# The CDS cross-match service

Thomas Boch, François-Xavier Pineau, Sébastien Derrière

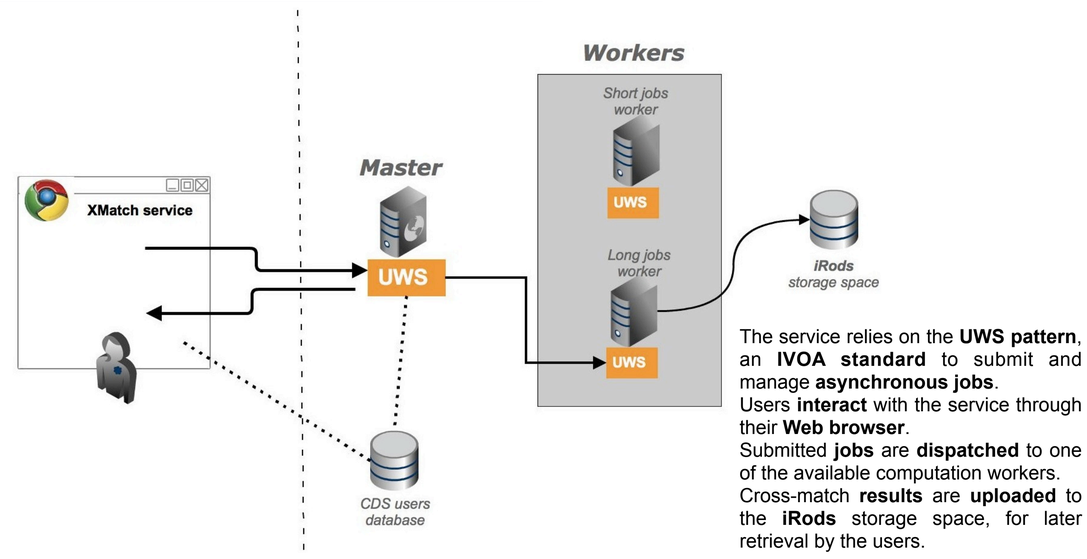
Centre de Données astronomiques de Strasbourg

## 1. Summary

The **CDS** has **released a cross-match service** allowing astronomers to **efficiently cross-identify** sources between **very large catalogues** (up to 1 billion rows) or between a user-uploaded list of positions and a large catalogue.
Cross-match jobs can be submitted through a Web application.
**Popular cross-identifications**, such as SDSS vs. 2MASS, are **pre-computed** in order to accelerate these queries.

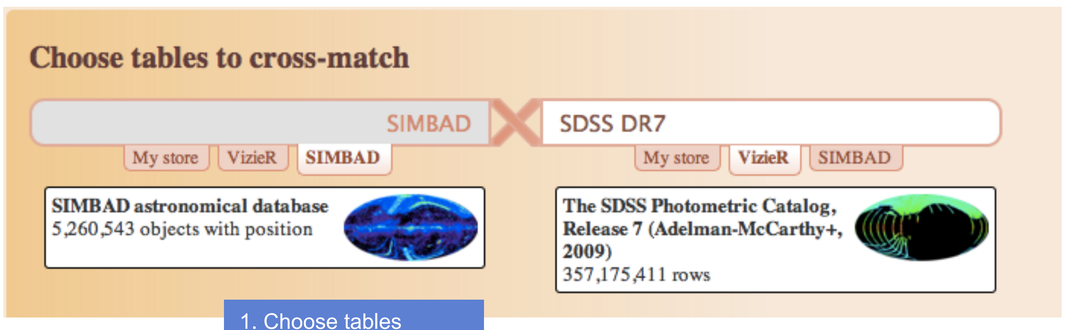
This service has been built on top of the methods described at *ADASS 2010* by Pineau et al., which are based on a **dedicated binary table file format** and the **HEALPix pixelisation** associated to **specialized KD-trees**.

## 2. Service architecture

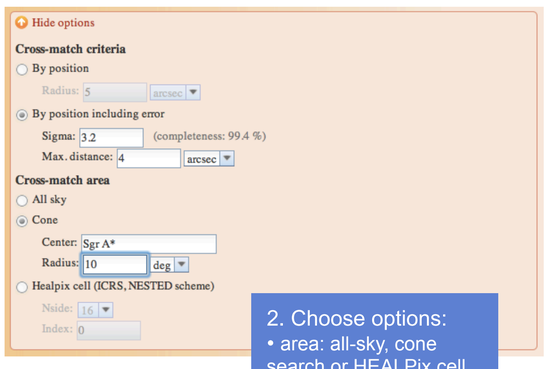



The service relies on the **UWS pattern**, an **IVOA standard** to submit and manage **asynchronous jobs**.
Users **interact** with the service through their **Web browser**.
Submitted **jobs** are **dispatched** to one of the available computation workers.
Cross-match **results** are **uploaded** to the **iRods** storage space, for later retrieval by the users.

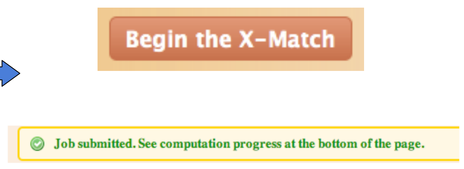
## 3. Web Interface




1. Choose tables to cross-match among:
- 9,000+ VizieR tables
- SIMBAD
- Uploaded tables

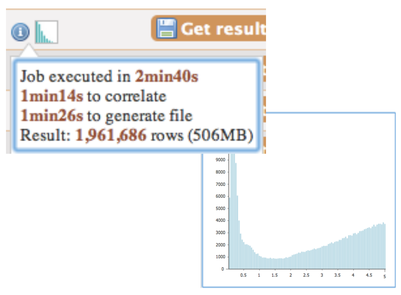



2. Choose options:
- area: all-sky, cone search or HEALPix cell
- mode : position only or position + errors




3. Submit job


4. Monitor progress




5. View results summary and distance histogram


6. Download result
- Distance (or sigma)
+ fields of table 1
+ fields of table 2
- Format: CSV or VOTable

## 4. Performances

| Table 1 | Table 2 | Computation time | Result generation time | Result size | Total time |
|---|---|---|---|---|---|
| **SDSS** *357M rows* | **2MASS** *470M rows* | 7 min | 11 min | **13 GB** | **18 min** |
| **DENIS** *355M* | **2MASS** *470M* | 11 min | 51 min | **58 GB** | **62 min** |
| **GLIMPSE** *104M* | **NOMAD** *1.1 billion* | 6 min | 17 min | **19 GB** | **23 min** |
| **SIMBAD** *5M* | **USNOB1** *1 billion* | 3 min | 1 min | **1 GB** | **4 min** |
| **SIMBAD** *5M* | **PPMX** *18M* | 20 seconds | 20 seconds | **440 MB** | **40 seconds** |

Summary of **cross-match performances at 5 arcsec**, running on a bi-quadcore (see details in panel 5)

## 5. Hardware

- **Fast RAM**
  - Worker 1 : 24 GB@**1333MHz**
  - Worker 2 : 32 GB@1333MHz

- **Fast HDD**
  - 6TB RAID5 array with **15,000 rpm disks**
    - Read speed : **570 MB/sec**
    - Write speed : 130 MB/sec

- **Multiple core CPUs**
  - Worker 1 : 2x Quad Core 2.27 GHz
    ⟶ **16 simultaneous threads**
  - Worker 2 : 2x Six Core 2.27 GHz
    ⟶ 24 simultaneous threads

http://cdsxmatch.u-strasbg.fr/

Come and see us at booth D2 for more info