BIG BANG MEETS BIG DATA

AST-06

SIP Interns: Sruti Vutukury & Ogni Goswami
Mentor: Dr. Guillermo Barro

What field do you see the tech industry moving towards over the next decade?

Filtering

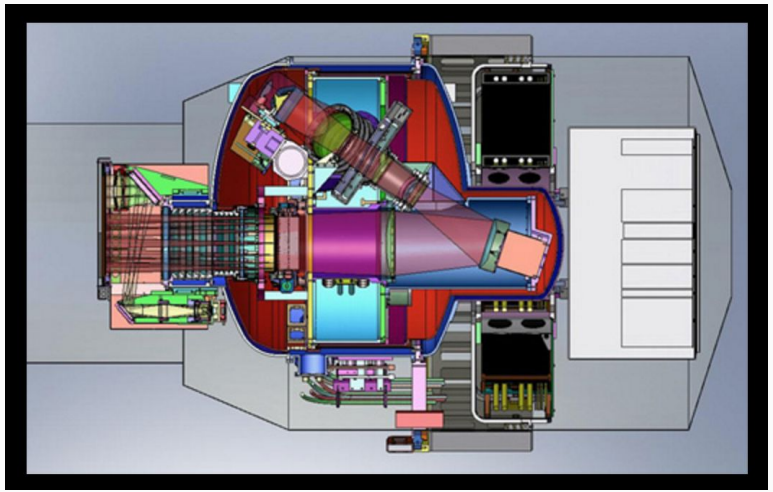
- To go through large amounts of material and extract things
- Examples

Big Data

- Storage usage per year of largest astronomical data center (bytes)?
- Square Kilometer Array: 100 petabytes/year
- UCSC UDS system
- 2025 SKA Projection
- Relative size

Keck Observatory

- Mauna Kea, Hawaii
- Contains most productive telescope instruments on Earth
- MOSFIRE and IR study



The Keck Archive:

A “dumpsite” for data...

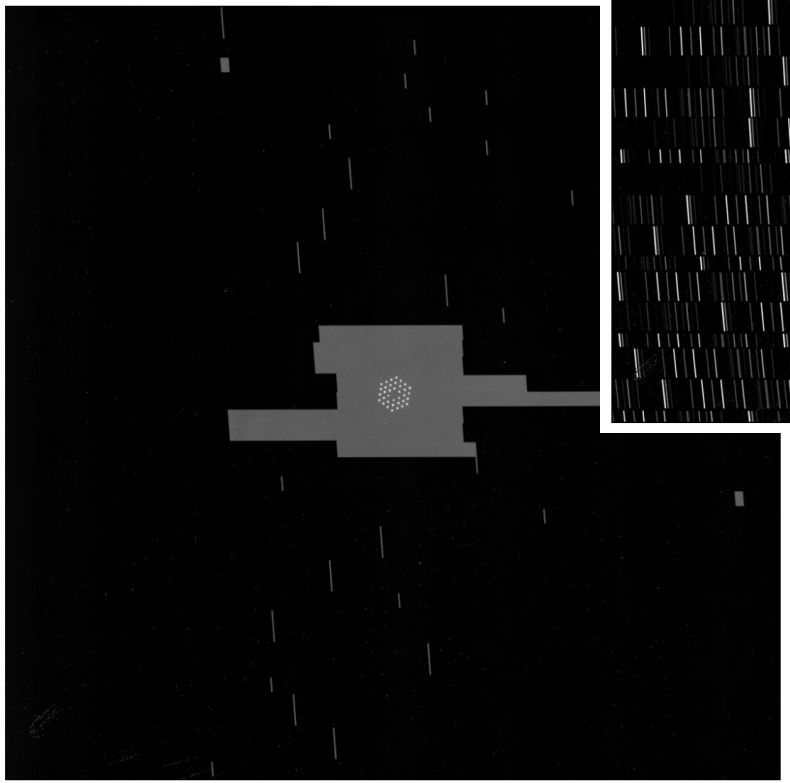
Row ID	KOAID	Instrument	Target Name	Associated Calibration Files	Frame Number	Quicklook Previews	RA (J2000)	Dec (J2000)	Observation Date (UT)	Observation Time (UT)	Exposure Time (sec)
1	MF.20120915.47826.fits	MOSFIRE	03323261-273754	MF.20120915.47826.caliblist	403	[Hdr] [Raw]	03:32:32.6	-27:37:54.1	2012-09-15	13:17:06.05	1.455
2	MF.20120915.47847.fits	MOSFIRE	03323261-273754	MF.20120915.47847.caliblist	404	[Hdr] [Raw]	03:32:32.6	-27:37:54.1	2012-09-15	13:17:27.45	1.455
3	MF.20120915.47899.fits	MOSFIRE	03323261-273754	MF.20120915.47899.caliblist	405	[Hdr] [Raw]	03:32:32.6	-27:37:54.1	2012-09-15	13:18:19.80	1.455
4	MF.20120915.47915.fits	MOSFIRE	03323261-273754	MF.20120915.47915.caliblist	406	[Hdr] [Raw]	03:32:32.6	-27:37:54.1	2012-09-15	13:18:35.00	1.455
5	MF.20120915.48223.fits	MOSFIRE	gds1209_H1	MF.20120915.48223.caliblist	407	[Hdr] [Raw]	03:32:29.6	-27:45:11.5	2012-09-15	13:23:43.10	8.729
6	MF.20120915.48264.fits	MOSFIRE	gds1209_H1	MF.20120915.48264.caliblist	408	[Hdr] [Raw]	03:32:28.8	-27:45:11.5	2012-09-15	13:24:24.10	8.729
7	MF.20120915.48311.fits	MOSFIRE	gds1209_H1	MF.20120915.48311.caliblist	409	[Hdr] [Raw]	03:32:28.8	-27:45:10.9	2012-09-15	13:25:11.96	8.729
8	MF.20120915.48517.fits	MOSFIRE	gds1209_H1	MF.20120915.48517.caliblist	410	[Hdr] [Raw]	03:32:28.8	-27:45:09.0	2012-09-15	13:28:37.26	119.293
9	MF.20120915.48668.fits	MOSFIRE	gds1209_H1	MF.20120915.48668.caliblist	411	[Hdr] [Raw]	03:32:28.9	-27:45:12.9	2012-09-15	13:31:08.92	119.293
10	MF.20120915.48822.fits	MOSFIRE	gds1209_H1	MF.20120915.48822.caliblist	412	[Hdr] [Raw]	03:32:28.8	-27:45:09.0	2012-09-15	13:33:42.77	119.293
11	MF.20120915.48975.fits	MOSFIRE	gds1209_H1	MF.20120915.48975.caliblist	413	[Hdr] [Raw]	03:32:28.9	-27:45:12.9	2012-09-15	13:36:15.42	119.293
12	MF.20120915.49128.fits	MOSFIRE	gds1209_H1	MF.20120915.49128.caliblist	414	[Hdr] [Raw]	03:32:28.8	-27:45:09.0	2012-09-15	13:38:48.12	119.293
13	MF.20120915.49281.fits	MOSFIRE	gds1209_H1	MF.20120915.49281.caliblist	415	[Hdr] [Raw]	03:32:28.9	-27:45:12.9	2012-09-15	13:41:21.77	119.293

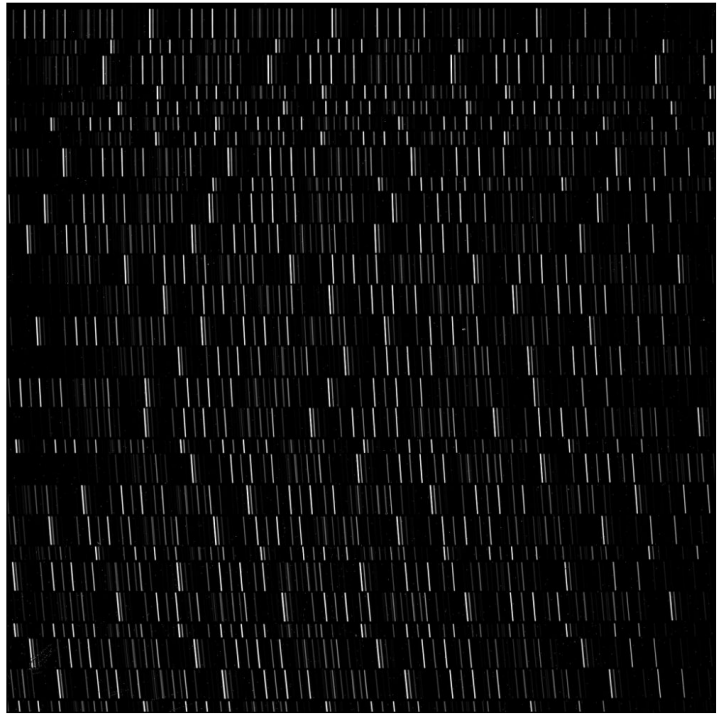
\ \fixlen = T\RowsRetrieved = 3008\ \ |
semid| propmin| prop|| char| char| koaid| in
ch
null| null| null| null| null|
ect 403 53.13591 -27.63169 2012-09-15 13:17:06.05 1.4547900000 2000.0000000000
9 2012-09-15 13:17:27.45 1.4547900000 2000.0000000000 U048M
000000 U048M UCSC Faber
UCSC Faber
Faber
A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies
A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies /koadata16/
/koadata16/MOSFIRE/20120915/lev0/MF.20120915.48517.fits 2012b_u048m
adata16/MOSFIRE/20120915/lev0/MF.20120915.48668.fits 2012b_u048m 18 f MF.20120915.48822.fits MOSFIR
18 f MF.20120915.48975.fits MOSFIRE gds1209_H1
MOSFIRE gds1209_H1 object
object 415 53.12021 -27.75359 2012-09-15
416 53.12006 -27.75249 2012-09-15 13:43:58.17 119.2927800000 2000.0000000000
15 13:46:28.13 119.2927800000 2000.0000000000 U048M UCSC
U048M UCSC Faber

How it looks on the Keck Archive website

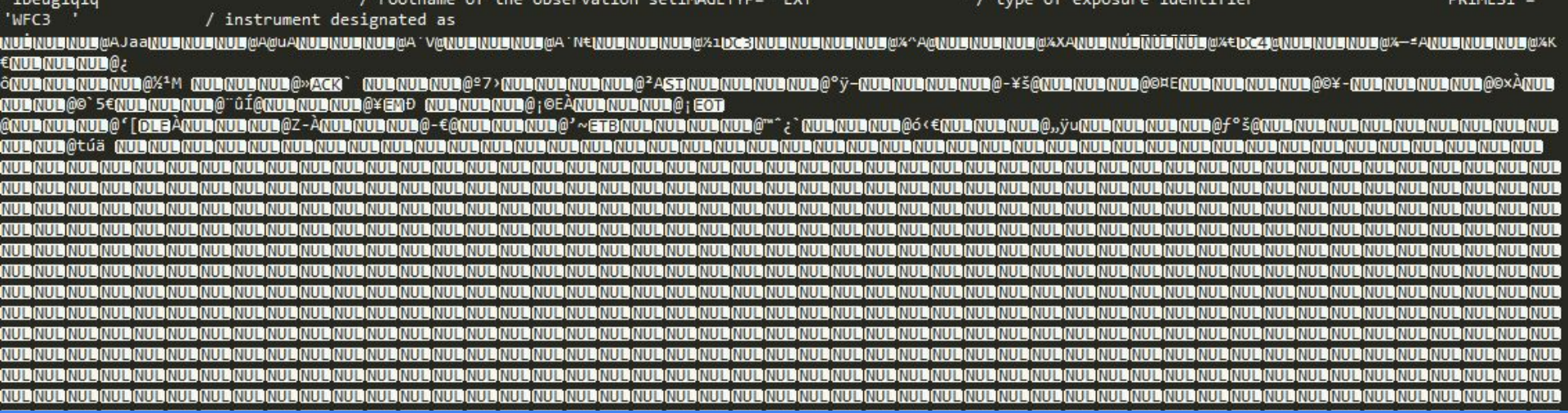
MF.20120915.47826.fits Header

```
SIMPLE  =                    T / C# FITS: 9/15/2012 3:17:13 AM
BITPIX  =                  -32
NAXIS   =                    2 / Dimensionality
NAXIS1  =                 2048
NAXIS2  =                 2048
LONGSTRN= 'OGIP 1.0'           / The OGIP Long String Convention may be used.
EXTEND  =                    T / Extensions are permitted
BUNIT   = 'ADU per coadd'      / Units of the pixel value
TIME-OBS= '13:17:05.953'       / UT start time of exposure
TIME-END= '13:17:10.921'       / UT end time of exposure (after last read)
ITIME0  =              2000000 / Requested itime per coadd
COADDS0 =                    1 / Requested number of coaddi
SMPMODE0=                    2 / Requested samp mode (1:Sir
READS0  =                    1 / Requested sampling mode re
TRUITIME=              1.45479 / True integration time in s
RMPRESET=                    1 / Number of resets in ramp
RMPREADS=                    1 / Number of reads in ramp
RMPDROP =                    0 / Number of drop frames in r
RMPGROUP=                    2 / Number of groups in ramp
RDITIME =              1.45479 / Time between sampling grou
ABORTED =                    F / Exposure was stopped befor
COADDONE=                    1 / Number of coadds actually
READDONE=                    2 / Reads actually completed i
GRPSDONE=                    2 / Groups actually completed
IGNORDRD=                    0 / Number of discarded initia
VRPSTYPE= 'channel '           / Vertical reference pixel s
HRPSTYPE= 'none    '           / Horizontal reference pixel
SATURATE=                33000 / Detector saturation level
DTGAINNM=                    9 / Detector gain setting [0-1
DTGAINDB= '15.05 db (low)'     / Detector gain in dB
DETGAIN =              5.65685 / Detector gain factor
SYSGAIN =                 2.15 / System gain in e-/ADU
VRESET  =               0.3011 / Reset voltage in V, obtair
DSUB    =               0.6002 / Substrate voltage in V, ob
VBIASGAT=               2.0507 / Gate voltage of SCA in V,
VBIASPWR=                  3.3 / Source voltage of SCA in V
```





```
SIMPLE  =                    T / conforms to FITS standard                      BITPIX  =                  -64 / array data type
NAXIS   =                    2 / number of array dimensions                     NAXIS1  =                  360
NAXIS2  =                  432                                                  OBJECT  = 'GS_ALL_CANDELS_ERS_UDF_F160W_V0.5_EXP[1/1]'
ORIGIN  = 'KPNO-IRAF'           /                                               DATE    = '2012-02-28T11:03:50'
IRAFNAME= 'gs_all_candels_ers_udf_f160w_v0.5_exp.hhh'  /  NAME OF IRAF IMAGE FILIRAF-MAX=          0.000000E0  /  DATA MAX
IRAF-MIN=           0.000000E0  /  DATA MIN                                     IRAF-BPX=                  32  /  DATA BITS/PIXEL
IRAFTYPE= 'REAL    '            /  PIXEL TYPE                                   DATAMIN =                  0.
DATAMAX =                    0.                                                 CRPIX1  =              184.01
CRVAL1  =             53.122751                                                 CTYPE1  = 'RA---TAN'
CD1_1   =         -0.0008333335                                                 CD2_1   =                  0.
CRPIX2  =                220.01                                                 CRVAL2  =           -27.805089
CTYPE2  = 'DEC--TAN'                                                           CD1_2   =                  0.
CD2_2   =          0.0008333335                                                ORIGIN  = 'NOAO-IRAF FITS Image Kernel July 2003' / FITS file originator
DATE    = '2012-02-07T13:22:02'                                                NEXTEND =                  5 / Number of standard extensions
FILETYPE= 'SCI       '          / type of data found in data file
TELESCOP= 'HST'                 / telescope used to acquire data               INSTRUME= 'WFC3  '             / identifier for instrument used to acquire data
EQUINOX =                2000.0 / equinox of celestial coord.
system                                                                                   / DATA DESCRIPTION
KEYWORDS                                                                                                   ROOTNAME=
'ibeuglqlq                      ' / rootname of the observation setIMAGETYP= 'EXT             ' / type of exposure identifier                  PRIMESI =
'WFC3  '              / instrument designated as
```



How the downloaded files look

The Problem

- Read data in Keck archive (automatically)
- Categorize and interpretable
- “Big Picture”

The Solution: Over 600 lines of code

Main Features

- Search for Essential Database Information (Project Investigator, Project, Mask, Filters, Exposure Times)
- Develop Summary Tables
- Object Categorization
- Plot mask footprints in field of interest

Essential Database Information: PI's and Projects

- Decodes ASCII file
- Gets all essential PI and project info out of the 3000+ entries
- Easily searches through database and presents relevant information

```
URL of ASCII table off Keck Archive
->https://tinyurl.com/astropy
Press 1 for information about PIs:
Press 2 for information about projects:
Press 3 for information about specific mask/target:
Press 4 to generate a summary table:
Press 5 for Object Information or to Search:
Press 0 to Quit:
->1
Press 1 for a list of PIs
Press 2 for the number of PIs
Press 5 to quit:
->2
22
Press 1 for a list of PIs
Press 2 for the number of PIs
Press 5 to quit:
->1
['Ellis', 'Kriek', 'Chu', 'Dale.Kocevski', 'Mobasher', 'Casey', 'Siana', 'Illingworth', 'Tanaka', 'Civano', 'L
.Cowie', 'Urry', 'Faber', 'Vieira', 'Hasinger', 'M.Dickinson', 'Scoville', 'vanDokkum', 'Shapley', 'Kartaltepe
', 'Marchesini', 'Cooray']
Press 1 for a list of the projects under a certain PI
Press 5 to quit:
->1
Enter the name of the PI whose projects you want listed: Faber
['Spatial Maps of SFR, Kinematics, and Line Ratios of z~1.5 Galaxies from Deep Multi-PA / Multi-Slit MOSFIRE S
pectroscopy', 'A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies']
Press 1 for the number of projects under this PI
Press 5 to quit:
->1
2
Press 1 for a list of PIs
Press 2 for the number of PIs
Press 5 to quit:
->5
```

Read FITS Files

- Decodes ASCII file
- Gets all essential info out of the 3000+ entries
- Reads FITS Files (option) and gets filters and exposure times info
- Presents summarized information in a table

```
Enter the name of the specific mask(target name) of which you want to know the total exposure time: CDFs
26347.70215 seconds
```

```
Press 2 to generate summary table WITHOUT filter info
Press 3 to generate summary table based on files downloaded WITH filter info
->3
Initializing
Reading 1 of 10 files: MF.20120916.47922.fits.gz
Reading 2 of 10 files: MF.20120915.47826.fits.gz
Reading 3 of 10 files: MF.20141106.33188.fits.gz
Reading 4 of 10 files: MF.20121223.34210.fits.gz
Reading 5 of 10 files: MF.20130117.30396.fits.gz
Reading 6 of 10 files: MF.20131014.40824.fits.gz
Reading 7 of 10 files: MF.20140102.27621.fits.gz
Reading 8 of 10 files: mosfire_DRP.log
Reading 9 of 10 files: data.txt
Reading 10 of 10 files: MASTER.py


Saved

   Date          PI                              ProjectName                                  Mask/TargetName Filter ExposureTime(sec)
---------- ------------ ------------------------------------------------------------ --------------- ------ -----------------
2012-09-15        Faber           A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies 03323261-273754      J       00:00:5.172
2012-09-16        Faber           A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies 03323261-273754      J       00:00:4.968
2012-12-23      Shapley                                The MOSDEF Survey: A Pilot Program    goodss_z2_01      K      00:03:25.422
2013-01-17     Scoville    Spectroscopic Confirmation of 7.5&amp;lt;z&amp;lt;10 Ly-alpha  HighzEllis_new      H      00:02:24.922
2013-10-14 Dale.Kocevski                A MOSFIRE Survey of z~2 Emission-Line Galaxies GDS_H1_wphotoz2      H       00:00:4.968
2014-01-02  Illingworth Lyman Alpha Emission at z~8: Probing the Ionization State of the       Y-CDFSYD1      Y      00:03:24.36
2014-11-06     Mobasher  Spectroscopy at z &amp;gt; 7 in the Hubble Frontier Fields: TMT      CDFS-kband      J      00:03:24.125
```

PI	Project Name	Mask/Target Name	Exposure Time (seconds)
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	03323261-273754	36.3697
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	gds1209_H1	5871.53
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	gds1209_H2	5990.83
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	gds1210_H1	7183.75
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	gds1210_K1	13935.4
Faber	A MOSFiRE Survey of Emission Lines in z ~ 2 Galaxies	gds1210_K3	5490.38

Object Categorization

- Decodes FITS files to get info of objects within that file
- Categorizes objects by mask
- Creates neat tables and excel files conveniently named after the mask

```
        Mask                    File                  Object  Right Acension Declination
---------------------- ------------------------ -------- -------------- -----------
GDS_H1_wphotoz23_1_wSusan2 MF.20131014.40824.fits.gz     9245     53.09391667 27.82572222
GDS_H1_wphotoz23_1_wSusan2 MF.20131014.40824.fits.gz    12473      53.119875 27.79873889
     Mask                File                  Object       Right Acension Declination
------------- ------------------------ --------------- -------------- -----------
HighzEllis_new MF.20130117.30396.fits.gz       CDFS21724H         53.1485 27.71948889
HighzEllis_new MF.20130117.30396.fits.gz       CDFS25094H     53.16808333 27.72350833

    Mask                File                  Object  Right Acension Declination
------------ ------------------------ -------- -------------- -----------
 gds1209_H2 MF.20120916.47922.fits.gz    em9767       53.194125     27.7333
 gds1209_H2 MF.20120916.47922.fits.gz    nug8158     53.17458333   27.753375
```

Name	Date modified	Type	Size
gds1212_H1_Mask_Objects.csv	8/8/2016 12:10 PM	Microsoft Excel C...	
goodss_z2_01_Mask_Objects.csv	8/8/2016 12:10 PM	Microsoft Excel C...	
HighzEllis_new_Mask_Objects.csv	8/8/2016 12:10 PM	Microsoft Excel C...	
gds1210_K1_Mask_Objects.csv	8/8/2016 12:10 PM	Microsoft Excel C...	
CDFS_Bdrops_m1_wstars_Mask_Objects.c...	8/8/2016 12:10 PM	Microsoft Excel C...	
HighzEllis_Mask_Objects.csv	8/8/2016 12:10 PM	Microsoft Excel C...	

gds1212_H1	MF.20130101.18558.fits.gz	nug25952	53.12113	27.69808
gds1212_H1	MF.20130101.18558.fits.gz	c2_18258	53.12958	27.70142
gds1212_H1	MF.20130101.18558.fits.gz	em12278	53.16042	27.70775
gds1212_H1	MF.20130101.18558.fits.gz	nug22883	53.14217	27.70743
gds1212_H1	MF.20130101.18558.fits.gz	nug23382	53.16229	27.71214
gds1212_H1	MF.20130101.18558.fits.gz	em12285	53.11004	27.70786
gds1212_H1	MF.20130101.18558.fits.gz	c2_19227	53.13483	27.71336
gds1212_H1	MF.20130101.18558.fits.gz	em11534	53.12317	27.71558
gds1212_H1	MF.20130101.18558.fits.gz	nug22603	53.107	27.71823
gds1212_H1	MF.20130101.18558.fits.gz	em10817	53.12283	27.72281

Object Search

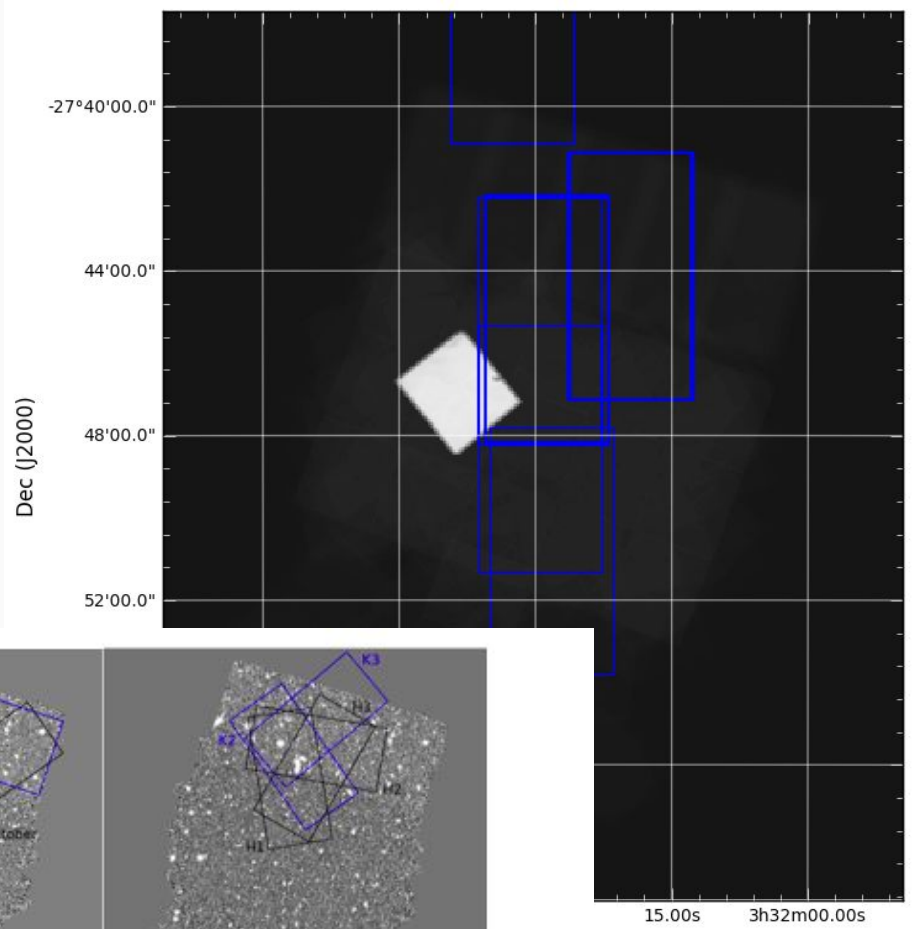
- Using object information from previous slide, runs searches
- Can search for any aspect of a FITS file (file name, mask name, object name, RA/DEC (coordinates))
- Creates neat tables and excel files conveniently named after the search query

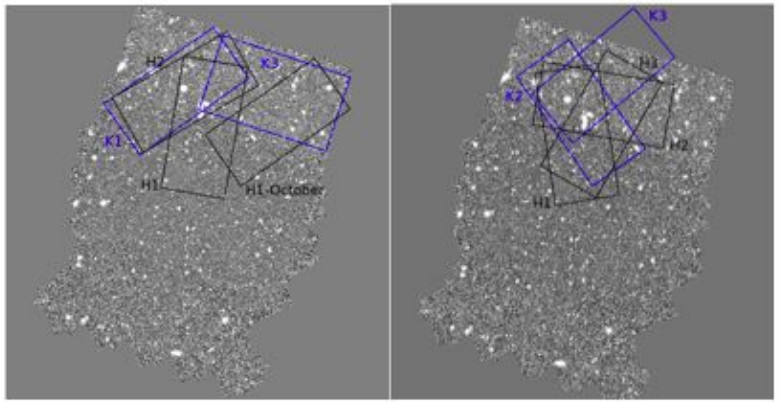
```
Do you want to search for a specific item? Press 1 for Yes. Press 2 for No: 1

Search: 53.1485

     Mask                File                  Object    Right Acension Declination
------------- ---------------------- ---------- -------------- -----------
HighzEllis_new MF.20130117.30396.fits.gz CDFS21724H        53.1485 27.71948889
    CDFS-kband MF.20141106.33188.fits.gz   BBG_6248        53.1485 27.77728611

Do you want to save this table? Press 1 for Yes. Press 2 for No: 1

Files Saved
```

CDFS-kband	MF.20141106.33188.fits.gz	BBG_17749	53.19688	27.76045
CDFS-kband	MF.20141106.33188.fits.gz	BBG_6357	53.17079	27.77541
CDFS-kband	MF.20141106.33188.fits.gz	BBG_16671	53.19017	27.76914
CDFS-kband	MF.20141106.33188.fits.gz	GS_LBG_2	53.15813	27.78641
CDFS-kband	MF.20141106.33188.fits.gz	GS_L700G_18	53.18879	27.77709
CDFS-kband	MF.20141106.33188.fits.gz	M_1431	53.16508	27.79366

Mask Footprints Plot

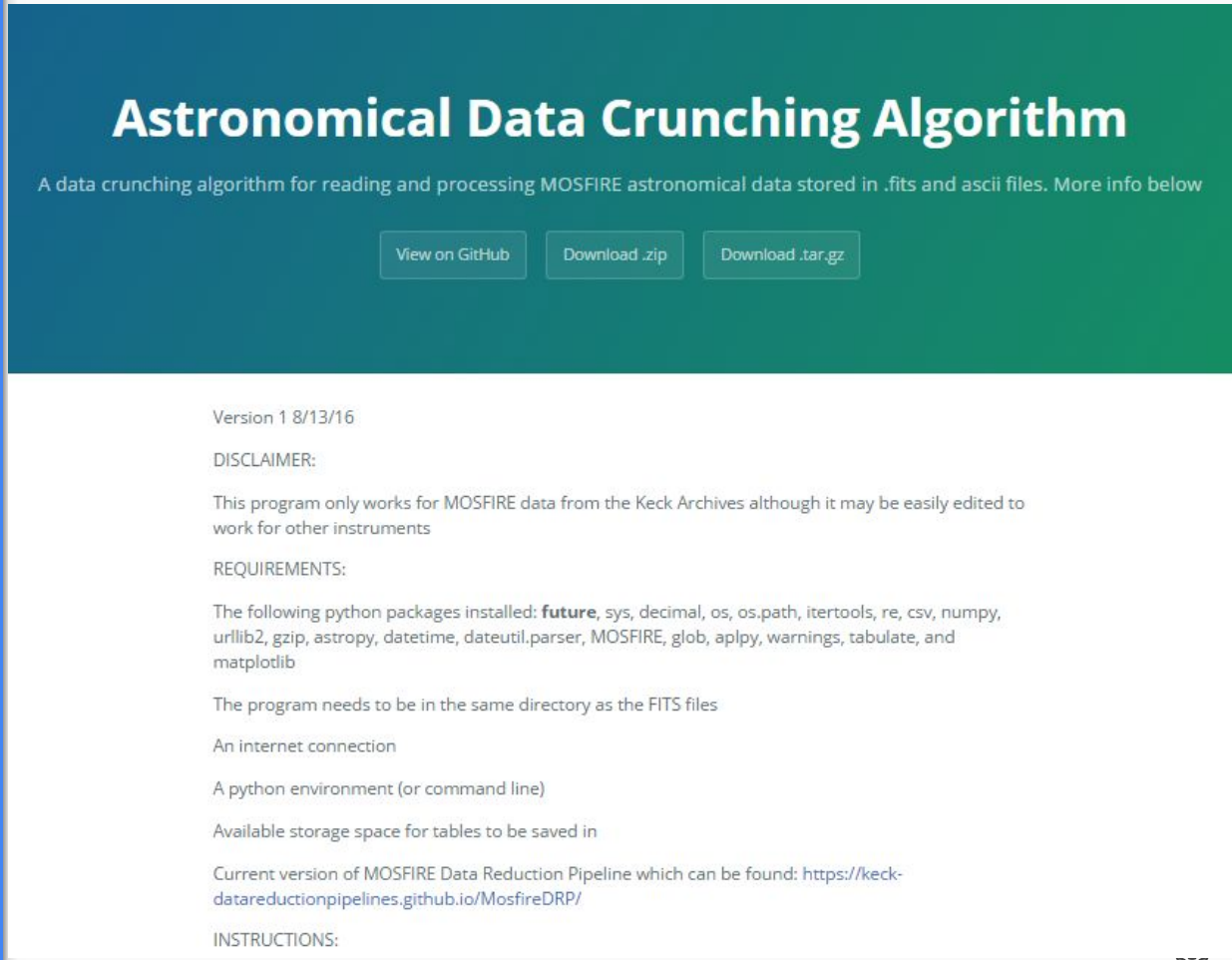
- Reads FITS headers for coordinates of masks
- Generates a plot of mask footprints in field of interest
- Other aesthetic features can be added:
  - colorscale
  - 3-color background
  - field contours





Publishing:

- Available to everyone for free on GitHub
- To look at the “big picture” of various Keck Observatory surveys
- Download at http://tinyurl.com/astrodatacrunching

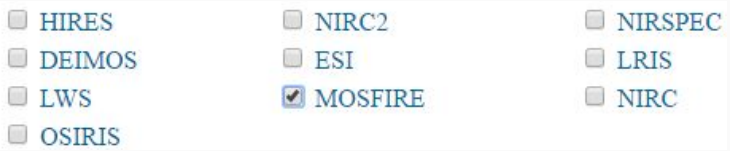


Future Work

- Add a GUI

```
Do you want to save this table? Press 1 for Yes. Press 2 for No: 2
```

- Work for instruments other than MOSFIRE
- Make the program more “flexible”




```
['Ellis', 'Kriek', 'Chu', 'Dale.Kocevski', 'Mobasher', 'Casey',
.Cowie', 'Urry', 'Faber', 'Vieira', 'Hasinger', 'M.Dickinson',
', 'Marchesini', 'Cooray']
Press 1 for a list of the projects under a certain PI
Press 5 to quit:
->1
Enter the name of the PI whose projects you want listed: Fabre
That PI is not part of this database. Please try again
```

Acknowledgements

- Dr. Guillermo Barro
- Dr. Raja GuhaThakurta
- Keck Observatory Archive
- SIP interns and families

Questions?