The CDS Information System

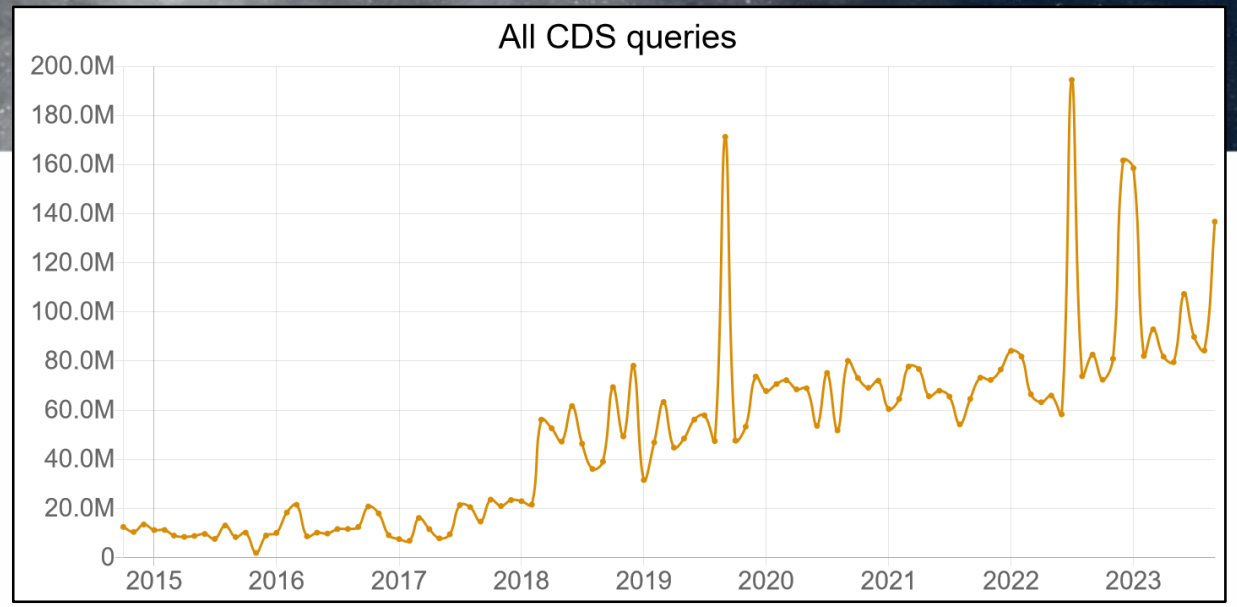
Overview & stats 2023

CDS Council – November 2023

Pierre Fernique
on behalf of all the CDS staff

Global usage 2023

- **3.0 million requests** per day
- **>390K unique IPs** per month

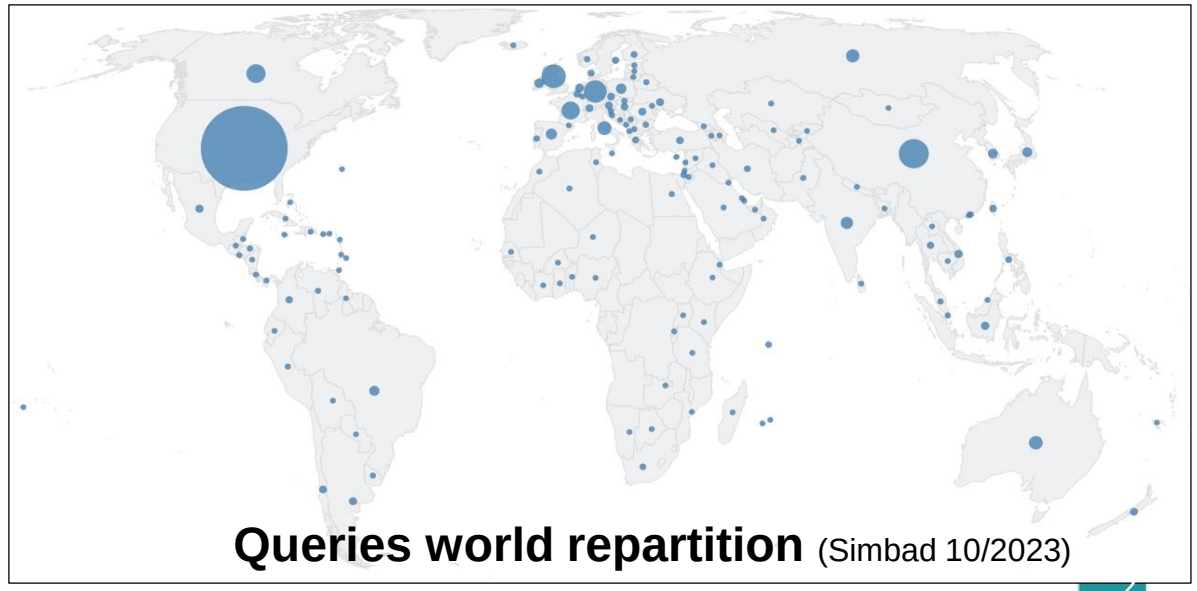


CDS main metrics *(Oct 2022 - Sep 2023)*

compared to Oct 2021 - Sep 2022

Main services	**Simbad** *Astro.object DB*	**Vizier** *Catalog service*	**Aladin** *Image service*	**Total**
users/month	190.3K +23% Adv.users: 1088	61.1K +35%	390.2K +8%	**> 390.2K**
queries/day	448.9K +30%	727.1K +100%	1.8M +5%	**3.0M**
load/day	6.1GB -13%	no data	263.3GB +62%	**> 269.4GB**
data volume	44.2GB +28%	77.0TB -3%	713.1TB +39%	**790.1TB**
data content	16.9M obj +17%	24.3K cats +6%	1203 HIPS +16%	
reliability	99.82% +0.08%	97.94% -1.2%	99.84% -0.09%	**> 97.94%**

(Bots have been removed of these stats)
users/month Based on unique IP (Adv.Users = only based on "advanced scientifical queries") on the main site
queries/day Based on effective DB queries (object queries, catalog queries, HiPS tile queries...)
load/day Downloaded from CDS servers
data volume Volume of service (data + index)
data content Number of service "items"
reliability Percentage of service availability (mirrors included)

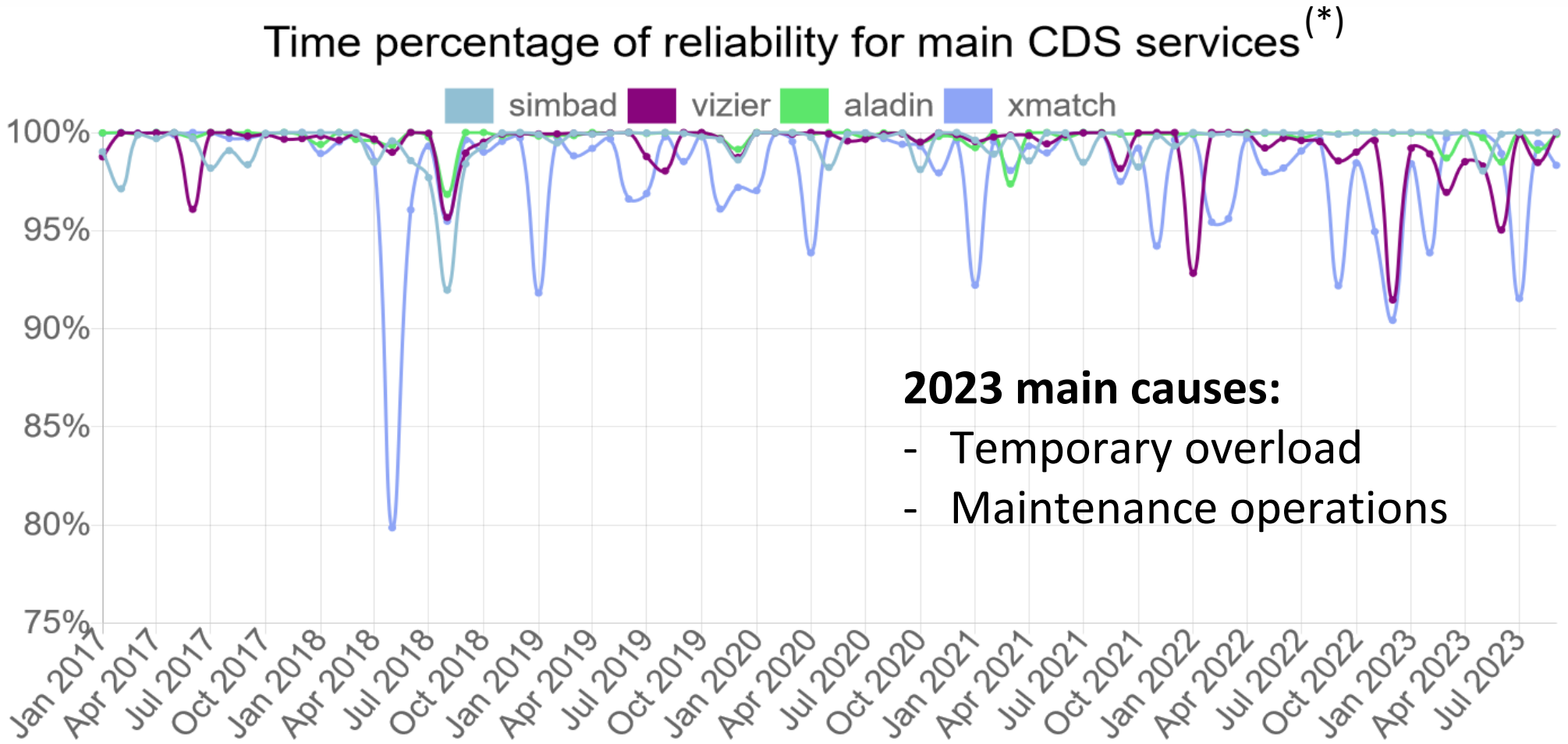


Queries world repartition (Simbad 10/2023)

Constraints and consequences

- **Our goal** : **permanent availability** 24 hours a day, 7 days a week, in a classic academic context (technical staff without on-call duty)
- **Our architecture**: **total duplication : services + data**, on several distinct geographical sites (still in progress)
- **It guarantees**:
  - the **performance** and **continuity** of services
  - the assurance of a **rapid resumption** of service in the event of a major accident at one of the sites (PRA)

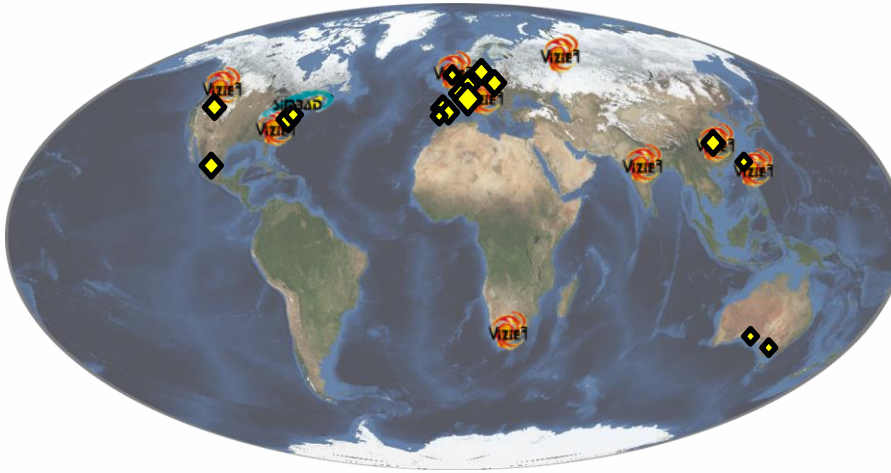
CDS S.I. reliability



(*) Based on tests every 9 minutes. Only partial inclusion of service redundancy (~30-minute delay).

CDS service architecture

- 2 **geographically distinct local machine rooms**
- **External mirror sites** (in France and in other countries)
- Server **virtualisation**, via a cluster of VMWare hypervisors, and Docker encapsulations
- Data storage equipment: **2 RAID6 bays synchronised on the 2 local sites** ("CDS Allsky data system")
- Archive storage equipment : **800To cheap disks**



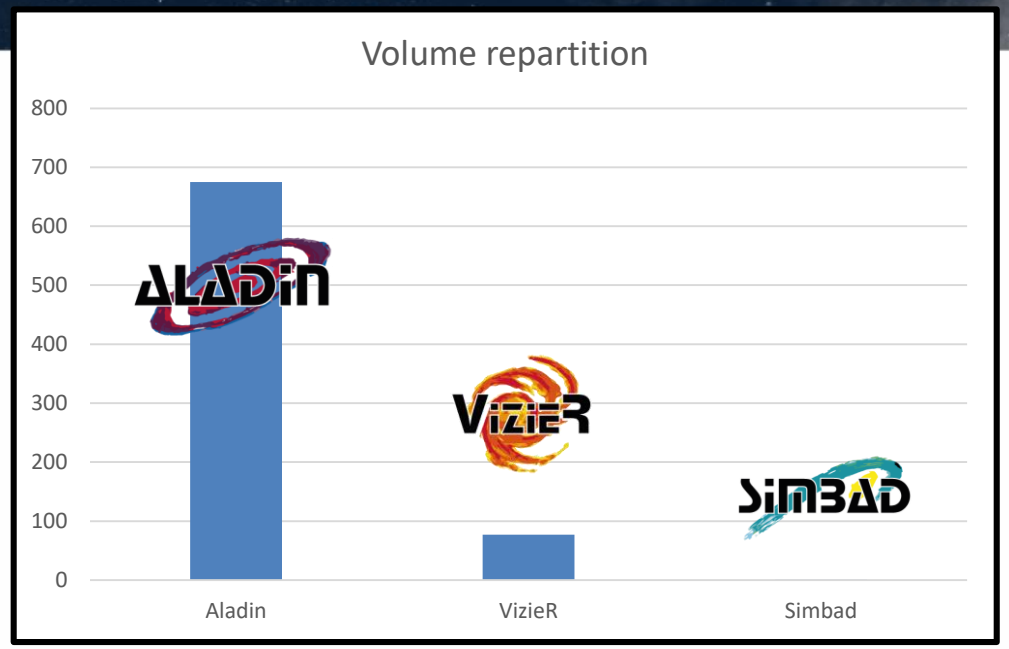
- **2 local computer rooms**
  => Unistra DC
  => Replicated elements in IPHC DC
- **7 external sites** (6 VizieR + 1 Simbad mirrors)
- **20 partner HiPS sites** (Aladin tiles)

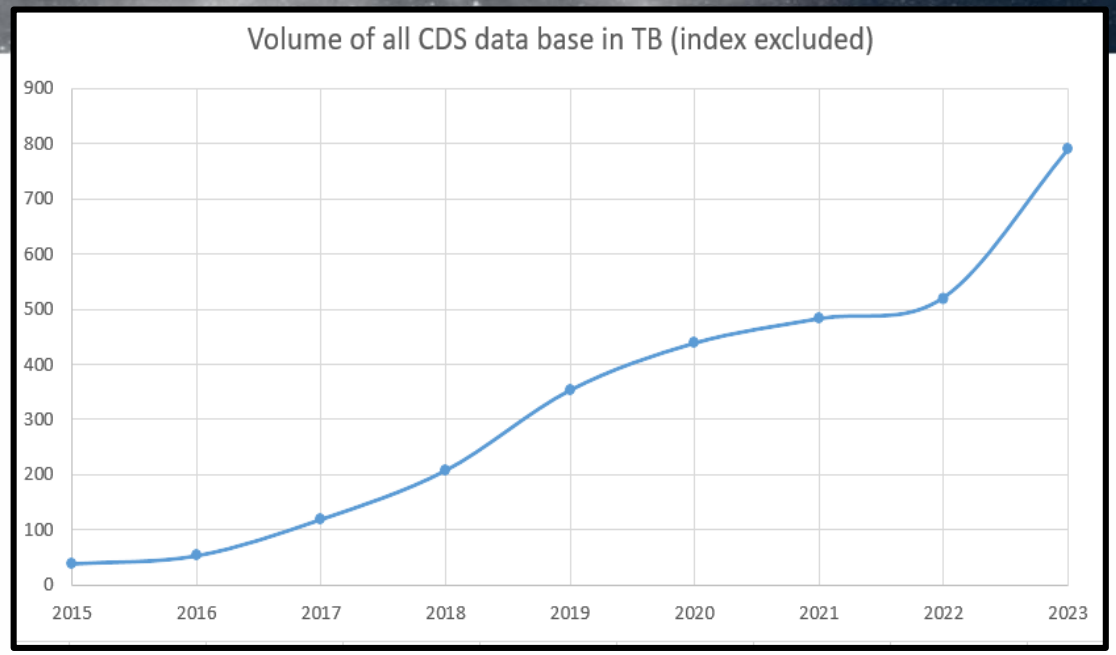
Local machine rooms evolution

- **Retirement** of CDS installations in the **local server room on the ObAS** site
  - Strong encouragement from the University to close this room (corresponding to University and national level policies of rationalization/reduction of small server rooms)
  - Technical limitations of the server room
- **Our solution: Partnership with the IPHC** (UMR7178) on the CNRS Cronenbourg campus.
  => Technical and functional conditions well adapted
- **University Data Centre will become the main local site** for our installations.

Relocation achieved during 2023 spring

Operational storage

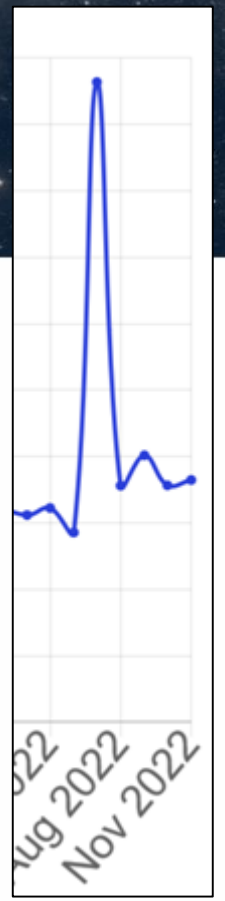




- Sept 2023: **790 TB** : 90% for Aladin, 9% VizieR
- Evolution relies to the volume of astronomical data published (bibliographic data, tables, catalogues and image surveys)
- Redundancy/wasteSpace factor : x5 (RAID6, mirror, snapshots, preparation, backup…)

- *Disk storage provisional plan : ~5 PB in 2025, ~9 PB in 2029 (including redundancy)*

Servers



- **CDS services** require **low CPU power** (except Xmatch service)
  But **a mistake to undersize the servers** because the availability constraint implies the capacity to absorb "peaks“
  => The technical solution: **virtualise the servers** as far as possible on VMWARE hypervisors (presently 5)

- **Computing machines** for data preparation
  => **3 big machines** are sufficient for the task (Aladin HiPS, VizieR big catalogs)

Software development strategy

- **Open source software** supported by a dynamic community (e.g. Postgres, astropy, etc.)
- **''In-house'' developments** for libraries and tools particularly specific to our activity (e.g. management of spatial indexes, astronomical libraries, bibliography processing, etc.)
  => Under Open licence (French minister recommendation)
- Use of paid software remains exceptional, justified by the absence of an equivalent free solution (e.g. VMWare). These paying solutions are considered as transitory

- At least **two developers per service**.

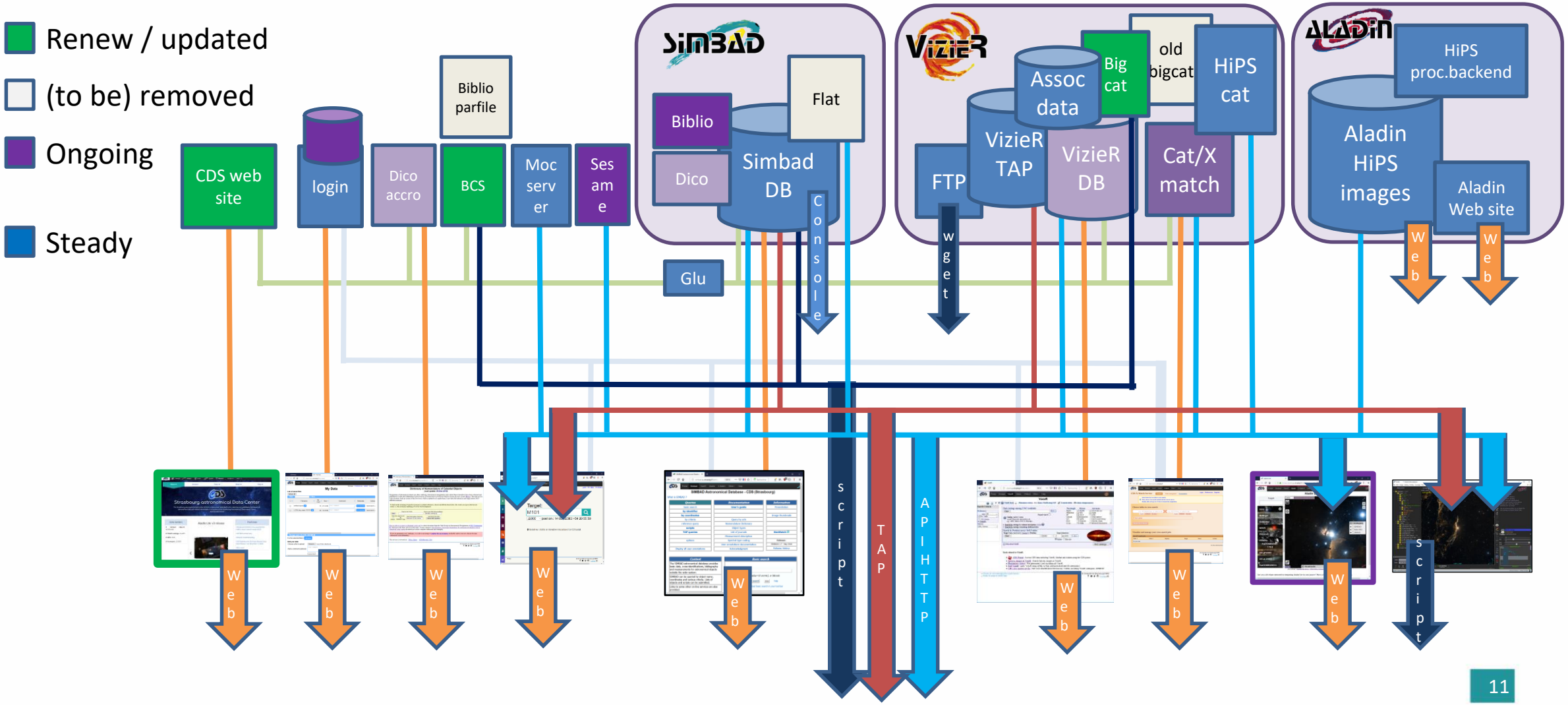
=> By the end of 2022, this strategy is assured partially with contractors

Information system

- Nov 2023: **CDS Information service** is based on **24 components**
- The permanent **evolution** of these components **is essential** to map:
  - The current needs (astronomy and data evolution)
  - The CDS human resources

- For each component, we evaluate its sustainability and we plan its renewal, evolution or withdrawal

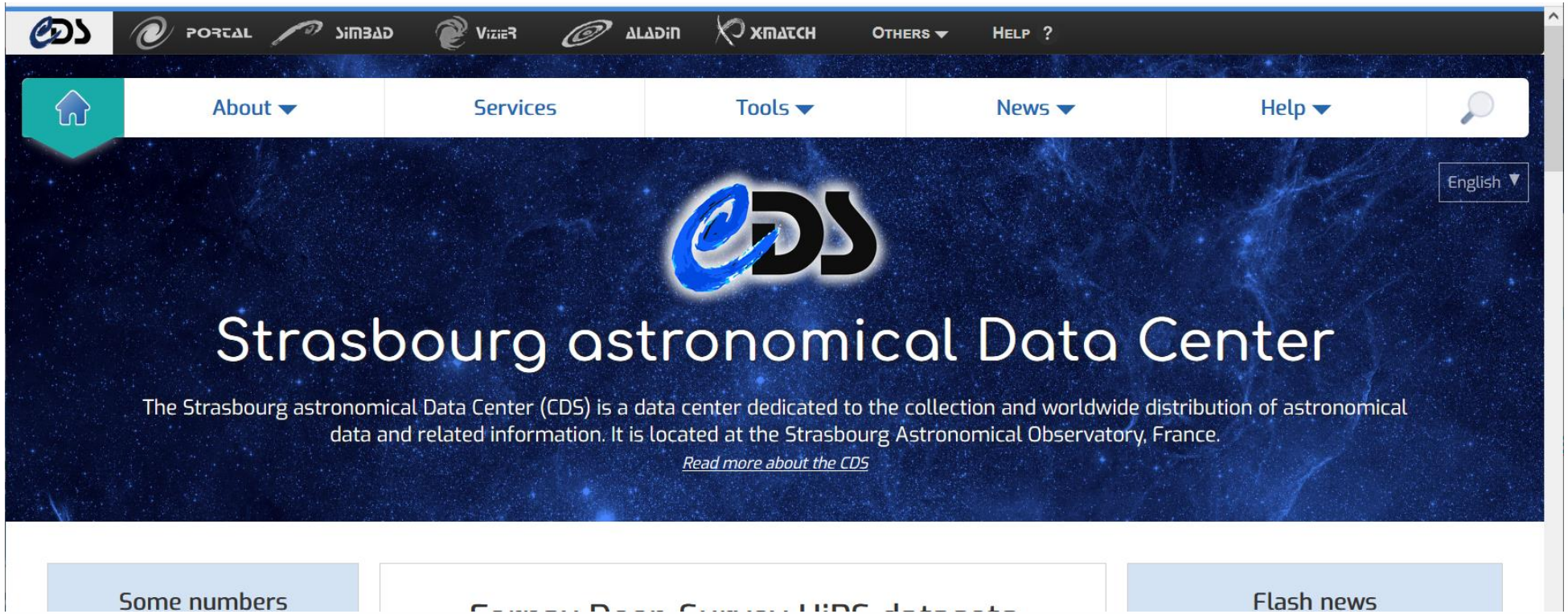
Evolution detail of IS components



Recent developments in the IS

- **BCS** - the deployment of the **new journal article management system**:
  => The "BCS" is replacing the old system (parfile based)
- **MOCset** - the evolution of the **global positional index of VizieR** for a MOC compatible solution (replacing Qbox).
- **QATSS** – new tool/lib for big catalog manipulations
- **Web site** - full redesign of **the CDS website** (deployement phase).

Questions ?


PORTAL
SIMBAD
VIZIER
ALADIN
XMATCH
OTHERS
HELP ?
About
Services
Tools
News
Help
English
Strasbourg astronomical Data Center
The Strasbourg astronomical Data Center (CDS) is a data center dedicated to the collection and worldwide distribution of astronomical data and related information. It is located at the Strasbourg Astronomical Observatory, France.
Read more about the CDS
Some numbers
Flash news