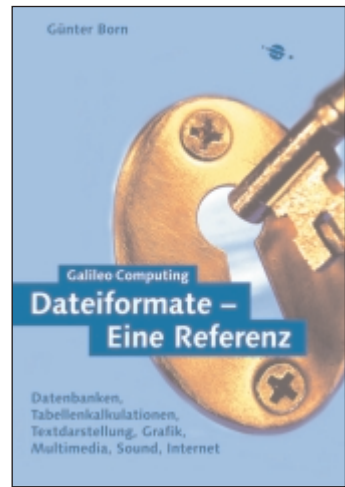


Dieses Kapitel stammt aus dem Buch
›Dateiformate – Eine Referenz‹
von Günter Born.

www.borncity.de

ISBN 3-934358-83-7
119,90 DM



Informationen zum Buch
mit Bestellmöglichkeit

www.galileocomputing.de

Galileo Computing

7 Super Data Interchange-Format (SDI)

Die Übertragung von Daten in DIF-Dateien bringt einige Einschränkungen bezüglich der übertragenen Informationen mit sich. So lassen sich Formelinhalte einzelner Zellen oder das jeweilige Format nicht im DIF-Format ablegen. Die Firma Computer Associates International (CA) erweiterte die DIF-Definition für das Produkt SuperCalc 3. Das Resultat wird als Super Data Interchange-Format (SDI) bezeichnet. Eine SDI-Datei lehnt sich im Aufbau an die DIF-Dateien an und besteht wieder aus einem Header, gefolgt vom Datenteil. Auch die Schlüsselwörter der DIF-Definition wurden übernommen und durch eigene Definitionen erweitert. Spalten bezeichnet man in SDI als *Vector* und Zeilen als *Tuple*. In SuperCalc (Version 2.0) werden die Zeilen von 1 bis 9999 durchnummeriert, während die Spalten von 1 bis 128 reichen. Diese Numerierung ist im SDI-Format abgebildet.

Die Sätze in Header und Datenbereich besitzen eine variable Länge, so daß jede Zeile einer SDI-Datei mit CR/LF (*Carriage Return* bzw. *Line Feed*) abschließt. Zur Definition des Zeichen- und Spaltenformats werden ASCII-Zeichen ebenfalls mit übertragen. Tabelle 7.1 enthält die Formatzeichen für solche Dateien:

Zeichen	Bedeutung
L	numerische Werte linksbündig justiert
R	numerische Werte rechtsbündig justiert
TL	Texte linksbündig justiert
TR	Texte rechtsbündig justiert
\$	numerische Werte mit zwei Dezimalstellen und eventuell anhängenden Nullen formatieren
*	numerische Werte als * darstellen, 3 wird dann als *** dargestellt, während 0 eine Leerstelle ergibt
I	numerische Werte als Integer ausgeben
G	Standardformat mit bestmöglicher Anpassung an die Breite einer Zelle
D	entferne alle Formate und ersetze diese durch das Standardformat
E	Exponentendarstellung

Tabelle 7.1 Steuerzeichen im SDI-Format

Der Header einer SDI-Datei

Der SDI-Header besteht aus minimal zwei Einträgen und verschiedenen optionalen Informationen. Jeder Eintrag im Header besitzt folgendes Format:

```
<Topic>
<Vektor>,<Wert>
"<String>"
```

Die Felder wurden hier mit den Zeichen <> markiert und besitzen eine besondere Bedeutung.

Topic enthält das Schlüsselwort zur Spezifizierung des Satztyps. Bei diesen Schlüsselwörtern handelt es sich um Namen mit einer maximalen Länge von 32 Großbuchstaben. Der Kopf einer SDI-Datei muß mindestens die beiden Records *Table* und *Data* enthalten. Außerdem wurden weitere fünf optionale Recordtypen definiert. Hierzu gehören – im Gegensatz zu den Satztypen bei .DIF-Dateien – auch *Vectors* und *Tuples*. Hier eine Aufstellung der möglichen Recordtypen:

- ▶ TABLE
- ▶ VECTORS
- ▶ TUPLES
- ▶ GDISP-FORMAT
- ▶ COL-FORMAT
- ▶ ROW-FORMAT
- ▶ DATA

Das Feld *Vektor* enthält einen numerischen Wert, der die Interpretation der nachfolgenden Daten spezifiziert. Ein Eintrag 0 signalisiert, daß sich der Satz auf die komplette Tabelle bezieht. Alle positiven Werte ungleich 0 spezifizieren die Spalte, auf die sich der Satz bezieht. Der numerische Wert im nächsten Feld muß, sofern nichts anderes spezifiziert wird, immer auf 0 gesetzt werden. In der dritten Zeile läßt sich optional ein Textstring einblenden. Falls kein Text übergeben wird, stehen dort nur zwei Hochkommata ("" ohne trennendes Leerzeichen dazwischen). Nachfolgend finden Sie eine Beschreibung des Aufbaus der verschiedenen Satztypen:

TABLE (erforderlich)

Dies ist der erste Record in einer SDI-Datei, der den Header-Anfang signalisiert. Der Satz besitzt das folgende feste Format:

```
TABLE
0,1
""
```

Das Schlüsselwort TABLE muß in Großbuchstaben erscheinen, die zweiten und dritten Zeilen sind mit den oben gezeigten Parametern auszugeben. Zwischen den Anführungszeichen der dritten Zeile darf kein Text stehen, womit das SDI-Format von der DIF-Beschreibung abweicht.

VECTORS (optional)

Dieser Satztyp im SDI-Header spezifiziert die Zahl der Vektoren (*Spalten*) der Tabelle. Es gilt folgendes Format:

```
VECTORS
0,<Count>
""
```

Das Feld *Count* spezifiziert die Zahl der Spalten der übertragenen Tabelle. Das Textfeld bleibt leer und muß durch zwei Hochkommata markiert werden. Ein Satz mit diesem Recordtyp ist im SDI-Format optional.

TUPLES (optional)

Die Zahl der Tuples (Zeilen) innerhalb der Tabelle wird mit dem Satztyp gemäß folgendem Format übertragen:

```
TUPLES
0,<Count>
""
```

Count spezifiziert die Zeilenzahl innerhalb der Tabelle. Das Textfeld bleibt leer und muß durch Hochkommata markiert werden. Der Satztyp ist im SDI-Format optional.

GDISPFORMAT (optional)

Hier handelt es sich um einen weiteren optionalen Recordtyp, der die Formatierung des Kalkulationsblatts spezifiziert. Es gilt folgender Satzaufbau:

```
GDISP-FORMAT
<width>,0
<Formatstring>
```

Der Satz besitzt kein Äquivalent im DIF-Format und ist daher auch etwas anders strukturiert. Das Feld *width* spezifiziert die globale Spaltenbreite für das gesamte Rechenblatt. In der dritten Zeile findet sich ein String mit den Steuerzeichen für die Formatierung, die möglichen Steuerzeichen wurden in Tabelle 7.1 beschrieben. Im Gegensatz zu den anderen Satztypen dürfen die Texte hier nicht in Anführungszeichen gestellt werden. Ein Beispiel:

```
GDISP-FORMAT
10,0
LTR$
```

Die Formel setzt die Spaltenbreite für alle Zellen des Kalkulationsblatts auf zehn Zeichen. Alle numerischen Werte sind am linken Zellrand (L) ausgerichtet, wohingegen Texte am rechten Zellrand (TR) justiert werden müssen. Über das Zeichen \$ werden alle numerischen Werte mit zwei Dezimalstellen ausgegeben.

Count spezifiziert die Zeilenzahl innerhalb der Tabelle. Das Textfeld bleibt leer und muß durch Hochkommata markiert werden. Der Satztyp ist im SDI-Format optional.

COL-FORMAT (optional)

Neben der globalen Formatangabe (GDISP-FORMAT) für alle Zellen des Kalkulationsblatts lassen sich über COL-FORMAT auch einzelne Spalten mit abweichenden Formaten belegen, und zwar mehrfach in einem Header. Jeder Satz besitzt folgenden Aufbau:

```
COL-FORMAT
<col>,<width>
<Formatstring>
```

Im Feld *col* wird die Nummer der jeweiligen Spalte angegeben, ihre Breite in Zeichen im nachfolgenden Feld *width*. In der dritten Zeile muß ein Formatstring mit den Zeichen aus Tabelle 7.1 erscheinen. Der String darf nicht durch Anführungszeichen eingefaßt werden.

ROW-FORMAT (optional)

Mit diesem Satztyp läßt sich das Darstellungsformat einer Zeile spezifizieren. Der Record besteht aus drei Zeilen mit dem folgenden Format:

```
ROW-FORMAT
<row>,<width>
<Formatstring>
```

Im Feld *row* wird die Nummer der betreffenden Zeile spezifiziert. Das nachfolgende Feld *width* definiert die Breite der einzelnen Zellen, in der dritten Zeile muß ein Formatstring mit den Zeichen aus Tabelle 7.1 erscheinen, wobei der String nicht durch Anführungszeichen eingefaßt werden darf. Der Satztyp ist optional und darf im Header mehrfach auftreten, wodurch sich das Format jeder Zeile beliebig einstellen läßt.

DATA (erforderlich)

Dieser Satztyp schließt den SDI-Header ab und ist auf jeden Fall erforderlich. Er signalisiert den Beginn des Datenbereichs und besitzt folgenden Aufbau:

```
DATA
0,0
""
```

Abbildung 7.1 enthält ein Beispiel für den Aufbau eines SDI-Headers:

TABLE	; Kopfsatz
0,1	; 0 = ganze Tabelle, Version 1
""	§
COL-FORMAT	; Spaltendefinition
0,5	; Spalte 0 5 Zeichen breit

TABLE	; Kopfsatz
""	
DATA	; Abschluß SDI-Header
0,0	
""	

Abbildung 7.1 Beispiel eines SDI-Headers

Die mit Semikolons markierten Kommentare gehören nicht zur SDI-Datei und wurden zur Erläuterung nachträglich hinzugefügt.

Der Datenteil einer SDI-Datei

Nach dem DATA-Record beginnt der Datenbereich, in dem für jede Zelle des Rechenblatts ein Eintrag vorhanden sein muß. Als erstes wird die Zelle A1 übertragen. Danach folgen die Zellen B1 bis X1 der ersten Zeile, wobei jede neue Zeile durch das Schlüsselwort BOT, eine Leerzeile (CR/LF) oder einen anderen Separator (beispielsweise ein Semikolon) eingeleitet wird.

Die eigentlichen Datensätze sind nach folgender Struktur aufgebaut:

<Typ>,<Wert>
<String>

In *Typ* wird die Satzart gemäß Tabelle 7.2 festgelegt:

Typ	Bedeutung
1	Text
0	numerischer Wert
-1	Datendefinition
-2	GOTO-Zielspezifikation
-3	Anzeigeformat für Eingaben
-4	Kalkulationsformel
-5	Wiederholungszähler

Tabelle 7.2 Satzarten für Daten im SDI-Format

Insgesamt sind im SDI-Format sieben verschiedene Datentypen festgelegt, die Bedeutung der restlichen Felder ist je nach Datentyp unterschiedlich definiert. Die entsprechenden Strukturen sind nachfolgend beschrieben.

Text Entry (Typ = 1)

Mit diesem Datentyp lassen sich Texte im SDI-Format übertragen. Es gilt dabei folgender Satzaufbau:

1,<Wert>
<String>

Die letzte Zeile enthält den zu übertragenden Text. Besteht er jedoch nur aus Leerzeichen, müssen diese in Hochkommata eingeschlossen werden. Im Feld *Wert* sind die Zahlen 0 und 1 zulässig. Diese Werte spezifizieren, auf welche Art und Weise der nachfolgende Text zu interpretieren ist. Beim Wert 0 liegt in der folgenden Zeile ein regulärer Text vor. Dieser Text wird in die adressierte Zelle gespeichert. Der Wert 1 spezifiziert, daß sich der nachfolgende Text in mehreren Zellen wiederholt. Die Information über die Zahl der Wiederholungen findet sich im Satztyp –5 (*Repeat Count Entry*).

Numeric Entry (Typ = 0)

Numerische Werte werden mit diesem Datentyp im SDI-Format übertragen. Es gilt dabei folgender Satzaufbau:

0,<Wert>
<Format>

Im Feld *Wert* steht die zu übertragende Zahl. Die Formatanweisung in der zweiten Zeile bestimmt dann die Darstellung in der betreffenden Zelle. In Anlehnung an die DIF-Definition sind in diesem Feld die Schlüsselwörter gemäß Tabelle 7.3 zulässig.

Indikator	Bedeutung	Wert
V	Value	beliebig
NA	not available	0
ERROR	Fehler	0
NULL	Zelle nicht belegt	0

Tabelle 7.3 Formate für numerische Werte in SDI-Dateien

Gültige numerische Werte wurden mit dem Buchstaben V markiert, während im Feld *Wert* der numerische Wert als ASCII-String steht. Beim Format NA (*not available*) steht kein numerischer Wert zur Verfügung, und das betreffende Feld enthält den Eintrag 0. Das gleiche gilt für die Formate NULL und ERROR. Die folgende Abbildung enthält einige SDI-Datensätze:

0,13.000	; gültiger Wert 13,0
V	
0,0	; Wert nicht verfügbar
NA	
0,0	; Zelle unbelegt
NULL	
0,0	; Wert fehlerhaft
ERROR	

Abbildung 7.2 Datensätze im SDI-Format

Die Kommentare gehören nicht zum SDI-Format und wurden nachträglich eingefügt.

Data Definition Entry (-1)

Dieser Recordtyp dient zur Markierung neuer Zeilen (*Begin of Tuple*) oder zur Endmarkierung (*End of Data*) des Datenbereichs. Die folgenden zwei Variationen sind zulässig:

-1,0 ; Zeilenanfang
BOT

-1,0 ; Ende Datenbereich
EOD

Das Schlüsselwort BOT (*Begin of Tuple*) steht vor dem ersten Datensatz, mit dem Daten einer neuen Rechenblattzeile übertragen werden. Dies ist insbesondere dann interessant, wenn am Ende einer Zeile nur leere Zellen vorhanden sind. Das Schlüsselwort EOD (*End of Data*) signalisiert das Ende des Datenbereichs. Nach diesem Satz können zwar noch Einträge folgen, sie werden jedoch nicht mehr ausgewertet. Im Feld *Werte* muß bei der Satzart -1 eine 0 eingetragen werden.

Origin Specifier (GOTO) (Typ -2)

Mit diesem Recordtyp lassen sich einzelne Zellen direkt adressieren. Der Satz besteht aus zwei Zeilen mit folgendem Aufbau:

-2,0
<String>

Das Feld für die numerischen Werte ist immer mit dem Wert 0 zu belegen. Die folgende Zeile enthält den Formatstring mit den Zellkoordinaten in diesem Format:

Spalte:Zeile

Die beiden Werte sind durch einen Doppelpunkt getrennt. Die Zelle mit den Koordinaten Spalte = 10 und Zeile = 3 würde also beispielsweise mit folgendem Satz adressiert:

-2,0
10:3

In SuperCalc 3 dürfen die Werte für die Spalten zwischen 1 und 127 sowie die Zeilennummern zwischen 1 und 9999 liegen. Durch diesen Recordtyp läßt sich die Position der nächsten adressierten Zelle direkt spezifizieren, was das Überspringen unbelegter Zellen ermöglicht.

Level Display Formatting Entry (Typ -3)

Mit diesem Format läßt sich das Format der vorhergehenden Zelle angeben. Es gilt folgender Satzaufbau:

-3,0
<Format-String>

Als *Format-String* sind die Zeichen gemäß Tabelle 7.1 erlaubt; dabei dürfen durchaus mehrere Zeichen kombiniert werden. Die Formatanweisung der vorhergehenden Zelle sollte allerdings übernommen werden. Falls keine vorhergehende Zelle existiert, erzeugt der SDI-Parser eine Fehlermeldung.

Formula Entry (Typ -4)

SuperCalc kann mit dem SDI-Format auch Formeln einlesen. Es gilt folgendes Format:

-4,0
<Formel>

Im Feld *Formel* muß eine gültige Kalkulationsformel übergeben werden.

Repeat Count Entry (Typ -5)

Die Satzart berücksichtigt den Wiederholungsfaktor, der bei *Text Entry* außer acht gelassen wurde – hier konnten lediglich Stringwiederholungen spezifiziert werden.

-5,<Wiederholungsfaktor>
<String-Format>

Der Wiederholungsfaktor bezieht sich auf den String in der vorhergehenden Zelle. Im Feld *String-Format* darf allerdings nur das Steuerzeichen *R* auftreten. Der Text der vorhergehenden Zelle wird in die folgenden Zellen n-mal kopiert.