

Retour d'expérience sur l'élaboration d'une politique de données au sein du Lesc

Hélène Gautier

Maté La Science Ouverte

Comment favoriser les pratiques de Science ouverte? Connaître et ouvrir les données de la
recherche

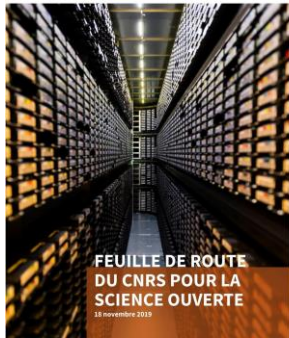
7/10/2021



Des politiques en faveur de la science ouverte à plusieurs échelles

Variété des politiques et des niveaux:

- organisations internationales et nationales
- agences de financement
- universités et institutions de recherche
- revues



Politique de données vs DMP

Politique de données

- Niveau: laboratoire
- Porte sur ce qui relève de la science ouverte (les données mais pas que!)
- Rubriques axées sur la science ouverte
- Communique une vision stratégique
- Révisable mais vocation à durer dans le temps

⇒ Scénario



Data Management Plan (DMP)

- Niveau: projet
- Porte sur les données
- Rubriques axées sur le cycle de vie des données
- Outil de planification/gestion
- Document évolutif (a minima 3 versions)

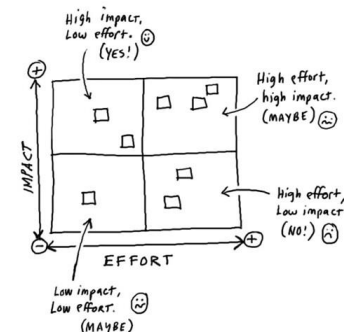
⇒ Plan d'action

Une expérience dans le cadre d'ateliers de l'InSHS

- Cycle d'ateliers animé par Lionel Maurel et Emmanuelle Morlock
- Objectif: rédiger un document résumant les choix du laboratoire en matière de science ouverte, de gestion et de partage des données ainsi que les responsabilités des différents acteurs impliqués
- Plusieurs unités de recherche dont le Lesc (UMR 7186, Nanterre), laboratoire d'anthropologie non spécialisé dans une aire géographique
- Binôme chercheur-ingénieur
- Travail entre les ateliers et en collaboration avec d'autres collègues (notamment ingénieurs)

Une première étape: l'inventaire des données

- Recensement (non exhaustif) des jeux de données
 - Avoir une idée des types de données, des volumétries, de leur localisation, etc.
 - Inclure tout ce sur quoi le laboratoire investit des ressources (matérielles et/ou humaines) ou qui est pris en charge par un membre du laboratoire
- Outil pour la FAIRisation
 - Matrice impact/effort
- Aide à la rédaction de la politique
 - Identifier les priorités
 - Définir le périmètre des “données” pour le laboratoire



Identification du jeu				Caractérisation des données produites				Informations complémentaires				Modes de diffusion / archives				Modalités de partage			
Programme de recherche	Désignation du jeu	Année de collecte ou de production	Nom du chercheur principal ayant collecté ou produit les données	Personne ou service en charge de la gestion des données	Caractérisation des données produites	Etat de finalisation (en cours / publié)	Format(s)	Taille (nombre de fichiers, espace disque)	Données brutes ou primaires (oui / non)	Type	Observations particulières (origine des données, réalisations potentielles, etc.)	Plan de gestion de données (oui/non/en cours)	Zones géographiques couvertes	Serveur et modalités de stockage / sauvegarde	Diffusion via le projet de publication numérique	Partage / archivage comme jeu de données réutilisables	Références/indicateurs par des méta-moteurs, des portails, etc.	Périmètre du partage potentiel	Licence attribuée
Projet NeurOptima	NeurOptima T2	2020		J. Pils Ingénieur	Transcriptomiques (transcriptifs et métabolomes) descriptifs des villes support communisme de République grecque numérisés.	en cours	CSV	1000 fichiers, moins de 2 Go	non	données brutes	Le dataset T2 des données brutes et des métabolomes est un jeu de données de séquençage de génomes dans 10 communautés des agrobiotes et en données quantitatives mesurées.	cf PSD (en plan cours)		Home-num (Hébergement web)	Home-num (Hébergement web)	Zenodo / Home-num (Datacite.org)		Tout le monde	CC-BY 4.0
	NeurOptima Images	2019		J. Pils Ingénieur	collection d'images.	en cours	TIFF, JPEG	1000 images, 200 Mo (données de Fichiers géographiques)		images		cf PSD (en plan cours)		Home-num (Hébergement web)	Home-num (Hébergement web)	Zenodo et Zenodo (en cours)		Tout le monde	cf Licence 2019 pour les images de licence CC-BY pour les images dont les attributions sont requises.
	NeurOptima système de publication	2020		J. Pils Ingénieur	application web personnalisée	en cours	JSON, CSV, XML, PDF, etc.	moins de 200 Mo	non	données informatiques, programmes		cf PSD (en plan cours)		Home-num (Hébergement web, GitHub)	Home-num (Hébergement web, GitHub)	Home-num (GitHub, Zenodo)		Tout le monde	cf Licence pour les codes modifiés par l'équipe de publication.
Plateforme NeurMetade	Plateforme NeurMetade	2017		le développement d'applications, de jeux documentaires	Photographies de terrain et reproduction de documents d'archives pour des projets de recherche sur les sites de Turin, de Genève et de Paris en collaboration avec des équipes associées.	publié	JPEG	1000 fichiers		images fixes	les fichiers et métabolomes publiés sur NeurMetade sont pour une partie importante de l'équipe.		Home-num (Hébergement web)	Home-num (Hébergement web)	Home-num (Hébergement web)	CC-BY-NC et CC-BY-NC pour les images, CC-BY pour les métabolomes.		Tout le monde	

<https://url.fr/gA8j>

L'inventaire de données au Lesc en pratique

- Sollicitation de l'ensemble des membres du laboratoire
 - Remplir l'inventaire ne constitue pas un engagement à ouvrir ses données
 - Informations de nature technique et documentaire
 - Aucune colonne obligatoire, donc aussi possible de compléter si toutes les informations ne sont pas disponibles
 - Possibilité d'avoir plusieurs lignes pour un même projet/jeu de données (selon le type de données, le périmètre de partage, l'état de finalisation, le lieu de stockage, etc.)
- Données surtout pensées en termes de projet (type ANR) ou de terrain, non numériques et "passives"
- Certaines colonnes peu (voire pas) remplies: périmètre de partage, licence
- Côté ingénieurs, prise de connaissance de jeux de données
- Côté chercheurs, prise de conscience de l'anticipation nécessaire des questions de gestion et de partage des données

Un modèle de politique à adapter

- Proposition du [modèle OpenAIRE](#) (2018) pour les infrastructures de recherche traduit en français et adapté par les animateurs de l'atelier
- Sections
 - Préambule
 - Périmètre d'application de la politique
 - Droits, obligations et responsabilités (laboratoire et membres)
 - Accès ouvert aux publications
 - Accès ouvert aux données de la recherche
 - Science ouverte
 - Infrastructure et services d'appui
 - Evaluation des travaux de recherche et des chercheurs
 - Formation
 - Application de la politique
 - Annexes

La politique de données en pratique

- Des questions
 - sur les tutelles (ont-elles des politiques? quels sont leurs engagements? sous quelle forme?...
 - sur l'environnement (y a-t-il des services dédiés dans l'environnement du laboratoire (MSH, université, etc.)?)
 - sur l'unité (les questions de science ouverte font-elles partie du plan quinquennal? comment sont gérées les données en interne?)
 - sur les pratiques (départ à la retraite, changement d'unité: que deviennent les données? Qui prend les décisions relatives aux données? Quand?)
- Peu de modifications mais ajout d'informations concrètes (personne ou service ressource sur telle question)
- Choix de ne pas inclure de mesures contraignantes

Bilan

- Outil le plus utile actuellement (et le plus simple à mettre en oeuvre):
l'inventaire de données
- Un point d'entrée particulièrement pertinent pour les anthropologues: la mission
- Des questions toujours en suspens
 - Faut-il faire valider et adopter cette politique par le laboratoire?
 - Avec quels outils la mettre en oeuvre?
 - Comment relier cette politique aux éventuels DMP?