Storage Inventory System, Object Store, and SODA

Patrick Dowler, Dustin Jenkins
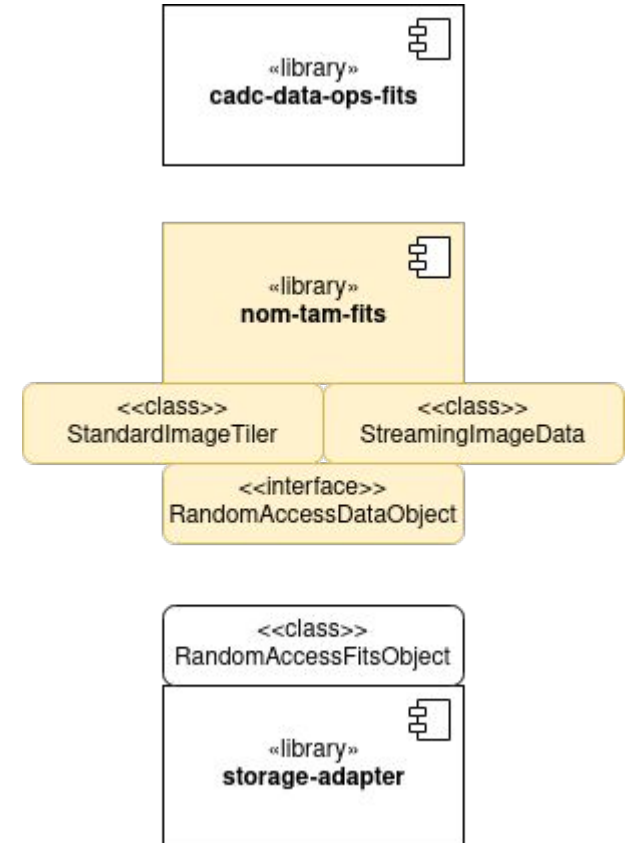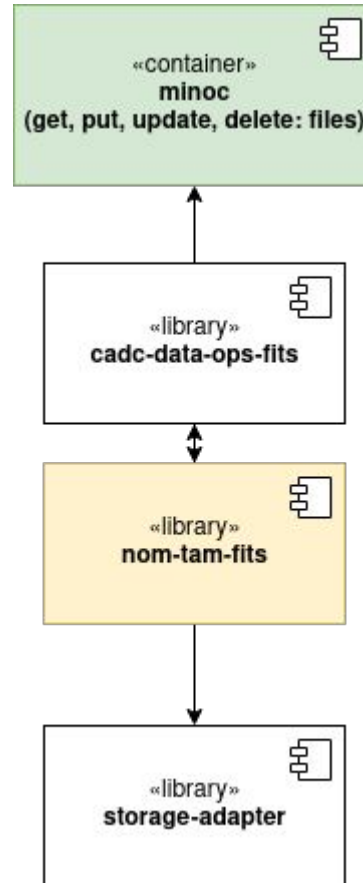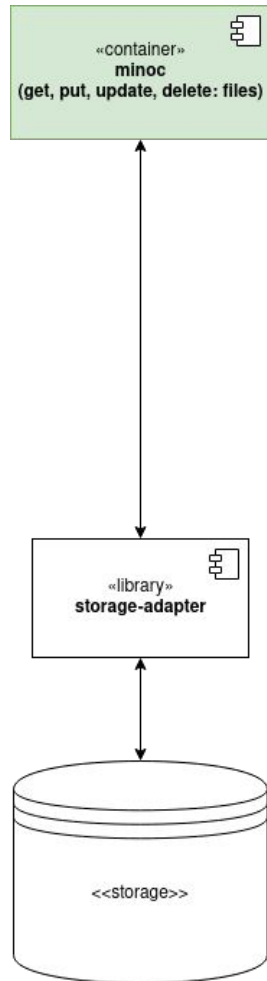**Canadian Astronomy Data Centre**

Background: Storage Inventory

- new distributed archive data management system developed by CADC
- minoc: data (file) access service

    - HTTP REST API: get, put, delete, post (update metadata)
    - GET supports several options:
        - range request for resumable download
        - metadata extraction (currently FITS header)
        - indexed data array subsets aka "pixel cutouts" (FITS image extensions using cfitsio-style syntax)
        - integrating now: SODA cutouts (WCS transformation + pixel cutouts as above)
        - using only metadata (headers) in the file

Storage Inventory

- minoc: data (file) access service

SODA implementation (inside minoc)

- cadc-dali, cadc-soda-server libs: parsing input params, data structure

- cadc-data-ops-fits lib:
  - navigate FITS HDUs
  - read headers
  - transform input sky coord request → local WCS → pixels

- nom-tam-fits lib (forked):
  - added RandomAccessDataObject interface
  - modified StandardImageTiler to stream (input → output)
  - all cutout operations occur while the tiler reads the input

- StorageAdapter impls support RandomAccessDataObject

- back end storage:
  - filesystem w/ extended attribute support
  - CEPH Object Store: using Swift API (reasons), S3 API could work

Preliminary Results & Future Work

- low level testing:
  - CEPH object store (via Rados Gateway REST API): ~< 10ms to seek and read 32-64KiB
  - independent of file size
  - independent of position in file

- TODO:
  - real world performance testing
  - testing & performance on segmented files (size > 5GiB)

- additional input params:
  - META=true|false
  - SUB=[<ext>][<n-dim pixel range>]&SUB=...

  - SODA-next?

Collaboration

https://github.com/opencadc/storage-inventory
- contains complete design and implementation of Storage Inventory system

https://github.com/opencadc/dal
- cadc-dali
- cadc-soda-server
- cadc-data-ops-fits
- cadc-data-ops-hdf in the near future?

https://github.com/opencadc/nom-tam-fits
- our fork of original upstream project (was ~dormant)
- upstream now has a maintainer (**Hi attipaci!**)
- planning to merge our streaming changes upstream asap

All OpenCADC libs now published in Maven Central: org.opencadc:{lib}