The PERSAM 3 spatial data binary file format

PERSAM reads the spatial input data from binary .bin files in the persamdata/bin directory. These files contain a copy of the EFSA Spatial Data Version 1.1, which is distributed in ASCII grid files. As a reference these original ASCII grid files are provided in the persamdata folder as well (arc_info_ascii_grid.zip).

The .bin files are optimised for performance by
. providing the data in binary format, which avoids parsing textual data into numbers, and
. cropping away data that is missing (nodata/-9000) in all spatial data sets (830 columns at the east side, 150 rows at the north side)

Format description

The data is stored using run-length encoding. This means the file contains a series of tuples (count, value). Such a tuple corresponds to a run of count numbers with the given value.
count is stored as a 32 bit integer. value is stored as a 64 bit floating point number.
For example:
the series of values
-9000 -9000 -9000 0.23 0.59 -9000 -9000 -9000 -9000
in the ASCII grid file, would be encoded as
3 -9000.0 1 0.23 1 0.59 4 -9000.0
and stored in binary in the .bin file.

The no_data value is -9000.0

The cells are ordered from west to east, north to south (row by row).
The files contain no metadata, nor row separators. A fixed grid is used:
. reference system: ETRS 89 LAEA
. reference unit: meter
. resolution: 1000
. min. X: 1500000
. max. X: 6570000 (= 7400000 (in the original ASCII grid files) - 830000 (east cropping))
. min. Y: 900000
. max. Y: 5350000 (= 5500000 (in the original ASCII grid files) - 150000 (north cropping))
. rows: 4450 (= 4600 (in the original ASCII grid files) - 150 (north cropping))
. columns: 5070 (= 5900 (in the original ASCII grid files) - 830 (east cropping))

Notes

The .bin files are read by PERSAM using the java.io.DataInputStream class.
To protect the data integrity, PERSAM validates the .bin files when they are read and refuses to start any calculations with modified data.