

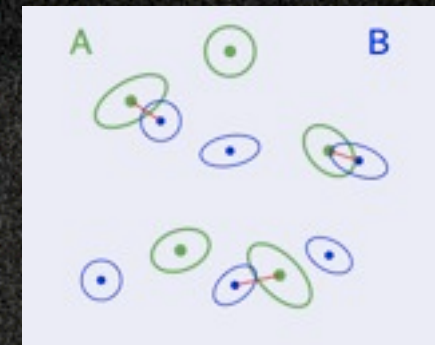
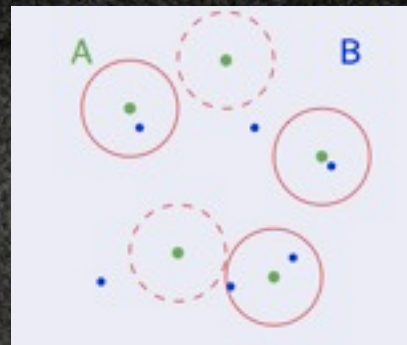
# CDS cross-match service

Thomas Boch  
François-Xavier Pineau



# Introduction

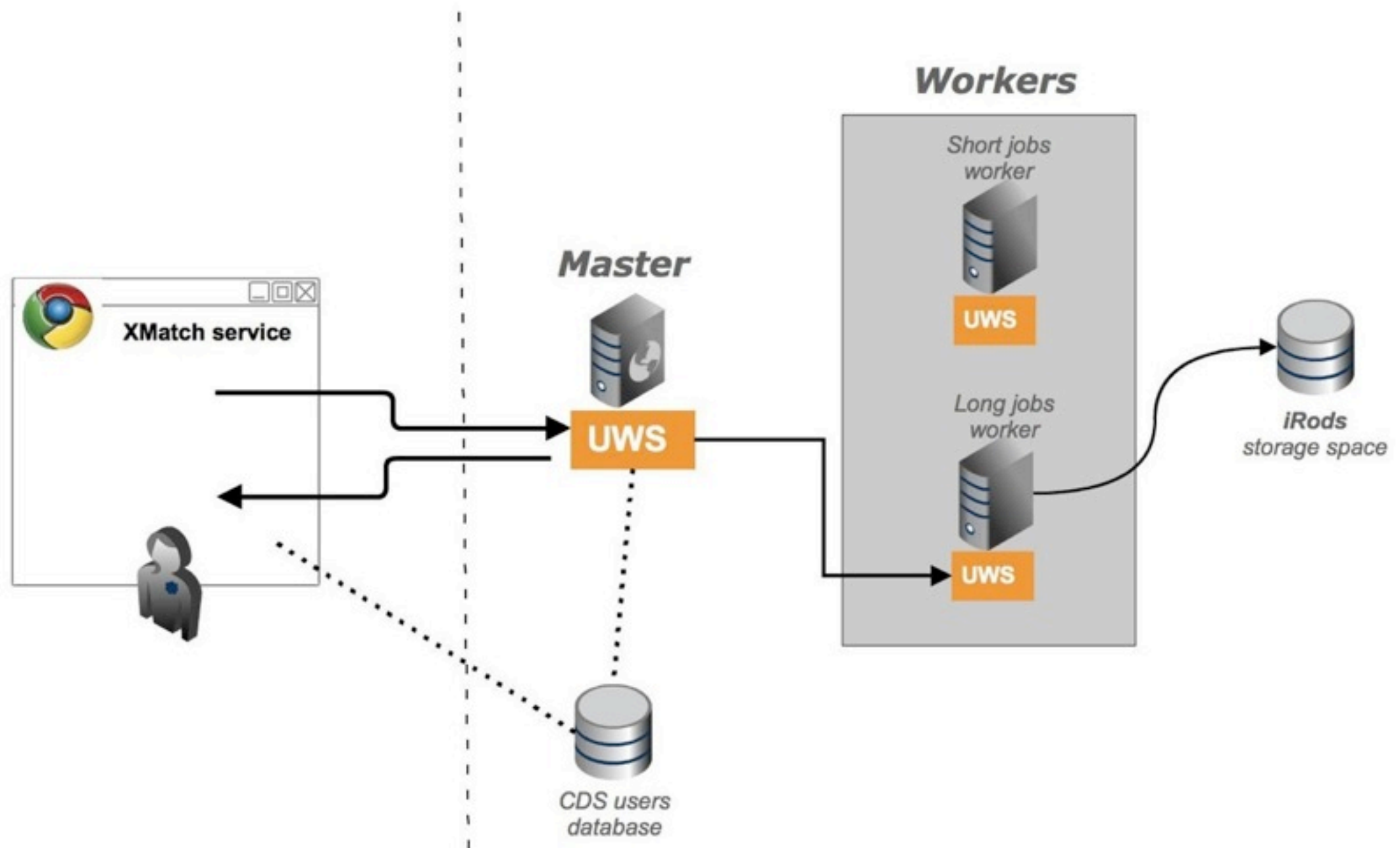
- ◆ New service, released on November 7th 2011 ([cdsxmatch.u-strasbg.fr](http://cdsxmatch.u-strasbg.fr))



- ◆ Purpose
  - ◆ Positional cross-identification of sources in 2 tables, among:
    - ◆ 10,000+ VizieR tables with positions (including large surveys)
    - ◆ a view of SIMBAD data
    - ◆ user-uploaded tables (in VOTable, FITS or CSV)
  - ◆ Users retrieve result in VOTable, CSV or ASCII format
- ◆ Focus on large-scale cross-match



# Architecture

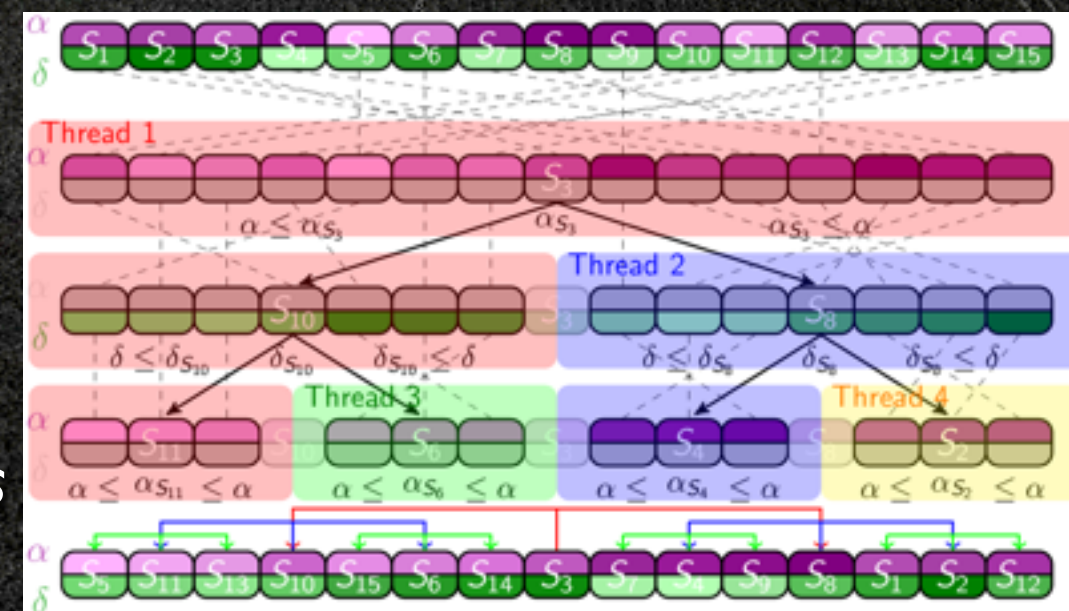
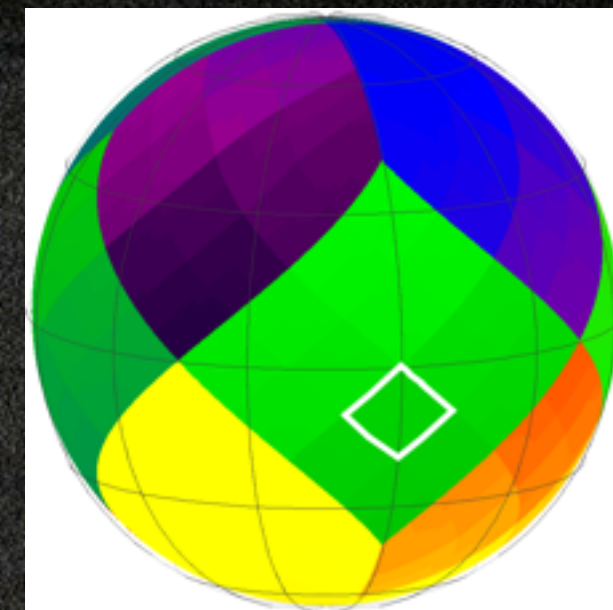


**UWS:** Virtual Observatory standard for execution of asynchronous jobs  
**iRods:** distributed storage system



# Algorithms

- ◆ Different strategies according to catalogs size
- ◆ For 2 large catalogs (>20M sources):
  - ◆ **HEALPix** partitioning (region of  $15^\circ$  by  $15^\circ$ )
  - ◆ XMatch done HEALPix cell by HEALPix cell
  - ◆ In memory creation of a **kd-tree** on catalog B equatorial coordinates
  - ◆ Multithreaded queries on the **kd-tree** from streamed catalog A sources
- ◆ **kd-tree** : data structure for partitioning data points in a  $k$ -dimensions space





# Performances

*Cross-match at 5 arcsec :*

<b>Table 1</b>	<b>Table 2</b>	<b>Computation time</b>	<b>Result generation</b>	<b>Result size</b>	<b>Total time</b>
<b>SDSS DR7</b> <i>357M rows</i>	<b>2MASS</b> <i>470M rows</i>	7 min	11 min	<b>13 GB</b>	<b>18 min</b>
<b>DENIS</b> <i>355M</i>	<b>2MASS</b> <i>470M</i>	11 min	51 min	<b>58 GB</b>	<b>1 hour 2 min</b>
<b>GLIMPSE</b> <i>104M</i>	<b>NOMAD</b> <i>1.1 billion</i>	6 min	17 min	<b>19 GB</b>	<b>23 min</b>
<b>SIMBAD</b> <i>7M</i>	<b>USNOB1</b> <i>1 billion</i>	3 min	1 min	<b>1 GB</b>	<b>4 min</b>
<b>List of</b> <i>40k positions</i>	<b>SIMBAD</b> <i>7M</i>	1 second	4 seconds	<b>10 MB</b>	<b>5 sec</b>



# Performances

*Cross-match at 5 arcsec :*

<b>Table 1</b>	<b>Table 2</b>	<b>Computation time</b>	<b>Result generation</b>	<b>Result size</b>	<b>Total time</b>
<b>SDSS DR7</b> <i>357M rows</i>	<b>2MASS</b> <i>470M rows</i>	7 min	11 min	<b>13 GB</b>	<b>18 min</b>
<b>DENIS</b> <i>355M</i>	<b>2MASS</b> <i>470M</i>	11 min	51 min	<b>58 GB</b>	<b>1 hour 2 min</b>
<b>GLIMPSE</b> <i>104M</i>	<b>NOMAD</b> <i>1.1 billion</i>	6 min	17 min	<b>19 GB</b>	<b>23 min</b>
<b>SIMBAD</b> <i>7M</i>	<b>USNOBI</b> <i>1 billion</i>	3 min	1 min	<b>1 GB</b>	<b>4 min</b>
<b>List of</b> <i>40k positions</i>	<b>SIMBAD</b> <i>7M</i>	1 second	4 seconds	<b>10 MB</b>	<b>5 sec</b>



# Performances

*Cross-match at 5 arcsec :*

<b>Table 1</b>	<b>Table 2</b>	<b>Computation time</b>	<b>Result generation</b>	<b>Result size</b>	<b>Total time</b>
<b>SDSS DR7</b> <i>357M rows</i>	<b>2MASS</b> <i>470M rows</i>	7 min	11 min	<b>13 GB</b>	<b>18 min</b>
<b>DENIS</b> <i>355M</i>	<b>2MASS</b> <i>470M</i>	11 min	51 min	<b>58 GB</b>	<b>1 hour 2 min</b>
<b>GLIMPSE</b> <i>104M</i>	<b>NOMAD</b> <i>1.1 billion</i>	6 min	17 min	<b>19 GB</b>	<b>23 min</b>
<b>SIMBAD</b> <i>7M</i>	<b>USNOBI</b> <i>1 billion</i>	3 min	1 min	<b>1 GB</b>	<b>4 min</b>
<b>List of</b> <i>40k positions</i>	<b>SIMBAD</b> <i>7M</i>	1 second	4 seconds	<b>10 MB</b>	<b>5 sec</b>



# Usage statistics

In 10 months :

- ◆ 2598 cross-match jobs submitted
- ◆ by 418 users
- ◆ 16.5 billion associations found (2.8 TB)

User loyalty :

- ◆ 65 users have submitted more than 10 jobs
- ◆ 17 users have submitted more than 30 jobs
- ◆ 1 user has submitted 101 jobs

Most popular datasets : 2MASS, SIMBAD, USNO-B1, SDSS DR7, WISE



# Planned developments

- ◆ Evolving service, taking into account users' feedback
- ◆ Programmatic access for external tools
  - ◆ through HTTP calls
  - ◆ will be used by TOPCAT
- ◆ Revamped web interface with new features :
  - ◆ selection of output columns
  - ◆ non-positional criteria to filter out output :
    - ◆ e.g: **J-H>2**



# Demonstration

The screenshot displays the CDS X-Match Service interface. At the top, there is a navigation bar with logos for CDS, SIMBAD, VizieR, Aladin, Catalogs, Dictionary, Biblio, Tutorials, and Resources. Below this is the 'CDS X-Match Service' header with user information 'thomas' and links for 'Preferences', 'Admin', and 'Logout'. The main content area is titled 'Choose tables to cross-match' and features two input fields: 'xmmsl1D\_total.fits.gz' (with source buttons 'VizieR', 'SIMBAD', 'My store') and 'SDSS DR8' (with source buttons 'VizieR', 'SIMBAD', 'My store'). Below these, two table cards are shown: one for 'xmmsl1D\_total.fits.gz' with 29,389 rows, and another for 'The SDSS Photometric Catalog, Release 8 (Adelman-McCarthy+, 2011)' with 794,013,950 rows. A 'Hide options' section is expanded, showing 'Cross-match criteria' with 'By position' selected (radius: 5 arcsec) and 'By position including error' (sigma: 0.43935, completeness: 99.73%, max distance: 5 arcsec). The 'Cross-match area' section has 'All sky' selected.