IRSA IRTF Archive Ops Concept, version 20161005b

List of product types[1]:
- Fits:
    - SpeX spectra
    - SpeX guider images
    - ISHELL spectra
    - ISHELL guider images
- Group label files
    - glbl text file with metadata for the group
- Summary information files
    - group target info files
    - group QA files
    - weather summary
- Program information
- log files
- weather data

Operations - IRTF:

IRTF will generate the above products and deliver them to the appropriate IRSA transfer filesystem (see **Data Delivery** below). Files should be organized as per the agreed directory structure. The fits files, program information, log files, and weather data will be delivered once. The files under the "Summary information files" heading may be delivered multiple times as IRTF improves their pipelines. When an update delivery is received, it will replace the previous version of that product entirely, i.e. a new QA delivery will replace all of the existing QA files, not add or modify them. File names should, in general, remain unchanged in update deliveries. For example, version information should be provided within the updated files, not in directory or file names (i.e. a png could have a "v2" or "created on DATE" label in one corner, but not have a file name *_qa_v2.png).

A delivery must include a manifest file containing a list of each file included in the delivery and checksums. Our current expectation is that the basic unit of deliveries will be observation nights. A delivery will contain all products for a given night. Initially, deliveries are anticipated to lag the observation date by up to a few weeks.

Updates to summary information files are anticipated roughly once a semester. Affected summary files from previous observation nights will be re-generated using the new summary pipeline, and delivered in the appropriate directory structure to the IRSA transfer filesystem. Detailed mechanics TBD.

Operations – IRSA:

IRSA runs a cronjob every [TBD times per day] that checks the filesystem for new deliveries. The manifests are used to verify when deliveries are complete. When a delivery is complete, IRSA will move the delivery directory to its final location, confirm that the delivery matches the manifest and checksums, send a confirmation message to IRTF, and parse the glbl file for the required metadata. If a delivery conflicts with existing products, (e.g. fits files, program info, log files, and/or group label files for an already existing directory are re-delivered), that will be an error condition and nothing will be done until IRTF and IRSA have confirmed what's going on and that a re-delivery was intended.

Update deliveries will arrive in a different part of the IRSA transfer filesystem and will replace the earlier versions of those products.

Any observations of solar system objects with no NAIF ID will be assigned a placeholder NAIF ID of 2,147,483,647 (2^{31} -1), the largest signed 32 bit integer. This avoids confusion with astrophysical observations where NAIF ID = NULL.

Data Delivery:

IRSA will create an account "irsairtf" on the machine irsaxfr.ipac.caltech.edu and install an ssh public key from IRTF. There will be a specified directory where IRTF will place the deliveries. We want to cleanly separate the products that might change from the products that cannot change. The following is an updated proposal based on discussions on 2016-10-05:

IRTF/$SEMESTER/$DATE/data/$PROGRAM_ID/$GROUP_ID/*.fits
IRTF/$SEMESTER/$DATE/data/$PROGRAM_ID/program_*.txt
IRTF/$SEMESTER/$DATE/data/ielog_$DATE.zip
IRTF/$SEMESTER/$DATE/data/weather_$DATE.zip   (WEATHER_FILE in glbl[2])
IRTF/$SEMESTER/$DATE/data/$PROGRAM_ID/$GROUP_ID/$GROUP_ID.glbl
IRTF/$SEMESTER/$DATE/summary/$PROGRAM_ID/$GROUP_ID/$GROUP_ID_target.*
IRTF/$SEMESTER/$DATE/summary/$PROGRAM_ID/$GROUP_ID/$GROUP_ID_qa.*
IRTF/$SEMESTER/$DATE/summary/weather_$DATE.png   (WEATHER_INFO in glbl[2])
IRTF/$SEMESTER/$DATE_manifest [TBD- research IRSA convention]

ISHELL only:
IRTF/$SEMESTER/$DATE/data/darks/*.fits
or a specific PROGRAM_ID reserved for ISHELL darks instead of "darks/"

The file "$DATE_manifest" is generated with the command

```
find $DATE/ -type f -print0 | xargs --null -n 1 sha256sum | sort > $DATE_manifest
sha256sum $DATE_manifest >> $DATE_manifest
```

Based on discussions on 2016-10-05, the weather summary file applies to the entire night, so there's no need to repeat it for each PROGRAM_ID and GROUP_ID. Is the WEATHER_FILE zip file also something that may be subject to updates? If so, that would move over to the summary/ part of the tree.

In this model, update deliveries would consist entirely of files under the summary/ subdirectory.

References:

1: http://irtfweb.ifa.hawaii.edu/~ida/dev/data_products/
2: http://irtfweb.ifa.hawaii.edu/~ida/dev/data_products/IDA-0012_IDA_glbl_Reference_v100.pdf