# caArray-caIntegrator – Status Meeting

**August 14, 2012**

## Attendees:

|  |  |  |  |
| --- | --- | --- | --- |
| Abe Evans-EL |  | Juli Klemm | Y |
| Cuong Nguyen | Y | Larry Brem |  |
| Deb Hope | Y | Mervi Heiskanen | Y |
| Don Swan | Y | Michael Benham |  |
| Eve Shalley |  | Quy Phung | Y |
| Greg Gurley |  | Rashmi Srinivasa | Y |
| Hemant Undale |  | Robert Shirley |  |
| Henry Schaefer | Y | Shine Jacob | Y |
| Jacob Mensah | Y | Tony Kerlavage |  |
| Jill Hadfield |  | Will Fitzhugh |  |
| JJ Pan | Y | Xiaopeng Bian |  |
| Ishwar | Y | Yeon Choi | Y |

## Discussion:

### Development and QA:

* Current focus:
  + High priority bug fixes that did not make it into the release.

|  |  |  |
| --- | --- | --- |
| ARRAY-2344, ARRAY-2585, ARRAY-2588, ARRAY-2590 | Unexpected error on Job Queue | Resolved |
| ARRAY-2584 | MAGE-TAB update of annotations fails when all referenced data files in a row are already imported | Resolved |
| ARRAY-2497 | Unexpected Error when deleting imported data file from imported/not parsed dataset | Resolved/Cannot Reproduce |
| ARRAY-2586 | Errors while uploading zipped cel files | Open |
| ARRAY-2589 | Experiment disappeared from workspace after changing permissions on caArray Stage tier. | Open |
| ARRAY-2053 | Cannot create new experiment when Agilent data is importing | Reopened due to side effects |
| CAINT-1166 | caIntegrator is not able to receive data from the AIM service at Emory | Resolved |
| CAINT-1163 | AIM Server Grid URL should be user configurable in the GUI | Resolved |
| CAINT-1175 | An error occurs if a user attempts to revert to an original mapping file after remapping | In Progress |
| CAINT-1132 | Attempting to edit or delete Multi Samples Per File data source that has already been mapped results in a HTTP500 error | In Progress |
| CAINT-1019 | Deleting query criteria can result in replicates not being returned | Open |

* Potential caArray Patch Release:
  + Assuming that ARRAY-2586 and ARRAY-2589 are the last two bugs that will be addressed in the patch, and assuming that no additional bugs or security vulnerabilities are found, the project plan shows a projected release date of September 25: <https://ncisvn.nci.nih.gov/svn/caarray2/trunk/docs/project_management/caArray_Project_Plan.mpp>
  + This falls in the Systems “Brown-out” period, and the estimated times for appscans and deployments may not be accurate as a result.
* Upcoming:
  + caIntegrator performance and usability improvements – current performance numbers need to be measured first.
  + The caIntegrator role-based permissions feature needs to be demo-ed to UCSF, but the team needs some time to set up the demo data and prepare a meaningful demo.
  + Support for any issues that come out of UAT by UCSF.
  + Start planning for migration of data from database to file system.

### Systems and DBA:

* caArray Curation and Training tier upgrade – Waiting for Doug Hosier’s decision.
* caIntegrator Training tier upgrade – In Progress
* Database and file system backup and recovery strategy – Yeon’s writeup after last week’s meeting:
  + Option 1. Schedule DB and filesystem backups to tape at the same time. If we needed to recover from tape, we can restore the DB to the time that the filesystem backup was taken (when the file list was generated) since filesystem backups would likely take longer. Potential downside is that there could be some discrepancy in blob files according to the database and what’s actually in the filesystem (i.e. - if files on the filesystem were modified or removed though the application while the filesystem tape backups are running). Hopefully this can minimized by running the backups at low application usage times. But it would be important to have a way of finding the discrepancies or differences through the application so that it can be reconciled if needed.
  + Option 2. Use pre and post scripts within Netbackup, shutdown the database while the backup is taken. This would get consistent backups of both the database and filesystem. Downside is that the database and application won’t be available while the backup is run. This can be anywhere from minutes to several hours, depending on if it’s an incremental or full backup and the size of the filesystem.
  + Option 3. Look into potential for filesystem snapshot backups. It would be similar to option 1 but it would get the DB and filesystem backup times to be very close, without having to roll forward the database to match the filesystem backup time difference of option 1. Downside is that it will utilize lot more to SAN storage to accomplish. Not sure if it’s even a possibility at this point. We’ll update you after we do the research.
  + For all of these options, there is a potential for data loss if a failure happens mid-day and we have to restore from previous night’s tape backups (which is not that different from the current setup). The real benefit of splitting out the blobs to the filesystem from a restoration point of view is that if we needed to restore from backups, we would be able to get the application back up right away in a limited functionality while the blobs are restored. Currently, it would take days to get the application available if we needed to restore from a tape backup.
  + Decision is to go with Option 1. The development team and Systems team will discuss ways to mitigate the impact to the user, e.g., by reducing the window for potential inconsistency, tracking inconsistencies so that the data owner can be informed, etc.

### Documentation:

* caIntegrator User’s Guide changes in progress.

### Support and Knowledge Center:

* caIntegrator installation issue is resolved! Installed v1.4. Don will follow up with the user re: access permissions/groups and set up a meeting with Mervi. They have Agilent gene expression and methylation data that they want to use the caIntegrator analytical tools on.
* Orphaned experiments in caArray can be handled better if a feature was implemented to allow a sysadmin to search by experiment (rather than just by user) to change ownership. ARRAY-1610 has been moved up to the top of the feature backlog.
* caArray user issue (Derrick Lin) – no update yet.

## Open Action Items:

* Full list of closed and current action items are available here: <https://ncisvn.nci.nih.gov/svn/caarray2/trunk/docs/project_management/caArray_caIntegrator_Action_Items.xlsx>

| **Item #** | **Description** | **Responsible Party** | **Date Identified** | **Disposition** |
| --- | --- | --- | --- | --- |
| 53 | Upgrade caArray Curation tier. | Quy Phung | July 24, 2012 | Not Started |
| 54 | Upgrade caArray Training tier. | Don Swan | July 24, 2012 | Not Started |
| 55 | Upgrade caIntegrator Training tier. | Henry Schaefer and Don Swan | July 24, 2012 | In Progress |
|  |  |  |  |  |