CDS Plans and Challenges 2017-2018

Mark Allen - Director CDS

Plans 2017-2018

- **Continuation of core work of building CDS content and operating services**
- Special attention to service architectures and internal tools
- Internal training and maintenance of processes
- **Major release:** Aladin v10
- **Users:** Feedback on CDS Portal, advertise VizieR associated data
- **Interactions with Journals:** A&A, AAS, MNRAS
- **VO:** Leadership roles, plans for Euro-VO
- ***Certification:*** DSA, ***Publication:*** CDS reference papers
- **Projects** — ASTERICS, RDA, AENEAS, Europlanet
- **R&D and new capabilities:** Python, Containers, user interfaces
- **Events:** IAU GA, ADASS, EWASS, AAS, SF2A
- **Facing the challenges:** Staffing and Infrastructure

Service architectures and internal tools

- SIMBAD architecture renewal
- DJIN2 tool for processing journal articles
- Dictionary of Nomenclature database
- Review of tools for process tracking (use of redmine and twiki)
- X-Match disks to be replaced (hopefully SSD)

Training and Team-work

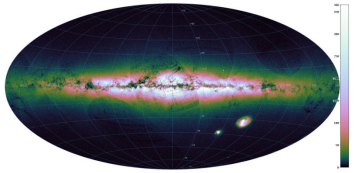
- **Documentalists** — define program for new arrivals
  - Astronomy
  - Documentation
  - Technology: unix/linux + python + CDS tools

- **Software Engineers / Scientists** : Explore needs for training on user interface design and/or seek external expertise
- Capitalise on very strong **R&D training program**, make it more visible
- Use of **social networks**… Science, technical, societal aspects

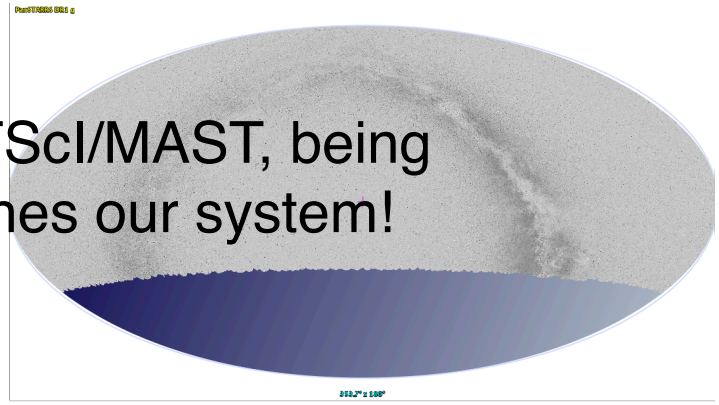
Advertise major releases and new capabilities

- **Aladin version 10** - release planned for Oct 2017
- Highlight new capabilities to scientists and also ADASS/VO developer community and data providers
- **CDS Portal** - user feedback phase, supported by CDS Tutorial, video mini-tutorials
  - Will influence plans for CDS service interfaces, and also discussions of generic VO-Portals
- **X-Match support for projects** — e.g. Gaia, CFIS, + , with attention to requirements for future CDS login and user-space
- **Video tutorials** — more to be made, and released
- Advertisement of the **VizieR associated data** service to authors

Important data sets in 2017-2018



- **Gaia DR2** — April 2018
  - Following the 'preliminary' release DR1 in Sept 2016, this will be the first real release of a billion stars…
  - CDS is ready based on excellent preparation done for DR1
  - Significant follow-on task of ingesting into SIMBAD



- **Pan-STARRS DR1** in coordination with STScI/MAST, being processed now. HiPS images— really pushes our system! DR1 catalogue — in VizieR very soon
- **SkyMAPPER** — HiPS images computed in preparation for DR1.1 (Global release) ~Nov 2017. Higher quality imaging expected in 2018.

- **Radio Astronomy**
  - Support of ingestion and use of Radio data in CDS services
  - Specfind — Radio catalogues in VizieR
  - Participation in AENEAS project: design study of SKA data centre
  - CDS-RADIO-POSTDOC — arriving February 2018

- **Time Domain Astronomy**
  - Eventual goal — time metadata to enable query of CDS service by time
  - Interoperability of Time Domain data — via IVOA, and to address future very large transient object surveys

Interactions with Journals

- **A&A, AAS, MNRAS** - support of formal arrangements
- **Digital Object Identifiers** - follow through on agreements for use of DOIs with ADS, and Journals
- Improvement and advertisement of the instructions for authors
- Pursue links with events run by Journals/Publishers - writing and publishing workshops etc.
- **IAU General Assembly** — communicate with Journals and Publishers

Virtual Observatory

- **Leadership roles at IVOA**
  - Exec, DAL*, Apps*, Semantics, Time Domain IG, DCP*
- Approach of **engagement with big projects** working, but needs to be maintained
- ***Interoperability -*** *a keyword everywhere*, but approaches/ support to VO in different countries varies. IVOA continues to have high level of participation (e.g. >150 attending in Santiago).
- Timely to define future of **Euro-VO** in particular for next ASTRONET Roadmap

- CDS / ASTERICS / Euro-VO participation in **United Nations UNOOSA "Open Universe"** meetings with other IVOA partners, and the astronomy data providers
  … *but with some significant reservations*

**until May 2018*

Projects

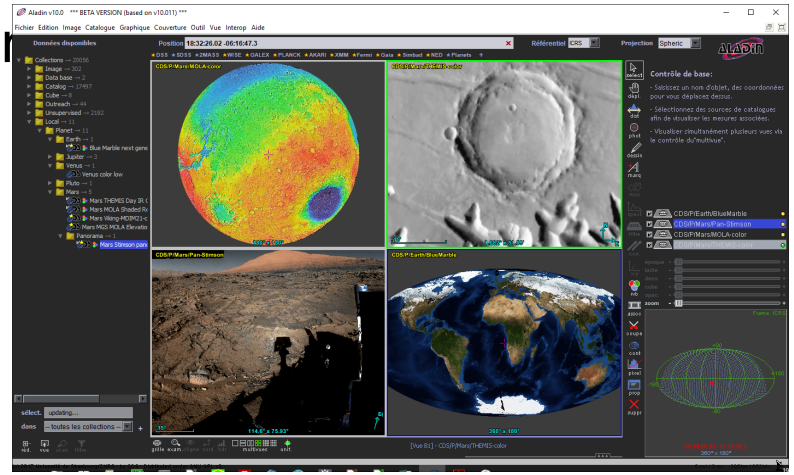
ASTERICS H2020 project, May 2015 - May 2019

In 2017-2018:

- 3rd DADI VO School, Madrid, November 2017
- DADI ESFRI Forum & Training Event, December 2017
- ASTERICS Policy Forum, January 2018
- 4th DADI Technology Forum, 2018
- CDS-ASTERICS Postdoc - advertised, start ~early 2018
- Ref. implementations of IVOA standards, Time Domain priority
- Planning for 4th DADI VO School, Strasbourg, late 2018
- Cooperation with partners to define the next project!

Projects

- **RDA** (2018 - 2020) — Important role for global data sharing
- **AENEAS** (2016-2019)
  - Minor partner, VO aspects, related to SKA initiative in France
- **Europlanet/VESPA** (2015-2019)
  - Virtual European Solar and Planetar
  - Planetary coordinate systems
  - EPN-TAP, HiPS



Staffing

- **Scientific Support of CDS services**
  - Critical need for scientific support of SIMBAD on 2019-2020 timescale
  - VizieR associated data service needs support to operate at its potential
  - Voluntary external effort being sought for “project mode” support for specific topics
  - Need to accommodate the absences and shifts expected
  - Contributions are being made by CDS Postdocs

Staffing

- **Documentalists**
  - A specific profile, with long training times
  - Contractors necessary to manage the current volume (2 + 1 arriving in Nov 2017)
  - Documentalist position is sought as first priority of ObAS
- **Software Engineers**
  - SIMBAD and VizieR - requires duplication of effort
  - Expert contract engineer arriving May 2018 (G. Mantelet)
  - Long term solutions needed

Big Data

CDS in the era of Big Data *…developing our approach*

- **Continue our role for reference data and reference services**
- Interoperability with very large data sets is key
- Reference data sets - volume issues not as critical as for the processing data centres. e.g. LSST 37 billion source “classic” catalogue feasible on the ~2025 timescale for public release. Annual image data releases also feasible.
- CDS does not seek to be a Euclid / LSST / SKA … archive
- At present we do host the largest reference data sets — we aim to do this in the immediate future
- Longer term, all the data may not be held at CDS, but it can operate as if it were. We have the metadata and the interoperability **— future agreements are critical**

Proprietary data projects

- We expect demand for use of CDS systems in support of projects even during their proprietary phase (Euclid?, CFIS?, …LSST??)
- **Dilemma** — CDS services are funded for open access
  - A ‘project’ model could be investigated
  - CDS X-Match an easy first step, projects can up-load their catalogues and use our X-Match compute power
  - Public vs. proprietary VizieR catalogues?
    - No… different models need to be considered
  - Would require building on CDS login with stronger Authorization & Authentication
- Internal discussions needed…

Infrastructure

- **CDS is undergoing a change in scale due to Big Data**
  - Currently ~1 Petabyte to store ~300 TB
  - The fastest growing collection is the HiPS image surveys
  - A study has been done to estimate needs for next ~5 yrs
  - The result is 2 Petabytes (1 PB replicated)
  - System configurations have been considered and quoted (DELL) with estimate ~ 350 k€ for system over 2 sites
- CDS responded to the MENESR call for *"Expression of needs"* [PIA-3] in 2017, with proposal ***"CDS - All Sky Data"***
- Unclear whether this will work… various reasons…
- Looking for other solutions
- Clear path forward needed in early 2018

Infrastructure

- The local landscape is changing:
  - Data Centre being constructed by Unistra (~2019?)
  - *…no real information on the cost of use, constraints, organisation of this centre*
  - 2 racks have been pre-reserved, for ‘back-up’ systems as a first step, and as a solution to ad-hoc current back-up system
- CDS has strong operational record, and control and flexibility are essential to our operations — Simplifications of mutualisation of server infrastructures can be dangerous
- Current approach — insist on local Master system, and use our new expertise, and long experience to guide developments