Data Access Hangout 2015-02-23

Date

23 Feb 2015

Attendees

- IPAC: Tatiana Goldina, Trey Roby, Loi Ly, Serge Monkewitz. Gregory Dubois-Felsmann. Unknown User (xiuqin), Unknown User (zhang)
- SLAC: John Gates, Fritz Mueller, Fabrice Jammes, Unknown User (danielw), Andy Salnikov, Jacek Becla, Kian-Tat Lim

Discussion items

Return JSON values

- json dumps vs flask.jsonify
    - both almost equivalent
    - flask.jasonify also creates some extra header information
- now returning a quoted json string
- build structure:
- an array of metadata objects, one per column, and an array of the data, and dump to json
- need to document what will really be returned, and have a test case that shows it
- what is header? title, description, units, precision. Have it (for db only) now in the API
- better to have arrays of columns instead of arrays of rows? And send as multiple json objects

ipac table

- was planning to use IPAC table
- lots of internal debates at ipac
- Complication: if we have wide columns, ipac table has to feel it with spaces. Also metadata in ipac table poor
- We should just go ahead and use json. If we can overcome streaming issues in json, use json

Web Service

- image cutout works now for any raw stripe 82
- we should have coadds from S13, they are served online (we think).
- John will add support for coadds, Yusra knows where to find coadds
- want to pass to few scientists to show (firefly)
- need to document how to go from db to butler to exposure, it is not based on exposureId because it is SDSS.
- show example: how to produce an identifier.
- Numeric identifiers are computed in obs_sdss/sdssMapper

Terminology

- Level = DC, L1, L2, L3, dev
- (Image) Collection = some label (e.g. DR1, DR2, ktl/test20150202)
- (Image) Kind = raw, fpC, deepCoadd, diffim, template, etc.
- same applies to database: e.g., a DR collection, with kinds (like "efd", etc)

efd

- btw let's not call it efd, because it will be restructured. Maybe "conditions db"? Need to start to design temporal data. For each individual final data release it is just one temporal dimension (it corresponds to the cannonial set used for production). Also, scientists will come up with after-burner corrections to existing data release, will add to personal workflows. They start as L3, at some these corrections might become official, policy question if it is L2 or L3. Need to be able to load new tables (without modifying existing tables), or appending columns.