



HAL
open science

Poster "Cycle de vie de la donnée" - printemps de la données 2023 à l'IUEM

Pauline Chauvet, Cédric Chauvel, Laurence Lebourg

► To cite this version:

Pauline Chauvet, Cédric Chauvel, Laurence Lebourg. Poster "Cycle de vie de la donnée" - printemps de la données 2023 à l'IUEM. 2023. hal-04160053v2

HAL Id: hal-04160053

<https://hal.univ-brest.fr/hal-04160053v2>

Preprint submitted on 7 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike | 4.0 International License

Cycle de vie de la donnée de recherche

PLANIFIER

Concevoir son projet avec ses données

Prendre un peu de recul sur le projet en le pensant au travers du prisme de la donnée.

Rédiger un Plan de Gestion de Données en utilisant les modèles à disposition (OPIDOR)

Anticiper la phase de publication et diffusion

CRÉER COLLECTER

Réutiliser ou créer de nouvelles données

Le protocole d'échantillonnage inclut une section « récolte des métadonnées ».

Ces nouvelles données sont accompagnées et stockées avec leurs métadonnées

Pour la réutilisation, vérifiez les licences, les limitations d'utilisations et citez correctement les auteurs.

TRAITER

Curer ses données brutes

C'est le processus visant à vérifier, nettoyer, valider et décrire les données

Les données peuvent se voir apposer un qualificatif de qualité

Les biais et erreurs dus à la machine sont corrigés. Les données personnelles peuvent être anonymisées.

Les données sont stockées et échangées de manière sécurisée pour la durée du projet

Le processus de curation est décrit et vient enrichir le fichier de métadonnées

ANALYSER

Produire des résultats

Les méthodes et outils appropriés sont utilisés pour produire des résultats et (éventuellement) des données secondaires. Les étapes d'analyses doivent être tracées et décrites pour enrichir le fichier de métadonnées.

STOCKER ARCHIVER

Trier ses données

Dans certains cas, il faut supprimer les données brutes à la fin du projet (p.ex. données personnelles non anonymisées)

Réfléchir aux données pertinentes à conserver sur le moyen terme

(stockage) et long terme (archivage). Cela permet de choisir une infrastructure de stockage/archivage et de structurer et organiser les jeux de données à préserver.

DIFFUSER

Valoriser son travail et en faire profiter la communauté

Selon l'adage « aussi ouvert que possible, aussi fermé que nécessaire »

Donner accès aux jeux de données communicables au travers d'entrepôts et de catalogues adaptés au type de jeu de données.

Attribuer des identifiants pérennes (DOI) et des licences adaptées

Fournir les métadonnées les plus précises possibles