



# Accès ouvert aux données scientifiques

défi 14

Strasbourg, 20 & 21 jan 2020

# DÉFI 14 – CONTEXTE



## PROBLÉMATIQUES ET ENJEUX



- Les données sont au cœur de la démarche scientifique de l'INSU
  - longue tradition de distribution ouverte des données via les SNOs et IR
  - une politique volontariste de soutien au partage de données. Axe transverse fort.
- Implication bien en amont du développement des principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) et Science Ouverte



## VERROUS SCIENTIFIQUES



### 5 grands thèmes:

- Contexte FAIR: national, européen, international
- Gestion des données et services: Observatoire Virtuel en AA, Infrastructures de Recherche, OSUs et SNOs
- Les données “hors observatoire”: données de longue traîne, attachées aux publications, hors SNOs, non-numériques, etc
- Harmonisation des modèles de données et des métadonnées: description des données, formats
- Certification des entrepôts de données et de services

Document complet avec toutes les recommandations sur le site CORE, défi 14 + vidéo des 2 jours

# DÉFI 14 – RECOMMANDATIONS



## ORGANISATION



- [R1] Pilotage par les besoins scientifiques
- [R2] Pilotage national
- [R3] Financement
- [R4] Soutien aux infrastructures thématiques
- [R5] Les Infrastructures de Recherche dans un paysage en évolution
- [R6] Mise en priorité
- [R7] Analyse du paysage
- [R8] Ressources humaines et gestion des personnels/reconnaissance des compétences
- [R9] Évolution de la culture du milieu scientifique, organismes, laboratoires et personnels
- [R10] Aspects techniques, standards
- [R11] Certification des entrepôts et services de données

## DES POINTS RÉCURRENTS

- **La science est au cœur** du processus. Elle doit rester le moteur des évolutions et des choix stratégiques
  - Il est essentiel de soutenir les centres de données et de services, de fournir aux communautés les outils adéquats définis par les besoins scientifiques
- **État des lieux :**
  - Services FAIR et entrepôts de données: faire l'inventaire de l'existant.
    - Participer à l'inventaire CNRS des services
    - Faire l'inventaire des autres services utilisés par les chercheurs
    - Documenter le besoin d'un entrepôt de données de longue traîne CNRS
  - Diversité des métiers liés aux données et de leur spécificités: faire l'inventaire des profils et des compétences

**Des recommandations complètes et opposables  
à prendre en compte dans chaque rubrique**

# DÉFI 14 – RECOMMANDATIONS



## DES POINTS RÉCURRENTS

- Notions de **coûts** (humains, temps, financiers...) / définitions de **priorités**
  - FAIRisation, partage interdisciplinaire, certification
  - besoins de financements pour la gestion des données pour des durées supérieures à celles des projets → overheads?
  - reconnaissance temps dévolu à la bonne gestion des données
- Evolution rapide de l'écosystème international/européen/national du partage des données scientifiques et de la Science Ouverte: poursuivre les réflexions sur son évolution et l'articulation des différents niveaux, y compris EOSC



## INSTRUMENTATION, MODÈLES ET DONNÉES



- Pilotage national par l'INSU et structuration en OSUs: un atout.
- Rôle majeur des IR au niveau national, européen et international. Deux exemples phares :  
- Gestion et harmonisation des données :
  - continuer à soutenir l'AS OV et renforcer le rôle de l'IR Data Terra (InterPôle).
  - Renforcer les échanges techniques et de bonnes pratiques.
- Documenter le besoin national d'un entrepôt de données longue traîne
- La certification des entrepôts de données et services implique aussi la gouvernance.
  - L'auto-évaluation est utile en soi.
  - Soutien via ANRs et RDA France/CoSO Données





# DÉFI 14 – RECOMMANDATIONS



## COMPÉTENCES ET INTERDISCIPLINARITÉ



- Complémentarité de plusieurs métiers et compétences différentes (chercheurs, informaticiens, documentalistes, juristes, ...) .
- Nouveaux métiers sur les données ("Data analyst", "Data architect", "Data steward", "Data scientist", ...) à définir.
- Identifier, reconnaître et valoriser les métiers liés à la gestion des données. En faire la liste, en identifiant les compétences.

## LIENS AVEC LES ODD ET LA SOCIÉTÉ



- Reconnaissance de l'impact des technologies de l'information dans les émissions de gaz à effet de serre. Existence de labels/certifications reconnaissant la conception responsable de service numérique.
- L'impact de la science ouverte vers la société, c'est aussi de donner accès aux données à tous, ce qui ouvre des opportunités égales pour les chercheurs de tous les pays, y compris les pays à revenus faibles et intermédiaires.

## LIENS AVEC LES INDUSTRIES



- Les accords ou MoU sont critiques. Besoins forts de clarifications et de faciliter l'accès aux informations juridiques.
- Proposition d'une concertation entre les services juridiques des établissements de manière à déboucher à une charte ou un guide synthétique.

## COMMUNICATION ET DIFFUSION



- Sciences participatives : les données qu'elles fournissent sont très diverses et de qualité très inégales -> Nécessité de faire des choix.
  - Chiffrer les coûts globaux de ce que pourrait représenter la gestion de ces données. Identifier des exemples.
- Diffusion auprès du grand public des méthodes scientifiques sur de vraies données et avec des outils utilisés et développés par les communautés INSU.