CDS Scientific Council, 13-14 November 2013

Research & Development

CDS Team

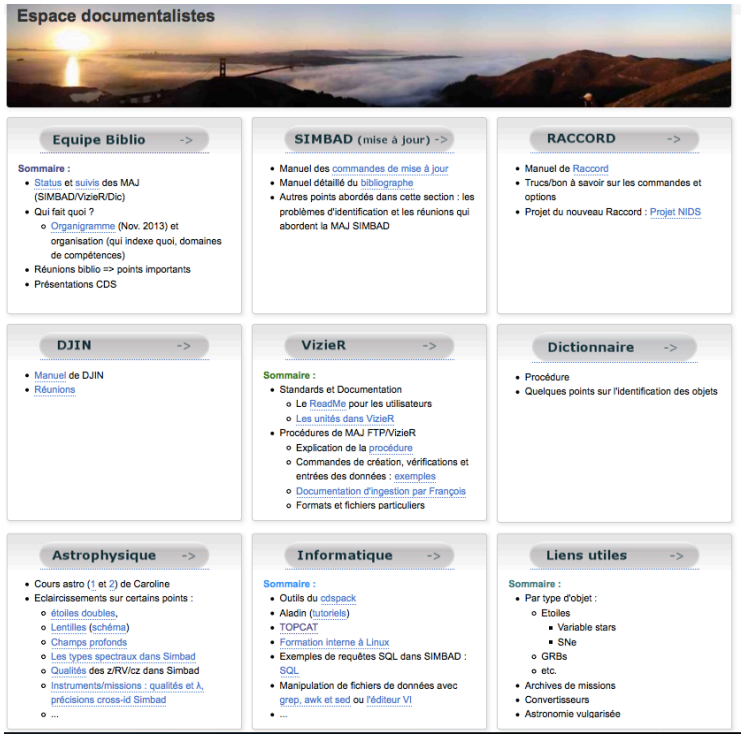
Introduction

- A significant time is spent every year on R&D
  - to facilitate the team everyday work
  - to evaluate emerging and/or interesting technologies
  - to evaluate or to propose IVOA standards (definition, implementation, prototyping)
  - to develop software components useful for the services (and potentially re-usable by the community)

facilitate the team
everyday work

Everyday work of the team is not forgotten !

- Improvement of the Collaborative environment
  - a tool for the daily work
  - and… finally a long term preservation (text format) of a part of the CDS work (meetings, “documentalistes” shared documentation, tutorials, projects, internships, … since 2003)



Everyday work of the team is not forgotten ! (2)

- Support to the services (external) users, cds-question
  - current tool is a simple Webmail
  - involves many people (needs time for discussion/redirection, history following, etc.)
  - ticket management systems (osTicket and RT) are under evaluation to replace the Webmail at the end of 2013
    - expected improvements : instant confirmation ticket creation sent to the user, switch to the people in charge of the concerned service, management of the answer delay, access to user previous questions, quick statistics, etc.

evaluate
emerging and/or
interesting
technologies

Technology evaluation

- Each year we experiment several emerging (or interesting) technologies to increase in-house skills, during internships and short contracts
  - this year : information retrieval, Clouds, CMS (Content management systems), android
  - in the past : android & iOS applications, HTML5 / Javascript, WebGL, Workflows, iRODS, High availability (DRBD), etc.
- Illustrations…

Illustration : Clouds

- Clouds are now common
  - used by people to store their music, pictures, movies, etc.
  - mainly based on « Pay what you use »
  - it is possible to use Public clouds (external provider) to extend Private clouds (if we install a cloud locally) on demand
- A solution to maintain a good availability during the rush hours ?
- Our use case : HEALPix image server in the clouds accessed by Aladin
  - performances were correct.
  - cost depends on the provider (from 1000 to 4000 euros per month for a “virtual server” equivalent to the infrastructure dedicated to the HEALPix image server (25 TB) at CDS) and most of the cost is linked to the storage and not to the data upload / download
- Conclusion : not adapted just to ensure availability

Illustration : android application SkyObjects

- A small android “app” (user interface in english and french) : search an object with the Sesame name resolver and point it

Illustration : Content Management Systems

- Test and prototyping of 3 CMS followed by the development (released in June) with Drupal of the new Euro-VO website (hosted and maintained by CDS since one year)

Illustration : Information retrieval

- The aim of this work started in April 2013 on an internship is to go further than simple queries based on a keyword.
- It will certainly be pursued because of its potential impact.
- The study was based on ElasticSearch, Solr and Lucene.
  - Solr (Apache open source) search server includes powerful full-text search, hit highlighting, faceted search, near real-time indexing, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search.

Illustration : Information retrieval (2)

Size and popularity of the VizieR catalogues used during the experiment

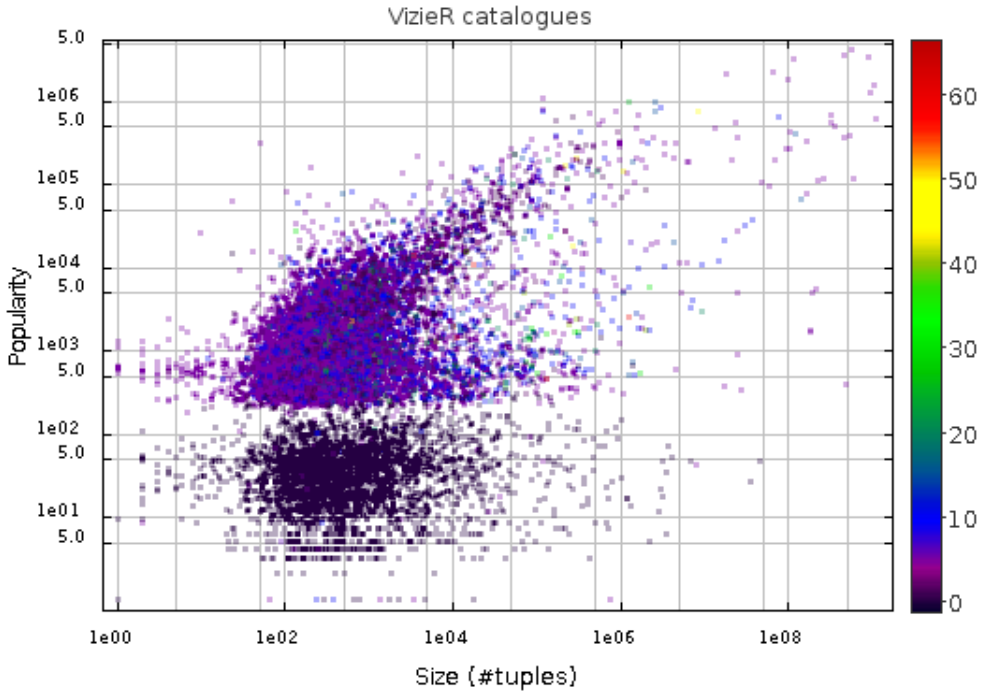
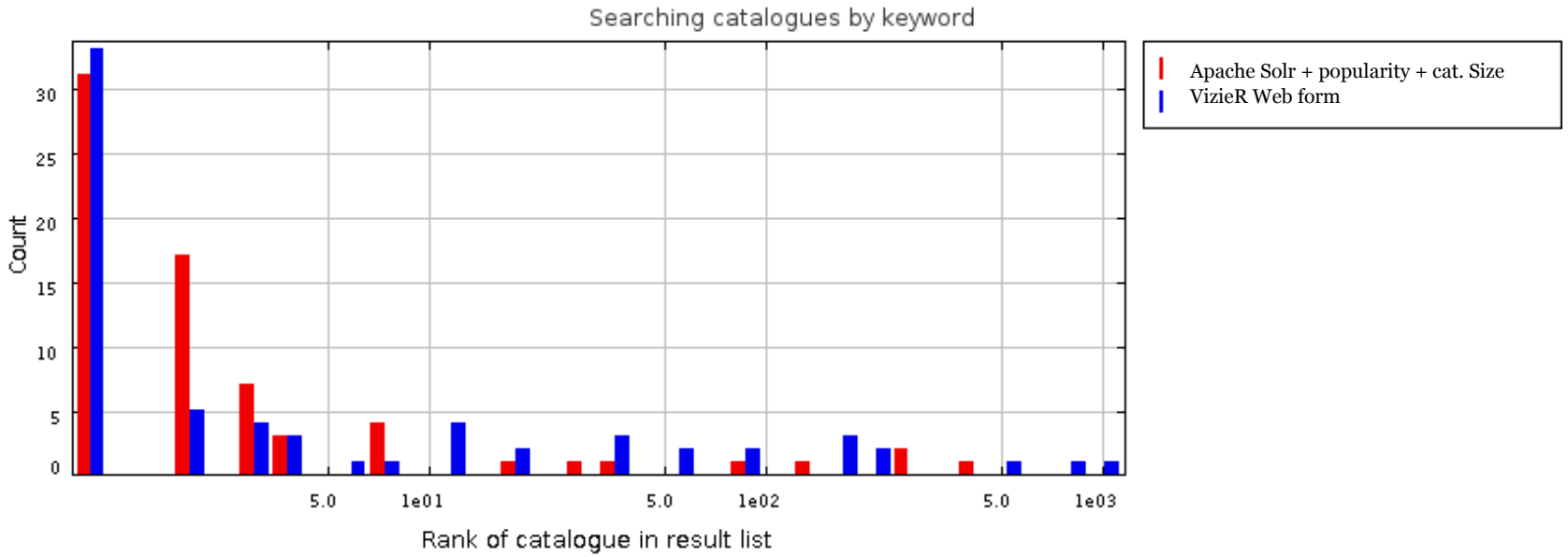


Illustration : Information retrieval (3)

Set of 50 common keywords used to query 70 relevant VizieR catalogues

2mass	iras	NOMAD	Kharchenko
2MASS	Tycho2	renson et al	parallax
hipparcos	cutri	temperature	abundance
usno	kharchenko	Salim	rc3
Hipparcos	gsc	houk	diameters
USNO	rosat	distance	MSX
IRAS	NVSS	nvss	Glushneva
ucac2	ucac	msx	Cutri
ascc	SDSS	hip	hde
tycho	henry	renson	UCAC2
nomad	gezari	radial velocity	nordstorm
sdss	GSC	landolt	
veron	Tycho	RC3	

Illustration : Information retrieval (4)

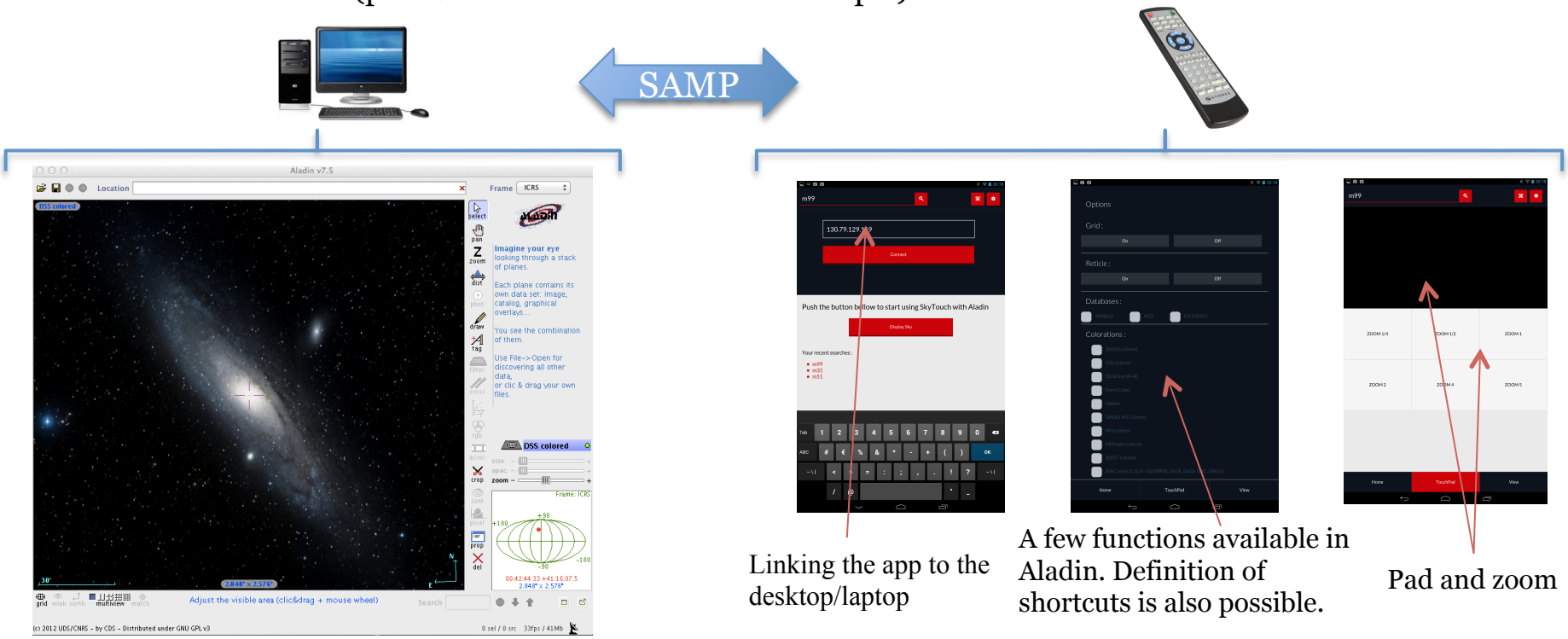


It would allow in the short/medium term queries
with multiple keywords (object names, coordinates, concepts),
and on a longer term queries in natural language.

evaluate or propose IVOA standards (definition, implementation, prototyping)

SkyTouch / SAMP (IVOA Simple Application Messaging Protocol)

- SkyTouch (running on iOS and android) is a study of a bridge between desktop / laptop applications (Aladin in our use case) and smartphones/tablets based (used as remote control) on the IVOA SAMP standard (possible use in Education for example).

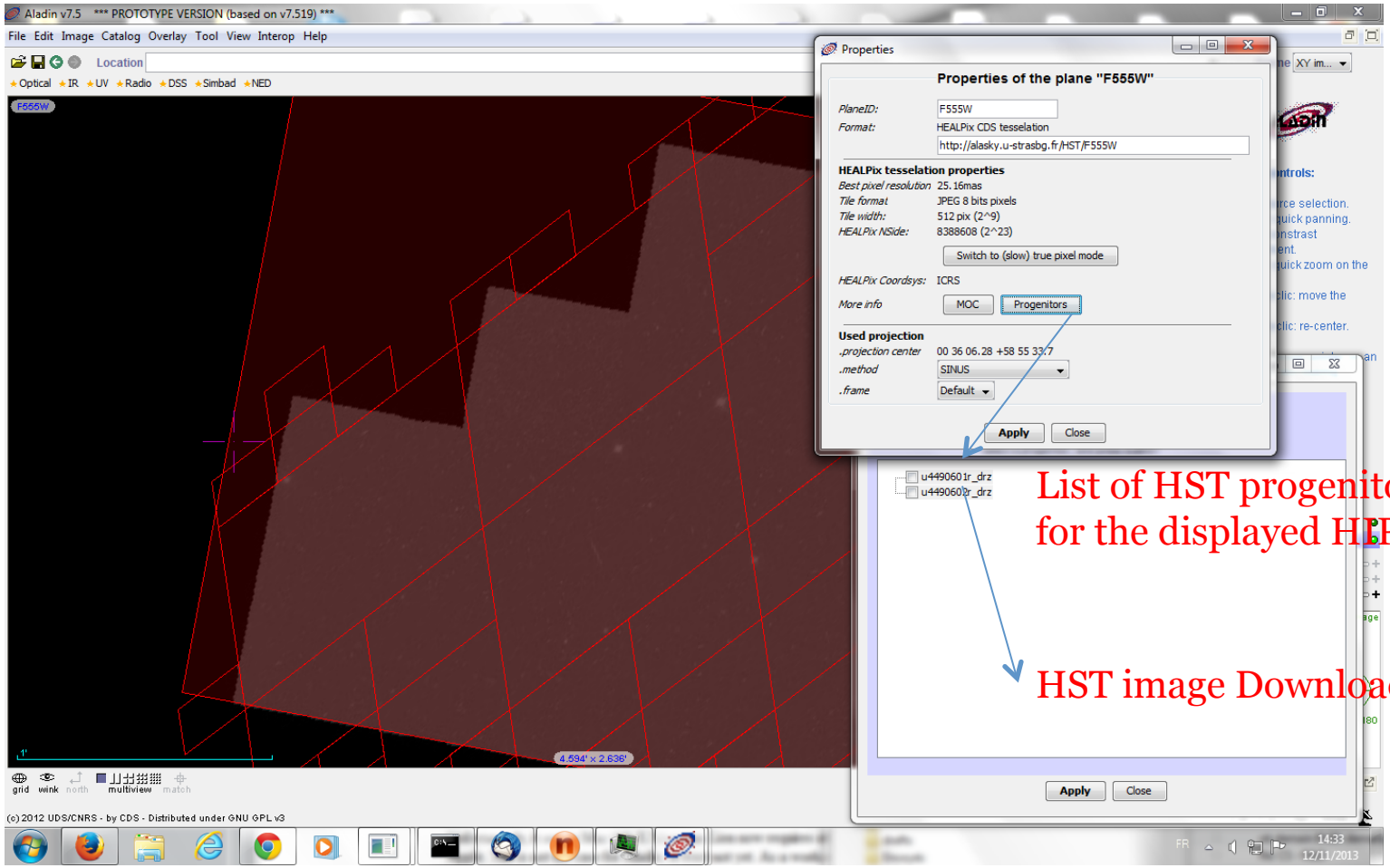


Prototyping DataLink and Cube AccessData IVOA protocols

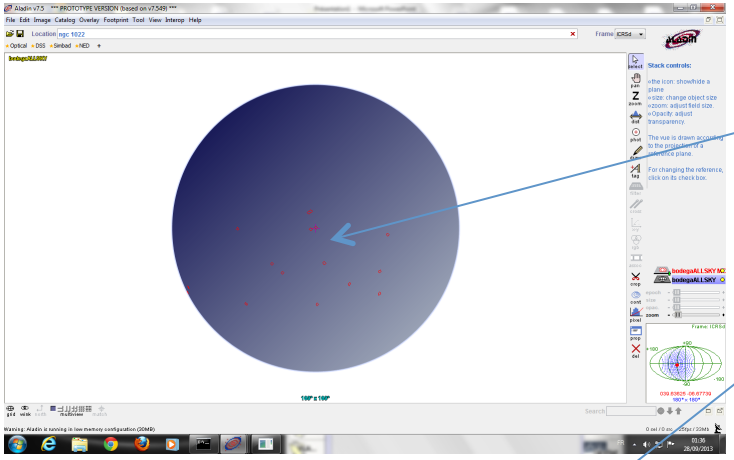
- DataLink is an emerging protocol for services helping to link miscellaneous resources to DataSets.
  - working draft status. CDS and Strasbourg observatory (XMM) involved
  - prototyped at CDS. Accessible via Aladin Allsky Healpix mode. Link to HST (or other) progenitors, cube metadata and cube access methods

- Cube AccessData is part of a set of protocols (with ImageDM and SIAV2) designed for interoperability of data cubes
  - working draft is emerging. CDS prototype implements 3D cutout, cube averaging and regridding. Built on Bodega collection.
- HST progenitor prototype in beta version
- Datalink for Cubes and Cube AccesData prototype bounded together.

HST progenitor



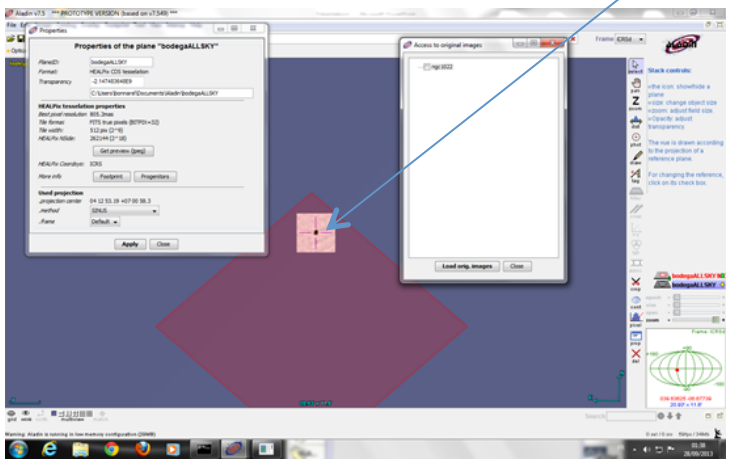
CDS prototype for DataLink and Cube access within Aladin

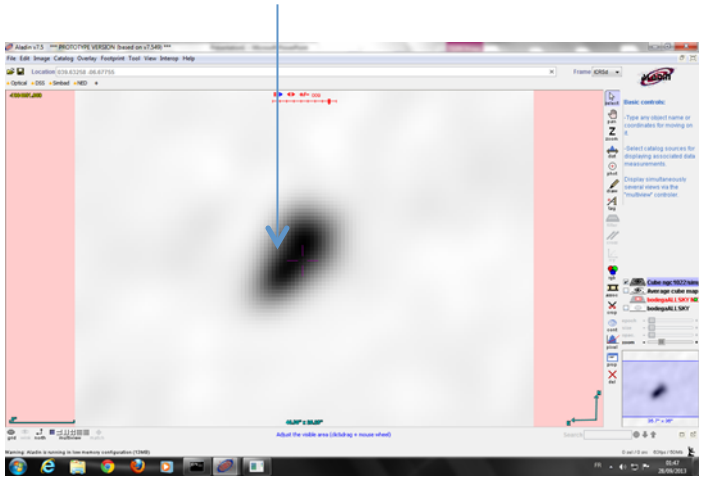


1) Centring the view on NGC 1022

2) Zoom on the NGC 1022 average image in Healpix representation. Progenitor button makes ObsID appear in Access window. Links will be retrievable

3) Cube cutout in action





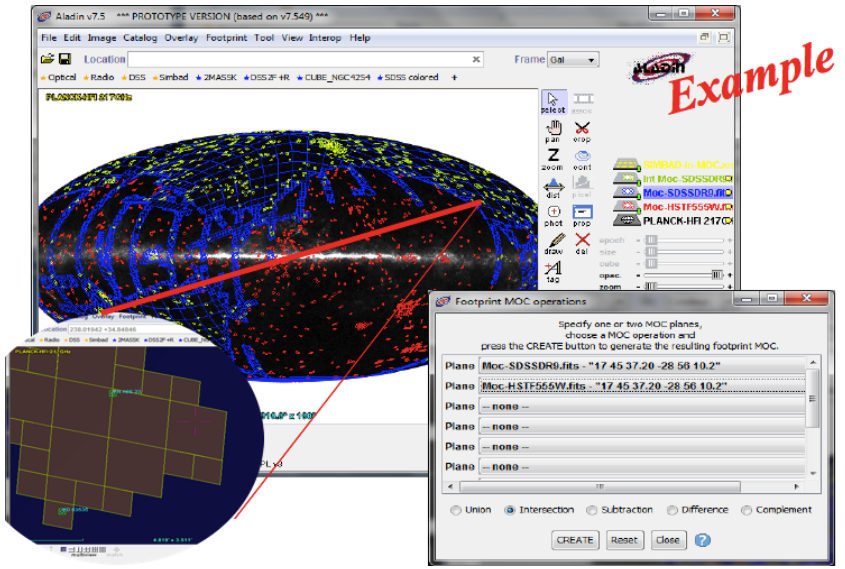
MOC : Multi Order Coverage Map

- A MOC : a simple and powerful method to specify sky regions
- Goals : provide very fast comparisons and data access methods
- Designed by CDS in 2011

« Please give me all Simbad objects observed by HST F606W and SDSS DR9 »

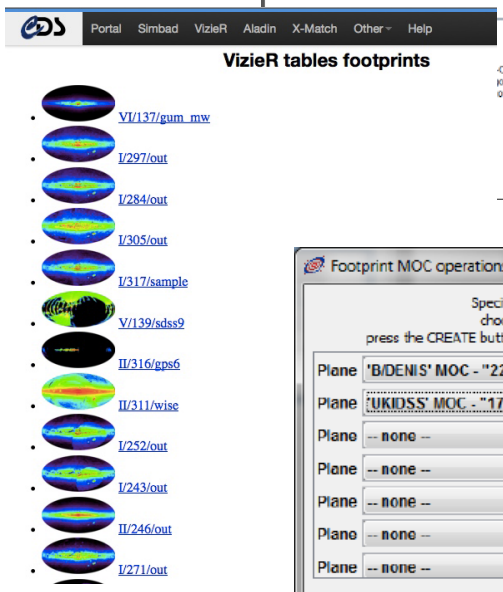
1. Load
   MOC-HST F606W
   MOC-SDSS DR9
2. Compute
   MOC intersection
3. Query Simbad
   by MOC

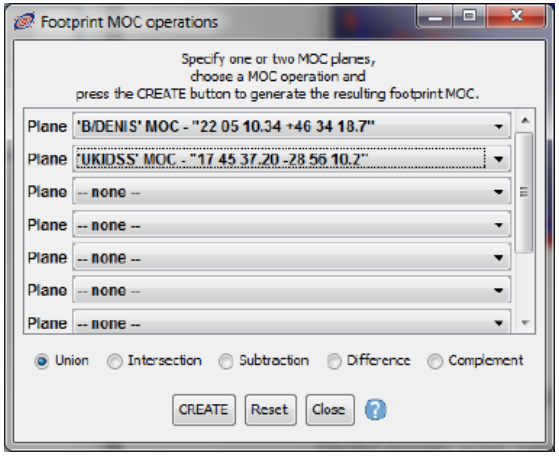
=> Realized in **5s**



2013 MOC progress

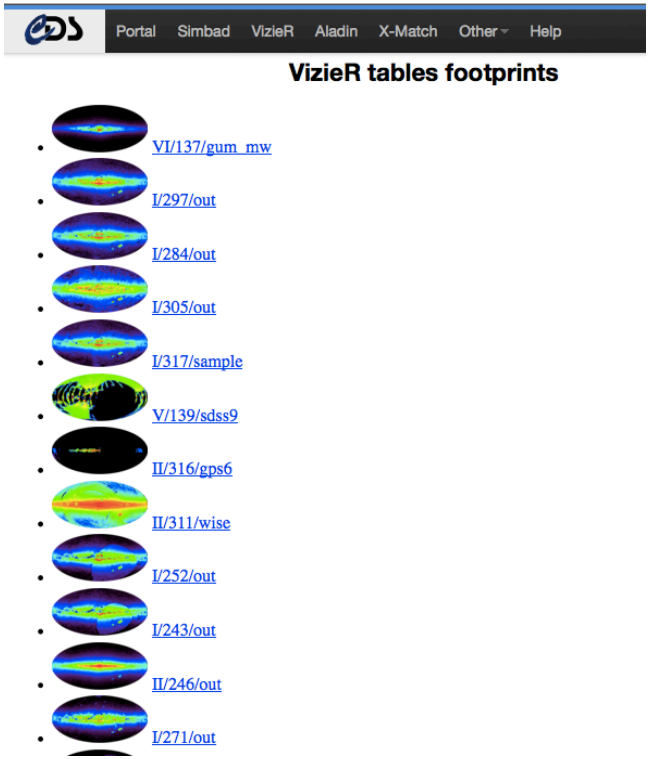
- Oct 2013: IVOA draft ready to **Proposed Recommendation** (Hawaii Interop)
- Sep 2013: s/w release of the java MOC library => on IVOA site
- Dec 2013: Aladin V8 fully MOC compatible
- MOC generation :
  - all CDS data sets (catalogues & pixel surveys)
  - ROE => UKIDSS, VVV, VMC, VHS, VIDEO,
  - CADC => HST
- Three MOC clients : Aladin, TOPCAT, MIZAR/CNES





MOC and density maps of VizieR tables

- Available from http://alasky.u-strasbg.fr/footprints/tables/vizier/



- Updated every week to compute MOC and density maps for new/updated VizieR tables

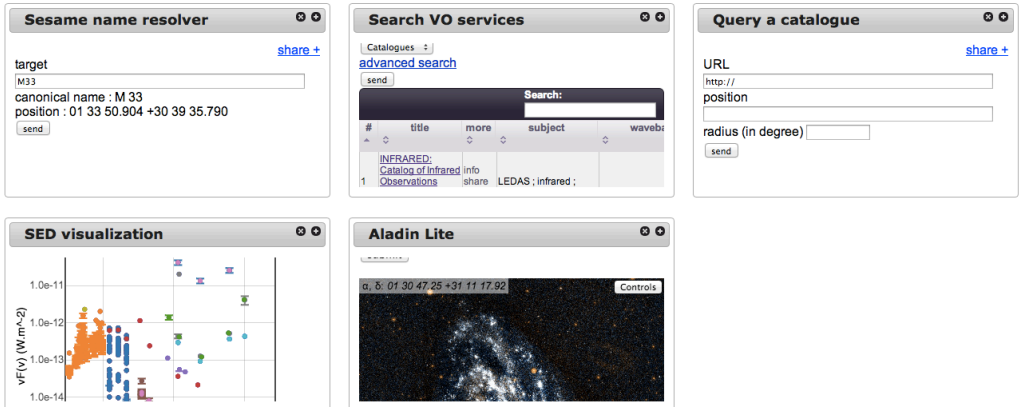
develop software components useful for the services (and potentially re-usable by the community)

Modular widgets

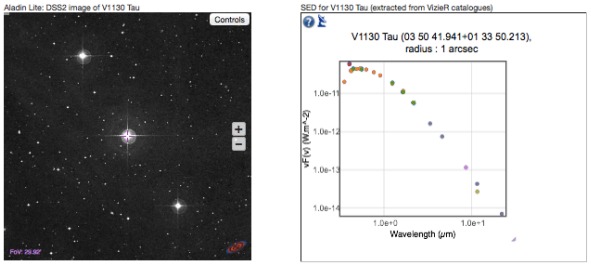
- Build a set of widgets
  - running in the browser
  - easily shareable
  - easily embeddable in a Web page
  - inter-connected
- Each widget targets one specific need: search for a service, query a service, visualize results

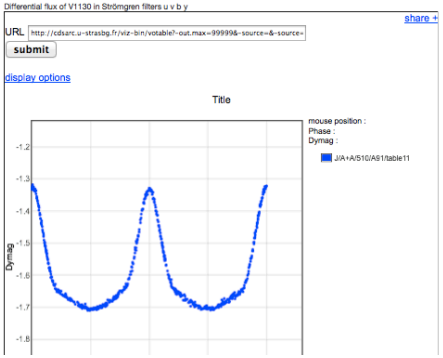
Usages

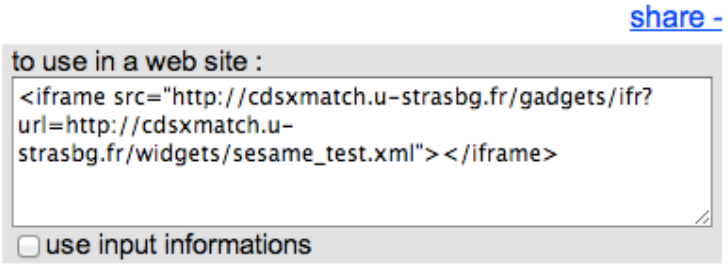
- Customizable user portal *(iGoogle for astronomy)*
- Portal dedicated to one astronomical object



Multiple widgets create a portal







Aladin Lite

- Assessment of Javascript platform
  - which technology: HTML5 canvas vs. WebGL
  - performance tests
  - no direct access to local disk
- HEALPix library converted to Javascript
- More information:
  - see earlier presentation
  - demos later this afternoon

Future

- Big Data (involved technologies around storing, querying, …), learn to surf the data wave
- Augmented reality (first try this year but no result, second try next year with Oculus VR)
- Other topics to define… technologies are evolving very quickly…

All the contributors

- Yann Bisch, Thomas Boch, François Bonnarel, Sébastien Derriere, Pierre Fernique, Kilian Hett, Maxime Heckel, Romain Houpin, Gilles Landais, Mireille Louys, Laurent Miguel, Emmanuelle Perret, André Schaaff, Anne-Camille Simon, Thomas Streiff, Nicolas Viard