Manage Experiment Data Files Use Case Specification  
caArray

|  |  |
| --- | --- |
| Last Revised: | Feb 29, 2012 |
| Produced By: | CBIIT Development |
| Version: | 2.2 |

# Description

Initiated by a Data Owner, this use case allows for data files, MAGE-TAB files and/or supplementary files to be uploaded to an experiment in preparation for import.

# Preconditions

A Data Owner, typically a Lab Scientist, has logged into the System, and has an experiment available to load data into.

# Basic Flow: Upload Files

The Data Owner chooses to upload data files, MAGE-TAB files and/or supplementary files to an existing experiment. To do this, the Data Owner selects one or more individual files or compressed (.zip) files, and specifies whether the .zip archive should be unpacked by the System.

The System starts the process of saving the files to the repository and keeps the Data Owner updated of progress. During the upload, the Data Owner is free to navigate to other portions of the System in a different window.

Once the upload is complete, the System shows the uploaded files along with their inferred type, and provides the Data Owner the ability to change the file type, validate, import, make supplemental or delete any subset of the files.

# Post Conditions

The uploaded files have been stored in the repository, and their types have been inferred (from the extension) if possible. The Data Owner can validate or import the files (see Use Case: Validate Data Files and Use Case: Import Experiment Data). The Data Owner can change the file type, make files supplementary, delete files or select a set of related files.

# Alternative Flow: Change File Type

If the System could not infer the type of the file from the extension, or if the System-inferred type is incorrect, the Data Owner can explicitly select the type of a file(s) from a list of supported types.

# Alternative Flow: Make Supplemental

The Data Owner can select any of the files (including ones with an unknown file type) and mark them as supplemental. These are files that are associated with the entire experiment rather than with specific samples, and are not parsed or validated.

# Alternative Flow: Delete Files

The Data Owner can delete any files, but imported files can be deleted only if they are not associated to any hybridization. Note that “Imported, not Parsed” files can be deleted even if they are associated with a hybridization.

# Alternative Flow: Select Referenced Files

This flow enables easy selection of a set of related files for validation or import as a whole. Out of the set of uploaded files, the Data Owner can pick a MAGE-TAB IDF file and ask for the MAGE-TAB SDRF file referenced from that IDF and all data files referenced from that SDRF to be selected. The System examines the IDF and associated SDRF and selects all files referenced, allowing the ability to validate or import the selected set of files as a whole. If the referenced files cannot be selected due to an error in the MAGE-TAB files, the System indicates the error to the Data Owner.

# Alternative Flow: Resume Interrupted Upload

If an upload is interrupted (either deliberately by the Data Owner pressing the "Pause" button, or by a network connection failure), the upload will be stopped.

The System will allow the Data Owner to resume the upload at the point it was interrupted.

# Exception Flow: Invalid zip

Trying to upload .zip archives with directories or corrupted zips will result in an error.

# Special Requirements: Parseable Types

The System provides parsers for a few supported parseable file types:

* Affymetrix CEL, CHP, CNCHP
* Illumina Sample Probe Profile TXT, Genotyping Processed Data Matrix TXT, CSV
* Agilent Raw TXT
* Genepix GPR
* Nimblegen Pair Report TXT
* Copy Number Data Matrix (a form of MAGE-TAB Data Matrix)

The System supports importing but not parsing for a few currently “unparseable” types:

* Affymetrix DAT, EXP, RPT, TXT
* Illumina IDAT, TXT
* Agilent TSV, Derived TXT
* MAGE-TAB Data Matrix (not Copy Number Data Matrix)
* ImaGene TXT, TIF
* Nimblegen TXT, GFF
* UCSF SPOT SPT
* GEO SOFT, GSM
* ScanArray CSV

The intention is to add parsers for the currently “unparseable” types in future releases.