Storing GREAT Data in the DCS Science Archive: Proposal to Change from FITS to CLASS

Ed Chambers (USRA/SOFIA), on behalf of the
Data Processing Team

SOFIA Users Group Meeting   Apr 15, 2015

Background

- "The SSMO shall store all imaging and spectroscopic data from science instruments and all telescope imagers in Flexible Image Transport Standard (FITS) files that adhere to a SOFIA keyword list documented within the DCS- SI-01 ICD." - *SOFIA Science & Mission Operations System Specification*

- GREAT Level 3 data are provided in CLASS format.
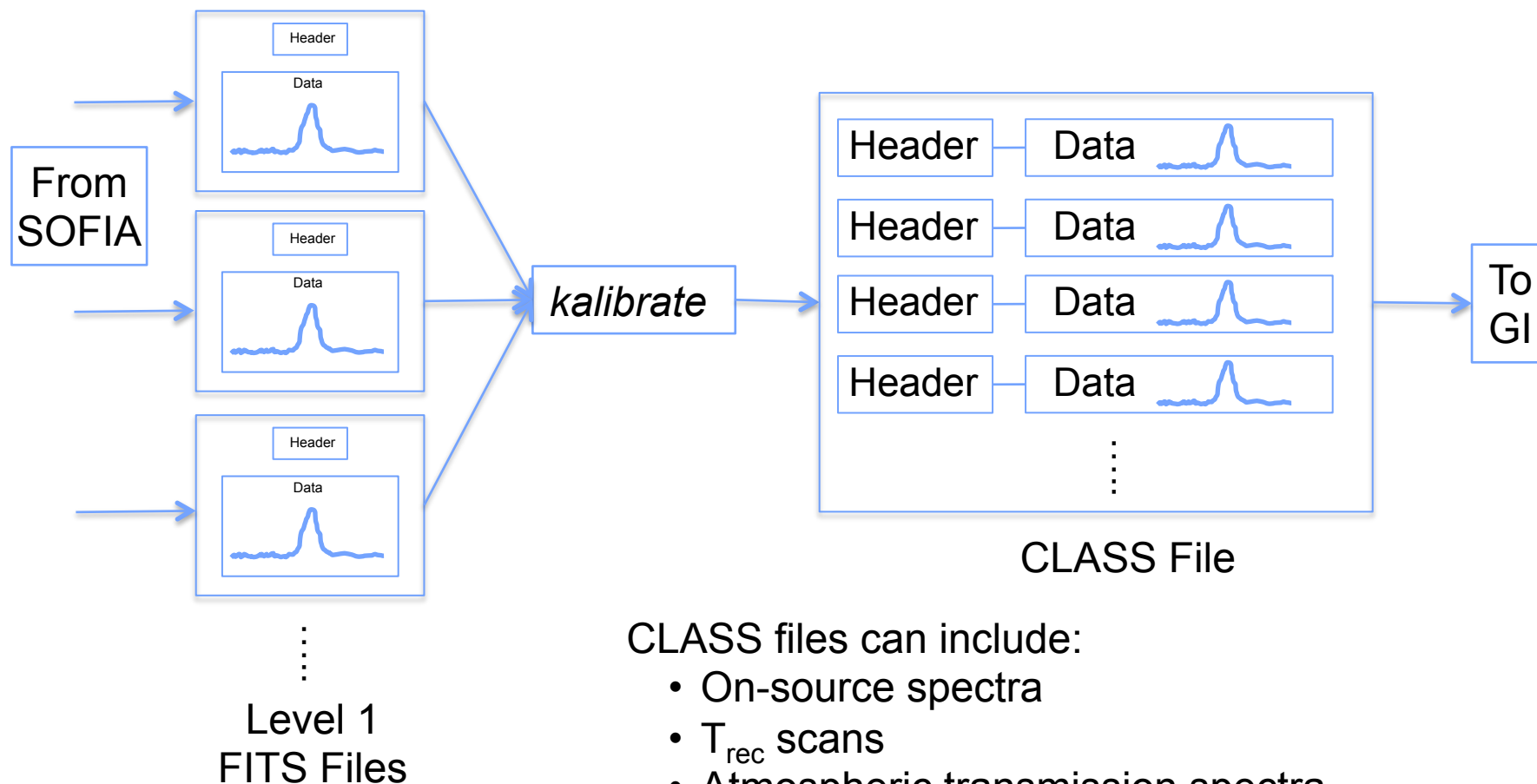
Questions for the SUG:

Does the astronomy community have enough familiarity with the CLASS file format that we could remove the FITS file requirement for the SOFIA archive?

Would doing so cause the data to only be used by spectroscopy experts?

<u>C</u>ontinuum and <u>L</u>ine <u>A</u>nalysis <u>S</u>ingle-dish <u>S</u>oftware

- "CLASS is a software package for reducing spectroscopic data obtained on a single-dish telescope."
  -- http://www.iram.fr/IRAMFR/GILDAS/

- Actively maintained by IRAM and LAB as a module of GILDAS (Grenoble Image and Line Data Analysis Software)

- Long heritage (over 30 years)

- Widely used in the spectroscopy community, for analysis of data obtained at both ground-based (IRAM 30-m, NOEMA, APEX, Effelsberg 100m, CSO, JCMT…) and space-based (SWAS) observatories
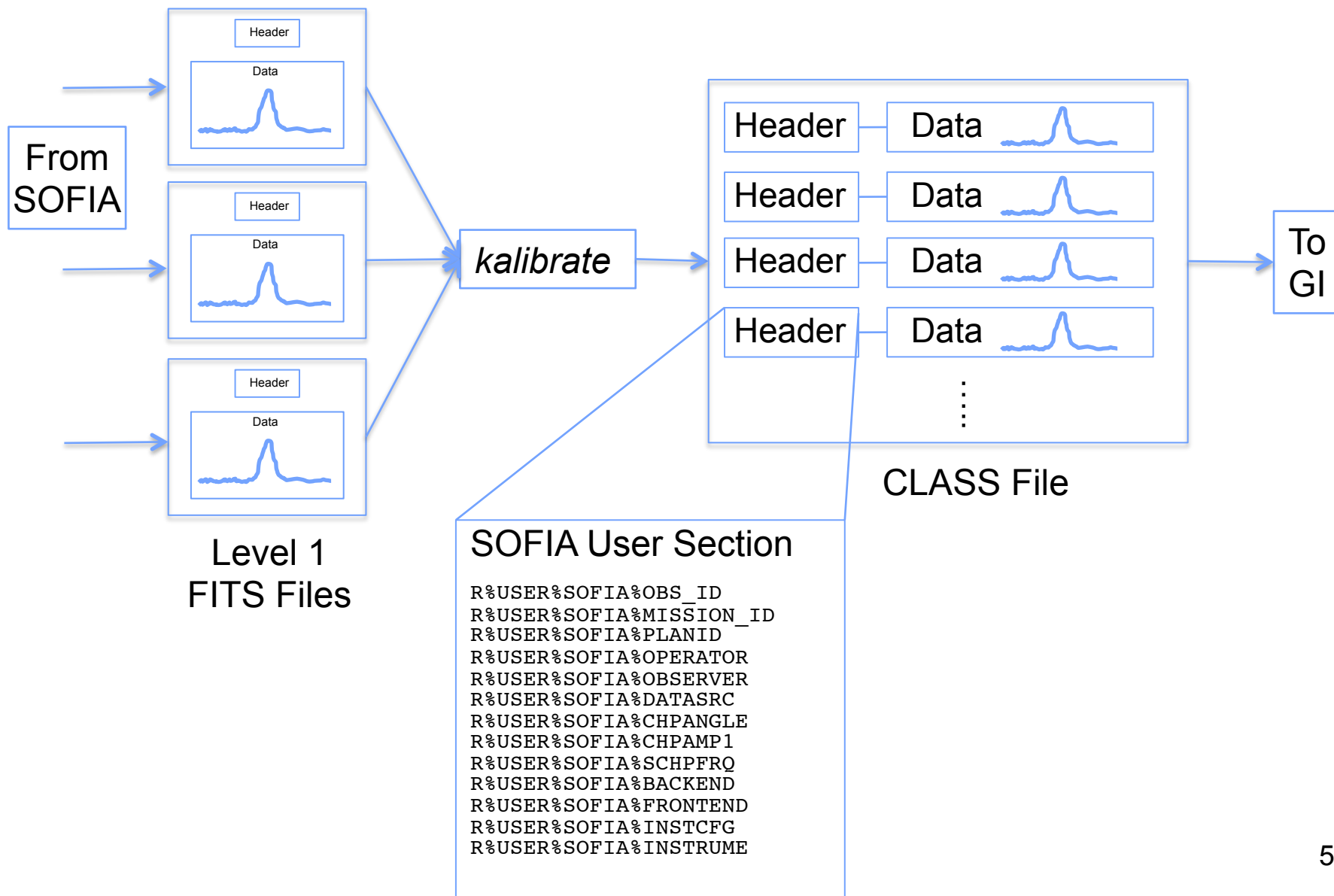
- Available for Linux, Mac OSX, and Windows

iram
Institut de
Radioastronomie
Millimétrique

3

GREAT and CLASS

From SOFIA → Level 1 FITS Files (Header / Data) → *kalibrate* → CLASS File (Header — Data repeated) → **To GI**
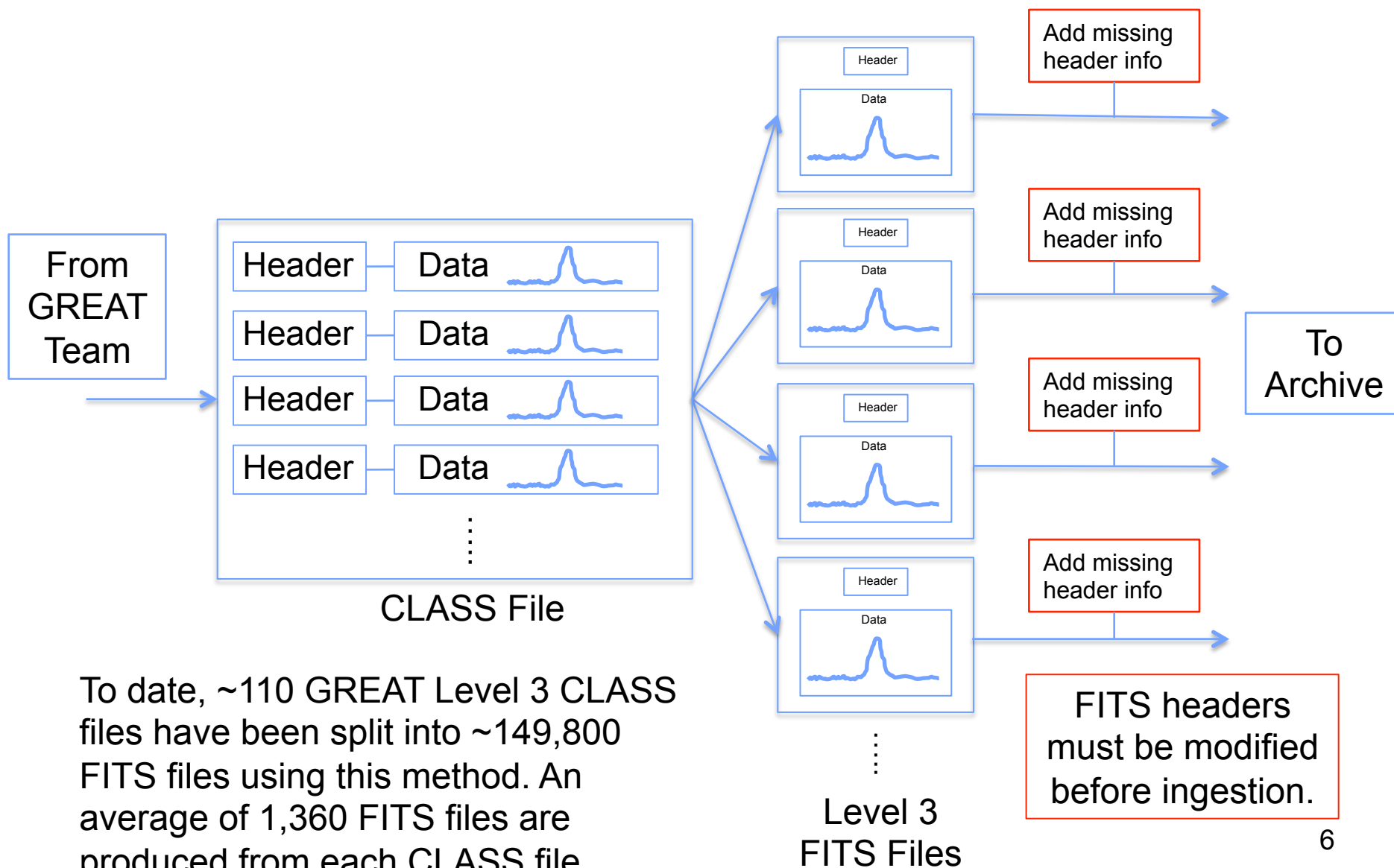
CLASS files can include:
- On-source spectra
- T_{rec} scans
- Atmospheric transmission spectra
- Model fits to atmospheric transmission
- Heterogeneous data—multiple frequencies, backends

GREAT and CLASS

Current Level 3 Archive Procedure

From GREAT Team

CLASS File
- Header — Data
- Header — Data
- Header — Data
- Header — Data
⋮

Level 3 FITS Files
- Header / Data — Add missing header info
- Header / Data — Add missing header info
- Header / Data — Add missing header info
- Header / Data — Add missing header info
⋮

To Archive

FITS headers must be modified before ingestion.

To date, ~110 GREAT Level 3 CLASS files have been split into ~149,800 FITS files using this method. An average of 1,360 FITS files are produced from each CLASS file.

6

Current GREAT Archive

DCS Archive Usage

2015-04-10 06:18:45 (UTC)

Archive Usage Summary

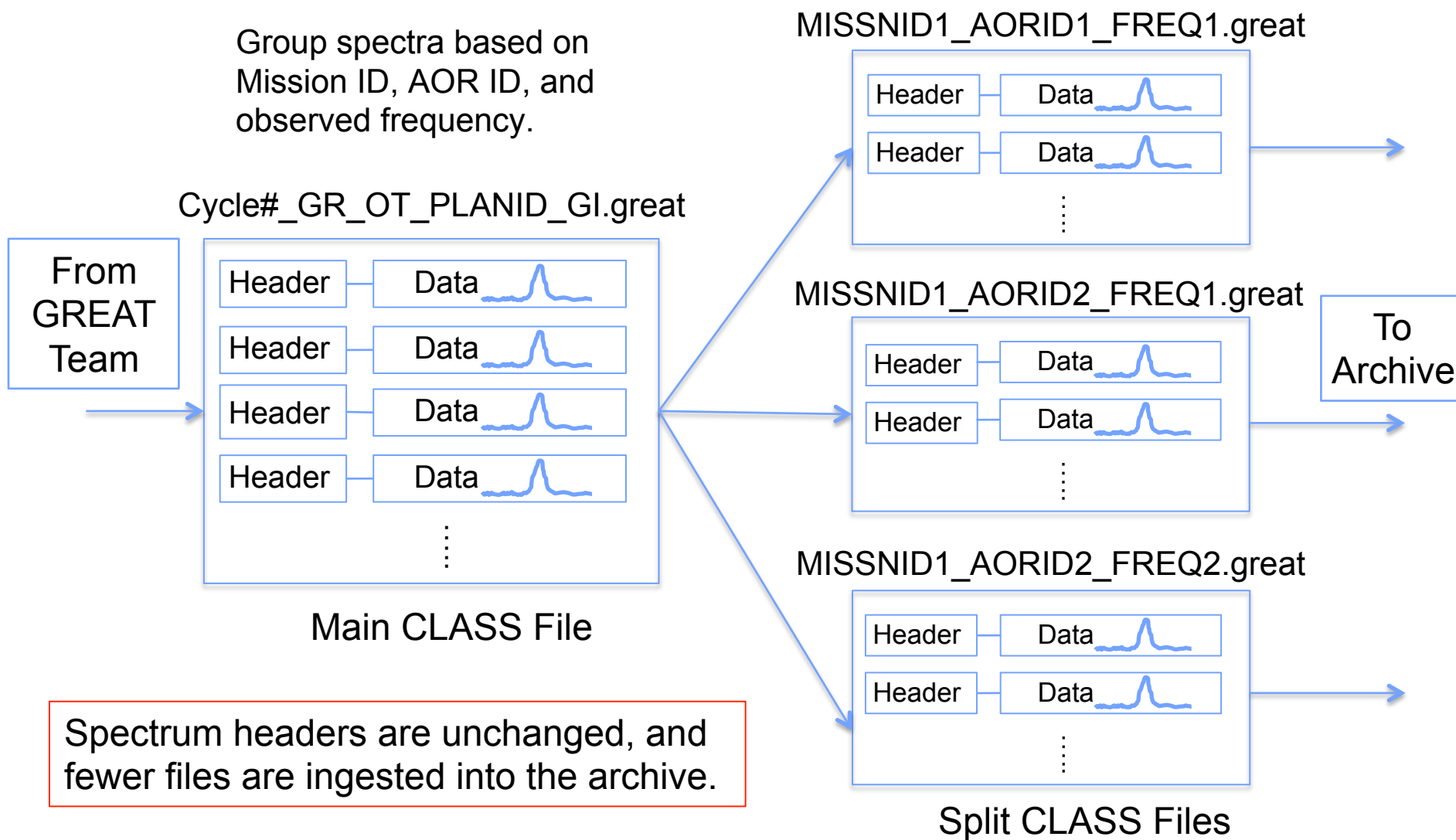|	File Type	Number Of Archived	Size of Archived (GB)	Number Of Downloads[*]	Size of Downloaded[*] (GB)
SI	EXES	890	171.84	121	36.56
	FIFI-LS	17,701	88.25	1,767	6
	FLITECAM	7,327	22.51	11,029	34.3
	FORCAST	87,510	289.82	58,586	134.5
	GREAT	309,478	51.6	8,623	1.07
	HIPO BLUE	5,987	38.44	265	3.08
	HIPO RED	8,081	65.69	296	3.41
	TOTAL	436,974	728.15	80,687	218.92
ARK		137,044	6,377.56	375	9.52
ANCI		75,581	1,281.73	298	5.39

[*] Collected since July 2013 (DCS 2.3.4)
[*] Repeated downloads on the same file are counted as multiple downloads.

In the last 8 months, 1 external user has downloaded GREAT data. Why is there such low utilization of the GREAT archive?

- Number of files?
- Data format?

Proposed Level 3 Archive Procedure for CLASS Files

Group spectra based on Mission ID, AOR ID, and observed frequency.

Cycle#_GR_OT_PLANID_GI.great

From GREAT Team

Main CLASS File

Spectrum headers are unchanged, and fewer files are ingested into the archive.

MISSNID1_AORID1_FREQ1.great

MISSNID1_AORID2_FREQ1.great

To Archive

MISSNID1_AORID2_FREQ2.great

Split CLASS Files

Example: GREAT OC2G

Observations for 7 projects were carried out in January 2015 over 3 flights.

There are two Level 3 CLASS files associated with each project (units of T_A^* and T_{MB}).

Number of archive files created from these 14 files using:

- current FITS procedure: 23,127 FITS files.
- proposed CLASS procedure: 58 CLASS files.

Archiving data in CLASS format results in many fewer data files and therefore a more user-friendly archive.

Summary

The current GREAT data archive is not being used.

The GREAT data archive would be improved by storing CLASS files rather than FITS files. Benefits include:

– No header translation required

– More manageable archive search results (fewer files)

– Files in a format of a widely used spectroscopy software

Questions for the SUG:

Does the astronomy community have enough familiarity with the CLASS file format that we could remove the FITS file requirement for the SOFIA archive?

Would doing so cause the data to only be used by spectroscopy experts?

Supplemental Slides

Ingesting CLASS Files—Use XML

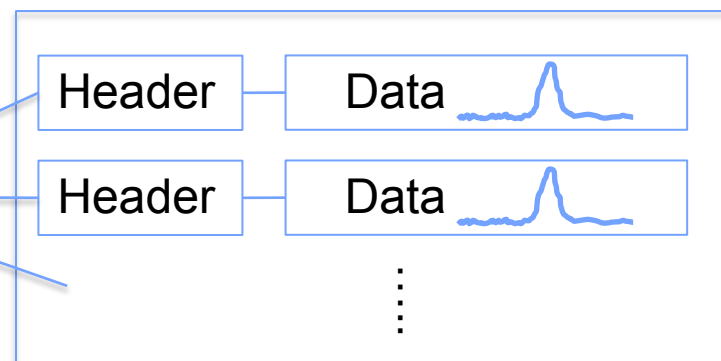XML Header File

```
<?xml version="1.0" encoding="utf-8"?>
<ClassHeader classfile="missnid_aorid_freq.great">
  <Keyword name='DATASRC' value='ASTRO' />
  <Keyword name='MISSN-ID' value='2015-01-22_GR_F188' />
  <Keyword name='DATE-OBS' value='2015-01-22T09:37:39' />
  <Keyword name='INSTRUME' value='GREAT' />
  <Keyword name='SPECTEL1' value='GRE_L2' />
  <Keyword name='SPECTEL2' value='NONE' />
  <Keyword name='OBS_ID' value='2015-01-22_GR_F188_GREAT_....'/>
  <Keyword name='PLANID' value='01_0001' />
  <Keyword name='AOR_ID' value='01_0001_1' />
    .
    .
    .
</ClassHeader>
```

Split CLASS File
MISSNID1_AORID1_FREQ1.great

Header — Data

Header — Data

New Archive
Ingestion
Module

To Archive

(file link)

12