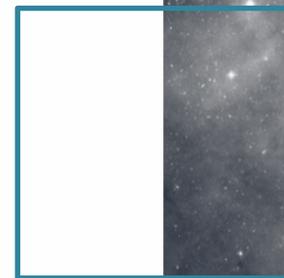


La mission et l'histoire du CDS

Françoise Genova et l'équipe du CDS



1972 – le matériel informatique

- Un ordinateur unique de l'INAG à Meudon : IBM 360/65
 - 512Ko mémoire rapide, 2Mo de mémoire lente
 - Disques amovibles
 - Bandes magnétiques



1972-1981 – L'équipement local

Perforatrice de cartes



Console ASR 33



Requêtes:

- Par la poste
- Par réseau avec un paquet de cartes

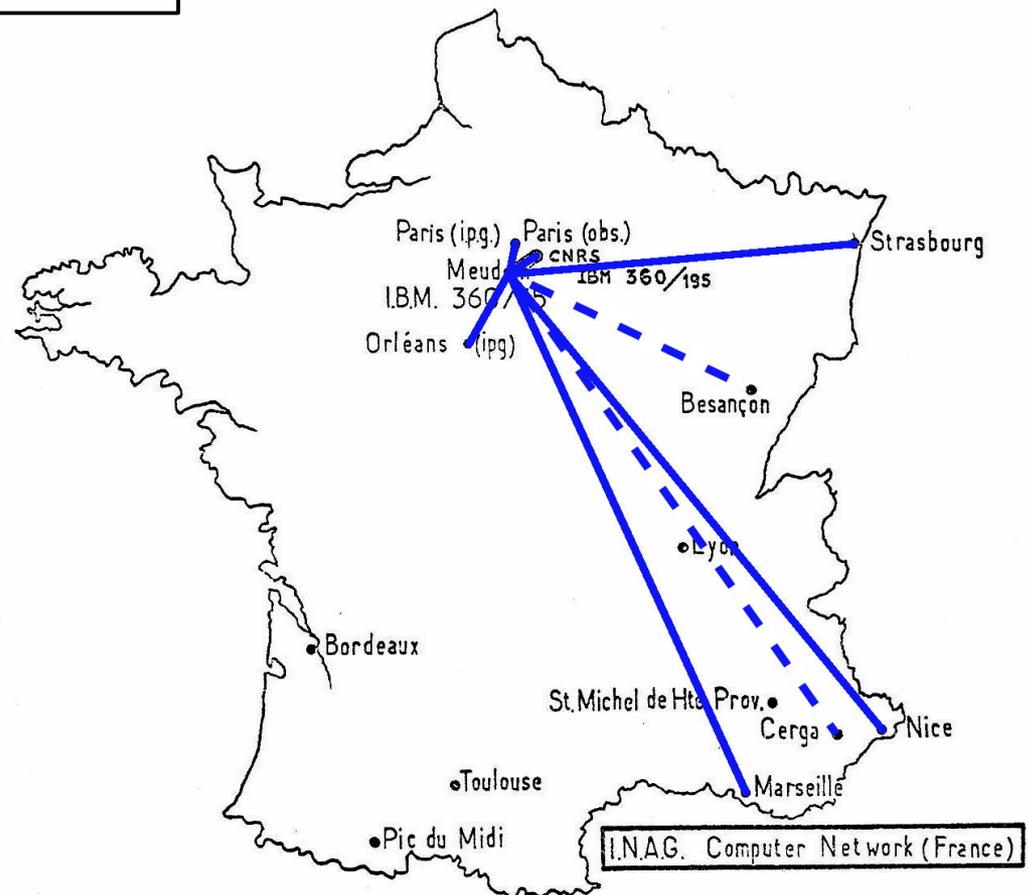
Inspiré du séminaire de Marc Wenger, juin 2014

1974 – le réseau

État du réseau INAG en 1974

Centre de calcul à Meudon

Observatoire de Paris
Institut d'Astrophysique de Paris
I.P.G de Paris
I.P.G. d'Orléans
Observatoire de Strasbourg
Observatoire de Besançon
Observatoire de Nice
Observatoire de Marseille
Cerga



□ La mission du CDS

- Le CDS a été créé par l'INAG (devenu l'INSU du CNRS) avec l'Université Louis Pasteur (devenue l'Université de Strasbourg)
- Depuis 1972
 - Collecter les données utiles sur les objets astronomiques, sous forme électronique
 - Les améliorer en les évaluant de façon critique et en les combinant
 - Distribuer les résultats à la communauté internationale
 - Conduire des recherches en utilisant les données
- Les mots-clé
 - Données sous forme électronique
 - Curation et qualité des données
 - Rôle international
 - Objectif: Servir la recherche

□ Le CDS est et a été pionnier

- Un des pionniers absolus des centres de données scientifiques, toute disciplines confondues
 - ICPSR, Sciences Politiques et Sociales, 1962
- Pionnier du partage des données numériques en astronomie avec la base de données du satellite IUE (1985)
- L'astronomie est elle-même une discipline pionnière de la Science Ouverte

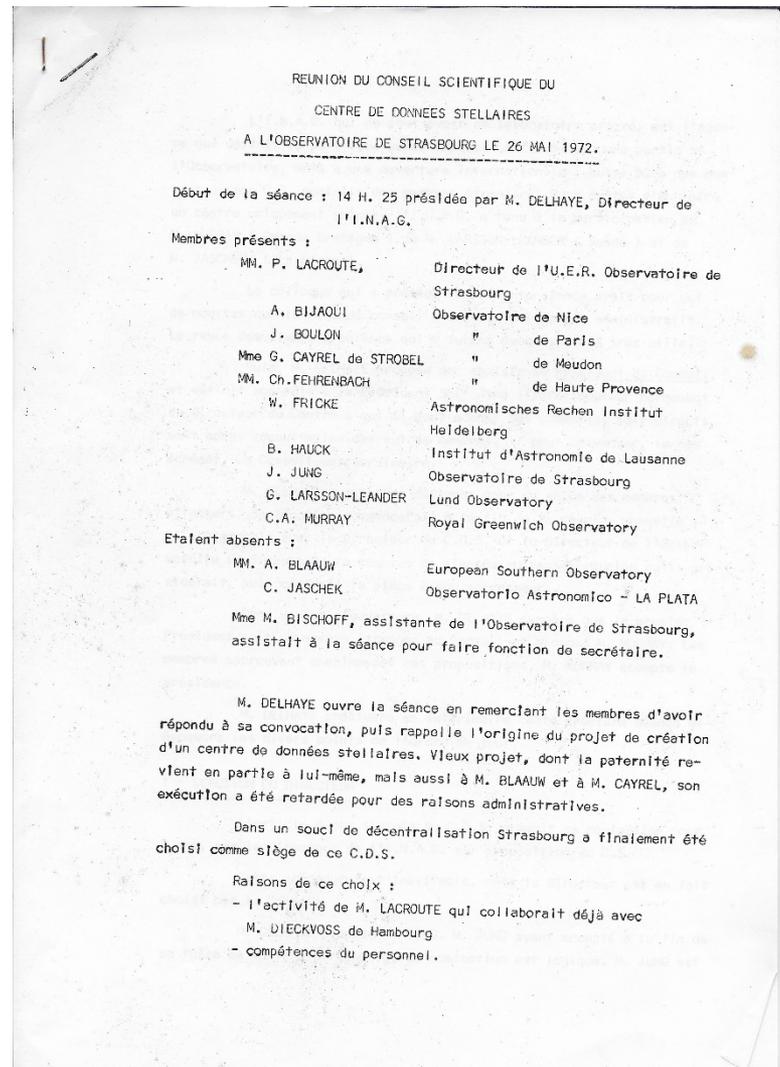
□ Les impératifs catégoriques

- Servir la recherche
 - Prendre en compte les besoins des scientifiques
 - Au meilleur niveau technologique
- Définir et mettre à jour la stratégie - le contexte évolue constamment
 - L'astronomie évolue – progrès scientifique, nouveaux instruments
 - La technologie évolue – la révolution du web! mais en fait évolution continue
 - Le contexte 'politique' évolue – la science ouverte, FAIR, l'EOSC

□ Le CDS est une infrastructure de recherche

- Une gouvernance identifiée, effective
- Une instance de pilotage stratégique et scientifique, le Conseil Scientifique
 - International depuis l'origine
 - Représentants de l'ESA, de l'ESO et de la NASA
- Un des piliers de la politique et de la pratique de la science ouverte internationale de l'astronomie
- Sur la Feuille de Route nationale depuis sa première version (2008)

Le premier Conseil Scientifique du CDS le 26 mai 1972



□ Ecrit en 2000

[10.1051/aas:2000333](https://doi.org/10.1051/aas:2000333)

The CDS information hub

On-line services and links at the Centre de Données astronomiques de Strasbourg

F. Genova, D. Egret, O. Bienaymé, F. Bonnarel, P. Dubois, P. Fernique, G. Jasniewicz*, S. Lesteven, R. Monier, F. Ochsenbein, and M. Wenger

CDS, Observatoire astronomique de Strasbourg, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France
e-mail: question@simbad.u-strasbg.fr

Received December 20, 1999; accepted January 17, 2000

- In practice, **the main challenge in the CDS activity is to constantly tune the contents and the services to the rapid scientific and technical evolution, to be able to deal with ever increasing volumes of information, and to be ready to respond to the new projects and to the evolution of the policy of the national and international Agencies.**
- The **guidelines in prioritizing the tasks** are several: offer the best service to the users, ensure the quality of contents, and make the best of technological innovation. This implies that new developments are begun neither too early, to rely on techniques which are as secure as possible, nor too late, to offer the best possible services and improve the functionalities - hence the importance of technical watch.
- A **balance has constantly to be kept**, between very long term activities, on time scales of several ten years, to build up the contents; the development and maintenance of database systems and user interfaces, on time scales of a few months to a few years; R&T activities, on similar times scales; and operational constraints on a day-to-day basis. Hence the importance of careful strategy definition and activity scheduling.

□ Continuer et progresser sur la (longue) durée

- Définir une stratégie explicite et évolutive
- Réunir et conserver les compétences et les faire évoluer
- Construire et tenir sa place dans le concert des nations
 - Fourniture de services essentiels
 - Réseau de collaboration avec les acteurs majeurs (agences, journaux, ADS)
 - Participation forte à la définition des standards
- Obtenir le soutien sur le long terme de la gouvernance

Le CDS dans l'infrastructure globale des données en astronomie

- Les données en astronomie
 - Sont FAIR - format FITS (**R**, 1981) et standards de l'Observatoire Virtuel (OV) Astronomique (**F, A, I**) définis par l'IVOA depuis 2002
 - Vu de l'utilisateur: une seule infrastructure de données mondiale, multipolaire
- Le rôle du CDS
 - Fourniture de services essentiels inclus dans l'OV
 - Plus de 2 millions de requêtes/jour
 - Augmente la visibilité et l'utilisabilité des services
 - Participation active à la définition des standards
 - 48 standards IVOA, 24 avec auteur(s) ObAS (dont éditeurs: 13)
 - Le CDS avait développé plusieurs précurseurs de l'OV
 - Pilotage du 1^{er} Groupe de Travail international *Interoperability*, pré-IVOA, dans le cadre du Réseau OPTICON (2000-2002) ⇒ 1^{er} standard de l'OV en 2002
 - En Europe, coordination de projets ou de WP
 - Soutien à l'OV dans des projets financés par la CE
 - L'OV est inclus dans les projets Cluster qui regroupent les ESFRI
 - Interfaçage de l'OV avec l'EOSC
 - En France
 - Coordination de l'Action Spécifique OV France de 2004 à 2020
 - Participation active à la diffusion des connaissances et au partage d'expérience

□ SIMBAD

- La base de données de référence pour la nomenclature et la bibliographie des objets astronomiques hors système solaire
- Bibliographical Star Index + Catalogue of Stellar Identifications ⇒ SIMBAD (1981)
- 30/03/22: 13 189 583 objets, 52 228 260 noms d'objets, 401 726 références bibliographiques, 29 354 128 citations, 15 093 acronymes
- La nomenclature
 - Un objet astronomique peut avoir de nombreux noms (Mrk 421 en a 138 dans SIMBAD)
 - Employé comme 'résolveur de nom' par les archives et ADS - décidé en 1992
- Les publications
 - Un identifiant partagé en 1990, bien avant le web et l'accord sur le DOI
 - Un objet astronomique peut être cité dans de nombreuses publications (15 855 pour le LMC)
 - Construction semi-automatique du contenu
 - Pour faire face à l'augmentation du volume à traiter
 - Tri des références par pertinence (dans quelles parties du texte l'objet est cité)

□ VizieR

- La base de données de référence pour les grands catalogues et les tables publiées dans les articles
 - Collection de catalogues depuis 1972
 - Les tables des articles d'*Astronomy & Astrophysics* en ligne au CDS en 1993. Les chiffres imprimés deviennent des données réutilisables!
 - VizieR a démarré en 1996 – d'Hipparcos à Gaia!
- 31/03/22: 22 144 « catalogues » dans VizieR
- Description détaillée du contenu
 - Une vision homogène de données hétérogènes
 - Pipeline d'entrée dans la base à partir des descriptions
 - En 1998, 100 000 colonnes de tables dans VizieR
 - ⇒ le standard Unified Content Descriptors
 - ⇒ standard IVOA en 2005

□ Aladin

- Atlas interactif du ciel
 - Images stockées au CDS (400 TB) et dans les archives des observatoires
 - Catalogues stockés au CDS et ailleurs
- Interopérabilité avec les archives d'observatoires dès 1998
- Interopérabilité avec les autres outils du VO: standard SAMP, 2009
- Portail de l'Observatoire Virtuel astronomique depuis 2002
- Une application à installer (Aladin Desktop, 1999) ou Aladin Light (2013), un élément à inclure dans une page web
- Les standards HiPS et MOC permettent de gérer les très grands volumes de données et d'y accéder de façon efficace (zoom)

□ Cross-identification

- Identification croisée des très grands catalogues
- A démarré en 2011
- Très efficace pour la cross-identification des très grands catalogues mais aussi utilisé pour des catalogues beaucoup plus petits
- Interfacé avec TopCat

□ D'autres hauts faits du CDS

- La première connexion internet en France, le 28 juillet 1988, pour permettre aux astronomes américains d'accéder à SIMBAD à l'occasion d'une réunion de l'Américain Astronomical Society à Baltimore
- La mise en réseau précoce des ressources bibliographiques en ligne de l'astronomie
- L'interopérabilité, très tôt

From “Et Dieu Créa l’Internet”
Christian Huitema (1995)



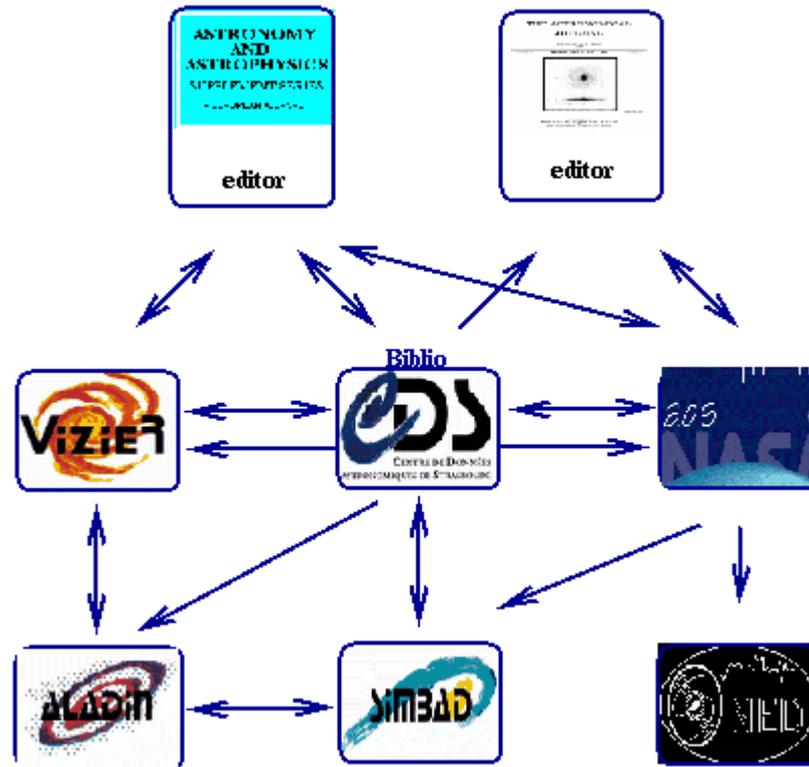
ensuite de commander la liaison à France Télécom et à la société américaine de communications MCI : le 28 juillet 1988, la France était raccordée à l’Internet.

Néanmoins, le 8 août, tout était prêt. Nous fîmes une belle démonstration, nos partenaires de la NASA purent accéder à la base de données qui, à Strasbourg, récapitulait des siècles d’observations astronomiques.

40^{ème} anniversaire du CDS

<https://interstices.info/au-coeur-de-la-premiere-connexion-francaise-a-larpanet/>

Navigation entre données bibliographiques



Lesteven et al.

LISA III, 1998

[1998ASPC..153...61L](#)

VO standards and tools

- Interoperability in 1998
 - CDS GLU : registry of distributed resources which keeps track of the web addresses and allows automated modification in all web pages by changing the registry record
 - Aladin: interactive sky atlas giving access to distributed image archives
 - Unified Content Descriptors: describing quantities, derived from 100 000 Vizier columns, ESO/CDS Data Mining Project, will be one of the first VO standards
- 2001 IDHA project: precursor of Characterization Data Model

40^{ème} anniversaire du CDS

□ R&D

- Le Centre de Données doit travailler sur la longue durée avec des contraintes opérationnelles fortes
- La technologie peut évoluer très vite dans le domaine
 - Des buzz sans lendemain
- Des solutions ‘suffisamment pérennes’ pour améliorer la manière de répondre aux besoins et aux attentes des utilisateurs
- Veille technologique et R&D à visées opérationnelles
 - Nombreux stages pour tester des solutions possibles
 - Décision explicite d’implémentation après évaluation

□ Le CDS comme centre d'expertise

- Pour l'astronomie
 - « Comment partager des données en astronomie »
 - Connaissance des données et des besoins scientifiques et techniques des utilisateurs
 - Compétences techniques et sur la curation des données
 - Compétences pour définir les standards et pour les utiliser
- Pour la science ouverte
 - « Comment partager des données scientifiques »
 - Partage d'expérience avec les autres disciplines (Data Terra, Humanum, ...), l'INSU, le CNRS, l'UNISTRA, le MESRI
 - Participation à des groupes d'experts (données, FAIR, EOSC)
 - Certification des entrepôts de données

□ L'évolution du contexte

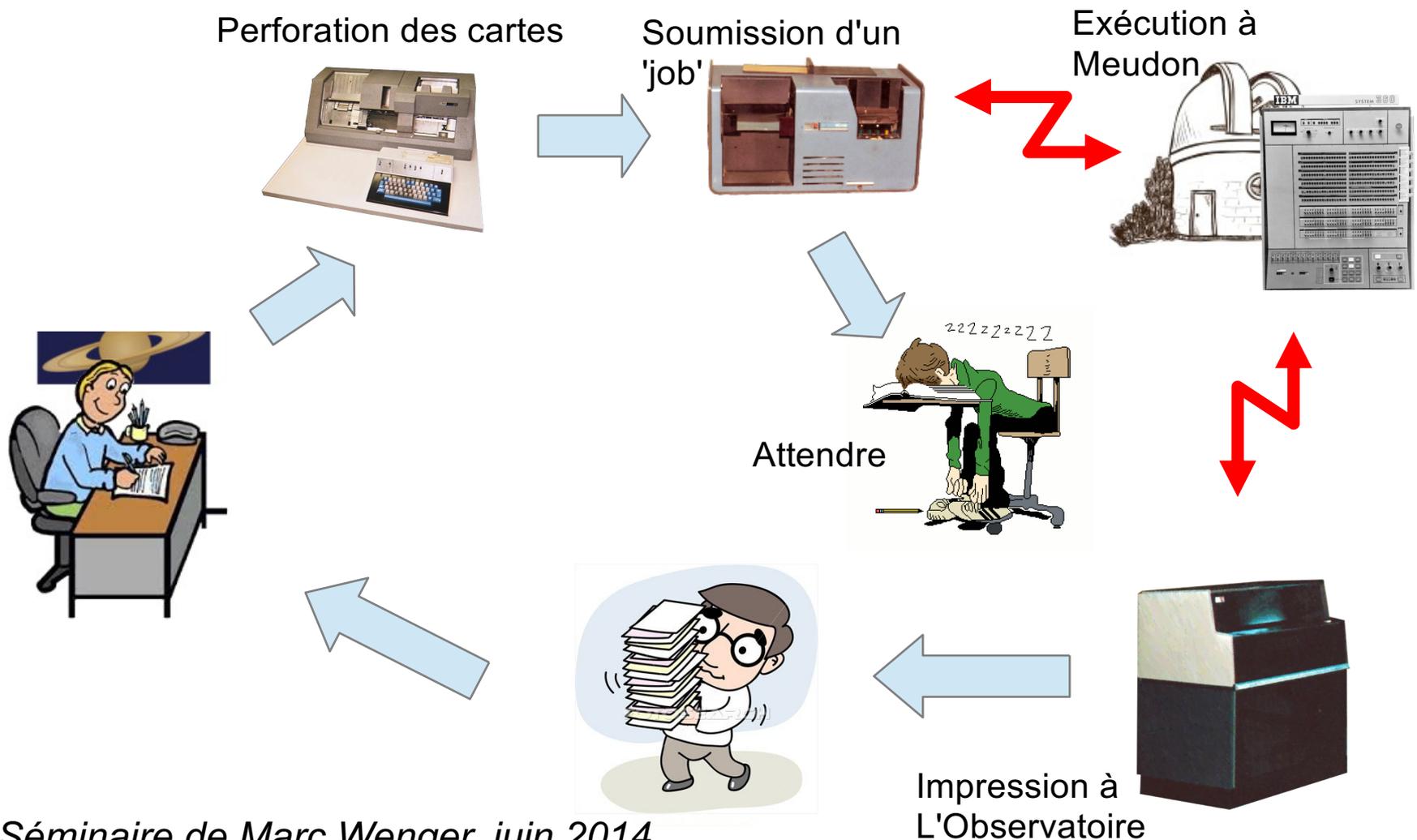
- Rôle international
 - Il reste la raison d'être du CDS: le soutien aux activités de recherche de la communauté astronomique mondiale
 - Contribution de haut niveau de la France à l'infrastructure mondiale de l'astronomie
- Au niveau national
 - Prospective INSU 1995: l'exploitation maximale des données est une priorité
 - En 2022: à la confluence du CoSO et du CoSIN
 - Pionnier de l'évolution vers la science ouverte
 - Contexte PNSO: Centre thématique, certification
 - *Recherche Data Gouv* se construit, tout comme la participation du CDS
 - Mutualisation – de l'IBM 360 de Meudon au datacentre local
- Au niveau européen
 - Participation à la construction de l'EOSC
 - Construction de la participation du CDS et de l'astronomie à l'EOSC
- Définition et partage des standards et des bonnes pratiques à tous ces niveaux, en astronomie et au-delà

□ Ce qui reste sur la durée

- Une charge de travail très lourde, qui continue à augmenter
- Les bases de la confiance des utilisateurs
 - La pertinence des services par rapport aux besoins des utilisateurs
 - Le souci constant de la qualité du contenu et des services
- La veille technologique et la R&D comme partie intégrante du travail du CDS
- Le réseau de collaboration internationale et le choix de construire sur la collaboration même s'il y a des risques (cf client-serveur, interopérabilité)
- Le soutien du CNRS/INSU et de l'Université
- Le risque que tout le monde pense que tout va bien au CDS alors qu'il relève constamment des défis complexes
- L'équipe CDS

1972 – Le développement des logiciels

- Interactions avec le Centre de Calcul uniquement en batch



1972 – CSI : procédure de mise à jour

