

# The spssDDI Package

June 25, 2008

**Type** Package

**Title** Read SPSS System files and produce valid DDI Version 3.0 documents

**Version** 0.1.1

**Date** 2008-05-28

**Author** Guido Gay

**Maintainer** <gay@irer.it>

**Description** Read SPSS System files and produce valid DDI Version 3.0 documents

**License** GPL-2

**LazyLoad** yes

## R topics documented:

readSpssSav . . . . .	1
spssDDI-package . . . . .	9
writeDDI . . . . .	10

<b>Index</b>	<b>12</b>
--------------	-----------

---

readSpssSav	<i>Read a SPSS Data File</i>
-------------	------------------------------

---

## Description

readSpssSav reads the metadata inside a SPSS System file and optionally the actual data.

## Usage

```
readSpssSav(filename, readdata = FALSE)
```

## Arguments

<code>filename</code>	Character string: the name of the SPSS System file to read from.
<code>readdata</code>	reads the actual data?

## Details

A system file encapsulates actual data and dictionary information (metadata). `readSpssSav` reads the metadata inside a SPSS System file (and optionally the actual data) and returns a list with either three or four elements (*Header, Variables, Mixed, Data*).

### *Header*

**RecordType** Record type code, set to \$FL2

**ProductName** Product identification string.

**LayoutCode** Normally set to 2, rarely 3

**CaseSize** Number of variables per case

**Compress** Set to 1 if the data in the file is compressed, 0 otherwise

**WeightIndex** If one of the variables is used as a weighting variable, set to the value describing its position within the file dictionary, otherwise 0

**NumCases** Number of cases

**Bias** Ordinarily set to 100

**CreationDate** Date of creation of the system file

**CreationTime** Time of creation of the system file

**FileLabel** File label declared by the user

### *Variables*

**Number missings** If the variable has no missing values, set to 0. If the variable has one, two, or three discrete missing values, set to 1, 2, or 3, respectively. If the variable has a range for missing variables, set to -2; if the variable has a range for missing variables plus a single discrete value, set to -3

**Print fmt** Number of decimal places, field width, format type (see below), last value set to zero

**Write fmt** Number of decimal places, field width, format type (see below), last value set to zero

**Varname** Variable name. The variable name is padded on the right with spaces

**Varlabel** Variable label

**Missing values** It has the same number of elements as the absolute value of Number missings. For discrete missing values, each element represents one missing value. When a range is present, the first element denotes the minimum value in the range, and the second element denotes the maximum value in the range. When a range plus a value are present, the third element denotes the additional discrete missing value.

**Typecode** Set to 0 for a numeric variable, for a string variable set to its width

Format types are defined as follows (see 'pspp-dev.pdf' (<http://www.gnu.org/software/pspp/>) for details):

**String** 1, 2

**Numeric** 3 - 12, 15 - 17, 31 - 37

**Date** 20, 22 - 24, 28 - 30, 38 - 39. Dates are stored as the number of seconds since midnight, October 14, 1582

**Time** 21, 25. Times are stored as a number of seconds that represents a time interval

**Other** 0, 13, 14, 18, 19, 26, 27

#### *Mixed*

**Value labels** Labels for discrete variable's values. Link to variable through the value describing the variable's position within the file dictionary

**Symmis, highest, lowest** System missing value; largest possible positive number; second-largest negative number

**Major, minor, revision, floating, endianness, character** Three software version numbers; floating point representation code (IEEE 754 = 1; IBM 370 = 2; DEC VAX = 3); endianness (big-endian = 1, little-endian = 2); character set (EBCDIC = 1, 7-bit ASCII = 2, 8-bit ASCII = 3, DEC Kanji = 4, plus windows codes)

**Short and long varnames** Short and long variables names

**Variable sets** Set name and variables names

**Very long variables** Very long string variable's width

**Document** Notes

**Trend** Trend

**Display** Display parameters. (see 'pspp-dev.pdf' (<http://www.gnu.org/software/pspp/>) for details)

**Strings' value labels** String variables' value labels

**Encoding** UTF-8, windows-1252, ...

**Other** Empty

#### *Data*

Character vector with length equal to CaseSize times NumCases.

#### **Note**

Invoking `str(readSpssSav("benchmarkSpss16.sav", readdata=T))` generates the following output:

```
List of 4
 $ Header      :List of 11
  ..$ RecordType : chr "$FL2"
  ..$ ProductName : chr "@(#) SPSS DATA FILE MS Windows 16.0.2
  ..$ LayoutCode  : int 2
  ..$ CaseSize    : int 20
  ..$ Compress    : int 1
  ..$ WeightIndex : num 0
```

```

..$ NumCases      : int 3
..$ Bias          : num 100
..$ CreationDate  : chr "23 Jun 08"
..$ CreationTime  : chr "16:21:00"
..$ FileLabel     : chr "Test file"
$ Variables:List of 20
..$ V1 :List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt      : int [1:4] 2 8 5 0
.. ..$ Varname        : chr "V1      "
.. ..$ Varlabel       : chr "Integer"
.. ..$ Missing values : NULL
.. ..$ Typecode       : int 0
..$ V2 :List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 2 8 5 0
.. ..$ Write fmt      : int [1:4] 2 8 5 0
.. ..$ Varname        : chr "V2      "
.. ..$ Varlabel       : chr "Decimal"
.. ..$ Missing values : NULL
.. ..$ Typecode       : int 0
..$ V3 :List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt      : int [1:4] 2 8 5 0
.. ..$ Varname        : chr "V3      "
.. ..$ Varlabel       : chr "Integer, sysmis"
.. ..$ Missing values : NULL
.. ..$ Typecode       : int 0
..$ V6 :List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 2 8 5 0
.. ..$ Write fmt      : int [1:4] 2 8 5 0
.. ..$ Varname        : chr "V6      "
.. ..$ Varlabel       : chr "Decimal, sysmis"
.. ..$ Missing values : NULL
.. ..$ Typecode       : int 0
..$ V7 :List of 7
.. ..$ Number missings: int 2
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt      : int [1:4] 2 8 5 0
.. ..$ Varname        : chr "V7      "
.. ..$ Varlabel       : chr "Integer, discrete user missings"
.. ..$ Missing values : num [1:2] 1 2
.. ..$ Typecode       : int 0
..$ V8 :List of 7
.. ..$ Number missings: int -2

```

```
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt     : int [1:4] 2 8 5 0
.. ..$ Varname       : chr "V8      "
.. ..$ Varlabel      : chr "Integer, range user missings"
.. ..$ Missing values : num [1:2] 1 5
.. ..$ Typecode      : int 0
..$ V9 :List of 7
.. ..$ Number missings: int -2
.. ..$ Print fmt      : int [1:4] 2 8 5 0
.. ..$ Write fmt     : int [1:4] 2 8 5 0
.. ..$ Varname       : chr "V9      "
.. ..$ Varlabel      : chr "Decimal, range user missings"
.. ..$ Missing values : num [1:2] 0.1 10.7
.. ..$ Typecode      : int 0
..$ V10:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt     : int [1:4] 2 8 5 0
.. ..$ Varname       : chr "V10     "
.. ..$ Varlabel      : chr "Integer, value labels"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 0
..$ V11:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 5 0
.. ..$ Write fmt     : int [1:4] 2 8 5 0
.. ..$ Varname       : chr "V11     "
.. ..$ Varlabel      : chr "Integer, value labels, variable attribute"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 0
..$ V12:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V12     "
.. ..$ Varlabel      : chr "String"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 8
..$ V13:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 14 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V13     "
.. ..$ Varlabel      : chr "Long string"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 14
..$ V14:List of 7
.. ..$ Number missings: int 0
```

```

.. ..$ Print fmt      : int [1:4] 0 -1 1 0
.. ..$ Write fmt     : int [1:4] 0 -1 1 0
.. ..$ Varname       : chr "V14      "
.. ..$ Varlabel      : chr "Very long string"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 255
..$ V15:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V15      "
.. ..$ Varlabel      : chr "String, value labels"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 8
..$ V16:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 14 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V16      "
.. ..$ Varlabel      : chr "Long string, value labels"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 14
..$ V17:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 -1 1 0
.. ..$ Write fmt     : int [1:4] 0 -1 1 0
.. ..$ Varname       : chr "V17      "
.. ..$ Varlabel      : chr "Very long string, value labels"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 255
..$ V18:List of 7
.. ..$ Number missings: int 1
.. ..$ Print fmt      : int [1:4] 0 8 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V18      "
.. ..$ Varlabel      : chr "String, discrete user missings"
.. ..$ Missing values : chr "aaaa    "
.. ..$ Typecode      : int 8
..$ V19:List of 7
.. ..$ Number missings: int 1
.. ..$ Print fmt      : int [1:4] 0 14 1 0
.. ..$ Write fmt     : int [1:4] 0 14 1 0
.. ..$ Varname       : chr "V19      "
.. ..$ Varlabel      : chr "Long string, discrete user missings"
.. ..$ Missing values : chr "bbbbbb  "
.. ..$ Typecode      : int 14
..$ V20:List of 7
.. ..$ Number missings: int 1

```

```

.. ..$ Print fmt      : int [1:4] 0 -1 1 0
.. ..$ Write fmt     : int [1:4] 0 -1 1 0
.. ..$ Varname       : chr "V20      "
.. ..$ Varlabel      : chr "Very long string, discrete user missings"
.. ..$ Missing values : chr "ccccccc"
.. ..$ Typecode      : int 255
..$ V21:List of 7
.. ..$ Number missings: int 0
.. ..$ Print fmt      : int [1:4] 0 8 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V21      "
.. ..$ Varlabel      : chr "String, variable attribute"
.. ..$ Missing values : NULL
.. ..$ Typecode      : int 8
..$ V22:List of 7
.. ..$ Number missings: int 1
.. ..$ Print fmt      : int [1:4] 0 14 1 0
.. ..$ Write fmt     : int [1:4] 0 8 1 0
.. ..$ Varname       : chr "V22      "
.. ..$ Varlabel      : chr "Long string, discrete user missings,copia"
.. ..$ Missing values : chr "bbbbbb  "
.. ..$ Typecode      : int 14
$ Mixed      :List of 12
..$ Value labels                                :List of 3
.. ..$ :List of 3
.. .. ..$ Value, label : chr [1:2] "1" "male"
.. .. ..$ Value, label : chr [1:2] "2" "female"
.. .. ..$ var number(s): int 8
.. ..$ :List of 3
.. .. ..$ Value, label : chr [1:2] "1" "male"
.. .. ..$ Value, label : chr [1:2] "2" "female"
.. .. ..$ var number(s): int 9
.. ..$ :List of 2
.. .. ..$ Value, label : chr [1:2] "aaaaaaaa" "a8"
.. .. ..$ var number(s): int 13
..$ Sysmis, highest, lowest                    : num [1:3] -1.80e+308 1.8
..$ Major,minor,revision,floating,endianness,character: int [1:6] 16 0 2 1 2 1252
..$ Short and long varnames                    :List of 20
.. ..$ Short, long: chr [1:2] "V1" "V1"
.. ..$ Short, long: chr [1:2] "V2" "V2"
.. ..$ Short, long: chr [1:2] "V3" "V3"
.. ..$ Short, long: chr [1:2] "V6" "V6"
.. ..$ Short, long: chr [1:2] "V7" "V7"
.. ..$ Short, long: chr [1:2] "V8" "V8"
.. ..$ Short, long: chr [1:2] "V9" "V9"
.. ..$ Short, long: chr [1:2] "V10" "V10"
.. ..$ Short, long: chr [1:2] "V11" "V11"
.. ..$ Short, long: chr [1:2] "V12" "V12"

```



## References

**DDI** <http://www.ddialliance.org/>

**PSPP** <http://www.gnu.org/software/pspp/>

**SPSSReader** [http://forge.opendatafoundation.org/gf/project/ukda\\_dext/frs/](http://forge.opendatafoundation.org/gf/project/ukda_dext/frs/)

## Examples

```
## Not run:
benchmark<-readSpssSav("benchmarkSpss16.sav")

essSample<-readSpssSav("essSample.sav")

essSample<-readSpssSav("essSample.sav", readdata=T)

## End(Not run)
```

---

spssDDI-package      *A package to read SPSS files and produce valid DDI Version 3.0 XML documents*

---

## Description

There are two functions included in this package:

`readSpssSav` reads the metadata inside a SPSS System file and optionally the actual data.

`writeDDI` produces a valid DDI Version 3.0 XML document from `readSpssSav` output.

## Details

The SPSS System file format is described in ‘pspp-dev.pdf’ (<http://www.gnu.org/software/pspp/>). PSPP is a program for statistical analysis of sampled data.

SPSS System files are also described in SPSSReader ([http://forge.opendatafoundation.org/gf/project/ukda\\_dext/frs/](http://forge.opendatafoundation.org/gf/project/ukda_dext/frs/)).

The Data Documentation Initiative (<http://www.ddialliance.org/>) is an international effort to establish a standard for technical documentation describing social science data.

spssDDI is a work in progress. The author welcomes questions, comments and code submissions.

## Author(s)

Guido Gay

Maintainer: <gay@irer.it>

## References

**DDI** <http://www.ddialliance.org/>

**PSPP** <http://www.gnu.org/software/pspp/>

**SPSSReader** [http://forge.opendatafoundation.org/gf/project/ukda\\_dext/frs/](http://forge.opendatafoundation.org/gf/project/ukda_dext/frs/)

## Examples

```
## Not run:
benchmark<-readSpssSav("benchmarkSpss16.sav")
writeDDI(benchmark, "benchmarkSpss16.sav")

essSample<-readSpssSav("essSample.sav")
sink("essSample.xml")
writeDDI(essSample, "essSample.sav")
sink()
## End(Not run)
```

---

writeDDI

*Write a valid DDI 3.0 XML document*

---

## Description

writeDDI produces a valid DDI Version 3.0 XML document from readSpssSav output.

## Usage

```
writeDDI(l, spssSavFile, studyunit="ID", maxNumberMissings=100)
```

## Arguments

`l` readSpssSav output.

`spssSavFile` Character string: the name of the SPSS System file.

`studyunit` Character string: a unique string that identifies the study.

`maxNumberMissings` Number: maximum number of missing values in the missingValue DDI's attribute.

## Details

writeDDI generates the following DDI content: Citation, Abstract, UniverseReference, Purpose, LogicalProduct (CategoryScheme, CodeSchemes, VariableScheme), PhysicalDataProduct (including ProprietaryRecordLayout for SPSS), PhysicalInstance.

## Value

The function doesn't return a value.

**Note**

To validate the DDI instance, please enable `physicaldataprotect_proprietary:3_0_Beta`.

**Author(s)**

Guido Gay

**References**

**DDI** <http://www.ddialliance.org/>

**PSPP** <http://www.gnu.org/software/pspp/>

**SPSSReader** [http://forge.opendatafoundation.org/gf/project/ukda\\_dext/frs/](http://forge.opendatafoundation.org/gf/project/ukda_dext/frs/)

**Examples**

```
## Not run:
benchmark<-readSpssSav("benchmarkSpss16.sav")
writeDDI(benchmark,"benchmarkSpss16.sav")

essSample<-readSpssSav("essSample.sav")
sink("essSample.xml")
writeDDI(essSample,"essSample.sav")
sink()

## End(Not run)
```

# Index

\*Topic **file**

readSpssSav, 1

writeDDI, 10

\*Topic **package**

spssDDI-package, 9

readSpssSav, 1

spssDDI-package, 9

writeDDI, 10