Coverage of astronomical datasets

Summer School “Geometry and data”

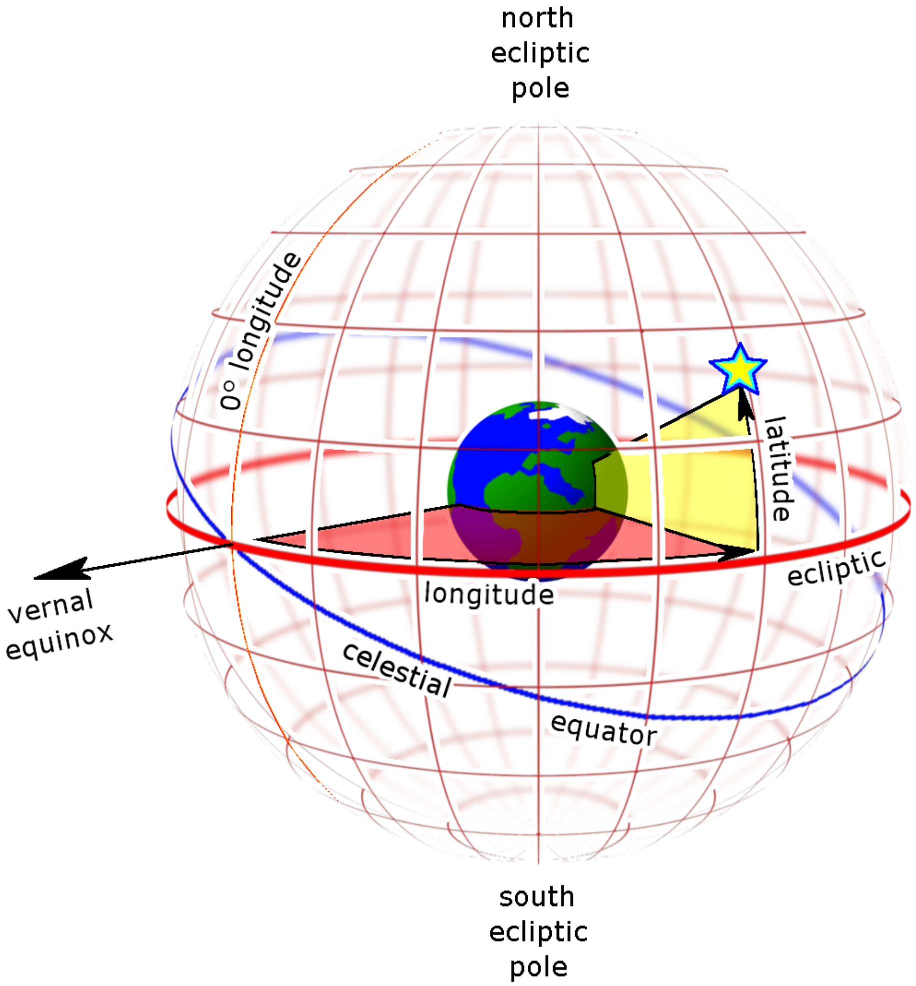
Sébastien Derriere
sebastien.derriere@astro.unistra.fr

Centre de Données Astronomiques de Strasbourg

Observatoire astronomique de Strasbourg | ObAS

Astronomical observations

- Spherical coordinates
- Often, distances are not known



JWST image

- Neptune & satellites
- Stars
- Galaxies

Astronomical images

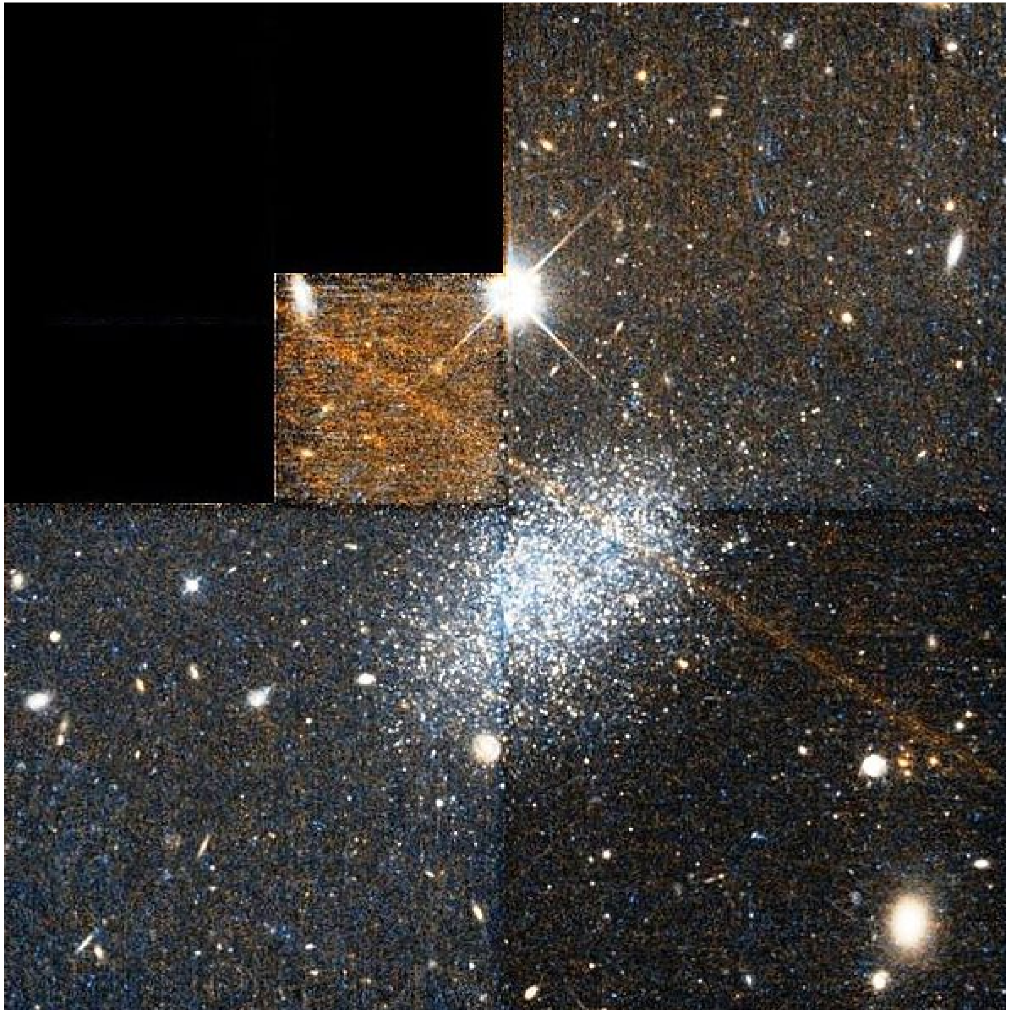
- 2D spatial sampling (field of view)

… but also

- Time sampling (exposure time)
- Frequency sampling (wavelength/energy)
- Limits in sensitivity
- …

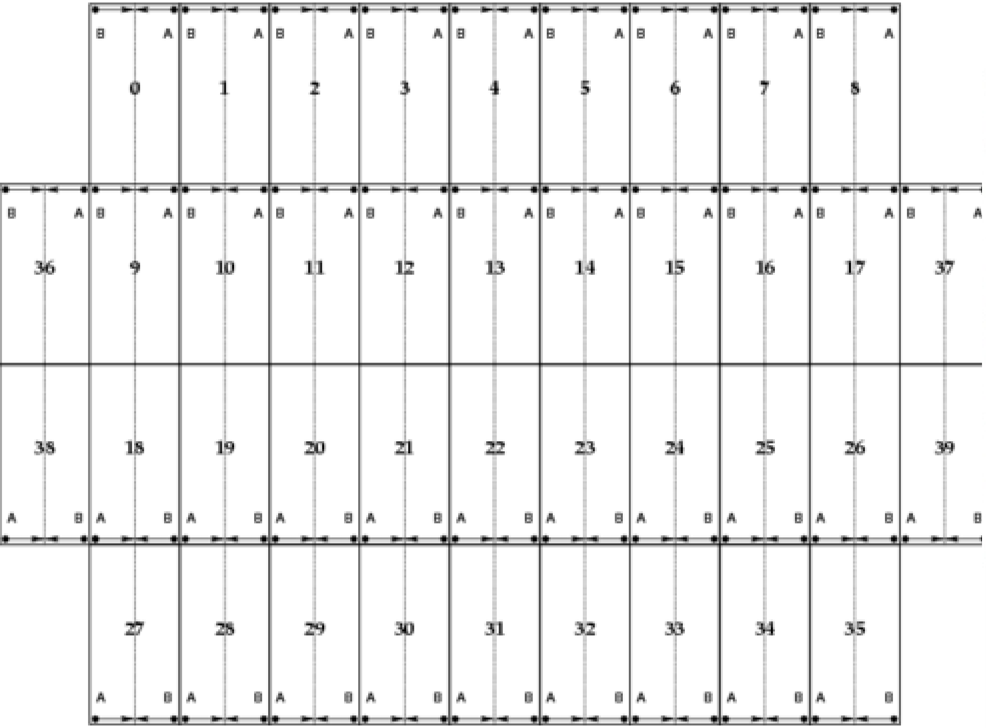
Astronomical images

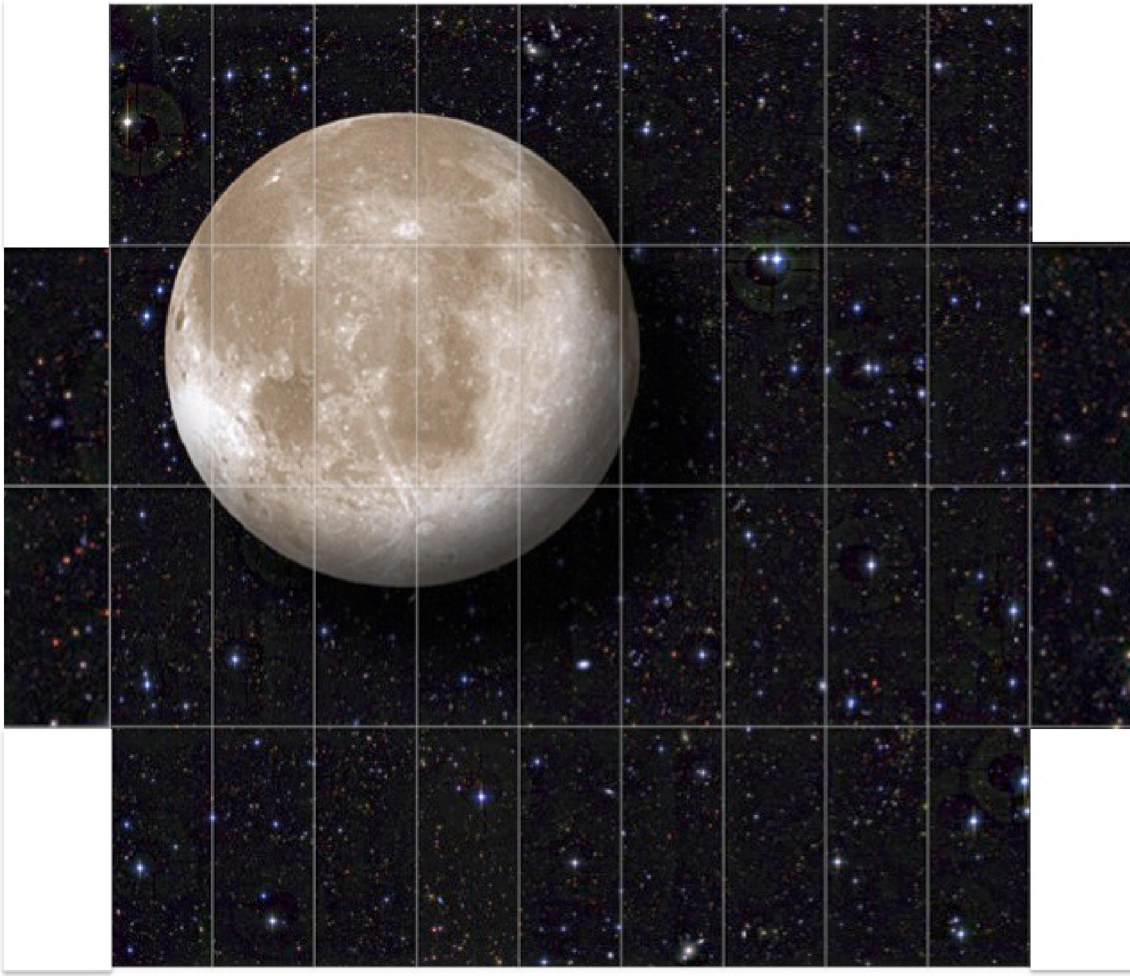
- Shape can be complex ! (HST WFPC2)



Astronomical images

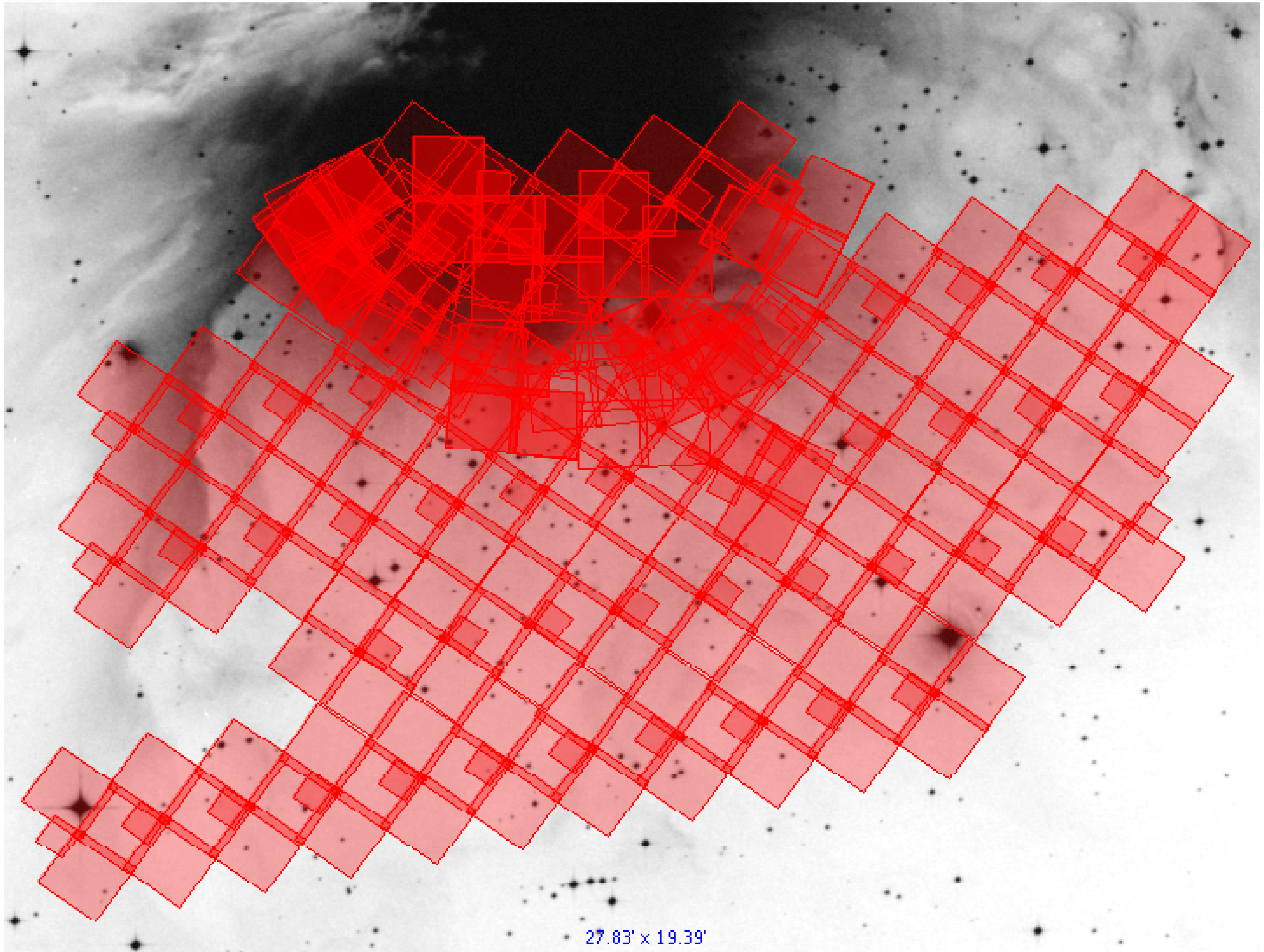
- Mosaics of detectors (MEGACAM, CFHT)





Sky surveys

- Collections of images, covering a fraction of the sky
- How to provide seamless access to image sky surveys ?
- How do you describe the spatial coverage ?
  - Shape ?
  - Sky fraction ?

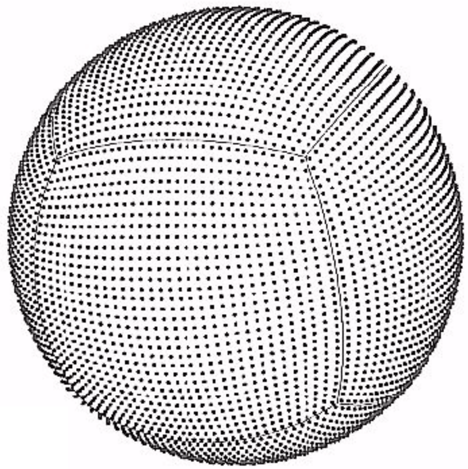


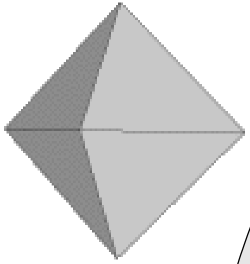
Sky surveys : all-sky view and coverage ?

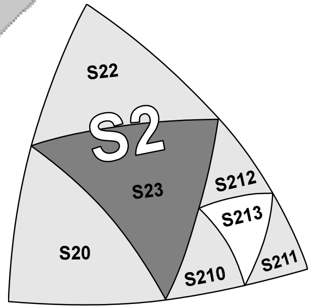
- Describe complex regions on the celestial sphere
- Compare, combine coverage of several datasets (intersect, ...)
- Global properties
  - Sky fraction (full-sky, ...)
  - Shape (multiple regions, gaps, holes, ...)

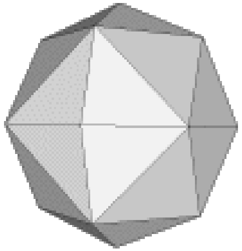
Sky partitions : several schemes

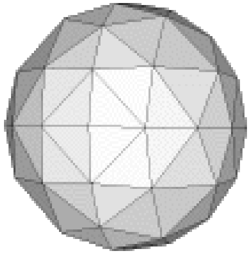
- Spherical-cubic projection, quad-cubes
  - 6 faces of a cube, subdivided recursively into 4 squares
- HTM : Hierarchical Triangular Mesh
  - 8 triangles, subdivided recursively into 4 triangles



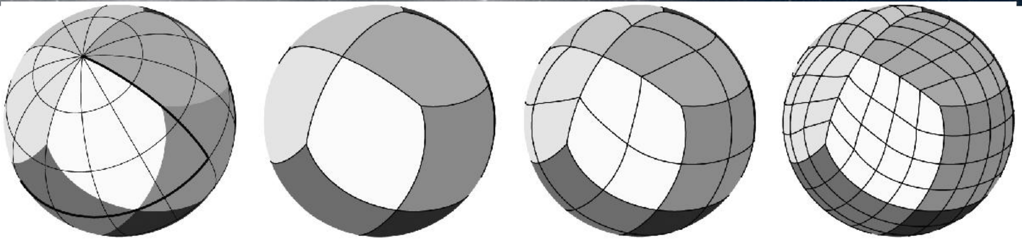




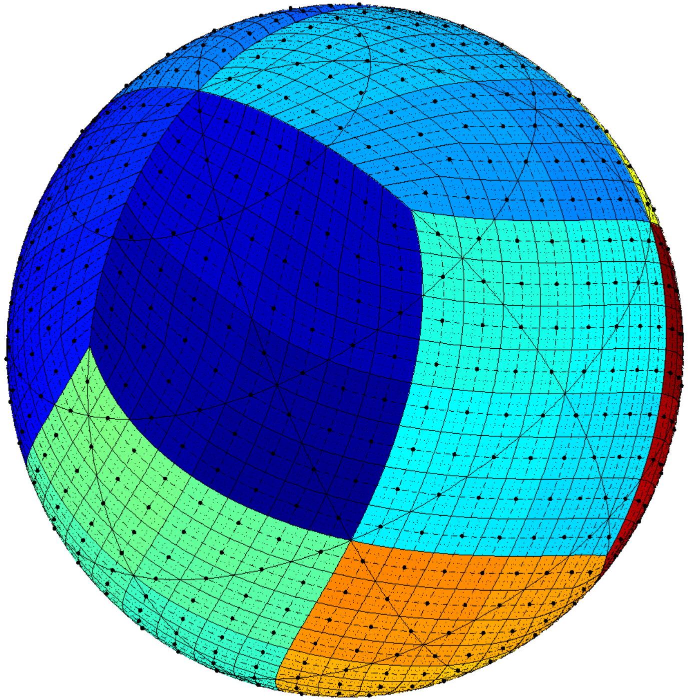




Sky partition : HEALPIX



- HEALPix (Gorski et al. 2005)
  - **12 quadrilateral pixels**
  - 2x2 division at each level
  - **Equal area,** Iso-latitude

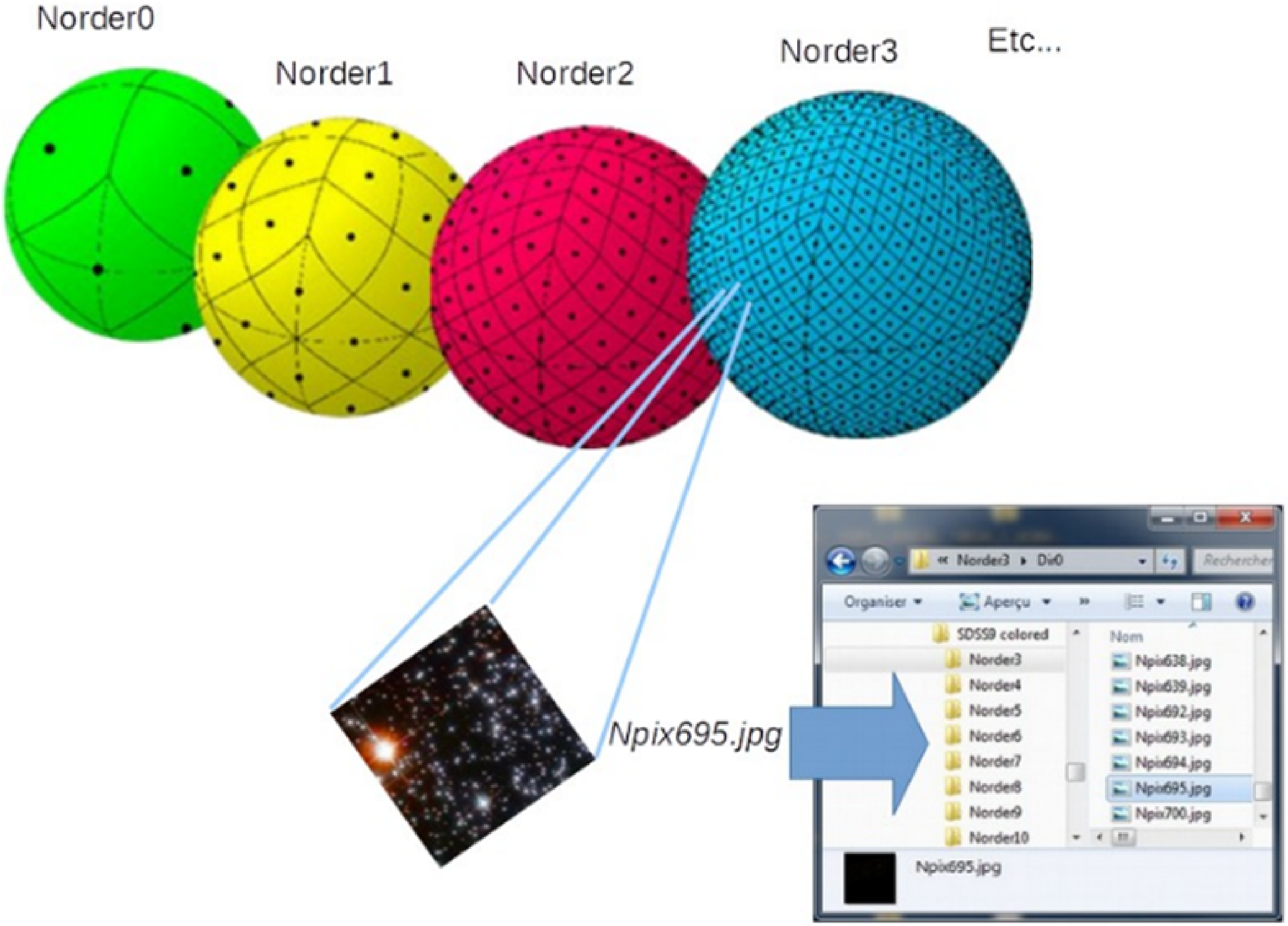


- All these partitions rely on quad-trees
  - Each cell is subdivided into 4 children, recursively

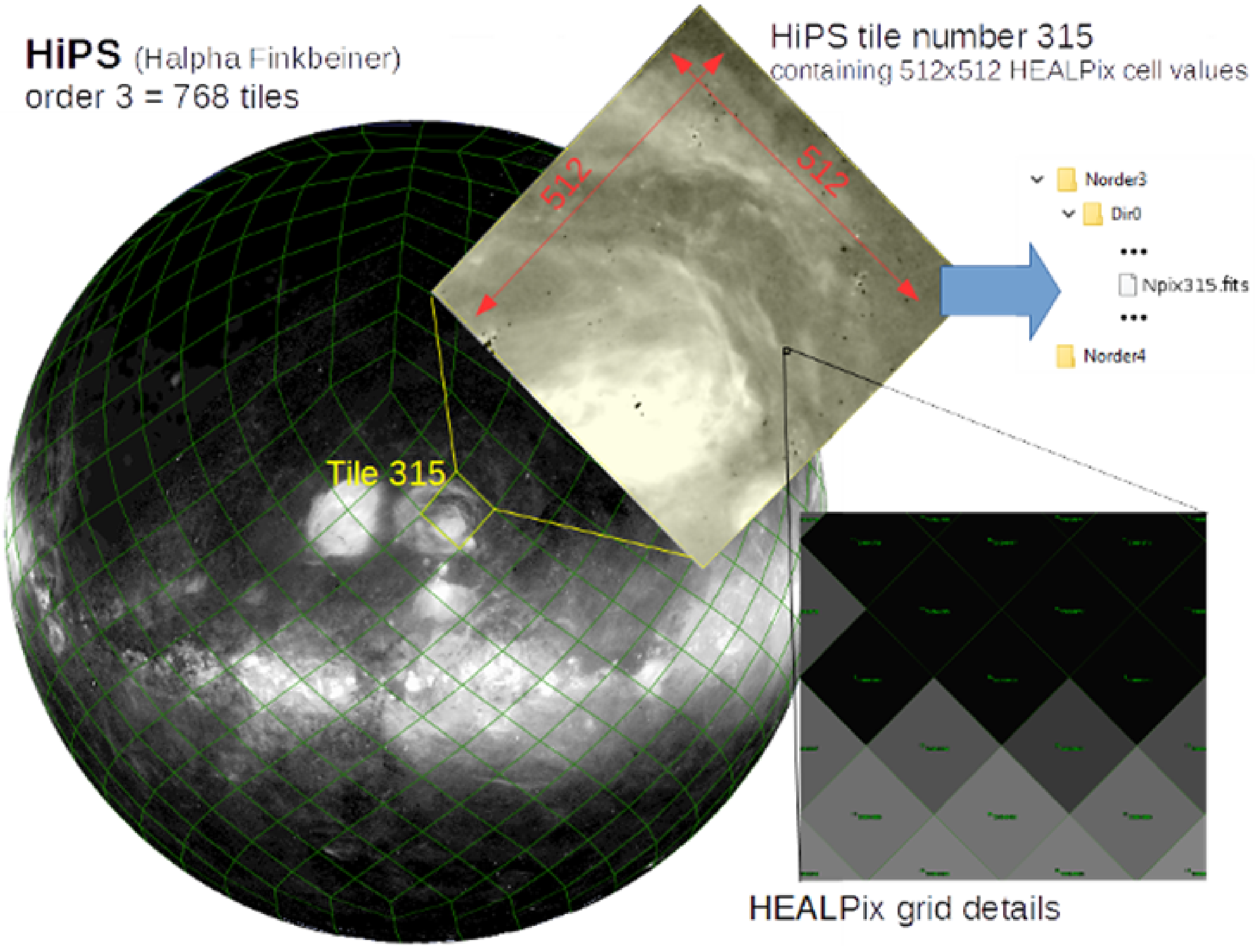
HiPS = Hierarchical Progressive Surveys

- Mosaic of HEALPix tiles (HEALPix pixel geometry)

  (Fernique et al. 2015)



HiPS = Hierarchical Progressive Surveys



HiPS as a standard

- International Virtual Observatory alliance standard – ivoa.net

International
Virtual
Observatory
Alliance

HiPS – Hierarchical Progressive Survey

Version 1.0
IVOA Recommendation
19th May 2017

This version:
1.0: Recommendation 2017-05-19

Previous version(s):
1.0: Proposed Recommendation 2017-04-06
1.0: Proposed Recommendation 2017-04-03
1.0: Proposed Recommendation 2017-02-07
1.0: Proposed Recommendation 2016-11-22
1.0: Working Draft 2016-06-23

Interest/Working Group:
Applications: http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications

Editor:
Pierre Fernique

Authors:
Pierre Fernique, Mark Allen, Thomas Boch, Tom Donaldson, Daniel Durand, Ken Ebisawa, Laurent Michel, Jesus Salgado, Felix Stoehr

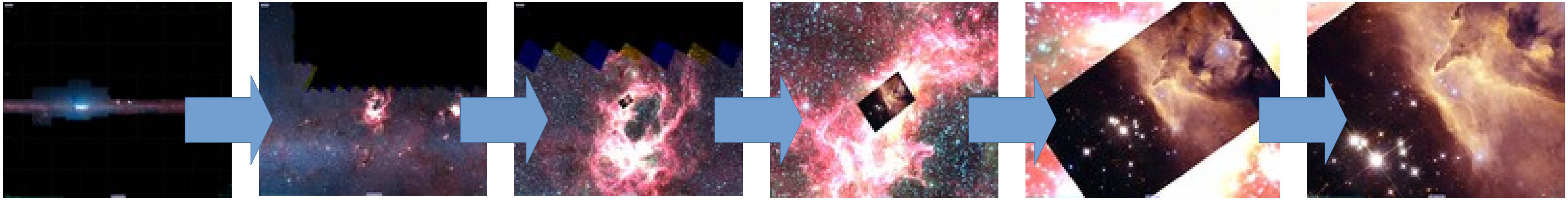
Abstract

This document presents HiPS, a hierarchical scheme for the description, storage and access of sky survey data. The system is based on hierarchical tiling of sky regions at finer and finer spatial resolution which facilitates a progressive view of a survey, and supports multi-resolution zooming and

1

HiPS : huge success

- Hierachical Progressive Survey
  - *"The more you zoom in on a particular area, the more details show up"*
- Multi-resolution HEALPix data structure for Images, Catalogues, 3-dimensional data cubes, ...
- Conserves scientific data properties alongside visualisation considerations
- No databases or servers, just HTTP

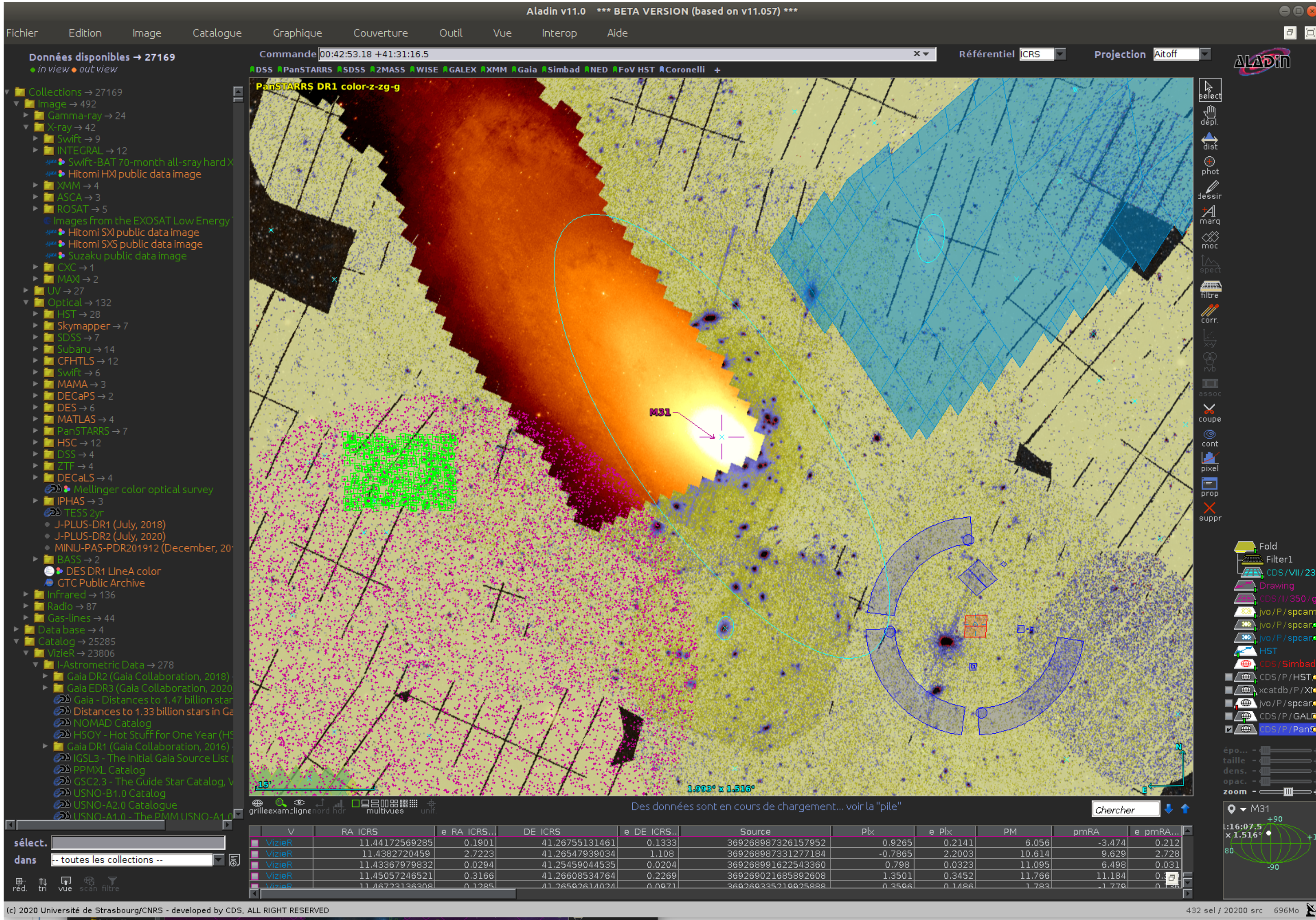


A few examples in Aladin

Aladin = sky atlas developed by CDS in Strasbourg.

Acces to 1200+ surveys !

http://aladin.cds.unistra.fr/hips/



MOCs

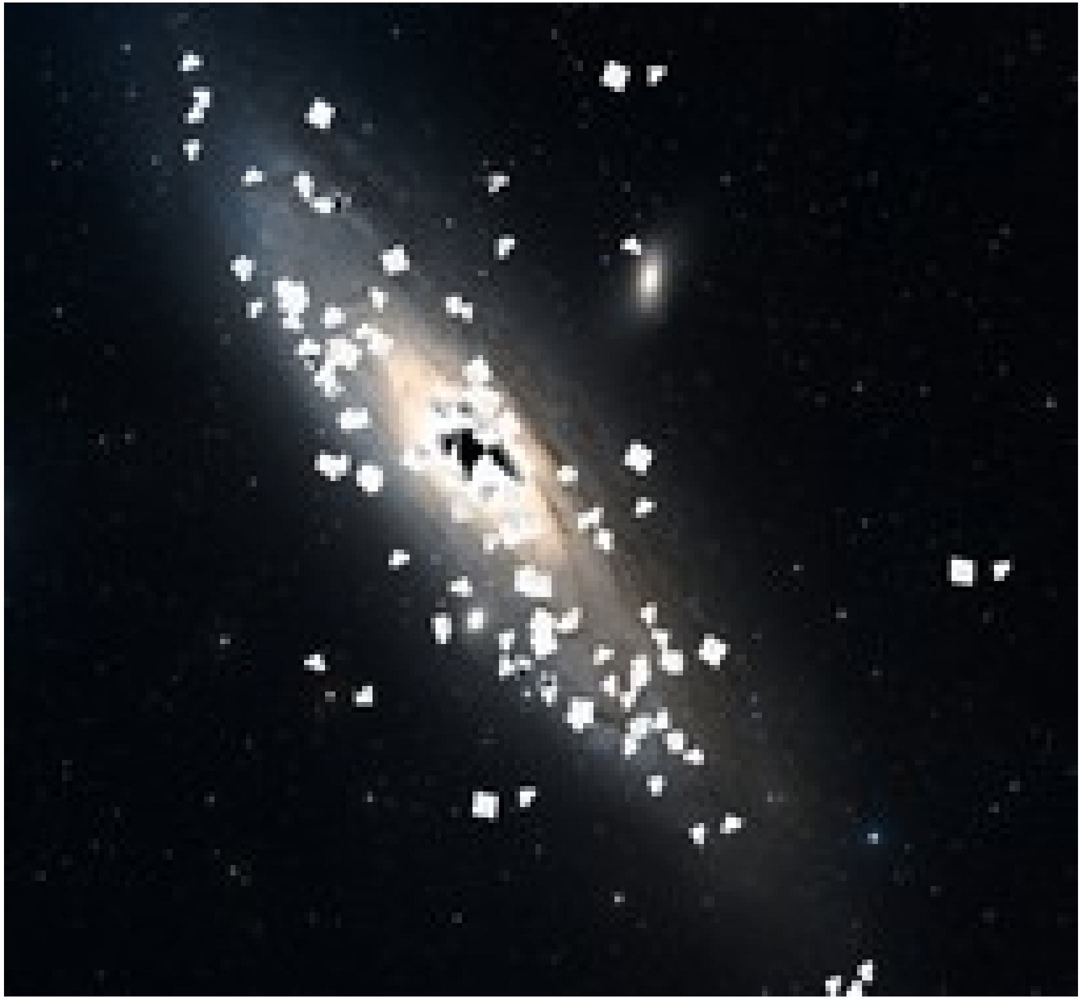
- Multi-Order Coverage map

  “Combine sky regions in few milliseconds”

- A simple and efficient method to specify any kind of sky regions
- Based on HEALPix tessellation
- Existing libraries: Java, C, python
- Used in VO tools (Aladin, TOPcat, …)

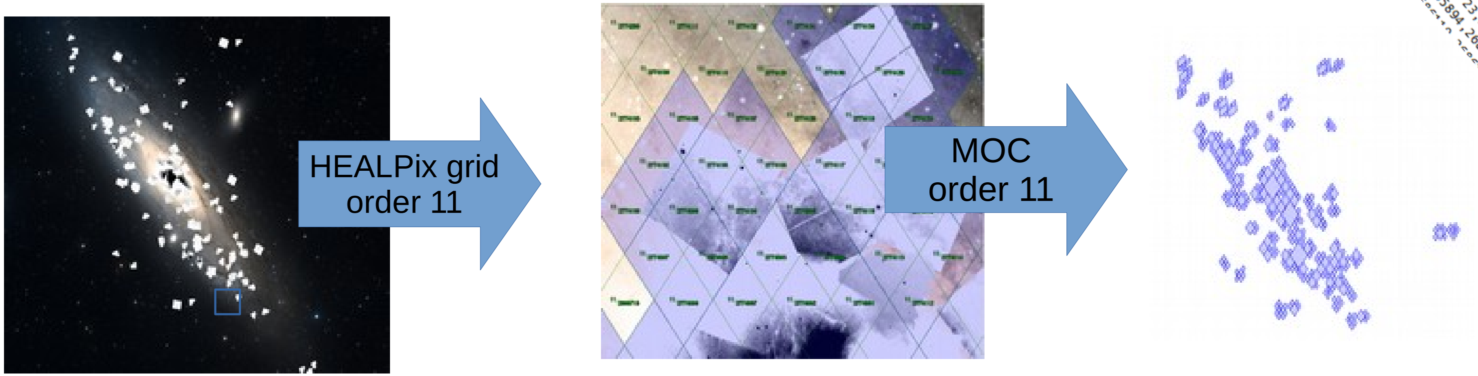
What is a MOC ?

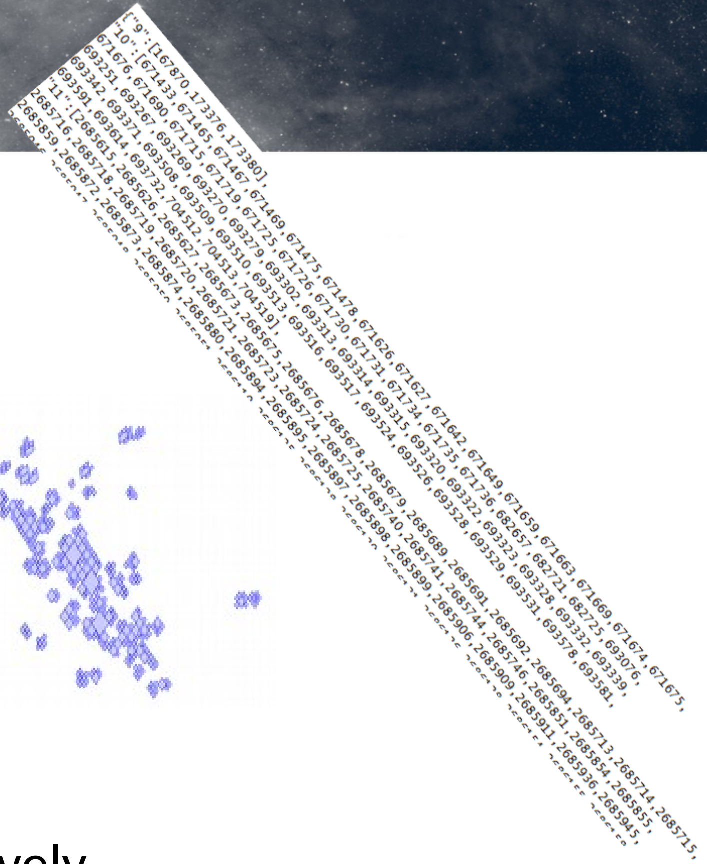
- How to describe the spatial coverage of HST observations around the Andromeda galaxy ?



What is a MOC ?

- “Just” the list of HEALPix cell numbers covering a region

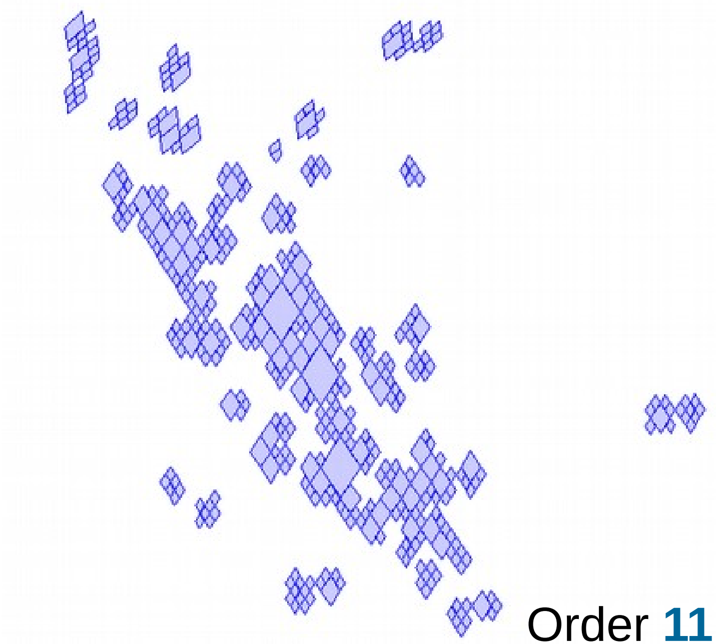




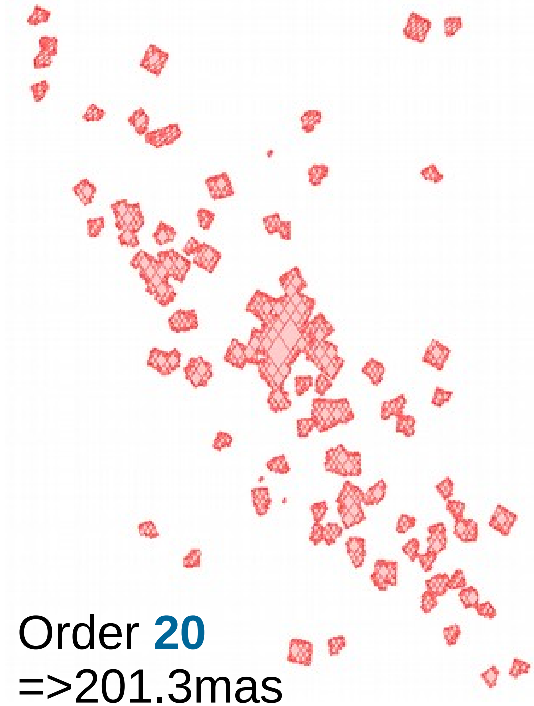
- 4 adjacent cells are replaced by the parent, recursively
  → intrinsic compression
- Store as a FITS table (or JSON)

What is MOC ?

- MOC accuracy depends of the MOC **order**
  = the smallest HEALPix cell level used in the MOC

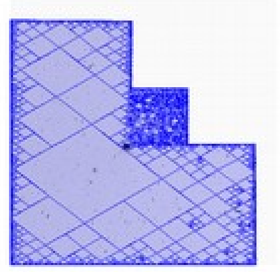


Order **11**
=>1.718 arcsec



Order **20**
=>201.3mas

. . .



Order **29**
=>391 µas

MOC standard

International Virtual Observatory Alliance

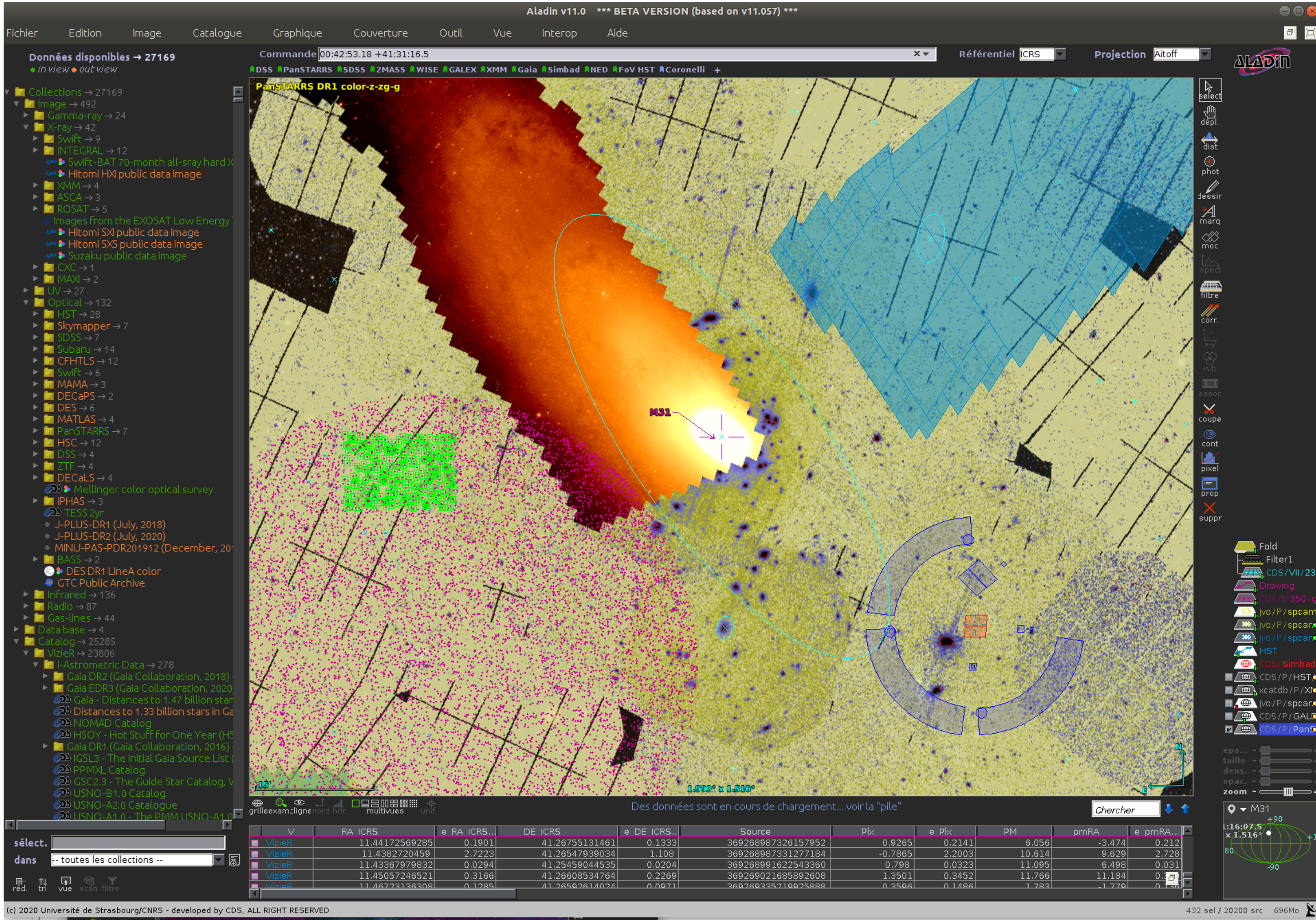
MOC: Multi-Order Coverage map
Version 2.0

IVOA Recommendation 2022-07-27

Working group
Applications
This version
http://www.ivoa.net/documents/moc/20220727
Latest version
http://www.ivoa.net/documents/moc
Previous versions
Version1.1
Version1.0
Author(s)
Pierre Fernique (CDS), Ada Nebot (CDS), Daniel Durand (CADC), Matthieu Baumann (CDS), Thomas Boch (CDS), Giuseppe Greco (EGO-Virgo), Tom Donaldson (STScI/NASA), Francois-Xavier Pineau (CDS), Mark Taylor (University of Bristol), Wil O'Mullane (Vera C. Rubin Observatory), Martin Reinecke (Max Planck), Sébastien Derrière (CDS)
Editor(s)
Pierre Fernique, Ada Nebot, Daniel Durand

A few examples in Aladin

Very easy to perform arithmetic operations between MOCs : intersections, union, ...



MOC 2.0

- Extension to describe not only space coverage, but also time coverage : ST-MOC.
- I'll come back to this at the end with an example
(it time allows) !

What about datasets coverage ?

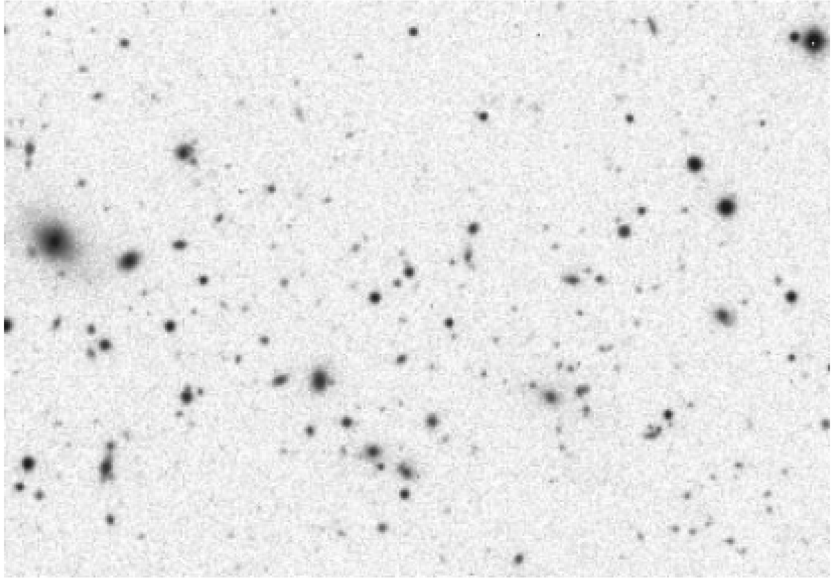
- MOCs seem to address the problem, but…
- Can I get general coverage properties ?
  - e.g. sky_fraction [0..1]

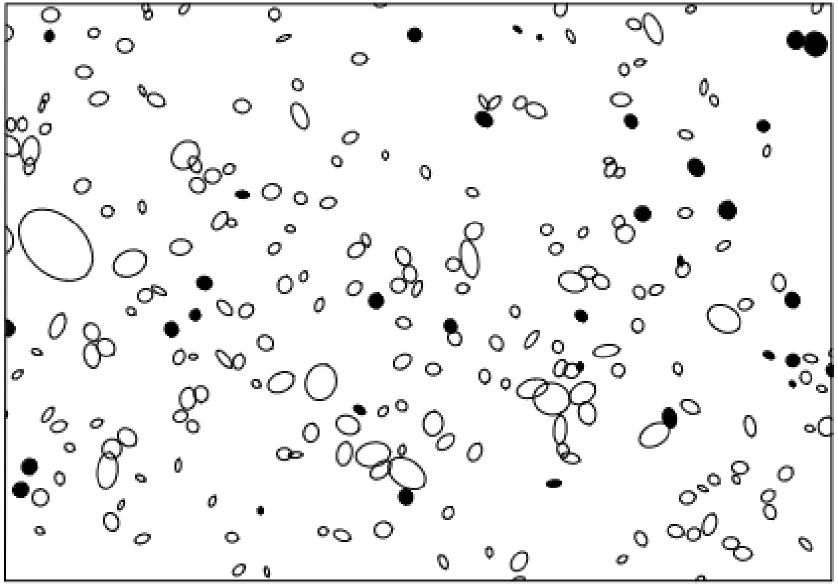
    (and attach these metadata to the dataset to facilitate filtering in discovery)
- What is the optimal HEALPix order ?
- What about coverage of astronomical catalogues ?

Astronomical catalogues

- Extraction of a list of sources from images

Bertin et al. 1996

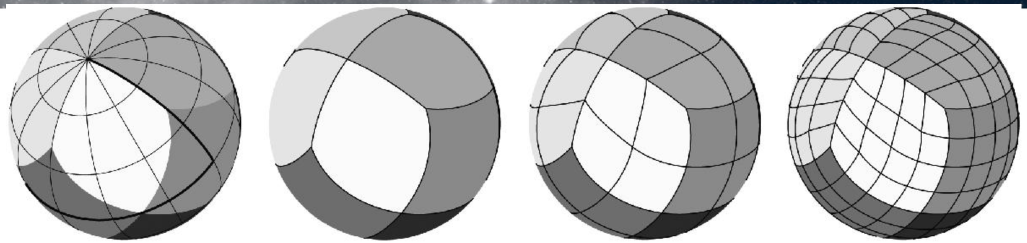


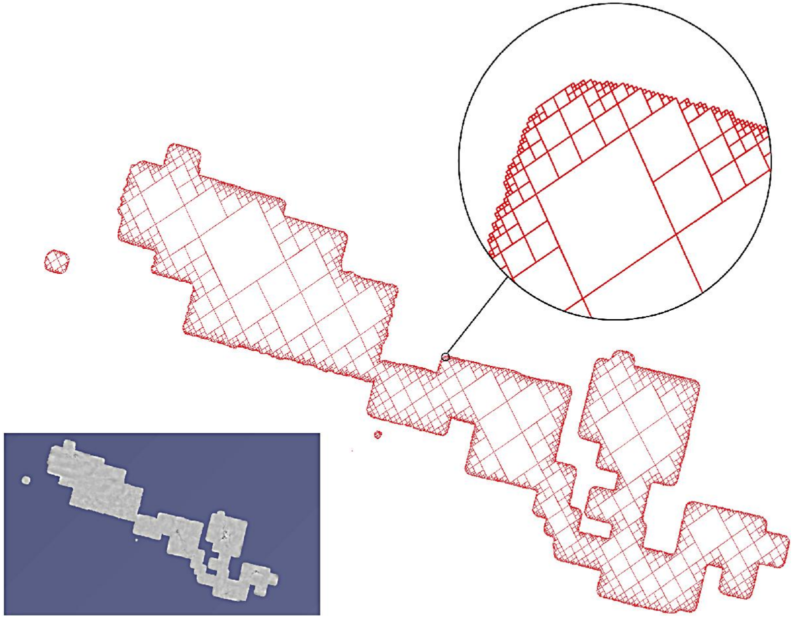


065.014917	+52.334263	0.07	0.07	90	04200358+5220033	15.621	0.068	0.070	18.4	14.980	0.095	0.096	13.2	14.923	0.126	0.126	8.8	AAB	222	221	ccc	665616	4.2	279	798312760
065.013035	+52.334438	0.06	0.06	135	04200312+5220039	14.232	0.038	0.040	66.0	13.789	0.047	0.048	39.7	13.642	0.049	0.050	28.5	AAA	222	221	cc0	665556	4.2	99	798312758
065.003712	+52.330948	0.33	0.29	169	04200089+5219514	16.580	0.147	0.147	7.6	16.271	0.272	0.273	4.0	15.681				BDU	220	110	000	060600	10.7	7	798312754
065.004288	+52.333900	0.06	0.06	90	04200102+5220020	14.073	0.033	0.035	76.4	13.137	0.038	0.039	72.3	12.818	0.029	0.030	60.8	AAA	222	111	000	666666	5.8	262	798312752
065.001656	+52.333672	0.23	0.22	108	04200039+5220012	16.442	0.133	0.133	8.6	15.859	0.192	0.192	5.9	15.003				BCU	220	110	cc0	061600	5.8	82	798312754
065.005662	+52.338749	0.06	0.06	90	04200135+5220194	11.717	0.021	0.024	669.3	10.729	0.028	0.029	664.2	10.412	0.022	0.023	558.0	AAA	222	111	000	666655	7.6	78	798312779
065.009027	+52.339203	0.06	0.06	90	04200216+5220211	13.832	0.029	0.031	95.4	13.381	0.039	0.040	57.7	13.181	0.041	0.042	43.6	AAA	222	111	ccc	666655	7.6	73	798312781
065.012324	+52.339809	0.21	0.19	45	04200295+5220233	16.607	0.160	0.161	7.4	15.859	0.178	0.178	5.9	15.630				CCU	220	110	cc0	060500	7.6	253	798312779
065.018176	+52.337391	0.27	0.27	135	04200436+5220146	16.448	0.127	0.128	8.6	16.087				15.541	0.209	0.209	5.0	BUD	202	101	000	060006	6.4	205	798312766

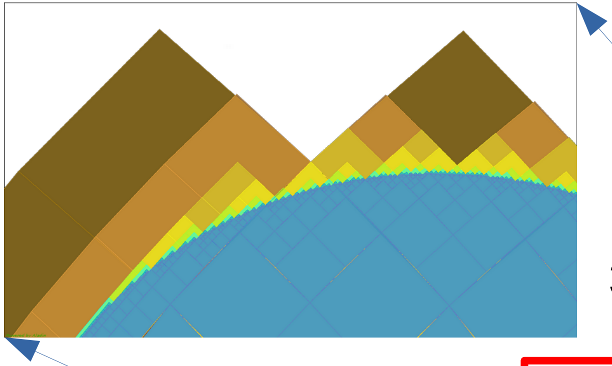
Optimal HiPS & MOC order ?

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	12 288	1°.83
15	2^{15}	1.29×10^{10}	6″.44	6	49 152	55′.0
16	2^{16}	5.15×10^{10}	3″.22	7	196 608	27′.5
17	2^{17}	2.06×10^{11}	1″.61	8	786 432	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
25	2^{25}	1.35×10^{16}	6.29 mas	16	5.15×10^{10}	3″.22
26	2^{26}	5.40×10^{16}	3.15 mas	17	2.06×10^{11}	1″.61

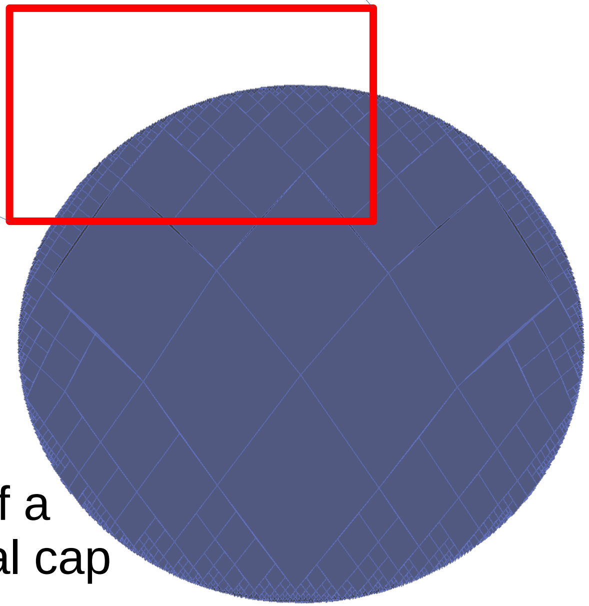




Optimal HiPS & MOC order ?



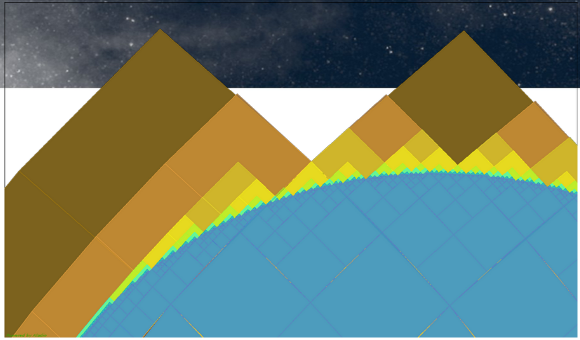
Detail of MOCs at orders 3 (brown) to 10 (blue)



MOC at order 10 of a 15° radius spherical cap

- MOCs can be used to estimate the sky coverage, by adding the areas of all HEALPix pixels.
- The total MOC area is overestimating the true area, because pixels spread on the perimeter.

Optimal HiPS & MOC order ?



- The highest the order of the MOC, the better the precision.

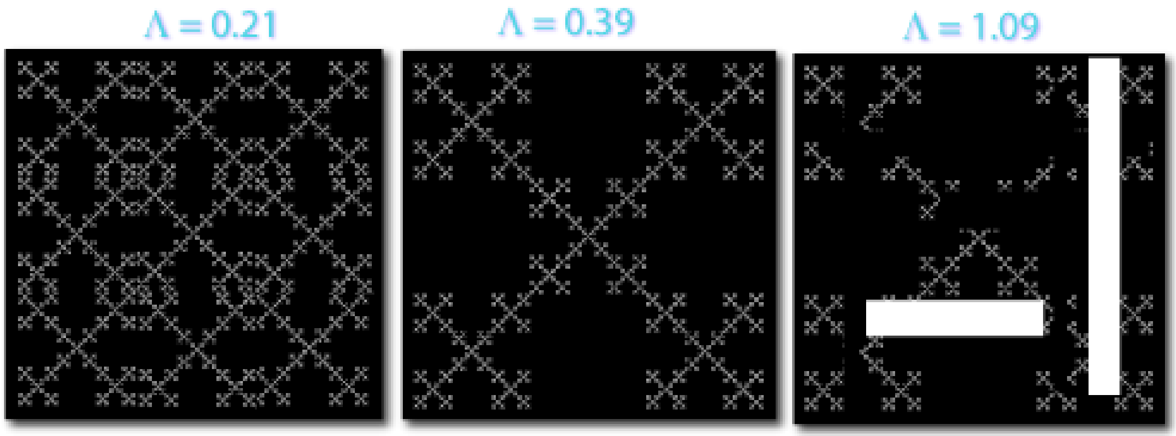
- The area of one HEALPix pixel at order *k* is $\Omega_{pix} = (\pi/3)\, 4^{-k}$
- The estimated area from a MOC at order *k* is related to the true area by a relation of the form $S_k = S_{true}(1 + L \cdot 2^{-k})$ where L is a parameter related to the perimeter of the MOC

- More complex MOCs will have larger values of L
- We define $\mathcal{L} = \log(L)$ as the lacunarity of the dataset

Derriere et al. 2020

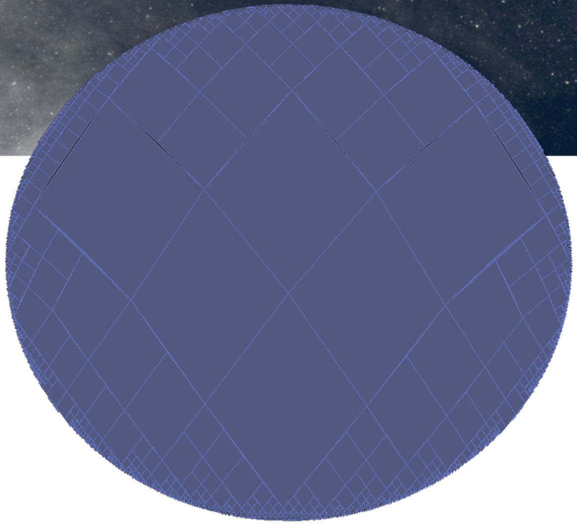
Lacunarity

- Lacunarity (Latin lacuna, meaning "gap") was introduced by Mandelbrot to describe the texture of fractals

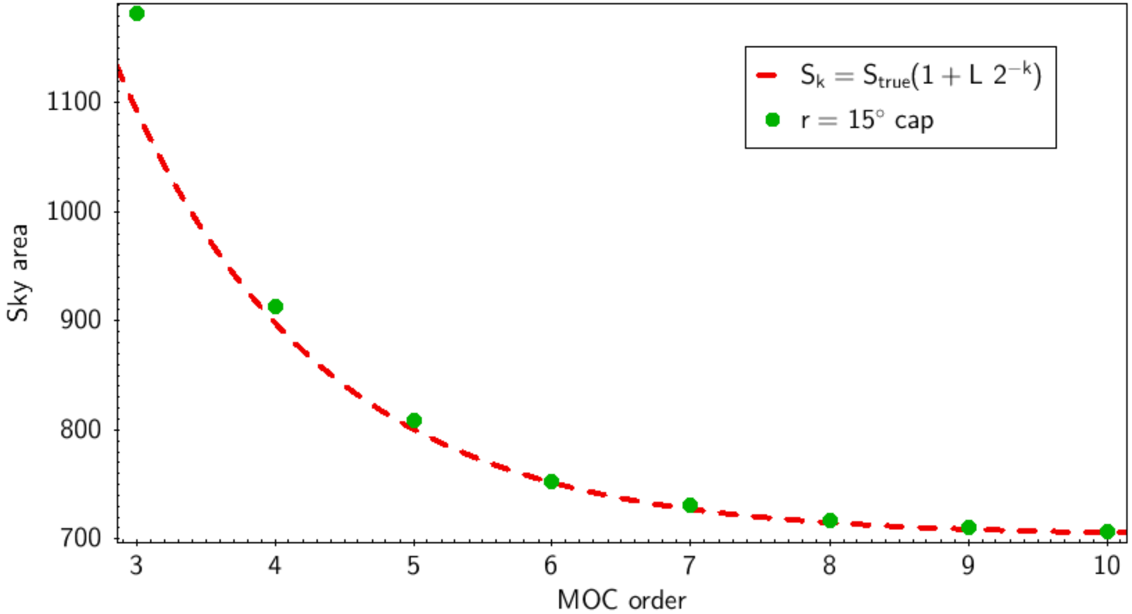


- Extended (see for example Dong (2000)) to describe the distribution of gap sizes in geometric objects :
  - homogeneous objects have low lacunarity
  - heterogeneous objects have high lacunarity

Example on spherical cap



- Fitting the formula $S_k = S_{true}(1 + L \cdot 2^{-k})$ over orders k=7 to 10 gives an estimated S_{true}=704.5deg² (exact area S=702.8deg²) with $\mathcal{L}$ = 0.648.



Coverage of catalogues ?!

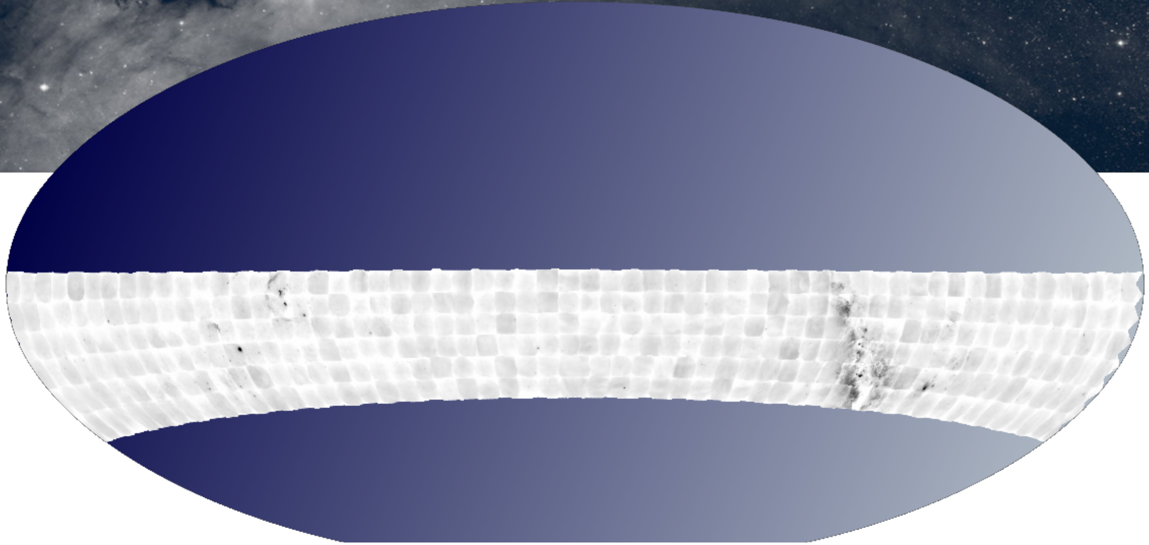
- More difficult than for image surveys.
- Beware of oversampling !!
  - If the density of sources per HEALPix pixel becomes too low (at high orders k), empty pixels cause an underestimation of the area.
- At very high orders k, asymptotic limit (1 source per pixel)

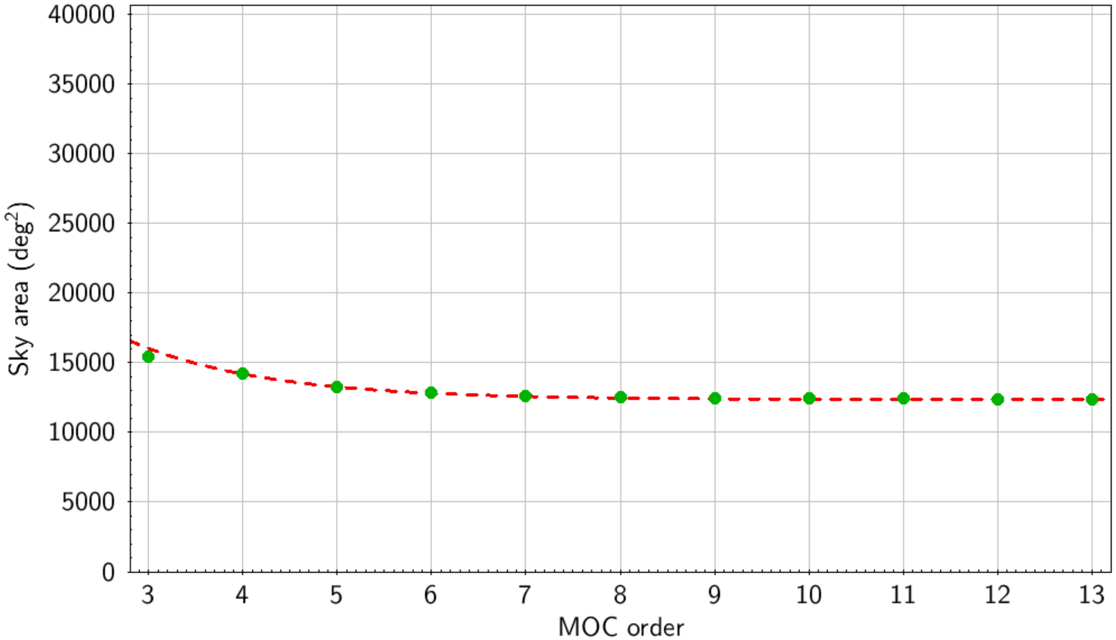
  $S_k = N(\pi/3)\, 4^{-k}$

  with *N* the number of catalogue sources

Example

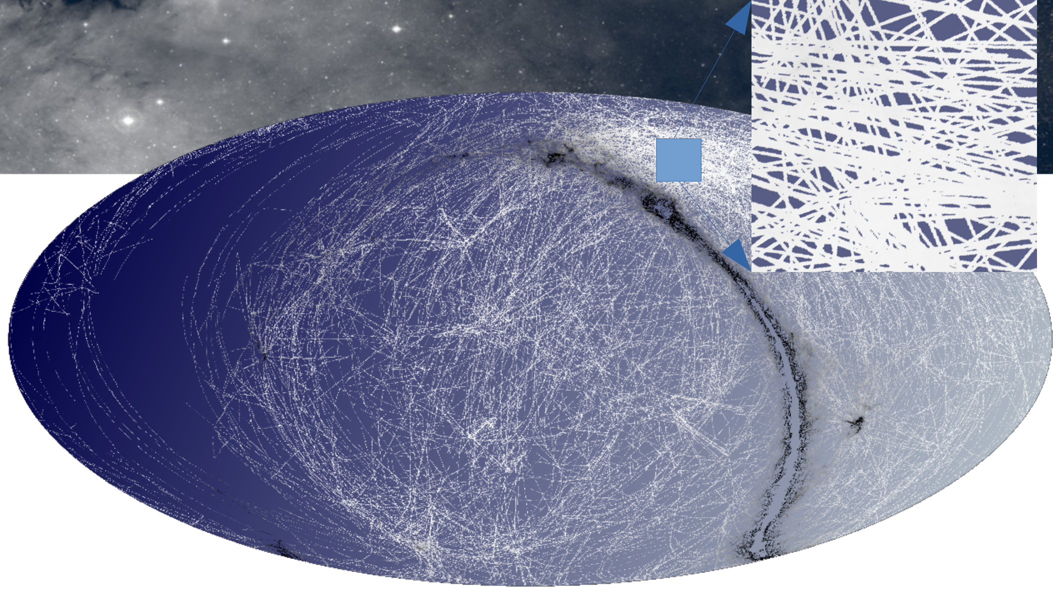
- CDS/P/MAMA/posse
- S_{true} = 12358 deg2
- $\mathcal{L}$ = 0.377

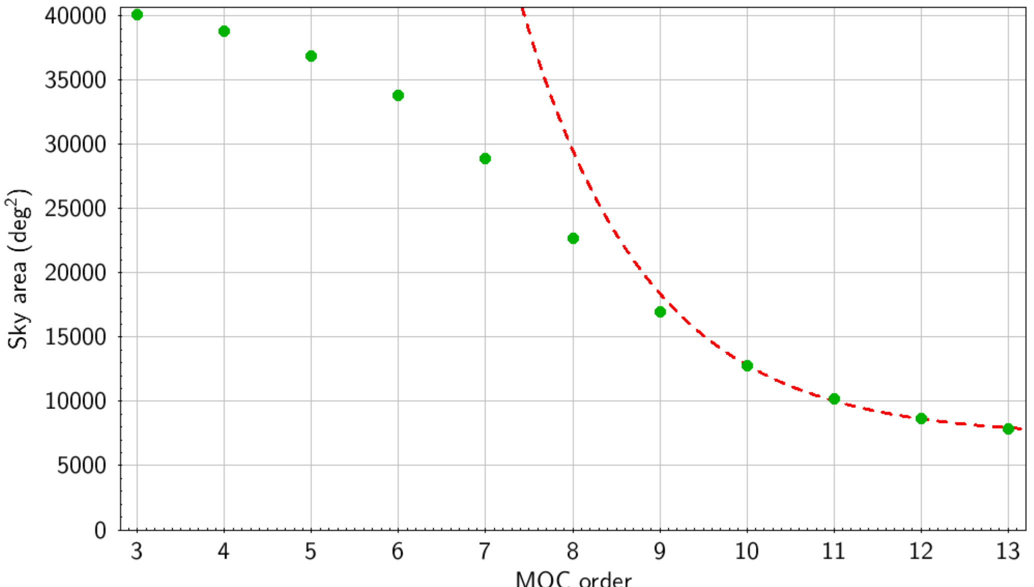




Example

- CDS/P/ISOPHOT/170
- S_{true} = 7252 deg2
- $\mathcal{L}$ = 2.896

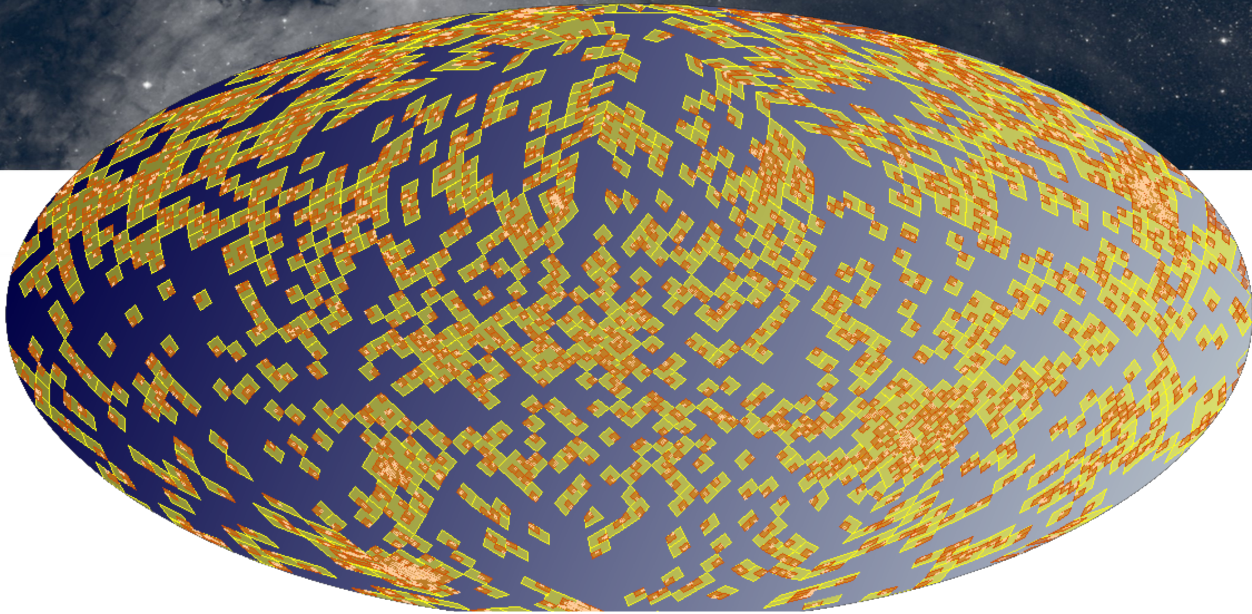


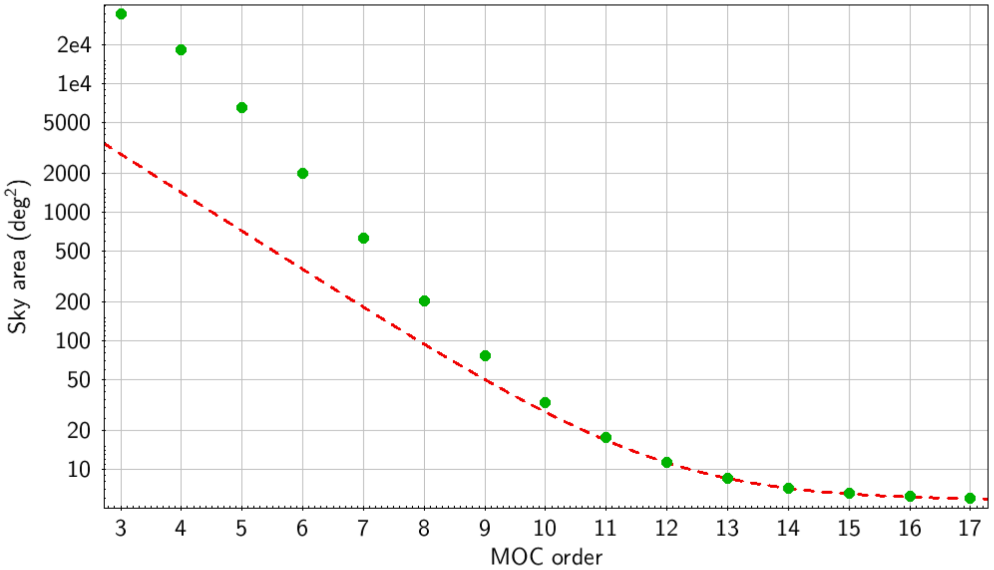


Example

- CDS/P/HST/V

- S_{true} = 5.745 deg2
- $\mathcal{L}$ = 3.595

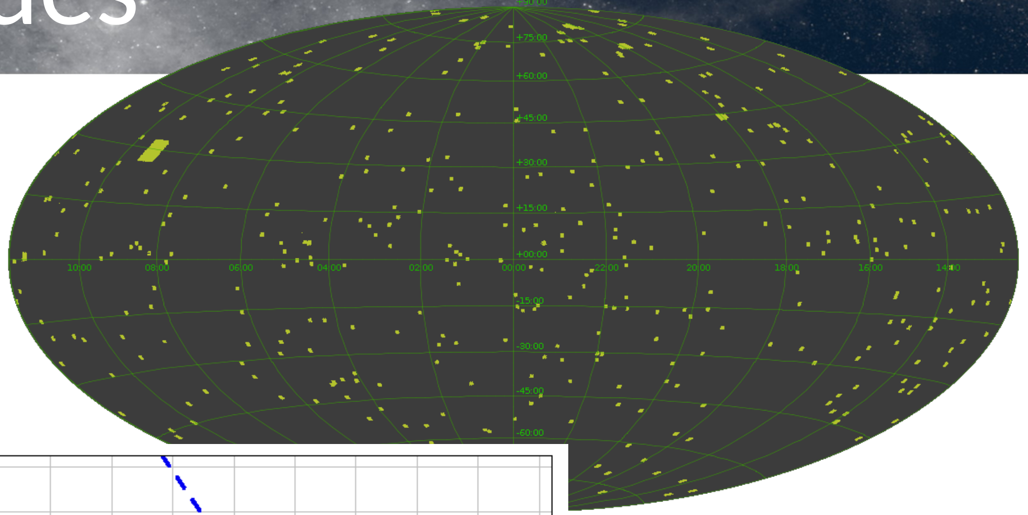


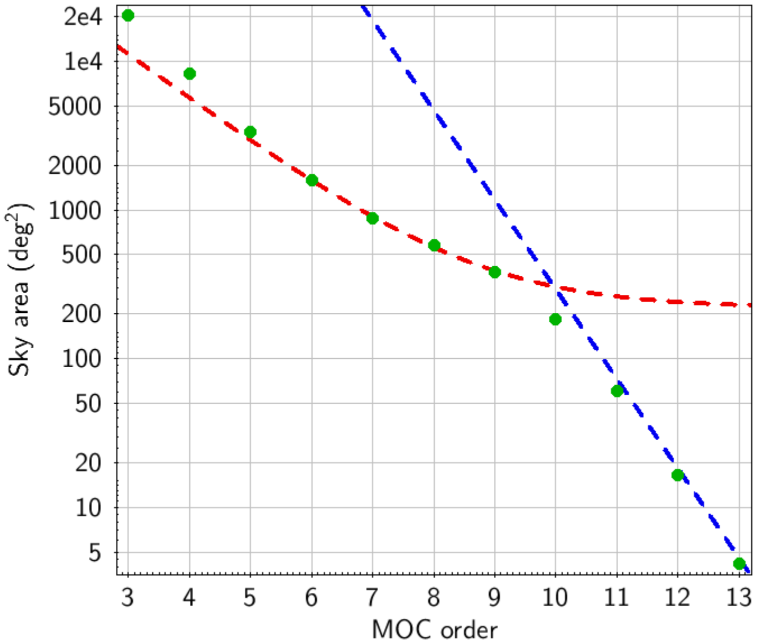


Example for catalogues

- CDS/I/273A/erlcat (de Vegt et al., 2001)
- N=89 422 sources

- S_{true} = 219.5 deg2
- $\mathcal{L}$ = 2.6

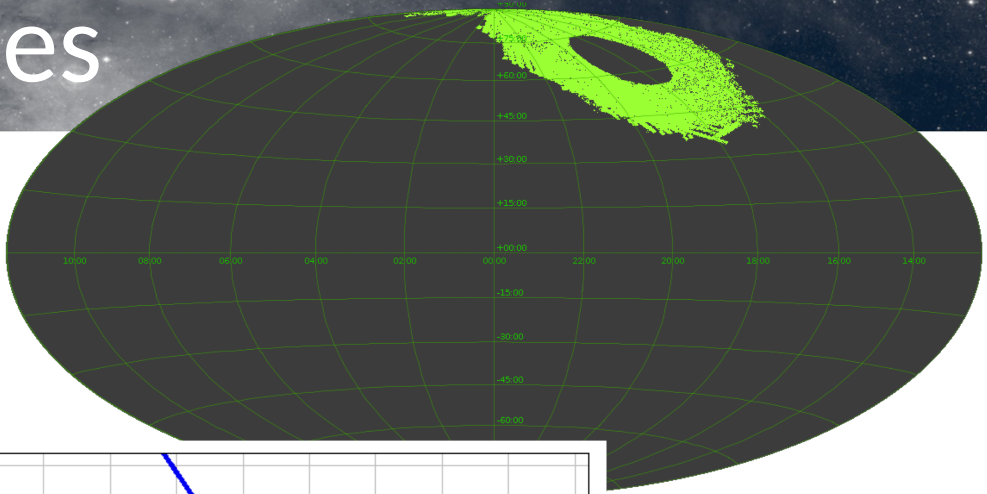


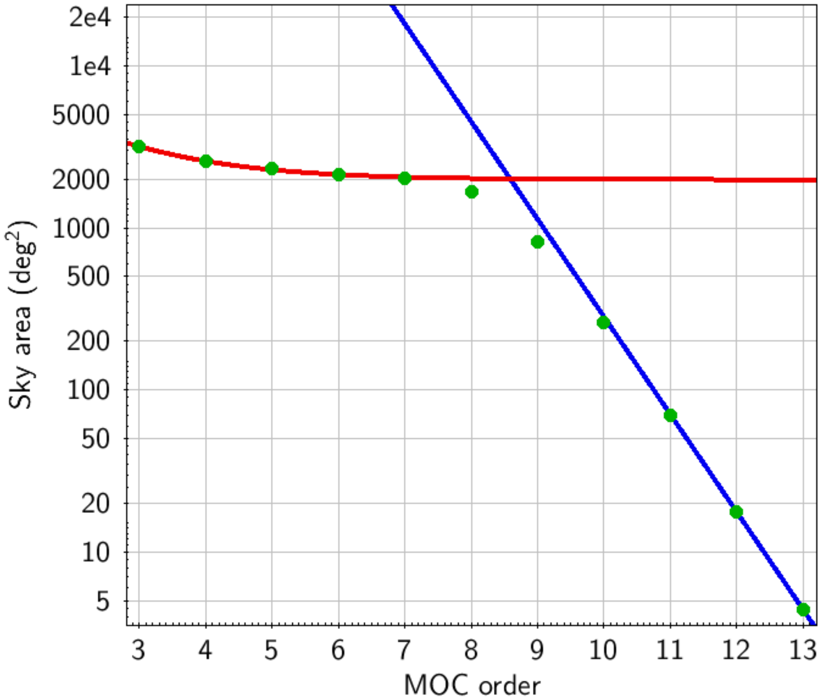


Example for catalogues

- CDS/I/335/table1 (Men et al., 2016)
- N=86 467 sources

- S_{true} = 1984 deg2
- $\mathcal{L}$ = 0.683





Summary

- MOC offer a convenient description of the coverage of astronomical image surveys or catalogues
- Global parameters such as total area (sky fraction) or lacunarity can provide additional insight on the data distribution on the sky
- Combining spatial MOCs with time (or frequency) coverage enables to easily solve difficult problems