

caArray 007 - MAGE-TAB Files

Question: What are MAGE-TAB Files?

Topic: caArray Usage

Release: caArray 2.X

Date entered: 02/12/2009

Topics on this page include the following:

- [Answer](#)
 - [MAGE-TAB and Data Files Relationships](#)
 - [Building MAGE-TAB Formatted files](#)
- [Have a comment?](#)

Answer

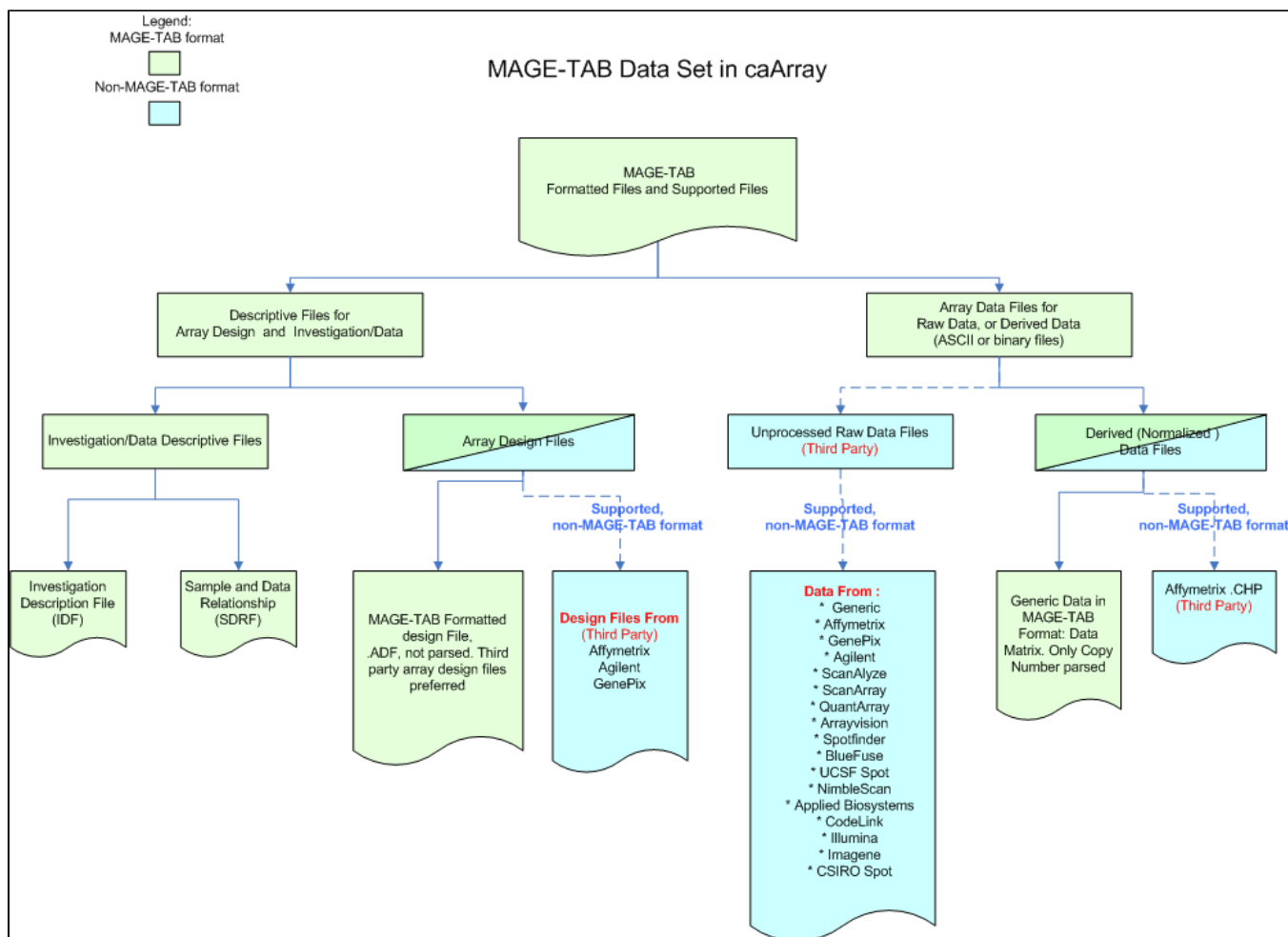
MAGE-TAB (MicroArray Gene Expression Tabular) files, considered the MAGE-TAB data set, refer to simple tab-delimited, spreadsheet-based files, which can be used for annotating and communicating microarray data in a [MIAME compliant fashion](#). MAGE-TAB specification is based on the [Microarray and Gene Expression Object Model \(MAGE-OM\)](#). The MAGE-TAB [specification document](#) and related [publications](#) provide more details on the format.

The MAGE-TAB specification defines four different types of files to fully describe a microarray investigation. The following table summarizes the definition for each of these file types.

File Type	File Extension	Description	Processed by caArray
Investigation Description Format (IDF)	.idf	A MAGE-TAB tab-delimited file providing general information about the investigation, including its name, a brief description, the investigator's contact details, bibliographic references, ontologies/databases referenced and free text descriptions of the protocols used in the investigation.  Warning In a tab-separated line in the IDF, there must be no empty columns, i.e. two tabs with nothing between them. Empty columns will result in import failure.	Yes. Parsed, validated before import
Sample and Data Relationship Format (SDRF)	.sdrf	A MAGE-TAB tab-delimited file (or files) describing the relationships between samples, arrays, data, and other objects used or produced in the investigation, and providing all MIAME information that is not provided elsewhere. An SDRF contains all of the information linking the samples to your data files, (for example, Affymetrix CEL, Affymetrix CHP). Each row in the table represents a hybridization channel, and the columns represent the steps of the experiment, read from left to right.	Yes. Parsed, validated before import
Array Design Format (ADF)	.adf	A tab-delimited file with standardized column names describing the design of an array, e.g., what sequence is located at each position on an array and what the annotation of this sequence is. MAGE-TAB ADF files are not required in a caArray experiment.	Can be uploaded, but are not parsed. Vendor-specific Array Design Files are recommended, if available.
Raw and processed data files	.txt or other	Raw data files: ASCII or binary files, typically in their native formats. Processed (derived) data: tab-delimited text format; can be MAGE-TAB data matrix files. Processed data can be binary as well, such as Affymetrix MASS generated CHP files.	Of data matrix files, only copy number data matrix files are parsed.

To load an experiment into caArray using the MAGE-TAB format, you must create the first two file types shown in the table above, IDF and SDRF. See [Building MAGE-TAB Formatted Files](#). You also must make sure that the corresponding Array Design (the third file type) for your experiment has been loaded into caArray. The fourth file type, raw and/or derived (processed) data files are created by the instruments reading the array results.

The files described in the table make up the MAGE-TAB data set. They can be categorized into two groups: descriptive files and array data files. Some of the files are in MAGE-TAB format (see the [Sourceforge overview](#)) and others are supported as part of the MAGE-TAB data set. See the following figure and its description below.



The descriptive files include IDF, SDRF and Array Design Files. Note that the MAGE-TAB format Array Design File (designated as ADF) is not mandatory in caArray, since array design files for the common arrays are usually available from their respective array providers. If an array design file (which may not be MAGE-TAB format) is available from its array provider, it should be chosen over an ADF file. MAGE-TAB ADF files are not parsed by caArray. Array design files, whether native or ADF, are uploaded using the [Manage Array Design](#) interface under Curation on the caArray left sidebar. The MAGE-TAB file types, IDF, SDRF and data matrix files, on the other hand, are uploaded on the [Manage Data page](#).

Array data files consist of raw array data and/or derived array data. A variety of raw data files, produced by several different scanner makes and models, are supported by the caArray MAGE-TAB parser even though these raw data may not be in MAGE-TAB format. Derived array data refers to either normalized array data, or a data file with data combined from more than one hybridization or scan. The caArray MAGE-TAB parser supports some of these non-MAGE-TAB files, as well. Derived array data files in MAGE-TAB format are called data matrix files. The only data matrix files that caArray parses are copy number data matrix files. For more information, see [File Types in caArray](#).

See also [About File Types in caArray](#) and [MAGE-TAB in caArray--Overview](#) in the caArray User's Guide.

MAGE-TAB and Data Files Relationships

MAGE-TAB files imported into caARRAY map to data files also imported into the same caARRAY experiment. The following figure illustrates the relationship.

