

Annuaire des dépôts et des
services de données du CNRS
Avancement des travaux du
groupe de travail DIST

Françoise Genova, Paolo Lai

2^{ème} Journée pour la Science Ouverte du CNRS

Le contexte: Feuille de Route Science Ouverte du CNRS



Action 5 de l'Axe 2 (Données):

Créer et afficher un répertoire des dépôts et des services de données dont le CNRS est responsable et auxquels il participe.

Ce répertoire permettrait d'afficher un élément important de la contribution du CNRS à la Science ouverte. Il regrouperait les services de données des infrastructures de recherche cités mis en place dans le cadre de l'action 3, les centres et services de données thématiques (action 4) et l'entrepôt de données généraliste éventuellement créé dans le cadre de l'action 4.

Groupe de Travail créé en janvier 2020

- Mandat du Groupe de Travail: *Mettre en œuvre l'Action 5 de l'Axe 2 de la Feuille de Route Science Ouverte du CNRS*
Identifier les dépôts de données et les services dont le CNRS est responsable et auxquels il participe
- Les discussions du GT montrent qu'il est aussi utile d'identifier les dépôts de données et les services « utiles aux membres de l'institut » (vision « bottom-up »)
Nous avons rassemblé des éléments, mais la priorité est de mettre en œuvre l'Annuaire

Le point de départ: Cat OPIDoR

- [Cat OPIDoR](#): Le wiki des services dédiés aux données de la recherche, maintenu par l'INIST
- Contenu: Information, Formation, Accompagnement, Outil de gestion de données, Plateforme d'acquisition, Plateforme de calcul, Entrepôt de données, Plateforme d'accès, Plateforme d'archivage
- Les actions menées et à mener
 - *Vérifier et compléter les informations*
 - *Identifier celles qui doivent apparaître dans l'Annuaire*
 - *Modifier éventuellement le modèle de données si des améliorations sont identifiées et travailler sur la visualisation de l'Annuaire*

Les objectifs et le contexte

- Afficher un élément important de la contribution du CNRS à la Science Ouverte
- Les données sont un point de départ et un produit du processus scientifique
- Ce sont les Instituts qui connaissent le paysage disciplinaire
 - ⇒ Implication des Instituts
- Point de vue disciplinaire complété par la vision « top-down » des Infrastructures de Recherche de la Feuille de Route Nationale
 - ⇒ Gros producteurs de données
 - ⇒ Point d'entrée stratégique vers le paysage national, européen et international
 - ⇒ Implication du Comité TGIR

Les membres du Groupe de Travail

Pilotes: Françoise Genova (DIST/Observatoire Astronomique de Strasbourg), Paolo Lai (Inist-CNRS)

Membres du GT: représentants des instituts et du Comité TGIR

- INSB: Marie Chabbert, Olivier Coulon, Laurent Heliot, avec Claudine Médigue
- INC: Marc Baaden, Stéphanie Castex
- INEE: Cécile Callou, Chloé Martin
- INSHS: Emmanuelle Morlock, Lionel Maurel
- INSIS: Benoît Pier
- INSMI: Christophe Berthon, Philippe Briand
- INP: Laurent Lellouch
- INS2I: Michel Daydé (COCIN)
- IN2P3: Volker Beckmann puis Sabine Crépe-Renaudin
- INSU: Aude Chambodut, Jean-Pierre Vilotte
- TGIR: Gilles Ohanessian

En support au GT: INIST: Anne Ciolek-Figiel, DIST: Christine Hadrossek

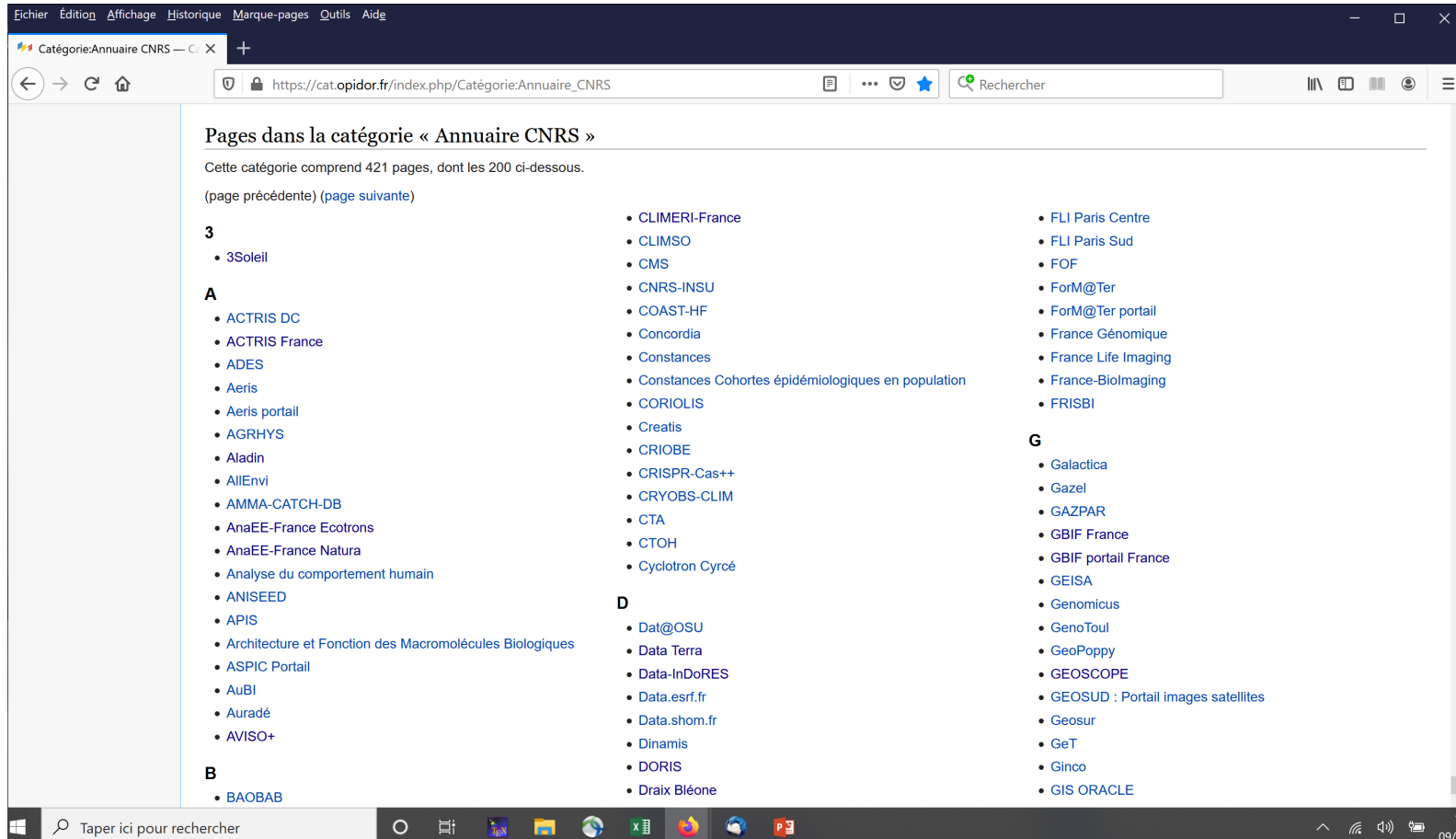
Les principes de base

- L'objectif: Identifier les dépôts et les services de données dont le CNRS est responsable ou auxquels il participe
- La confiance est au cœur de la Science Ouverte
- Obtenir une cartographie de qualité
 - ⇒ Gros travail sur le contenu et sa validation
- Retenir pour l'Annuaire les entités ayant une certaine qualité et une certaine pérennité
 - ⇒ Entités « revendiquées » par les Instituts
- Le cadre est celui de la Science Ouverte
 - ⇒ Entités qui partagent des données ou au moins des métadonnées sur les données
- Prendre en compte les liens « hiérarchiques » entre les entités
 - ⇒ Y compris par exemple la contribution du CNRS aux ESFRI

Où nous en sommes

- Construction du contenu dans Cat OPIDoR en collaboration étroite avec les Instituts - un très gros travail
 - Le 9 novembre: 421 entrées identifiées dans l'Annuaire attachées à au moins un Institut
 - En cours: Pour chaque entrée, une fiche détaillée
- En cours: Finalisation de la validation du contenu par les Instituts
 - Validation de la liste des entités attachées à l'Institut
 - Identification et discussion de nombreux cas particuliers
- En cours: Vérification avec le point de vue des Infrastructures de Recherche
- Fait: Ajout des informations nécessaires à l'identification des éléments de l'Annuaire dans Cat OPIDoR

L'Annuaire « à plat »



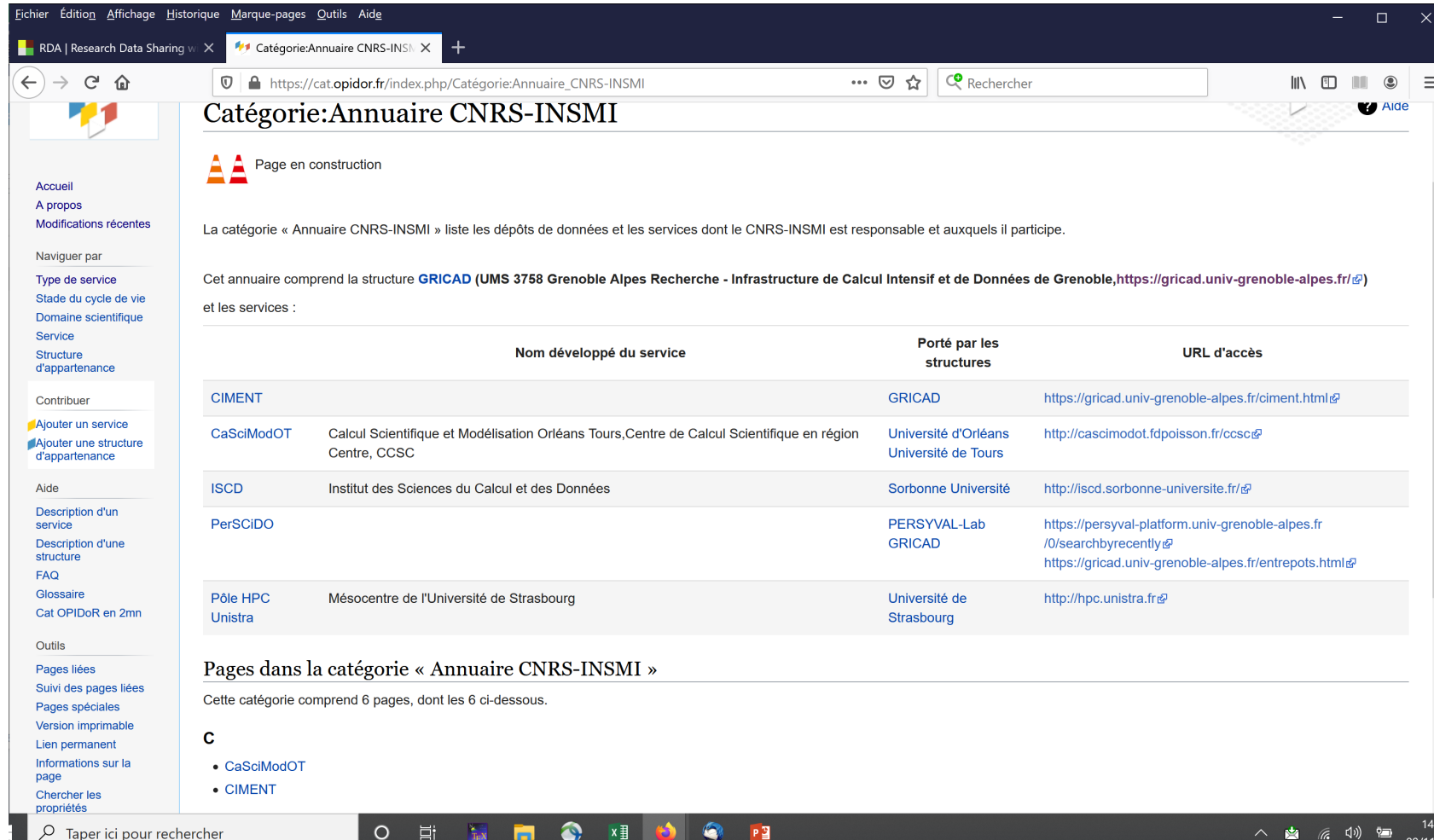
The screenshot shows a web browser window with the following elements:

- Browser Title Bar:** "Catégorie:Annuaire CNRS — C. X"
- Address Bar:** "https://cat.opidor.fr/index.php/Catégorie:Annuaire_CNRS"
- Search Bar:** "Rechercher"
- Page Content:**
 - Section Header:** "Pages dans la catégorie « Annuaire CNRS »"
 - Text:** "Cette catégorie comprend 421 pages, dont les 200 ci-dessous." and "(page précédente) (page suivante)"
 - Alphabetical List:**
 - 3**
 - 3Soleil
 - A**
 - ACTRIS DC
 - ACTRIS France
 - ADES
 - Aeris
 - Aeris portail
 - AGRHYS
 - Aladin
 - AllEnvi
 - AMMA-CATCH-DB
 - AnaEE-France Ecoltrons
 - AnaEE-France Natura
 - Analyse du comportement humain
 - ANISEED
 - APIS
 - Architecture et Fonction des Macromolécules Biologiques
 - ASPIC Portail
 - AuBI
 - Auradé
 - AVISO+
 - B**
 - BAOBAB
 - C**
 - CLIMERI-France
 - CLIMSO
 - CMS
 - CNRS-INSU
 - COAST-HF
 - Concordia
 - Constances
 - Constances Cohortes épidémiologiques en population
 - CORIOLIS
 - Creatis
 - CRIOBE
 - CRISPR-Cas++
 - CRYOBS-CLIM
 - CTA
 - CTOH
 - Cyclotron Cyréc
 - D**
 - Dat@OSU
 - Data Terra
 - Data-InDoRES
 - Data.esrf.fr
 - Data.shom.fr
 - Dinamis
 - DORIS
 - Draix Bléone
 - F**
 - FLI Paris Centre
 - FLI Paris Sud
 - FOF
 - ForM@Ter
 - ForM@Ter portail
 - France Génomique
 - France Life Imaging
 - France-Bioluminescence
 - FRISBI
 - G**
 - Galactica
 - Gazel
 - GAZPAR
 - GBIF France
 - GBIF portail France
 - GEISA
 - Genomicus
 - GenoToul
 - GeoPoppy
 - GEOSCOPE
 - GEOSUD : Portail images satellites
 - Geosur
 - GeT
 - Ginco
 - GIS ORACLE

Ce qui est en cours

- La réflexion sur l'évolution de Cat OPIDoR
 - Besoins en visualisation de l'information
 - Passer d'un annuaire « à plat » à un annuaire qui permette de visualiser des informations sur les entités
 - Identification des besoins du CNRS, des Instituts et des chercheurs
 - Réflexion sur les aspects sémantiques
 - Typologie des entités
 - Relations entre entités (« appartient à », et autres)
- Nous avons rassemblé beaucoup d'information utile sur ces sujets avec la réflexion sur le contenu de l'Annuaire et les discussions avec les Instituts

Un exemple de présentation possible: l'Annuaire de l'INSMI



The screenshot shows a web browser window displaying the 'Annuaire CNRS-INSMI' page. The page is under construction, as indicated by the 'Page en construction' notice. The main content area lists the services provided by the CNRS-INSMI, including the GRICAD structure and several specific services. A table lists these services with their names, descriptions, the structures they are supported by, and their access URLs. Below the table, there is a section for 'Pages dans la catégorie < Annuaire CNRS-INSMI >' which lists 6 pages, including 'CaSciModOT' and 'CIMENT'.

Page en construction

La catégorie « Annuaire CNRS-INSMI » liste les dépôts de données et les services dont le CNRS-INSMI est responsable et auxquels il participe.

Cet annuaire comprend la structure **GRICAD (UMS 3758 Grenoble Alpes Recherche - Infrastructure de Calcul Intensif et de Données de Grenoble)**, <https://gricad.univ-grenoble-alpes.fr/>) et les services :

Nom développé du service		Porté par les structures	URL d'accès
CIMENT		GRICAD	https://gricad.univ-grenoble-alpes.fr/ciment.html
CaSciModOT	Calcul Scientifique et Modélisation Orléans Tours, Centre de Calcul Scientifique en région Centre, CCSC	Université d'Orléans Université de Tours	http://cascimodot.fdpiisson.fr/ccsc
ISCD	Institut des Sciences du Calcul et des Données	Sorbonne Université	http://iscd.sorbonne-universite.fr/
PerSCIDO		PERSYVAL-Lab GRICAD	https://persyval-platform.univ-grenoble-alpes.fr/0/searchbyrecently https://gricad.univ-grenoble-alpes.fr/entrepots.html
Pôle HPC Unistra	Mésocentre de l'Université de Strasbourg	Université de Strasbourg	http://hpc.unistra.fr

Pages dans la catégorie « Annuaire CNRS-INSMI »

Cette catégorie comprend 6 pages, dont les 6 ci-dessous.

C

- [CaSciModOT](#)
- [CIMENT](#)

Exemple d'extraction d'information: les Zones Atelier dans l'Annuaire INEE

The screenshot shows a web browser window with the URL https://cat.opidor.fr/index.php/Catégorie:Annuaire_CNRS-INEE. The page content includes a sidebar with navigation links and a main area with a list of 'Zones Atelier'.

Structure de rattachement [+]
Label ou certification [-]
 Zone Atelier

Tutelles [+]
Est en relation avec [+]
Impliqué dans un projet international [+]

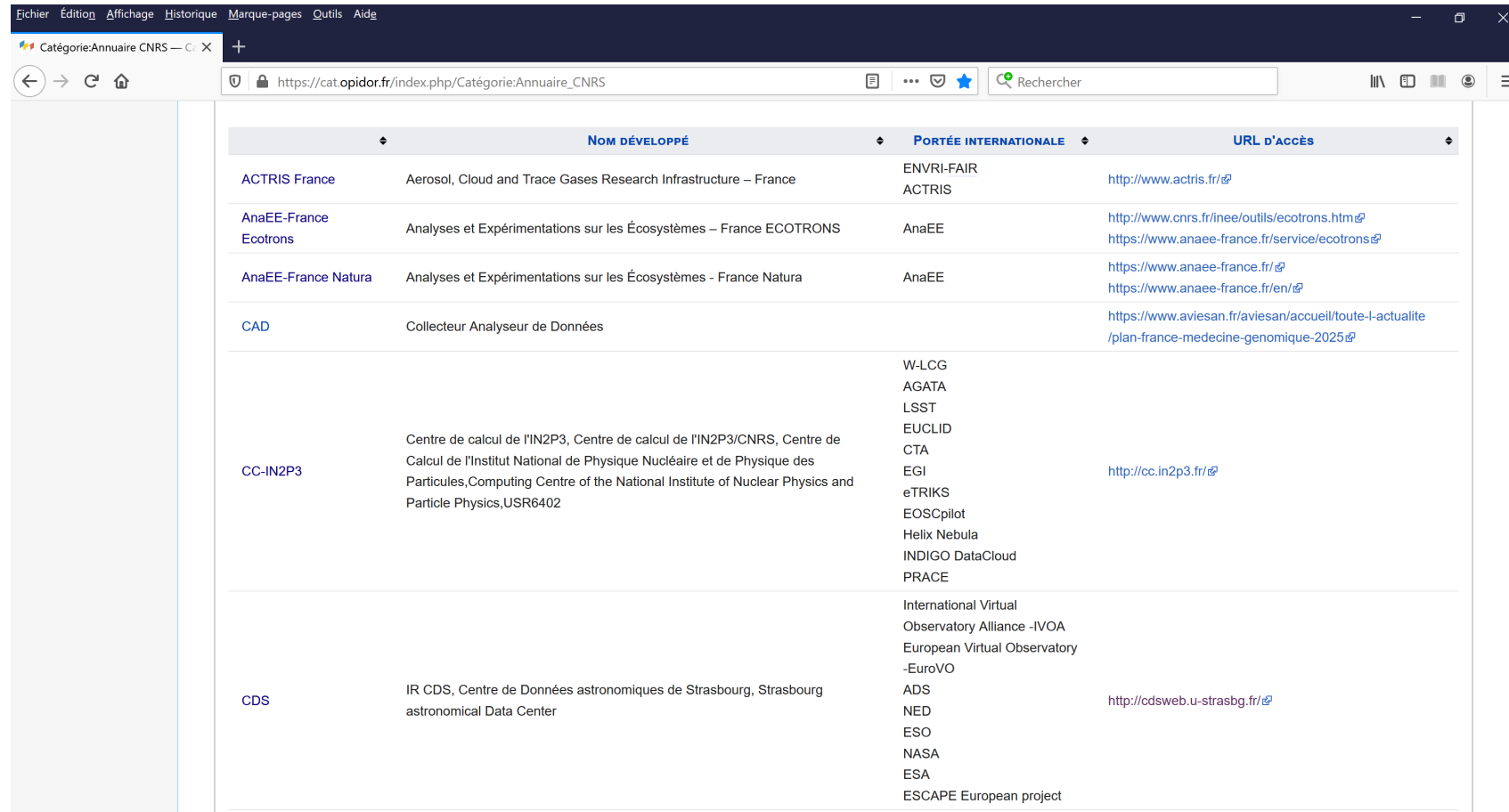
	AUTRES FORMES DE NOM	URL d'Accès
ZA Alpes	Zone Atelier Alpes, ZAA	http://www.za-alpes.org/
ZA Antarctique	ZA Antarctique et Subantarctique,ZATA,Zone Atelier Antarctique et Terres Australes	https://zaantarctique.org/
ZA Arc Jurassien	Zone atelier Arc Jurassien, ZAAJ	https://zaaj.univ-fcomte.fr/
ZA Armorique	Zone Atelier Armorique, ZARR	https://osur.univ-rennes1.fr/ZA-Armorique/
ZA Bassin du Rhône	Zone Atelier Bassin du Rhône, ZABR	http://www.graie.org/zabr/index.htm
ZA Brest Iroise	Zone Atelier Brest Iroise, ZABrI	http://www-ium.univ-brest.fr/zabri/fr/
ZA Environnementale Urbaine	Zone Atelier Environnementale Urbaine Strasbourg, ZAEU	https://zaeu-strasbourg.eu/
ZA Hwange	Zone Atelier Hwange,ZAHWG	http://www.za-hwange.cnrs.fr
ZA Loire	Zone atelier Loire, ZAL	http://www.za-loire.org
ZA Moselle	Zone Atelier Moselle, ZAM, Zone Atelier Bassin de la Moselle	https://zam.univ-lorraine.fr/
ZA PYGAR	Zone Atelier PYGAR, Zone Atelier Pyrénées-Garonne,	https://pygar.omp.eu/
ZA Plaine et Val de Sèvre	Zone Atelier Plaine et Val de Sèvre, Zone Atelier Plaine & Val de Sèvre, ZAPVS	http://www.za.plainevalsevre.cnrs.fr/
ZA Seine	Zone atelier Seine, ZAS	https://www.za-seine.fr
ZA Territoires Uranifères	Zone Atelier Territoires Uranifères dans l'Arc Hercynien, ZATU	https://zatu.org

Exemple d'extraction d'information: les Services Nationaux d'Observation dans l'Annuaire INSU

ci pour rechercher

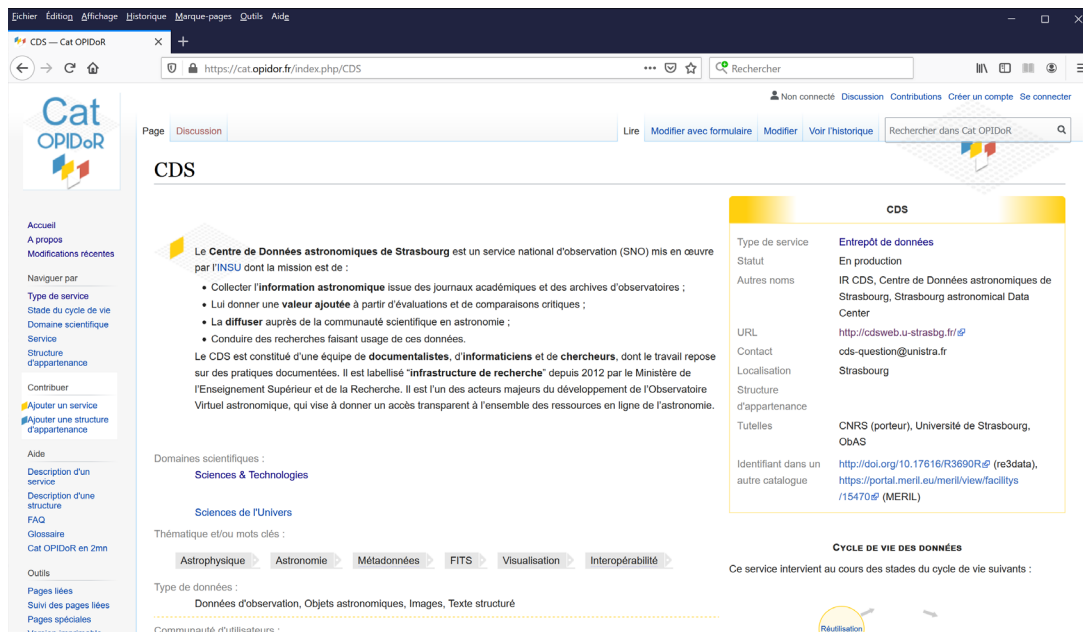
	NOM DÉVELOPPÉ	URL d'accès	PORTÉ PAR LES STRUCTURES	Sous-DISCIPLINE
3Soleil	Surveillance sol du Soleil et du rayonnement solaire (3Soleil)	https://observations-solaires.obspm.fr/	Obs. Paris	Sciences de l'Univers
AMMA-CATCH-DB	Observatoire éco-hydro-météo en Afrique de l'Ouest, Observatoire hydrologique, écologique et météorologique de long terme en Afrique de l'Ouest	http://bd.amma-catch.org/	OSUG	Sciences du Système Terre
APIS	Auroral Planetary Imaging and Spectroscopy	http://apis.obspm.fr/	Obs. Paris	Sciences de l'Univers
ASPIC Portail	Archives Spectrophotométriques publiques intégrées au CeSAM, Archive of Spectrophotometry Publicly available In Cesam	http://cesam.lam.fr/aspic/	Institut Pythéas	Sciences de l'Univers
BCMT	Bureau Central de Magnétisme Terrestre	http://www.bcmt.fr/	IPGP EOST	Sciences du Système Terre
BCSF-RENASS	Bureau Central Sismologique Français- Réseau National de surveillance sismologue	http://www.franceseisme.fr/donnees/autres.html	EOST	Sciences du Système Terre
CDPP	Centre de données pour la physiques des plasmas naturels du système solaire, Plasma Physics Data Center	http://www.cdpp.eu/		Sciences de l'Univers
CDS	IR CDS, Centre de Données astronomiques de Strasbourg, Strasbourg astronomical Data Center	http://cdsweb.u-strasbg.fr/		Sciences de l'Univers
CESBIO	Centre d'étude spatiale de la biosphère	https://www.cesbio.cnrs.fr/donnees/	OMP	Sciences du Système

Exemple d'extraction d'information: Les Infrastructures de la Feuille de Route Nationale auxquelles le CNRS contribue en termes d'entrepôts ou services de données



NOM DÉVELOPPÉ	PORTÉE INTERNATIONALE	URL d'accès
ACTRIS France	Aerosol, Cloud and Trace Gases Research Infrastructure – France	ENVRI-FAIR ACTRIS http://www.actris.fr/
AnaEE-France Ecotrons	Analyses et Expérimentations sur les Écosystèmes – France ECOTRONS	AnaEE http://www.cnrs.fr/inee/outils/ecotrons.htm https://www.anaee-france.fr/service/ecotrons
AnaEE-France Natura	Analyses et Expérimentations sur les Écosystèmes - France Natura	AnaEE https://www.anaee-france.fr/ https://www.anaee-france.fr/en/
CAD	Collecteur Analyseur de Données	https://www.aviesan.fr/aviesan/accueil/toute-l-actualite/plan-france-medecine-genomique-2025
CC-IN2P3	Centre de calcul de l'IN2P3, Centre de calcul de l'IN2P3/CNRS, Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules, Computing Centre of the National Institute of Nuclear Physics and Particle Physics, USR6402	W-LCG AGATA LSST EUCLID CTA EGI http://cc.in2p3.fr/ eTRIKS EOSCpilot Helix Nebula INDIGO DataCloud PRACE
CDS	IR CDS, Centre de Données astronomiques de Strasbourg, Strasbourg astronomical Data Center	International Virtual Observatory Alliance -IVOA European Virtual Observatory -EuroVO ADS NED http://cdsweb.u-strasbg.fr/ ESO NASA ESA ESCAPE European project

Exemple de fiche: le Centre de données astronomiques de Strasbourg



CDS

Le **Centre de Données astronomiques de Strasbourg** est un service national d'observation (SNO) mis en œuvre par l'**IINSU** dont la mission est de :

- Collecter l'**information astronomique** issue des journaux académiques et des archives d'observatoires ;
- Lui donner une **valeur ajoutée** à partir d'évaluations et de comparaisons critiques ;
- La **diffuser** auprès de la communauté scientifique en astronomie ;
- Conduire des recherches faisant usage de ces données.

Le CDS est constitué d'une équipe de **documentalistes, d'informaticiens et de chercheurs**, dont le travail repose sur des pratiques documentées. Il est labellisé "**infrastructure de recherche**" depuis 2012 par le Ministère de l'Enseignement Supérieur et de la Recherche. Il est l'un des acteurs majeurs du développement de l'Observatoire Virtuel astronomique, qui vise à donner un accès transparent à l'ensemble des ressources en ligne de l'astronomie.

Domaines scientifiques :
[Sciences & Technologies](#)
[Sciences de l'Univers](#)

Thématique et/ou mots clés :
[Astrophysique](#) [Astronomie](#) [Métadonnées](#) [FITS](#) [Visualisation](#) [Interopérabilité](#)

Type de données :
Données d'observation, Objets astronomiques, Images, Texte structuré

CDS	
Type de service	Entrepôt de données
Statut	En production
Autres noms	IR CDS, Centre de Données astronomiques de Strasbourg, Strasbourg astronomical Data Center
URL	http://cdsweb.u-strasbg.fr/
Contact	ods-question@unistra.fr
Localisation	Strasbourg
Structure d'appartenance	
Tutelles	CNRS (porteur), Université de Strasbourg, ObAS
Identifiant dans un autre catalogue	http://doi.org/10.17616/R3690R#(re3data) , https://portal.meril.eu/meril/view/facilities/15470#(MERIL)

CYCLE DE VIE DES DONNÉES
Ce service intervient au cours des stades du cycle de vie suivants :

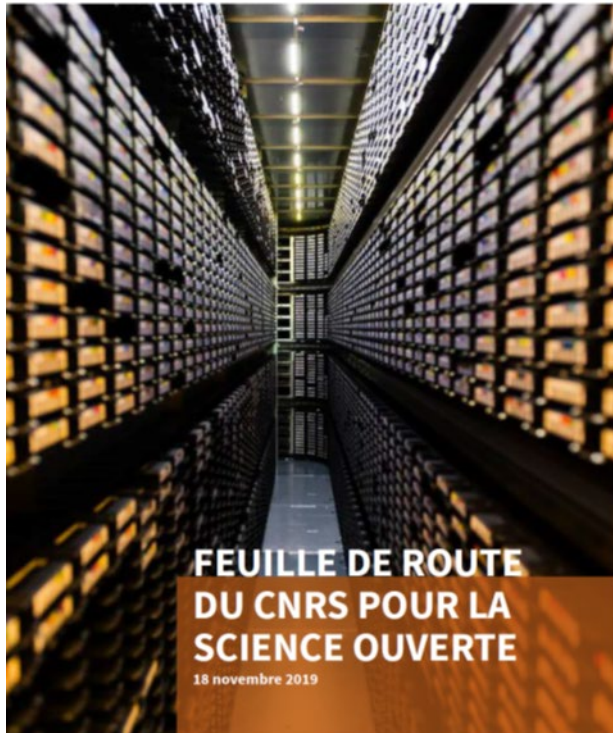
La fiche contient aussi des informations entre autres sur

- les conditions d'usage,
- les certifications et labels,
- la portée internationale,
- les domaines d'intervention dans le cycle de vie des données,
- une carte pour les entités localisées

Conclusions

- C'est un « work in progress », mais nous avons beaucoup progressé
 - ⇒ Clarification du périmètre de l'Annuaire
 - ⇒ Les Instituts ont fait le bilan des dépôts et des services de données qu'ils mettent à disposition, une de leurs contributions à la Science Ouverte
 - ⇒ Entrées de nombreuses données dans Cat OPIDoR, y compris pour des informations qui ne seront pas directement dans l'Annuaire
- L'exercice n'est pas facile mais il est utile
- Les membres du GT apportent la connaissance du paysage et le lien pour reboucler avec leur institut si besoin

Pour la Feuille de Route Science Ouverte



Plusieurs contributions à l'Axé Données

- Construction de l'Annuaire (Action 5)
- Point aussi sur les services qui diffusent les données des infrastructures de recherche (éléments pour l'Action 3)
- Grande diversité des approches (organisation, culture) selon les instituts
 - ⇒ Nous apprenons tou.te.s quelque chose dans l'exercice sur la culture de la gestion de la donnée (Action 1)!

La période est difficile, le travail à accomplir est complexe
mais nous avançons de façon significative

Merci à tous ceux qui sont impliqués et qui font le maximum,
dans les Instituts et à l'INIST et à la DIST