

Recherche.data.gouv, c'est quoi ?

Café Science Ouverte PAMIR

-
1^{er} mars 2024

Laurent Romary

Inria – Direction de la culture et de l'information
scientifiques



Rappel des épisodes précédents

- Pourquoi rédiger un plan de gestion de données ?
 - Anticiper sur les contraintes portant sur la gestion et la diffusion des données
- Quelles données publier ?
 - Périmètre de la notion de données de la recherche et des devoirs associés
- Aujourd'hui, on adresse le comment
 - [recherche.data.gouv](https://recherche.data.gouv.fr) comme solution d'hébergement générique pour les jeux de données

Qu'attend-on d'une solution d'hébergement ?

- Mission
 - Délimitation claire des objets hébergés
- Pérennité
 - Gouvernance et modèle économique
 - Archivage à long terme
- Services techniques de base
 - Traçabilité des dépôts, identifiants pérennes, licences, métadonnées précises (référentiels), versions, mise à disposition suffisante d'espace
- Curation/modération
 - Assurer un niveau de qualité/conformité (métadonnées, contenus)
- L'ensemble de ces contraintes fait l'objet d'une possible certification :
CoreTrustSeal

recherche.data.gouv.fr

- Un ensemble de services autour des données de la recherche
 - Ateliers de la donnée : accompagnement des chercheurs sur les sites de l'ESR
 - En Ile-de-France: Data Univ Eiffel, ADN (Nanterre), DatASaclay, Données Condorcet, ADSSD (P8 – en cours)
 - Centres de références thématiques : services thématiques
 - IFB (BioInfo), PROGEDO (sciences sociales), DATA TERRA (sciences de la terre), CDS (astronomie), PNDB (biodiversité), Huma-Num (sciences humaines)
- Une solution d'hébergement et d'ouverture générique des données, articulée autour d'espaces institutionnels
- Espaces institutionnels – modération par les institutions concernées
 - Organismes de recherche : CIRAD, CNRS, INRAE, Inria, INSERM, Institut Pasteur, IRD, ONERA
 - Universités et écoles (Ile-de-France) : CYU, Ecole des Ponts ParisTech, ENSAM, ESA, INSHEA, Sciences Po, SU, Gustave Eiffel, P8, Paris Nanterre, Paris-Saclay

Quelles alternatives ?

- Importance de se poser la question de l'hébergement dans un contexte plus large du cycle de vie du jeu de données, d'où le plan de gestion des données
- Solutions institutionnelles fermées
 - Gestion des données au cours de leur création, transformation : avant l'ouverture
- Cas spécifique de jeux de données ponctuels associés à une publication
 - Possibilité de les associer à un dépôt dans HAL ; rappel : danger à n'utiliser que les sites des éditeurs pour cela
- Solutions thématiques
 - Ex. Nakala (Huma-Num) – à privilégier quand elles existent. On peut/pourra référencer de tels jeux de données sur recherche.data.gouv.fr et s'appuyer sur les centres de références thématiques
- Autres solutions génériques
 - Ex. Zenodo – mission, gouvernance, pérennité

Bilan

- Un paysage moins structuré que pour les publications
 - Conséquence de la complexité même de la notion de donnée de la recherche
- Toujours choisir en priorité les solutions souveraines de sa communauté de recherche
- S'appuyer au maximum sur les structures support de ses institutions
 - Ateliers de la donnée
- Référencer les jeux de données en lien avec les publications
 - Citations dans le texte et liens dans HAL
- Participer à la gouvernance et la gestion de recherche.data.gouv.fr
 - Mise en place d'une unité d'appui à la recherche
 - Groupes de travail d'accompagnement