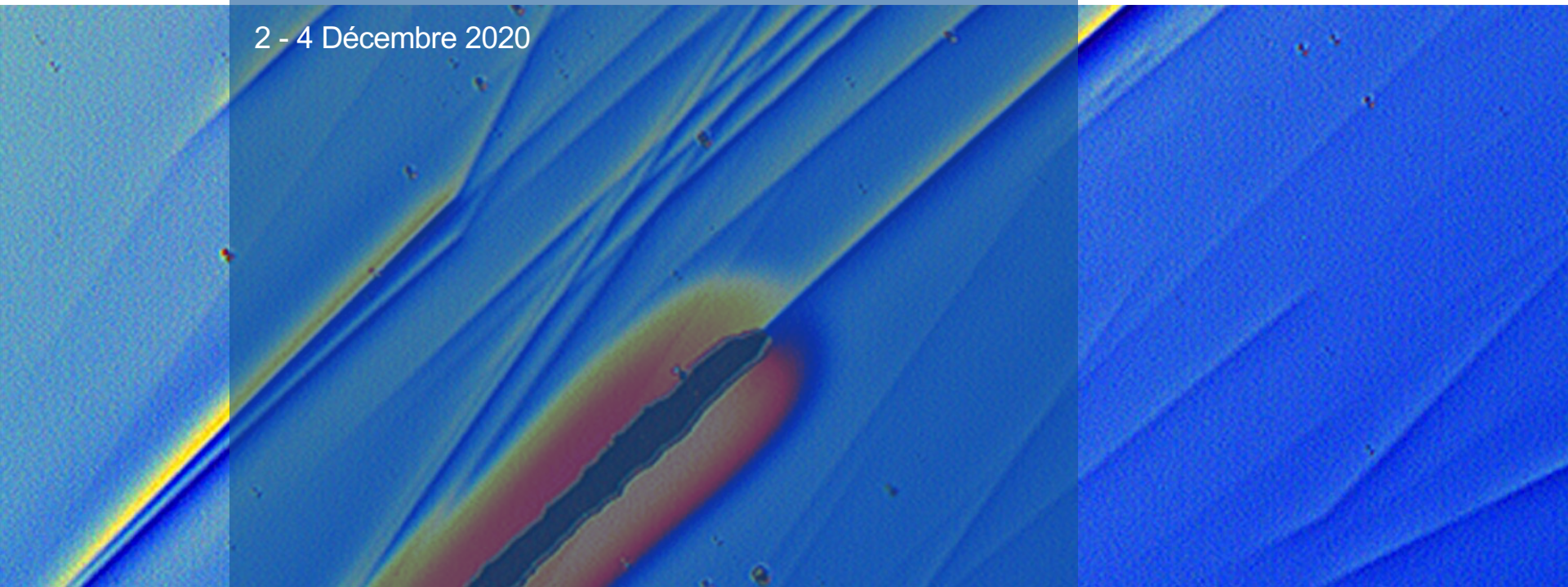




Plan Données de la Recherche – Direction des Données Ouvertes de la Recherche (DDOR)

Journées Calcul - Données

2 - 4 Décembre 2020



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 ...)

LES DONNÉES DE LA RECHERCHE DANS LE COP 2019 – 2023

- **Action 25** : Développer une culture de la gestion et du partage des données chez tous les acteurs du cycle de vie de la donnée basée sur la mise en oeuvre des principes FAIR (Faciles à trouver, Accessibles, Interopérables et Réutilisables).
- **Action 30** : Se doter en 2020 d'un plan global « Données » en s'appuyant sur une gouvernance spécifique dédiée à la donnée.

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
- 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 DONNÉES 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.”*

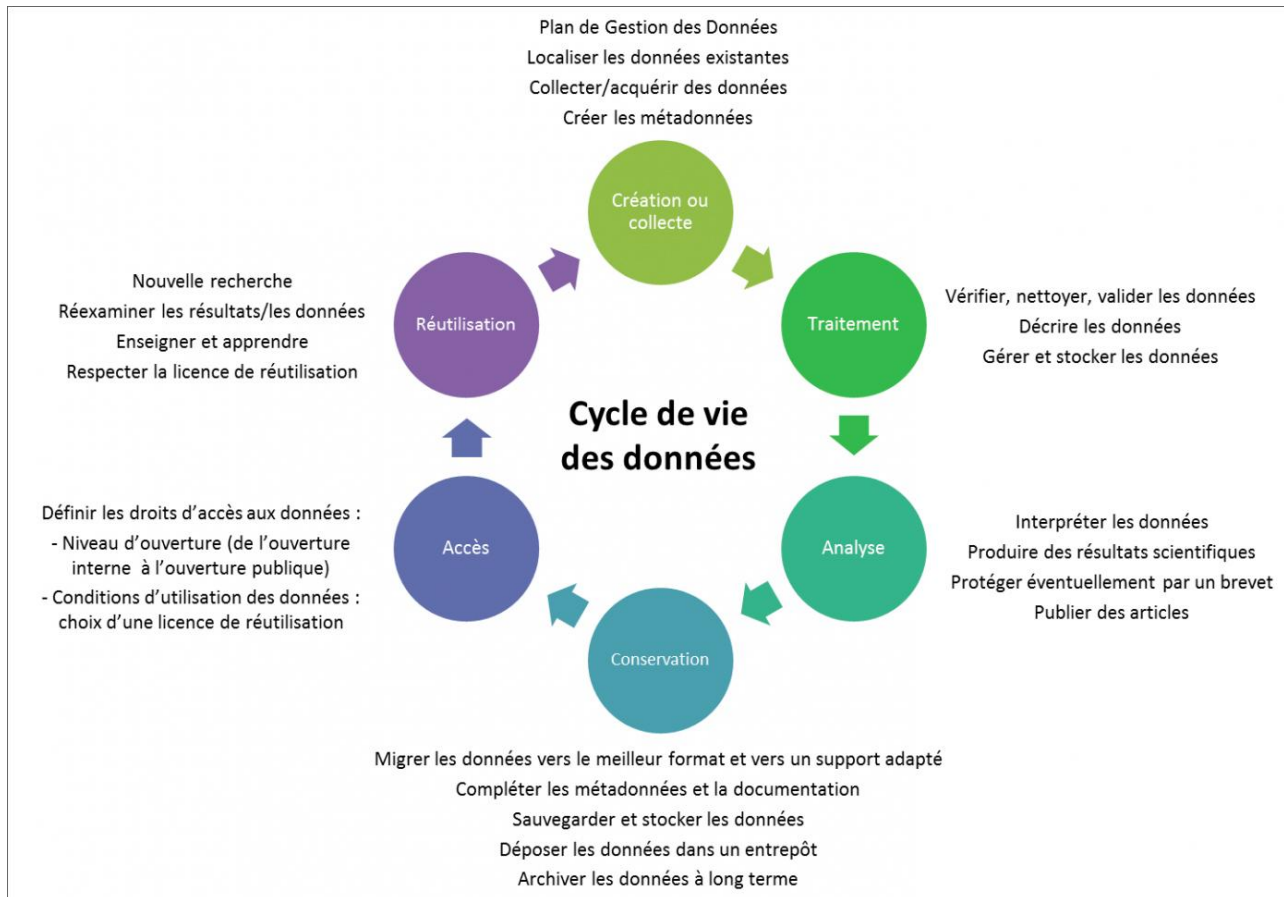
Tous les formats des données de la recherche sont concernés y compris les textes, les documents, les logiciels, les algorithmes et les protocoles.

Trois types de données de la recherche :

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

Ecosystème des données de la recherche : Infrastructures numériques, Big Data, services et principes FAIR, Science Ouverte.

LES DONNÉES DE LA RECHERCHE



POURQUOI UN PLAN DONNEES MAINTENANT ?

Constat de la mission MICADO dans le Livre Blanc 2018 de promouvoir une « culture de la donnée » et de doter le CNRS d'une stratégie des données

http://www.cocin.cnrs.fr/IMG/pdf/livre_blanc_donnees_2018.pdf

Accélération des actions pour développer la Science Ouverte :

- Plan National Science Ouverte 2018 et demande d'un administrateur des données dans les établissements
- Feuille de Route Science Ouverte CNRS 2019 avec un volet données de la recherche

Mise en place d'une gouvernance pour EOSC (European open science cloud)

- au niveau international, AISBL avec le CNRS comme membre (2020),
- au niveau français un « EOSC français » (DGRI)

Enjeu scientifique fort pour la gestion et le partage des données au CNRS

- Voir les motivations scientifiques, juridiques et de moyens

https://www.cnrs.fr/sites/default/files/pdf/Plaqueette_PlanDDOR_Nov20.pdf

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

4 **GOVERNANCE DES DONNEES : constat sur l'existant**

Lieux de pilotage et de discussion actuels au CNRS

- MICADO : comité MICADO et comité directeur COCIN
- EOSC : GT EOSC, EOSC France et Europe
- Direction Information Scientifique et Technique
 - Corr_IST
 - Corr_Données
- INIST : axe “valorisation des données” DMP OPIDoR, CAT OPIDoR, DOI DATA CITE, formation en ligne DORANUM
- CCSD : projet PIA 3 HALliance incluant l'accès aux données liées aux publications
- CC-IN2P3 : projet PIA3 FITS d'une infrastructure distribuée de stockage, de traitement et de mise à disposition de données pour répondre aux besoins des infrastructures de recherche (TGIR/IR) en fédérant CC-IN2P3 et IDRIS
- MESRI : Réflexion sur la faisabilité d'un entrepôt générique des données pour la longue traîne des données ; Proposition d'une gouvernance pour un “EOSC français”

La Direction des Données Ouverte de la Recherche

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

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)



**Merci de votre
attention**