

caArray 017 - Meaning of the caArray Status of Importing - Imported versus Imported Not Parsed

Question: What is the meaning of the caArray Status of Importing: Imported versus Imported Not Parsed?

Topic: caArray Usage

Release: caArray 2.0 and above

Date entered: 04/01/2009

Details about the Question

Browsing through NCI's caArray experiments, I noticed that the status for array design files and for experiment data files can be either "Imported" or "Imported (Not Parsed)". What are the differences between these two states?

Status Designation for Files Imported into caArray

Figure 1. Experiment Data and Array Design Status

Imported Data									
experiment-id-1015897590474521_curated.sdrf	Mage Tab SDRF	 Imported	1,008 Bytes	13 KB					
experiment-id-1015897590474521.idf	Mage Tab IDF	 Imported	2 KB	3 KB					
GSM2204.txt	GEO GSM	 Imported (Not Parsed)	72 KB	237 KB					
GSM2205.txt	GEO GSM	 Imported (Not Parsed)	72 KB	237 KB					
AD7.CEL	Affymetrix CEL	 Imported	1.9 MB	6.5 MB					
AD8.CEL	Affymetrix CEL	 Imported	2.1 MB	6.7 MB					

Array Designs									
HumanRef-8 v2 sequence	Illumina	Gene Expression	2	in_situ_oligo_features	Homo sapiens			 Imported	
HumArray 1.14	UCSF Spot	aCGH	1.14	spotted_ds_DNA_features	Homo sapiens			 Imported (Not Parsed)	

In the screenshot from caArray shown above, some files listed in the 'Imported Data' table, like 'GSM2205.txt', show a status of 'Imported (Not Parsed)', while others, like 'AD7.CEL', show a status of 'Imported'.

Answer

Imported and **Imported (Not Parsed)** describe the import status for array design files and for sample-related array data or annotation files (See [caArray 016 - Which file can be parsed into caArray? What is the benefit of file parsing?](#) File Supported by caArray). All images and experiment-level related information should be added to Supplemental Files ([caArray 014 - Which files can be uploaded into caArray?](#)).

File Import versus File Parsing

Depending on the type of import, there are two stages to the adding a file to caArray: **importing** and **parsing**.

All files being imported pass through the "importing" state. It is a pure data storage process: the data file uploaded will be stored into the caArray database in its original data format (as it is).

If the files submitted are supported by caArray (see [caArray 016 - Which file can be parsed into caArray? What is the benefit of file parsing?](#) File Supported by caArray), they will be further parsed into a caArray table. During the file parsing, values from each data column, such as signal values, or log ratio values from an Affymetrix's CHP file, are extracted into a file-type specific table structure in caArray and associated to probe sets. These parsed files will be marked as **Imported** at the end of process.

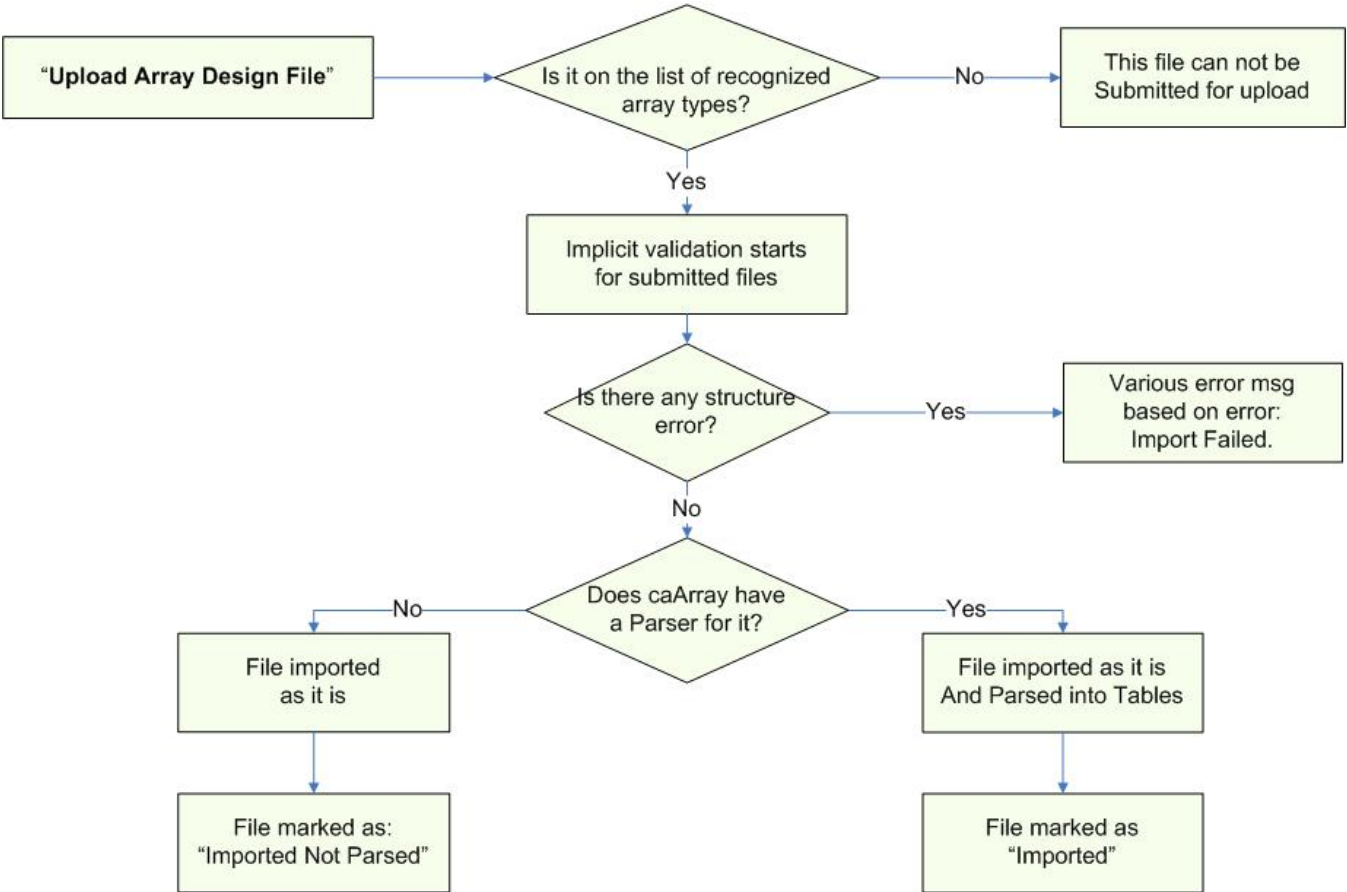
caArray has the ability to upload the array design files or experiment data from many array providers, even if it doesn't have a parser available yet. Those files will ONLY be imported into caArray, without being validated or parsed. The end state of those files is **imported not parsed** as shown in the illustration with the question. Even if a file is not parsed, the user will still be able to download its contents (through caArray's user interface as well as through the programmatic API), and associate the file to samples, extracts, and hybridizations. This caArray feature allows data to be shared and helps the system identify which new parsers need to be developed in the future.

File Upload Process # 1: Array Design files

The following illustration shows the data flow of array design file processing in caArray.

Data Flow of Array Design Processing in caArray

Figure 2: Upload Array Design File



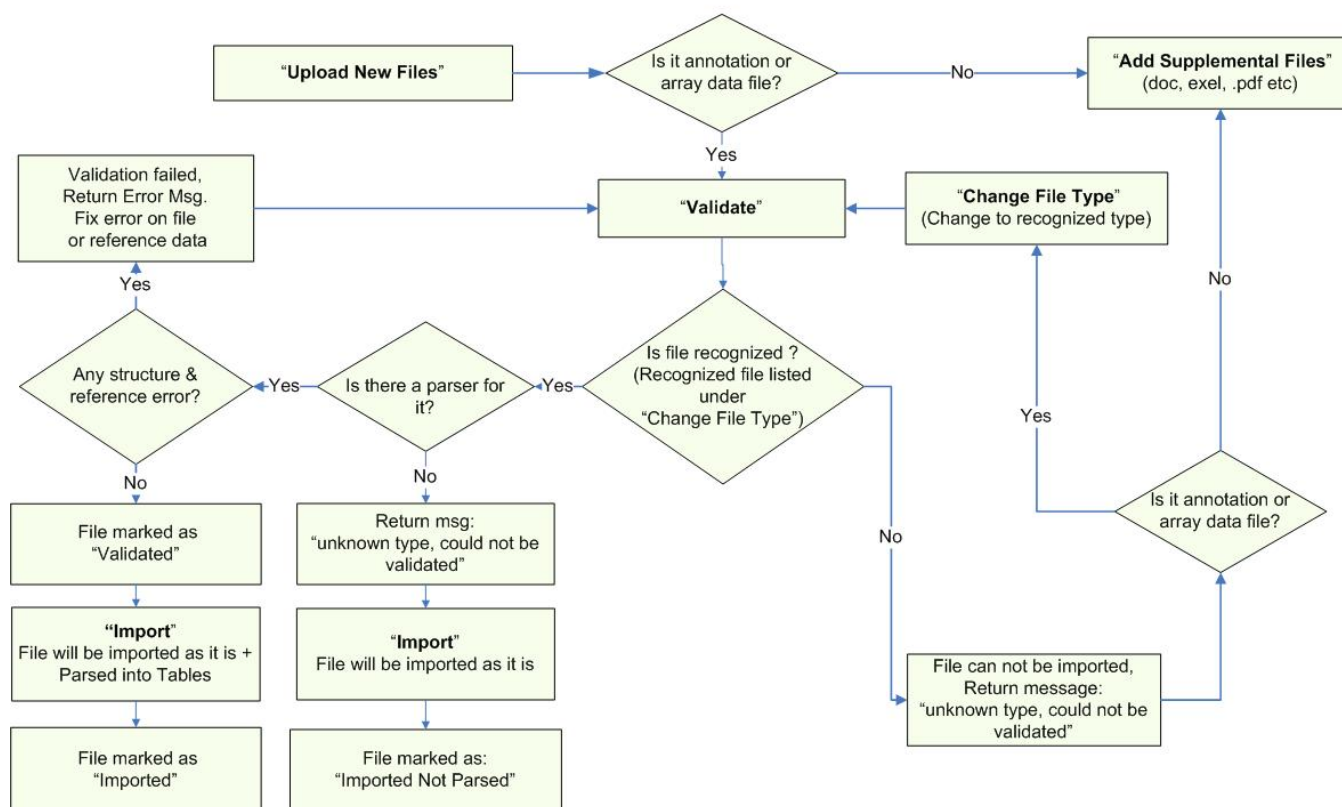
According to the figure above, the first step in processing an uploaded array design file in caArray is to check whether the file on the list of recognized array types. If it isn't, the user is notified that the file can not be submitted for upload. If it is, the file is then checked for any structure errors. If errors are found, an 'Import Failed' error message is displayed with the specific structure error found. If errors aren't found, the file is then checked to see if caArray is capable of parsing it. If caArray can't parse it, the file is imported as it is and marked with a status of 'Imported Not Parsed'. If, however, caArray can parse it, the file is parsed into tables and marked with a status of 'Imported'.

File Upload Process # 2. Non Array Design Files

The following illustration summarizes the general data flow for uploading a file into caArray. It depicts when a file should be loaded as a supplemental file and when it would be imported, either with a final status of **Imported** or **Imported Not Parsed**.

Data Flow for Uploading a File into caArray

Figure 3: File Upload: Supplemental File, Imported, Imported Not Parsed



According to the figure above, the first step in processing an uploaded file other than an array design file is to check whether the file contains annotation or array data. If it doesn't contain either, the user is prompted to 'Add Supplemental Files' such as .doc, .xls, and .pdf.

If the file does contain annotation or array data, the next step is to validate the file. Specifically, the file is checked to see whether it is a recognized file as listed under 'Change File Type'. If it isn't recognized, then the file can't be imported, and the user receives an error message stating, 'Unknown type, could not be validated'.

If the file is recognized, the next step is to check whether it can be parsed. If it can't, then the user receives an error message stating, 'Unknown type, could not be validated'. The file will still be imported as it is and marked with a status of 'Imported Not Parsed'.

If, on the other hand, the file can be parsed, the next step is to check it for any structural or reference errors. If it's free of these errors, the file will be imported as it is, parsed into tables, and marked with a status of both 'Validated' and 'Imported'.

If the file does have any structural or reference errors, the user receives an error message stating, 'Fix error on file or reference data'.

Have a comment?

Please leave your comment in the [caArray End User Forum](#).