

INTRODUCTION A LA SCIENCE OUVERTE

Florence THIAULT





URFIST

rennes

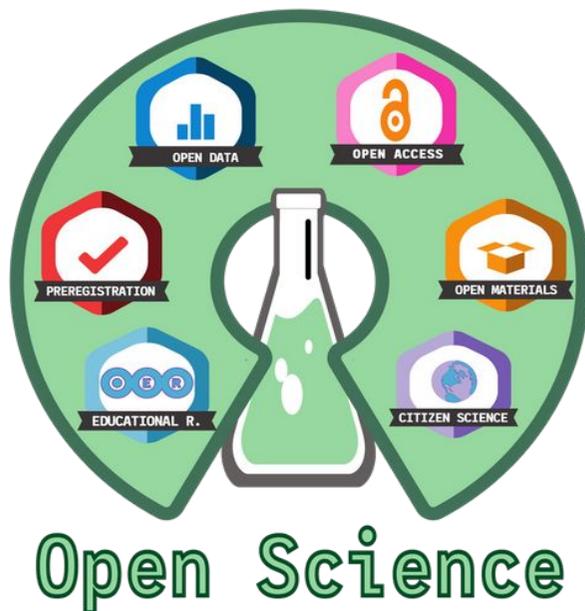


Plan

- Définir la science ouverte
- Les principes de la science ouverte
- Les grandes dates de la science ouverte
- L'accès ouvert
- L'ouverture des données et matériaux de recherche
- Les logiciels libres
- Ouvrir la recherche entrain de se faire
- L'évaluation ouverte
- Ouvrir la science avec et pour la société

Définir la science ouverte

Une définition parmi d'autres



« La science ouverte est un mouvement visant à rendre la recherche, ses outils et ses productions **accessibles** à tous les niveaux de la société. C'est pratiquer la science de manière à ce que d'autres puissent collaborer et contribuer, où les données de recherche, les notes de laboratoire et autres processus de recherche sont librement disponibles, selon des conditions qui permettent la **réutilisation, la redistribution et la reproduction** de la recherche et de ses données et méthodes sous-jacentes. »

Source: FOSTER

La science ouverte en France

OUVRIR

LA SCIENCE

[SCIENCE OUVERTE](#) [COMITÉ](#) [GROUPES](#) [BLOG](#) [AGENDA](#) [RESSOURCES](#) [JNSO 2019](#)

[FR](#) - [EN](#) 



POLITIQUES PUBLIQUES

La science ouverte est la diffusion sans entrave des publications et des données de la recherche. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données de la recherche.



Télécharger

2^{ème} plan national Science Ouverte 2021-2024

LIEN

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
*Liberté
Égalité
Démocratie*

Deuxième Plan national pour la science ouverte



Premier axe

Généraliser l'accès ouvert aux publications

8

Mesures

10

- 1 Généraliser l'obligation de publication en accès ouvert** des articles et livres issus de recherches financées par appel à projets sur fonds publics
- 2 Soutenir les modèles économiques d'édition en accès ouvert** sans frais de publication pour les auteurs (modèle « diamant »)
- 3 Favoriser le multilinguisme** et la circulation des savoirs scientifiques par la traduction des publications des chercheurs français

2^{ème} plan national Science Ouverte 2021-2024

LIEN

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Démocratie

Deuxième Plan national pour la science ouverte



Deuxième axe

Structurer, partager et ouvrir les données de la recherche

12

Mesures

14

- 4 Mettre en œuvre l'obligation de diffusion des données de recherche financées sur fonds publics**
- 5 Créer Recherche Data Gov, la plateforme nationale fédérée des données de la recherche**
- 6 Promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR)**

URFIST

rennes

2^{ème} plan national Science Ouverte 2021-2024

[LIEN](#)

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Démocratie

Deuxième Plan national pour la science ouverte



Troisième axe

Ouvrir et promouvoir les codes sources
produits par la recherche

16

Mesures

18

- 7 **Valoriser et soutenir la diffusion sous licence libre des codes sources** issus de recherches financées sur fonds publics
- 8 **Mettre en valeur la production des codes sources** de l'enseignement supérieur, de la recherche et de l'innovation
- 9 **Définir et promouvoir une politique en matière de logiciels libres**

2^{ème} plan national Science Ouverte 2021-2024

LIEN


MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
*Liberté
Égalité
Démocratie*

Deuxième Plan national pour la science ouverte



Quatrième axe

Transformer les pratiques pour faire
de la science ouverte le principe par défaut

20

Mesures

22

- 10 Développer et valoriser les compétences de la science ouverte** tout au long du parcours des étudiants et des personnels de la recherche
- 11 Valoriser la science ouverte et la diversité des productions scientifiques dans l'évaluation** des chercheurs et enseignants-chercheurs, des projets et des établissements de recherche
- 12 Tripler le budget de la science ouverte** en s'appuyant sur le Fonds national pour la science ouverte et le Programme d'investissements d'avenir


URFIST

rennes

Les principes de la science ouverte

Les principaux concepts de la science ouverte

- ❑ **Transparence** : mieux comprendre comment la recherche avance, réduire les marges d'erreur en permettant à d'autres chercheurs de reproduire des expériences ou des raisonnements
 - ❑ **Intégrité scientifique** : lutter contre le plagiat, la fabrication de données ou falsification de données
 - ❑ **Démocratisation** : permettre à la société (décideurs, praticiens, citoyens...), aux scientifiques (quel que soit leur institution ou leur pays d'appartenance), d'avoir accès à la connaissance scientifique et d'en réutiliser les outils et résultats
- ❑ *La gratuité n'est pas une fin en soi mais un moyen*
- ❑ *L'accès ouvert et gratuit aux résultats scientifiques n'est pas suffisant, c'est la **réutilisation sans entrave** qui est au cœur de la science ouverte*



Portée par les progrès sans précédent de notre monde numérique, la transition vers la science ouverte permet aux informations, aux données et aux résultats scientifiques d'être plus largement accessibles (**accès ouvert**), d'être exploités de manière plus fiable (**données ouvertes**), avec la participation active de toutes les parties prenantes concernées (ouverture vers la société).

« Vers une recommandation de l'Unesco sur la science ouverte : établir un consensus mondial sur la science ouverte » ([en ligne](#)).



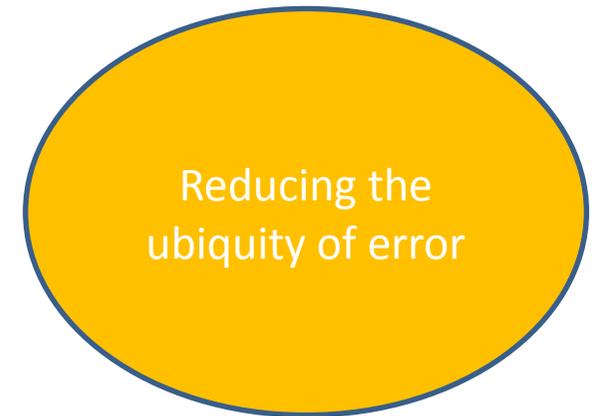
Principes et pratiques de la science ouverte

❖ Principes :

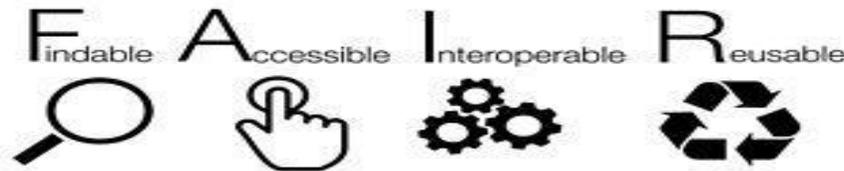
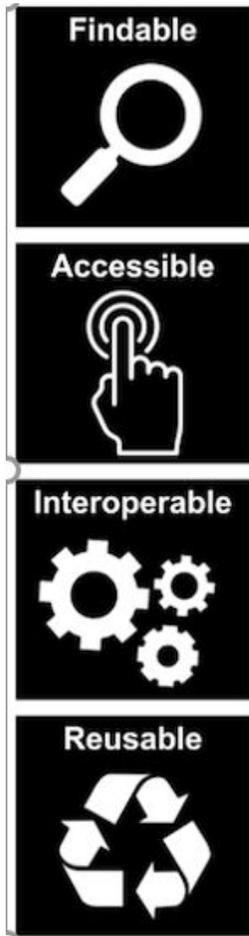
- transparence, réutilisation, participation, coopération, responsabilité, reproductibilité
- impartialité, équité, partage

❖ Pratiques :

- accès ouvert aux travaux de recherche
- partage des données de recherche
- transparence de l'évaluation de la recherche
- transparence des méthodes de recherche
- ouverture des codes-sources
- science citoyenne
- ressources éducatives ouvertes



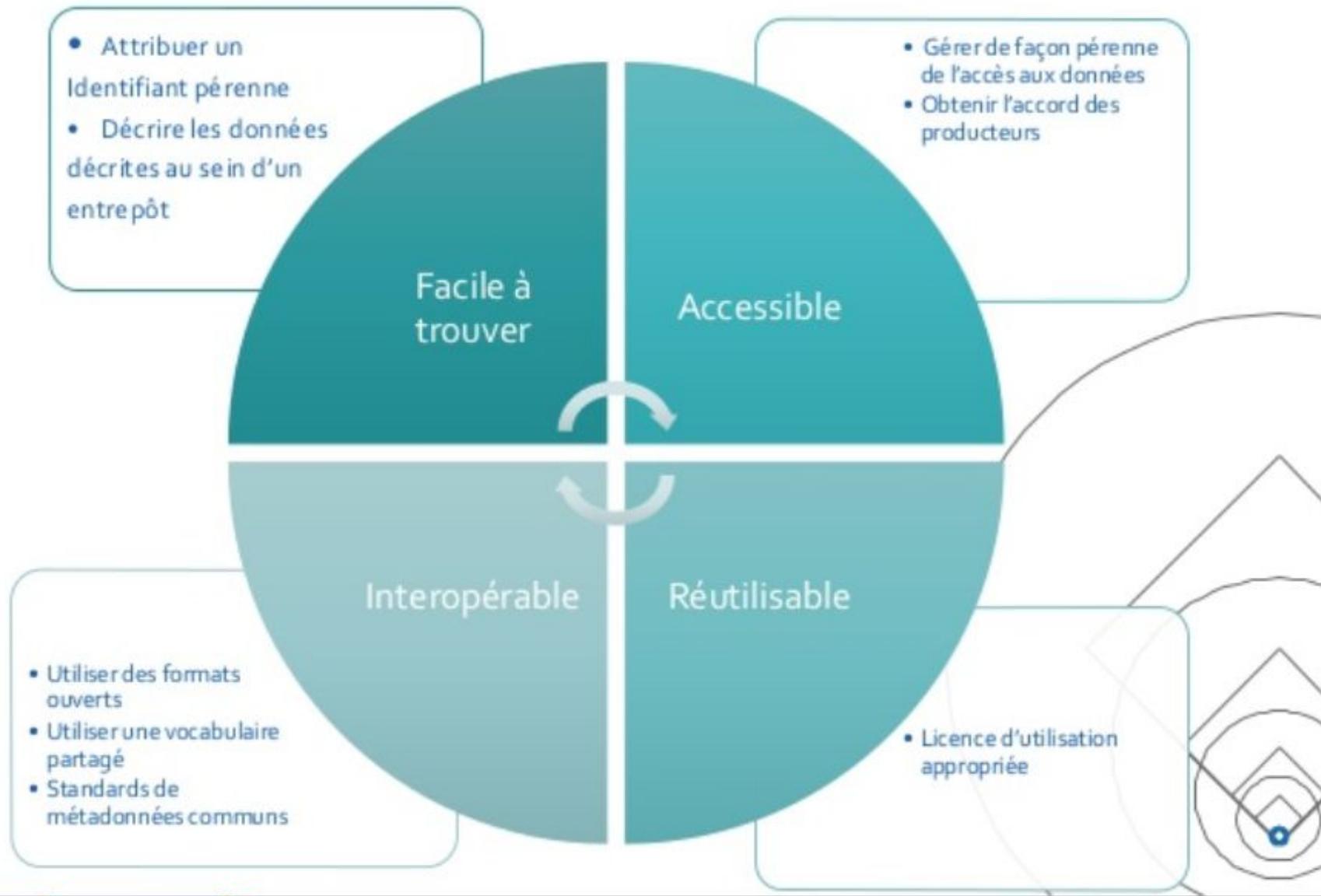
Ouvrir, c'est bien, diffuser, c'est mieux



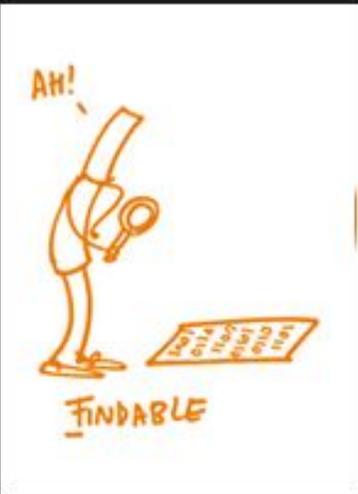
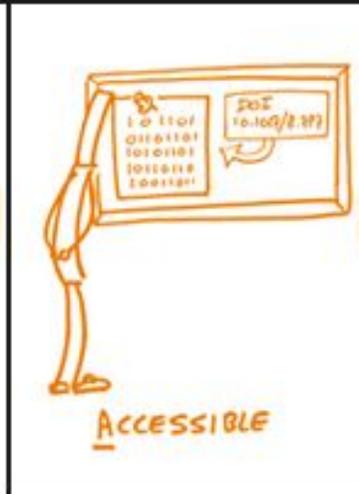
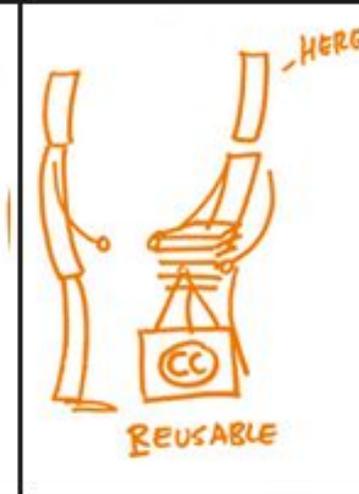
Faciles à trouver, repérables par des moteurs de recherche ;
Accessibles, atteignables au moyen d'outils et processus répandus (interrogeables par les protocoles standards d'internet) et selon des politiques claires (appartenance à un établissement...);
Interopérables, consultables et réutilisables par un maximum de monde sans barrière technique ou linguistique (ex.: traduction automatisée, conversion de fichiers, traitement simultané de données...);
Réutilisables, dont les conditions de collecte, d'analyse, de traitement et de rediffusion soient clairement définies.

□ Les ressources scientifiques doivent être exploitables par des humains, mais surtout, par des machines / robots informatiques.

Principes FAIR : Augmenter le potentiel des données



Pratiques diverses suivant les disciplines / Respect du droit

			
<ul style="list-style-type: none"> identifiant pérenne métadonnées enrichies données recherchables et trouvables en ligne 	<ul style="list-style-type: none"> données récupérables en ligne par des protocoles standardisés restrictions si nécessaire 	<ul style="list-style-type: none"> formats et standards communs vocabulaires contrôlés 	<ul style="list-style-type: none"> données documentées (pex. avec un fichier README) licence et information de provenance claires
<p>Si le jeu de données ne peut pas être mis à disposition librement, il est recommandé qu'au minimum sa description soit librement accessible en ligne.</p>	<p>Dans un repository qui stocke les données de manière sécurisée</p>	<p>Pour être compatibles et combinables avec d'autres jeux de données</p>	<p>Afin que les données soient interprétées correctement</p>



Les actions de la science ouverte

Libre accès aux publications	Données scientifiques ouvertes	Communication scientifique ouverte
<ul style="list-style-type: none">• Revues en libre accès• Archives ouvertes• Licences libres• Politiques de dépôt des publications• Loi(s) en faveur du libre accès• Gouvernance des communs scientifiques	<ul style="list-style-type: none">• Infrastructures de recherche• Web sémantique• Extraction automatisée de textes et de données• Dépôt de données• Dépôt de code	<ul style="list-style-type: none">• Évaluation ouverte• Plateformes collaboratives• Blogs scientifiques• Science citoyenne• Évaluation transparente

La science ouverte pour le chercheur

Des principes

- Transparence des pratiques
- Visibilité / Citabilité
- Réutilisation / Reproductibilité
- Coopération
- Partage
- Lutter contre les fraudes scientifiques
- Confiance des citoyens dans la science

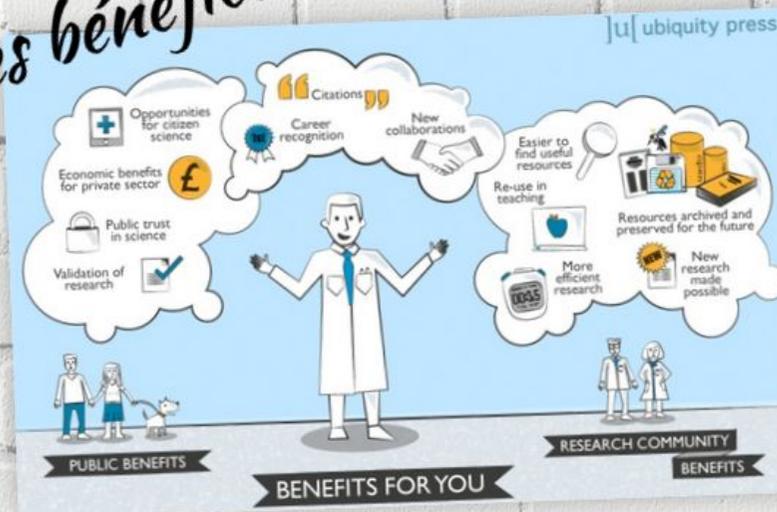
Des pratiques

- Libre accès aux publications scientifiques
- Partage / Accessibilité des données
- Réflexe professionnel
- Méthodologie transparente
- Évaluation ouverte
- Création de formations et services

Ma stratégie

- Organiser mon travail
- Acquérir de nouvelles méthodes
- Préserver / Protéger mes travaux
- Gagner du temps
- Construire mon identité numérique
- Me spécialiser / Me positionner
- Développer de nouvelles compétences
- Maîtriser le paysage de la recherche

Mes bénéfices



Les grandes dates de la science ouverte

- ✓ [DATES CLÉS DE LA SCIENCE OUVERTE](#)
- ✓ [HISTORIQUE DU LIBRE ACCÈS](#)

« Par “accès libre” à cette littérature, nous entendons sa mise à disposition gratuite sur l’Internet public , permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s’en servir de données pour un logiciel, ou s’en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l’accès et l’utilisation d’Internet. » (BOAI)

Années 1990 :
une naissance
pleine de
promesses

Prises de position (ex.:
Proposition subversive)
Premières plateformes
Protocole OAI-PMH

Années 2000 :
la fougue de la
jeunesse

Licences de réutilisation
Engagement politique
Répertoires d’outils
Participation des financeurs

Vers un écosystème
open global ?
Législation *open data*

Années 2010 :
l’entrée dans l’
âge adulte

De plus en plus de soutien
institutionnel et (inter)national
Premières contestations
 Crainte de dérives possibles

- Budapest Open Access Initiative (2002)
- Déclaration de Berlin (2003)

Années 2020 :
la maturité ?



Les années 1990 : une naissance pleine de promesses



Essor d'internet + volonté de donner un accès en ligne et gratuit aux publications scientifiques :

- **1991** : création d'arXiv par Paul Ginsparg : entrepôt centralisé pour les échanges de papiers entre physiciens
- **1994** : “*Subversive Proposal*” de Stevan Harnad : les articles devraient être gratuitement accessibles en ligne dès leur publication
- **1997** : naissance de SciELO, archive ouverte d'Amérique latine
- **1999** : réunion de Santa Fé :
 - mise au point de l'OAI-PMH (*Open Archive Initiative Protocole for Metadata Harvesting*)
 - création du premier logiciel d'auto-archivage gratuit, Eprint

Les années 2000 : la fougue de la jeunesse

- 2000 : naissance de PubMed Central
- 2001 : naissance de HAL, Creative Commons et PLOS
- 2002 : Budapest Open Access Initiative (BOAI)
- 2003 : Déclaration de Berlin sur le libre accès à la connaissance, Déclaration de Béthesda pour l'édition scientifique en libre accès
- 2004 : déclaration OCDE sur le libre accès aux données de la recherche
- 2006 : Lyon-1 en faveur de l'open access. Recommandations de la Commission européenne et projet DRIVER/OpenAIRE
- 2007 : L'ANR se prononce en faveur de l'OA, incite au dépôt dans HAL
- 2008 : 1er « mandat de dépôt » européen, à l'université de Liège

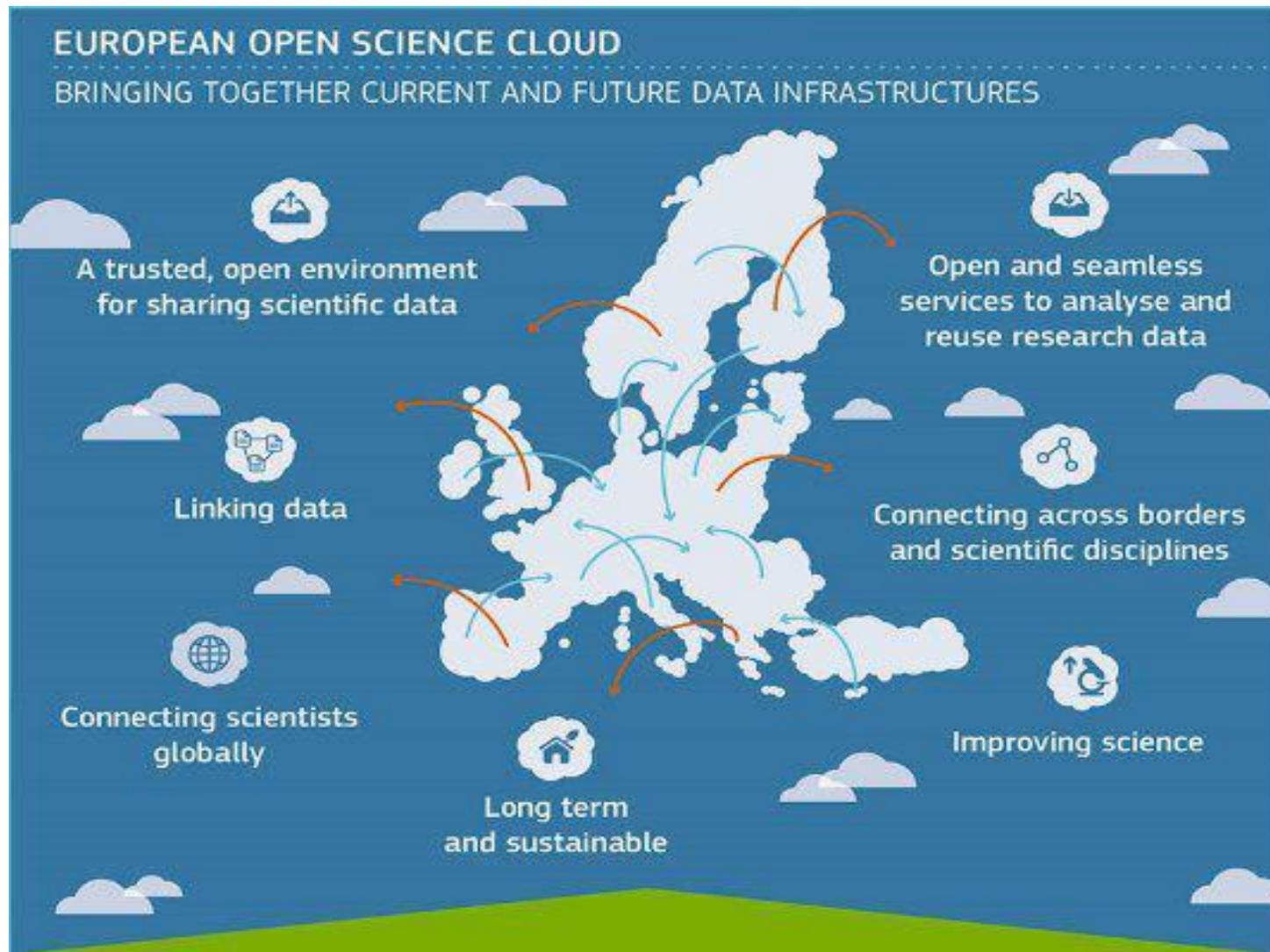


Les années 2010 : l'entrée dans l'âge adulte

- **2010** : premier « mandat de dépôt » français, à l'IFREMER
- **2011** : *Research Works Act* (USA)
- **2013** : archive ouverte avec « mandat de dépôt » à l'univ. d'Angers
- **2013** : *Fair Access to Research Act* (USA)
- **2013** : Réitération des demandes de la Commission européenne (modèle général de convention de subvention)
- **2016** : Loi pour une République numérique (France)
- **2017** : Appel de Jussieu pour la bibliodiversité
- **2018** : Plan national pour la science ouverte (France)
- **2018** : Plan S



Les années 2020 : la maturité ?

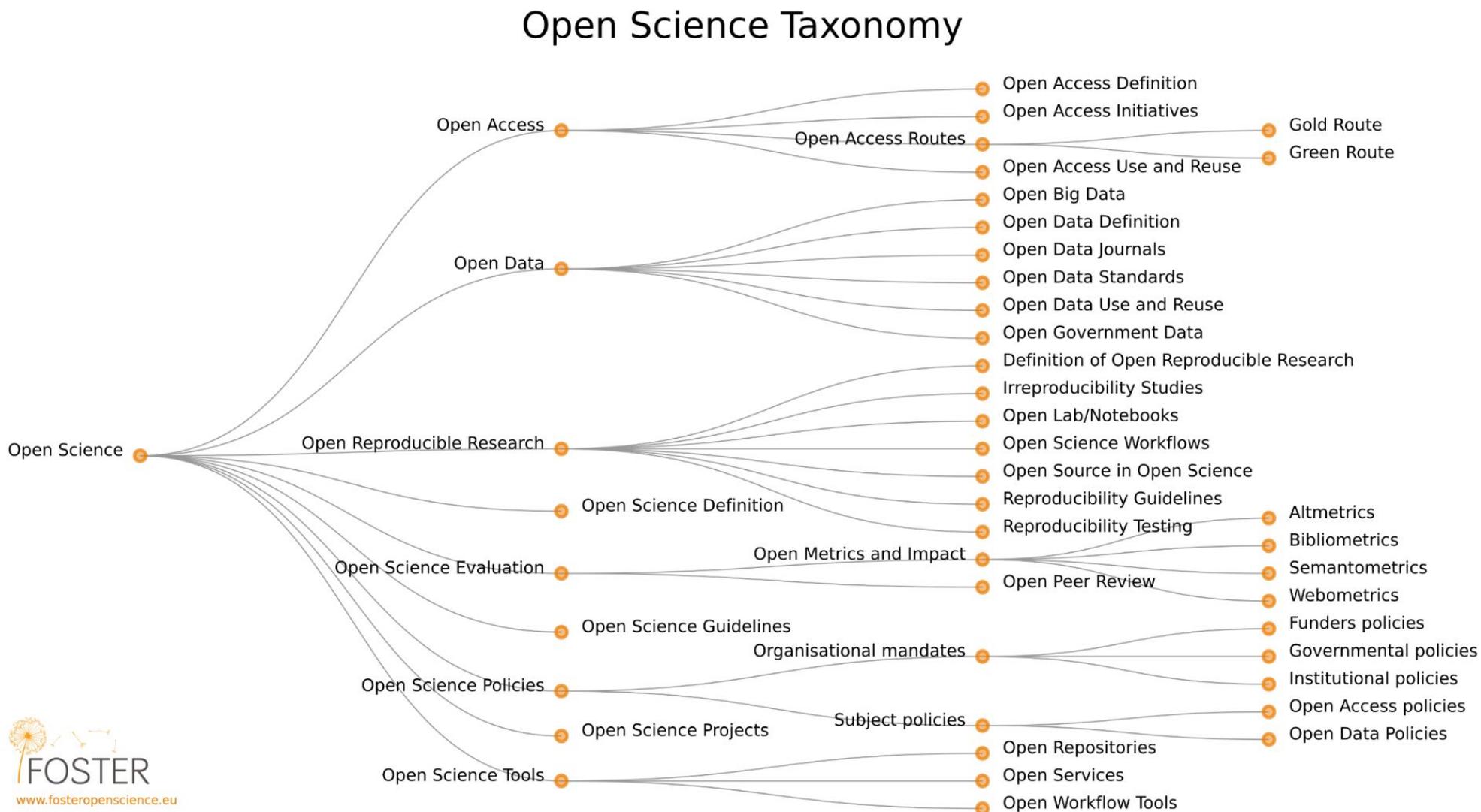


L'open science au cœur des préoccupations scientifiques actuelles : « **Open science is just science done right** »
(Jonathan Tennant)

Les composantes de la science ouverte



La terminologie de la science ouverte



Qu'est-ce que l'accès ouvert ?



« Le libre accès à la littérature scientifique est sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. »

Source

Génèse de l'Open Access

Définition



Selon l'[Initiative de Budapest pour l'Accès Ouvert](#), « le libre accès à la littérature scientifique est sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer (...), ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. » (d'après [Libre Accès Inist](#)).

Le **libre accès** (*open access*) à l'information scientifique recouvre deux réalités distinctes : **les revues en libre accès** et **les archives ouvertes** (*open archive*). Il repose sur des standards, des outils et des plates-formes en Open Source (logiciels libres).

Les textes fondateurs

- [Budapest Open Access Initiative \(BOAI\)](#) - 14 février 2002.
- [The Bethesda Statement on Open Access Publishing](#) - 20 Juin 2003.
- [La Déclaration de Berlin sur le Libre Accès à la connaissance en Sciences et Humanités](#) - 22 Octobre 2003.

Le problème de la double addition



Les éditeurs commerciaux à but lucratif augmentent le prix des abonnements aux revues depuis les années 2000. Les établissements n'ont plus les moyens d'avoir accès aux résultats pourtant produits par leurs propres chercheurs...

Les voies de l'accès ouvert

Voie dorée (*Gold Open Access*) : publier des travaux scientifiques qui sont immédiatement rendus disponibles en libre accès. C'est l'éditeur commercial (le *publisher*) qui se charge de cette dissémination.

Voie verte (*Green Open Access*) : Rendre des travaux scientifiques disponibles par une voie parallèle à l'édition scientifique traditionnelle. C'est le chercheur lui-même (ou son institution) qui se charge de cette dissémination : on parle aussi d'auto-archivage.



LA SUBVENTION

FINANCEMENT PRIS EN CHARGE PAR UNE INSTITUTION, UNE SOCIÉTÉ SAVANTE, UN ORGANISME DE RECHERCHE, UNE ASSOCIATION OU UNE FONDATION.



CONCERNE 66% DES REVUES FINANÇÉES PAR LE **Doaj** (Pris de 10.000 revues!)

LE FREEMIUM

CE SONT LES SERVICES COMPLÉMENTAIRES QUI SONT COMMERCIALISÉS



EXEMPLE: **Open Edition**

LA SOUSCRIPTION

LA DIFFUSION EN OA SE FAIT QUAND LE SEUIL FINANCIER EST ATTEINT
ex: *Knowledge Unlatched*, *Open Library of Humanities*

OU ENCORE LE

CROWDFUNDING



LE MODÈLE AUTEUR - PAYEUR



A.P.C

ARTICLE PROCESSING CHARGES

PAVÉS PAR L'AUTEUR OU SON INSTITUTION



COÛT À L'ARTICLE OU ADHESIONS INSTITUTIONNELLES OU INDIVIDUELLES

DE QUELQUES CENTAINES À PLUSIEURS MILLIERS D'EUROS

exemple: **Plos**



DÉRIVES TARIFAIRES DE CERTAINS ÉDITEURS



FONT PAYER LES AUTEUR MAIS N'OFFRENT AUCUNE EXPERTISE



REVUES DITES "HYBRIDES" DONT L'ACCÈS EST SOUMIS À UN ABONNEMENT

WTF ?!
L'INSTITUTION PAYE PLUSIEURS FOIS POUR L'ACCÈS AU DOCUMENT

LA VOIE D'ORÉE

FINANCEMENT EN AMONT → LA PUBLICATION EST EN OPEN ACCESS



DÉFINISSANT UNE POLITIQUE D'ÉTABLISSEMENT CONCERNANT LES APC

REFUSER LES MODÈLES HYBRIDES

FIXER UN MONTANT D'APC MAXI



+ RÉSERVE FINANCIÈRE CENTRALISÉE

EN ...

SOUTENANT DES MODÈLES OUVERTS ET DURABLES



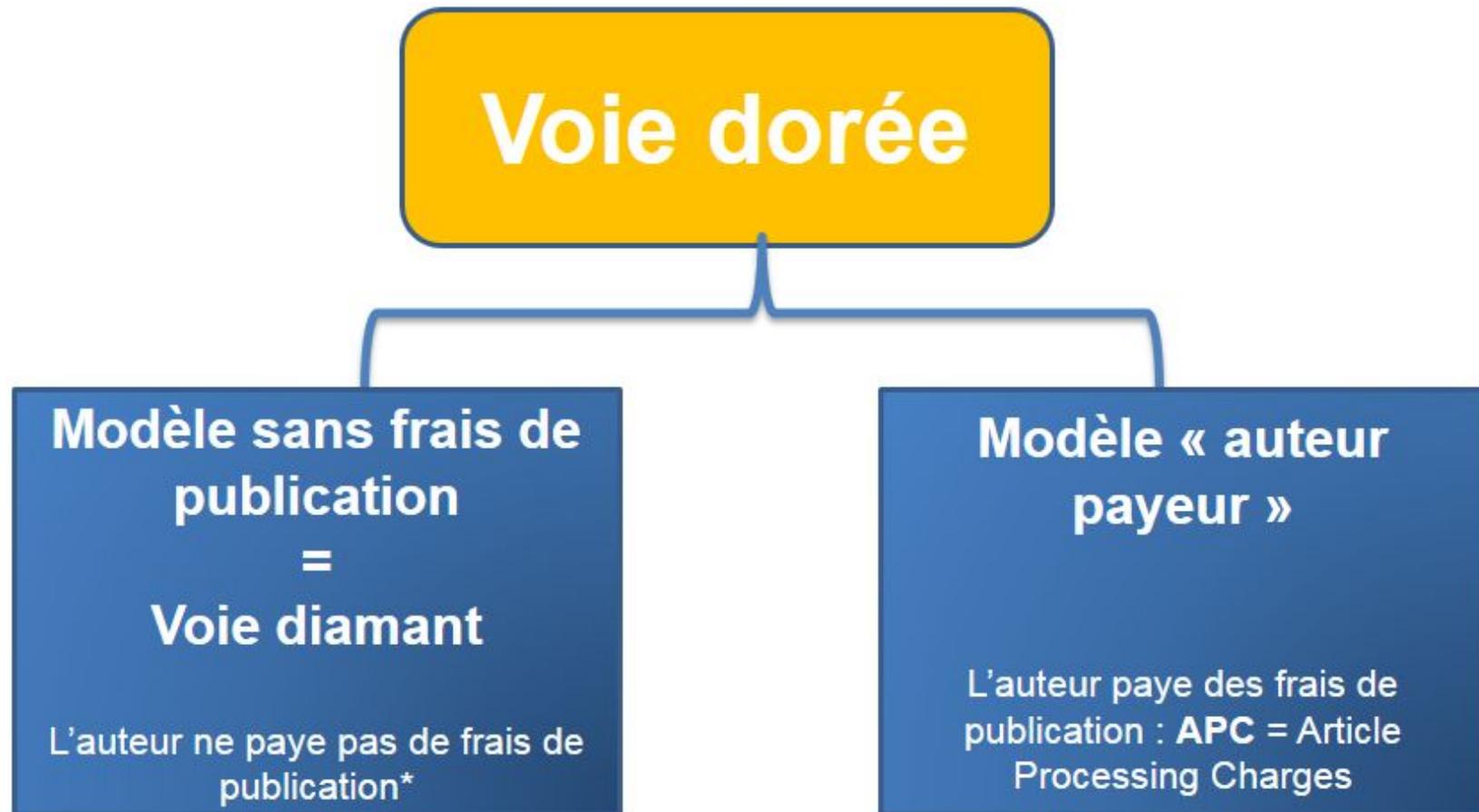
IDENTIFIER LES € DÉPENSES POUR LES APC!

ŒUVRANT À LA TRANSFORMATION DU MODÈLE DES PRESSES UNIVERSITAIRES

exemple: **Collège de France**, **l'université de Göttingen**

@magalufall

Les déclinaisons de la voie dorée



* Près de 75% des publications en libre accès ne facturent aucun frais de publication aux auteurs. Source : DOAJ.

Les déclinaisons de la voie dorée

Voie platine ou modèle *freemium* : l'accès au texte des publications est libre, mais les services complémentaires fournis par les éditeurs (texte mis en page ou enrichi de liens, statistiques de consultation...) sont payants.

Voie bronze : les publications sont librement accessibles, mais on ignore sous quelles conditions elles peuvent être réutilisées. Il s'agit d'accès ouvert, mais pas d'accès libre.



Identifier les ressources pour la voie dorée

Le répertoire DOAJ (*Directory of Open Access Journals*) recense et donne des indications (présence d'un comité de lecture, montant des APC, droits des auteurs...) sur plus de 16 000 revues en libre accès

Le répertoire DOAB (*Directory of Open Access Books*) fait de même pour plus de 40 000 ouvrages académiques relus par les pairs
Les bases de données bibliométriques traditionnelles (Web of Science, Scopus...) se sont récemment munies d'un **filtre *Open Access*** permettant d'identifier les revues concernées

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS



Les voies de l'accès ouvert

Voie dorée : publication (gratuite pour l'auteur) dans des revues en libre accès (gratuite pour le lecteur) . Abonnement des biblio

Voie verte (auto-archivage) : dépôt des articles par les chercheurs dans des réservoirs dédiés (archives ouvertes)

Voie hybride ou auteur-payeur \$ publication (payante pour l'auteur) dans des revues en open access (gratuite pour le lecteur)





AMI CHERCHEUR,
C'EST VOUS QUI DÉPOSEZ

#AUTO-ARCHIVAGE

VOUS GARDEZ
Tous
VOS DROITS SUR
LE DÉPÔT

#DROIT DAUTEUR

ACCÈS
LIBRE
GRATUIT

LA VOIE
Verte

VOUS POUVEZ DÉPOSER
UNE LARGE VARIÉTÉ DE
DOCUMENTS DE RECHERCHE
SCIENTIFIQUE

#CONTENUS



SI VOTRE DOCUMENT
A ÉTÉ PUBLIÉ PAR UN ÉDITEUR,
UNE PÉRIODE
D' **EMBARGO**



ENTRE 6 ET 12 MOIS
PEUT ÊTRE IMPOSÉE

#DIFFUSION

UNE FOIS DÉPOSÉS DANS
UNE ARCHIVE OUVERTE,
VOS DOCUMENTS
SONT ARCHIVÉS À
LONG-TERME

**#ARCHIVAGE
PERENNE**

@magablogall

Trouver une plateforme d'auto-archivage

2 outils principaux pour identifier les plateformes d'auto-archivage et prendre connaissance de leur fonctionnement :

- OpenDOAR (Directory of Open Access Repositories)
- ROAR (Registry of Open Access Repositories)

Sur HAL liste des portails institutionnels ou disciplinaires qui composent cette archive ouverte



HAL
Documentation

Les voies « pirates » de l'accès ouvert

Échange de fichiers : des chercheurs empruntent les identifiants de collègues pour accéder à des ressources sur abonnements, ou partagent + ou - publiquement sur internet leur besoin d'accéder à telle ressource payante pour qu'ils leur envoient le fichier PDF.

Sci-Hub : cette plateforme propose des ressources scientifiques ordinairement en accès payant, mais qui ont été piratées sur les serveurs des éditeurs à l'aide d'identifiants légalement obtenus par des établissements d'ESR.

Partage via des réseaux sociaux académiques : des chercheurs font le choix de mettre leurs propres publications en accès ouvert à partir de leur profil utilisateur sur des réseaux sociaux académiques (ResearchGate ou Academia). Ces pratiques ne sont pas toujours autorisées par les éditeurs.



L'ouverture des publications

Avec **la loi pour une République numérique de 2016** (article 30), tous les chercheurs ont le droit de mettre leurs publications en libre accès si :

- Ces publications sont issues d'une activité de recherche financée au moins pour moitié par des fonds publics ;
- Les écrits concernés ont été publiés dans un périodique paraissant au moins une fois par an ;
- Un délai minimal entre la publication de ces écrits et la diffusion en libre accès est respecté (délai de 6 mois pour les sciences, techniques et médecine, et délai de 12 mois pour les sciences humaines et sociales).

□ Les publications ne peuvent pas faire l'objet d'une exploitation commerciale

□ Les éditeurs ne peuvent pas s'opposer à ce « droit d'exploitation secondaire »



Financeurs publics et Open Access

cOAlition S | 4

Pourquoi le Plan S ? Agences de recherche



- ⊙ Les agences de recherche de la cOAlition S veulent:
 - ⊙ accélérer la science en en mettant les résultats de la recherche immédiatement à la disposition de la plus grande audience possible globalement.
 - ⊙ une transparence plus grande dans la communication scientifique
 - ⊙ une transition économique du modèle d'abonnement (non-durable) vers un modèle de publication en accès ouvert.
 - ⊙ utiliser leurs fonds de recherche pour diriger la publication scientifique vers l'accès ouvert complet et immédiat.



Financeurs et science ouverte

10 principes du Plan S



1. Les auteurs conservent le droit d'auteur sur leurs publications, qui doivent être publiés sous une licence libre de droits (Creative Commons, de préférence CC-BY) ;
2. Les membres de la cOAlition établissent des critères et prérequis pour déterminer la conformité du libre accès des revues et des plateformes ;
3. Ils doivent inciter à la création de revues et plateformes en libre accès conformes, si elles n'existent pas encore, et les soutenir financièrement ;
4. Les frais de publication doivent être payés par les bailleurs de fonds ou les universités, et non par les chercheurs ;
5. Les membres de la cOAlition soutiennent toutes les formes d'accès ouvert, y compris les archives et dépôts ouverts ; en cas de paiement de frais de publication, ceux-ci doivent être normalisés et plafonnés, proportionnels au service offert et transparents ;

Financeurs et science ouverte

10 principes du Plan S



6. Les universités, les organismes de recherche et les bibliothèques doivent aligner leurs politiques et stratégies en vue d'une plus grande transparence ;
 7. Les livres et chapitres de livres sont traités à part, avec un calendrier distinct ;
 8. Les revues en libre accès "hybrides" ne doivent être qu'une solution transitoire, dans le cadre d'accords de transformation ;
 9. Les membres de la coalition doivent surveiller et sanctionner la conformité avec le plan
 10. Les organismes de financement s'engagent à évaluer les résultats de recherche en fonction de la valeur intrinsèque du travail et non du canal de publication
- À partir de 2021, tous les articles de revue financés par les membres de la cOAlition devront être en accès libre.

L'ouverture des données et matériaux de recherche

Qu'est-ce qu'une donnée de la recherche ?

Périmètre très large :

Tout ce dont un chercheur aurait besoin pour vérifier un résultat scientifique, refaire un cheminement de pensée, reproduire une expérience

Tout ce qu'un chercheur manipule au cours de ses recherches, et dont la perte ou le vol aurait une incidence grave sur son travail

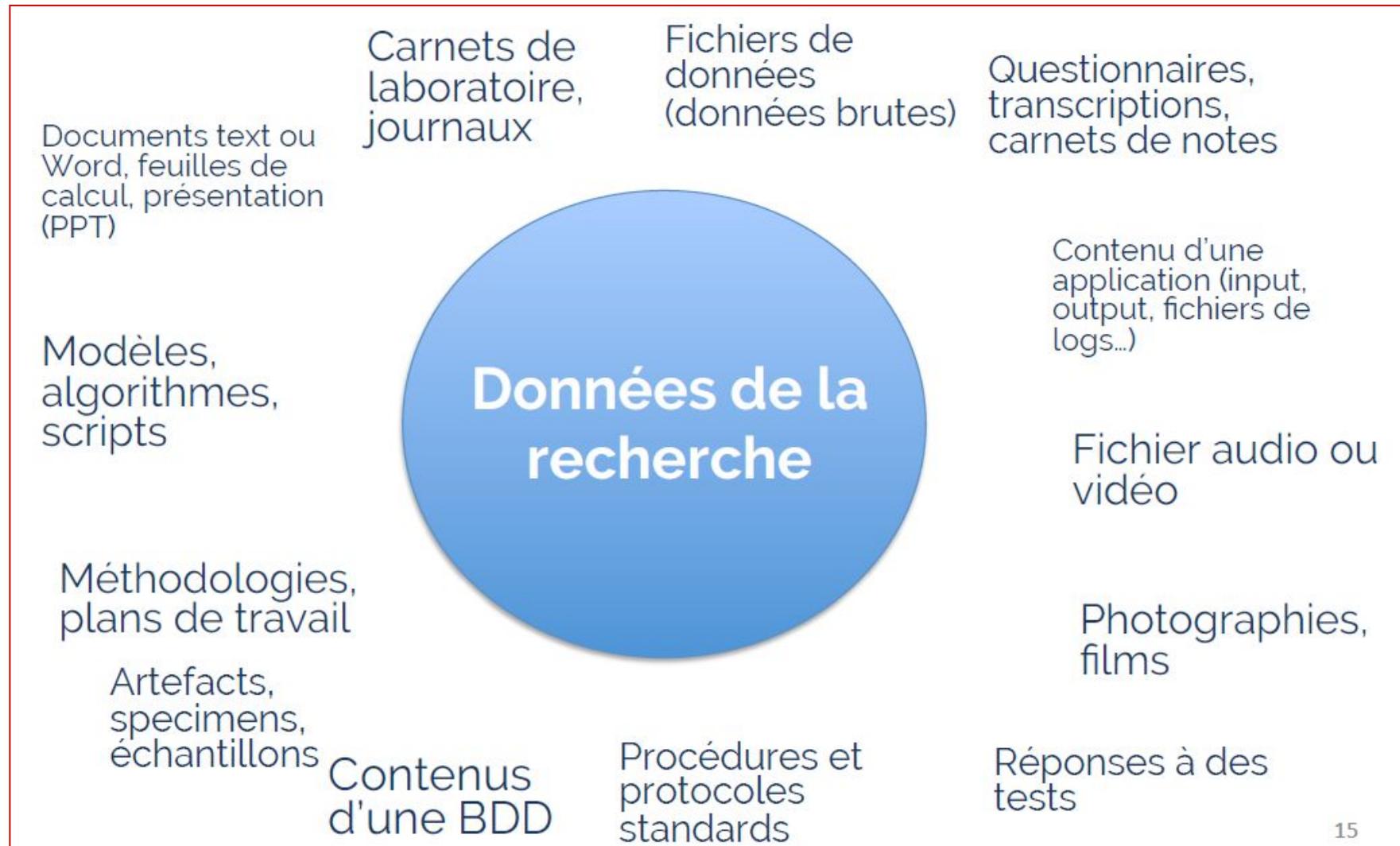
Il peut s'agir de :

Fichiers nativement numériques (textes, tableurs, images, vidéos, sons, programmes informatiques, objets 3D...)

Documents numérisés (archives, ouvrages, documents iconographiques, cartes...)

Supports analogiques de tout type (manuscrits, échantillons, prélèvements, réactifs, matériaux, œuvres d'art...)

Diversité des données de la recherche



Source : Marie Puren. *Gérer les données de la recherche*, 2018

Qu'entend-on par données de recherche ?

La définition de l'OCDE :

« Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche »

([OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007](#))

1

produites : ce sont les données créées, élaborées, générées lors d'activités de recherche (observations, mesures...)

2

collectées : ce sont des données déjà existantes (corpus, archives...) qui sont utilisées pour le projet. Les données utilisées peuvent avoir été recueillies initialement dans un autre contexte que celui de la recherche mais elles sont utilisées comme données de recherche dans le cadre du projet.

3

produites et collectées



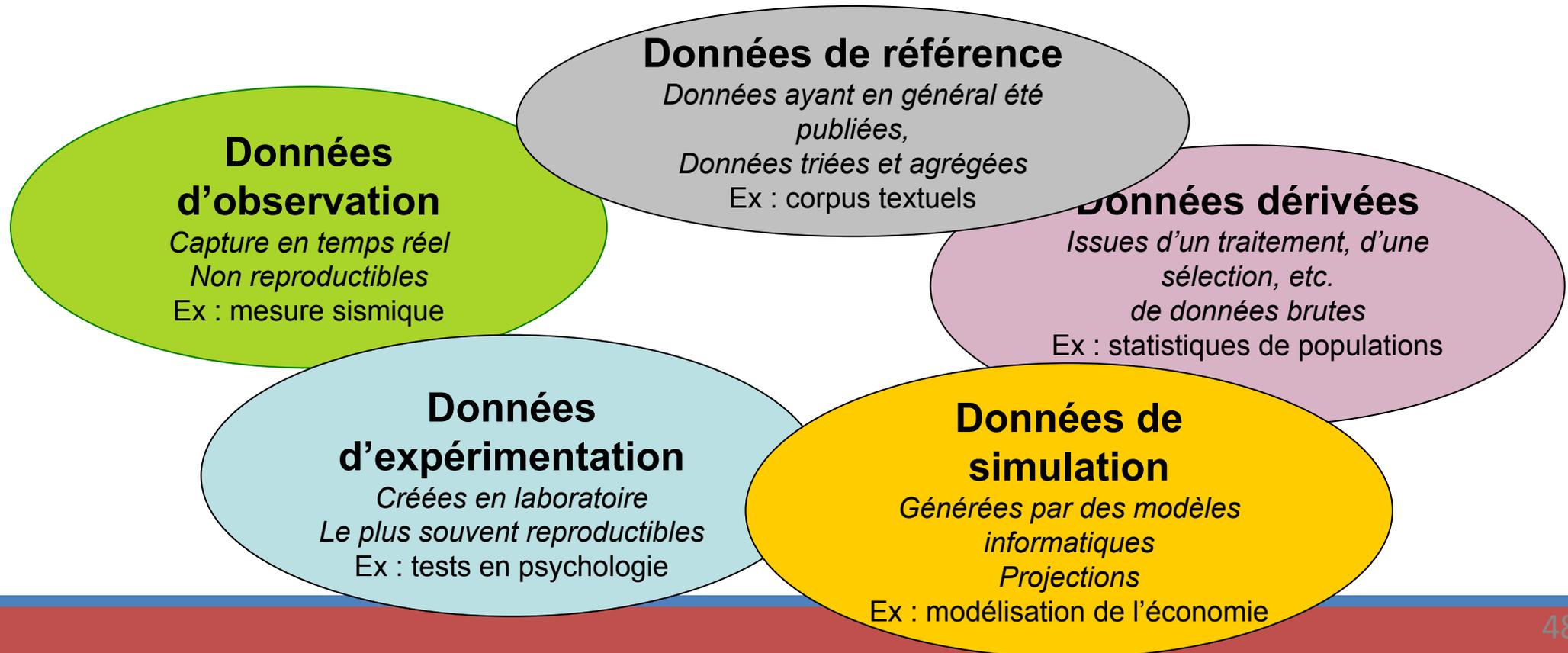
Typologie des données de la recherche

• Approches classificatoires

- Selon leur méthode d'obtention

(André, Francis. Déluge des données de la recherche ? Petit manuel d'immersion. Curation, infrastructures et partage. *Big Data*, 2014)

= Selon les « méthodes et instruments qui les ont générées »



Typologie des données de la recherche

Selon le stade de leur cycle de vie

- Données préliminaires ou préparatoires
- Données « brutes » (acquises lors du processus de recherche, « neutres »)
- Données traitées et analysées

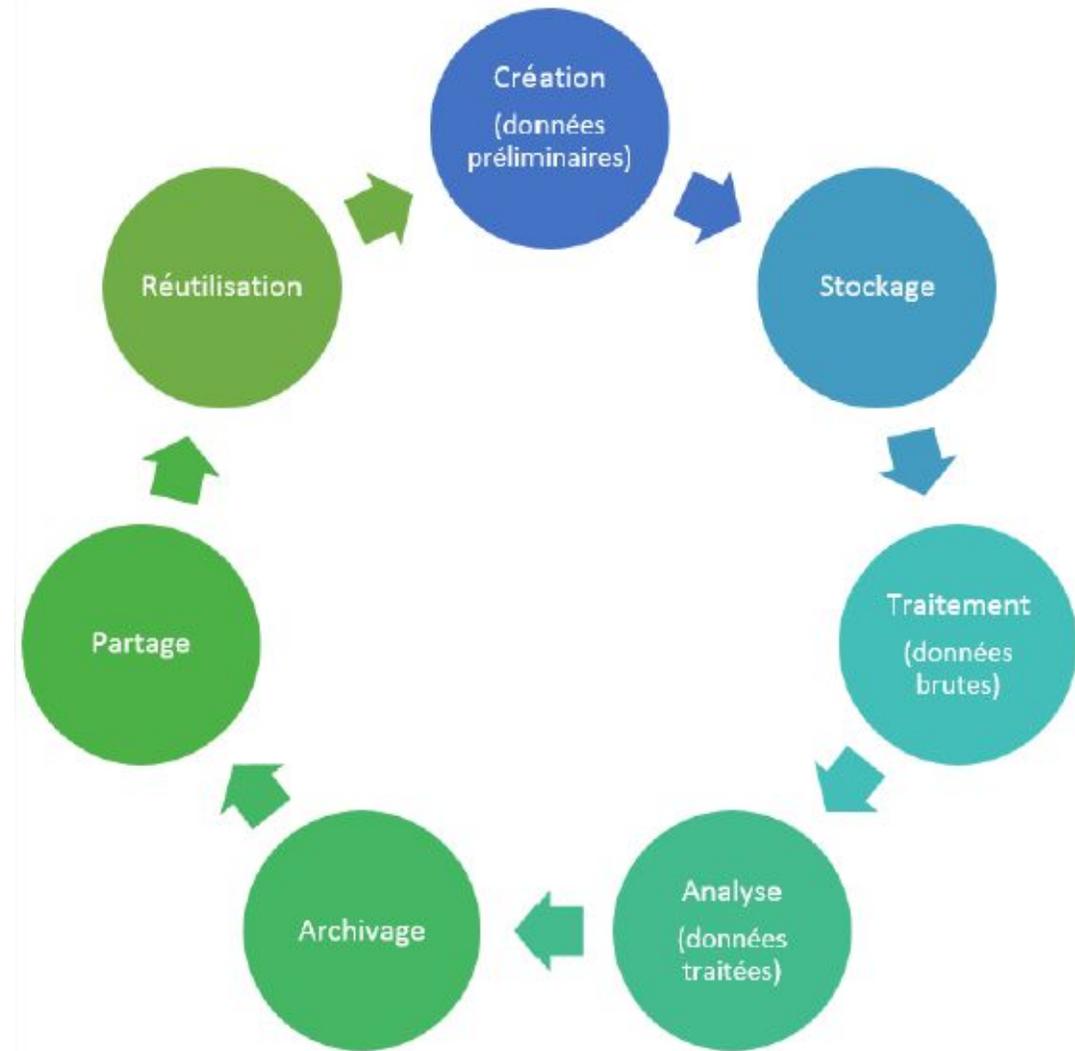


Figure 3 : Le cycle de vie des données de recherche³³

(Pain, Marielou. « Les données de la recherche et leurs entrepôts, de la documentation à la réutilisation : étude de cas pour l'archive HAL », 2016 et Schöpfel, Kergosien, Prost, 2017)

Typologie selon le cycle de vie des données

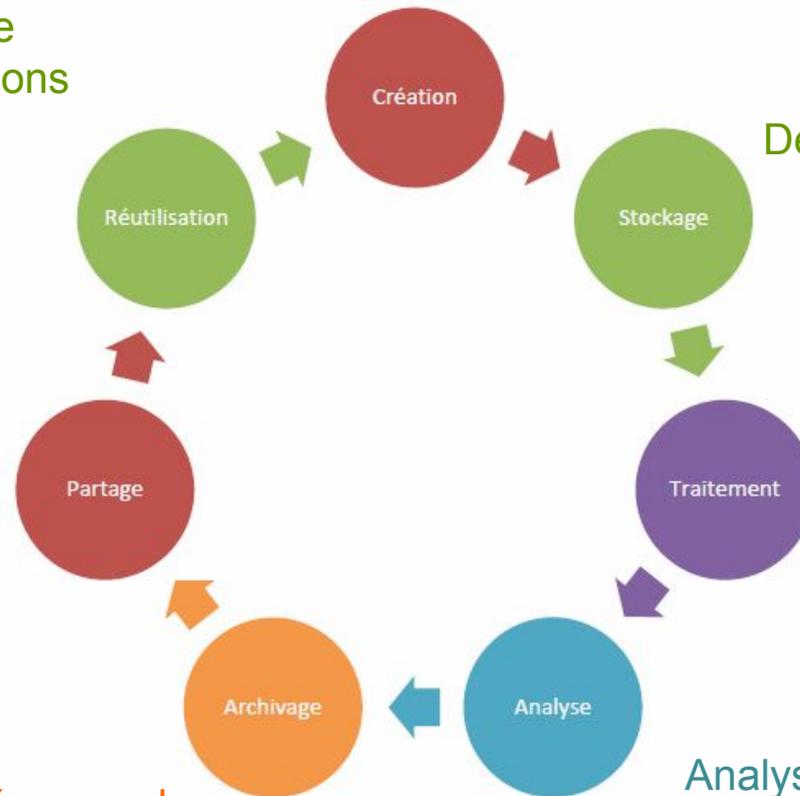
Réaliser des analyses secondaires, un suivi de recherche, une nouvelle recherche, des évaluations de la recherche...

Collecter / créer les données; capturer les données avec les métadonnées; acquérir des données existantes d'un tiers.

Définir un espace de stockage des données

Définir les droits d'auteur et d'utilisation; contrôler les métadonnées et la documentation ; faire des copies de sauvegarde; publier et partager les données.

Entrer, numériser, transcrire, traduire les données ; vérifier, valider nettoyer, anonymiser ; dériver des données ; décrire et documenter les données ; gérer et sauvegarder les données.



Préparer les données pour la conservation; migrer les données vers le meilleur format et un support adapté; les sauvegarder et les décrire dans un entrepôt de données; définir les accès et promouvoir les données archivées

Analyser et interpréter les données ; produire les résultats de recherche ; citer les sources de données.

Pourquoi ouvrir les données de la recherche ?

- Permettre leur **réutilisation par autrui**, développer de nouvelles hypothèses de travail ou des travaux de grande envergure
- Permettre une **évaluation plus fine des travaux scientifiques**, en reproduisant des expériences ou un raisonnement : renforce l'intégrité scientifique en mettant en évidence d'éventuelles fraudes
- **Mettre en valeur le travail des chercheurs**, qui ont souvent consacré beaucoup de temps à l'élaboration de leurs jeux de données

Enjeux liés à leur partage et leur réutilisation

- Permettre et faciliter l'accès le plus large
- Eviter les phénomènes de privatisation, d'appropriation des données
- Améliorer la qualité, la visibilité, l'impact de la recherche
- Favoriser les collaborations scientifiques demandant le partage ou l'élaboration collective de données

Il importe de :

- Documenter les données et les décrire de manière très précise
- Définir des conditions d'accès, des droits de réutilisation
- Contrôler la compatibilité, l'interopérabilité des formats, la lisibilité
- Procéder à une Dé-contextualisation – re-contextualisation

**« Aussi ouvert que possible, aussi fermé que nécessaire »
Les données peuvent relever de droits multiples**



Comment partager des données de recherche ?

DES DONNÉES PEUVENT ÊTRE PARTAGÉES

En étant incluses dans des **publications scientifiques** (figures, tableaux...)

En étant jointes en **annexes à des publications scientifiques** – on parle de *supplementary materials*)

En étant déposées dans des plateformes dédiées, qui vont s'occuper de la conservation et de la valorisation des données qui leurs sont confiées : des **entrepôts de données**

RENDRE SES DONNÉES ACCESSIBLES ET/OU RÉUTILISABLES

Par tous / Sous certaines conditions (demande écrite au producteur des données, réservé à un établissement...)
Uniquement par une ou plusieurs personnes (données relevant du secret industriel et commercial)

□ Ces options sont à préciser au moyen de licences de diffusion (outils juridiques), et à paramétrer au moyen d'outils techniques précis (fournis par les entrepôts)



Cadre juridique de l'ouverture des données

DIFFUSION DES DONNÉES PUBLIQUES

2 principes dans la loi française :

Open data : les données doivent obligatoirement être rendues publiques, ou du moins communiquées à tout citoyen qui en ferait la demande

Libre réutilisation : les données rendues publiques peuvent être réutilisées sans aucune condition.

EXCEPTIONS POUR CERTAINS TYPES DE DONNÉES

- données personnelles,
- données sous droit d'auteur
- données concernées par le secret industriel et commercial, (etc.).

Les contrats signés par les chercheurs avec des éditeurs sur leurs données ne peuvent pas aller à l'encontre de ces principes

Les logiciels libres

Logiciels et recherche scientifique

Le logiciel peut être à la fois :

- **Moteur de la recherche** :
- **Résultat de la recherche** : preuve d'existence d'une solution algorithmique à un problème donné ;
- **Objet de recherche** : dans des disciplines qui étudient les logiciels
- **Production de recherche** : création de logiciels venant en appui ou en démonstration de leurs publications.

[Recommandations Ouvrir la science](#)

Enjeux de la conservation et de la mise à disposition des logiciels

- **Droit d'auteur** : contributions apportées dans l'élaboration d'un logiciel.
- **Reproductibilité** : conditions d'obtention d'un résultat scientifique
- **Patrimoine** : conserver et valoriser les productions scientifiques que sont les logiciels



Missions du groupe Logiciels libres et open source du COSO

Inciter la communauté scientifique

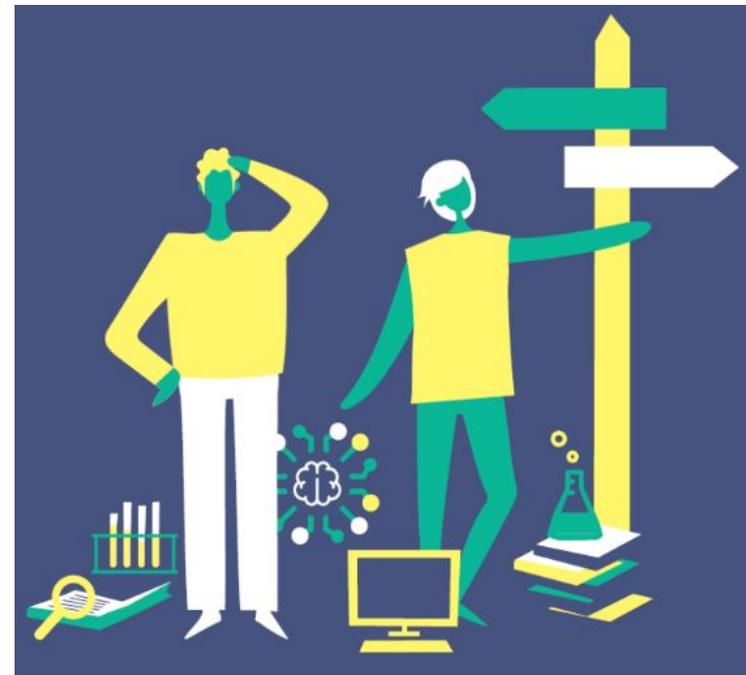
Former aux bonnes pratiques tout au long de la production de logiciels
Promouvoir le développement libre et ouvert dans les actions de soutien à la recherche (appels à projets, etc.);

Reconnaître l'implication des chercheurs :

Améliorer le référencement des logiciels libres ;
Produire de nouveaux indicateurs pour mesurer l'impact de ces logiciels.

Enjeux affichés par le CoSO :

Ouverture des résultats au plus grand nombre
Facilitation de la reproductibilité des résultats
Capitalisation du travail réalisé
Possibilité d'améliorer sans refaire



Ouvrir la recherche
en train de se faire

« Carnets de notes » ouverts

Ouverture des carnets de notes
(open-notebook science) :

- rendre accessible intégralité processus de recherche en train de se faire, et non à la fin du projet.

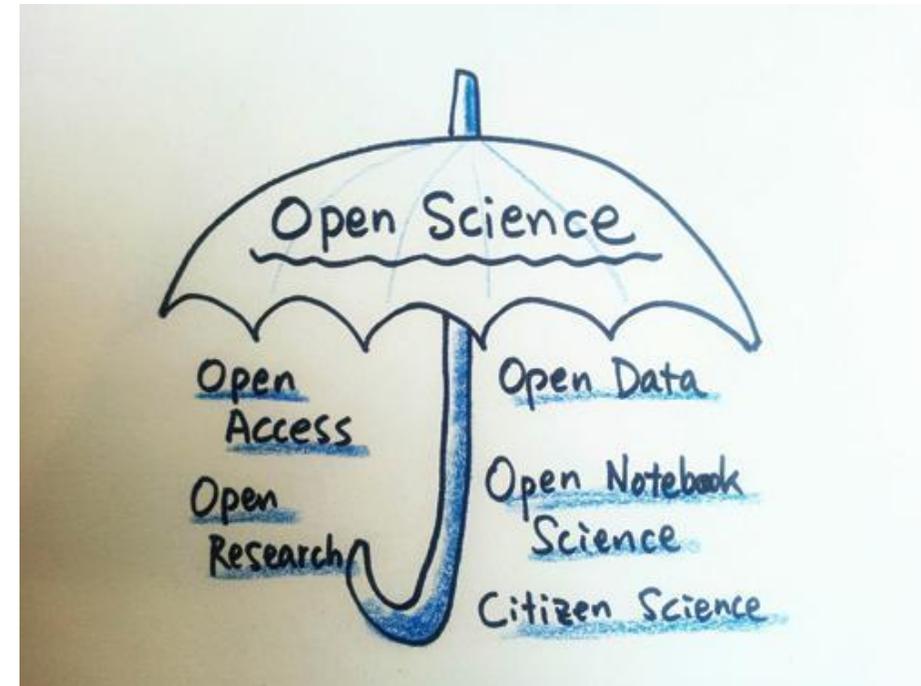
Rendre public au fur et à mesure de la recherche :

Les protocoles de recherche

Les carnets de notes personnels ou collectifs (laboratoires) des chercheurs

Les données brutes puis traitées

Les matériels annexes générés (codes, supports physiques...)



Transparence
Reproductibilité

Les cahiers de laboratoire ouverts : définitions

Le cahier de laboratoire est un outil de traçabilité des travaux de recherche, utilisé par les unités de recherche comme par les entreprises. Journal de bord quotidien, il permet d'enregistrer le détail précis du mode opératoire des expériences réalisées au sein d'un labo, de l'idée de départ au résultat de recherche (date, objectifs, contributions individuelles, paramètres...) ([source](#) : réseau CURIE).



C'est un outil documentaire au service de la reproduction scientifique, mais aussi de la protection juridique, puisqu'il peut servir de preuve dans la revendication d'une invention ([source](#) : [DATAACC](#)).

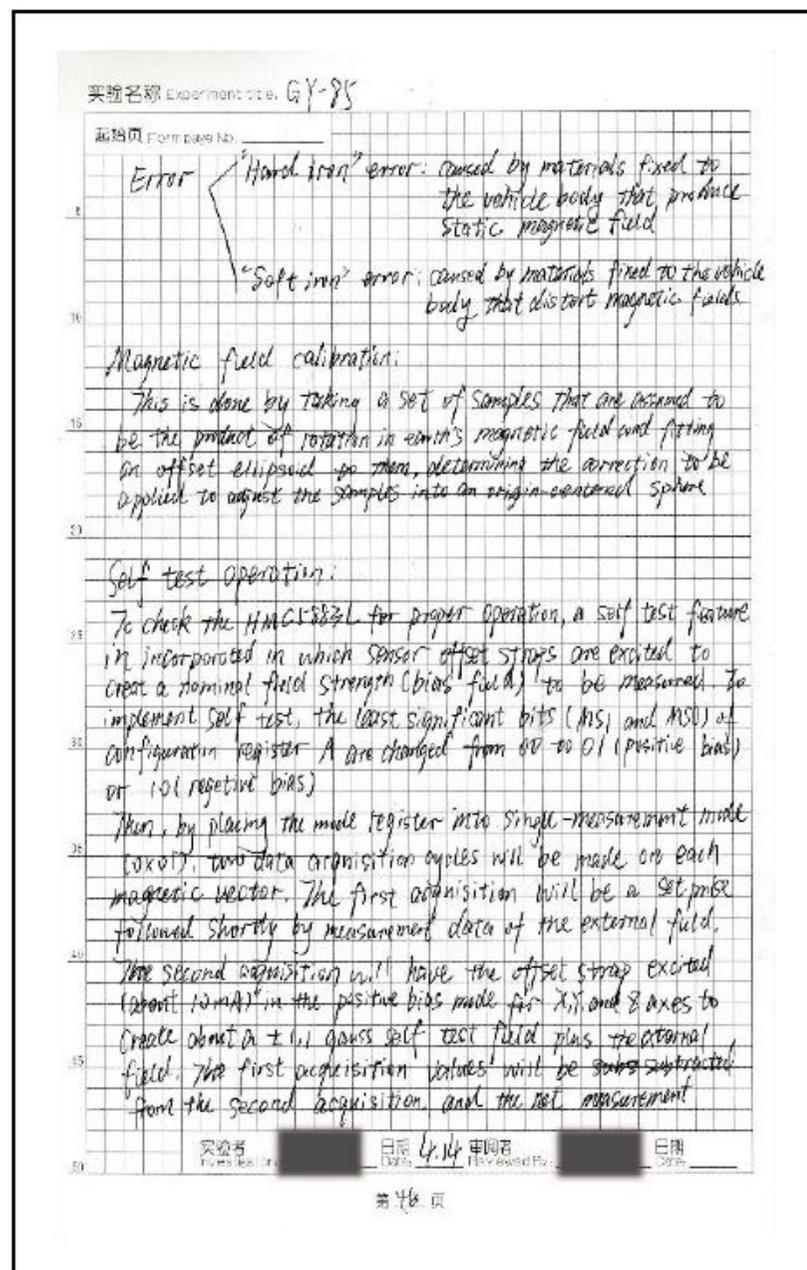


Fig. 2 Lab work ought to be recorded in a well maintained notebook. (a) An extract from a student's paper notebook.

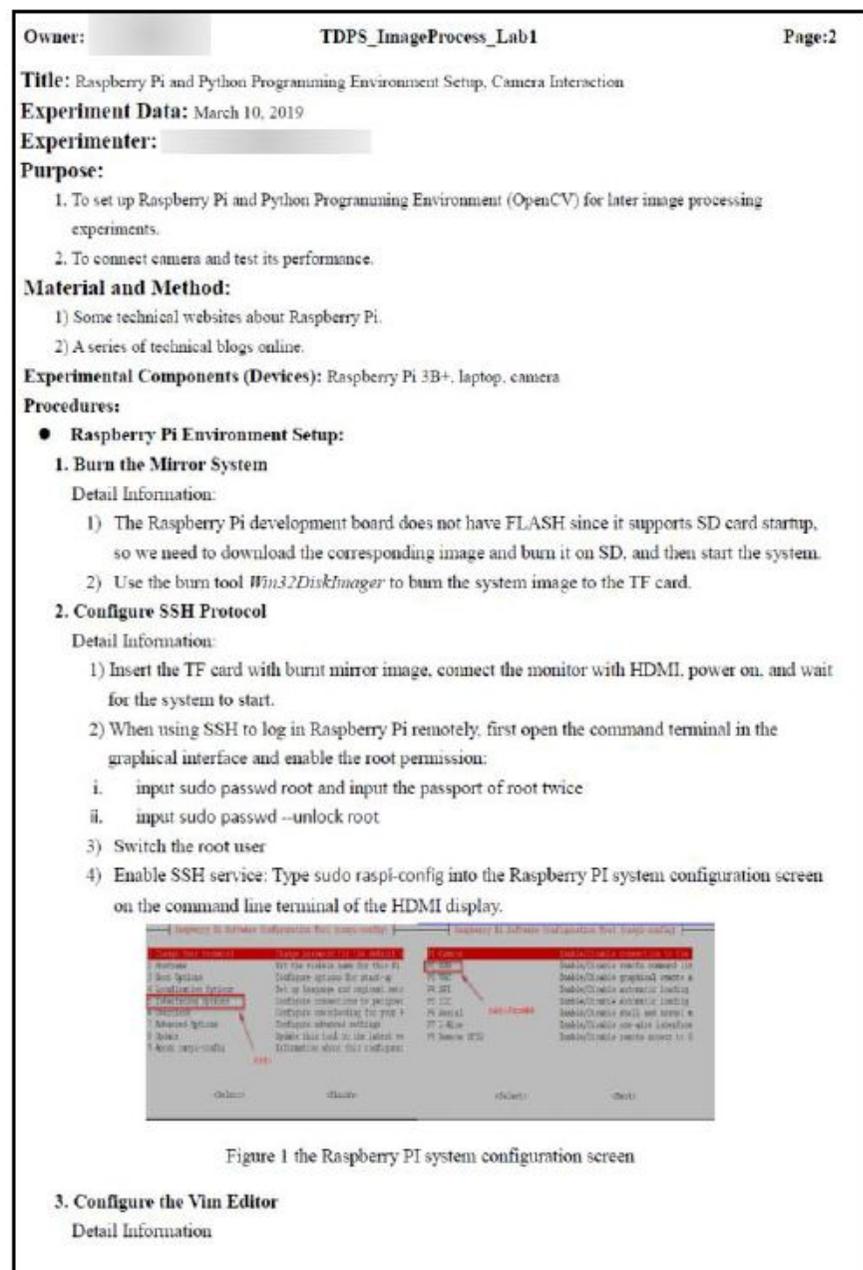


Fig. 3 An extract from a student lab notebook, which was developed using a dedicated ELN software product.

Les carnets de recherche ouverts

L'édition électronique ouverte Le carnet d'OpenEdition

Cléo – Revues.org | Édition électronique | Internationalisation | Publications | Tutelles et partenaires | Veille

Accueil | À propos | Calendrier du Cléo | Chroniques | Crédits | La Deuxième Université d'été en direct | TGE Adonis

hypotheses.org

Bienvenue sur le blog du Cléo

L'édition électronique ouverte est un blog animé par l'équipe du Centre pour l'édition électronique ouverte. Il vous informe de l'actualité du centre, de l'édition électronique, de l'information scientifique et des formations proposées.

L'affiche du manifeste des Digital Humanities



Des ISSN pour les carnets d'Hypotheses.org

22 juin 2011
Par Frédérique Muscinesi



Read this blog in English!

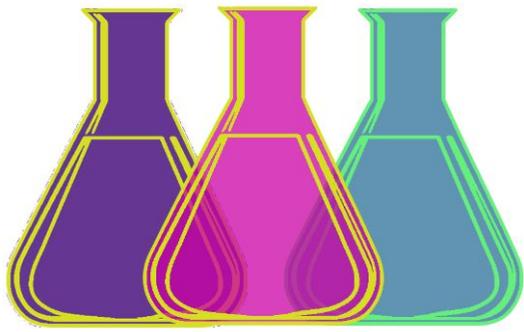


Open electronic publishing is the blog of the OpenEdition platform, developed by the Centre for Open Electronic Publishing (Cléo, France). This tool is designed to distribute information about the evolutions in and updates to the OpenEdition portal, about electronic publishing in the humanities and social sciences and about academic and technical issues. It also deals with the Cléo team's daily activities, its projects and profession.

Read/Write Book. Le livre inscriptible



Un carnet de recherche est un « mode de communication qui s'apparente au blog, mais dans un contexte professionnel de recherche et qui est complètement intégré aux pratiques de communication liées à la recherche » (P. Mounier, cité par M. Faury)



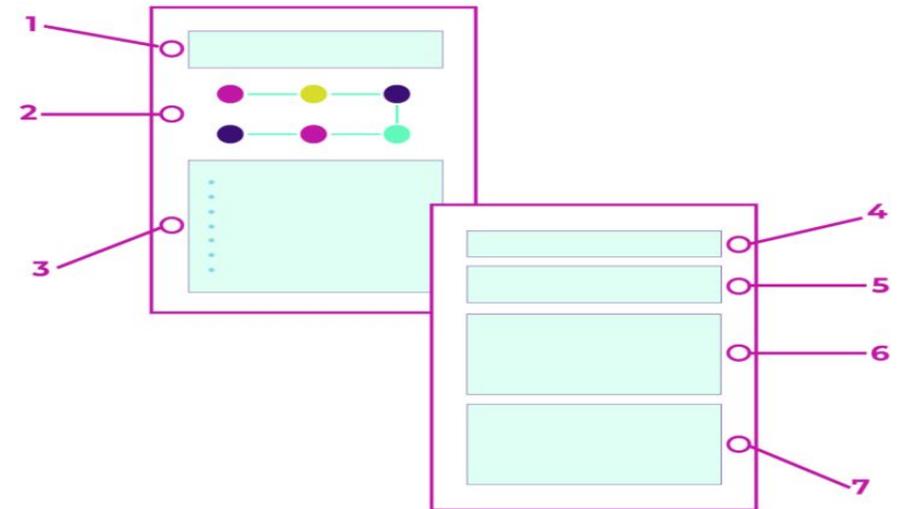
Laboratoires ouverts

□ Rendre public documents, matériels, informations produits/ collectés par chercheurs d'un laboratoire au cours de leurs recherches :

Projets possibles ou envisagés

Listes de références pour les demandes de financement, planification des travaux ou interprétation des résultats ;
Protocoles expérimentaux mis au point au sein d'une équipe ou d'un laboratoire ;

Données brutes, traitées ou analysées ;



1. Protocol title
2. Interactive features
3. Step-by-step instructions with measurements, reagents, and more

4. Article title
5. Introduction
6. Materials & methods
7. Expected results

□ communication d'informations sur les activités menées dans les projets :
Vulgarisation scientifique (revues ou sites)
Blogs scientifiques et carnets de recherches

Evaluation ouverte

Comment la science est-elle évaluée ?

Évaluation qualitative : lecture attentive de travaux scientifique.
Evaluation par les pairs (*peer review*)

Avant publication dans une revue ou au sein d'une collection d'ouvrages, afin de vérifier que les travaux sont originaux, intéressants et bien menés

Suppose de recueillir l'avis d'au moins deux chercheurs spécialistes du domaine

[Lien](#)



Évaluation quantitative : études des productions scientifiques sous l'angle mathématique et statistique.

Intéressant pour étudier de grands ensembles, pas de sens à l'échelle individuelle.

Beaucoup de raccourcis, de généralisations, pas toujours pertinentes

Ex. de métriques calculées : dénombrement des productions scientifiques, des citations reçues...⁶⁵

Une évaluation opaque

Évaluation qualitative : processus d'examen d'une publication par d'autres chercheurs experts du domaine (secret) :

Noms des relecteurs inconnus

Remarques sur le texte soumis ne sont pas publiques

Des abus (conflits d'intérêts non déclarés, vol d'idées...)

Évaluation quantitative : données de calcul des indicateurs sont produites par des acteurs privés

- Difficile de contrôler la qualité de ces données, et de reproduire les calculs

Step 1 (rate, 20 seconds)

open evaluation

EVALUATE A PAPER

paper: Smith D, Warden EL (2010) Replicated-template mod

desired impact: [impact factor]

Enter the impact factor of the hypothetical journal you feel this paper would be appropriate for.

Step 2 (sign, 10 seconds)

EVALUATE A PAPER

paper: Smith D, Warden EL (2010) Replicated-template mod

desired impact: 12

signed: [your name]

Step 3 (disclose, 30 seconds)

EVALUATE A PAPER

paper: Smith D, Warden EL (2010) Replicated-template mod

desired impact: 12

signed: Daniel B Baker

copublished with: [number of authors]

friends with: [number of authors]

acquainted with: [number of authors]

findings support my work: [yes, no]

Step 4 (review, open ended)

EVALUATE A PAPER

paper: Smith D, Warden EL (2010) Replicated-template mod

desired impact: 12

signed: Daniel B Baker

copublished with: 0

friends with: 0

acquainted with: 1

findings support my work: no

Ouvrir l'évaluation qualitative

Des éditeurs et revues proposent de :

Rendre publics les noms des évaluateurs des publications (*open identities*)

Rendre publics les rapports d'évaluation et les échanges avec les auteurs (*open reports*)

Permettre à ceux qui le souhaitent d'évaluer un manuscrit (*open participation*)

Déléguer le processus d'évaluation à un service tiers, indépendant (*open platforms*)



Des plateformes débattent publiquement de la teneur de travaux scientifiques :

Avant publication : discuter de documents « de travail » sur des plateformes comme [PeerCommunityIn](#)

Après publication : poursuivre la discussion après examen par les pairs, pour élargir le débat (plateformes comme [PubPeer](#))



Ouvrir la science
avec et pour la
société

Accès libre aux ressources éducatives

Les ressources éducatives libres (REL) sont des matériaux d'enseignement, d'apprentissage ou de recherche appartenant au domaine public ou publiés avec une licence de propriété intellectuelle permettant leur utilisation, adaptation et distribution à titre gratuit (Unesco).

- Supports de cours mis à disposition en libre accès
- MOOCs (*Massive Open Online Courses*), formations en ligne ouvertes



Exemples (BU de Montréal) :

- Vidéos, images, musiques
- Manuels et autres textes
- Outils d'évaluation
- Scénarios d'activités d'apprentissage
- Applications ou codes

Sciences citoyennes

Les sciences citoyennes (ou **recherche participative**, sciences avec et pour la société...) sont des formes de production de connaissances scientifiques auxquelles des acteurs non-scientifiques-professionnels, qu'il s'agisse d'individus ou de groupes, participent de façon active et délibérée (rapport Houllier, 2016).

Plusieurs étapes de la recherche scientifique, en lien avec les données, peuvent être ouvertes à la collaboration des citoyens :

Collecte de données (*ex.: comptage d'oiseaux sur une période donnée*)

Analyse de données (*ex.: identification de structures de galaxies à partir de l'examen de photographies astronomiques*)

Traitement de données (*ex.: pliage de protéines*)

**SCIENCES
CITOYENNES**



Production participative

La production participative (externalisation ouverte ou *crowdsourcing*), est l'utilisation du travail, de la créativité, de l'intelligence et du savoir-faire d'un grand nombre de personnes, en sous-traitance, pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur.

Prolongation des sciences participatives, elle peut permettre à des structures de recherche de disposer de matériel de recherche ouvert ou d'instruments développés en mode ouvert.



Avantages :

- Pour structures de recherche : accélérer processus, réduire des coûts
 - Pour citoyens : acquisition de compétences, création de lien avec la communauté scientifique, rémunération
- Pour tous : produire des ressources ouvertes

Ressources

2020
2021

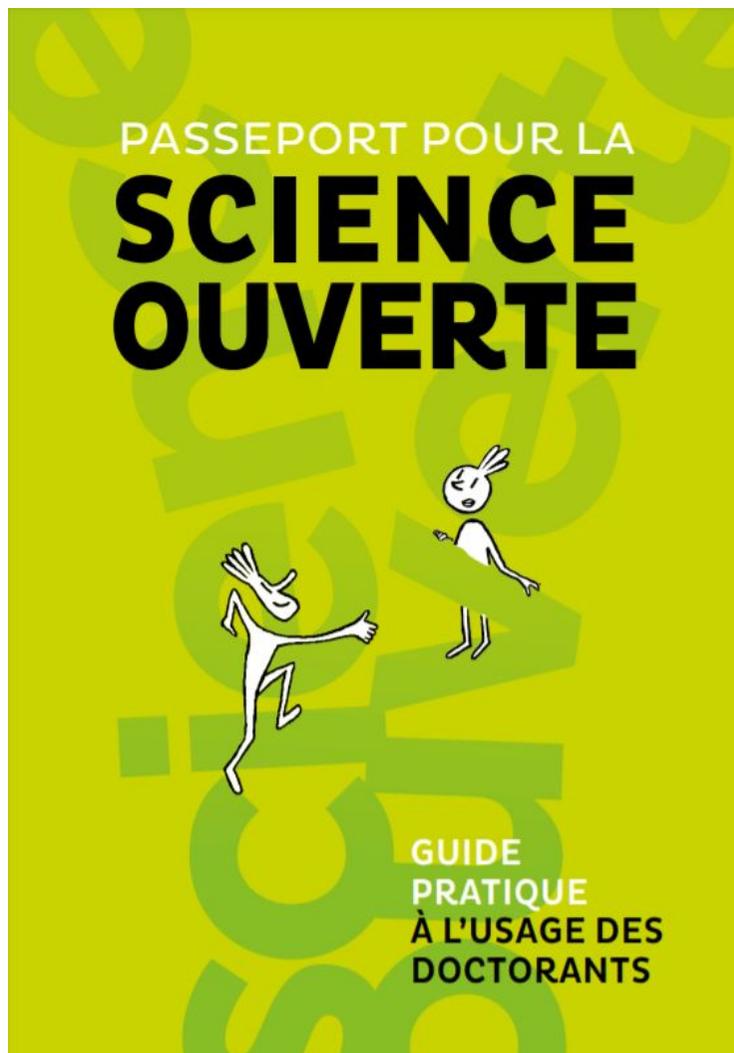


Table des matières

1. Concevoir sa démarche scientifique de manière ouverte

- Utiliser des ressources librement accessibles p. 6
- Prévoir la gestion des données p. 8
- Travailler de manière reproductible : pour soi, pour les autres p. 11

2. Diffuser sa recherche

- Diffuser ses publications en accès ouvert p. 16
- Rendre sa thèse librement accessible p. 21
- Ouvrir les données de recherche p. 25

3. Préparer l'après-thèse, rejoindre le mouvement

- Des politiques publiques enracinées p. 30
- Évaluer la recherche autrement p. 32

Agir dès maintenant p. 34

Aller plus loin p. 35

Glossaire p. 36

Sources p. 38

Légende

Le texte souligné renvoie au glossaire.

▼ signale des outils donnés en exemple.

☒ indique un lien externe

Version numérique disponible sur le site www.ouvrirlascience.fr

Guides Formadoct

- Les revues scientifiques en libre accès
- Publier un article dans une revue scientifique
- Déposer un texte dans une archive ouverte
- Connaître les archives ouvertes

Vidéos Doranum

What is data? - Interviews with French researchers

Merci de votre attention

Contact :

Adresse mel :

Florence.thiault@univ-rennes2.fr

Site web :

<https://urfist.univ-rennes2.fr/>