

# DataCat for LSST Data Access 2014-11-26

## Date

25 Nov 2014

## Attendees

- Brian Van Claveren
- Tony Johnson
- [Jacek Becla](#)
- [John Gates](#)
- [Fabrice Jammes](#)

## Goals

- To discuss DataCat

## Discussion items

- have mysql instance on slac machine, datacat talking to it
- install tomcat locally on db machine at slac (Brian/Tony need access)
- new python client in works
- need mechanism to load metadata about files residing at NCSA into our catalog
- crawler can't be remote yet (it is on todo)
  - run at NCSA and transfer into via ssh
- datacat key features we like best and want to use: mix of structured and unstructured, crawler, rest api
- use case to implement in prototype: extract metadata from fits headers using afw, load to datacat, then reproduce fits files using metadata from datacat, without reading fits headers.
- would be useful to have notion of hot and cold data in datacat
- java packages that datacat is using for rest api: jersey, jackson
- think more about "foreign tables"
  - foreign tables are tables defined via plugins, then joined during query, only touched when query explicitly asks for it
  - foreign tables still have to be in local db, e.g., didn't explore true functionality of oracle, postgres or mysql that'd access remote tables (planning to)
  - unclear where dividing line for us should be: exposures produced by pipelines, metadata in datacat etc
  - will work on that over the next few weeks
- some relevant pages we looked at:
  - <http://lsst-web.ncsa.illinois.edu/schema/index.php?sVer=baseline>
  - <https://dev.lsstcorp.org/trac/wiki/db/Provenance>
  - <http://dev.lsstcorp.org/htmldocs/DBSchema.pdf>