



HAL
open science

Rédaction d'un plan de gestion de données : retour d'expérience

Lauriane Locatelli

► **To cite this version:**

Lauriane Locatelli. Rédaction d'un plan de gestion de données : retour d'expérience. 2020. hal-02781765

HAL Id: hal-02781765

<https://hal.science/hal-02781765>

Submitted on 4 Jun 2020

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.

Dans le cadre d'un projet ANR ou d'un ERC qui a déjà débuté (un DMP doit être déposé au plus tard six mois après le début du projet), il faut également démontrer que ces principes sont appliqués et expliquer comment. Des modèles sont disponibles sur la plateforme OPIDoR. Dans notre cas, j'ai choisi d'utiliser le modèle ANR proposé sur OPIDoR, alors que je devais rédiger un DMP pour un projet ERC car le modèle de l'ANR contient de précieuses recommandations.

Le DMP est rédigé en français pour les dossiers ANR et en anglais pour les dossiers ERC.

Les ressources

Afin de rédiger notre DMP, j'ai cherché toutes les informations et formations qui pouvaient m'être utiles. J'ai ainsi appris l'existence de journées d'informations ou des formations autour du DMP, comme l'OPIDoR Tour, ou encore les formations dispensées par les URFIST (Unité Régionale de Formation à l'Information Scientifique Technique) ou les MSH (Maison des Sciences de l'homme). Pour ma part, j'avais trouvé les informations de l'OPIDoR Tour très générales, la formation de l'URIFST très courte (2 heures), celle de la MSH de Lyon ayant été annulée en raison de la grève du 10 décembre 2019.

En tant que chercheuse en sciences de l'Antiquité, et pour les chercheur·e·s de ma communauté, un des réflexes serait d'aller chercher les réponses à ces questions dans les livres, mais il n'y a pas à notre connaissance d'ouvrages qui pourraient constituer une aide notable. Cela m'a contraint à parcourir de nombreux PowerPoint plus ou moins bien faits sur la question.

Pour notre part, j'ai pu contacter la MSH (Maison des Sciences de l'homme), mais aussi le service Ingénierie de Projets de l'Université. Le personnel des Bibliothèques Universitaires peut aussi être contacté pour aider en la matière. D'après nos observations, la meilleure solution semble toutefois être l'entraide entre personnes du même laboratoire, rédigeant ou ayant rédigé ce type de document.

Contenu du Plan de gestion de données

Le DMP doit être conforme aux principes du FAIR, pour *Findable Accessible Interoperable Reusable*. En somme, les données collectées et produites doivent pouvoir être réutilisées et, pour cela, il faut que l'on parvienne à les trouver, à y accéder et qu'elles soient interopérables.

Mais qu'est-ce qu'une donnée ?

Cette question est fondamentale pour la rédaction du plan de données. Dans notre cas, il s'est agi de photos (dans ce cas, le format privilégié est le TIFF), mais aussi des tableurs (de préférence en .csv), de documents textes dans un format libre, de fichiers xml, de script en Python, de la bibliographie Zotéro, etc.

Nos sources et nos productions sont des données.

Le choix de l'entrepôt Nakala

Dans le cadre du DMP, il est demandé au porteur de projet d'expliquer comment et où vont être stockées les données du projet (collectées et produites). Pour les besoins de notre projet, nous avons

choisi de stocker les données dans l'entrepôt numérique Nakala, ce que nous avons renseigné dans le DMP. Recommandé pour les SHS, Nakala nous a semblé l'outil le plus adapté. En effet, cet entrepôt attribue un identifiant à chaque donnée : un handle Nakala. Les données qui sont sur Nakala peuvent également être retrouvées *via* Isidore, qui peut « moissonner » Nakala sur demande.

Concernant les questions éthiques et les données personnelles, il faut dire que le projet n'est pas concerné par ces questions.

Licence

Le DMP invite aussi à préciser l'utilisation de licences. Nous avons identifié les licences CC BY-NC (Creative Commons) et la licence Etalab. Cette dernière est une licence libre française, mais peu connue en dehors de France. La question du volume des données doit aussi être traitée. Dans notre cas, il s'agit de quelques Go. Les conventions de nommage des fichiers doivent être abordées, de même que la question du management du versioning (il existe un logiciel libre de gestion de versions : Git). Ce logiciel sera peut-être utilisé pour la suite du projet. L'utilisation des standards du *Dublin Core* et de vocabulaire(s) contrôlé(s) favorisent l'interopérabilité. Concernant l'utilisation de vocabulaire contrôlé, les standards communautaires du TLG (*Thesaurus Linguae Graecae*) pouvaient être mobilisés pour notre part, de même que PACTOLS, etc.

Afin de s'assurer de la qualité des formats de fichier, l'outil FACILE du CINES (Centre Informatique National de l'Enseignement Supérieur) et le site du W3C pour la validation des formats xml peuvent s'avérer utiles.

Cette expérience m'a appris à adopter les bonnes pratiques de gestion de données et m'a appris à réfléchir à l'accessibilité des données, à leur stockage et leur réutilisation. En somme, cette expérience m'a permis de me poser les bonnes questions quant à la gestion des données et a permis au projet de se rapprocher des recommandations en la matière. Le DMP a été rédigé en trois mois, il a évolué de jour en jour, jusqu'à atteindre 16 pages afin de contenir des informations non essentielles pour le dépôt de projet, mais nécessaires à l'équipe pour comprendre le document après mon départ. Puis, il a fallu retravailler le DMP afin qu'il soit moins long, qu'il ne contienne que les informations nécessaires. Des schémas explicatifs ont ensuite été ajoutés pour une meilleure compréhension. Le DMP a ensuite été traduit en anglais comme il est d'usage pour les projets ERC.

La rédaction de plan de gestion de données pour les chercheurs en SHS n'est pas chose aisée, mais les ressources se développent, d'autant que les DMP sont un des éléments clefs des projets de recherche de demain.

>> Télécharger le DMP sur OPIDoR en français

>> Télécharger le DMP sur OPIDoR en anglais

SITOGRAFIE

Nathalie Reymonet, Magalie Moysan, Aurore Cartier, Renaud Délémontez. Réaliser un plan de gestion de données "FAIR" : modèle . 2018. sic_01690547v2

<https://anr.fr/>

<https://dmp.opidor.fr/>

<https://doranum.fr/plan-gestion-donnees-dmp/>

<https://erc.europa.eu/>

<https://www.cines.fr/>

<https://www.huma-num.fr/>

<https://www.inist.fr/>

<https://www.nakala.fr/>

<https://www.ouvrirlascience.fr>

