

Humanités Numériques

Cours 11 : Humanités numériques et sciences ouvertes

Loïc Grobol

2021-11-23

Sciences ouvertes

- La (les) science(s) ouverte(s), *open science* ou *open research*
- Mouvement visant à rendre accessibles
 - La recherche scientifique
 - Les données et leur diffusion accessibles
- *Open data*
- Outils de travail collaboratif (Wikipédia, Wikiversité, Wikispecies...) et de science participative (Plant-net...)
- E-learning

Accessible à qui ?

Tout le monde, amateurices et professionnel es.

-
- Multidisciplinarité
 - Multilinguisme
 - Multiculturalité

On considère la science et les données comme un « bien commun ».

Un cadre juridique

(en France)

Article 30 de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique :

Donne un nouveau **droit** aux auteurs de publier en accès ouvert leurs articles après un embargo de 6 (sciences de la nature et sciences formelles) à 12 (LSHS) mois, pour des recherches financées au moins à 50% sur fonds public.

L'article 6 définit le principe d'ouverture par défaut pour les données administratives : données construites dans le cadre des missions de service public, auxquelles les données de recherche appartiennent : il crée une **obligation**

En pratique : suivant les champs de recherche, les implémentations sont variables

- *Gold open access*
- *Green open access*
- Publications « hybrides » : modèle « auteur payeur », à des prix très élevés (Van Noorden, 2013)

Pour la plupart des champs de recherche, le modèle *publish or perish* sert de levier aux éditeurs pour capter indirectement des financements de recherche.

L'ouverture des données reste un champ de bataille (Stérin et Noûs, 2019)

Une politique scientifique

(aussi en France)

Plan science ouverte : La France s'engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement.

Généraliser l'accès ouvert aux publications scientifiques

L'ouverture des publications scientifiques doit devenir la pratique par défaut aussi vite que possible. [...] les publications issues de recherches financées au moyen d'appels à projets sur fonds publics seront obligatoirement mises à disposition en accès ouvert, que ce soit par la publication dans des revues ou ouvrages nativement en accès ouvert, soit par dépôt dans une archive ouverte publique comme HAL.

Structurer et ouvrir les données de recherche

Faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes **FAIR** (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes.

S'inscrire dans une dynamique durable, européenne et internationale

Le succès de la science ouverte implique le développement de nouvelles pratiques quotidiennes pour les chercheurs. Cela nécessite la définition de nouvelles compétences, le développement de nouvelles formations et l'adoption de nouveaux services.

À l'échelle de l'UE

Plan S :

- Obligation de publication en libre accès pour la recherche financées par des subventions publiques
- Coalition de financeurs de la recherche (en France, l'ANR)

Le mouvement pour la science ouverte

Les chercheureuses emploient aujourd’hui les outils numériques dans toutes les facettes de leur activité.

- Outils
 - Plateformes
 - Données
 - Leurs publications nativement numériques.
-

Cependant, ces productions scientifiques numériques de la recherche sont souvent difficilement accessibles :

- Ressources non publiques ou non documentées
- Nécessité d’être abonné ou d’avoir acheté la ressource (*paywall*)
- Formats incompatibles, voire incohérents

→ L’utilisation pleinement efficace du numérique est entravée.

La science ouverte vise à corriger cette insuffisance :

- Approche transversale de l’accès au travail scientifique, des visées et du partage des résultats
 - une nouvelle façon de faire de la science, en ouvrant les processus, les codes, les méthodes, les protocoles...
-

Accès ouvert

- Rend possible le mouvement de diffusion de la production scientifique sur internet
- Lève les barrières de l’accès en maintenant toutes les protections du droit d’auteur. (relativement)
- Aussi ouvert que possible, « aussi fermé que nécessaire » (Rapport commission européenne 2016)

Humanités numériques et sciences ouvertes

Cycle de vie des données

La diversité des sujets autour des données de la recherche est illustrée par leur cycle de vie, constitué de 6 étapes :

- Création ou collecte
- Traitement
- Analyse
- Préservation
- Partage
- Réutilisation

Les données FAIR

Constat :

- Le taux de réutilisation des données (scientifiques) est extrêmement faible

Objectifs :

- Améliorer les infrastructures pour permettre aux machines de retrouver les données, et aux chercheurs de les réutiliser
 - Promouvoir des principes concis et mesurables
 - Bonnes pratiques pour la gestion des données.
-

13 principes qui se déclinent en un ensemble de caractéristiques que doivent présenter les données et les métadonnées pour faciliter leur découverte et leur utilisation par les humains **et** par les machines

Référence aux trois types d'entités :

- Données (ou tout objet numérique),
- Métadonnées (informations sur l'objet numérique),
- L'infrastructure de support (moteur de recherche, plateforme de stockage...).

Facile à trouver (*Findable*)

Faciliter la découverte des données par les humains et les systèmes informatiques par une description et une indexation des données et des métadonnées.

- **F1** Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne.
- **F2** Les métadonnées décrivant les données sont riches.
- **F3** Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher.
- **F4** Les métadonnées spécifient l'identifiant de la donnée.

Accessible

Stocker durablement les données et les métadonnées et à faciliter leur accès, spécifier les conditions d'accès (ouvert ou restreint) et d'utilisation (licence) :

- **A1** Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.
 - **A1.2** Le protocole utilisé permet l'authentification et l'autorisation si besoin.
 - **A1.1** Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.
- **A2** Les métadonnées sont accessibles même quand les données ne le sont plus.

Interopérable

Intégrer les données téléchargeables, utilisables, intelligibles et combinables avec d'autres données :

- **I1** Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
- **I2** Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR.
- **I3** Les données et les métadonnées incluent des liens vers d'autres (méta)données.

Réutilisable

Rendre les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproductibilité, transparence) :

- **R1** Les données et les métadonnées ont des attributs multiples et pertinents.
- **R1.1** Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible.
- **R1.2** Les données et les métadonnées sont associées à leur provenance.
- **R1.3** Les données et les métadonnées correspondent aux standards des communautés indiquées.

Bibliographie

Anne-Laure Stérin et Camille Noûs. 2019. Ouverture des données de la recherche : les mutations juridiques récentes. *Tracés. Revue de Sciences humaines*(#19, #19):37-50, décembre.

Richard Van Noorden. 2013. Open access : The true cost of science publishing. *Nature*, 495(7442, 7442):426-429, mars.