



CeDONA : Centre de Données pour l'Observation en Nouvelle Aquitaine
Pascal CALVAT - OASU Bordeaux



OASU – PESSAC - 11 mars 2019



Plan

- Présentation de CeDONA
- Services informatiques de CeDONA
- FAIR data, RDA, plan de gestion de données
- CeDONA au service des pôles du système terre et de l'IVOA



Présentation de CeDONA

- CEDONA (Centre de DOnnées en Nouvelle Aquitaine) est le centre de données de l'OASU à Bordeaux
- Missions :
 - Renforcer la visibilité nationale et internationale des projets portés dans les équipes grâce au partage des technologies et des compétences.
 - Assurer la cohésion des services dans un contexte de recherche multidisciplinaire
 - Répondre aux standards internationaux de mise à disposition des données et procurer des services à valeur ajoutée sur les données issues des équipes de recherche de l'OASU.





Unités Constitutives OASU depuis 2003

- UMR LAB (~65p, 5 équipes) :
 - Astrophysique : Astrochimie, Origine Systèmes Planétaires, Exoplanètes, Planètes, Formation stellaire, Métrologie de l'espace, Astrodynamique
 - SNO : AA-ANO₁ (IVS), AA-ANO₂ (JUICE, ALMA, SKA), AA-ANO₄ (GAIA), AA-ANO₅ (KIDA)
- UMR EPOC (~150p, 7 équipes) :
 - Ecotoxicologie et Chimie de l'environnement, Océanographie Côtière, Géosciences Marines
 - SNO: OA-ANO₂-SOMLIT, SIC-(OA)-ANO₃-DynaLIT, AO-ANO₁-PHOTONS/AERONET
 - Autres SO : MAGEST (qualité eau estuaire) → une partie dans DynaLIT, Molluscan Eye (qualité des milieux aquatiques), SOERE (AllEnvi – Trait de Cote)
- UMS POREA (~20p, 1 équipe) :
 - Services Techniques
 - Services InformatiqueS

Unités Constitutives de IRSTEA et Université de La Rochelle



- EABX (IRSTEA, Bx, ~68p) : Ecosystème aquatique, Equipements (bateaux, plateformes de chimie, etc..) (3 équipes)
- ETBX (IRSTEA, Bx, ~52p) : Environnement, Territoire, Infrastructure (2 équipes)
 - Disciplines: sciences économiques, sciences politiques, agronomie, sociologie, hydraulique, modélisation, statistiques
- LIENSs (La Rochelle, ~150p) : Etude du littoral
 - biologie, écologie, géophysique, chimie, sciences humaines (4 équipes)
 - SNO: OA-ANO₂-SOMLIT, SIC-(OA)-ANO₃-DynaLIT, SIC-(OA)-ANO₃-SONEL



CeDONA : les services informatiques

- Développement web via campagnes d'appel à projet
- Déploiement de sites web métier (SNO, SO, ...)
- Aide à l'utilisation de services nationaux de diffusion de données (pôle de données?)
- Backup et/ou réplication de données
- Versionning de codes (gitlab)
- Calcul en local et sur le meso-centre MCIA (10 000 coeurs)
- Accompagnement rédaction de plan de gestion de données



Architecture informatique pour les sites web

- Cluster PROMOX VE 5.1 :
 - 2 machines (DELL Power Edge 740, 24 cœurs, 128 Go de RAM) pour les sites web (SNO)
 - 1 machine pour le backup des VM (100 To)
 - Virtual Machine (VM) en Cent OS 7
- Pas de file system partagé (à mettre en place?)
- Mysql en maitre/esclave pour éviter perte de données sql en cas de crash du maitre

Appel à projets



- Une ou deux campagnes annuelles d'appel à projets
- Evaluation par le comité scientifique de l'OASU
- Un pourcentage de FTE est attribué aux différents projets par le comité
- 1 développeur permanent et 1 prestataire sont ensuite affectés aux projets
- Spécifications fonctionnelles obligatoires pour réaliser les développements

CeDONA promeut la production de données FAIR

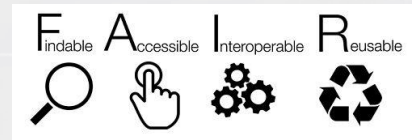


Dans le contexte de la science ouverte, la notion de FAIR data recouvre les manières de construire, stocker, présenter ou publier des données.
Le FAIR DATA est imposé dans les projets financés par l'Europe.



Plus d'infos voir :

<https://www.force11.org/group/fairgroup/fairprinciples>

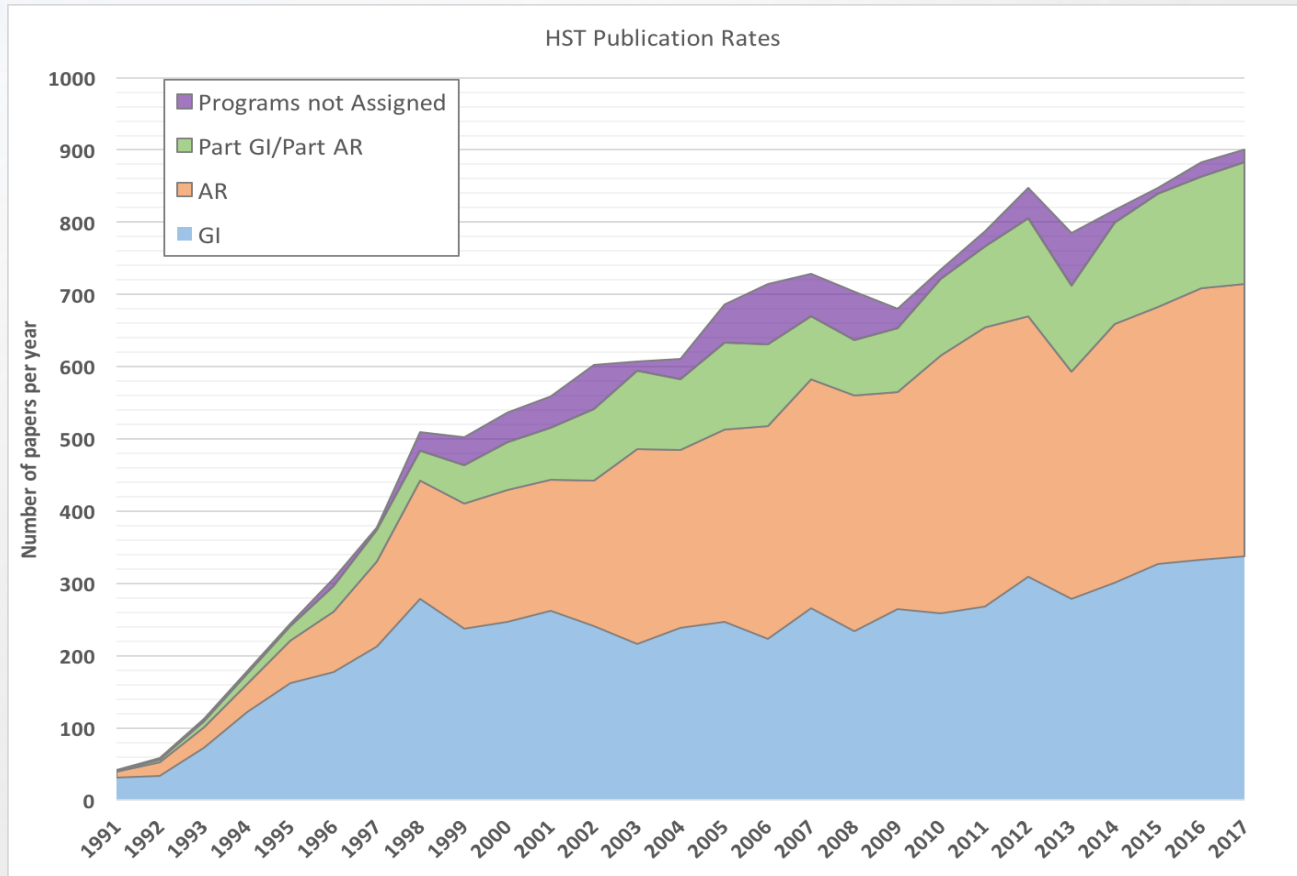


Les principes du FAIR DATA

- FAIR DATA principes :
 - **Findable**: Facile à trouver par l'utilisation *d'identifiant global unique et pérenne*, des *métadonnées riches* et le signalement dans un *catalogue*
 - **Accessible** : Accessible via *des protocoles de communication ouverts*, *accès permanent aux métadonnées*
 - **Interoperable** : Interoperable par l'application de *standards pour les formats* de représentation des données et métadonnées, utilisation de *vocabulaires et ontologies*
 - **Reusable** : Réutilisable grâce à *une description riche*, l'utilisation de *standards communautaires* et de *licences*



Réutilisation des données Hubble (HST)



Crédit: Robert J. Hanisch, Space Telescope Science Institute

Guest Investigator (GI) paper: At least one author from the paper was an investigator on the original proposal that obtained the data.

Archival paper: No overlap between the paper authors and investigators on the GI proposal that obtained the data.

Part GI Part Archive: Multiple programs were used. Overlap of authors with some programs and not with others.

Programs not Assigned: Insufficient information given in the paper to determine the programs of the data used.

60 % des publications réalisées grâce à la réutilisation des données



CeDONA et la *Research Data Alliance*

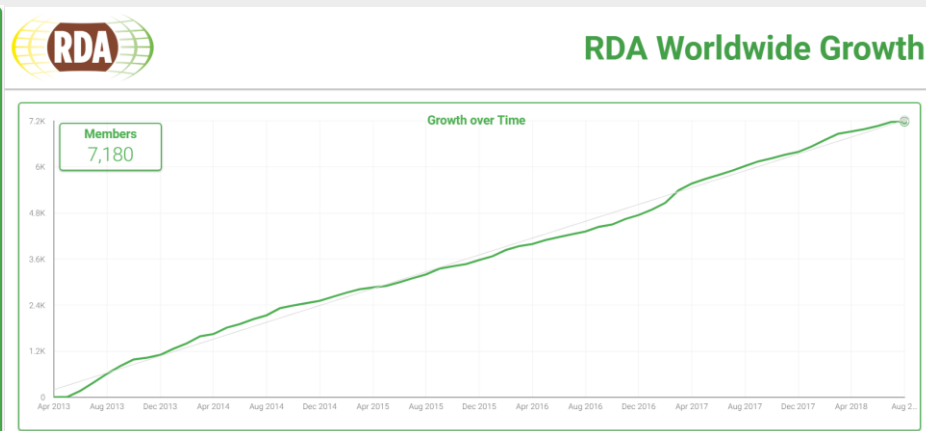
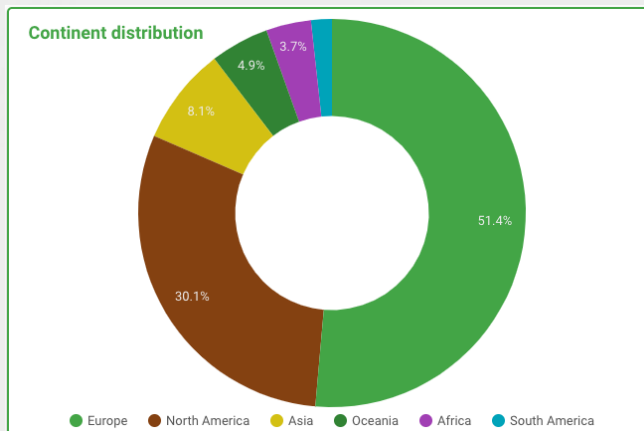


CeDONA sera attentif aux standards et recommandations issus de la RDA (Research Data Alliance)

La RDA est une organisation internationale dont les membres développent des activités et des infrastructures communes pour réduire les obstacles au partage et aux échanges de données.

RDA :

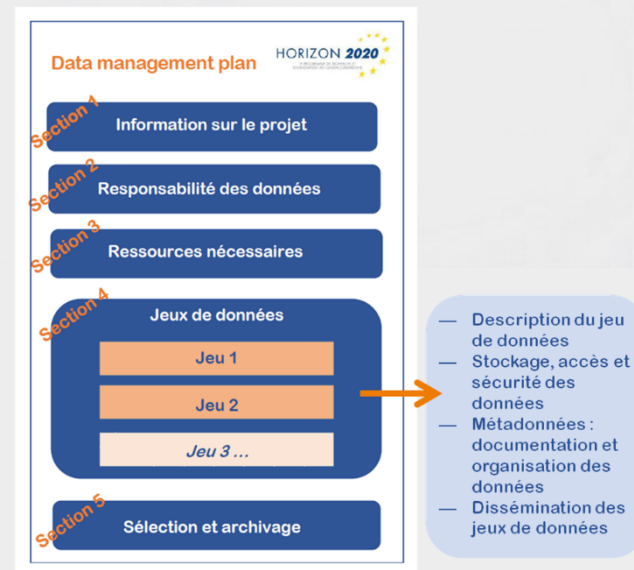
- 7 100 membres venant de 137 pays, 100 groupes de travail
- regroupe **chercheurs et professionnels des données scientifiques** travaillant dans de multiples disciplines et organisations à travers le monde.



Rédaction de plan de gestion de données pour les projets CeDONA













- Un **plan de gestion des données**, *Data management plan* est un document évolutif qui aide le(s) chercheur(s) à définir un plan pour gérer leurs données
- Le plan aborde les éléments suivant:
 - *Description du projet , responsabilité des données*
 - *Stockage, partage, protection et diffusion au cours du projet*
 - *Dissémination et archivage à l'issue du projet*
- Contact avec l'INIST pour utiliser les modèles de DMP OPIDoR



maDMPs : plan de gestion de données « machine actionnable »



- maDMPs : DMP Machine Actionnable.
 - But : Extraire les informations du DMP pour les exploiter par des machines
- 10 règles à mettre en œuvre pour passer du DMP au maDMP
 - <https://zenodo.org/record/1172673>
- API REST opidor en cours de développement

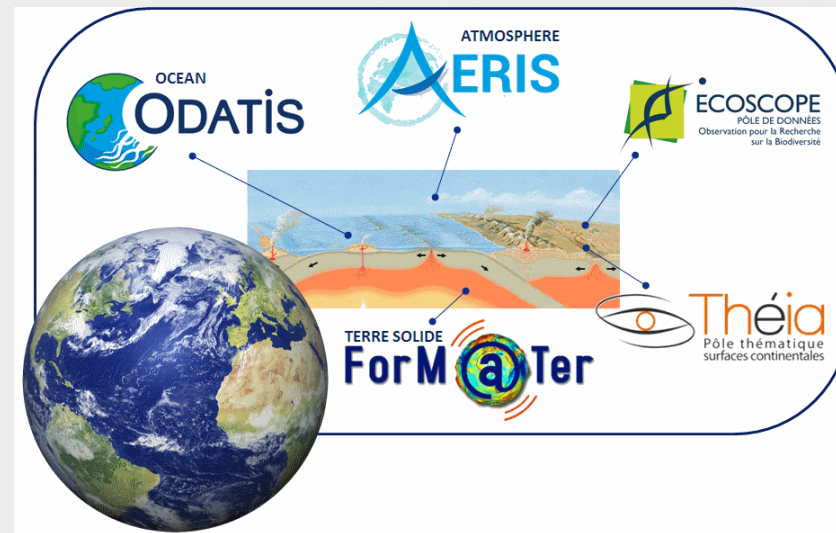
-  1 Integrate DMPs with the workflows of all stakeholders in the research data ecosystem
-  2 Allow automated systems to act on behalf of stakeholders
-  3 Make policies (also) for machines, not just for people
-  4 Describe—for both machines and humans—the components of the data management ecosystem
-  5 Use PIDs and controlled vocabularies
-  6 Follow a common data model for maDMPs
-  7 Make DMPs available for human and machine consumption
-  8 Support data management evaluation and monitoring
-  9 Make DMPs updatable, living, versioned documents
-  10 Make DMPs publicly available



CeDONA collabore avec les pôles d'observation du système Terre



- Pour répondre aux questions que se posent nos sociétés sur leur environnement, la recherche doit appréhender le « Système Terre » dans son ensemble du noyau terrestre jusqu'aux limites de l'atmosphère.
- 5 pôles de données :
 - [FORM@TER](#) : Terre solide
 - [ODATIS](#) : Océans
 - [THEIA](#) : Surfaces continentales
 - [AERIS](#) : Atmosphère
 - [ECOSCOPE](#) : Biodiversité
- Attente de CeDONA sur les outils, les standards et bonnes pratiques à mettre en œuvre dans les projets

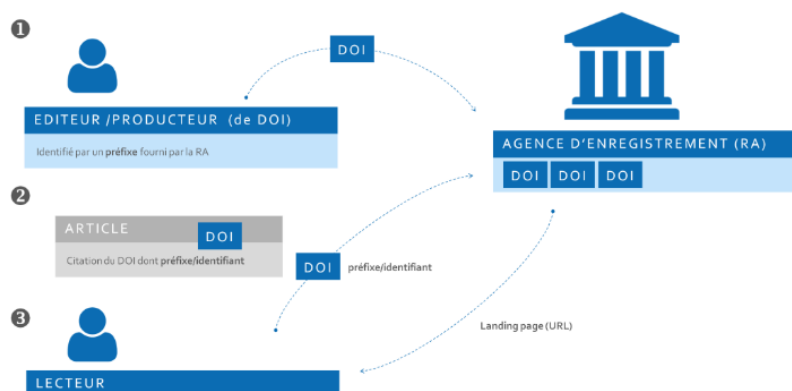


Suivi des recommandations du groupe technique Interpole



- Une fois à deux fois par an, se tiennent sur 2 jours, des ateliers permettant de faire un point sur l'avancement de différents points techniques qui concernent les pôles de données et de services.
- Recommandation sur l'attribution d'un DOI pour l'ensemble des pôles de données : <https://www7.obs-mip.fr/interpole/doi/>

1. Sa création et son enregistrement dans une R.A.
2. Sa citation dans un article (papier ou autre)
3. Sa résolution - via la R.A. - afin de trouver la landing page.



En astrophysique, cadre commun : IVOA



- L'astrophysique utilisent les standards et protocoles développés dans le cadre de l'IVOA (International Virtual Observatory Alliance)
- Déploiement de bases dans tous les OSU - ces bases doivent être interopérables en utilisant les standards IVOA



MERCI

Question?