caArray 2.5.0 and 2.6.0

# Introduction

The purpose of this document is to define the high-level vision for the next major release of caArray – version 2.5.0. It focuses on the capabilities needed by the users of caArray, and the features planned in 2.5.0 to address those needs.

The scope of caArray 2.5.0 is being driven by TRANSCEND requirements, user community needs to support new assay types, the desire to scale to higher volumes of data, current limitations of the upload/import process, the user community’s desire to better manage user roles and collaboration groups, the NCI CBIIT infrastructure requirements, and the need to stay current with the technologies used by caArray.

# Problem Statement

As the number of experiments in caArray grows, there is a critical need for the application to be able to scale to support large volumes of data. The architecture for data storage, import and download must be examined and updated in order to support large data volumes. In addition, there is a need for the user community to be able to easily add support for new assay/data types.

There are current limitations in the user experience during upload and import of data to caArray, and there is a need to make it fast and easy to perform these tasks without a lot of manual effort to “chunk” large data sets.

A requirement from the TRANSCEND team is to be able to restrict permissions to samples and data in caArray based on the Treatment Arm that the patient corresponding to that sample belongs to. This involves updating the caArray permissions model in order to allow this ability.

Single Sign-on is also a TRANSCEND requirement, and the desire is to have a user be able to login to one of the 2 applications (caArray and caIntegrator), and navigate to the other application without having to login again.

Another TRANSCEND requirement is the automatic refresh in caIntegrator of data in caArray. This is expected to be supported by a new/enhanced programmatic interface between caArray and caIntegrator.

The TRANSCEND team would also like an audit log in caArray that shows relevant changes to an experiment – sample/data additions and deletions.

Finally, several of the technologies used by caArray are becoming dated, and there is an urgent need to upgrade them. Key technologies being targeted for 2.5.0 are Jboss 5.1, Java 6, and build and deployment infrastructure.

# Product Overview

## Needs and Features

|  |  |  |
| --- | --- | --- |
| **Need** | **Priority** | **Features** |
| Allow easy addition of support for new assay types | *High* | \* Implement a plugin architecture that allows easy addition of support for new assay types or data parsers. |
| Support storage of large volumes of data | *High* | \* Allow storage of data files on the file system instead of the database.  \* Allow a local installer to configure file system storage. |
| Make it easy to upload and import large sets of data | *High* | \* Eliminate the 2GB upload limit and allow resumable uploads.  \* Split large imports into smaller transactions, thus eliminating the 2GB import limit. |
| Support a permissions model that allows restriction of permissions to samples by Treatment Arm | *High* | \* Update the caArray permissions feature to allow the Data Owner to search for samples by Treatment Arm and set permissions appropriately. |
| Support Single Sign-on | *High* | \* Support the ability for a user to login to caArray or caIntegrator and navigate to the other application without having to log in again. |
| Upgrade the tech stack | *High* | \* Upgrade the application server to Jboss 5.1  \* Upgrade to Java 6  \* Upgrade to MySQL 5.1  \* Upgrade to Nexus (New)  \* Upgrade to BDALite and AHP3 (New) |
| Support automatic refresh in caIntegrator of data in caArray | *Medium* | \* Provide an API method that caIntegrator could call to query if a given experiment has changed in the given time period. |
| Audit log requirements | *Medium* | \* Show relevant changes to an experiment – sample/data additions and deletions. |