



HAL
open science

Les bonnes pratiques de la science ouverte appliquées aux thèses de doctorat

Sofia Papastamkou

► **To cite this version:**

Sofia Papastamkou. Les bonnes pratiques de la science ouverte appliquées aux thèses de doctorat. Doctorat. France. 2020. hal-03106243

HAL Id: hal-03106243

<https://cel.hal.science/hal-03106243>

Submitted on 11 Jan 2021

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.

Les bonnes pratiques de la science ouverte appliquées aux thèses de doctorat

Sofia Papastamkou, IE CNRS, IRHiS (CNRS/UdL)
Formation doctorale, IRHiS, 7 février 2020
CC-BY

Sommaire de la présentation

Introduction: qu'est-ce la science ouverte

1. Comprendre l'émergence et les politiques pour la science ouverte
2. Pourquoi appliquer les principes de la science ouverte lors d'une thèse de doctorat
3. Comment appliquer les principes de la science ouverte lors d'une thèse de doctorat

Conclusion: la pratique de la science ouverte, une affaire de sociabilité scientifique?

Introduction

- Science ouverte: brève vue d'ensemble
- Science ouverte: quelle utilité pour un(e) doctorant(e)

La science ouverte, qu'est-ce?

La science ouverte est le **mouvement** qui vise à rendre la recherche, les données et la diffusion scientifiques accessibles à tous les niveaux d'une société d'apprentissage.

Il s'agit d'un ensemble de **principes** et de **pratiques**.

=> définition Foster

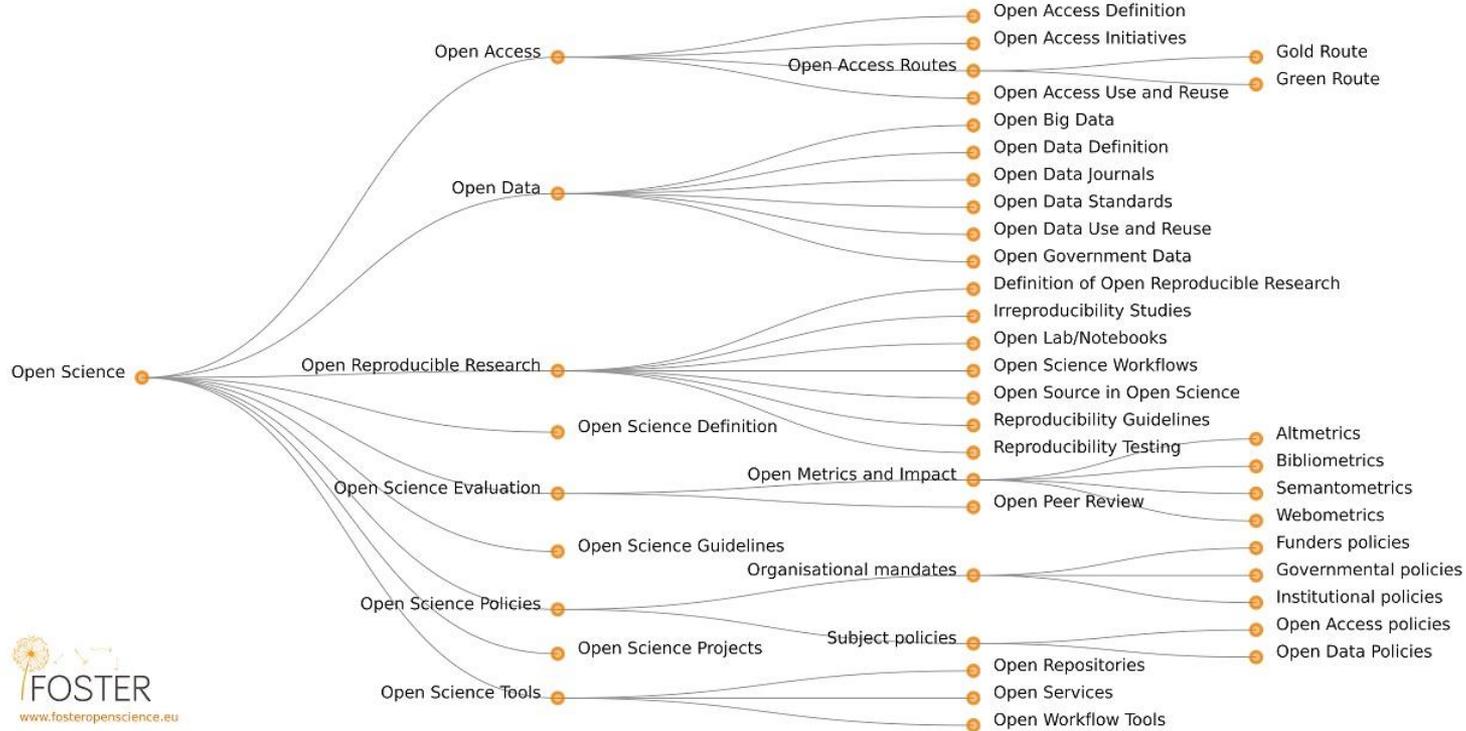
Sonja Bezjak et al. (2018). Open Science Training Handbook (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.1212496>

... mais aussi:

Un ensemble de **politiques** pour la recherche:

- Espace européen de la recherche (EER)
- Cadres juridiques nationaux
- Organismes de financement

Open Science Taxonomy



Principes et pratiques de la science ouverte

Principes

- Transparence
- Réutilisation
- Reproductibilité
- Coopération
- Inclusion
- Partage

Pratiques

- Libre accès aux publications scientifiques
- Partage des données
- Evaluation ouverte
- Reproductibilité de la recherche
- Méthodologie transparente
- Open source
- Science citoyenne
- Ressources éducatives libres

En quoi ces pratiques me sont-elles utiles?

Je préserve mon travail -> protection, pérennité

Je facilite mon organisation -> efficacité; gain de temps

Je réfléchis par la pratique sur mes matériaux de recherche -> réflexion épistémologique

Je me forge une identité numérique -> positionnement + participation dans son champ de spécialisation

Je me prépare pour le futur paysage de recherche -> je sais pratiquer la SO

Comprendre la science ouverte

- Mise en perspective
- Le mouvement pour le libre accès
- Les politiques pour la science ouverte

Petit mémo utile sur une tension de base

Bien commun -> ce qui appartient à la communauté

Biens communs informationnels: Biens communs qui peuvent être créés, échangés et manipulés sous forme d'information, et dont les outils de création et le traitement sont souvent eux-mêmes informationnels (logiciels). Il peut s'agir de données, de connaissances, de créations dans tous les médias, d'idées, de logiciels.

Philippe Aigrain, *Cause commune : L'information entre bien commun et propriété*, Paris, Fayard, 2005

Cité dans: "[Bien communs informationnels](#)", P2P Wiki

Propriété intellectuelle -> droits exclusifs sur des créations intellectuelles

- Propriété littéraire & artistique (droits d'auteur et voisins/CR)
- Propriété industrielle (brevets etc.)

Financiarisation des connaissances

Les communs informationnels, une affaire de longue durée?

Moyen-Âge: “Scientia donum Dei est, unde vendi non potest”

Venise: Instauration des privilèges d’invention (1474)

18e s.: Diderot (conception libérale) ⇔ Condorcet (conception “communautaire”)

- Copyright Act, Royaume-Uni 1710; USA 1770; (prémisses du) droit d’auteur [France révolutionnaire](#) 1793

19e s.: Instauration de brevets et du droit d’auteur

Le mouvement pour le libre accès (*open access*), 1990-2000

Besoins de communication / circulation / partage d'informations à l'intérieur des communautés scientifiques - *longue durée*

Concentration du marché mondial de l'édition scientifique et hausse des prix d'abonnements (1975-1995: +200-300%, Dewatripont et al, 2006) - *conjoncture*

Internet / WWW (1991) - *événement*

// Mouvement pour le logiciel libre/open source

Une mobilisation des acteurs académiques

arXiv (1991)

Rencontre de Sante Fé, NM (1999)

Déclaration de Budapest (2002, Open Society Institute): cadre théorique pour l'OA

Charte de l'European Cultural Heritage Online (2002, Institut Max Planck)

Déclaration de Bethesda, MD (2003)

Déclaration de Berlin (2003, Institut Max Planck)

Définir le libre accès

“La mise à disposition en ligne gratuit et sans restriction” des publications scientifiques (Budapest OA Initiative, 2002)

Accès ouvert = accès au texte intégral

Libre accès = accès au texte intégral et possibilité de réutilisation

Libre accès: formes

Gold: revues en libres accès (DOAJ)

Green: auto-archivage

Bronze: revues à accès ouvert

Hybride: à accès ouvert dans une revue à accès payant

Fermé: toute autre forme (y compris réseaux sociaux académiques)

Déclaration de Budapest, 2002

"accès libre" à [la littérature de recherche validée par les pairs] = mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou créer un lien vers le texte intégral de ces articles, les analyser automatiquement pour les indexer, s'en servir comme données pour un logiciel, [...] sans barrière financière, légale ou technique autre que celles indissociables de l'accès et de l'utilisation d'Internet. Seule contrainte: garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être cités. (Site web Budapest Open Access Initiative)

Déclaration de Berlin, 2003

Libre accès = une source universelle de la connaissance humaine et du patrimoine culturel

Conférence qui marque un engagement plus actif européen

Appelle à des mesures concrètes en faveur de la généralisation de l'OA

Berlin 3, 2005

Feuille de route qui pose les bases de politiques institutionnelles:

- définir un cadre organisationnel (une politique précise)
- créer des infrastructures interopérables (archives ouvertes)
- formuler un cadre juridique pour les dépôts aux archives ouvertes vis-à-vis des auteurs, des éditeurs et des utilisateurs des contenus
- soutenir les revues en libre accès

Des politiques d'en haut: l'Espace européen de la recherche (EER)

Initiative de la **Commission européenne**, 2000, pour favoriser l'excellence scientifique, la compétitivité et l'innovation.

Un "marché intérieur" européen de la recherche: coordination des politiques au niveau européen et des initiatives conçues et financées par l'UE.

Priorités (en lien avec l'OA): infrastructures de la recherche; accès aux publications et aux données de la recherche.

Des politiques d'en haut: l'Espace européen de la recherche (EER)

European Strategy Forum on Research Infrastructures (2002); feuille de route (2006) en faveur d'une politique de libre accès pour la recherche

Recommandation de la Commission (2012): l'OA = un des principes de H2020

H2020: obligation de libre accès aux publications - incitation à diffuser en OA toutes les données de la recherche (obligatoire pour projets pilotes Open Research Data)

Du libre accès à la science ouverte

Carlos Moedas, commissaire européen à la recherche, à l'innovation et à la science (2014-2019): les 3 O (juin 2015)

- Open Innovation
- Open Science
- Open to the World

La science ouverte, un modus operandi pour l'Horizon Europe

- accès ouvert immédiat aux publications
- accès ouvert aux données et aux plans de gestion des données issues de la recherche: *"aussi ouvert que possible, aussi fermé que nécessaire"*

Objectif: "constitution d'un écosystème de communication universitaire innovant et durable, permettant à l'Union européenne d'accroître son efficacité et sa compétitivité en matière de recherche et d'innovation."

Présentation du programme Horizon Europe (09/2019), 17/09/2019, <http://www.horizon2020.gouv.fr>

<http://www.horizon2020.gouv.fr/cid145004/presentation-du-programme-horizon-europe-09-2019.html>

Cadres juridiques nationaux (liste non-exhaustive)

2011: Espagne

2012: engagement du gvt GB en faveur de l'OA

2013: Allemagne, Italie

2014: Danemark, Pays-Bas

2016: Loi pour une République numérique, France

Source: MESRI

https://cache.media.enseignementsup-recherche.gouv.fr/file/Edition_scientifique/23/1/Tableau_compare_des_politiques_de_l_OA_en_Europe_-_Gayoso_-_CSES_-_septembre_2018_1044231.pdf

La science, un bien commun

Science ouverte - notion de communs: 2013 European e-Infrastructure Reflection Group (e-IRG), White Paper, 2013

“La connaissance est un bien commun de l’humanité et [...] les données de la science doivent être érigées en données d’intérêt général” CPU, octobre 2015

“La place grandissante des technologies de l’information dans la recherche scientifique en tant qu’outil de recherche, de partage et de collaboration a réactivé la notion de « bien commun » associée à la Science”. Direction de l’IST, CNRS, Livre blanc.
Une science ouverte dans une République numérique, mars 2016

“La science est un bien commun qui doit profiter à tous” F. Vidal, 2 avril 2019

La science ouverte, un *modus operandi*
de la recherche

Loi pour une République numérique, art. 30

Lorsqu'un **écrit scientifique** issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un **périodique** [...], son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, **du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique**, sous réserve de l'accord des éventuels coauteurs, **la version finale de son manuscrit acceptée pour publication**, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique ou, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est au maximum de six mois pour une publication dans le domaine des sciences, de la technique et de la médecine et **de douze mois dans celui des sciences humaines et sociales**.

Loi pour une République numérique, art. 30

Dès lors que les **données** issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par **un droit spécifique ou une réglementation particulière** et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur **réutilisation** est libre.

Plan national pour la science ouverte (Frédérique Vidal, 4 juillet 2018)

3 piliers:

- généralisation de l'**accès ouvert aux publications**
- **structuration et ouverture des données de la recherche**
- inscription dans une dynamique durable, européenne et internationale

Plan d'action ANR, 2019

" L'Agence demande à ce que les publications consécutives aux projets qu'elle finance, soient déposées en texte intégral dans une archive ouverte, et imposera de plus **un plan de gestion des données (DMP)** pour les projets financés à partir de 2019."

Plan d'action 2019, anr.com, <https://anr.fr/fr/lanr-et-la-recherche/plan-daction-2019/>

Feuille de route du CNRS pour la SO, 18/11/2019

- 100% des publications CNRS accessibles et réutilisables (green OA)
- Partage et réutilisation des données de la recherche
 - “Aussi ouvert que possible, aussi fermé que nécessaire”
 - Principes FAIR
- Fouille et analyse de textes et de données (infras, outils, loi)
- Repenser l'évaluation dans le contexte de la SO (archives ouvertes, data papers, partage de données, ouverture des processus de la recherche)

CoSO, novembre 2019: SO et évaluation

Typologie de documents et d'activités éligibles à l'évaluation

- Valorisation des bonnes pratiques SO pour l'ensemble des productions scientifiques (publis, données, projets, DMP, brevets...)
- Reconnaissance de l'ensemble des activités de la recherche (pas seulement publications!)

“Types de documents, productions et activités valorisées par la science ouverte et éligibles à une évaluation”. Comité pour la science ouverte, version 1, novembre 2019. *Ouvrir la science*,

<https://www.ouvrirelascience.fr/types-de-documents-productions-et-activites-valorisees-par-la-science-ouverte-et-eligibles-a-une-evaluation/>

Pour résumer...

- Plan de gestion de données
 - Opidor
- Structuration et ouverture de données scientifiques
 - Principes FAIR
 - Dépôts de données
- Dépôt des **articles** scientifiques sur une archive ouverte
 - HAL

Principes FAIR

- Findable
- Accessible
- Interoperable
- Reusable
- Identifiants uniques, indexation
- Protocoles
- Formats et métadonnées ouverts
- Provenance, licence

FORCE11, “The FAIR Data Principles”

<https://www.force11.org/group/fairgroup/fairprinciples>

Les pratiques de la science ouverte et ma thèse

Pourquoi les appliquer?

- Diffuser les résultats de sa recherche en libre accès
- Développer une stratégie d'organisation et de préservation des données de sa thèse

Diffuser ma thèse en libre accès ou pas?

La réglementation en vigueur

Arrêté du 25 mai 2016 sur la délivrance du diplôme national de doctorat

Star (ABES) - Sudoc - Identifiant national - Cines

“Sauf si la thèse présente un caractère de confidentialité avéré, sa diffusion est assurée dans l'établissement de soutenance et au sein de l'ensemble de la communauté universitaire. La diffusion en ligne de la thèse au-delà de ce périmètre est subordonnée à l'autorisation de son auteur, sous réserve de l'absence de clause de confidentialité”.

... ce qui signifie:

- Obligation de déposer un fichier unique .pdf de sa thèse (pour la délivrance du diplôme)
- Pas d'obligation de diffuser sa thèse en libre accès (autorisation requise)
 - Si autorisation: diffusion sur la nouvelle plateforme de l'UdL (à partir d'avril 2020)
 - Si autorisation: possibilité de définir une période d'embargo avant diffusion
 - Si pas d'autorisation: diffusion sur Intranet et possibilité de prêt entre bibliothèques
 - Si confidentialité: signalement dans le catalogue mais pas de diffusion du tout
- En droit de déposer sa thèse sur son initiative dans l'archive ouverte Thèses en ligne <<https://tel.archives-ouvertes.fr/>> (autoarchivage)

Comment fait-on à l'UdL SHS?

Votre interlocuteur: la **BU SHS**

Point de contact: liluth@univ-lille.fr

Formations: "Déposer, signaler et diffuser sa thèse : ce qu'il faut savoir" (C. Malleret)

Une nouvelle plateforme de diffusion en 2020

Une nouvelle charte de diffusion de thèses à l'UdL en 2020

Pourquoi diffuser ma thèse en libre accès?

- Sociabilité scientifique: participer et se positionner dans son champ de recherche
- Aboutissement d'un cycle de travail défini dans le temps (avec ou sans monographie)
- Visibilité: faire connaître son travail et son auteur(e) via une diffusion en ligne
- Protection contre le plagiat
- Si allocataire de recherche: en accord avec l'esprit de l'article 30 de la loi pour une République numérique

Consulter deux témoignages:

Émilien Ruiz, "Pourquoi déposer sa thèse dans TEL ?", *La lettre de l'EHESS*, 69, février 2014 <http://lettre.ehess.fr/index.php?7094>

Sébastien Goulard, "Le dépôt de thèse en ligne : une évidence", *La lettre de l'EHESS*, 80, mars 2015 <http://lettre.ehess.fr/index.php?8636>

Pourquoi diffuser mes articles en libre accès?

- Réglementation en vigueur: aucune - pas d'obligation pour les doctorant(e)s
- Si allocataire de recherche / membre d'un projet: vous êtes en droit de suivre l'esprit de l'art. 30 de la loi pour une République numérique
- Pourquoi diffuser mes articles en libre accès: même avantages que pour la diffusion d'une thèse

Une stratégie pour organiser et
préserver ses données de recherche

Les données de recherche dans les thèses de doctorat: éléments de contexte

La réglementation en vigueur

Aucune. [Mais cela peut évoluer](#) à l'heure d'avancement rapide des politiques autour de la science ouverte.

Le contexte de l'UdL SHS

Un fort positionnement (exploratoire) autour des données des thèses de doctorat:

- Deux projets sur les pratiques et la définition d'un workflow de dépôt des données des thèses (J. Schöpfel), [2015](#), [2017-2018](#).
- Un séminaire de l'ED SHS: "[Les données de la recherche dans les thèses de doctorat](#)"
- Un livre blanc produit en 2015: Stéphane Chaudiron, Catherine Maignant, Joachim Schöpfel, Isabelle Westeel. Les données de la recherche dans les thèses de doctorat - Livre blanc. [Rapport de recherche] Université de Lille 3. 2015. <hal-01192930>

⇒ vers des thèses composées de deux livrables (1 fichier texte: résultats de la recherche; 1 dépôt séparé de données, plutôt qu'un .pdf pour tout)?

Données de recherche dans les
thèses de doctorat: de quoi
parle-t-on au juste?

Image sous droits d'auteur - consulter la version disponible en ligne:
Jorge Cham, “Data by the numbers”, PhD Comics,
<http://phdcomics.com/comics.php?f=462>

Données de recherche: définition

“Les fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche.”

Code de la propriété intellectuelle, L 122-5; art. 38 de la loi pour une République numérique

Données de recherche: définition

“Les données sont **des valeurs factuelles enregistrées** (chiffres, textes, images et sons) qui servent de **sources primaires pour la recherche scientifique** et font le consensus au sein de la communauté scientifique en tant que des **éléments nécessaires pour la validation des résultats scientifiques**. **Un jeu de données est une représentation systématique d'un sujet de recherche.**”

Définition de l'Organisation de coopération et de développement économiques

Données de la recherche: définition

“Une donnée est **une description élémentaire [...] d’une réalité**. C’est par exemple une **observation** ou une **mesure**.

À partir de données collectées, **de l’information** est obtenue en organisant ces données, en les structurant pour en dégager du sens.

En comprenant le sens de l’information, nous aboutissons à **des connaissances**, c’est-à-dire à des « faits » considérés comme vrais dans l’univers d’un locuteur, et à des « lois » (des règles logiques) de cet univers.”

Exemple

“Des mesures de température relevées chaque jour dans une station météo, ce sont des **données**. Une courbe donnant l'évolution dans le temps de la température moyenne dans un lieu, c'est une **information**. Le fait que la température sur Terre augmente en fonction de l'activité humaine, c'est une **connaissance**.”

Serge Abiteboul, *ibid.*

Ce que je sais du doctorat – Acquis de conscience

https://consciences.hypotheses.org/413

des émotions, séminaire "Façons d'être : corps, émotions, récit de soi du Moyen Âge à nos jours" (Aix en Provence)
— 9 et 10 décembre — Colloque "Le goût de l'archive audiovisuelle à l'ère numérique", INHA (Paris)

ON PARLE DE...

archives archives privées autorité bibliographie
biographie catholicisme clergé confession conscience
correspondance correspondances couple direction
direction spirituelle doctorat **genre** histoire
historiographie hommes idée intimité jeunesse
journal lettre lettres littérature mariage masculinité
Michelet mystique numérique recherche Religion
Rome réflexivité sexualité souffrance source thèse
vacances violence virilité écrire écriture éducation

NOURRIR LA PENSÉE EN POURSUIVANT LA PROMENADE AILLEURS :

MES ENSEIGNEMENTS

Supports de cours sur Prezi
Supports de cours sur Slides

CARNETS QUE J'AIME

5. Il passe de longues heures dans les bibliothèques, centres d'archives, sur le terrain, en entretien, à la machine à café
6. Il lit, lit, lit toute la bibliographie qui concerne son sujet (il essaie) ; prend des notes et établit des fiches ; oublie ce qu'il a lu et refait trois fois la même fiche ;
7. Il tente de s'organiser : il achète des classeurs et télécharge Zotero, se renseigne sur les outils de veille en ligne ; est terrifié à l'idée qu'un incendie puisse détruire... ses données ;
8. Il écrit : son cahier de thèse, ses idées, sa thèse quand il rédige, des mails, des mails, des mails....
9. Il traite des données : statistiques, logiciels plus ou moins compliqués, transcriptions ; il y gagne une belle paire de lunettes !

découvre son univers professionnel...

10. Il propose des interventions à des colloques ou journées d'étude
11. Il écrit ses interventions, bien souvent l'avant veille car le doctorant est organisé (voir point 7)
12. Il voyage pour aller parler de sa thèse et manger des glaces en Italie
13. Il essaie de décoder les règles non écrites de l'université: qui apprécie qui, quel sujet éviter, les questions de hiérarchie et de pouvoir: il connaît des maladroites bien sûr !

Taper ici pour rechercher

Chargement terminé (100 %) 2020

Caroline Muller, « Ce que je sais du doctorat », Acquis de conscience, 4 septembre 2013 <https://consciences.hypotheses.org/413>

“Ce que je sais du doctorat”, extraits

[Le doctorant]...

passé de longues heures dans les bibliothèques, centres d'archives, sur le terrain, en entretien, à la machine à café => *collecte de données primaires*

[...] achète des classeurs et télécharge Zotero, se renseigne sur les outils de veille en ligne ; est terrifié à l'idée qu'un incendie puisse détruire... ses données ; => *organisation, sauvegarde de données*

[...] écrit : son cahier de thèse, ses idées, sa thèse quand il rédige [...] => *documentation; résultats de recherche*

traite des données: statistiques, logiciels plus ou moins compliqués, transcriptions [...] => *traitement de données primaires; production de données secondaires*

Typologie de données sur lesquelles portent les recherches en SHS

(Prost et Schöpfel, 2015)

(Corpus de) textes

... que nous qualifions de plusieurs termes:

Enquêtes et entretiens

données primaires, données sources, données brutes (raw data), données organiques (organic data), données issues de collectes supervisées, données issues de collectes non supervisées...

Observations

Expériences

Archives

Statistiques

Photos (documents, objets...)

Enregistrements audiovisuels

Typologie de données que produisent les recherches en SHS

(Prost et Schöpfel, 2015)

Textes

... que nous qualifions de plusieurs termes:

Tableaux

Données secondaires

Bases de données

Données d'analyse

Visualisations, modèles multidimensionnels

Données résultats

Enregistrements audiovisuels

Données dérivées

Photos

Cartes et plans

Programmes et applications

... par conséquent:

Les données sont au coeur de votre travail de thèse, c'est pourquoi elles sont précieuses!

La préservation de vos données de thèse

Objectifs

- Accès
- (Ré)utilisation

Moyens

- Sauvegarde
- Structuration
- Description et organisation
- Documentation

S'inspirer des principes de la science ouverte

Reproductibilité: open source, open notebooks

Open Data: standards ouverts

Open Tools: dépôts de données, archives ouvertes

A noter: il ne s'agit pas de tout ouvrir, si vous ne le voulez pas. Il s'agit d'abord de partir sur les bonnes bases qui garantissent l'accessibilité et la pérennité de votre travail pour vous; et, le cas échéant, de l'ouvrir quitte à définir des périmètres.

Les pratiques de la science ouverte et ma thèse

Comment les appliquer
concrètement?

Avant et après tout: la sauvegarde

Image sous droits d'auteur - consulter la version disponible en ligne:

Jorge Cham, “The Four Stages of Data Loss”, PhD Comics,

<http://phdcomics.com/comics/archive.php/images/archive.php?comiconid=382>

Sauvegarder: la règle 3-2-1

3 (1 + 2) copies

2 supports de stockage (disque externe, cloud...)

1 sauvegarde hors site

+ Régularité

Il est plus facile et intelligent de sauvegarder des données bien structurées et organisées.

Documenter: garder la mémoire vive de sa recherche

- Préserver le contexte de la création des données
- Expliciter la méthodologie de création des données et des choix épistémologiques
- Garder trace de l'évolution d'une recherche
- Créer les conditions pour se réapproprier une recherche (intervalles de temps, travail à plusieurs...)
- Faciliter la rédaction
- Poser les bases d'une publication (data paper)

Documenter: quelques exemples

Notes sur la consultation des documents d'archives

Notes sur la génération de données tabulées

Notes sur la collecte de sources audiovisuelles

Notes sur la préparation d'entretiens, d'enquêtes...

Pour aller plus loin sur la documentation des données

James Baker, «Préserver ses données de recherche,» traduction par Anne-Sophie Bessero-Lagarde, *The Programming Historian en français* (2020),

<https://programminghistorian.org/fr/lecons/preserver-ses-donnees-de-recherche>

2 exemples de dépôts de données tabulées accompagnées d'une documentation:

Galvez-Behar, Gabriel. (2020). Patents and certificates of addition granted in France from 1904 to 1921 by applicant's country (Version 1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3627575>

Crymble, A., Falcini, L., & Hitchcock, T. (2015). Vagrant Lives: 14,789 Vagrants Processed by Middlesex County, 1777-1786 (version 1.1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.31026>

Structurer les données de recherche

A comprendre: les manières d'organiser et de décrire les données en choisissant les formats, le nommage et l'organisation de nos fichiers

1. Formats

Format informatique: une convention sur la représentation d'une donnée numérique sur un support de stockage. Le format est le support de nos fichiers.

Fichier: le fichier est un ensemble de données numériques regroupées sous un même nom dans un support de stockage. Un fichier se caractérise notamment par un nom, une taille, un format, des dates (création, modification) et autres **métadonnées**.

Formats d'édition, formats de préservation

Formats d'édition / publication courants

doc, docx

odt

pdf

xls, xlsx

ods

Formats de préservation

txt: texte brut Unicode

csv: comma-separated values

tsv: tabulation-separated values

xml

Formats: recommandation

Qu'il s'agisse de formats d'édition ou de préservation, préférer les formats ouverts non propriétaires.

Pour préserver les données de recherche, préférer le format

- .txt (pour les textes);
- .csv ou .tsv pour les tableurs
- .png pour les photos

Consulter: le service FACILE du Cines <https://facile.cines.fr/>

2. Nommer les fichiers

Buts:

- Fournir des métadonnées (données descriptives de vos données).
- Permettre de repérer, identifier, classer les documents et données vite et facilement

Votre convention de nommage doit d'abord parler à vous-même!

Consulter:

James Baker, «Préserver ses données de recherche,» traduction par Anne-Sophie Bessero-Lagarde, *The Programming Historian en français* (2020),

<https://programminghistorian.org/fr/lecons/preserver-ses-donnees-de-recherche>

Nommer les fichiers: bonnes pratiques

- Éviter les signes diacritiques (accents, caractères spéciaux)
- Éviter les mots vides et les articles (par ex: mesnotesdelajourneeetc.doc => #fail)
- Éviter les espaces, utiliser _ pour les remplacer
- Nom court, significatif, unique (par ex: enquete_2017.csv)
- Versionnage explicite (par ex: nomFichier_v1.odt, nomFichier_v2.odt)
- Respecter une convention du début à la fin d'un travail de thèse

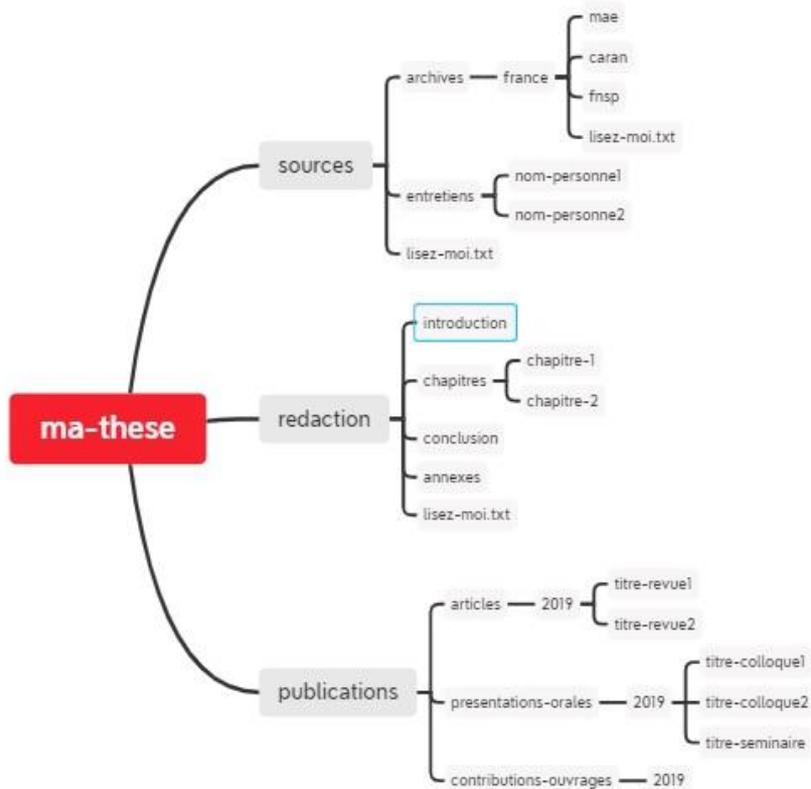
3. Organiser les fichiers

Objectifs:

- Accéder facilement à ses données
- Faciliter la gestion et le repérage des données
- Fournir du contexte
- Permettre la reconnaissance des fichiers dans des environnements techniques différents

Organiser les fichiers: quelques idées

- La meilleure organisation de fichiers et de répertoires est celle qui parle le plus à leur(s) créateur(s)
- Peut s'inspirer des règles du records management ou de l'archivage, mais la thèse fournit un cadre souple car personnel
- Pour le nommage, combiner des éléments significatifs (par exemple une thématique) et des éléments d'identification (par exemple une date, un nom...) selon sa recherche
- Éviter une arborescence profonde (3 niveaux max)
- Mener une réflexion au début de sa thèse, concevoir un cadre de classement évolutif, adapter / modifier au fur et à mesure que la recherche avance.



Esquisse d'un plan basique et imparfait d'organisation des fichiers de sa thèse

Le plan de gestion de données

“Le plan de gestion de données ou Data Management Plan (DMP) est un document formalisé explicitant les modalités de production, de traitement, de protection et de diffusion des données, la propriété intellectuelle, les choix réalisés en termes d’archivage”.

Guide de l’université Descartes, <https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>

Consulter:

Marie Puren, Créer un plan de gestion de données (DMP), ANF Participer à l’organisation du management des données de la recherche, gestion de contenu et documentation des données, juillet 2017, Vandoeuvre-lès-Nancy,

https://anfdonnees2017.sciencesconf.org/data/pages/20170706_dmp_puren.pdf

Le plan de gestion de données et la thèse

Conçu pour les projets de recherche financés sur fonds publics, il peut être **un instrument de rationalisation de l'organisation de son travail de thèse.**

Exercice:

Écrire un DMP inspiré de sa thèse avec l'application Opidor <https://dmp.opidor.fr/>

Déposer / partager ses données de recherche

Pas d'entrepôts spécifiques pour les étudiant(e)s en thèse

Zenodo: www.zenodo.eu (un service de l'Organisation européenne pour la recherche nucléaire (CERN))

Exemples de dépôts de jeux de données:

Galvez-Behar, Gabriel. (2020). Patents and certificates of addition granted in France from 1904 to 1921 by applicant's country (Version 1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3627575>

Papastamkou, Sofia. (2017). 2015 Greek Referendum (#greferendum) tweets [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.887993>

Mémo: attention aux questions juridiques!

Question de la protection des données à caractère personnel selon le RGPD.

Pour se renseigner, consulter:

Réseau supDPO, Quinze recommandations aux chercheurs sur la protection des données dans le cadre de leurs activités de recherche, v1, 16 janvier 2020

<https://reseau.supdpo.fr/wp-content/uploads/2020/01/SupDPO-Recommandations-chercheurs-v1.pdf>

Plusieurs communications de Thomas Soubiran (IE CNRS, Ceraps, CNRS/UdL) disponibles sur

<https://pro.univ-lille.fr/thomas-soubiran/publications/>

Auto-archivage des résultats de recherche

Toute publication peut être déposée sur l'archive HAL, le dépôt Zenodo... à condition de respecter l'esprit de l'art. 30 de la loi pour une République numérique.

Possibilité d'utilisation de **licences CC** pour protéger la paternité de son travail et définir les termes d'utilisation.

Pour en savoir plus, consulter: <https://creativecommons.org/licenses/?lang=fr-FR>

Identifiant contributeur/contributrice

Qu'est-ce?

Un code unique et pérenne qui identifie de manière non équivoque

Pour qui?

Toute personne impliquée à des activités de recherche et aux productions scientifiques

À quoi cela sert?

Associer un(e) auteur(e) aux résultats de ses recherches (données, articles...)

Identifiant contributeur/contributrice

A générer soi-même:

ORCID: identifiant global (<https://orcid.org/>)

IdHAL: identifiant local HAL (<https://doc.archives-ouvertes.fr/identifiant-auteur-idhal-cv/>)

Sofia Papastamkou - Publications

Nombre de documents

11

ARTICLE DANS UNE REVUE

2 documents



Sofia Papastamkou. Greece between Europe and the Mediterranean, 1981-1986: The Israeli-Palestinian Conflict and the Greek-Libyan Relations as Case Studies. *Journal of European integration history*, Nomos-Verl.-Ges., 2015, Europe and the Mediterranean in the long 1980s, 21 (1), pp.49-69. (halshs-01214944)



Sofia Papastamkou. De la crise au Liban au mémorandum du 17 septembre 1958: la politique étrangère de la France entre deux Républiques et une guerre. *Matériaux pour l'histoire de notre temps*, 2010, La France et la Méditerranée : ambition de puissance, perceptions, interactions, pp.76-83. (halshs-01294083)

COMMUNICATION DANS UN CONGRÈS

3 documents



Antonio Rojas Castro, Sofia Papastamkou, Anna-Maria Sichani. Three Challenges in Developing Open Multilingual DH Educational Resources The Case of The Programming Historian. *Digital Humanities 2019*, ADHO, Jul 2019, Utrecht, Netherlands. (halshs-02277639)

Sofia Papastamkou. Twitter as a source for the history of the present : the 2015 Greek referendum as a case study. *Digital Humanities 2017*, Aug 2017, Montréal, Canada. (halshs-01689289)

Exemple de page générée avec l'IdHAL

Bâtir son identité numérique chercheur(e)

- Déposer ses publications scientifiques sur une archive ouverte
- Déposer ses données de recherches sur un dépôt de données
- Obtenir un identifiant auteur(e)
- Communiquer autour de ses recherches sur un carnet de recherche ou blog

⇒ Construire une identité numérique cohérente pour se positionner dans son domaine de recherche dans un environnement numérique.

Exercices

1. Ouvrir un compte sur HAL <https://halshs.archives-ouvertes.fr/> et déposer une publication, si pertinent, ou à défaut les métadonnées d'une publication.
2. Ouvrir un compte sur Zenodo et explorer l'interface. Déposer une publication ou un jeu de données, si pertinent (<https://zenodo.org/>).
3. Obtenir un identifiant Orcid sur <https://orcid.org/>
4. Rédiger sur place un plan de gestion de données selon votre travail de thèse: à l'aide de l'application Opidor (<https://dmp.opidor.fr/>) ou sur un tableur à l'aide du guide du SCD de l'Université de Paris (le guide et le formulaire type sont disponibles ici: <https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>).

À titre de
conclusion...

(Re)lisons Eco:

“Faire une thèse signifie donc apprendre à organiser des données et ses propres idées: c’est une expérience de travail méthodique pour construire un “objet” qui, en principe, serve aussi aux autres. Ce qui importe est donc moins le sujet que l’expérience de travail qu’elle implique”.

Umberto Eco, *Comment écrire sa thèse*, Paris, Flammarion, 2016, p. 29

“[...] montrer qu’on “fait partie du club” n’est pas une question de snobisme, d’appartenance à une élite: cela renvoie plutôt au fait que le discours du chercheur, mais aussi de l’étudiant qui écrit une thèse, [...] s’inscrit toujours déjà dans un ensemble assez vaste d’autres chercheurs, qui ont commencé à chercher avant lui et continueront après lui – et en partie grâce à lui. On ne cherche jamais seul, même dans la solitude d’une bibliothèque, même en lisant un livre ou en rédigeant son propre texte. [...] “Ne jouez pas les génies solitaires”, recommande Eco”. Et ce n’est pas seulement une question de diffusion des résultats ou de circulation du savoir: [...] faire une thèse c’est se plonger dans un monde d’échanges [...], en faire partie, y contribuer [...].”

Que retenir?

S'inspirer des bonnes pratiques de la science ouverte permet de travailler de manière méthodique et intelligente sur son sujet. Cela permet aussi d'acquérir des compétences qui sont appréciées dans une évolution de carrière, que ce soit dans le monde de la recherche ou ailleurs.

Plus que tout, ce sont les sociabilités scientifiques qui sont inscrites au coeur des métiers de la recherche. Indépendamment des politiques de financement de la recherche, les principes de la science ouverte s'inscrivent essentiellement dans ce type de sociabilités.