



# PLAN DONNEES DE LA RECHERCHE DU CNRS

SYLVIE ROUSSET, DIRECTRICE DE LA DDOR

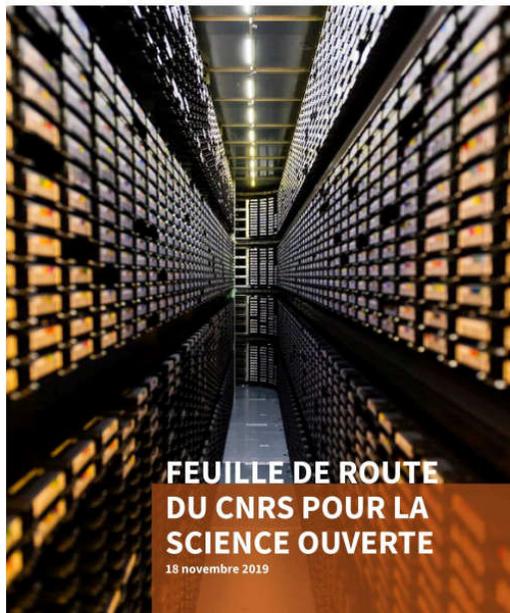
DIRECTION DES DONNEES OUVERTES DE LA RECHERCHE

[Sylvie.rousset@cnrs.fr](mailto:Sylvie.rousset@cnrs.fr)

ANF RNBM – 4 et 5 Octobre 2021 – CIRM - Marseille

# ACCÉLÉRATION DE LA STRATÉGIE DU CNRS DEPUIS 2018

## SCIENCE OUVERTE ET DONNÉES DE LA RECHERCHE



Novembre 2019  
Français ou English version



Novembre 2020  
Français ou English version

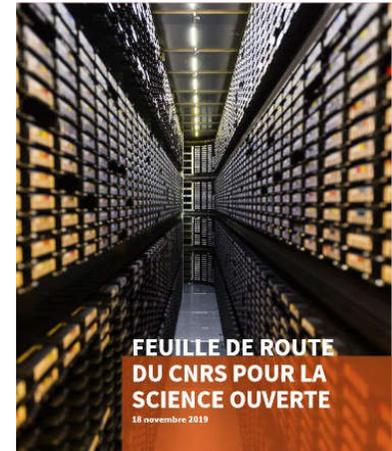


Pour en savoir plus –recueil  
des interventions de la journée  
du 8 Octobre 2019

<https://www.science-ouverte.cnrs.fr/>

# Feuille de route du CNRS pour la Science Ouverte

- 100% de publications en accès ouvert et ré-utilisables
- Des données de la recherche FAIR-isées
- Développer et promouvoir les outils pour l'analyse et la fouille des textes et des données
- L'évaluation individuelle des chercheurs
- Site WEB <https://www.science-ouverte.cnrs.fr>
  - Feuille de route Science Ouverte du CNRS
  - CNRS Roadmap Open Science



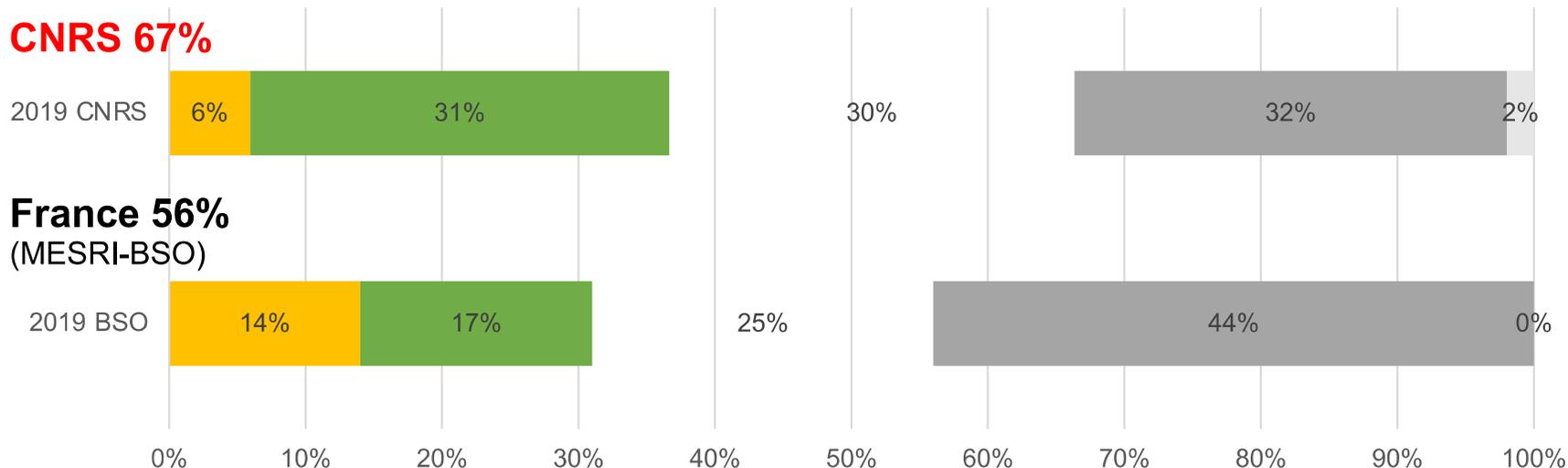


**OBJECTIF  
100% DE PUBLICATIONS EN  
ACCES OUVERT**

Quelques actions phares 2019 et  
2020

# BAROMETRE DE LA SCIENCE OUVERTE AU CNRS

## CNRS 67%



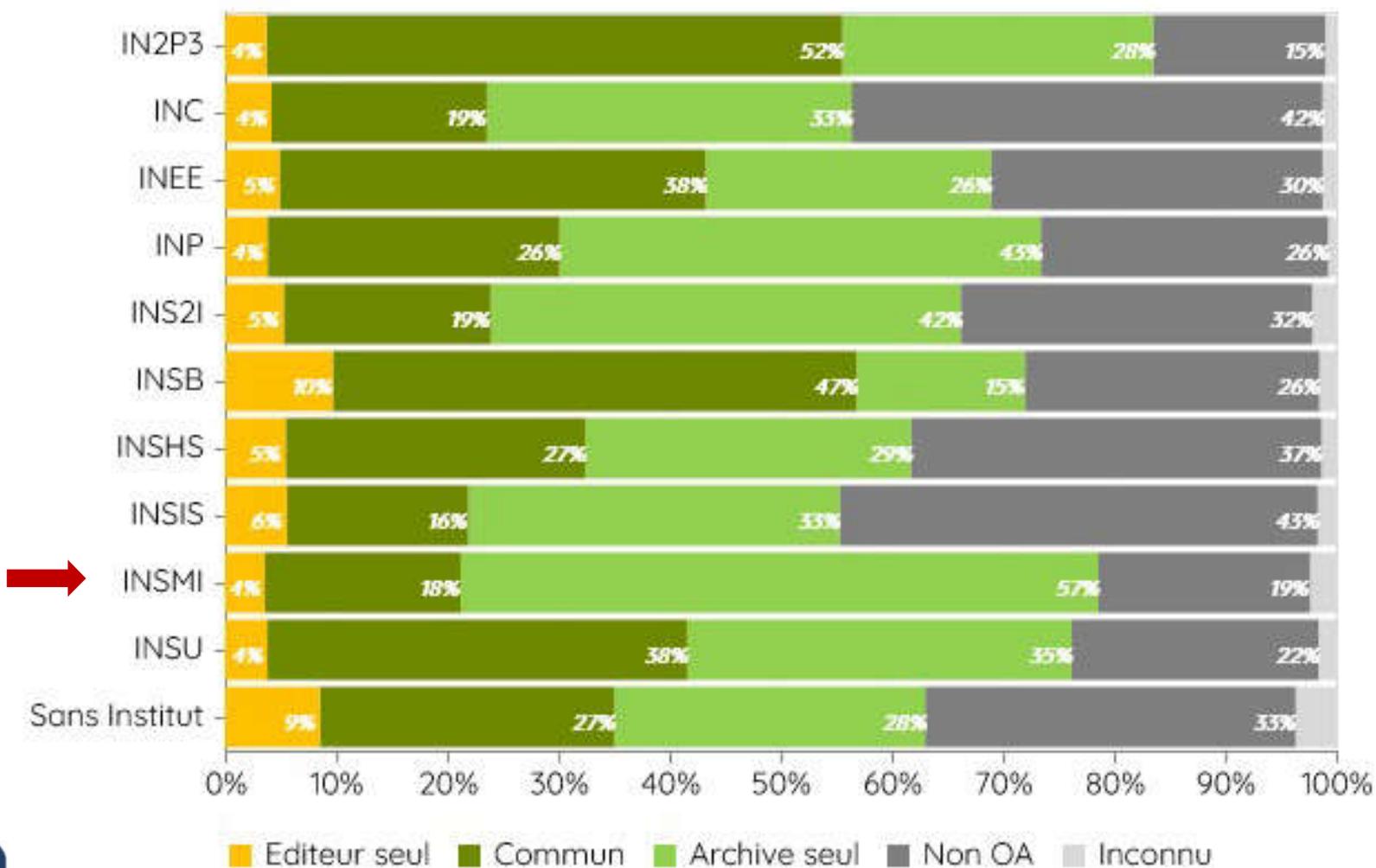
## France 56% (MESRI-BSO)

- AO chez l'éditeur seulement
- AO sur les archives ouvertes seulement
- AO commun archives et chez l'éditeur
- Accès fermé
- Inconnu

- ✓ **PUBLICATIONS 2019 DU CNRS EN ACCES OUVERT : 67%**
- ✓ **PROGRESSION DE L'ACCES OUVERT DE 49% À 67% EN 3 ANS**
- ✓ **VOIE POUR L' ACCES OUVERT : 91% VIA OPEN ARCHIVES**

Travail commun DDOR - INIST - DAPP & BSO (MESRI)

# PUBLICATIONS 2019 EN ACCÈS OUVERT SELON LES INSTITUTS



**INSMI**  
79% d'accès ouvert

# ARCHIVES OUVERTES HAL PILOTÉES PAR LE CCSD A LYON

The screenshot displays the HAL website interface. At the top, there is a navigation bar with the CCSD logo, language options (fr, en), and a user profile for Bénédicte Kuntziger. Below this is a dark blue header with the HAL logo and the URL archives-ouvertes.fr. A central illustration shows a stylized figure holding an open book, surrounded by floating mathematical symbols and geometric shapes. A secondary navigation bar includes links for Home, Submit, Browse, Search, Documentation, and My space. The main content area is divided into several sections: a 'Home' breadcrumb, a title 'The open archive HAL', a brief description of HAL as an open archive for scholarly documents, and two sections for authors and readers. On the right side, there are three summary boxes: 'SEARCH' with a search input field, 'NUMBER OF RESOURCES' showing 1 779 279, and 'NUMBER OF FULLTEXTS' showing 584 387. Below these are sections for 'NEWS' with two recent articles listed.

CCSD HAL Episclences.org Sciencesconf.org Support fr en Bénédicte Kuntziger

**HAL**  
archives-ouvertes.fr

Home Submit Browse Search Documentation My space

Home

### The open archive HAL

HAL is an open archive where authors can deposit scholarly documents from all academic fields.

**For the attention of the authors**

- The deposit of the fulltext should be made in agreement with the co-authors and in the respect for the policy of the publishers.
- The deposit is subject of a control, HAL reserves the right to refuse Items that do not meet the criteria of the archive.
- Any deposit is definitive, no withdrawals will be made after the on-line posting of the publication.
- Text files in pdf format or image files are sent to CINES for long-term archiving.

**For the attention of the readers**

- In a context of electronic distribution, each author keeps their intellectual property rights.

**LATEST SUBMISSIONS IN HAL**

- Mohamed El Boudani. Suivi non destructif de l'altération de marbres par méthodes acoustiques. Mécanique des matériaux [physics.class-ph]. Université de Cergy Pontoise, 2017. Français. (NNT : 2017CERG0947). (tel-02118867).
- Daichi Kashino, John Silverman, David Sanders, Jeyhan Kartaltepe, Emanuele Daddi, et al.. The FMOS-COSMOS Survey of Star-forming Galaxies at  $z \sim 1.6$ . VI. Redshift and Emission-line Catalog and Basic Properties of Star-forming Galaxies. Astrophysical Journal Supplement, American Astronomical Society, 2019, 241 (1), pp.10. (10.3847/1538-4365/ab06c4). (hal-02110856)
- G. Misson, J. Rousseau, Pierre Sagout. Investigating the stability of the LRM RGK through an extended von Neumann

**SEARCH**

Termes de recherche (\* pour tous)

**NUMBER OF RESOURCES**

1 779 279

**NUMBER OF FULLTEXTS**

584 387

**NEWS**

**PUBLISHERS ALLOWING AAM ('POSTPRINT') POSTING TO REPOSITORIES WITHOUT EMBARGO** (5/3/19)

**INVESTIGATING YOUR PUBLISHING AND OPEN ACCESS PRACTICES** (12/2/19)

## **POLITIQUE FORTE EN FAVEUR DES ARCHIVES OUVERTES HAL AVEC UNE OBLIGATION DE DÉPÔT DANS HAL**

- **Politique d'incitation à déposer les publications scientifiques dans HAL**
- **Message du DGDS le 29 avril 2019 pour que le CRAC ne soit alimenté que par HAL en 2019 (Ribac en 2020)**
- **Message du DGDS le 29 avril 2020 pour que la campagne CRAC 2020 s'applique au texte intégral Interopérabilité et compatibilité avec ArXiv et BioRxiv**
- **Moyens exceptionnels en 2019 (650 k€) pour l'accompagnement du dépôt dans HAL**
- **Création du portail HAL-CNRS pour accueillir la production CNRS, ouvert le 18 nov. 2019 à l'URL, <https://hal-cnrs.archives-ouvertes.fr/>**
- **Refonte de la Gouvernance du CCSD avec toutes les tutelles (INRIA, INRAE, CNRS) et le CoSO et nouveau modèle économique de soutien via les établissements partenaires du CCSD qui ont un portail HAL (135 établissements : Ecoles, Universités, Organismes)**

# DÉPÔTS SELON LA DATE DE PUBLICATION DE 2002 À 2021 VISIBLES DANS LE PORTAIL HAL-CNRS

Nombre et type de dépôts CNRS CRAC

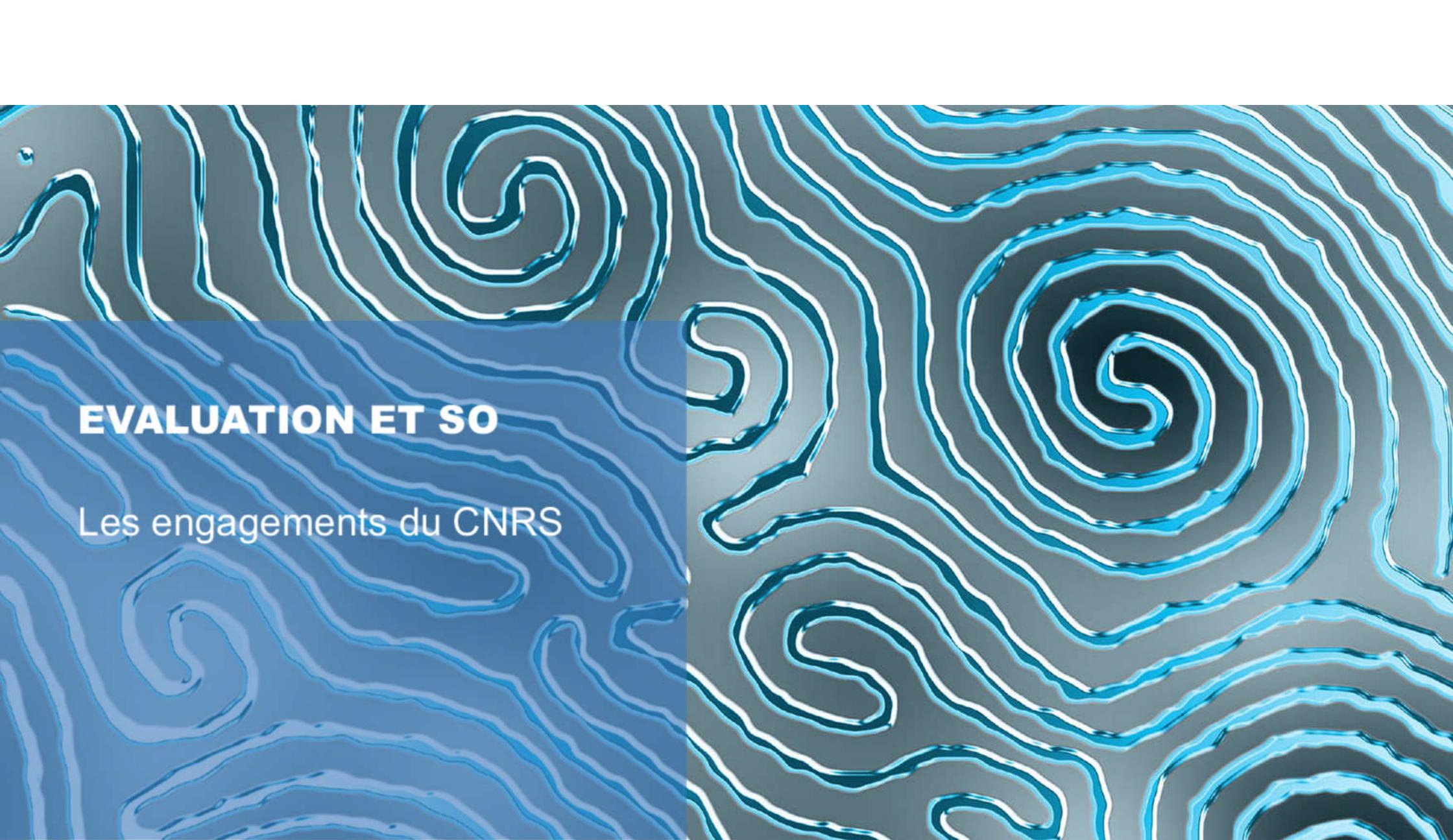


Pour la première fois en 2020, le nombre de dépôts avec le texte intégral est plus important que le nombre des dépôts de notices uniquement.

## **POLITIQUE DE SOUTIEN A LA BIBLIODIVERSITÉ ET EN PARTICULIER À L'ÉDITION SCIENTIFIQUE EN ACCÈS OUVERT**

*Développer la bibliodiversité, créer de nouvelles revues en accès ouvert ou reprendre les revues existantes pour les publier en accès ouvert, en partenariat avec des institutions françaises et européennes et de façon adaptée à chaque discipline.*

- **Abondement du fond FNSO en 2019 : 500 k€**
- **Soutien en 2019 pour la transformation des Comptes Rendus de l'Académie des Sciences qui ont quitté Elsevier pour être publié sur la plateforme du Centre Mersenne (1ers numéros début 2020) : 150 k€**
- **Soutien en 2019 à Open Edition : 200 k€**
  
- **Soutien SHS en 2020 pour l'attribution de DOI et l'édition diamant : 25 k€**
- **Soutien au Centre Mersenne : 42 k€**
  
- **Soutien en 2020 à d'autres modèles vertueux :**
  - le modèle Subscribe to Open (S2O) via des revues de la SMAI : 116 k€**
  - Les PCI (Peer Community In) : 70 k€**
  - La base de données zBlatt (en discussion)**



## EVALUATION ET SO

Les engagements du CNRS

# **DORA** (San Francisco Declaration on Research Assessment)

(Signée par le CNRS en juillet 2018)

- 18 **engagements**. Pour chacun des acteurs et des actrices de l'évaluation. Trois extraits:
- **Pour les établissements :**
  - (4) *Afficher explicitement les critères utilisés dans les décisions de recrutement, de titularisation et de promotion, en soulignant clairement, surtout pour les chercheurs débutants, que le contenu scientifique d'un article est beaucoup plus important que les indicateurs de publication ou l'image de marque de la revue (...)*
  - (5) *Aux fins de l'évaluation de la recherche, tenir compte de la valeur et de l'impact de tous les résultats de travaux de recherche (...)*
- **Pour les chercheurs et chercheuses (évaluateurs - évaluatrices) :**
  - (15) *Lors d'une participation à des commissions exerçant des décisions de financement, d'embauche, de titularisation ou de promotion, produire des évaluations fondées sur le contenu scientifique plutôt qu'en fonction des indicateurs de publication.*



# UNE ÉVALUATION PLUS QUALITATIVE ET MOINS QUANTITATIVE

Généralisation de 4 principes communs pour les évaluations individuelles

- 1. Ce sont les résultats eux-mêmes qui doivent être évalués**, et non pas le fait qu'ils aient pu être publiés dans une revue prestigieuse ou autre media réputé
- 2. Pour chacune des productions citées dans les dossiers d'évaluation, les chercheurs doivent en expliquer la portée, l'impact, et la contribution personnelle qu'ils y ont apporté.**
- 3. Tous les types de production doivent pouvoir être des éléments de l'évaluation.**
- 4. Toutes les productions citées dans les dossiers d'évaluation doivent être accessibles dans HAL ou éventuellement dans une autre archive ouverte\* :**

\*Trois exceptions à cette règle sont recevables :

1. Les résultats trop récents peuvent être sous embargo. Auquel cas ils doivent quand même avoir été déposés dans HAL, avec une durée d'embargo ne dépassant pas ceux prévus par la loi (6 mois en STM, et 12 mois pour les SHS). Ils sont alors fournis par un lien privé dans HAL (ou alors dans le dossier).
2. Pour les recrutements, cette règle ne peut pas être absolue pour les candidats exerçant à l'étranger dans des institutions étrangères ou internationales, ou des institutions privées.
3. Le type de production peut ne pas être accepté dans HAL.



## MODIFICATION DES DOSSIERS DE CANDIDATURE POUR LES CHERCHEURS ET CHERCHEUSES CANDIDAT-E-S

- La notice de titres et travaux n'est plus demandée
- Ni la liste complète des publications
- Mais d'expliquer ses activités, pas seulement de production de résultats scientifiques (formation, innovation, management etc.)
- Et de ne citer qu'un nombre limité de « produits », non réduits aux seules publications ou conférences (données, logiciels, cours, etc.) en appui aux explications
- Productions à rendre disponibles en accès ouvert dans HAL ou éventuellement dans une autre archive vertueuse, sauf exceptions
- Faire valoir ses engagements pour la Science Ouverte le cas échéant



## **LE PLAN DONNEES DE LA RECHERCHE DU CNRS**

Lancé en novembre 2020  
Création d'une direction des  
données ouvertes de la recherche  
DDOR

## I. ENQUETE « SCIENCE OUVERTE 2020 »

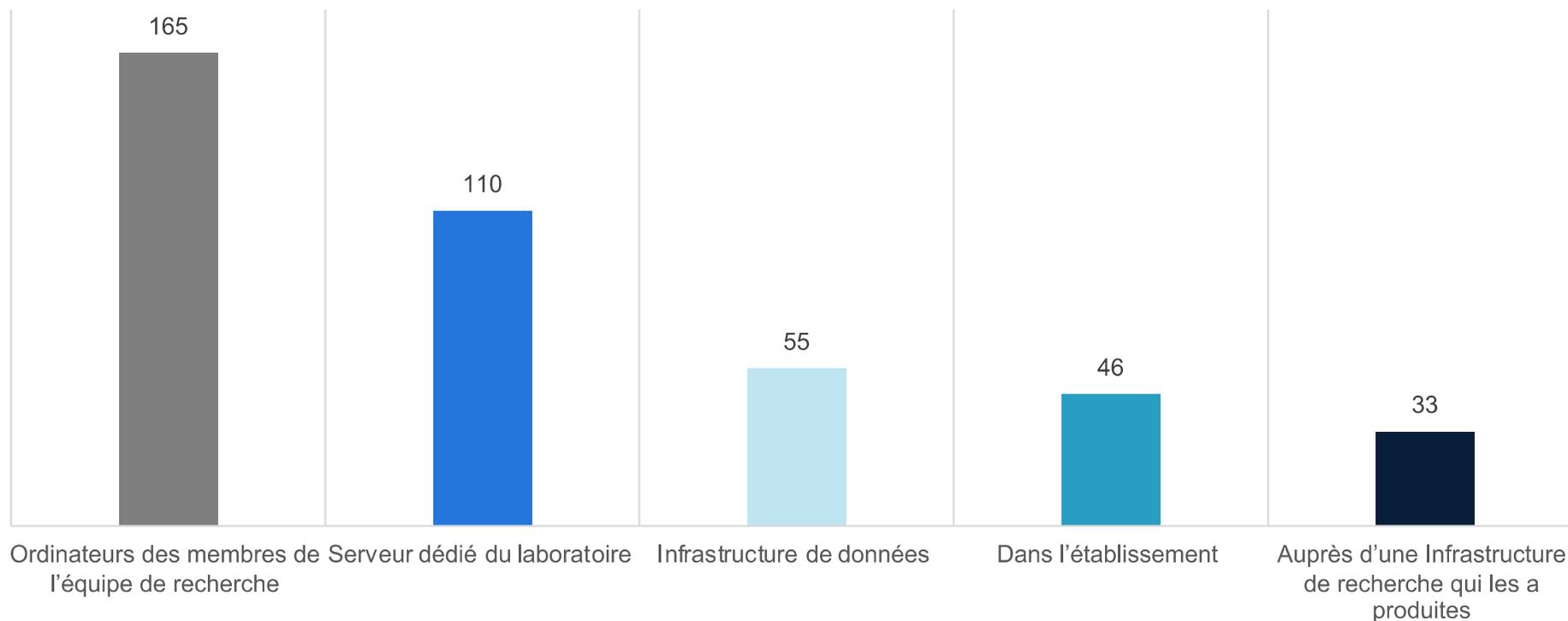
- Lancée suite aux réunions avec les DU des sept délégations visitées par la DIST

	Institut de rattachement	Nombre d'unités sollicitées	Nombre d'unités ayant répondu
• DR01 - Île-de-France Villejuif			
• DR02 - Paris-Centre	▪ INSHS	139	57
• DR04 - Gif-sur-Yvette	▪ INP	47	24
• DR12 - Provence et Corse	▪ INSIS	58	21
• DR14 - Occitanie Ouest	▪ INSB	106	21
• DR17 – Bretagne et pays de Loire	▪ INEE	40	16
• DR20 - Côte d'Azur	▪ INSU	45	10
	▪ INC	70	21
▪ <u>191 unités</u> ont répondu	▪ INSMI	25	13
▪ Taux de réponse moyen : <b>34 %</b>	▪ INS2I	22	7
▪ 19 questions posées	▪ IN2P3	9	1
	<b>Total</b>	<b>561</b>	<b>191</b>

## II. OÙ SONT PRODUITES LES DONNÉES DE LA RECHERCHE ?

Uniquement dans le laboratoire	112	<b>58,6 %</b>
Dans le laboratoire et auprès des infrastructures de recherche	65	<b>34 %</b>
Uniquement auprès des infrastructures de recherche	6	<b>3,1 %</b>
Ne produisent pas de données de la recherche	5	<b>2,6 %</b>
Autre	3	<b>1,6 %</b>
Total	<b>191</b>	<b>100 %</b>

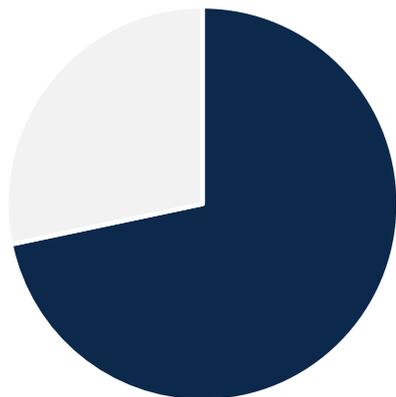
### III. OÙ SONT STOCKÉES LES DONNÉES DE LA RECHERCHE ?



Attention : Réponse à choix multiples.

Le nombre d'unités répondantes est de 191. La somme totale des réponses ici est supérieure à 191

## IV. PRATIQUES DE PARTAGE DES DONNÉES DE LA RECHERCHE



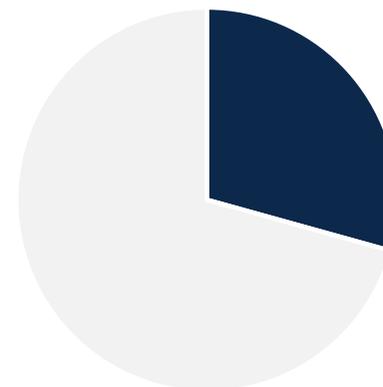
**71,7 %**

**PROJET COLLABORATIF**



**60,2 %**

**DONNÉES LIÉES  
À UNE PUBLICATION**



**29,3 %**

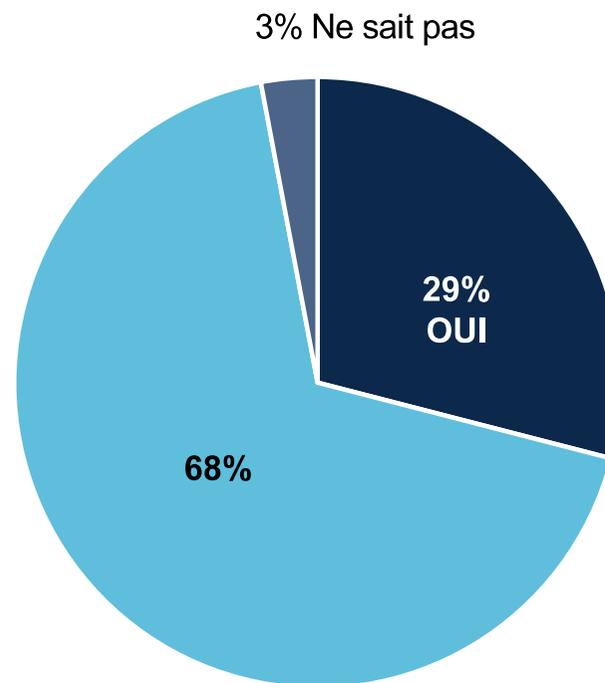
**RÉUTILISATION OUVERTE  
EN GÉNÉRAL**

Attention : Réponse à choix multiples.

Le nombre d'unités répondantes est sur un total de 191.

## V. GESTION DES DONNÉES DE LA RECHERCHE

- **Ressources humaines pour la gestion des données de la recherche**
- Avez-vous des personnels pour assurer les fonctions associées à la gestion de vos données de recherche ?



## LES DONNÉES DE LA RECHERCHE

Une donnée doit être ouverte ou protégée. L'ouverture des données s'entend selon l'expression « **ouvert autant que possible, fermé autant que nécessaire** ».

**Toutes les données de la recherche n'ont pas vocation à être ouvertes ou divulguées.** Il existe des exceptions évidentes telles que les données spécifiques à caractère confidentiel, que cela soit du fait de leur caractère personnel, pour des raisons de concurrence industrielle ou pour des intérêts fondamentaux ou réglementaires des États.

La décision d'ouverture ou de protection des données de la recherche doit être prise avec les services compétents du CNRS :

- les Services Partenariat Valorisation pour la propriété intellectuelle,
- la Délégation à la protection des données pour les données à caractère personnel et de
- la Direction de la sûreté pour les questions relatives à la souveraineté.

Les actions proposées ici traitent des données ayant vocation à être ouvertes, mais pas que !

# LES DONNEES DE LA RECHERCHE

## Définition des données de la recherche :

“information, en particulier des faits et des nombres, collectés afin d’être utilisés à des fins de raisonnement, discussion ou calculs.”

## Trois types de données de la recherche :

- Les données brutes
- Les données “curées”
- Les données reliées aux publications scientifiques

## Tous les formats des données de la recherche sont concernés :

y compris les textes, les documents, les logiciels, les algorithmes et les protocoles.

## Ecosystème des données de la recherche :

Infrastructures numériques, Big Data, services et principes FAIR, Science Ouverte.

## MOTIVATIONS

Rendre la recherche plus efficace et non redondante (pas de duplication inutile)

Assurer l'intégrité scientifique (reproductibilité et validation des résultats)

Etre en capacité de réutiliser les données même sans en être à l'origine

Croiser les données pour favoriser de nouvelles analyses, voire faire émerger de nouvelles thématiques

Satisfaire le cadre légal d'ouverture des données a priori : « Ouvert autant que possible, fermé autant que nécessaire » (loi pour une république numérique 2016)

Mutualiser et rationaliser les infrastructures informatiques, les moyens RH et identifier les nouveaux métiers (datastewarship ...)

## **PLAN D' ACTIONS : VALORISER LES DONNEES**

Promouvoir une « culture de la donnée », ainsi que les infrastructures, les outils et services, en incluant la question des moyens (humains et financiers) et de la formation

Travailler avec les instituts en prenant en compte les spécificités disciplinaires et l'existant, et en s'appuyant sur une gouvernance transverse des données

Soutenir les communautés scientifiques pour définir les éléments spécifiques de la gestion de leurs données à FAIRiser

Encourager le dépôt des données dans des entrepôts thématiques en priorité

Valoriser les données en lien avec les publications, et les « datapapers »

## PLAN D' ACTIONS : OUTILS ET SERVICES

Proposer **un annuaire** pour identifier les outils et services existants (Cat-OPIDoR)

Accompagner la réalisation des **Plan de Gestion des Données** (PGD- OPIDoR)

Aider à **la certification des dispositifs** de prise en charge des données (ex. CoreTrustSeal)

S'assurer que toutes les données à FAIRiser trouvent **un entrepôt des données** (métadonnées + jeux de données) et **un espace physique de stockage**

Fournir des **identifiants pérennes** pour les jeux de données (DOI – Datacite – INIST)

Encourager la FAIRisation des données avec les **Cahiers de Laboratoire Electroniques**

Travailler sur les **portails de référencement** pour accéder aux données disponibles

## **PLAN D' ACTIONS : RH ET FORMATION**

Recenser les nouveaux métiers liés au traitement des données

Modifier les référentiels de postes IT (Referens) pour inclure les nouvelles compétences

Nourrir les référentiels d'évaluation pour inclure l'implication dans la gestion et le partage des données

Former les personnels à la culture des données et aux nouveaux outils et services par la formation permanente notamment (très forte demande)

Faire connaître les succès du partage des données aux communautés réticentes

# Une direction fonctionnelle des données ouvertes de la recherche : la DDOR

Continuum du Calcul aux Publications : données massives, longue traîne des données, infrastructures matérielles et logicielles, services aux utilisateurs, référencement de la documentation.

Rapprocher deux communautés IST et Calcul : des grandes différences y compris le vocabulaire mais les sujets sont liés et on ne peut pas les dissocier (même problématique au MESRI, rapprochement entre les infrastructures numériques - Laurent Cruzet et la science ouverte - Marin Dacos).

Interlocuteur pour les relations avec les partenaires de l'ESR et au niveau européen

## Direction DDOR à l'intérieur de la DGDS avec le périmètre suivant:

- Calcul (calcul haute performance, grilles, "cloud", données massives, représentation du CNRS dans les "objets" nationaux : INFRANUM (Infrastructures numériques, DGRI), GENCI, RENATER, ... )
- Données (Stockage, FAIRisation, Annuaire, DMP, DOI, Aide à la certification des entrepôts, représentation CoSO, RDA, OPenAIRE)
- Publications (Négociations et politique documentaire, Archives ouvertes, Nouvelles plateforme d'édition, Pilotage GIS FNSO, CoSO)

## Instance de décision au sein de la DDOR:

Comité de pilotage avec la direction de la DDOR et les 10 représentants des instituts (niveau DAS)



## A VENIR

- 19 et 20 Octobre 2021 ANF Science Ouverte à Meudon
- 15 novembre 2021 Journée Science Ouverte et Evaluation du CNRS à Paris
- 4 et 5 février 2022 Open Science European Conference, Paris, PFUE



**MERCI DE VOTRE ATTENTION**

<https://www.science-ouverte.cnrs.fr/>

<https://www.cnrs.fr/fr/cnrsinfo/science-ouverte>