Affymetrix� CEL Data File Format

⊞ CEL FILE

Description
The CEL file stores the results of the intensity calculations on the pixel values of the DAT file. A single representative intensity value is stored per cell (feature) of the image. The information below will describe versions 3 and 4 of the CEL file format. Version 3 files were generated by the MAS software while version 4 files are generated by the GCOS software.

Version 3 Format
The format of the CEL file is an ASCII text file similar to the Windows INI format.

The file is divided up into sections. The start of each section is defined by a line containing a section name enclosed in square braces. The section names are: "CEL", "HEADER", "INTENSITY", "MASKS", "OUTLIERS" and "MODIFIED". The data in each section is of the format TAG=VALUE.

The "CEL" section contains the version number of the file. The TAGS are:

TAG	Description
Version	The version number. Always set to 3.

The "HEADER" section contains miscellaneous header information. The TAGS are:

TAG	Description
Cols	The number of columns in the array (of cells).
Rows	The number of rows in the array (of cells).
TotalX	Same as Cols.
TotalY	Same as Rows.
OffsetX	Not used, always 0.
OffsetY	Not used, always 0.
GridCornerUL	XY coordinates of the upper left grid corner in pixel coordinates.
GridCornerUR	XY coordinates of the upper right grid corner in pixel coordinates.
GridCornerLR	XY coordinates of the lower right grid corner in pixel coordinates.
GridCornerLL	XY coordinates of the lower left grid corner in pixel coordinates.
Axis-InvertX	Not used, always 0.
Axis-InvertY	Not used, always 0.
swapXY	Not used, always 0.
DatHeader	The header from the DAT file.
Algorithm	The algorithm name used to create the CEL file.
AlgorithmParameters	The parameters used by the algorithm. The format is TAG:VALUE pairs separated by semi-colons or TAG=VALUE pairs separated by spaces.

The "INTENSITY" section contains intensity information. The TAGS are:

TAG	Description
NumberCells	The total number of cells in the array (Rows*Cols).
CellHeader	The header for the remainder of the data in this section. The header is always set to: "X Y MEAN STDV NPIXELS"
NA	The remaining lines in this section contain the intensity, standard deviation value and the number of pixels used to compute the intensity value for each cell in the array. The order is defined by the header.

The "MASKS" section specifies which cells have been masked by the user. The TAGS are:

TAG	Description
NumberCells	The number of masked cells.
CellHeader	The header for the remainder of the data in this section. The header is always set to: "X Y".
NA	The remaining lines in this section contain the XY coordinates of those cells masked by the user.

The "OUTLIERS" section specifies which cells were called outliers by the software. The TAGS are:

TAG	Description
NumberCells	The number of outlier cells.
CellHeader	The header for the remainder of the data in this section. The header is always set to: "X Y".
NA	The remaining lines in this section contain the XY coordinates of those cells called outliers by the software.

The "MODIFIED" section specifies which cells were modified by the user. This feature was dropped in MAS 4 thus the number of cells in this section should always be 0. The TAGS are:

TAG	Description
NumberCells	The number of outlier cells.
CellHeader	The header for the remainder of the data in this section. The header is always set to: "X Y ORIGMEAN".
NA	The remaining lines in this section contain the XY coordinates and the original intensity value (calculated by the software) of those cells modified by the user.

Version 4 Format
The format of the CEL file is an binary file were values are stored in little-endian format.

The file contents are define by:

Item	Description	Type
1	Magic number. Always set to 64.	integer
2	Version number. Always set to 4.	integer
3	Number of columns.	integer
4	Number of rows.	integer
5	Number of cells (rows*cols).	integer
6	Header length.	integer
7	Header as defined in the HEADER section of the version 3 CEL files. The string contains TAG=VALUE separated by a space where the TAG names are defined in the version 3 HEADER section.	char[length defined above]
8	Algorithm name length.	integer
9	The algorithm name used to create the CEL file.	char[length defined above]
10	Algorithm parameters length.	integer
11	The parameters used by the algorithm. The format is TAG:VALUE pairs separated by semi-colons or TAG=VALUE pairs separated by spaces.	char[length defined above]
12	Cell margin used for computing the cells intensity value.	integer
13	Number of outlier cells.	DWORD
14	Number of masked cells.	DWORD
15	Number of sub-grids.	integer
16	Cell entries - this consists of an intensity value, standard deviation value and pixel count for each cell in the array. The values are stored by row then column starting with the X=0, Y=0 cell. As an example, the first five entries are for cells defined by XY coordinates: (0,0), (1,0), (2,0), (3,0), (4,0).< /p>	(float, float, short)

17	Masked entries - this consists of the XY coordinates of those cells masked by the user.	(short, short)
18	Outlier entries - this consists of the XY coordinates of those cells called outliers by the software.	(short, short)
19	Sub-grid entries - This is the sub-grid definition. There are as many sub-grids in the file as defined by the number of sub-grids above. Each sub-grid is defined as: - row number (integer) - column number (integer) - upper left x coordinate in pixels (float) - upper left y coordinate in pixels (float) - upper right x coordinate in pixels (float) - upper right x coordinate in pixels (float) - lower left x coordinate in pixels (float) - lower left y coordinate in pixels (float) - lower right x coordinate in pixels (float) - lower right x coordinate in pixels (float) - left cell position (integer) - top cell position (integer) - right cell position (integer) - bottom cell position (integer)	(integer, integer, float, float, float, float, float, float, float, float, integer , integer , integer , integer)

Types used are defined as: integer (A 32-bit signed integer), DWORD (32-bit unsigned integer), float (An 32-bit floating-point number), short (16-bit signed integer).