Long-term management of 1000s of All-Sky reference data sets using the HiPS network

ADASS – October 2016 - Trieste
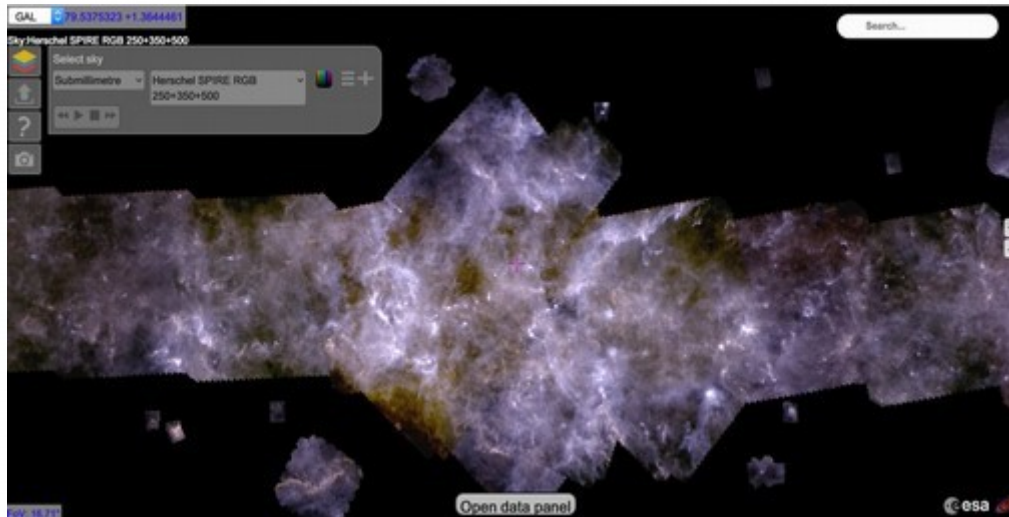
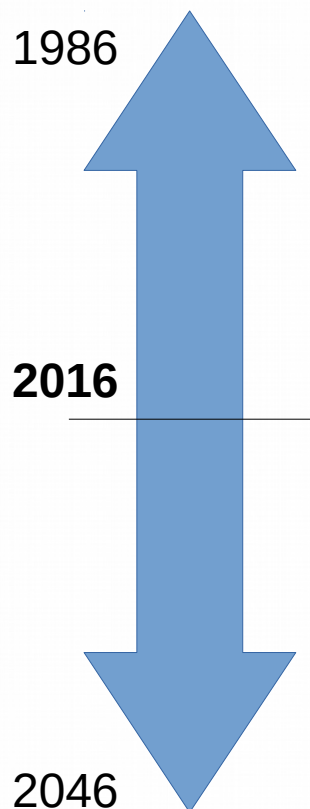P.Fernique, T.Boch, A. Oberto, M. Allen, D. Durand, K. Ebisawa, B. Merin, J. Salgado

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

Asterics

□ What's the plan ?

1) What does "**long term**" mean ?

2) Why **HiPS** could be **a good candidate** for long term management ?

☐ Long term ?

1986

2016

2046

- Which 30 years old files are you able to load today on your laptop ?

  FM2 ? TGA ? DOC ? GIF ? COD ? Fits ? JPEG ? ASCII ? JPEG2000 ? HDF4 ? PDF ? PCT ? CR1 ? PDS ? XML ? PPM, CDR ? WKS ? …

- Which existing files will be still **loadable/usable in 30 years** ?

□ The long term problems

1) Not just a data storage issue
   => The goal: **use the data** as long as possible

2) In fact, a couple of issues: **Data + Metadata**

3) Observation: *"More the data+metadata are easy to read/understand, greater is their life time"* (ex: ASCII tables)

4) A basic metric: **How hard is to (re)write the data reader ?**
   (indication: how depth is the documentation?).

□ Long term "enablers"

1) **Open solution** => proprietary solution can always disappear

2) **Basic & Standard coding** => avoid unusable or complex bit codings, dedicated compression and/or encryption

3) **Meta data easily readable, easy to associate to the data** => no binary, avoid subtle data model

4) **Avoid dependencies** to a medium, a DB, or other specific tools

5) **Extensible** => avoid package barriers such as data size or pixel depth limit (will force to a future format evolution)

☐ Two opposit approaches

1) Describe the data as they have been produced (ex: PDS)
=> The best for scientific analysis,
=> Easier for data producers,
=> Difficult for the data readers (have to support all collections of data+metadata)

2) Translate the data in a common basic package (ex: FITS)
=> Impact on future scientific analysis,
=> Difficult for the producers (translation required),
=> Easier for the readers (only one package to support)

Illustration of these approaches

1) ...the data as they have been produced...(ex:PDS)

...The collection consists of Mariner 10 images ... file consists of 22 blocks containing 31,944 bytes per block. Eack block is composed of 33 logical records of 968 bytes each. The first logical record of the first block contains a single set of VICAR label information. The tapes were created on an IBM system, so the VICAR Standard Format labels are written in EBCDIC. ...

2) ...the data in a common basic package...(ex:FITS)

*...FITS...consisted of **a binary array** preceded by **an ASCII text header** with information describing the organization and contents of the array ...*

☐ Consequences…

PDS: ~2 readers, 1 lib, only unix

reply to **post by BuzzDengue**

I couldn't find any program besides NASA View and the ISIS program (ISIS is a suite made up of dozens of programs) that can read PDS version 3 images. ☹

30 years later...

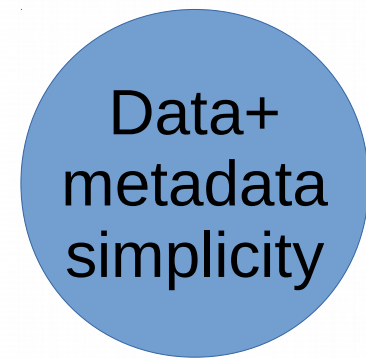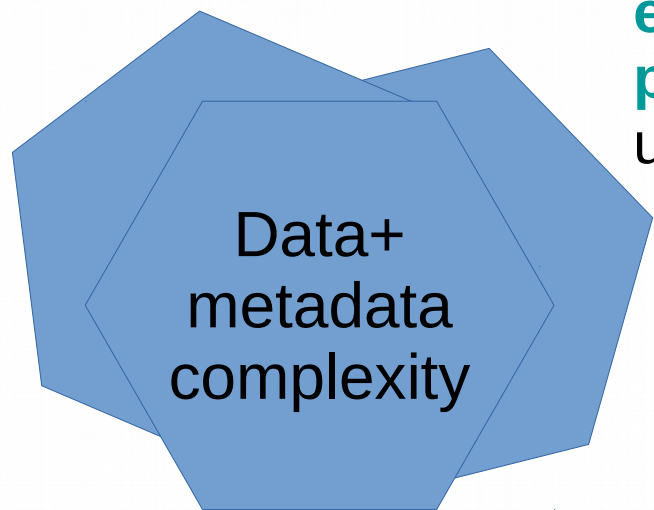**FITS : hundreds of readers + libs, multiple OS**

The dilemma

Data+ metadata complexity

Data+ metadata simplicity

A few dedicated readers
=> small audience
=> Probably short term usage

Lots of readers
=> large audience
=> Probably long term usage

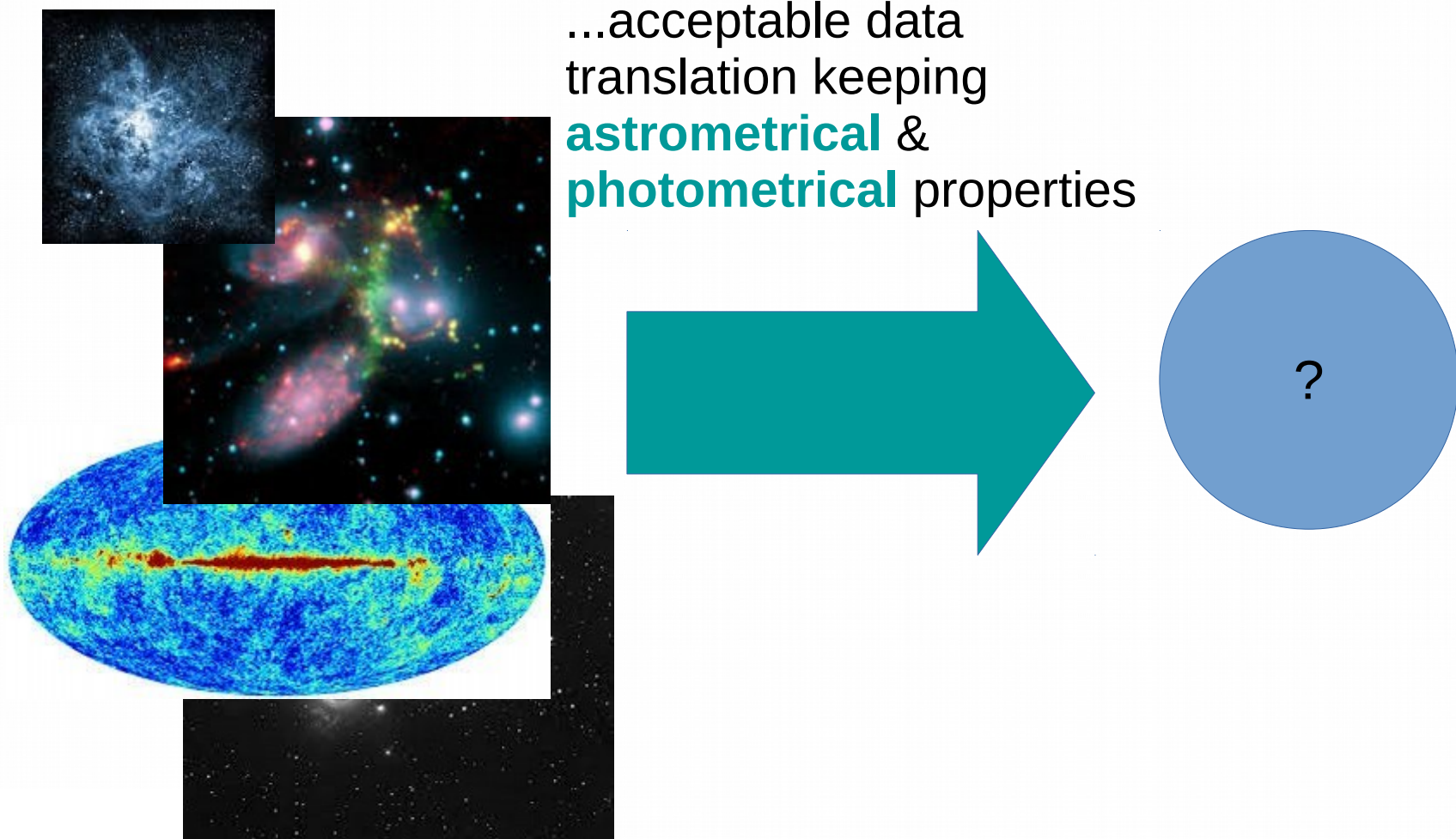# □ The method

Find **an acceptable data translation keeping enough scientific properties** for future usages

Data+ metadata complexity

Data+ metadata simplicity

...acceptable data translation keeping **astrometrical** & **photometrical** properties

?

☐ Which solution for astronomical images?

...keeping **astrometrical** & **photometrical** properties

?

FITS

HEALPix

?
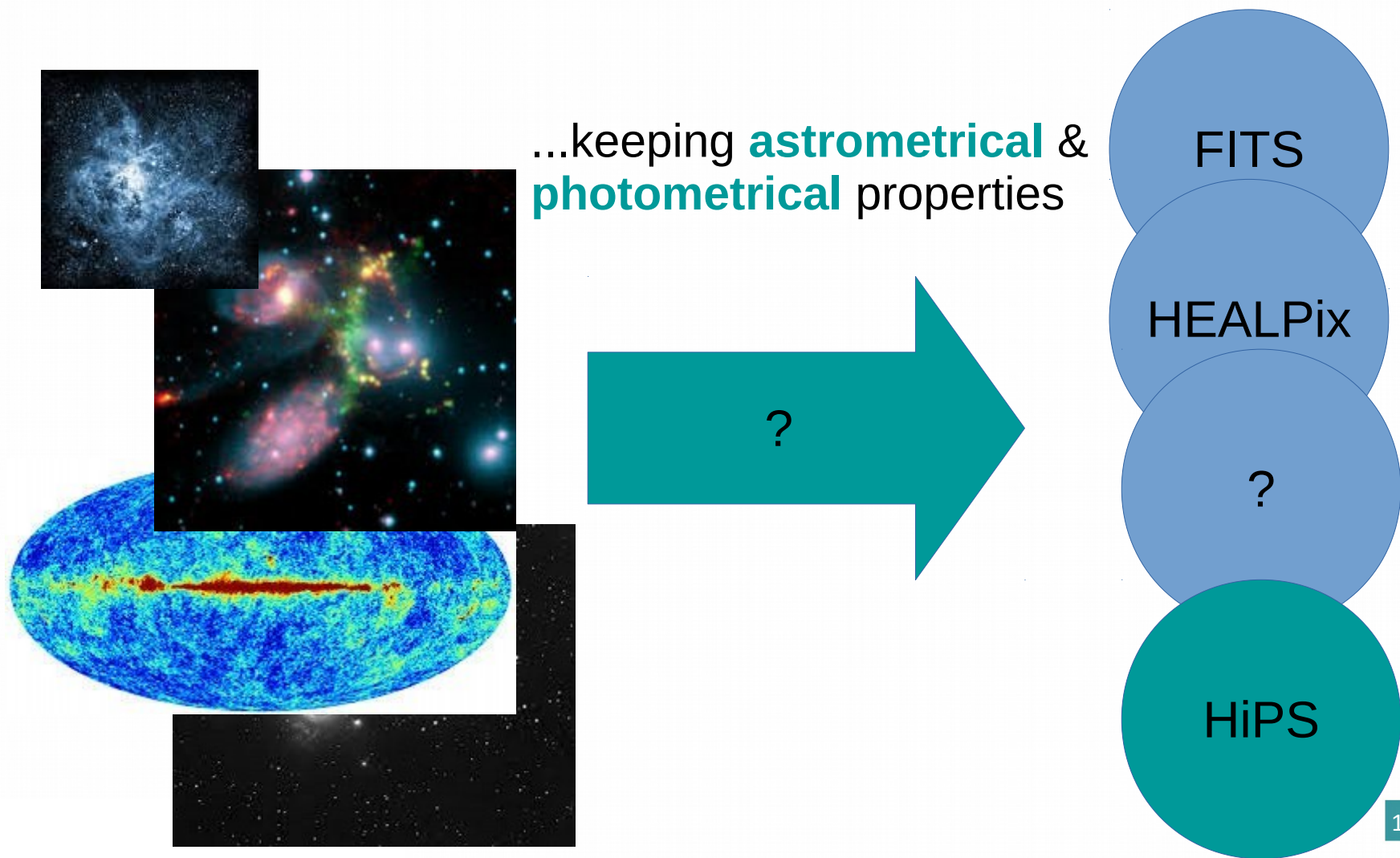
HiPS

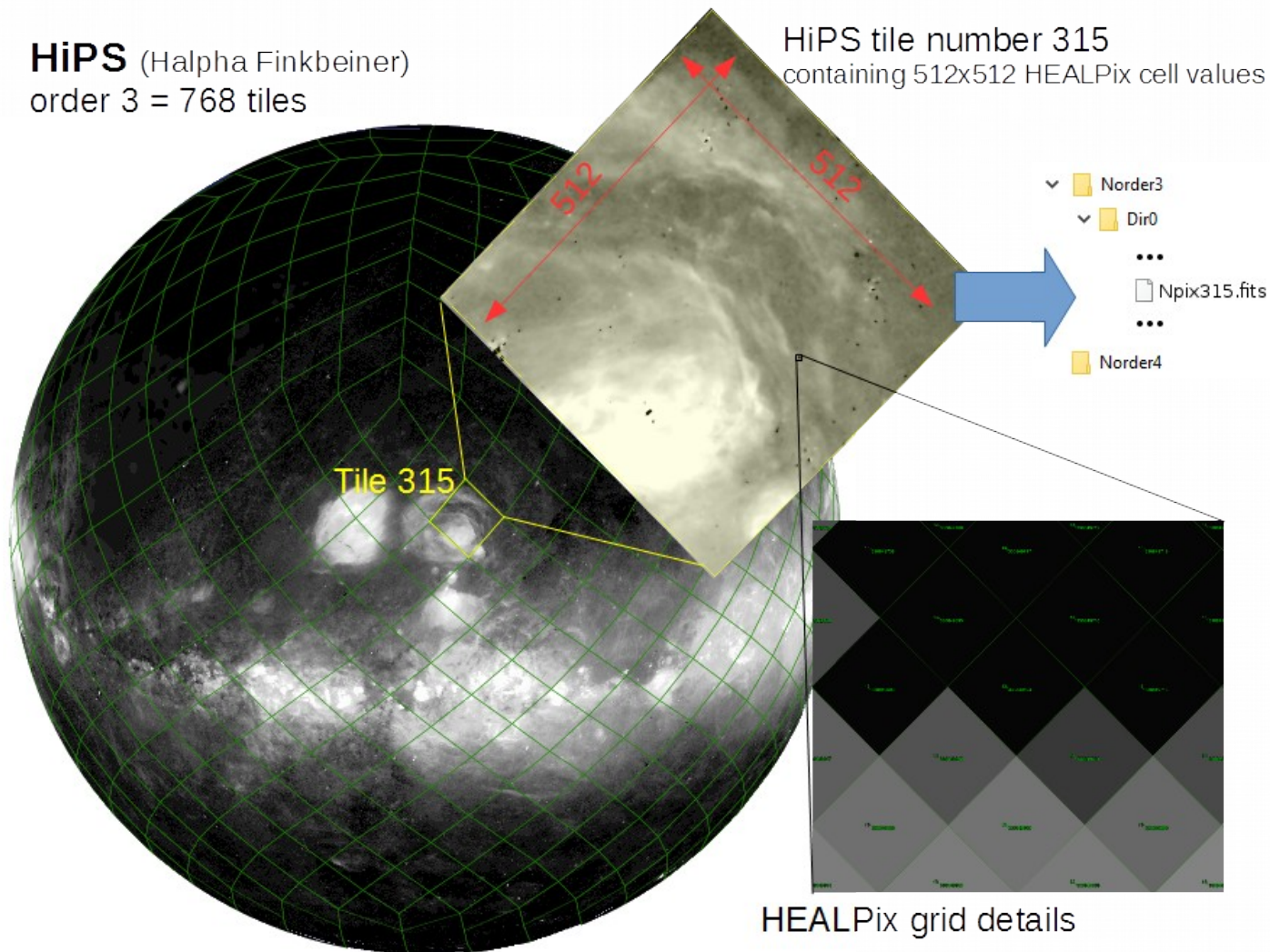# HiPS – What is it ?

Hierachical Progressive Survey

"The more you zoom in on a particular area,the more details show up"
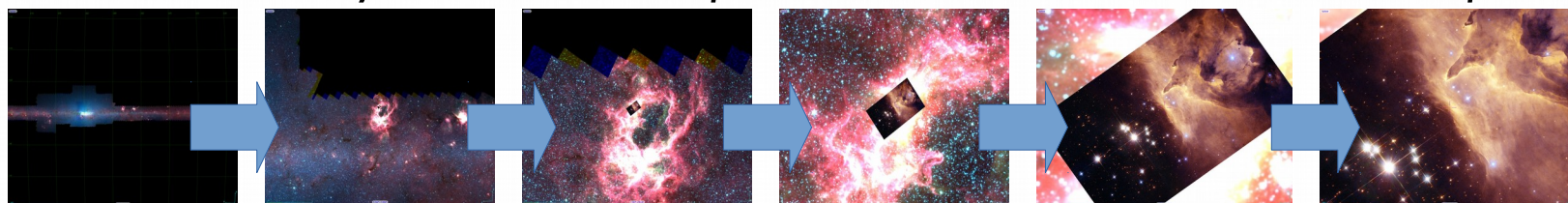
- Conserves **scientific data properties** (pixel localization & value)

- Based on the **HEALPix geometry**

- Dedicated to **any kind of image collections** (surveys, pointed observations, and more…)

- Stored as a **basic hierarchy of regular files**

HiPS = Mosaic of HEALPix (spherical) tiles…
… stored in a hierarchy of files

HiPS (Halpha Finkbeiner)
order 3 = 768 tiles

Tile 315

HiPS tile number 315
containing 512x512 HEALPix cell values

512
512

Norder3
  Dir0
    ...
    Npix315.fits
    ...
  Norder4

HEALPix grid details

Each tile/file can be coded in FITS, JPEG or PNG

Npix89913.fits   Npix89913.jpg   Npix89913.png
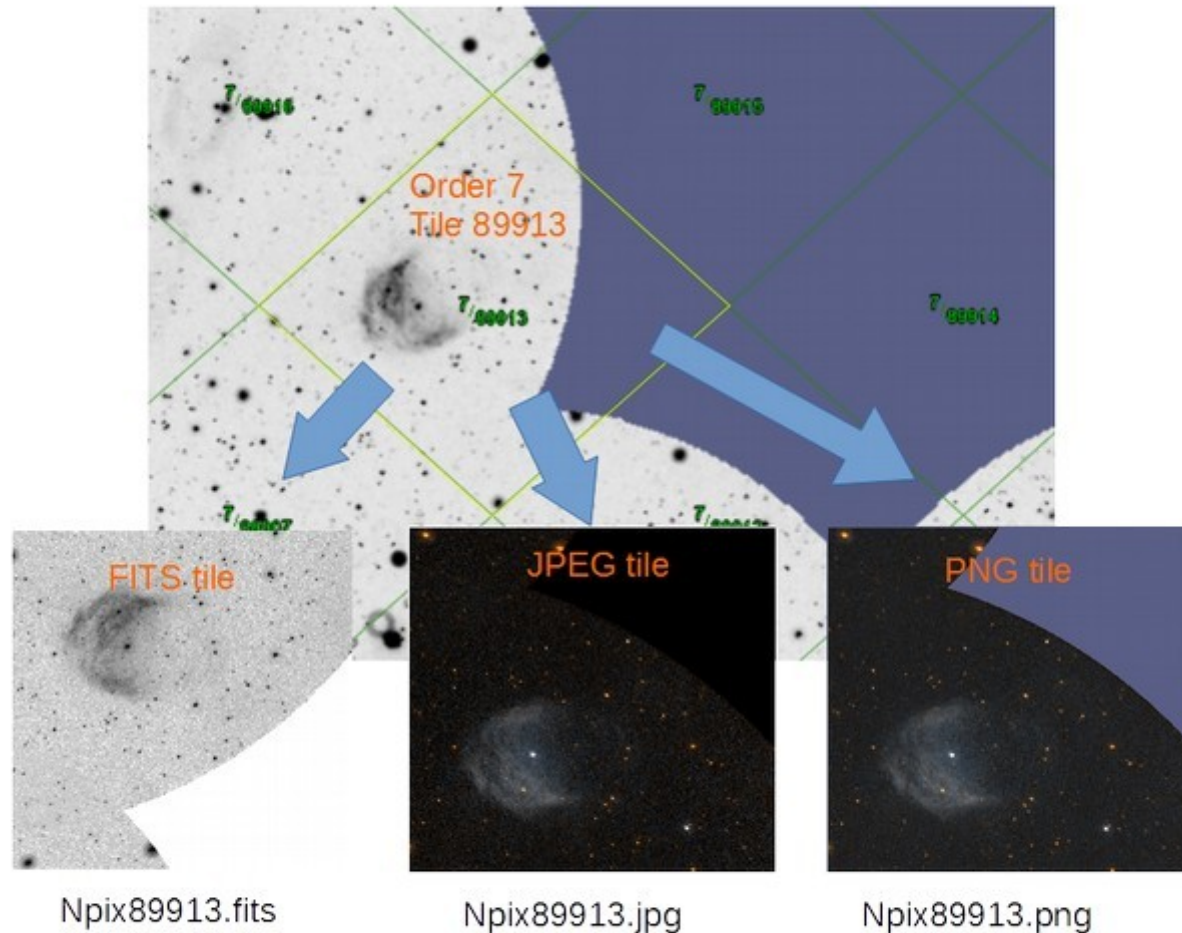
State of art
(November 2016)

- **300+** HiPS for **100TB** data

- HiPS clients :

  → Aladin Desktop (CDS), Aladin Lite (CDS), MIZAR (CNES)
  + in dev: STScI portal (NASA), openWWT (Microsoft), proto (China), …

  + Aladin Lite implementation: ESAsky (ESAC), JUDO2 (JAXA), Skymap (LIGO/virgo) …
  + Aladin Lite web page inclusion: Simbad, VizieR, GLIMPSE360, CADE, ADS allsky, CASSIS, Akari-Viewer, VistaOrion, AstroDEEP, CDS portal, Gamma Sky, MOPRA, …
  + Aladin Desktop usage "diversion": Arches walker

- **300 000+ HiPS tiles** requested / day

State of art (November 2016)

- **12+** HiPS servers

  → CDS, SSC-XMM, IAS, IRAP/CADE, IPAC, ADS, ESAC, JAXA, AMIGA, Spanish-VO, Vista-Orion, TGSSADR...

- HiPS network infrastructure for registration & publication (see IVOA standard)

- **IVOA standard** (WD stage)

  → http://www.ivoa.net/documents/HIPS/20160623

- HiPS generator→ Aladin/Hipsgen (perf: 10h/1Tpix),

- Paper → 2015A&A...578A.114F

☐ Is HiPS a good candidate for long term usage?

- Today…
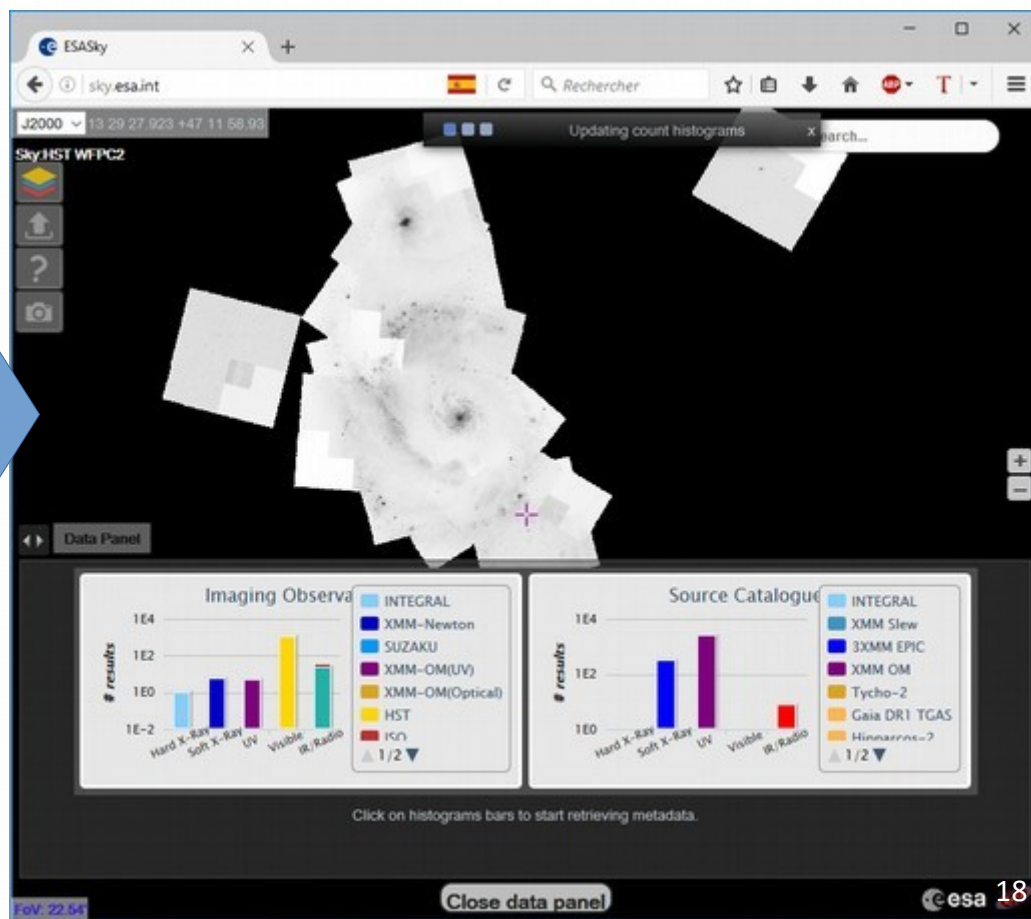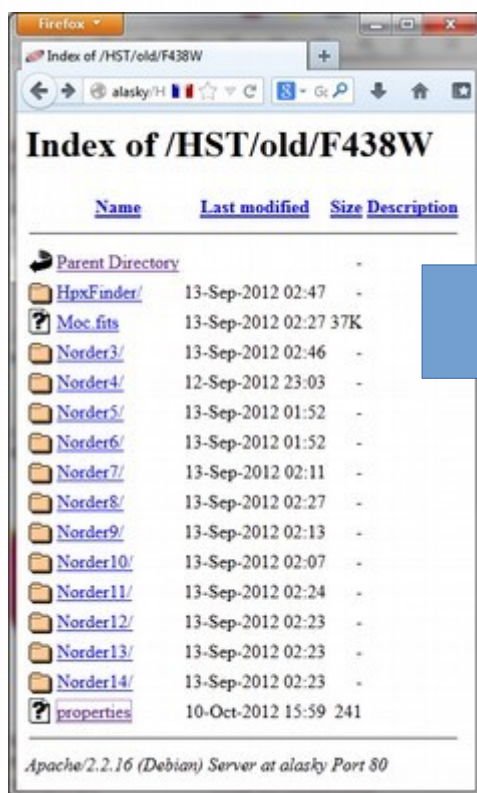
HiPS characteristics

☑ Astrometry: HEALPix based (400 μarcsec)

☑ Photometry: FITS tiles HEALPix based

☑ Coding: basic files (FITS | JPEG | PNG, ASCII)

☑ Metadata: simple representation (key=value)

☑ Big data ready: hierarchical (=> no limit)

☑ Open: IVOA standard

☑ Dependencies: HEALPix lib

☐ The gamble…

- Tomorrow !

Thanks ! Questions ?