

La production et la diffusion de l'information scientifique et technique (IST)

NB : Cette fiche est à lire en regard des fiches 1-7 « La science ouverte » et 1-8 « Les données de la recherche »

A voir aussi : fiches Biblio 1-5 « La documentation numérique en bibliothèque » et Biblio 2-6 « Les métadonnées ».

L'information scientifique et technique est au cœur de l'activité de recherche.

En 2012, le CNRS organise une journée intitulée *Données de la recherche : enjeux, perspectives, politique(s)*¹. En 2013, le chantier Bibliothèque scientifique numérique (BSN) consacré aux données de la recherche est lancé. Les journées d'étude sur le thème se développent. La Commission européenne lance un projet pilote sur le libre accès des données financées sur fonds publics², Horizon 2020 (H2020) et « constitue ainsi un jalon politique essentiel »³.

Aujourd'hui, H2020 est remplacé par le projet Horizon Europe⁴ dont le programme a été publié au Journal Officiel de l'Union Européenne le 12 mai 2021⁵.

Cette science ouverte (open science, science 2.0 ou e-science) place les données de la recherche au premier plan. « La science ne serait plus le fruit d'observations expérimentales, d'hypothèses ou de théories, mais naîtrait aussi désormais de la manipulation et de l'agrégation des données »⁶.

Dans son discours lors du congrès annuel de la Ligue des Bibliothèques Européennes de Recherche (LIBER) le 4 juillet 2018, Frédérique Vidal, ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation (de mai 2017 à mai 2022), présente son plan pour la

1 <https://fredoc.hypotheses.org/124> (consulté le 13/07/2024).

2 Plus d'info : <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020> (consulté le 13/07/2024).

3 Rémi Gaillard, mémoire d'accès au diplôme de conservateur, *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche*, p.11, disponible sur : <https://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche> (consulté le 13/07/2024).

4 Pour aller plus loin : <https://www.horizon-europe.gouv.fr/> (consulté le 13/07/2024).

5 Voir les textes officiels : <https://www.horizon-europe.gouv.fr/horizon-europe-les-textes-officiels-24560> (consulté le 13/07/2024).

6 Rémi Gaillard, op.cit, p.13.

science ouverte⁷. Ce plan s'articule autour de 3 axes : généraliser l'ouverture des publications, structurer et ouvrir les données de la recherche, inscrire la France dans une dynamique durable européenne et internationale.

1. Les enjeux de l'IST

1.1 Le paysage de l'IST

Le paysage de l'IST est en permanente évolution et les enjeux sont forts.

On remarque :

- Une généralisation du phénomène de numérisation de la production scientifique (édition numérique, bases de données, corpus documentaire...)
- Une accélération de la circulation des travaux scientifiques, augmentation de la quantité d'informations et de données scientifiques et des besoins d'accès à ces savoirs
- Une multiplication des infrastructures numériques (bibliothèques numériques, outils collaboratifs, réseaux sociaux...)

1.2 Quelles actions à mener ?

Elles sont essentiellement de deux ordres :

- Valorisation et diffusion de la production scientifique par la collecte, le traitement, l'archivage pérenne et la diffusion, notamment sous forme d'archives ouvertes.
- Politique d'acquisition des ressources numériques (exemple des licences nationales) et dispositifs de signalement des ressources numériques.

1.3 Les enjeux

- Ouverture et réutilisation des données (données ouvertes, interopérabilité, ...).
- Construction d'un écosystème propice au partage et à la réutilisation des données.
- Utilisation de langages communs : standard de métadonnées, terminologies (thésaurus, classifications...), identifiants pérennes, formats ouverts, protocoles.

Les métadonnées sont importantes dans ce contexte de partage et d'interopérabilité⁸.

⁷ Disponible sur : <https://www.enseignementsup-recherche.gouv.fr/fr/plan-national-pour-la-science-ouverte-discours-de-frederique-vidal-49326> (consulté le 13/07/2024).

⁸ Voir la fiche de bibliothéconomie TC 2-6 sur les métadonnées.

Il est alors primordial de bien connaître la provenance et le contexte des données (évaluation et fiabilité des données), d'en faire jaillir des mots clés (pour la découverte et la réutilisation), et d'y attribuer des identifiants (permettre l'accessibilité, faire le lien entre les données et les publications, pouvoir les citer).

2. Les acteurs de l'IST en France

La question de l'ouverture des données implique plusieurs acteurs. Elle ne peut se faire que par les établissements. Toute la chaîne doit être concernée : les organismes de recherche, les communautés scientifiques, les financeurs, les éditeurs...

Nous pouvons distinguer trois « acteurs » principaux : le Ministère de l'Enseignement supérieur et de la Recherche, les établissements, le Centre national pour la recherche scientifique (CNRS).

2.1 Le ministère de l'Enseignement supérieur et de la Recherche

Le CODORNUM (Comité d'orientation du numérique) est l'organisme décisionnel mis en place en 2014, auquel participent quatre comités de pilotage :

- InfraNum : définit les orientations en matière d'infrastructures techniques nationales.
- SIESR (service interacadémique de l'enseignement supérieur et de la recherche) : coordonne les évolutions et l'interopérabilité des systèmes d'information de l'ESR.
- Formation : coordination des portails ESUP et Fun-Mooc.
- Le Comité pour la science ouverte⁹.

2.2 Le Centre National de la Recherche Scientifique (CNRS)

2.2.1 La DDOR

La Direction des données ouvertes de la recherche (DDOR) a été créée¹⁰ en 2020. Cette direction résulte de la fusion de la Direction de l'information scientifique et technique (Dist) et de la mission Calcul Données (MiCaDo). Elle élabore et met en œuvre la politique

⁹ <https://www.ouvrirlascience.fr/presentation-du-comite/> (consulté le 13/07/2024).

¹⁰ Plus d'infos : <https://www.science-ouverte.cnrs.fr/les-acteurs-au-cnrs/> (consulté le 13/07/2024).

du CNRS relative aux données dans une optique d'ouverture définie dans sa feuille de route pour la science ouverte¹¹ et son plan « Données de la recherche »¹².

Les trois unités du CNRS (l'INIST, le CCSD et Persée) sont rattachées à la DDOR.

2.2.2 L'INIST-CNRS

L'Institut de l'information scientifique et technique assure la logistique des portails CNRS et la référence française de Datacite (pour l'attribution des DOIs aux objets de la recherche) notamment.

2.2.3 Le CCSD

Le Centre pour la Communication Scientifique Directe gère les plateformes :

HAL (Hyper Articles en Ligne) : plateforme nationale de dépôts et de consultation d'articles scientifiques en texte intégral.

TEL (Thèses en Ligne) : serveur promouvant l'autoarchivage en ligne des thèses de doctorat et d'HDR (habilitations à diriger des recherches).

Sciencesconf.org : plateforme s'adressant aux organisateurs de colloques, workshops ou réunions scientifiques.

Episciences.org : plateforme technique d'examen par les pairs (« peer reviewing ») afin de permettre l'émergence d'épérevues (revues électroniques en libre accès, alimentées par des articles déposés dans les archives ouvertes).

2.2.4 Persée¹³

Plateforme offrant un accès libre à des collections complètes de publications scientifiques, de la première parution à la plus récente, en SHS, SVT et STM (plus de 650000 documents en libre accès).

11 Feuille de route du CNRS pour la Science ouverte, novembre 2019. Disponible sur : https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquelette_Science-Ouverte_18112019.pdf (consulté le 13/07/2024).

12 Plans de données de la recherche du CNRS, novembre 2020. Disponible sur : https://www.science-ouverte.cnrs.fr/wp-content/uploads/2021/01/Plaquelette-Plan-Donnees-Recherche-CNRS_nov2020.pdf (consulté le 13/07/2024).

13 Site Persée : <http://www.persee.fr/> (consulté le 13/07/2024).

2.3 Les communautés scientifiques¹⁴

2.3.1 Les équipes de recherche

Au sein des laboratoires, les chercheurs sont impliqués dans tout le cycle des données¹⁵, de la production à la réutilisation.

2.3.2 Les professionnels d'appui

On peut compter parmi ces professionnels d'appui :

- ✓ Des professionnels de l'IST : tâches d'indexation, d'enrichissement des métadonnées, de conservation des données... qui permettent de conserver de manière pérenne et lisible des données de qualité.
 - Les services IST des universités, les Services inter-établissements de coopération documentaire (SICD) et les Services communs de la documentation (SCD) peuvent être amenés à gérer l'IST de leur périmètre¹⁶. Ces services gèrent ainsi les : archives ouvertes (Hal ou autre plateforme comme Dissemin, Arxiv, ...), la bibliométrie, les libellés des affiliations auteurs/unité, etc.
- ✓ Des informaticiens : développement de plateformes pour entreposer les données.
- ✓ Des juristes pour les questions éthiques et juridiques.
- ✓ Mais également des chargés de projet, des médiateurs, des ingénieurs...

2.3.3 Les éditeurs

Certains éditeurs jouent un rôle dans le partage des données, au moment de leur diffusion. Ils peuvent :

- ✓ Demander un accès aux données pour évaluer et valider les articles scientifiques.
- ✓ Demander de déposer les données dans leur propre entrepôt.
- ✓ Publier des articles décrivant les données (data papers¹⁷) et attribuer des identifiants pérennes aux articles.

14 URFIST Méditerranée, Les acteurs de la gestion et du partage des données de la recherche, DOI : [10.13143/1gjk-hs7Z](https://doi.org/10.13143/1gjk-hs7Z) : . Disponible sur DoraNum : <https://doranum.fr/enjeux-benefices/acteurs-gestion-donnees-recherche/> (consulté le 13/07/2024).

15 Voir le schéma de la fiche 1-8 « Les données de la recherche ».

16 Voir la fiche 3-12 « Les services d'appui aux chercheurs ».

17 Plus d'infos sur la plateforme DoraNum par exemple : <https://doranum.fr/data-paper-data-journal/> (consulté le 13/07/2024).

3 La valorisation et la diffusion de l'IST

3.1 Comment valoriser ?

Les avantages de l'open access :

- ✓ Transparence : confiance des citoyens en la recherche.
- ✓ Nouvelles approches scientifiques : science de données (web de données, moteurs de recherche... = diffusion la plus large possible) et science collaborative.
- ✓ Innovation : renforce la liberté de l'investigation scientifique, encourage la diversité des études et des opinions, favorise de nouveaux domaines d'activité.
- ✓ Retours sur investissement.
- ✓ Relever des défis sociétaux d'envergure mondiale : santé, sécurité, changement climatique, énergie renouvelable...
- ✓ Eviter la fraude.

3.2 Comment diffuser ? Où partager ?

3.2.1 Où partager ?

- ✓ En les publiant dans des data papers ou data journals. Les data papers sont des publications dont le but est de décrire un ou plusieurs jeux de données brutes (à l'aide de métadonnées). L'accès en ligne se fait par lien pérenne (doi ou url) vers l'entrepôt de données où elles ont été déposées. Elles sont parfois regroupées dans des data journals (ou épirevues).
- ✓ En les déposant dans des entrepôts de données de recherche (data repository ou digital repository) : réservoirs constitués de données brutes ou élaborées et décrites par des métadonnées permettant ainsi de les retrouver. Des identifiants pérennes leur sont attribués ainsi que des numéros d'accès. Intérêts : conservation des données, réutilisabilité par le déposant ou les pairs, citabilité, visibilité accrue, transparence... Exemples de répertoires : OpenDOAR ou ROAR (répertoires d'archives), re3data (recommandé par la Commission européenne - H2020) ou OpenAccess Directory (annuaires d'entrepôts multidisciplinaires) ...
- ✓ En la publiant en ligne via le site web d'un projet ou d'un réseau.

3.2.2 Comment diffuser ?

- ✓ Protection des données : penser à l'anonymisation des données.
- ✓ Propriété intellectuelle : données littéraires et artistiques (attention à la propriété détenue par un tiers, pas de propriété intellectuelle sur les données brutes !), droit sui generis des bases de données, propriété industrielle.

- ✓ Définition des permissions d'accès : embargo, restriction de l'accès à un groupe de personnes...
- ✓ Attribution de licences aux données : licences ouvertes¹⁸ (etalab) licences creative commons¹⁹ (droit international, droit des bases de données pris en compte) ...
- ✓ Attribution d'identifiants pérennes pour les données. Cela permet de rendre visible et accessible les données au même titre que les publications, permet de lier les données aux publications. Exemple : doi, handle, PURL, ARK...
- ✓ Attribution d'identifiants pérennes pour le chercheur. Cela permet de gérer les homonymies et d'harmoniser les noms, permet de trouver et citer toute la production d'un chercheur, évite au chercheur de ressaisir toutes les informations le concernant. Exemple : ORCID, ISNI, VIAF, IdHAL, ResearcherID...

4 Quelques exemples d'initiatives et d'infrastructures en France

- ✓ CollEx-Persée (Collections d'excellence pour la recherche) : infrastructure qui vise à contribuer à la performance de la recherche française en développant des services documentaires au plus près des besoins des chercheurs²⁰.
- ✓ Charte nationale de déontologie des métiers de la recherche politique d'éthique et d'intégrité de l'ANR²¹. La première mise en œuvre de cette Charte est la création, en mars 2017 et au sein de l'HCERES (Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur) de l'OFIS²² (Office français de l'intégrité scientifique) (éthique et juridique).
- ✓ HAL : Archive ouverte pluridisciplinaire française, développée par le Centre pour la Communication Scientifique Directe (CCSD), destinée au dépôt et à la diffusion d'articles scientifiques (publiés ou non) et de thèses. Cette archive est incontournable pour la diffusion des savoirs de l'ensemble de la communauté scientifique et universitaire française.
- ✓ TGIR Huma-Num : Très grande infrastructure de recherche en sciences humaines et sociales, développé par l'UMS des Humanités Numériques, met à disposition un ensemble de services pour les données numériques produites en SHS.
- ✓ OpenEdition : Infrastructure de recherche sur l'édition électronique ouverte en

18 Plus d'infos sur le blog d'Etalab : <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (consulté le 13/07/2024).

19 Plus d'infos : <https://creativecommons.org/licenses/?lang=fr-FR> (consulté le 13/07/2024).

20 Plus d'infos : <https://www.collexpersee.eu/> (consulté le 13/07/2024).

21 Disponible sur : <https://comite-ethique.cnrs.fr/charte/> (consulté le 13/07/2024).

22 Plus d'infos : <https://www.hceres.fr/fr/ofis> (consulté le 13/07/2024).

sciences humaines et sociales, développée par l'USR OpenEdition Center (CLEO). Elle met à disposition quatre plateformes de publication et d'information en accès ouvert : revues, livres, carnets de recherche et annonces d'événements académiques internationaux en SHS.

- ✓ Métopes : Infrastructure de recherche sur les méthodes et d'outils pour l'édition structurée, porté par le Pôle « Document numérique » de la maison de la recherche en sciences humaines (MRSH). Elle est conçue à l'usage des éditeurs publics et au service de l'activité éditoriale de l'ensemble des établissements publics d'enseignement supérieur et de recherche. Elle a pour missions principales, de fournir un ensemble d'outils et de méthodes afin d'organiser la production et la diffusion des produits éditoriaux en libre accès.

Pour aller plus loin :

Rémi Gaillard, *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*, mémoire d'étude pour l'accès au diplôme de conservateur, disponible sur : <https://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche> (consulté le 13/07/2024).

Le site dédié au Plan national pour la science ouverte : <https://www.ouvrirlascience.fr/> (consulté le 13/07/2024).