An hierarchical approach to Big Data

Hierarchical Progressive Surveys (HiPS)
and Multi-Order Coverage (MOC) Maps

Mark G. Allen (CDS)
Pierre Fernique (CDS)
Thomas Boch (CDS)
Daniel Durand (CADC)
Anais Oberto (CDS)
Bruno Merin (ESA)
Felix Stoehr (ALMA)
Francoise Genova (CDS)
Francois-Xavier Pineau (CDS)
Jesus Salgado (ESA)

Introduction

- Scientific exploitation of Big Data requires practical solutions for:
  - Data Access
  - Visualisation
  - Analysis
- Hierarchical sky tessellation techniques
  - Working now!
  - Scalable to Big Data

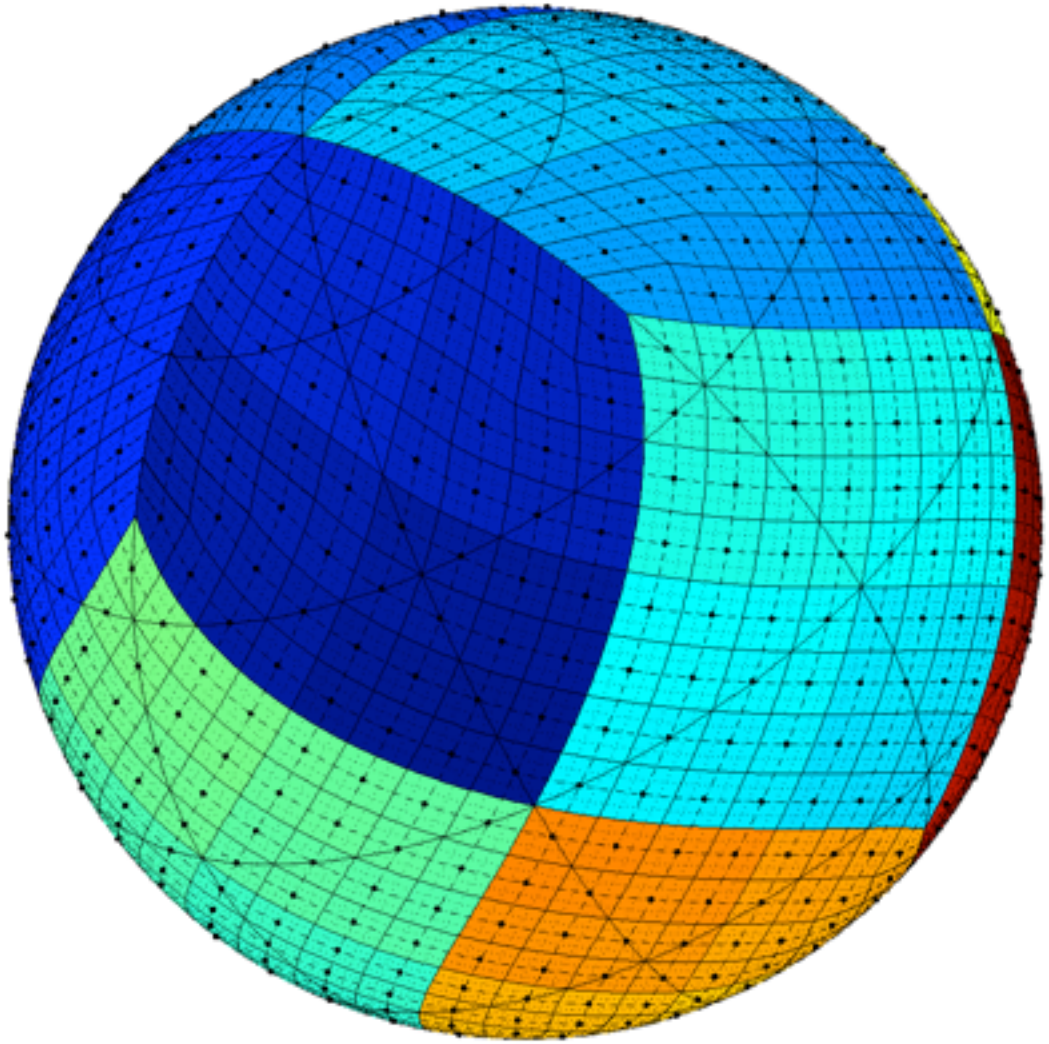
Introduction

- Starting points
  - Hundreds of image surveys, and thousands of catalogues at CDS
    - ~2 TB per band for all-sky surveys
    - Catalogues up to ~2 billion rows
  - Need for interactive display/comparison
  - Need to maintain scientific integrity of data, and links to original data
  - Ease of use and interoperability

HiPS

- HiPS: **Hierarchical Progressive Surveys**
  - Multi-resolution HEALPix data structure for
    - *Images*
    - *Catalogues*
    - *3-dimensional data cubes*
  - Conserves scientific data properties alongside visualisation considerations
  - Implemented for ~250 data sets and growing
  - New levels of interoperability - *images, catalogues, cubes, coverage maps*

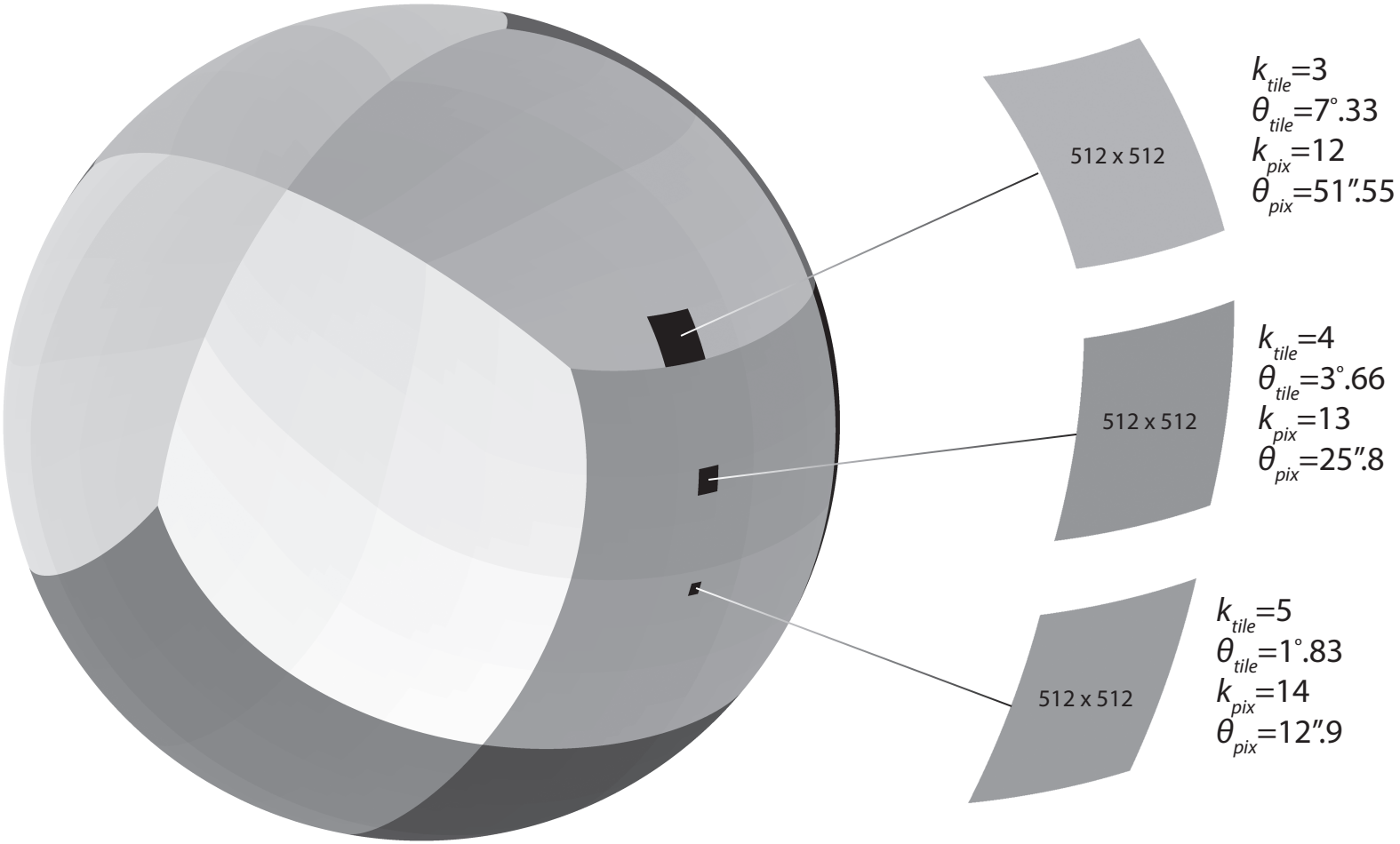
HEALPix*



- 12 quadrilateral pixels
- 2x2 division at each level
- Equal area
- Iso-latitude
- Nested index scheme encodes inheritance
- Libraries

*Gorski et al. 2005

HiPS – Tiles and Pixels

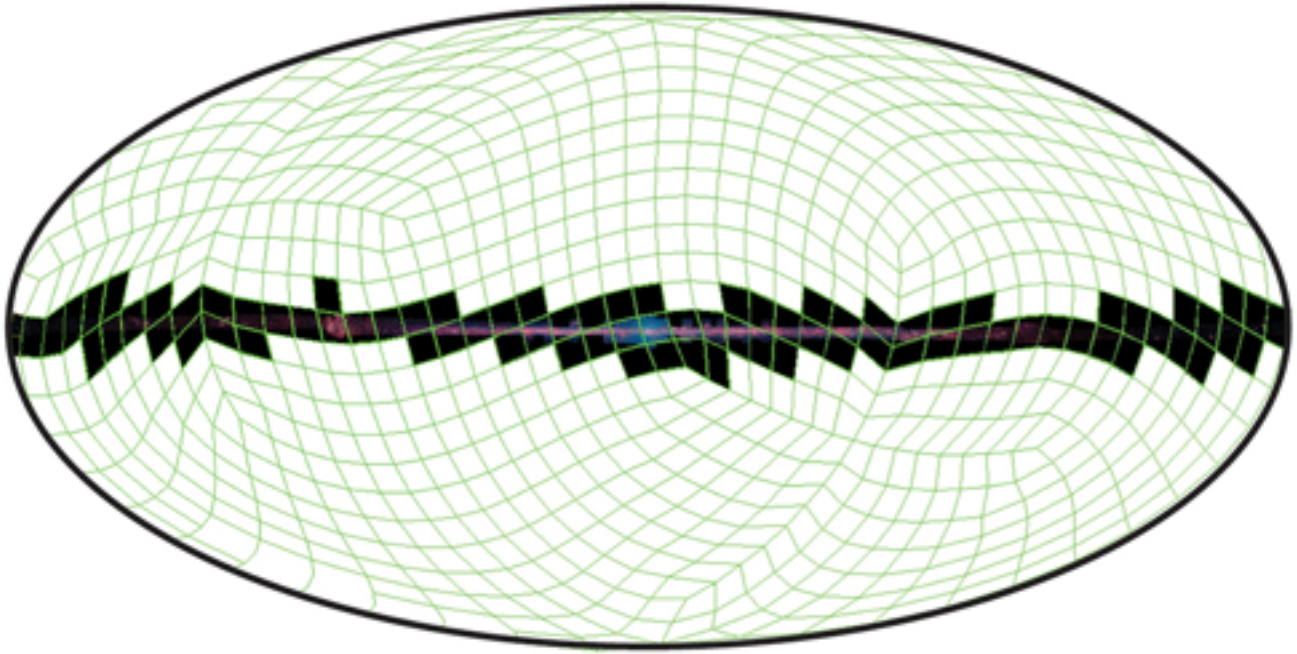


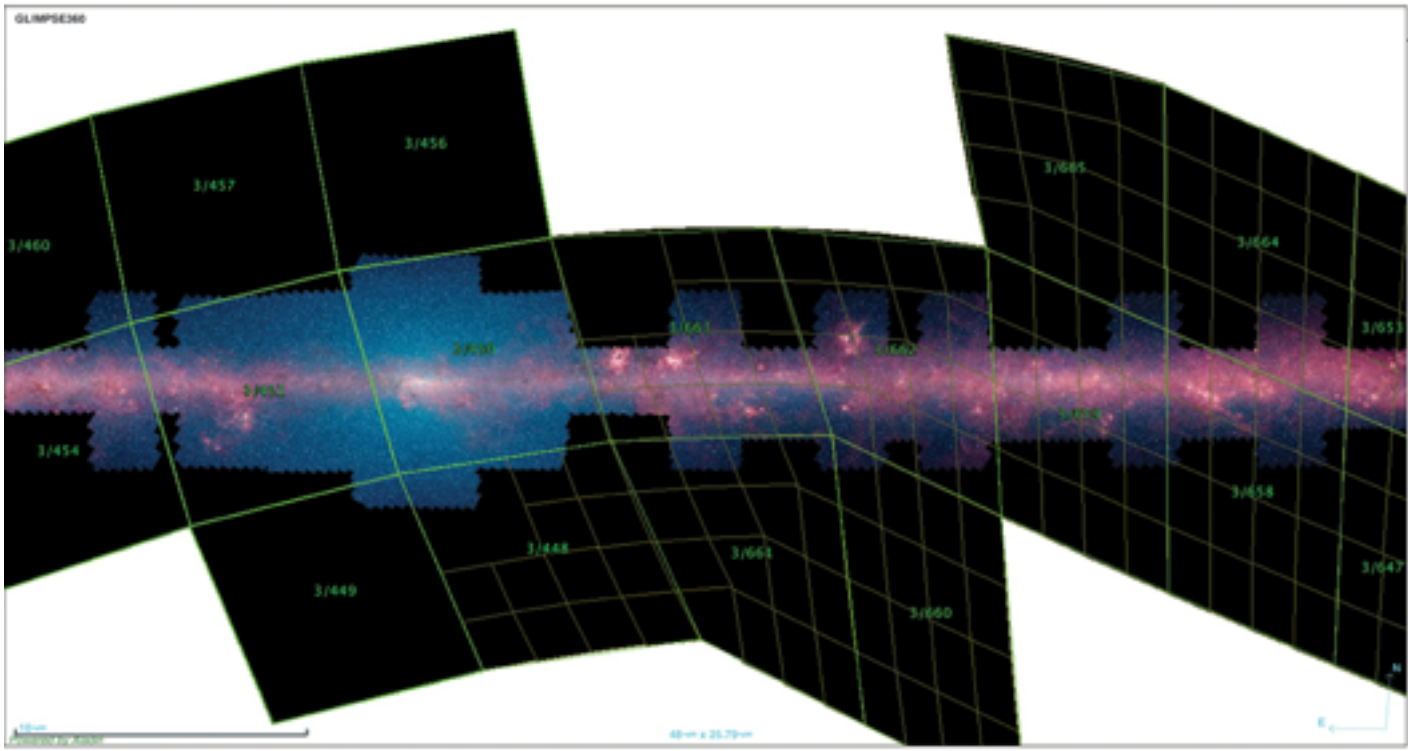
k	$N_{side} = 2^k$	N_{pix}	θ_{pix}	$k_{tile,512}$	$N_{tile,512}$	$\theta_{tile,512}$
0	1	12	58.°6			
1	2	48	29.°3			
2	4	192	14.°7			
3	8	768	7.°33			
4	16	3072	3.°66			
5	32	12,288	1.°83			
6	64	49,152	55.′0			
7	128	196,608	27.′5			
8	256	786,432	13.′7			
9	512	3,145,728	6.′87	0	12	58.°6
10	1024	12,582,912	3.′44	1	48	29.°3
11	2048	50,331,648	1.′72	2	192	14.°7
12	4096	201,326,592	51.″5	3	768	7.°33
13	8192	805,306,368	25.″8	4	3072	3.°66
14	2^{14}	3.22×10^9	12.″9	5	12288	1.°83
15	2^{15}	1.29×10^{10}	6.″44	6	49152	55.′0
16	2^{16}	5.15×10^{10}	3.″22	7	196608	27.′5
17	2^{17}	2.06×10^{11}	1.″61	8	786432	13.′7
18	2^{18}	8.25×10^{11}	0.″81	9	3,145,728	6.′87
19	2^{19}	3.30×10^{12}	0.″40	10	12,582,912	3.′44
20	2^{20}	1.32×10^{13}	0.″20	11	50,331,648	1.′72
21	2^{21}	5.28×10^{13}	0.″10	12	201,326,592	51.″5
22	2^{22}	2.11×10^{14}	50.3 mas	13	805,306,368	25.″8
23	2^{23}	8.44×10^{14}	25.1 mas	14	3.22×10^9	12.″9
24	2^{24}	3.38×10^{15}	12.6 mas	15	1.29×10^{10}	6.″44

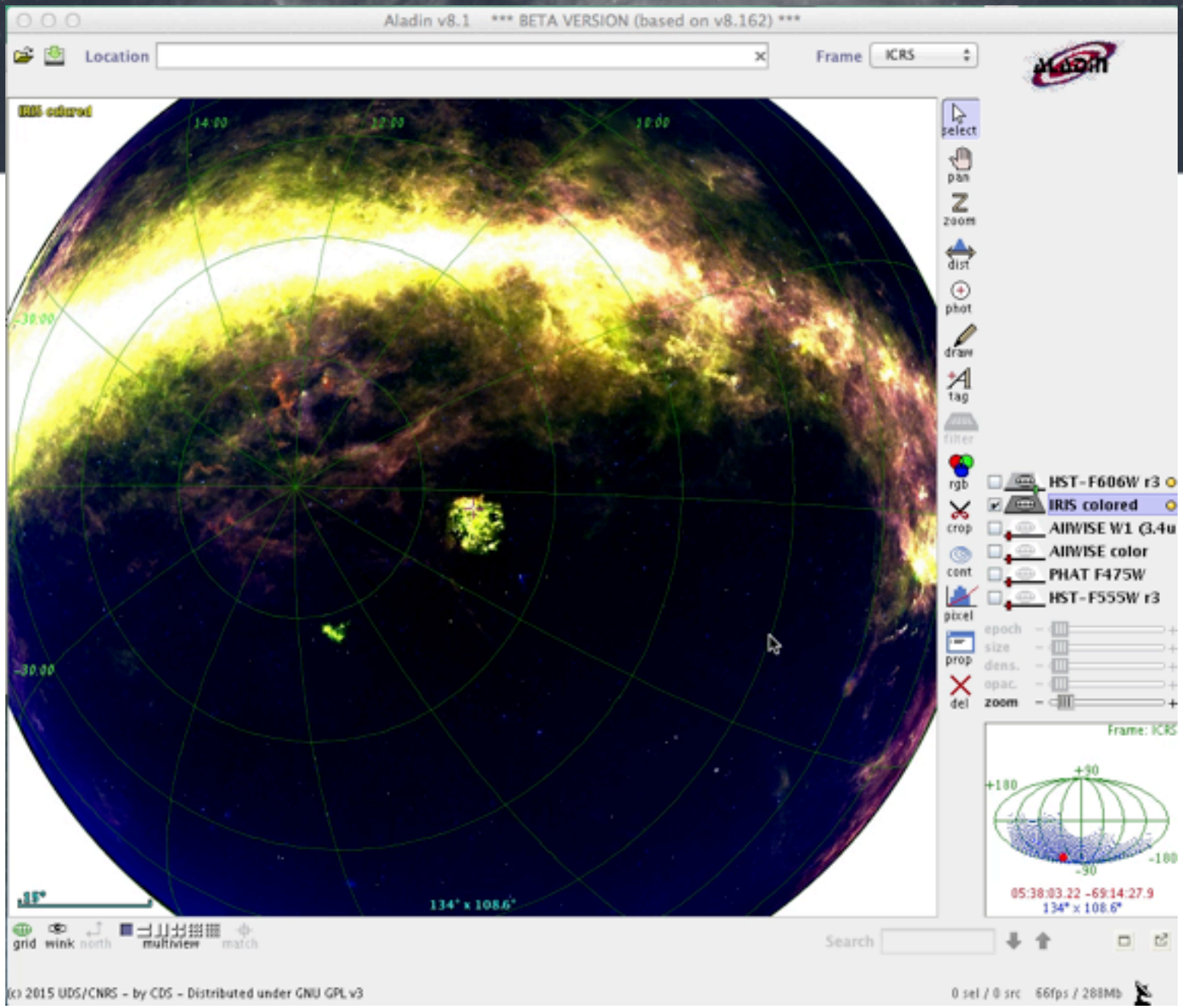
- WMAP
- PLANCK HFI
- IRAS
- NVSS
- SCUBA
- DSS
- SDSS
- CFHTLS
- HST ACS

e.g.

Spitzer GLIMPSE 360






Aladin v8.1 *** BETA VERSION (based on v8.162) ***
Location
Frame ICRS
IRIS colored
14 00
12 00
10 00
-30 00
-30 00
select
pan
zoom
dist
phot
draw
tag
filter
rgb
crop
cont
pixel
prop
del
HST-F606W r3
IRIS colored
AllWISE W1 (3.4u
AllWISE color
PHAT F475W
HST-F555W r3
epoch
size
dens.
opac.
zoom
Frame: ICRS
+90
+180
-180
-90
05:38:03.22 -69:14:27.9
134' x 108.6'
15'
134' x 108.6'
grid
wink
north
multiview
match
Search
(c) 2015 UDS/CNRS - by CDS - Distributed under GNU GPL v3
0 sel / 0 src
66fps / 288Mb

Going beyond zoom and pan

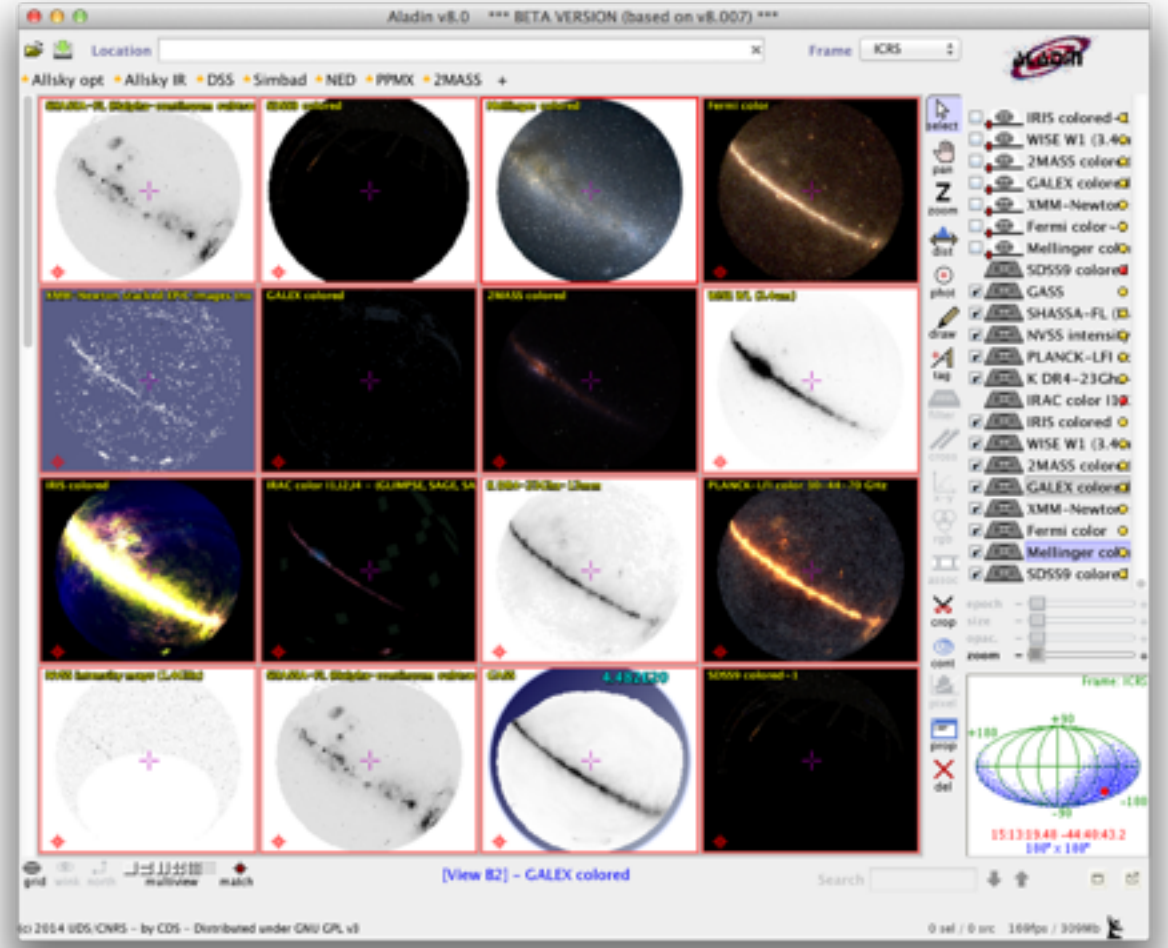
- Visualisation and preservation of science data
  - JPEG/PNG and FITS (preserves dynamic range)
  - Link mechanism to original data
  - HEALPix is scientifically robust

- Ease of implementation (generation & publication)
  - No databases or servers, just HTTP
  - Leads to fast take-up and innovation

```
java -Xmx16000m -jar Aladin.jar -hipsgen in=Fits_directory
```

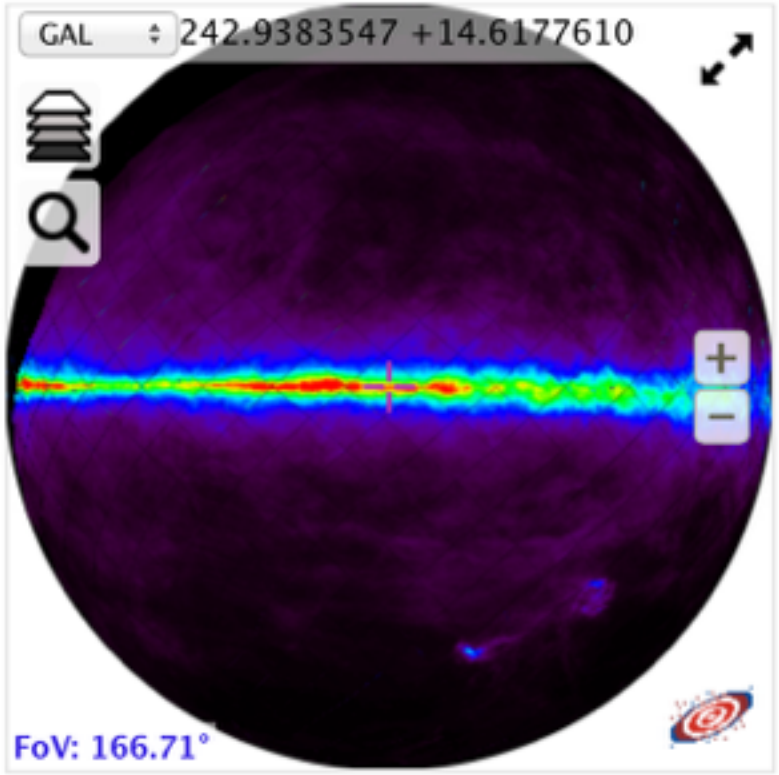
HiPS Image Data Sets

- **250+** HiPS
- ~ 50 TB
- Providers: **9**
- 4+ clients: Aladin, Aladin-lite, Mizar (CNES), MAST Discovery Portal

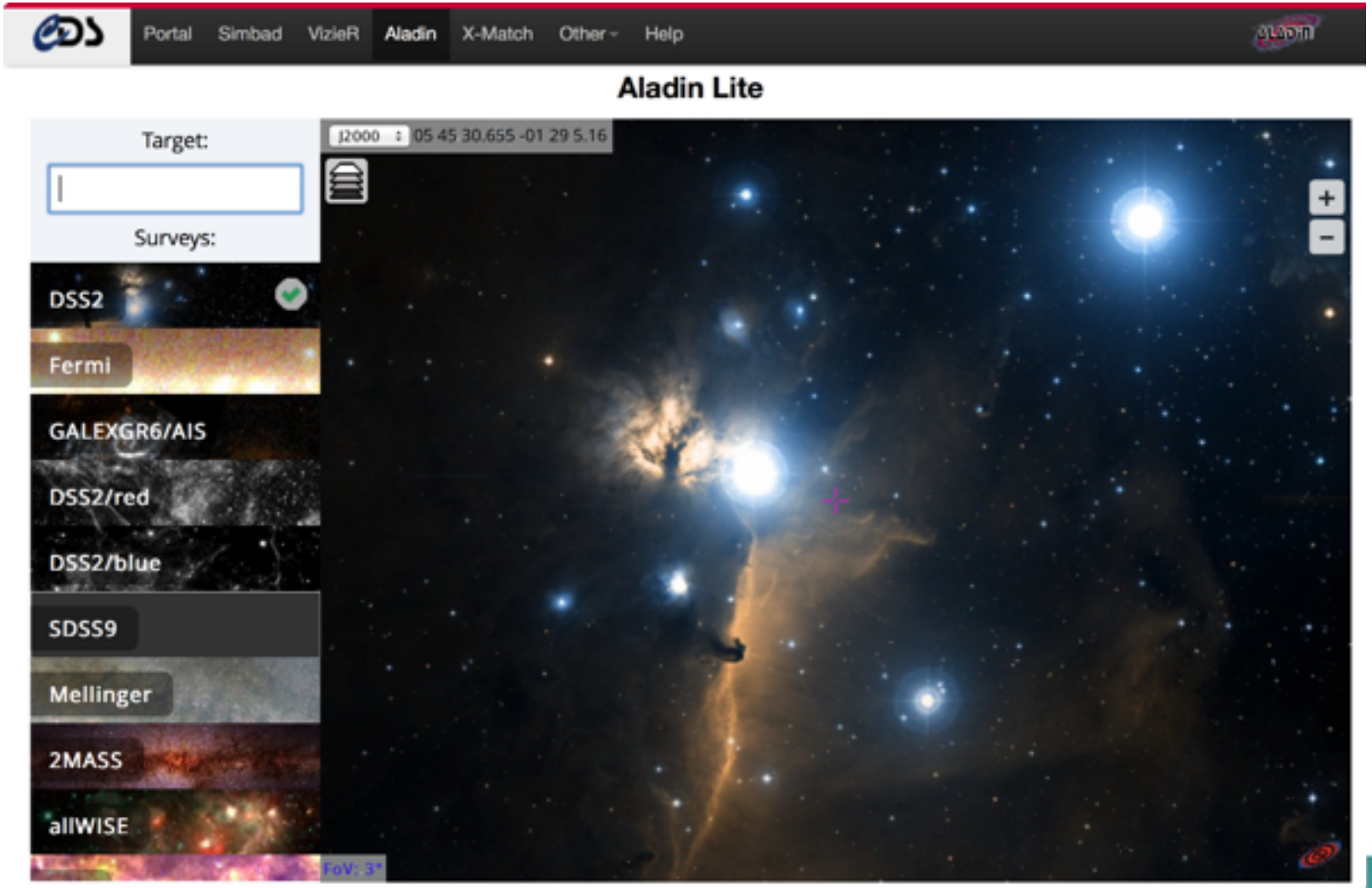


Aladin Lite

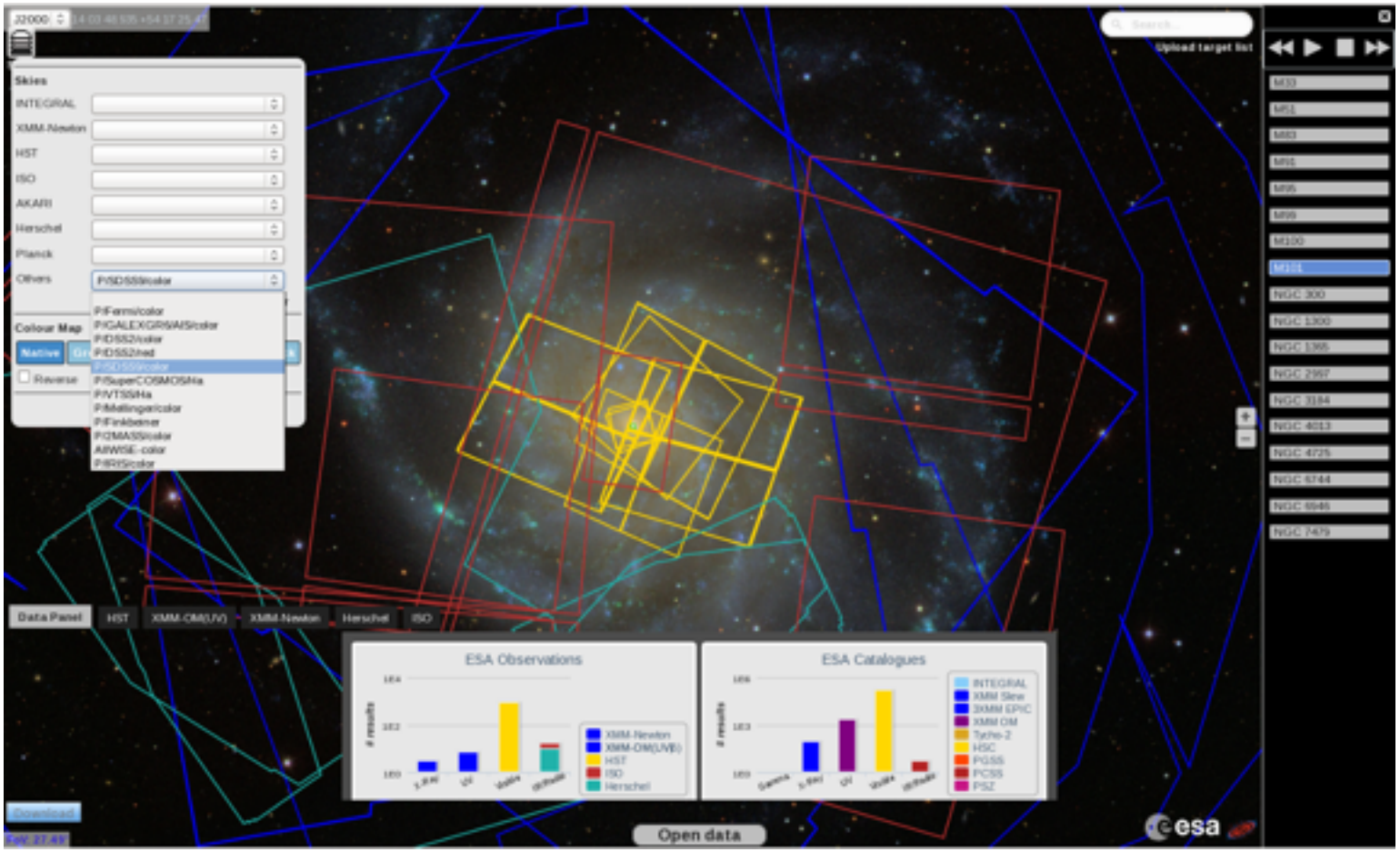
- Aladin running in browser
  - Javascript embeddable widget
  - Customisable
  - Open source GPL3
  - Very easy to implement
- Examples (outside CDS)
  - ESA Sky
  - Akari Explore Tool
  - GLIMPSE360
  - CADE *(Centre d'Analyse de Données Etendues)*
  - ADS *All-sky Survey*



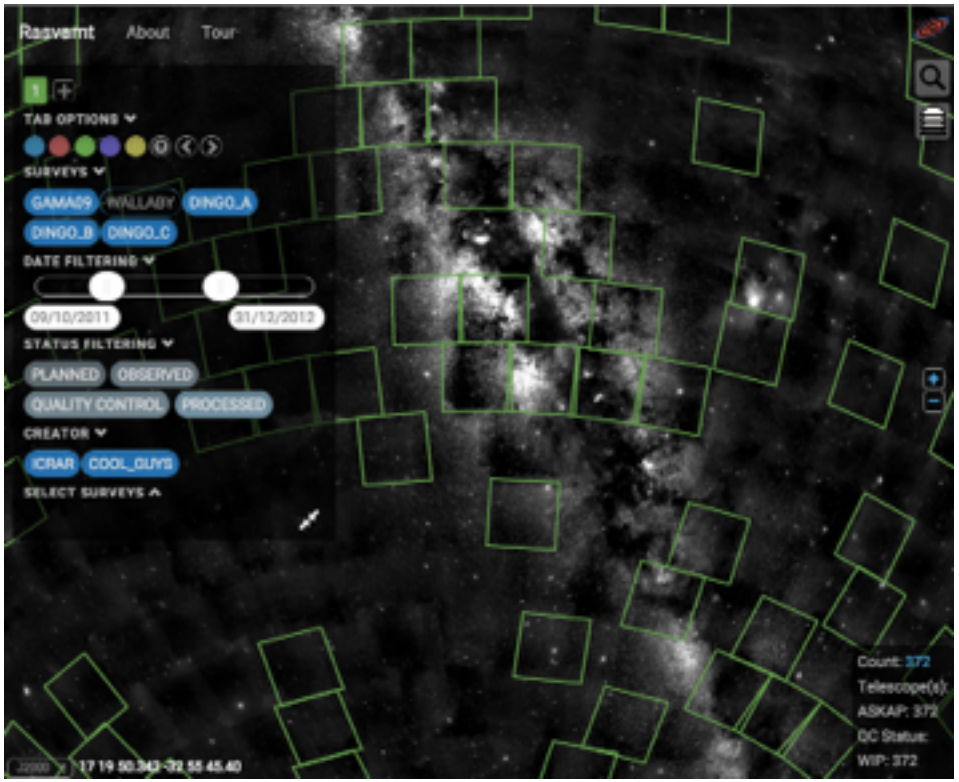
Aladin Lite


Portal
Simbad
VizieR
Aladin
X-Match
Other
Help
Aladin Lite
Target:
Surveys:
DSS2
Fermi
GALEXGR6/AIS
DSS2/red
DSS2/blue
SDSS9
Mellinger
2MASS
allWISE
J2000
05 45 30.655 -01 29 5.16
FoV: 3°

ESA sky - built on Aladin Lite



Radio Astronomy Survey Visualisation Monitoring Tool (ICRAR)

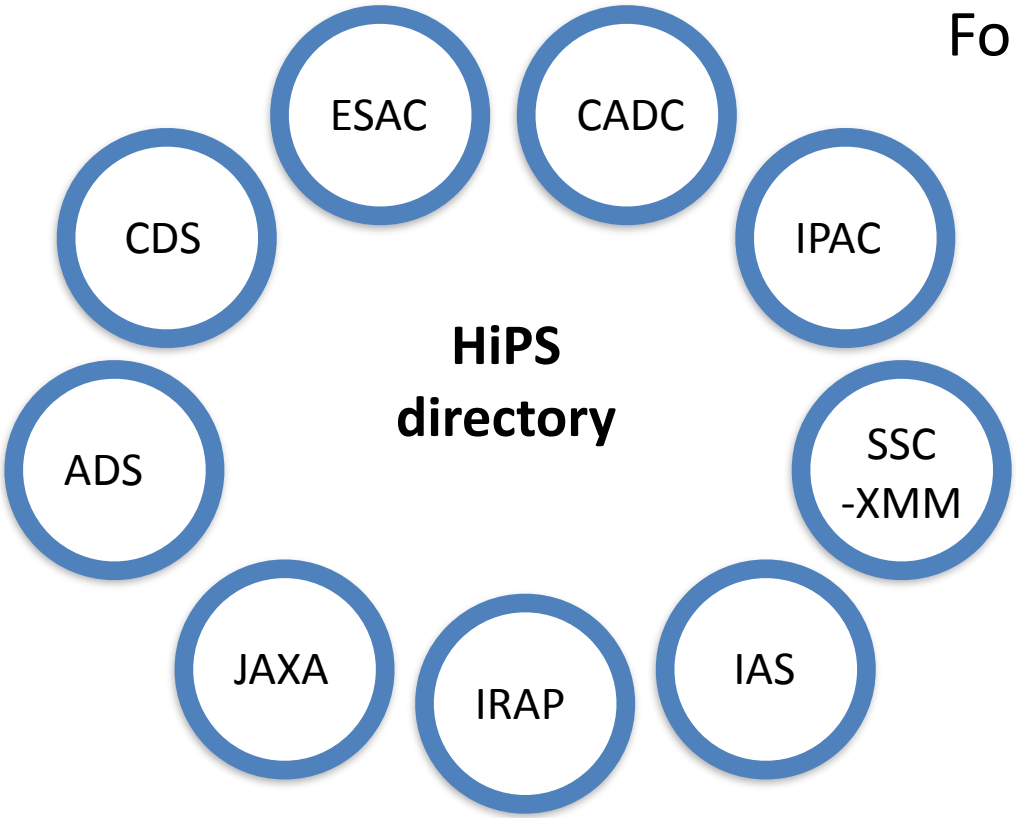


GLIMPSE 360

HiPS for Big Data

- ALMA: *recent tests with entire set of Cycle 0 cubes*
- HST: *use the whole archive as a survey*
- All sky at 0.8” (HEALPix level 18)
  - HiPS: 2TB *(16 bit FITS),* 100 GB *(JPEG)*
- ***Looking ahead:***
  - ~LSST: 18000 deg^2 , 0.3” resolution, each 3 days for 3 years
  - HiPS: ~5 PB cube *(16 bit FITS),* 256 TB *(JPEG)*
  - HiPS access would be feasible today

HiPS Network



Formed by HiPS providers for:

- Sharing
- Discovery
- Mirroring/Redundancy
- Interoperability

…coming to IVOA

Summary

- HiPS solutions for today's Big Data
- Enables new interoperability between images, catalogues and cubes

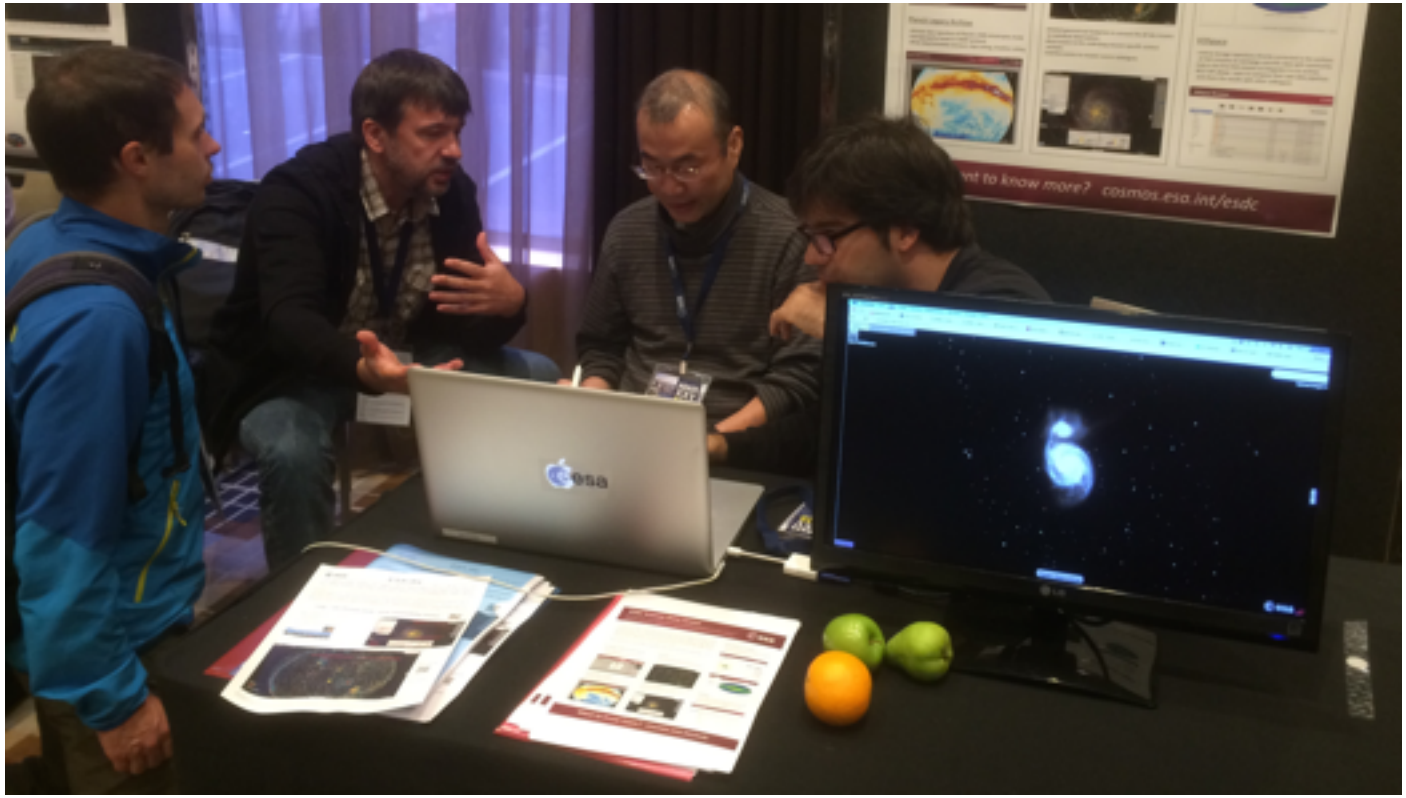
- Simplicity is key for success
  - interoperable by design
  - enables innovation
  - customisable to different needs
  - forming a community of HiPS providers

Links

- Hierarchical Progressive Surveys
  - *Fernique et al. 2015, A&A 578, 114*
  - HiPS on CDS web pages: http://aladin.u-strasbg.fr/hips
- Aladin http://aladin.u-strasbg.fr/AladinDesktop/
- Aladin Lite http://aladin.u-strasbg.fr/AladinLite/
- HiPS generation tools: http://aladin.u-strasbg.fr/hips/#tools

- ADASS: Visit us at the CDS booth

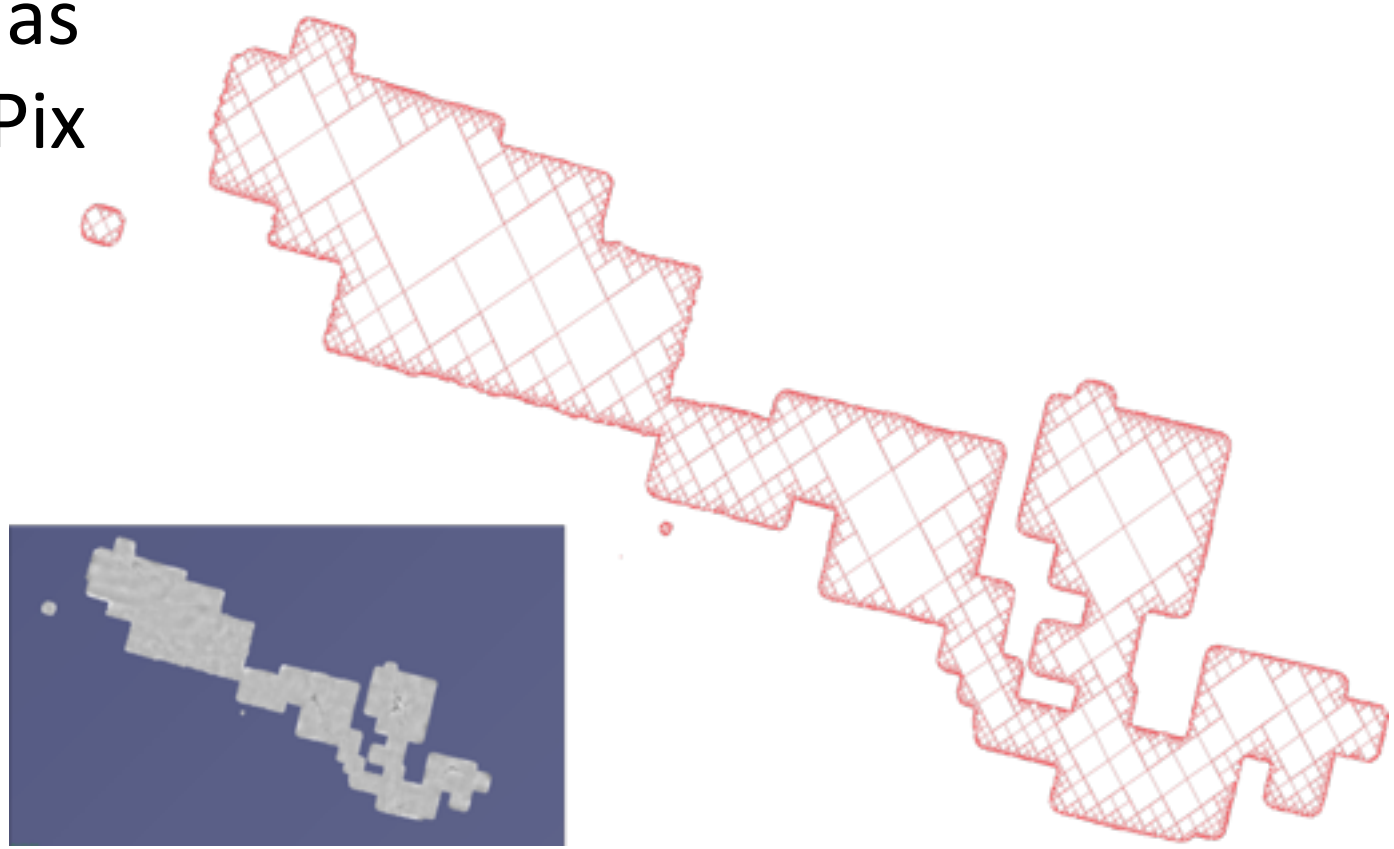
Extra Slides



- Fermi
- EGRET
- XMM
- INTEGRAL
- RASS
- GALEX
- DSS
- SDSS
- CFHTLS
- HST
- 2MASS
- ULTRAVISTA
- WISE
- ALLWISE
- DIRBE
- IRIS
- GLIMPSE360
- SPITZER
- AKARI
- SCUBA
- BGPS
- WMAP
- PLANCK
- NVSS
- CHIPASS
- SUMSS
- DWINGELOO
- HASLAM408
- VLSSR
- WENSS
- SHS
- SHASSA
- VTSS
- GASS
- CGPS
- GALFA
- CALIFA
- MUSE
- HARP/JCMT
- Gaia GUMS
- SIMBAD
- GOODS
- PHAT

Multi-Order Coverage (MOC*) Maps

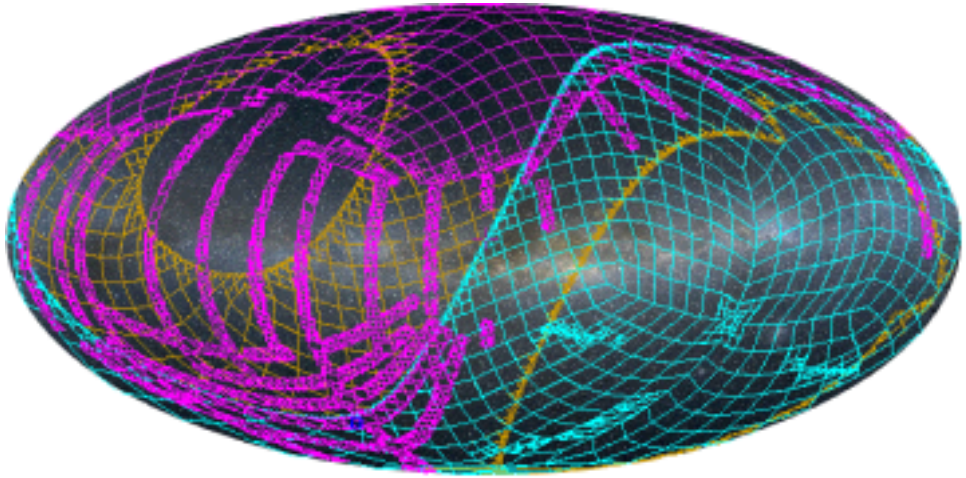
- Sky regions as list of HEALPix indices
- Multi-order
- Unique
- Fast



* IVOA Recommendation: MOC 1.0

HiPS and MOC

- Comparing coverage of 100s of large data sets
- Intersection/union/complement trivial
- Catalogue coverage
  - 1000s (Vizier)



- Queries based on coverage and catalogues
  - e.g. Veron quasars in HST, XMM and SDSS images

HiPS 3-dimensional cubes