\tmp
# VERBOSITY=TRUE
```

csv2siard „und los gehts“

Mit den beigelegten anonymisierten Testdaten aus dem Projekt „Archivierung von Gebäudeversicherungsdaten“ kann sofort losgelegt werden:

The image shows a Windows command prompt window on the left and the SiardEdit software interface on the right. The command prompt shows the execution of the csv2siard program, which has successfully converted a SIARD file. The SiardEdit interface displays a tree view of the database structure and a detailed view of the 'gv_gebaeude' table.

```
C:\WINNT\system32\cmd.exe

C:\Programme\csv2siard>bin\csv2siard.exe

csv2siard v 1.6, Copyright (C) 2011 Mar
This program comes with ABSOLUTELY NO W
This is free software, and you are welc
see GPL-2.0_COPYING.txt for details.

Process table (encoding: iso-8859-1) gv
Process table (encoding: us-ascii) gv_p
Process table (encoding: us-ascii) gv_a
Process table (encoding: us-ascii) gv_s
Process table (encoding: us-ascii) gv_s
Process table (encoding: us-ascii) gv_p
ZIP SIARD file .....
SIARD file created: C:/Programme/csv2si
Conversion completed

C:\Programme\csv2siard>
```

SiardEdit - P:\KOST\Pilotloesungen\Gebaeudeversicherung\4_GV-Viewer\gvtg.siard

File Edit Tools ?

gvtg.siard

- schemas (1)
 - schema0
 - tables (6)
 - gv_anlage
 - rows (3)
 - columns (4)
 - foreign keys (1)
 - gv_gebaeude
 - rows (9)
 - columns (24)
 - primary key
 - gv_person
 - rows (13)
 - columns (18)
 - foreign keys (1)
 - gv_position
 - rows (12)
 - columns (14)
 - foreign keys (1)
 - gv_schaden
 - rows (2)
 - columns (16)
 - foreign keys (1)
 - gv_schaetzung
 - rows (9)
 - columns (19)
 - primary key
 - foreign keys (1)
- users (1)

Table name gv_gebaeude
Table folder table0
Number of data records 9
Table description Umfasst die Basisdaten zu versicherten und fremdversicherten Gebäuden sowie zu abgebrochenen Gebäuden mit einem Schadenfall in der Beobachtungsperiode

Column position	Column name	SQL column type
1	id	CHARACTER VARYING(16)
2	grundstueck_id	CHARACTER VARYING(20)
3	police_nr	CHARACTER VARYING(20)
4	suchbegriff	CHARACTER VARYING(100)
5	status_code	CHARACTER VARYING(10)
6	status_text	CHARACTER VARYING(255)
7	zweck_code	CHARACTER VARYING(10)
8	zweck_text	CHARACTER VARYING(255)
9	baujahr	INTEGER
10	bauart_code	CHARACTER VARYING(10)
11	bauart_text	CHARACTER VARYING(255)
12	lage_code	CHARACTER VARYING(10)
13	lage_text	CHARACTER VARYING(255)
14	kanton	CHARACTER VARYING(2)
15	bezirk	CHARACTER VARYING(100)
16	gemeinde_bfs	CHARACTER VARYING(20)
17	gemeinde	CHARACTER VARYING(255)
18	strasse	CHARACTER VARYING(255)
19	strasse_nr	CHARACTER VARYING(255)
20	wohnung_nr	CHARACTER VARYING(20)
21	plz	CHARACTER VARYING(6)
22	plz_zusatz	CHARACTER VARYING(20)
23	ort	CHARACTER VARYING(255)
24	ort_zusatz	CHARACTER VARYING(20)

SIARD Archive P:\KOST\Pilotloesungen\Gebaeudeversicherung\4_GV-Viewer\gvtg.siard

csv2siard die wichtigsten Features

- Generiert ein *apache Torque.v4* Datenbankschema aus einer Sammlung von CSV Dateien
- Erkennt *Integer, Decimal, Date* und *Varchar* und die zugehörige Feldgrösse
- Implementiert viele DBSM Features des Torque Datenbankschema
- Bei vorgegebenen Schema wird die CSV Datenstruktur beim Import geprüft.

Etliche neue Features in v.1.7:

- Viele gängige *Date* Formate werden erkannt, spezielle Formate können definiert werden
- Binärdaten können ebenfalls umgewandelt werden