

1901—1960			1961—2020			2021—2080			2081—2140			2141—2200					
m	8u	+22°	m	8u	+22°	m	8u—9u	+22°	m	9u	+22°	m	9u—10u	+22°			
9-2	10	14.0	2.6	9-2	28	8.3	18.2	K	9-5	13	40.5	27.2	9-1	46	20.1	8.9	K
9-4		18.8	15.5	9-2		12.2	40.9	K	8-2		44.8	6.4	9-2		39.8	5.9	K
9-5		20.7	47.5	9-2		18.7	13.7	K	9-5		46.9	58.3	9-4		44.9	28.3	K
9-5		29.2	9.0	9-3		48.4	49.0	K	9-5		47.2	1.8	9-5	48	6.6	20.0	
9-5		51.1	55.4	8-6		54.3	29.0		8-9	14	8.5	8.6	9-4		39.3	11.3	
9-2		53.2	47.1	9-5	29	18.1	31.2		9-5		52.0	56.8	9-5		40.4	34.2	
9-5		56.4	1.3	9-5		32.9	44.5		9-2		11.5	59.6	8-2	49	40.4	34.2	
9-4	11	3.4	8.8	9-5		47.6	54.2		9-5		26.3	26.5	8-2	50	23.4	59.3	K
9-5		21.0	22.2	9-5	30	29.2	40.3		9-5		40.1	33.7	6-7	51	24.3	1.6	K
9-4		29.0	17.8	9-5		33.5	18.3		6-7		42.7	24.8	9-5		30.8	24.2	
									9-5	49	7.0	18.9	9-5		41.6	7.0	



# Les données de la recherche en astronomie

André Schaaff

Observatoire astronomique de Strasbourg & Centre de Données astronomiques de Strasbourg



# Avant-propos

Une présentation concernant les données de la recherche en astronomie sous différents angles avec une rapide introduction sur l'enjeu des données de la recherche en général et les projets internationaux transdisciplinaires récents.

Un focus sur le Centre de Données astronomiques de Strasbourg.



# Le CDS ?

# Le CDS

Le CDS (Centre de Données astronomiques de Strasbourg) existe depuis 1972, ses objectifs initiaux sont (toujours d'actualité):

- la collecte des données utiles sur les objets astronomiques
- l'enrichissement des données en les évaluant de façon critique et en les combinant
- la distribution des résultats à la communauté internationale
- la conduite de recherches utilisant les données

Pour ce faire, le CDS:

- dispose d'une équipe de documentalistes, d'astronomes et d'informaticiens (environ 30 personnes également réparties dans les 3 métiers)
- développe des outils et mène des actions de R&D
- participe à des projets dans son domaine de connaissances: CoRoT, Gaia, ..

Aujourd'hui, c'est environ 1,000,000 requêtes / jour sur ses services...

# Les principaux services du CDS



~8,000,000 Objets

Base de données d'objets astronomiques:  
identification, bibliographie, données, mesures...



~ 13,000 catalogues, 26,000 tables

Collection de données sous forme de catalogues:  
issus de journaux, de logs d'observation ou de  
grand relevés principalement constitués de  
données tabulaires (mais aussi spectres, images,  
séries temporelles)



50 To, 175 relevés au format HiPS

Atlas interactif du ciel: découverte, visualisation et  
manipulation de données, bases de données et  
archives d'images locales et distantes

# Les données de la recherche en général et l'exemple de l'astronomie

# Les infrastructures de recherche



Outil (ou dispositif) possédant des caractéristiques uniques identifiées par la communauté scientifique utilisatrice comme requises pour la conduite d'activités de recherche de haut niveau

Organisations internationales (CERN, EMBL, ESO)

Infrastructures nationales ou établies en coopération ou en réseau

La Feuille de route ESFRI a un fort effet structurant pour la recherche européenne

# Toutes les disciplines sont concernées



Des instruments, mais aussi des infrastructures « numériques », en particulier en Sciences humaines et sociales, dans la feuille de route nationale



- Huma-Num: faciliter le tournant numérique de la recherche en sciences humaines et sociales
- PROGEDO: favoriser la participation d'équipes françaises à de grandes enquêtes européennes et veiller à rendre accessible à tout chercheur la statistique publique



... et dans la feuille de route européenne



- DARIAH: Digital Infrastructure for the Arts and Humanities
- CLARIN: Common LAnguage Resources and technology INfrastructures

# La révolution numérique pour la science

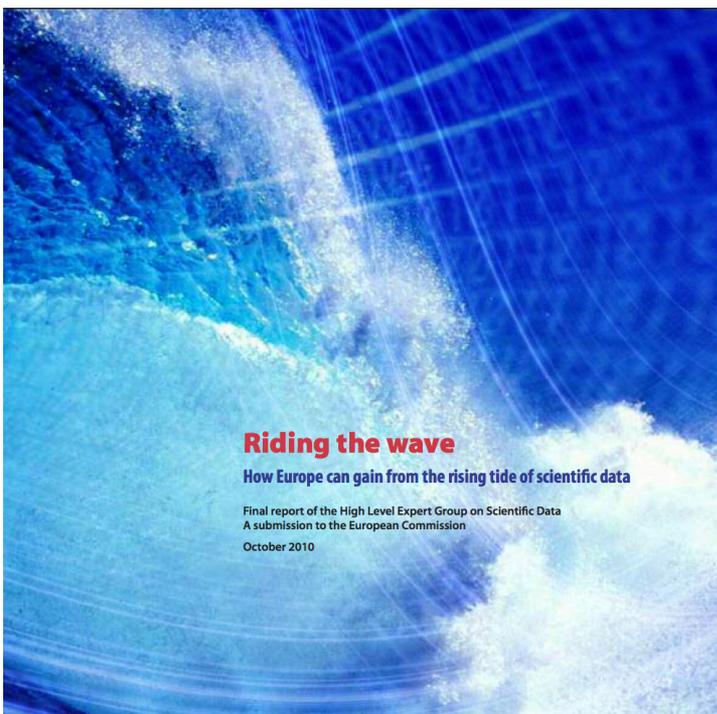
Réseaux de communication, moyens de calcul massif, acquisition des données, mais aussi **la mise à disposition des données**

Les données doivent devenir une des infrastructures de la recherche

Une révolution dans les méthodes de travail des chercheurs

*Découvrir les données, y accéder, les réutiliser, les combiner*

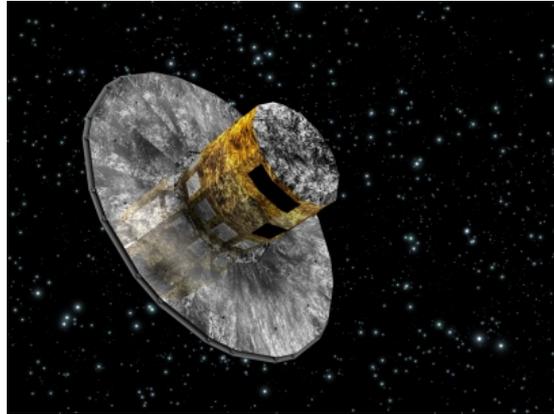
L'astronomie est l'exemple d'une discipline qui a effectué cette (r)évolution



# Les infrastructures de l'astronomie (exemples)



Vue d'artiste de SKA  
Crédit: SKA Organisation



Vue d'artiste de Gaia  
Crédit: ESA



Vue d'oiseau du Very Large Telescope  
Crédit: J.L. Dauvergne & G. Hüdepohl ([atacamaphoto.com](http://atacamaphoto.com))/ESO



Hubble, le télescope spatial  
Crédit: NASA, 2002



XMM-Newton  
Crédit: Image courtesy of ESA



Sous le charme des Nuages de Magellan, ALMA  
Crédit: ESO/C. Malin

# Les données en astronomie (I)

Des raisons scientifiques majeures de conserver les données et de les réutiliser

- Observations sur le long terme de phénomènes variables
- Un très grand nombre d'objets, des interactions complexes à différentes échelles (des phénomènes à l'oeuvre sur les grains du milieu interstellaire aux échelles cosmiques!)

Le changement de paradigme est effectif, les astronomes utilisant quotidiennement des données qu'ils trouvent dans des services distants (plus de 1 000 000 requêtes / jour au CDS)

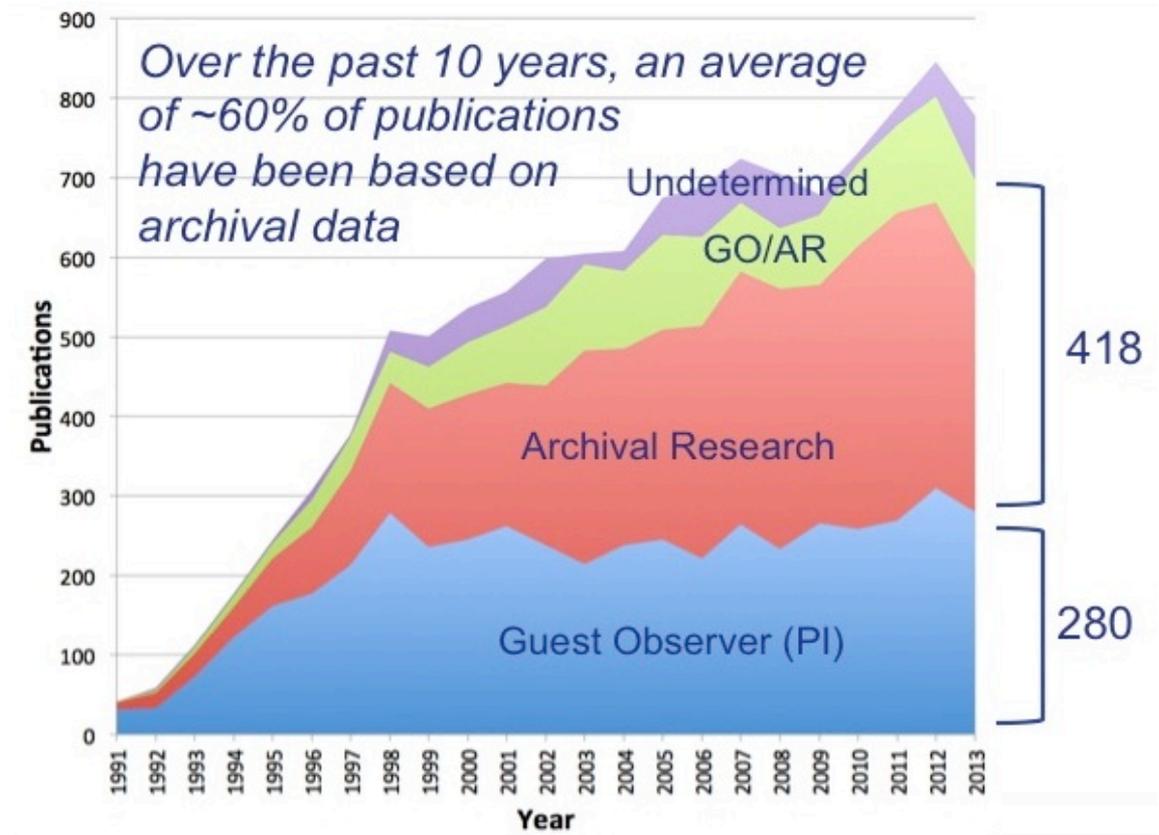
Les données sont réutilisées pour des objectifs différents des objectifs initiaux

Une augmentation significative du retour scientifique des gros investissements



# Illustration avec un instrument bien connu...

## L'exemple du HST (Hubble Space Telescope)



Remerciements:  
Robert J. Hanisch  
Space Telescope Science Institute

# Pourquoi ? Comment ?

L'habitude de travailler dans des collaborations internationales

Un format de données commun depuis les années 70, FITS (Flexible Image Transport System), toujours maintenu sous l'égide de l'Union Astronomique Internationale

Les données des observatoires sont « ouvertes », en général après une période « propriétaire » durant laquelle elles sont réservées aux personnes qui ont obtenu du temps d'observation sur appel d'offre – dans ces conditions les intérêts légitimes des personnes qui ont eu l'idée de l'observation sont préservés, et le fait que les données sont ouvertes est accepté par la communauté, qui en profite pour ses propres recherches

Excellente collaboration avec les journaux académiques de la discipline

# Mise en réseau des données

Elle a commencé dès le démarrage du Web (~1993), par la mise en ligne et en réseau des bases de données à valeur ajoutée et des publications, puis des archives

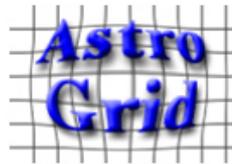
Depuis 2000 l'Observatoire Virtuel (OV) astronomique définit des standards d'interopérabilité et développe des outils pour permettre un accès unifié aux données

*Facilite l'accès aux données en éliminant l'étape d'apprentissage des interfaces propres à chacun des services de données*

## Mise en place de l'International Virtual Observatory Alliance

- Elle regroupe les projets OV d'une vingtaine de pays de tous les continents
- Utilisation de standards génériques quand c'est possible (registre des ressources OAI-PMH, vocabulaires SKOS / RDF), pour assurer la communication avec le monde extérieur

# L'International Virtual Observatory Alliance



# Le rôle de l'IVOA

L'IVOA est une « alliance » internationale qui coordonne les moyens à mettre en œuvre pour rendre interopérable les ressources des centres de données en astronomie

- Standard sur les formats utilisés en plus des standards disciplinaires (exemple: FITS) et la description des données
- Standards sur les protocoles de dissémination

Un fonctionnement « à la » W3C avec des groupes de travail et d'intérêt, des processus (discussions, RFC, etc.) permettant d'aboutir à des

Recommandations (exemple: VOTable 1.3)

<http://www.ivoa.net>

# L'Observatoire Virtuel, cela fonctionne comment ?

Pas de point central, un monde multipolaire

Un modèle ouvert

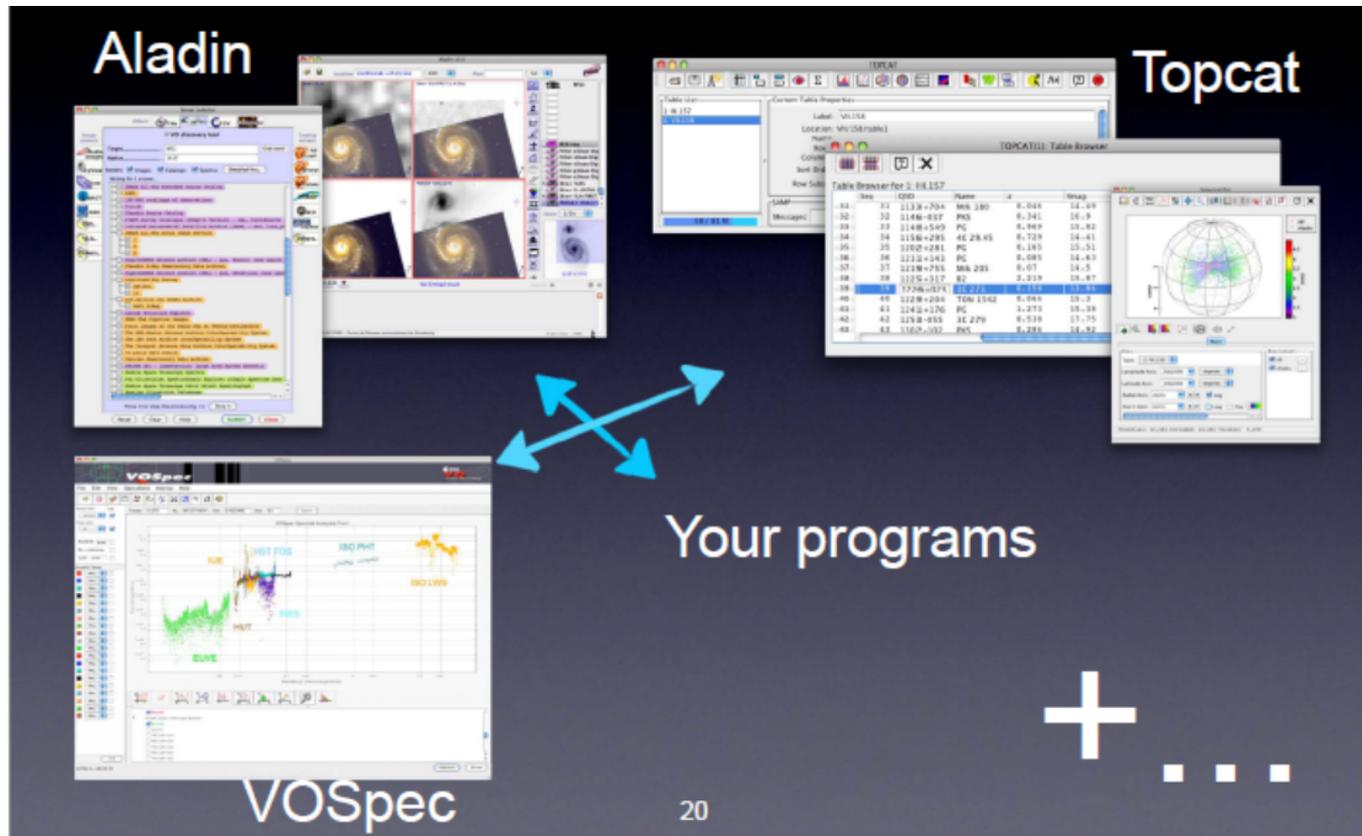
- Une fine couche assure l'interopérabilité mais chacun organise ses données comme il le souhaite
- Tout le monde peut enregistrer un service de données et développer un outil d'accès aux données

L'OV est « invisible », mais il est utilisé chaque fois que les chercheurs utilisent les outils ou les services

Les outils d'accès aux données « se parlent »

**LES DONNÉES SONT UNE DES INFRASTRUCTURES DE L'ASTRONOMIE**

# Une interopérabilité « transparente »



# Et pas seulement les données des grands instruments !

La définition en commun de standards et de métadonnées est la clé

Les données des grands observatoires, mais aussi les données « longue traîne » de la recherche

Le CDS accueille les données « attachées » aux publications: tables, images du ciel, séries temporelles, ..., dans un service de données (VizieR) qui accueille aussi les catalogues des grands relevés sol et spatiaux

Accord avec le journal « européen » Astronomy & Astrophysics dès 1993 sur la publication des tables (>~100 lignes) en ligne

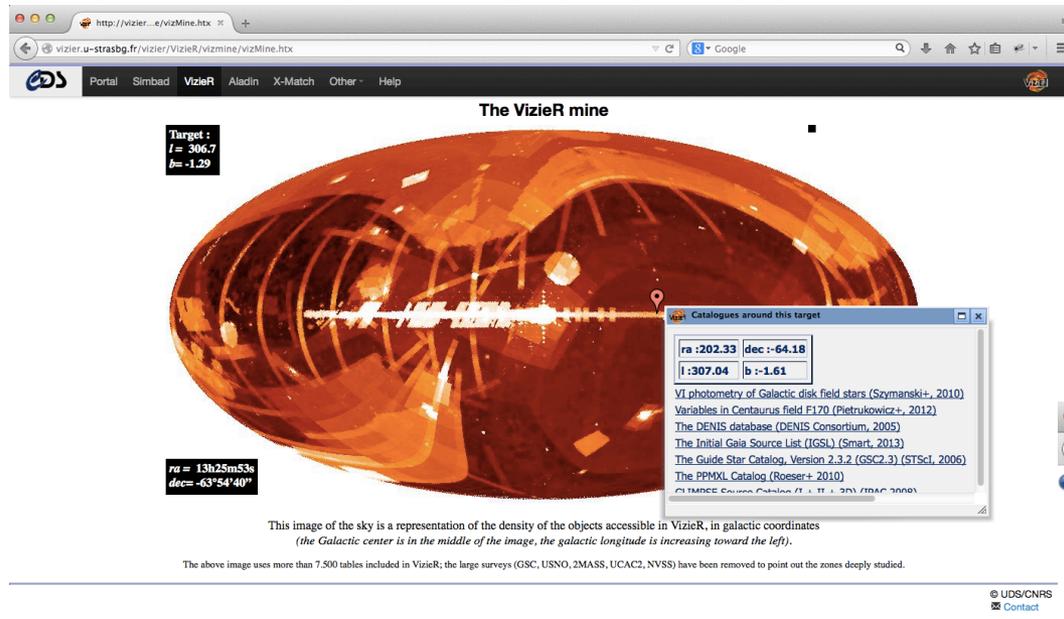
# Et pas seulement les données des grands instruments ! (2)

Les tables imprimées deviennent des données réutilisables

Description standard des données acceptée par tous les journaux, 13 000 « catalogues » incluant des données de tous les grands journaux internationaux de la discipline

Près de 300 000 requêtes extérieures par jour en moyenne sur ce service (VizieR), via le site Web ou par d'autres interfaces

# Illustration

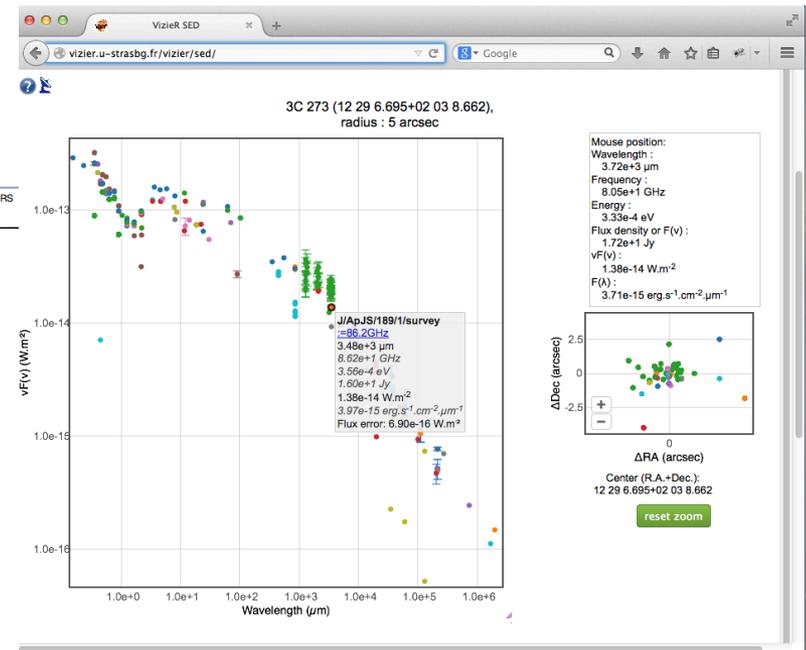


Des données validées par une publication que l'on peut découvrir et utiliser

Une vue homogène de données hétérogènes

Combinaison des données grâce aux métadonnées

Un spectre composite extrait de la collection



# Le rôle des documentalistes

Les documentalistes peuvent jouer un rôle important

Les données ne sont pas des publications, et **il faut des compétences disciplinaires**

Des documentalistes hautement spécialisés: 1/3 du personnel du CDS

Dans certains pays, les bibliothèques universitaires se positionnent comme lieu de dépôt et de diffusion des données

# Quelques mots à propos d'une autre « Alliance », RDA (Research Data Alliance)

# Research Data Alliance

Mise en place en mars 2013 par la CE, l'Australie et la NSF pour faciliter le partage des données par delà les frontières des états et les frontières disciplinaires

Travail sur les « blocs de base » nécessaires au partage des données

Un forum de discussions technologiques et disciplinaires

~2000 participants de 70 pays

# Research Data Alliance

RD | Research Data Shari... x +

https://rd-alliance.org

REGISTER LOGIN CONTACT US

**RDA** Research Data Sharing without barriers  
RESEARCH DATA ALLIANCE

ALL SOCIAL f t in

Home Organisation Working and Interest Groups Plenary Meetings News & Events Early Career Programmes About

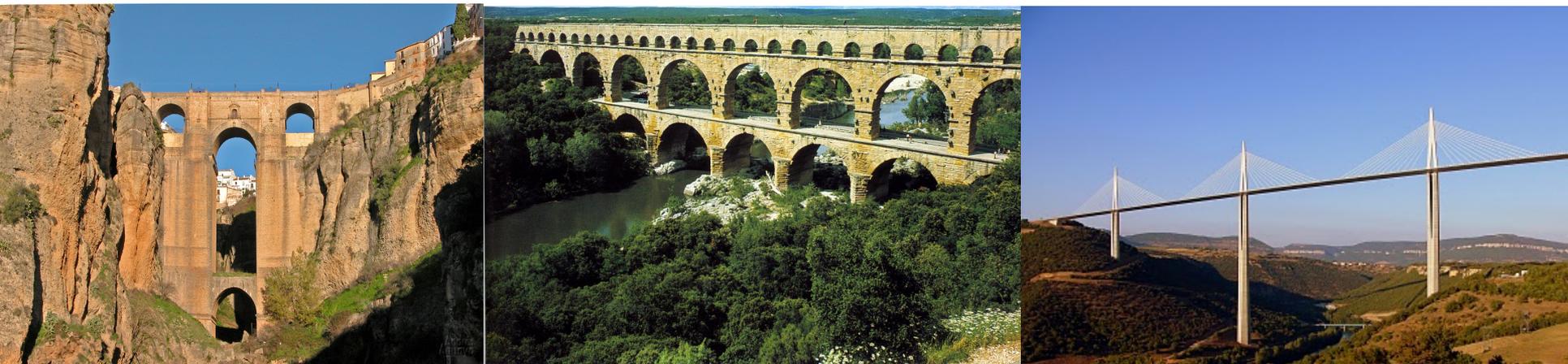
**RDA and global Data and Computing e-Infrastructure challenges Workshop**  
11&12 December 2014, Rome

**CALL FOR EXPRESSIONS OF INTEREST TO ORGANISE AN RDA PLENARY MEETING**

WORKING GROUPS	INTEREST GROUPS	BOF GROUPS
<b>RDA/WDS Publishing Data Bibliometrics WG</b> Kerstin Lehnert, Sarah Callaghan 23	<b>Domain Repositories Interest Group</b> George Alter, Ruth E. Duerr, Robert J. Hanisch, Peter Doorn 48	<b>Data re-use, share your experiences</b> Odile Hologne 21
<b>RDA/WDS Publishing Data Services WG</b> Hylke Koers, Adrian Burton 20	<b>Education and Training on handling of research data</b> Miguel-Angel Sicilia, Yuri Demchenko, Natasha Balac 54	<b>Sustainability of eResearch / Cyberinfrastructure</b> Stefanie Kethers, Andrew Treloar 7
<b>RDA/WDS Publishing Data Workflows WG</b> Sunje Dallmeier-Tiessen, Elizabeth Newbold, Theodora Bloom 25	<b>ELIXIR Bridging Force IG</b> Bengt Persson, Carole Goble, Rob Hooft 20	

# L'objectif ?

*Les chercheurs et les innovateurs partagent ouvertement les données par delà les frontières des technologies, des disciplines et des pays pour résoudre les grands défis..*

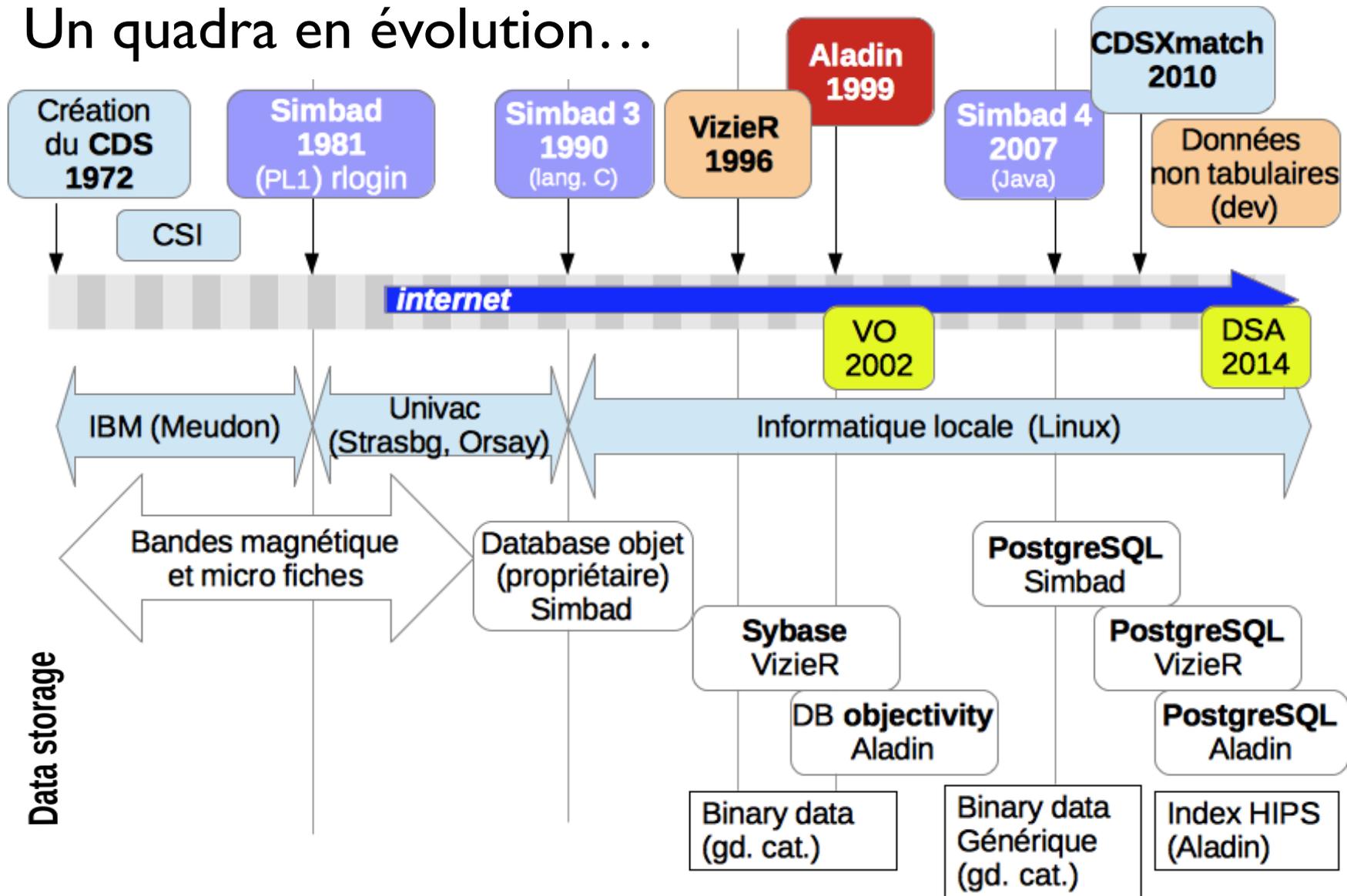


*... construire des passerelles (techniques, sociologiques) pour permettre le partage des données au niveau mondial*

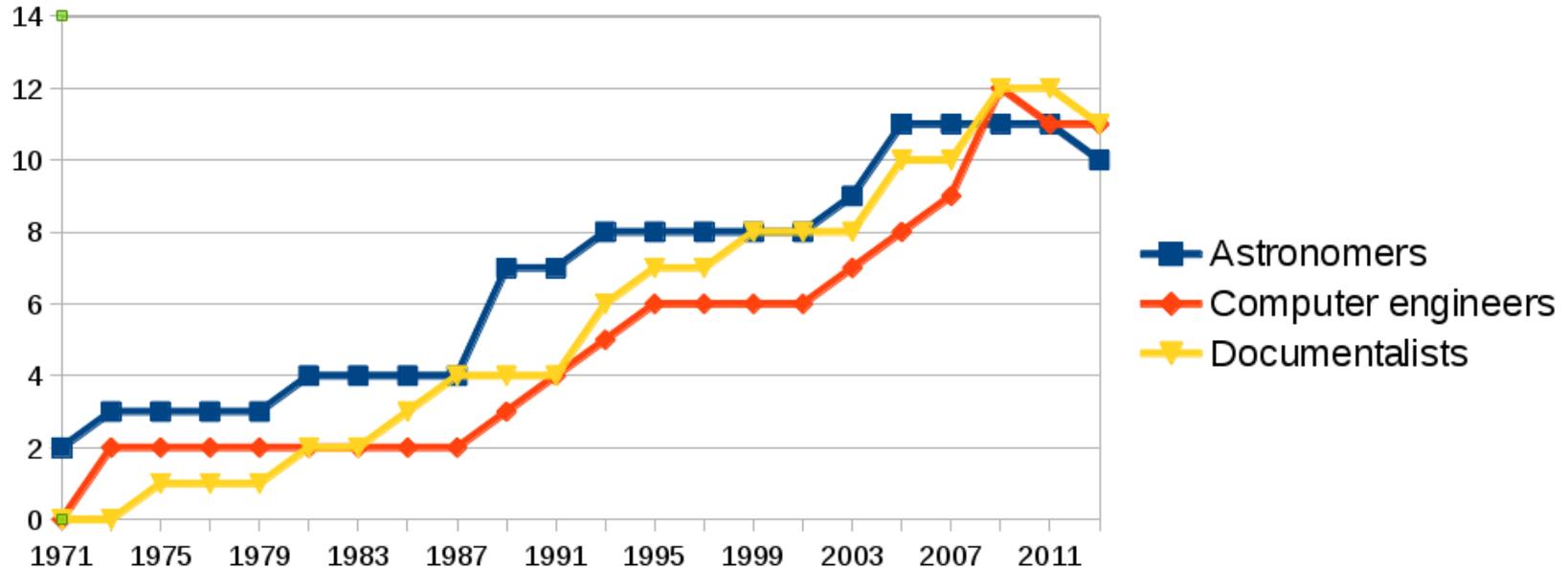
***Les chercheurs, les praticiens des données, les « technologistes » sont invités à travailler ensemble***

# Focus sur le Centre de Données astronomiques de Strasbourg

# Un quadra en évolution...



# Un personnel également en évolution



Un équilibre entre les 3 métiers...

# Illustration des migrations / évolutions du CDS



— 43 —

S<sup>o</sup>-10<sup>o</sup> +22<sup>o</sup>

1901-1960			1961-2020			2021-2080			2081-2140			2141-2200																	
m	8 <sup>u</sup>	+22 <sup>o</sup>	m	8 <sup>u</sup>	+22 <sup>o</sup>	m	8 <sup>u</sup> -9 <sup>u</sup>	+22 <sup>o</sup>	m	9 <sup>u</sup>	+22 <sup>o</sup>	m	9 <sup>u</sup> -10 <sup>u</sup>	+22 <sup>o</sup>															
9.2	14.0	2.6	9.2	28	8.3	18.2	K	8.1	47	37.7	14.5	K	9.5	13	40.5	27.2	9.1	46	20.1	8.9	K								
9.4	18.8	15.5	7.7	7.7	12.2	40.9	K	9.1	44.2	8.3	K	8.2	44.8	6.4	R	9.2	39.8	5.9	9.2	39.8	5.9	K							
9.3	20.7	47.5	9.2	18.7	13.7	K	9.5	45.4	58.9	9.2	46.9	58.3	B	9.4	44.9	28.0	9.4	44.9	28.0	9.4	44.9	28.0	K						
9.3	29.2	9.0	9.3	48.4	49.0	K	9.5	45.8	42.1	9.2	47.2	1.8	B	9.5	48	6.6	20.3	9.5	48	6.6	20.3	9.5	48	6.6	K				
9.2	51.1	55.4	8.6	54.3	29.0	K	8.9	48	0.7	24.8	K	9.5	14	8.5	8.6	9.4	39.3	11.3	9.4	39.3	11.3	9.4	39.3	11.3	K				
9.5	53.2	47.1	9.5	20	18.1	31.2	9.5	11.5	59.6	B	9.0	52.0	56.8	9.5	49	40.4	34.2	9.5	49	40.4	34.2	9.5	49	40.4	34.2	K			
9.4	56.4	1.3	9.5	32.9	44.5	9.2	26.3	26.5	k	9.1	15	1.5	49.2	K	8.2	50	23.4	59.3	8.2	50	23.4	59.3	8.2	50	23.4	59.3	K		
9.3	3.4	8.8	9.5	47.6	54.2	9.5	40.1	33.7	9.5	16	13.1	43.1	6.7	51	24.3	1.6	6.7	51	24.3	1.6	6.7	51	24.3	1.6	6.7	51	24.3	1.6	K
9.3	21.0	22.2	9.5	30	29.2	40.3	6.7	42.7	24.8	L	9.2	39.2	20.3	9.5	30.8	24.2	9.5	30.8	24.2	9.5	30.8	24.2	9.5	30.8	24.2	9.5	30.8	24.2	K
9.4	29.0	17.8	9.5	33.5	18.3	9.5	49	7.0	18.9	9.5	17	32.3	11.3	9.5	41.6	7.0	9.5	41.6	7.0	9.5	41.6	7.0	9.5	41.6	7.0	9.5	41.6	7.0	K

VizieR

vizier.u-strasbg.fr/viz-bin/VizieR-4

Portal Simbad VizieR Aladin X-Match Other Help

### VizieR Result Page

Send to VO tools

Search Criteria  
Save in CDSportal

Keywords: I/122/bd [Add]

Tables: I/122, ..bd, ..bdchg [Choose]

Constraints: [Modify Query]

Preferences: max: 20, HTML Table, All columns, Compute [Submit]

Mirrors: CDS, France

Show the target form  
Show constraint information

The 4 columns in color are computed by VizieR, and are *not part of the original data*. The precision of the *computed positions* has been increased compared to the original positions.

I/122/bd Bonner Durchmusterung (Argelander 1859-1903) ReadMe+ftp Post annotation The Bonner Durchmusterung catalog (325037 1903BD...C...0A)

start AladinLite

Full	RAJ2000	DEJ2000	zonesign	zone	num	suppl	mag	RA1855	DE1855	RA.icrs	DE.icrs
"h:m:s"	"d:m:s"	"d:m:s"	deg	deg	deg	mag	"h:m:s"	"d:m:s"	"h:m:s"	"d:m:s"	
1	11 49 28.2	+89 35 15	+	89	1		9.5	00 11 05.0	+89 36.2	11 49 28.2	+89 35 15
2	04 43 00.2	+89 37 52	+	89	2		9.2	01 17 35.0	+89 00.2	04 43 00.2	+89 37 52
3	09 46 02.7	+89 34 08	+	89	3		8.8	01 49 36.0	+89 29.2	09 46 02.7	+89 34 08
4	08 57 15.8	+89 35 58	+	89	4		9.4	01 50 57.0	+89 23.6	08 57 15.8	+89 35 58
5	07 17 48.9	+89 36 11	+	89	5		9.5	01 51 58.0	+89 13.3	07 17 48.9	+89 36 11
6	06 02 42.2	+89 25 50	+	89	6		9.4	02 16 43.0	+89 00.6	06 02 42.2	+89 25 50
7	07 06 10.6	+89 16 32	+	89	7		9.3	03 11 12.0	+89 04.3	07 06 10.6	+89 16 32
8	07 21 43.9	+89 05 41	+	89	8		9.5	03 53 49.0	+89 00.5	07 21 43.9	+89 05 41
9	09 27 57.8	+89 09 51	+	89	9		9.1	04 49 39.0	+89 27.1	09 27 57.8	+89 09 51
10	08 44 18.5	+88 49 27	+	89	10		9.5	05 43 44.0	+89 06.0	08 44 18.5	+88 49 27
11	08 46 46.8	+88 45 32	+	89	11		9.5	05 57 03.0	+89 03.6	08 46 46.8	+88 45 32
12	10 29 31.8	+89 04 55	+	89	12		9.1	06 17 38.0	+89 37.9	10 29 31.8	+89 04 55
13	09 21 51.3	+88 34 10	+	89	13		7.0	07 03 40.0	+89 01.8	09 21 51.3	+88 34 10
14	09 26 40.0	+88 32 05	+	89	14		9.5	07 14 02.0	+89 01.0	09 26 40.0	+88 32 05
15	09 35 19.5	+88 30 48	+	89	15		9.2	07 27 50.0	+89 01.6	09 35 19.5	+88 30 48
16	11 12 19.0	+88 55 13	+	89	16		9.5	08 58 47.0	+89 39.4	11 12 19.0	+88 55 13
17	11 30 54.7	+88 44 55	+	89	17		9.0	10 26 21.0	+89 32.0	11 30 54.7	+88 44 55
18	11 51 48.6	+88 55 43	+	89	18		8.9	11 08 27.0	+89 43.9	11 51 48.6	+88 55 43
19	11 46 50.0	+88 44 49	+	89	19		9.5	11 09 11.0	+89 32.9	11 46 50.0	+88 44 49
20	12 05 23.5	+88 37 52	+	89	20		9.5	12 00 23.0	+89 26.3	12 05 23.5	+88 37 52

Du papier (le catalogue Bonner Durchmusterung (F.W.A.Argelander 1799-1875)) au numérique (via l'interface Web de VizieR)

# La dissémination des données

## Des données en libre accès

- Des applications Web ou « standalone » (Aladin) dédiées à chaque service
- Accès via les outils de l'Observatoire Virtuel ou par scripts
- Les formats des résultats des requêtes reconnus par la communauté astronomique: TSV, VOTable, FITS
- Des recherches de ressources (catalogues / tables / images...) facilitées par des métadonnées permettant des recherches multicritères
- Une indexation efficace

## Une identification pérenne et reconnue par la communauté

- Un identifiant « bibcode » est attribué à chaque article
- Un identifiant est attribué à chaque catalogue

Le « bibcode » et l'identifiant catalogue sont reconnus et utilisés par l'application de bibliographie ADS qui indexe les articles parus en astronomie  
L'identifiant catalogue est présent dans le « registry » de l'Observatoire Virtuel pour être utilisé par des application externes

# Un effort de labellisation



Obtention du « **Data Seal of Approval** » en août 2014

Le DSA est un label international pour la préservation des données scientifiques sur le long terme

Il nécessite une réévaluation tous les 2 ans et il comporte 16 critères pouvant se résumer aux règles suivantes:

- Les données sont accessibles via Internet
- Les licences / droits d'utilisations des données sont clairement exprimés
- Les données sont réutilisables
- Les données sont fiables
- Les données sont identifiées de manière unique selon une nomenclature permettant un référencement externe

Organisation de l'infrastructure selon un modèle reconnu internationalement et adaptée au cas spécifique, « OAIS like »

# Mais une labellisation ciblée

Seuls « VizieR » et « Aladin » sont labellisés

« Simbad » ne peut pas être vu comme une archive car:

- c'est une base de données vivante construite au jour le jour à partir des informations contenues dans les articles publiés
- l'identification des objets n'est pas figée (l'amélioration continue de l'instrumentation peut faire évoluer la nature des objets, une étoile observée aujourd'hui peut se révéler comme étant une étoile double lors d'observations futures)

# Labellisation DSA et structure du Système d'Information

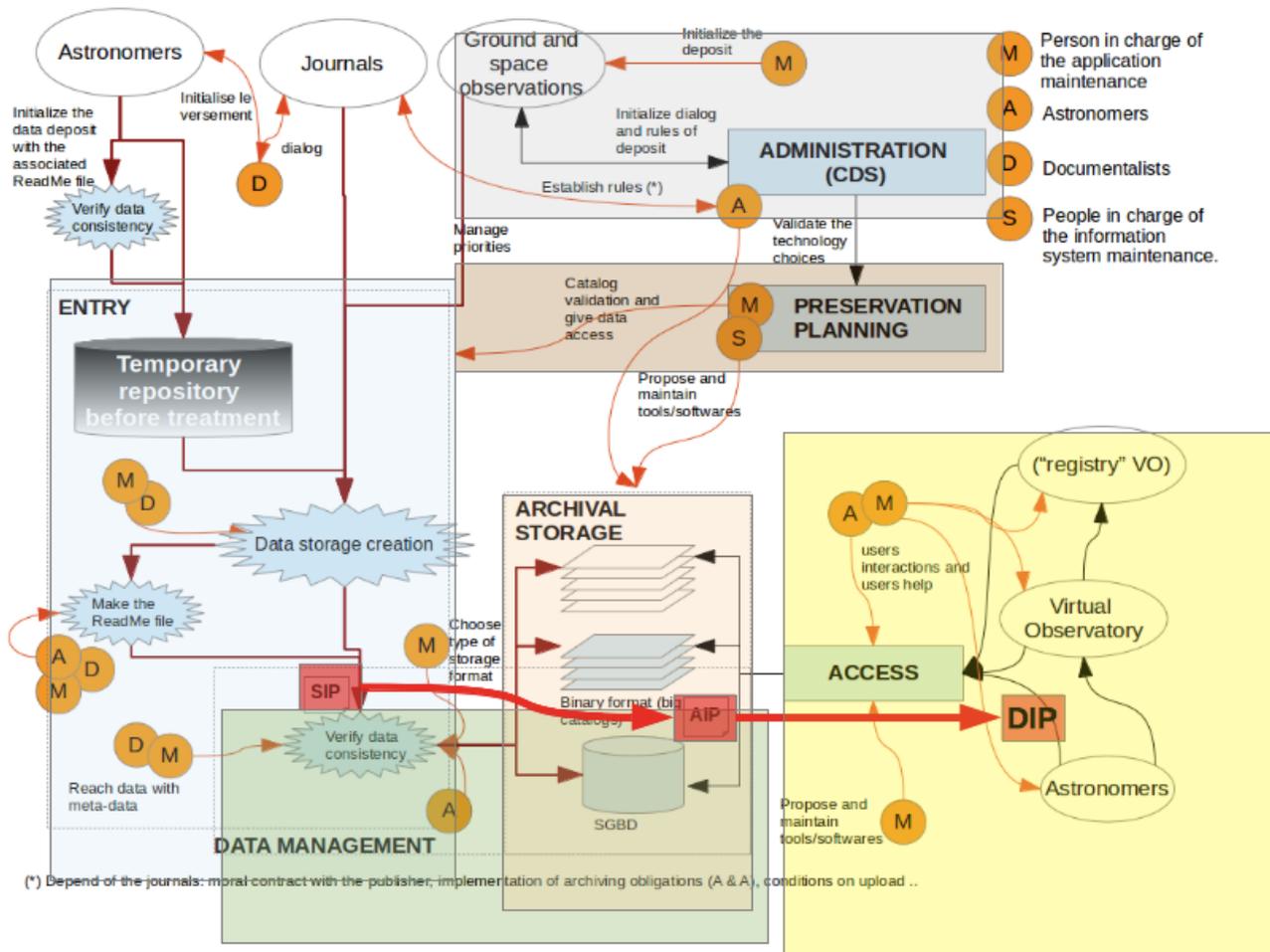
**OAIS** = **O**pen **A**rchive **I**nformation **S**ystem (ISO 14721)

C'est un modèle conceptuel destiné à la gestion, à l'archivage et à la préservation à long terme de documents numériques.

- Définit les acteurs / responsabilités dans le SI: producteur / utilisateur / manager
- Définit un modèle fonctionnel de l'archive en 6 composantes
- Définit les flux d'informations, les paquets OAIS: SIP (en entrée), AIP (archives), DIP (diffusion)
- Définit des normes / recommandations « ouvertes », exemple: modalité de versements des données



# Architecture OAIS-like de VizierR



# L'ingestion des données issues des journaux

J/A+A/424/545      Optically faint obscured quasars      (Padovani+, 2004)

Discovery of optically faint obscured quasars with Virtual Observatory tools.  
Padovani P., Allen M.G., Rosati P., Walton N.A.

<Astron. Astrophys., 424, 545-559 (2004)>  
=2004A&A...424..545P

ADC\_Keywords: QSOs ; Active gal. nuclei ; X-ray sources  
Keywords: astronomical data bases: miscellaneous - methods: statistical -  
galaxies: quasars: general - X-rays: galaxies

## Abstract:

We use Virtual Observatory (VO) tools to identify optically faint, obscured (i.e., type 2) active galactic nuclei (AGN) in the two Great Observatories Origins Deep Survey (GOODS) fields. By employing publicly available X-ray and optical data and catalogues we discover 68 type 2 AGN candidates.

## File Summary:

FileName	Lrecl	Records	Explanations
ReadMe	80	.	This file
table1.dat	90	47	Type 2 AGN candidates, NDF-N
table2.dat	90	21	Type 2 AGN candidates, CDF-S
table4.dat	90	3	Type 2 AGN candidates, UDF

## See also:

J/AJ/126/539 : The Chandra Deep Fields North and South (Alexander+, 2003)  
J/ApJS/155/271 : Chandra Deep Field-South: Optical spectroscopy (Szokoly+ 2004)  
II/258 : Hubble Ultra Deep Field Catalog (UDF) (STScI, 2004)  
II/261 : GOODS initial results (Giavalisco+, 2004)

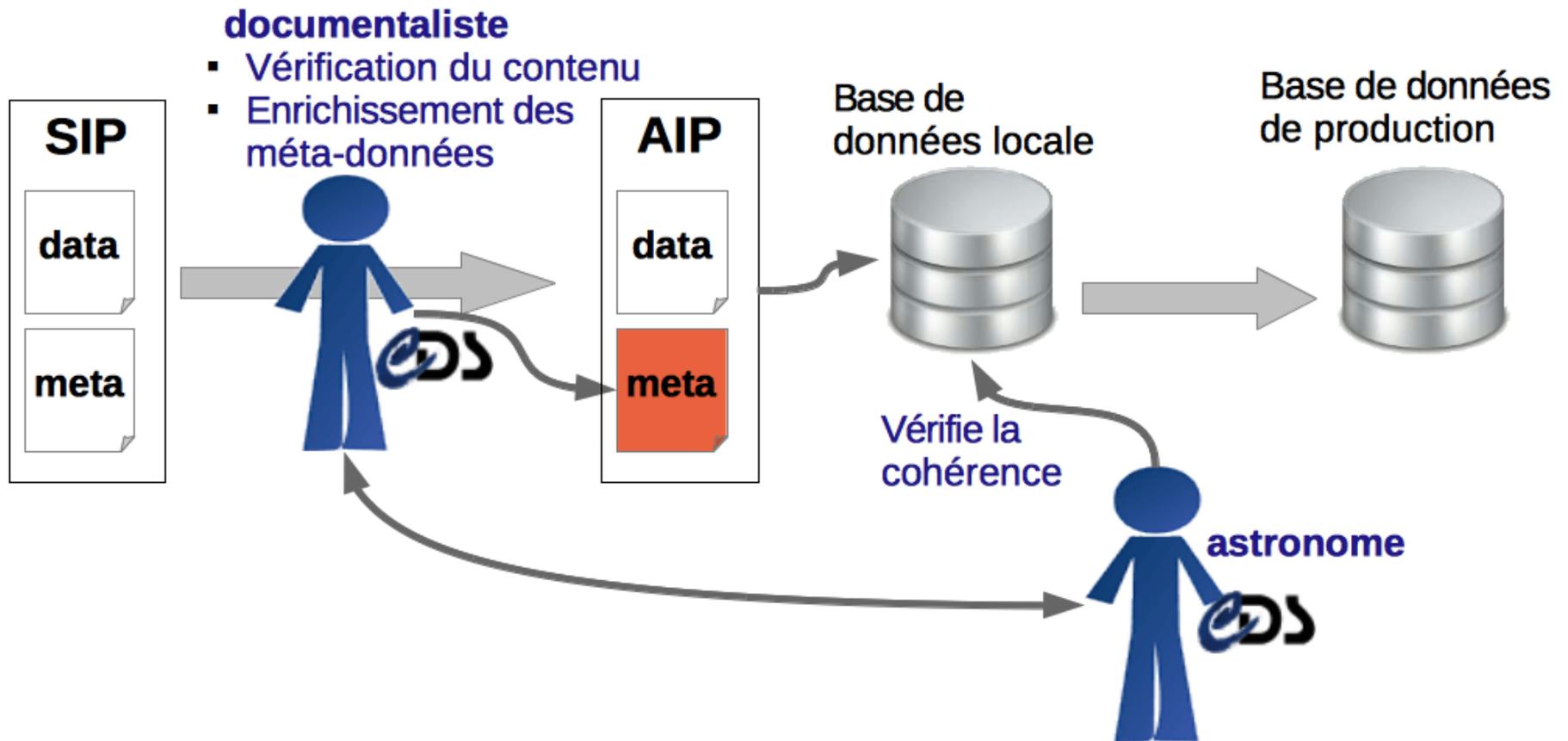
## Byte-by-byte Description of file: table\*.dat

Bytes	Format	Units	Label	Explanations
1- 19	A19	---	GOODS	GOODS designation (JHHMSS.ss+DDMMSS.s)
22- 25	I4	---	UDF	? UDF designation (Cat. II/258, table 4 only)
27- 29	I3	---	A03	Alexander et al. (2003, Cat. <J/AJ/126/539> sequential number, [ABB2003] CDFN NNN (table1 or [ABB2003] CDFS NNN (table2&3) in Simbad
31- 33	I3	---	S04	? Szokoly et al. (2004, Cat. <J/ApJS/155/271> sequential number, [SBH2004] XID NNNa in Simbad (table2 only)
	34	A1	m_S04	[a] Multiplicity index on S04
35- 36	I2	h	RAh	Right ascension (J2000.0)
38- 39	I2	min	RAm	Right ascension (J2000.0)
41- 45	F5.2	s	RA s	Right ascension (J2000.0)

- Les données tabulaires sont accompagnées d'un fichier « ReadMe » (en ASCII) décrivant les données, il constitue le paquet SIP, **construit par les éditeurs, les auteurs ou les documentalistes**
- Il est accompagné d'un fichier de configuration construit par les documentalistes pour enrichir les données / métadonnées
- Les données / métadonnées tabulaires sont stockées sous forme de tables dans le SGBD



# Une fiabilisation des données en entrée



# Le cas des grandes volumétries (VizieR)

Les grands catalogues issus des centres spatiaux et télescopes

- ~30 grands catalogues (parmi 13 000) stockés sur ~6Tb
- Souvent la totalité du ciel (et peuvent comporter des images ou des données tabulaires)
- La plus grande table Gaia (simulation Gaia) comporte ~2milliards de lignes de ~40 colonnes
- Les catalogues les plus interrogés

Utilisation de formats adaptés, propres au CDS, pour maintenir des temps d'accès faibles (garantie de l'interactivité des services)  
+ stockage récent en SGBD pour les besoins de l'Observatoire Virtuel

# L'accès aux métadonnées

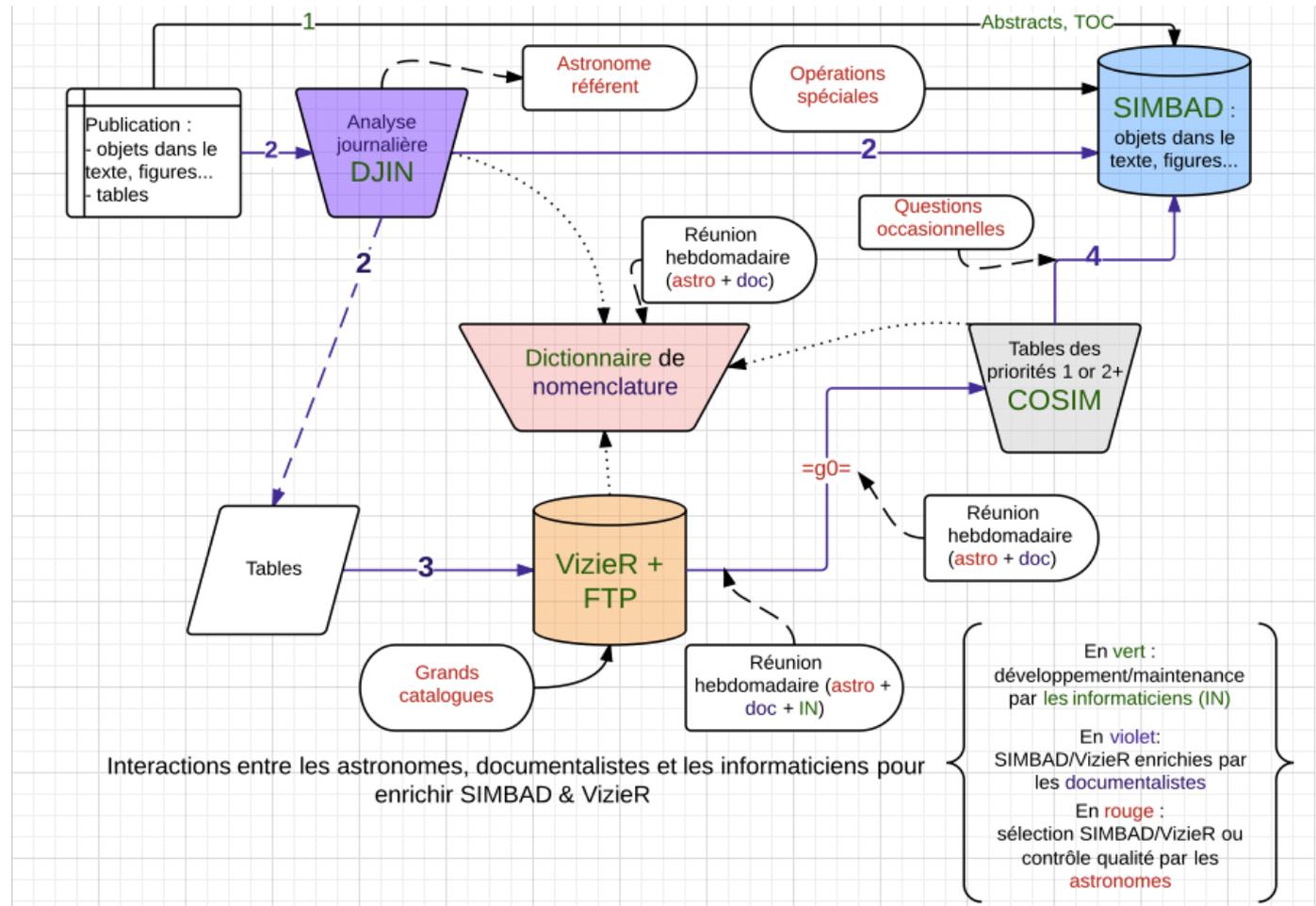
Les métadonnées sont distribuées dans les résultats des requêtes (VOTable)

Tous les catalogues contiennent un fichier « ReadMe » accessible via le site Web: il contient notamment un lien vers l'article original: littérature, licences ...

The screenshot shows the Astronomy & Astrophysics journal website. At the top, there is a search bar and navigation links. The main content area displays a list of articles with columns for issue information, article title, and links to abstracts, full HTML, PDF, and references. A specific article is highlighted with a blue box, showing its title and a link to the 'ReadMe' file. The 'ReadMe' file is described as a 'Standardized Description of the Catalogue ReadMe'.

The screenshot shows the 'Access to Astronomical Catalogues' website. It displays a detailed view of an article titled 'Companions to close spectroscopic binaries (Tokovinin+, 2006)'. The article includes a summary, abstract, and keywords. A table of related records is shown below the article, with columns for 'Objects', '2006A&A...', and 'ReadMe+ftp'. A blue box highlights the 'ReadMe+ftp' link, and a blue arrow points to it from the 'Standardized Description of the Catalogue ReadMe' text.

# Un travail d'équipe: exemple d'interaction entre les personnes



# « Working together »

## Interactions between documentalists and astronomer (Nomenclature team) : questions



### Is a new acronym needed ?

- « This supernova host galaxy is not really detected. Is an acronym necessary in that case? »
- « For these microlensing events, it is possible to retrieve the official EWS name (from the OGLE website). However would an acronym be useful to keep track of the nomenclature used here (since the names look like EWS names, but the positions don't match)? »

## Interactions between documentalists and astronomers (VizieR team) : example of questions

### Technical problems :

- What should I do with this kind of FITS file that cannot be handled with the regular commands ?
- UCD « PHOT\_SB\_IR » not understood by the program which stops there...
- The graphic does not work...

### General astronomical problems :

- Why do I have  $<-0.555$  in the online table whereas I have  $-1 -9 0.455$  in my prepared table ? (when the  $HR=-1$  or  $1$ , we have limit value intervals in the three columns HR and lower, upper errors so it is the same thing!)
- « tau » (cooling time) from this other reference (J/ApJS/156/47) or this work?

### Identification problems to link objects to SIMBAD :

- J/ApJS/190/233/table10 : 561 HII regions without coordinates like +264-037 or H13 to be retrieved from 54 different references not all with acronyms...

2014/09/09

CDS Scientific Council – based on LISA VII presentation  
CDS : the symbiosis between astronomers, documentalists & IT specialists

13

## Interactions between documentalists and astronomers (COSIM team) : example of questions

### Merge/scission discussions



- 2011A&A...530A..60M: NVSS J114454+195130 at 2.6" of a GiC, merge?
- In the Bootes and First Look Survey fields, can I merge the 2MASS and ELAIS IR sources?
- Should I create the BC component of this WDS system or do I add the reference on the existing WDS J06071+3228B and WDS J06071+3228C?
- Follow the cross-id with FIRST 132030.9+332608 and AX J132032+3326 given in table 7?

### Precisions about data to add in SIMBAD

- 2014ApJ...788...122P: OMC-2 FIR 4 = [NCM2003] MIR 24 = HOPS 108 whereas in 2014ApJ.786.26, [NCM2003] MIR 24 = HOPS 64?
- Is [SCK2004] G020.08-0.14 N, the same object as IRAS 18253-1130 & [ABJ2009] U20.08-0.14 also used in 2013MNRAS.431.2385X for this region?

### SIMBAD/article incoherence (scission/merge) :

- In SIMBAD : [MWZ90] OMC-2 FIR 4 = [NCM2003] MIR 24 = HOPS 108 whereas in 2014ApJ.786.26, [NCM2003] MIR 24 = HOPS 64?
- Is [SCK2004] G020.08-0.14 N, the same object as IRAS 18253-1130 & [ABJ2009] U20.08-0.14 also used in 2013MNRAS.431.2385X for this region?

### Is it for SIMBAD ?

- 2014ApJ.788.122, Tables 1 & 2 : for SIMBAD? (answer : tough one ! These are objects but part of molecular clouds with peculiar properties without correspondence in the SIMBAD object list...)

### Do we have to change data in SIMBAD ?

- 2014MNRAS.437.2017T, Table 1 : redshifts for Q0107-025A & B are not the same in Simbad : which one to keep (with which quality)?
- 2013MNRAS.431.1005D, Table 2: Is there a hierarchical link to add to IC 2391, and in that case, with which probability?

2014/09/09

CDS Scientific Council – based on LISA VII presentation  
CDS : the symbiosis between astronomers, documentalists & IT specialists

11

Illustrations extraites de présentations de LISA VII et du Conseil Scientifique du CDS (6&9/2014, E. Perret)

# Le rôle de l'OV dans la dissémination des données

La réutilisation des données est garantie par l'utilisation de standards

- En utilisant des sorties standards (VOTable / XML) comprises par les applications de l'OV comprenant les données + les métadonnées
- En utilisant une nomenclature standard pour désigner le contenu des colonnes des tables
- En implémentant les services OV pour accéder aux données :
- Recherche spatiale
- Interrogation des tables par un langage commun basé sur SQL et incluant des fonctionnalités nécessaires pour l'astronomie

Une visibilité des données du CDS à travers l'Observatoire Virtuel

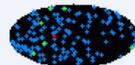
Le « registre OV » est un « dictionnaire de ressources », basé sur le standard OAI-PMH, est interrogé par les outils de l'OV

# Le rôle de l'OV dans la dissémination des données (2)

Send to VO tools Broadcast 

▶ [Show the target form](#)  
 ▶ [Show constraint information](#)

The 4 columns in *color* are computed by VizieR, and are *not part of the original data* (note that the *computed coordinates* are computed from the positions *and* the proper motions given in the table)  
 The precision of the *computed positions* has been increased compared to the original positions.

[J/A+A/450/681/table2](#) [Companions to close spectroscopic binaries \(Tokovinin+, 2006\)](#) [2006A&A...450..681T](#) [ReadMe+ftp](#) 

[Post annotation](#) Object list (160 rows)

 start AladinLite

Full	RAJ2000	DEJ2000	n	HIP	RAJ2000	DEJ2000	p
	"h:m:s"	"d:m:s"			"h:m:s"	"d:m:s"	
1	00 12 30.0	+14 33 49*		999	00 12 30	+14 33 49	
2	00 35 15.0	-03 35 34		2762	00 35 15	-03 35 34	
3	00 35 34.0	-39 44 45		2790	00 35 34	-39 44 45	

Aladin v7.5  
 File Edit Image Catalog Overlay Tool View Interop Help

Location: 07:14:03.24 -11:20:02.2  
 \*Optical \*IR \*UV \*Radio \*DSS \*Simbad \*NED



Search

n	HIP	RAJ2000	DEJ2000	pmRA	pmDE	Plx	Vmag	B-V	SpType	N	M	Z	Sj	Simbad	HD/BD
*	21395	04 35 34	+12 06 02	40	-17	13.5	8.99	0.59	F8	3	1	1	0	Simbad	286898
*	21482	04 36 48	+27 07 57	232	-147	56.0	8.24	1.10	K2	3	0	1	0	Simbad	283750
*	22394	04 49 13	+24 48 11	85	-52	19.0	9.58	1.05	K3V	2	1	1	1	Simbad	283882
*	22524	04 50 48	+16 12 38	84	-24	19.3	7.32	0.54	F8	2	1	1	1	Simbad	30738
*	24663	05 17 31	+20 07 55	-14	-94	13.7	6.74	0.52	F7V	3	0	1	0	Simbad	34335
*	30630	06 26 10	+18 45 26	-119	-164	88.2	6.91	0.94	K0	3	0	0	0	Simbad	45098
*	31850	06 39 31	+24 36 00	19	90	28.4	6.49	0.52	F8IV	2	1	0	0	Simbad	47415
*	36238	07 27 44	+21 26 44	-49	-123	29.4	5.29	0.46	F5IV-V	5	0	1	2	Simbad	58728

TIP: Adjust the zoom factor via the mouse wheel

95 sel / 50 src 158MB

# Traçabilité des données

L'origine des données est assurée par la provenance des données: articles de journaux, centres spatiaux, etc.

Historisation des modifications (disponible uniquement par le CDS)

Chaque ingestion / mise à jour est « signée » par la personne du CDS qui ingère / met à jour un catalogue

La correspondance mail (liée aux diverses opérations) est sauvegardée systématiquement

# Les grands partenaires du CDS

## Les relations du CDS avec les centres producteurs de données

- Relation avec les grands producteurs de données (ESO, ESA, NASA, ...)
- Les accords avec les journaux par contrat (A&A)
- Discussions avec les journaux (ex:AAS pour le choix des métadonnées et pour la mise en place des pipelines)

## Les relations du CDS avec des partenaires internationaux

- Le CDS est un acteur clé dans les prises de décisions de l'Observatoire Virtuel: propositions,...
- Des relations avec des partenaires comme celles avec le portail Web pour la bibliographie en astronomie (ADS)

## Relation avec les astronomes

- Les astronomes du CDS forment le 1/3 de l'équipe du CDS
- Les réunions annuelles de l'AAS permet de rencontrer les astronomes américains

Le conseil scientifique (composé de 6 membres français et 6 membres étrangers) évalue de manière critique le CDS et émet des propositions , ...

# Conclusion

Les disciplines scientifiques ont toutes vocation à partager leurs données (compte-tenu évidemment d'éventuelles restrictions légitimes)

Elles sont dans des états différents en terme de prise de conscience et de réalisation

Ce sont elles qui doivent faire l'effort initial, pour que les données soient correctement choisies et documentées

Le cadre du partage des données scientifiques est par essence international

Il faut aussi pouvoir stocker et distribuer les données, mais la mise en place d'infrastructures de données doit inclure un soutien aux communautés, y compris pour des discussions internationales

# Recommandations

Equipe CDS et plus particulièrement F. Genova, G. Landais et E. Perret pour leur contribution au contenu de cette présentation  
(« working together !! »)

# Des questions ?