

Comité de pilotage de Resinfo, 15/12/2021

GT Atelier Données

Marie-Claude Quidoz (CEFE/CNRS)



Ce(tte) œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International.



Vous êtes autorisé à :

Partager — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats

Adapter — remixer, transformer et créer à partir du matériel

Selon les conditions suivantes :



Attribution — Vous devez mentionner le nom de l'auteur de la manière suivante :
« Marie-Claude Quidoz, CEFE-CNRS, 2021 »



Partage dans les Mêmes Conditions — Si vous modifiez, transformez ou adaptez cette œuvre, vous n'avez le droit de distribuer votre création que sous une licence identique ou similaire à celle-ci.



Genèse du GT

- **Rencontres des réseaux professionnels du CNRS (13-14/01/2016)**
 - Atelier : le jeu « inter-réseaux » : la donnée de l'acquisition à la valorisation»
 - 19 inscrits
(Calcul, Cogiter, Cristec, DevLog, Loops, Médici, rBDD, RCCM, Renatis, Resinfo, RTMFM, QeR...)
 - Restitution : https://webcast.in2p3.fr/video/restitution_des_ateliers
 - De nombreux besoins recensés
- **Proposition de création du GT « Atelier Données »**
 - Mettre en place une journée thématique sur l'interopérabilité
 - Rédiger un guide pratique sur la traçabilité des activités de recherche
- **GT « Atelier Données»**
 - ✓ Une initiative des réseaux
 - ✓ Soutien de la Mission pour les Initiatives Transverses et Interdisciplinaires (MITI) du CNRS



Lettre de mission du GT

- Etablir et diffuser une vision transversale de la gestion des données afin d'enrichir la pratique de chaque réseau dans le domaine des données et permettre le développement de la complémentarité entre réseaux
- Identifier les problématiques concernant les données dans chaque réseau
- Valoriser l'apport des expériences et expertises « métier » entre les réseaux
- Sensibiliser les communautés professionnelles sur la gestion des données
- Mettre en commun et partager de nouvelles pratiques



Le GT aujourd'hui

- 20 membres
- Deux duo d'animateurs
 - ✓ Pierre Brochard & Marie-Claude Quidoz (Juillet 2019 – Juin 2022)
- Différentes « structures »
 - ✓ Réseaux MITI : Calcul, Devlog , Medici, QeR, rBDD, Renatis, **Resinfo**, RIS
 - ✓ Autres réseaux : Relier, SIST
 - ✓ Structures : DDOR, INIST
- Des moyens de communication
 - ✓ <https://gt-atelier-donnees.miti.cnrs.fr/>
 - ✓ donnees-inter-reseaux@services.cnrs.fr (création en décembre 2019)
 - ▶ 400 adhérents – 350 messages
 - ✓ Lettre d'information mensuelle (la première a été faite en juin 2020)



Actions réalisées par le GT

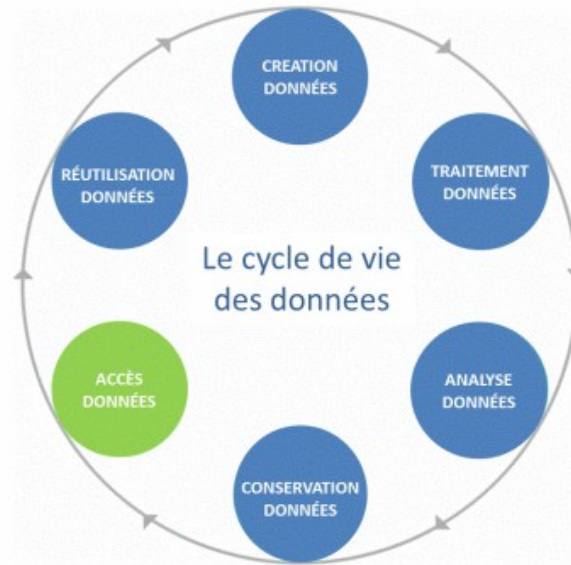
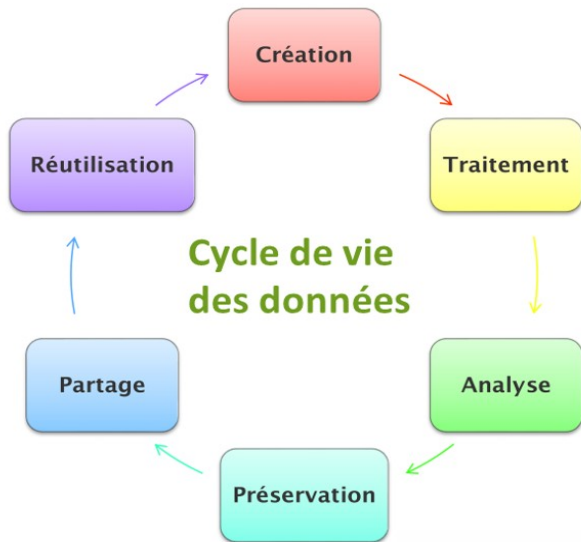
(séminaire, webinaire, hackathon, FAQ, guide, cartographie)

- **Cartographie des actions des réseaux métiers autour de la gestion des données** (2017)
- **Interopérabilité et pérennisation des données de la recherche. Comment FAIR en pratique ? Retours expériences** (2018)
- **Comment améliorer le dépôt de données de recherche** (2020)
- **Data paper, une incitation à la qualification et à la réutilisation des jeux de données** (2020)
- **Guide de bonnes pratiques sur la gestion des données de la recherche** (2021)
- **Qualité des données** (2021)
- **Foire Aux Questions** (entrepôts de données et les data papers)

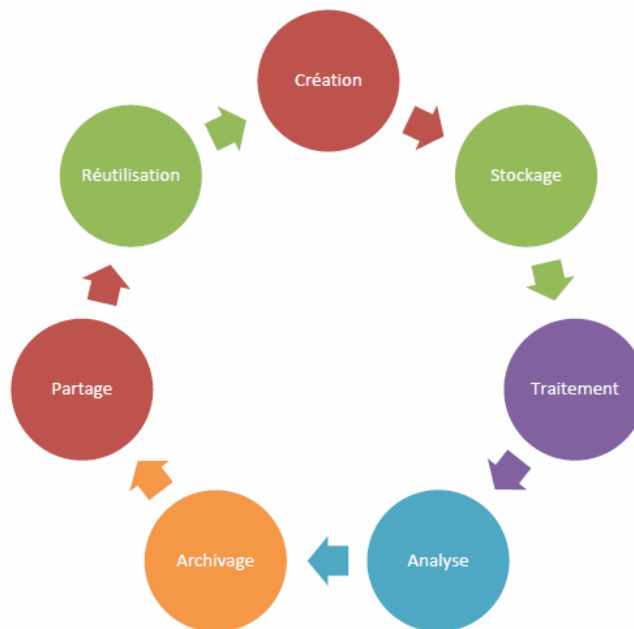
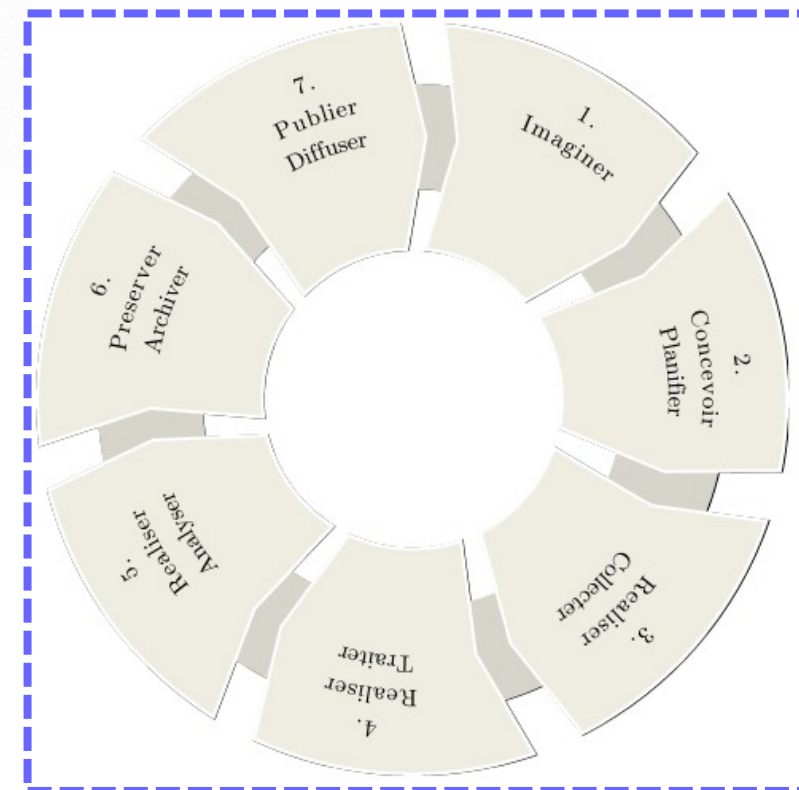


Cartographie des actions des réseaux métiers autour la gestion des données

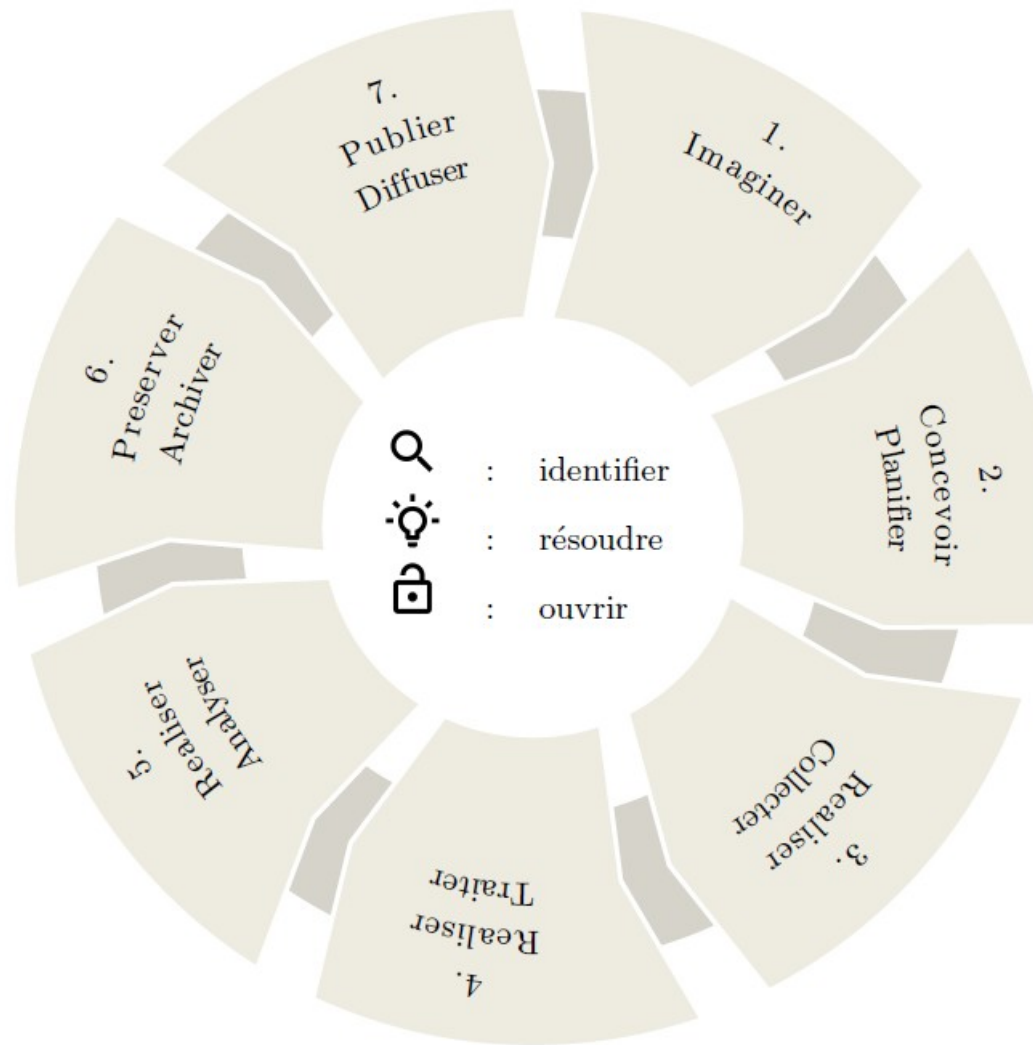
Définir notre cycle de vie des données



GT « Atelier Données »



Se définir sur chaque étape



identifier : problématique de l'étape telle que perçue par le réseau

résoudre : solutions trouvées par le réseau sur cette étape

ouvrir : manques à combler ou défis et perspectives à explorer

Positionnement des réseaux sur l'étape 5

5. Réaliser et analyser

CALCUL

- 🔍 Choix et mise en œuvre des techniques pertinentes pour l'analyse
- 💡 Conception de méthodes, d'outils et de chaînes de traitement adaptées
- 🔒 Croiser les pratiques entre informaticiens, statisticiens et numériciens dans le contexte de données massives et hétérogènes

DEVLOG

- 🔍 Définir les chaînes de traitement et les dispositifs à mettre en œuvre pour chaque projet
- 💡 Identifier les données les plus pertinentes, inscrire l'enrichissement des métadonnées associées dans des chaînes de traitement
- 🔒 Pouvoir connecter des systèmes de données externes

QeR

- 🔍 Disposer de données validées
- 💡 Qualification, étalonnage et suivi des équipements d'analyse, qualification des logiciels, renseignement du cahier de laboratoire

RBDD

- 🔍 Utilisation des outils de requêtage, de statistiques, de visualisation et les portails collaboratifs
- 💡 Définir des procédures de validation des données

RENATIS

- 🔍 Variété des méthodes d'analyse et des disciplines
- 💡 Analyser la cohérence et la qualité de la gestion des données et documenter les processus d'analyse produisant eux-mêmes des données
- 🔒 Améliorer la qualité des données en se servant des analyses réalisées dans le cadre des projets de recherche

RESINFO

- 🔍 Temps de traitements (grands volumes de données), gestion des valeurs manquantes ou aberrantes
- 💡 Contrôle qualité, mise en place de procédure d'étalonnage et de filtrage, outiller et formaliser des chaînes de traitement
- 🔒 Veille technologique sur les outils d'analyse



Intérêt de ce travail

- S'appropriier les étapes du cycle de vie des données
- Mieux connaître le coeur de métier de chaque réseau
- Identifier les manques et les zones de recoupement entre les réseaux
- Initier des collaborations entre réseaux
 - ✓ Interventions dans des formations



Cartographie et RESINFO



RESINFO

- 🔍 Savoir anticiper l'arrivée de nouveaux outils ou savoir identifier les technologies en émergence
- 💡 Traiter en priorité les problématiques liées à l'interopérabilité et au partage des données
- 🔒 Miser sur le développement des infrastructures

1. Imaginer

RESINFO

- 🔍 Obtenir une vision d'ensemble de la place des données dans les projets
- 💡 Veiller à communiquer suffisamment sur les objectifs et le déroulement du projet, instaurer un dialogue entre chercheurs et équipes d'accompagnement technique
- 🔒 Collaborer avec les équipes scientifiques dès la phase de définition des projets

2. Concevoir - Planifier

RESINFO

- 🔍 Résoudre les questions liées à la gestion des données : volumétrie, infrastructure de stockage, transfert de données, débits, interopérabilité
- 💡 Description des processus de production et de gestion des données dans des plans de gestion de données et mise en place de chaînes d'acquisition et de transfert "du capteur au serveur"
- 🔒 Participer à la définition et la rédaction des plans de gestion de données avec les chercheurs

3. Concevoir - Collecter

RESINFO

- 🔍 Problèmes liés au partage des données (accès, sécurisation, authentification) et de représentation spatio-temporelle
- 💡 Préparation et classification des jeux de données, reformulation des métadonnées, modélisation de protocoles et de schémas de bases de données
- 🔒 Veille technologique sur les méthodes d'analyse, les formats favorisant l'interopérabilité

4. Réaliser - Traiter

RESINFO

- 🔍 Temps de traitements (grands volumes de données), gestion des valeurs manquantes ou aberrantes
- 💡 Contrôle qualité, mise en place de procédure d'étalonnage et de filtrage, outiller et formaliser des chaînes de traitement
- 🔒 Veille technologique sur les outils d'analyse

5. Réaliser - Analyser

RESINFO

- 🔍 Stockage et sauvegarde de grand volumes de données
- 💡 Elaborer des politiques de sauvegarde
- 🔒 Mise en œuvre de l'archivage pérenne, veille technologique sur les standards et la sécurité (chiffrement) des données

6. Préserver - Archiver

RESINFO

- 🔍 Connaissance des moyens de gestion et de diffusion des données, notamment les plateformes et les chaînes de traitement
- 💡 Utiliser la visualisation de données, les référentiels de métadonnées, les identifiants pérennes, les grilles de services dédiés aux grands volumes de données, veiller à la sécurisation technique et juridique des accès
- 🔒 Veille juridique ; veille sur les infrastructures

7. Publier - Diffuser

