

# memo

dept. Bone Marrow Donors Worldwide  
author Jack Bakker, Europdonor Foundation, Leiden  
phone +31-71-5685300 fax +31-71-5210457  
e-mail Bakker@Europdonor.NL



our reference  
your reference

version 1.0 to BMDW web site  
date 15-Apr-2005  
subject BMDW Format for Data Export address  
cc

## Format for Data Export

### Introduction

This document describes the file format for the donor and cord blood data file as available *for download* from the BMDW web site.

### File Format

The data file is a plain ASCII file that will be compressed using standard compression (ZIP) software available for multiple platforms, removing the requirement of using a Windows or DOS system to extract the data.

The data file is written in CSV (Comma Separated Values) format. There are many tools available for parsing CSV files, for all sorts of platforms. The format of regular CSV files ([1], based on [2]) is specified as follows:

1. Each record of a CSV file is a set of ASCII values separated by a comma (,).
2. If a data value contains itself the separator (,), then it (the data value) has to be put between double quotes ("). If the value does not contain the separator character then quoting is optional.
3. If a data value contains a double quote ("), that double quote is represented by two double quotes ("").
4. Leading or trailing spaces are removed from data values.
5. The output string "" represents the value ". In other words, it is assumed that it was created through rule 3, and only this rule, i.e. that the value was not quoted.

A generic example: the record

```
123,"123,521.2","Mary says ""Hello, I am Mary""",""
```

contains the following four values:

- 123
- 123,521.2
- Mary says "Hello, I am Mary"
- an empty string

### Columns included in the BMDW data file

The following data columns, and specifically in this order, are included in the file:

Serology or search determinant values:	A1	HLA-A, first antigen
	A2	HLA-A, second antigen
	B1	HLA-B, first antigen
	B2	HLA-B, second antigen
	DR1	HLA-DR, first antigen
DNA values:	DR2	HLA-DR, second antigen
	DNA_A1	DNA-A, first allele
	DNA_A2	DNA-A, second allele
	DNA_B1	DNA-B, first allele
	DNA_B2	DNA-B, second allele
	DNA_C1	DNA-C, first allele
	DNA_C2	DNA-C, second allele
	DRB11	DNA-DRB1, first allele
	DRB12	DNA-DRB1, second allele
	DRB31	DNA-DRB3, first allele
	DRB32	DNA-DRB3, second allele
	DRB41	DNA-DRB4, first allele
	DRB42	DNA-DRB4, second allele
	DRB51	DNA-DRB5, first allele
	DRB52	DNA-DRB5, second allele
	DQB11	DNA-DQB1, first allele
	DQB12	DNA-DQB1, second allele
	DQA11	DNA-DQA1, first allele
	DQA12	DNA-DQA1, second allele
	DPB11	DNA-DPB1, first allele
	DPB12	DNA-DPB1, second allele
	DPA11	DNA-DPA1, first allele
	DPA12	DNA-DPA1, second allele
	Other values:	ID
NR		Number of donors or cord blood units for this phenotype
REGCODE		Registry code
Cord blood values:	NVC	Net volume collected (in ml)
	TNC	Total nucleated cell count (rounded number, in $10^7$ )
	CD34P	Collected number of CD34+ cells (numeric value with decimal point, in $10^6$ )
	MONONUC	Collected number of mononuclear cells (in $10^7$ )

Other data values may be added in the future, when they become available.

### **File details**

The data file name is BMDWDATA.CSV, and BMDW will use the PKWare ZIP format [3] for compressing the file. The file available for download will therefore be named BMDWDATA.ZIP.

The new format will become active once the new BMDW online match programs are released. We will keep the current download active for a couple of months.

Finally, the file will be available for download only through the BMDW web site, similar to the current download, and will only be made available to participating registries that have requested authorization for downloading the data. Authorization is required since enabling the download for a user requires system administration (setting up the user account, etc.).

### **Revision History**

- 15-Apr-2005 v1.0 Initial version (JB)

### **References**

- [1] <http://aspn.activestate.com/ASPN/docs/ActiveTcl/8.4/tcllib/csv/csv.html>, or  
<http://developer.apple.com/documentation/Darwin/Reference/ManPages/mann/csv.n.html>
- [2] <http://tools.ietf.org/html/rfc4180>
- [3] <http://datacompression.info/ArchiveFormats/zip.txt>