

# Une charte de réutilisation des données : pourquoi faire ? comment faire ?

Café Science Ouverte PAMIR

–

23 avril 2024

Laurent Romary, Inria



# Au programme

- Au bout du cycle de vie de la donnée
- Pourquoi une charte de réutilisation des données ?
- Les principes de la charte de réutilisation des données
- Mise en œuvre

# Regarder plus loin que le bout de son DMP

## Préparation du projet

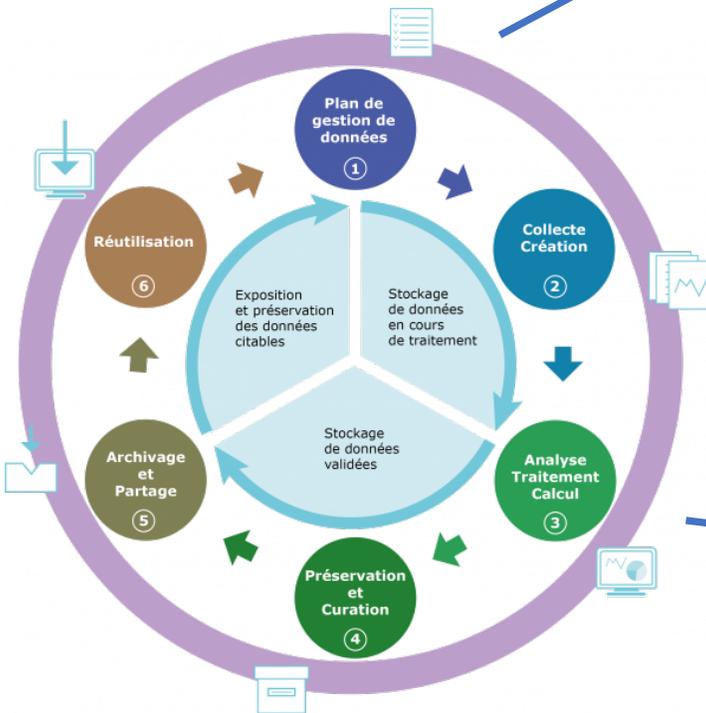
- Objectifs scientifiques (thématique, cf. hébergement)
- Porteurs, responsabilités – institutions, financement
- Contraintes, warnings (sensibilité des données, PI)
  - Actions menées (saisines), définition de la durée d'utilisation ?
- Identification des données entrantes (conditions de réutilisation)
- Initier un profil de métadonnées

## Début effectif du projet

- Instruments, infrastructures, logiciels d'acquisition
- Documentation associée à la collecte/création => liens
  - Annotations, questionnaires, description de la « population »
- Echantillonnage, formats — moment essentiel pour les métadonnées techniques (cohérence avec les corpus de référence)
  - Contrôle qualité / échantillonnage (curation)
- Planifier la réutilisation

## Au cœur du travail scientifique

- Changement de statut éventuel des résultats (sensibilité, ouverture)
- Responsabilités additionnelles (chercheurs, ingénieurs, annotateurs)
- Infrastructures de calcul
- Méthodes <=> références/outils
- Outils <=> référencement des logiciels



# Regarder plus loin que le bout de son DMP

## Le jeu de données change de contexte

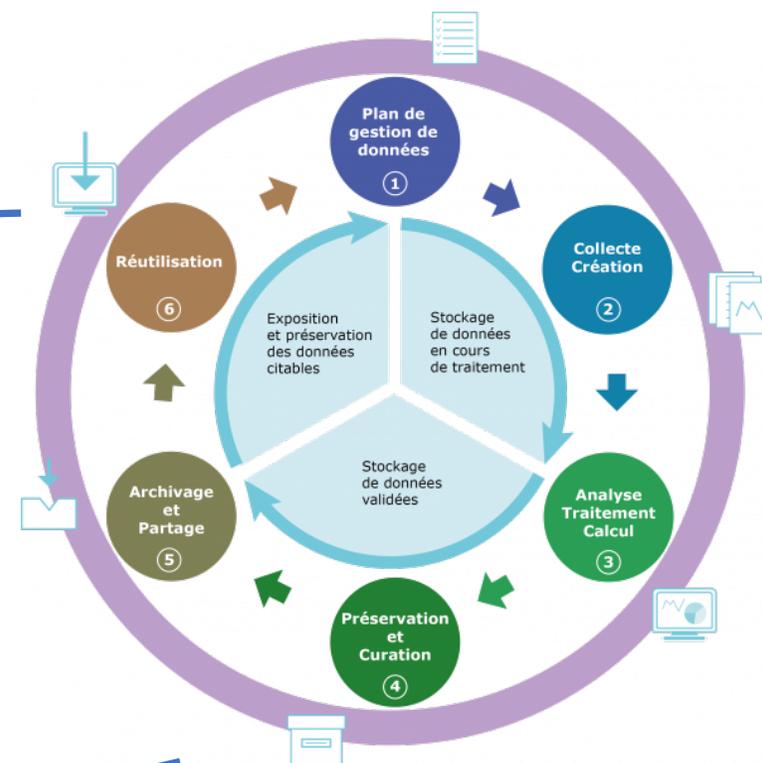
- Contrôle des diffusions
- Suivi des usages
- Retours des utilisateurs, corrections/compléments
- Importance de l'expression des attentes vis-à-vis de tiers
- Charte de réutilisation des données ?

## On coupe le cordon ?

- Décision d'hébergement
  - E.g. intégration à un corpus de référence
- Décision d'ouverture
  - Licence
- Responsabilités (curation)

## Refroidissement des données

- Phase de sélection
- Phase de documentation plus fine
- Phase de publication : lien publications  $\Leftrightarrow$  jeux de données
- Scripts de transformation/visualisation
- Data paper



# Pourquoi une charte de réutilisation des données ?

- Le trauma du projet Cendari (2012-2016)
  - Impossibilité de publier des travaux d'archives du fait de l'ignorance des conditions de réutilisation des sources
- Pour tous les acteurs de la recherche en sciences du patrimoine en particulier
  - Archives/musées, instruments, chercheurs
  - Quel lien de confiance peut-on créer entre un fournisseur d'un jeu de données et ses potentiels utilisateurs ?
- Un travail initial effectué entre infrastructures de recherche européennes
  - DARIAH, CLARIN, E-RIHS, APE, Europeana
  - Ex. la charte de DARIAH campus (formations en ligne) : <https://campus.dariah.eu/docs/reuse-charter>

# La charte de réutilisation des données

- <https://datacharter.hypotheses.org>
- Un déroulé en 6 principes – le projet, l’institution se positionne le long de chacun d’entre eux
  - Réciprocité : quelle attente ai-je vis-à-vis des utilisateurs de mes données (signalement des erreurs, enrichissement, redistribution) ?
  - Interopérabilité : quels sont les formats que j’utilise et que j’encourage ?
  - Citabilité : quel format de citation j’encourage vis-à-vis des jeux de données (publication de référence, DOI) ?
  - Ouverture : quelle licence j’associe à mes jeux de données ? Quelle politique je préconise pour les usages ou dérivation des données que je distribue ?
  - Maintenance/curation (stewardship) : quel engagement ai-je vis-à-vis de la mise à jour et la préservation de mes jeux de données ?
  - Confiance/traçabilité (trustworthiness) : quelle politique ai-je défini pour permettre de remonter aux sources de mes données (données primaires, instruments) ?

# Mise en œuvre

- Un affichage de sa propre politique
  - A associer aux jeux de données ou plus largement comme point d'entrée d'un site de projet, de laboratoire ou d'institution patrimoniale
  - Possibilité d'un affichage à différents niveaux
- Le contraire d'un contrat à valeur juridique
  - Une indication des attentes vis-à-vis de l'environnement
- Un document obligatoire pour les projets du DIM PAMIR
  - Entraîner notamment les jeunes chercheurs à penser leur relation à leur communauté sous l'angle de leurs données