

# Synchronizing Object Ids

LSST Database 362

We may need to keep objectIds in sync between data releases as well as in between nightly processing and deep detection runs. This will require doing non-trivial association, which we haven't thought about in the past...

Main reasons to keep objectIds in sync:

- Different groups/users will maintain "external" catalogs and private data sets which they will want to correlate with the central LSST catalogs. If we change objectIds in every release, all these external catalogs and private data sets will have to be recreated or re-associated with our central catalog.
- Released data products include DIASources and Sources. The former set is generated during nightly processing, and the latter during deep detection run. Both of these catalogs will need to be linked to Object Catalog, thus objectIds used in nightly processing and during deep detection runs has to agree.
- It is useful to be able to compare different deep detection runs (e.g. with different deblending thresholds).

## DataAccWG discussion on June 18, 2008 is summarized below:

We could maintain unique objectIds: master list of everything we had every seen

- we would need to maintain one "mapping table" for each auxiliary database
- issue: how to find which partition to use for joins? (for a given objectId we don't know which partition to go to)
- we would need to keep partitionId inside the mapping table

We need to maintain both Source and DIASources in the released catalogs

- are DIASources interesting if we have Sources?
  - ◆ yes, it is an easy way to pick up variability
- we can re-associate DIASource from nightly against Object Catalog produced by Deep Detection
  - ◆ this takes care of DIASources / Sources objectId synchronization

Ultimate generic solution: associate with "previous" catalog every time we generate a new catalog.

- The association should be more thorough than just spatial
- application people should tell us what criteria to use for association
- such re-association will ensure that corresponding object ids are the same

## Issues

Alerts may publish an associated objectId. It's possible that the object is "new"; that is, not present in the deep detection output used to seed the nightly pipelines for the data release cycle in progress. Associating with the previous catalog release will fail to catch such cases. Therefore it seems desirable to sync objectIds of an about to be released catalog with those of the up-to-date catalog immediately prior to release. However, since the association between *about-to-be-released* and *up-to-date* catalogs cannot be performed in zero time, it may still be possible for new objects to appear in the *up-to-date* catalog while the association step is performed (these objects won't participate in the association). This brings up another problem - the deep

detection run producing the *about-to-be-released* catalog may have been running for quite a while (months?) and is therefore unlikely to include images that would cause deep-detect to pick up many "new" objects. Furthermore, new objects created by the nightly pipelines might only show up in a very small number of images - if as a result they fall below the SNR threshold used for detection on coadds, then deep detect may never create corresponding objects and we will have published alerts referring to objects that don't exist in public catalogs. Is this a problem, and if so how should it be handled?

KTL: Data release processing will include difference image analysis, so transients that are not found in deep detection will still be added to the data release object and DIASource catalogs. Data release processing will *not* include the images taken during that processing, so some catch-up association will be needed to resynchronize the nightly pipeline with the data release object catalog.

One approach would be to have alerts include objectIds only for those objects which appear in a DR - for alerts on (or signaling the creation of) new objects, only tentative object attributes would be published. Another possibility is to extract "new" objects from the *up-to-date* catalog and append them to the *about-to-be-released* catalog right before release.

A final question: even if we don't add "new" objects to an official data release, do we want to carry them over to the catalog used by nightly processing for the next DR cycle (since deep detect may not have processed all the images relevant to them)?