



# LA LETTRE **CorIST**

Information bimestrielle de la DIST  
et des correspondants IST des instituts

N° 13 | ÉTÉ 2016

*Le partage  
de  
l'IST*



cnrs

[www.cnrs.fr](http://www.cnrs.fr)

Direction de l'information scientifique et technique

# Editorial

Bonjour,

L'été a apporté à la Dist une moisson de nouvelles, heureuses et malheureuses.

La disparition de Raymond Bérard a été pour tous l'occasion d'un rassemblement, par la pensée et pour beaucoup lors de ses obsèques le 3 août à Lyon. La Dist est évidemment au premier rang de ceux qui, après le Président du CNRS, ont exprimé leur tristesse et leur compassion à Edith Bérard et à ses filles. Qu'il me soit permis d'y ajouter, aux côtés de tous nos collègues, en particulier de l'Inist, une très grande reconnaissance pour l'effort de renouveau qui s'associe à toute son œuvre. Je parlerai volontiers, en pensant à lui, que j'ai eu l'honneur de côtoyer au quotidien, dans des responsabilités étroitement complémentaires, de son goût profond pour l'espérance : il la mêlait à toutes choses, avec un talent visionnaire que Dominique Wolf garde précieusement avec elle et qui nous lie.

L'été nous apporte aussi la confirmation des choix faits par la Commission mixte paritaire (CMP) pour la loi numérique, puisque l'Assemblée nationale vient de se prononcer, le 20 juillet, en confirmant à l'unanimité (moins une voix...), le texte de la CMP : il reste maintenant, avant publication de la loi en octobre, l'examen du texte par le Sénat, prévu pour la dernière semaine de septembre. Préparons nous, dès à présent, à réfléchir à toutes les directions nécessaires à l'application de ces textes importants pour la science, dès que nous connaissons le texte définitif.

L'été, c'est enfin, autour de Marie Jeanne Gaudoin, les beaux résultats du Copist, que nous allons pouvoir travailler dès la rentrée avec nos collègues des grandes Conférences (Universités et Ecoles) ainsi qu'avec nos amis d'Eprist, de l'ADBU et de Couperin...

La vie continue, donc, et avec elle, tous les bonheurs de l'été : bonnes vacances à toutes celles et ceux qui en prennent, et à très bientôt dans tous nos projets !

Bien cordialement,

Renaud FABRE

Directeur de la DIST

[renaud.fabre@cnrs-dir.fr](mailto:renaud.fabre@cnrs-dir.fr) / 01 44 96 46 57



# Actualités IST

## [Catalogue d'offres partagées d'IST \(COPIST\) : l'enquête auprès de l'ESR.](#)

L'enquête adressée le 20 avril dernier aux acteurs de la recherche publique, dans le cadre de l'appel à manifestation d'intérêt autour du Catalogue d'Offres Partagées d'IST (COPIST) est close depuis le 11 juillet.

La diffusion de cette enquête auprès des Présidents d'Universités, des Bibliothèques universitaires, des Présidents d'organismes de recherche et des responsables IST des organismes, des Directeurs des grandes écoles a permis de recueillir 103 réponses pour un panel de 163 sondés.

L'ADBU, EPRIST, la CPU et le CNRS s'associent pour remercier vivement les participants de leurs riches contributions.

Cet inventaire permettra à tous d'échanger sur l'état de leurs ressources d'IST. Il recense les pratiques et les usages à l'échelle nationale, en dépassant le cadre du CNRS.

Comme prévu les positions et les souhaits des établissements recueillis et classés permettront de lancer des études en souscription qui proposeront des modèles d'usages et des options ouvertes pour mieux partager l'IST.

Un groupe de travail (CPU, ADBU, Eprist et CNRS) sera constitué pour analyser les retours, après un premier examen effectué par les corIST et l'Inist. Les premiers résultats pourront être diffusés dès cet automne.

## [Une archive des bases bibliographiques Pascal et Francis en accès libre sur le web](#)

### **Une première version met en ligne plus de 14 millions de notices bibliographiques**

Dans une optique d'ouverture et de partage, l'Inist-CNRS met à la disposition de tous un patrimoine bibliographique et scientifique issu d'une quarantaine d'années de signalement de la production scientifique française et mondiale.

Produites par l'Inist-CNRS entre 1972 et fin 2015, les bases de données Pascal et Francis réunissent près de 23 millions de références bibliographiques :

- Pascal couvre le champ des sciences, technologies, médecine
- Francis couvre 15 disciplines en sciences humaines et sociales

La plateforme d'interrogation propose, dans un premier temps, un accès à plus de 14 millions de notices bibliographiques d'articles, de communications et d'actes de congrès (12 millions issues de Pascal et 2 millions issues de Francis). Elle s'enrichira progressivement d'autres types de documents, et avec des notices de partenaires ayant coopéré à

l'alimentation des bases de données PASCAL et FRANCIS, au fur et à mesure des accords obtenus.

Le site : <http://pascal-francis.inist.fr>

## Genèse, portée et application de la loi numérique sur l'information scientifique

Restés dans le flou pendant de longues années, alors qu'une dynamique internationale d'*open science* a été initiée, le partage des connaissances et le traitement des données dans un but scientifique ont été renforcés dans le projet de loi pour une « République numérique ». L'information scientifique va ainsi largement bénéficier de cette avancée juridique qui va pouvoir accompagner un développement soutenu des échanges documentaires et à travers ceci, de nouvelles possibilités d'investigation scientifique.

Fin juin 2016, la validation du texte de loi en commission mixte paritaire a permis d'en fixer la portée. La partie concernant l'information scientifique et technique (IST), Titre I<sup>1</sup> essentiellement, va enrichir directement le travail du CNRS, à travers trois points essentiels : l'utilisation des données, une nouvelle politique de publication, la validation juridique de la fouille de texte et de données (*text and data mining*, dite TDM).

### **1. Genèse. Comment la stratégie d'IST a-t-elle aiguillé la loi numérique ?**

#### **1.1. Les concepts initiaux et les étapes de travail**

La conception de la loi numérique a été aiguillée par des documents et travaux antérieurs, par le CNRS, l'EPRIST, le Consortium Couperin ou encore l'ADBU, avec un appui juridique décidé par le projet ISTEEX.

#### *La stratégie « Mieux partager les connaissances » du CNRS*

La stratégie du CNRS visait à promouvoir une recherche publique ouverte, en insistant sur la circulation et la disponibilité de l'information scientifique et technique. Dans le but de favoriser la compétitivité économique, de stimuler la recherche fondamentale et d'aiguiller la recherche appliquée, le CNRS souhaitait :

- Ouvrir l'accès à l'IST ;
- Partager les ressources des membres de l'IST, laboratoires, chercheurs... ;
- Offrir des outils de traitement de la connaissance pour tous ;
- Partager les projets, comme ISTEEX ou COPIST.

Comme le précisait cette stratégie : « L'émergence du numérique dans les pratiques d'IST provoquent un décalage entre le droit et la pratique. Si le Code de la recherche organise la recherche publique et définit ses objectifs, il n'affirme à aucun moment les valeurs communes de la Science. »<sup>2</sup>

Surtout, ce travail pour les chercheurs et les producteurs de la connaissance s'est constitué avec un objet de valorisation de l'IST. Ce fut cet argument de valorisation qui a marqué le plus les décisions du Parlement. Cette valorisation des résultats scientifiques et l'impératif

---

<sup>1</sup> Projet de loi pour une République numérique, Texte élaboré par la CMP, AN, 3902, Sénat 744, 30 juin 2016

<sup>2</sup> CNRS, *Mieux partager les connaissances*, 2014I, p.35

de compétitivité ont été une orientation principale dans les débats sur les contraintes de publication et sur la réduction des embargos.

#### *La mission relative au data mining et l'analyse de Couperin et de l'ADBU*

L'analyse spécifique du TDM a permis de montrer en quoi les bénéfices économiques, tant au niveau de la recherche publique que du secteur commercial, justifient une adaptation du droit : « les bénéfices pour l'ensemble de la société, qu'il s'agisse du secteur public ou commercial, sont très nettement supérieurs au peu probable préjudice que pourraient encourir les titulaires de droits du fait des usages attachés au TDM. »<sup>3</sup>

#### *Le Livre blanc « Une science ouverte dans une République numérique »*

L'accès aux données scientifiques mais surtout leur utilisation via des outils de traitement représentent le fer de lance de la politique scientifique au 21<sup>e</sup> siècle. L'objectif n'est plus la simple consultation de ses données, mais leur compréhension, leur gestion, leur organisation, leurs lectures et interprétations détaillées. C'est cette stratégie qui a animé les orientations du TDM dans la loi.

Le Livre blanc a permis d'insister sur ces points fondamentaux par des témoignages et de l'expertise. Face à la hausse du nombre d'articles, c'est l'importance de leur traitement qui a été confirmée dans le projet retenu. La loi va ainsi permettre de clarifier le droit de la recherche en matière d'utilisation des données.

## **1.2. Les formulations nationales et internationales de la défense de l'IST**

#### *Des stratégies nationales, notamment européennes*

La défense de l'IST et de son application libre via des outils comme le TDM est un processus national, qui se retrouve au niveau européen. L'action du Commissaire européen à la recherche, à l'innovation et à la science, M. Carlos Moedas, va en ce sens. En faisant de l'« *open science* » un des objectifs principaux de son action à la Commission<sup>4</sup>, il a souligné l'importance que le traitement des données scientifiques et leur partage avaient dans le cadre de la stratégie européenne d'« *excellence scientifique* » (*Horizon 2020*). Cette excellence scientifique a déjà été l'objet d'innovations juridiques en Europe et ailleurs, notamment dans le cadre du TDM.

Le Japon a été le premier pays à ainsi légalisé le TDM de façon indirecte en 2011. Le pays a validé la reproduction sur un ensemble de données dans un objectif d'analyse d'informations, mais sans délimiter les bonnes pratiques<sup>5</sup>. Le Royaume-Uni a été le deuxième pays à infléchir sa législation, en intégrant une « *exception for copying of works for use by text and data analytics* »<sup>6</sup>. Ces deux cadrages juridiques ouvrent à des précisions et des interprétations nationales<sup>7</sup>. C'est l'Irlande, en 2012, qui a effectivement donné droit à l'utilisation du TDM de façon directe : « *this section applies to (a) text-mining, data-mining, and similar analysis or reseach* »<sup>8</sup>. Notons enfin l'inspiration nord-américaine du projet, avec

<sup>3</sup> CSPLA – Mission relative au TDM (exploration de données) : l'analyse de Couperin et de l'ADBU, 2014

<sup>4</sup> <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>

<sup>5</sup> Loi sur le copyright, Section 5, Art. 47-7, Japon, Copyright Research and Information Center

<sup>6</sup> «Modernising Copyright: A modern, robust and flexible framework», Réponse du Gouvernement à la consultation sur les droits d'auteur, Annexe E, Août 2011, p.36

<sup>7</sup> De Wolf & Partners, *Study on the Legal Framework of TDM*, Mars 2014

<sup>8</sup> Comité sur le droit d'auteur, « Page de consultation », Ministère de l'emploi, de l'économie et de l'innovation, Dublin, 2012

les conclusions apportées à la bibliothèque *HathiTrust* en matière de fouilles de textes (ainsi que la jurisprudence *GoogleBook* en 2013) dans le cadre de la politique d'*open access* américaine<sup>9</sup>.

### *Les formulations théoriques et internationales de plus long terme*

La valorisation du partage des connaissances reflète au fond un mouvement plus profond, d'ouverture progressive des articles, de la science. Si elle est concrétisée dans des politiques, elle est surtout l'objet de déclarations internationales et de formulations théoriques implicites.

Cette ouverture de la science est avant tout rendue possible par la technique et le développement d'Internet, du numérique et des outils auxquels ils donnent forme. C'est notamment le cas des licences *creative commons*, des archives ouvertes (HAL) ou du réseau des DIST (COPIST). Ces outils sont amplement promus par l'*Open Research Society* ou l'*Internet Society*, voire de façon plus récente par l'Europe (*Open Science Policy Platform*). En plus des outils, les volontés institutionnelles ont débordé sur le monde politique : Déclaration de Berlin (2003), document *Principes directeurs pour le développement et la promotion du libre accès* par l'UNESCO (2013), travaux de Lawrence Lessig et du *Center for Internet and Society*, rapport *Reda* (2015), Déclaration de la Haye par LIBER (2015), l'appel d'Amsterdam (2016), etc. Tous ces travaux sont régis par une volonté scientifique bien plus profonde, que le sociologue R.K. Merton avait déjà invité à suivre en 1942<sup>10</sup>.

Tous ces travaux et des dynamiques internationales ont assurément rendu possible la partie du projet de loi Pour une République numérique sur l'information scientifique et technique. Adoptée en CMP, la version actuelle préfigure une validation définitive de trois points centraux.

## **2. Portée. Quels sont les trois champs de la loi numérique concernant l'IST ?**

### **2.1. Les données**

#### *Les articles du texte de loi*

L'utilisation des données renvoie à l'article 4 de la section 1 du chapitre Ier du titre I : « Circulation des données et du savoir », « Ouverture de l'accès aux données publiques ». Cette ouverture des données devient une contrainte légale : « *les administrations mentionnées au premier alinéa de l'article L. 300-2 [...] publient en ligne les documents administratifs suivants...* », exceptions faites de dispositions législatives contraires ou de dispositions précisées par décrets. Elle est spécifiée pour l'information scientifique à l'article 9. Les dispositions s'appliquent aux données brutes qui accompagnent une publication et en sont un des éléments de l'administration de la preuve.

#### *Quels objectifs ?*

Cette ouverture des données va faciliter la production de valeur scientifique, en élargissant les bases d'accès et en ouvrant la connaissance de champs certains, que ce soit en SHS

<sup>9</sup> Executive Office of the President, Office of Science and Technology Policy, "Increasing Access to the Results of Federally Funded Scientific Research", 2013

<sup>10</sup> Robert K. Merton, « The Normative Structure of Science », in *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, University of Chicago Press, 1942

comme en STM. Surtout, avec l'article 9, la vérification, la transparence de la recherche mais aussi le partage des données brutes et le renouvellement des expériences vont enrichir les méthodes scientifiques et la pertinence des conclusions.

## **2.2. La publication**

### *Les articles du texte de loi*

La publication scientifique renvoie à l'article 17 du chapitre 2 « Economie du savoir » du titre I. La modification du Code de la recherche tient ici aux délais de publication. La publication d'un article dont le financement est décrit à l'article L.533-4 n'empêche pas l'auteur de cet article de le « *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* ». Le délai qui était en place avant entre les publications de l'éditeur et de l'auteur est réduit. Il passe à six mois pour un article du domaine des STM et à douze mois dans le cadre des SHS. Cette disposition anticipe la directive européenne de révision du droit d'auteur.

### *Quels objectifs ?*

Cette possibilité de diffuser les connaissances dans des délais courts dans les archives ouvertes est un facteur essentiel de la dynamique des sciences.

## **2.3. La fouille de données (*text and data mining*)**

### *Les articles du texte de loi*

La fouille de données renvoie aux articles 18 et 18 bis. Elle modifie le Code de la propriété intellectuelle. Sont ainsi autorisées à être traitées « *les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale.* » La base de données et ses copies sont elles-mêmes concernées par la disposition.

### *Quels objectifs ?*

La loi supprime l'exclusivité des droits de recherche, de traitement et de fouilles des articles et documents qu'un éditeur possède (si le financement de cette recherche est assuré pour moitié par des fonds publics). L'objectif est ici de redonner une forme de souveraineté aux chercheurs, mais surtout de perdre l'exploitation des bases de données d'articles.

## **3. Application. Comment le CNRS peut-il mettre en place la loi numérique ?**

Les deux chambres devraient adopter en deuxième lecture le texte validé par la commission mixte paritaire. L'Assemblée nationale l'a fait le 20 juillet. Les chercheurs attendront le 27 septembre pour le Sénat. Cette validation législative va entraîner une étape d'application, dont les décrets sont l'enjeu fondamental. Au-delà, pour les personnels de la recherche, de nouvelles questions se posent.

### **3.1. Les décrets. Quelle mise en œuvre des textes est prévue ?**

L'écriture des décrets est un espace de liberté. Suite à la promulgation de la loi, des décrets d'application viendront fixer les modalités de cette loi. C'est notamment le cas de l'article 18 bis : « *Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche.* » Autrement dit, trois aspects méritent d'être interrogés de façon prospective : l'exploration, la conservation et la communication des fichiers et données.

Au niveau de la procédure, l'initiative de la programmation des décrets viendra du Ministère de la Justice. Le travail d'écriture devrait lier logiquement le Ministère de l'Enseignement Supérieur et de la Recherche, ainsi que le Ministère de l'Economie, de l'Industrie et du Numérique. Le calendrier devrait être précisé suite à l'adoption du texte de la CMP au Sénat.

### **3.2. Les changements pour la science et les chercheurs. Quelles perspectives la loi numérique offre-t-elle sur le plus long terme ?**

L'accès aux données ainsi que leur utilisation par le TDM ouvrent de nouveaux domaines d'application pour la recherche française. En plus des avancées métriques et calculatoires, ces utilisations vont ouvrir la voie à de nouvelles pratiques : vérification poussée pour les hypothèses de recherche, travail sur les données brutes... La contestabilité des articles va assurément renforcer le dispositif de preuves des chercheurs.

Une étude d'impact a été demandée par le Parlement pour mieux comprendre les retombées sur les activités d'édition mais aussi sur la circulation du savoir. Les bénéfices économiques sont néanmoins reconnus dans le cadre du TDM. Une récente étude de l'Université d'Amsterdam stipulait que la recherche académique nationale était stimulée par la pratique du TDM<sup>11</sup>. Cette influence pourrait être facilement démontrée auprès du public français. En effet, des documents d'application (soit des exemples d'utilisation, des mises en œuvre possibles du TDM) permettraient de renforcer la fiabilité et la puissance du TDM.

---

<sup>11</sup> Handke, Guibault & Vallbé, "Is Europe Falling Behind in Data Mining? Copyright's impact on Data Mining in Academic Research", Juin 2015



# Dossier

## [Le partage de l'IST](#)

### Appel à projet « Technologies innovantes au bénéfice de l'interdisciplinarité » - Mission pour l'Interdisciplinarité (MI)

Les réseaux métiers et technologiques du CNRS, par leur capacité à fédérer la réflexion des personnels d'horizons disciplinaires variés, sont des acteurs de premier plan pour identifier les émergences et les verrous à lever pour conduire ces avancées technologiques au meilleur niveau, au bénéfice des équipes de recherche.

Dans ce contexte la MI a choisi de soutenir à travers les réseaux, des projets technologiques qui devaient s'appuyer sur des synergies interdisciplinaires pour :

- Concevoir et mettre en œuvre de nouveaux dispositifs technologiques,
- Lever des verrous technologiques limitant les performances des instruments actuels ou domaines applicatifs,
- Développer des méthodologies de conception ou d'utilisation innovantes.

Les projets soumis devaient démontrer leur potentiel applicatif interdisciplinaire, l'intérêt de leur développement au sein d'un ou plusieurs réseaux et en se situant dans le cadre des réflexions issues des actions des réseaux (groupe de travail, ateliers, journées thématiques).

36 dossiers ont été examinés par le comité de suivi qui a sélectionné 6 dossiers technologiques et 2 dossiers plus méthodologiques pour les soumettre à l'avis définitif du comité de pilotage de la MI. Celui-ci, réuni le 22 juin dernier, a choisi 3 projets dont 2 technologiques : « INDUCHAUD » pour le réseau Hautes Pression et « PHYCHIMISOL » du réseau CMDO (Cristaux Massifs, Micro-nano structures et Dispositifs pour l'Optique) et le **projet méthodologique « ValorEx - Outil de valorisation des compétences et expertises des réseaux professionnels de la mission pour l'interdisciplinarité » porté par le réseau RENATIS qui associe plusieurs d'autres réseaux (Médecin, MateSHS, RNBM, Qualité).**

L'objectif du projet est de concevoir un outil commun qui identifie et recense et valorise les domaines d'expertise et métiers dans les réseaux professionnels du CNRS.

En effet, chaque réseau dispose peu ou prou d'une base de données ou d'un outil permettant de recenser ses membres ou activités. Une action transversale viserait à recenser ces plateformes, à les interconnecter et à identifier les domaines de compétences. L'objectif final étant de développer une base de compétences des membres de différents réseaux.

## Cartographie des plateformes de partage d'IST

Cette section présente différentes plateformes (bibliothèques numériques, bases de données, archives ouvertes, moteurs de recherche, etc.) utilisées pour le partage des résultats de la recherche. Les plateformes sont recensées et catégorisées afin de mettre en valeur les initiatives en faveur de la mutualisation des connaissances et du libre accès à la production scientifique.

### « Plateformes », une définition large

Une plateforme numérique est un objet technique accessible en ligne offrant un ensemble intégré ou relié de services et de ressources. Les plateformes de partage d'IST sont diverses entités numériques, telles que des archives ouvertes, des portails documentaires (de labos, de bibliothèques, d'éditeurs...), des réseaux sociaux, des bibliothèques numériques et d'autres bases de données.

### Contexte et objectifs

L'article 2 du « décret du 24 novembre 1982 portant sur l'organisation et le fonctionnement du CNRS » stipule que le CNRS doit contribuer à « développer l'information scientifique et l'accès aux travaux et données de la recherche, en favorisant l'usage de la langue française ». La création d'une typologie avancée des plateformes de partage d'IST mettant en valeur les plateformes francophones et en libre accès entre dans le cadre de cette mission nationale.

Le recensement et la catégorisation de plateformes est une première étape au sein d'un projet mené par la DIST de création d'un annuaire sur le partage de l'IST. Les résultats de l'enquête COPIST viendront bientôt enrichir ce premier document. Cette maquette doit permettre d'améliorer la compréhension d'un écosystème numérique en constante mutation, composé de plateformes, logiciels, et services de partage d'IST ; ainsi que de mieux identifier les acteurs qui y contribuent. Cette première cartographie a vocation à être enrichie<sup>12</sup>.

Ce projet contribuera à répondre aux objectifs d'Open Science au travers des étapes suivantes :

Étapes	Objectif	Actions à mener
1 (en cours)	<b>Transparence et visibilité</b>	Recenser et catégoriser les ressources et services d'IST existants. Analyser la diversité des plateformes de partage d'IST en termes de formats de données, modèles économiques, organismes gestionnaires, etc. Éventuellement voir si le concept de « biodiversité » peut s'appliquer aux plateformes.

<sup>12</sup> Vous pouvez à tout moment proposer vos ajouts et modifications à Stéphanie Dos Santos, chargée d'étude plateformes et réseaux à la DIST ([stephanie.dos-santos@cnrs-dir.fr](mailto:stephanie.dos-santos@cnrs-dir.fr) ; 01 44 96 45 79)

2	<b>Visualisation des réseaux</b>	Effectuer une cartographie des réseaux faisant apparaître des liens entre des communautés de chercheurs et des ressources numériques. Cette cartographie sera effectuée à l'aide de logiciels d'analyse et de visualisation des réseaux (ex : NodeXL <a href="http://nodexl.codeplex.com/">http://nodexl.codeplex.com/</a> ). Les résultats du COPIST seront pris en compte.
3	<b>Décloisonnement</b>	Valoriser certaines plateformes de partage d'IST, notamment celles dont le modèle économique est innovant. A terme, favoriser une ouverture de l'IST, un décloisonnement des plateformes, avec la naissance de nouveaux liens au sein de ces réseaux (par exemple au travers de partenariats).
4	<b>Nouveaux services</b>	Permettre l'émergence de nouveaux services aux chercheurs (ex : lecture en TDM).

## Exemple de catégorisation de plateformes francophones (notamment du CNRS).

### 1. Plateformes d'accès à l'IST

#### Bases de données en libre accès

- **HAL** (<http://hal.archives-ouvertes.fr>) (Hyper articles en ligne) est une archive ouverte pluridisciplinaire destinée au dépôt et à la diffusion d'articles scientifiques (publiés ou non) et de thèses. Les thèses sont disponibles dans TEL (thèses en ligne). L'acquisition des données se fait par auto-archivage. La diffusion d'HAL est sous la responsabilité du **CCSD** (Centre pour la communication scientifique directe), une unité mixte de services du CNRS (DIST), de l'INRIA et de l'Université de Lyon.
- **Gallica** (<http://gallica.bnf.fr/>) permet d'accéder librement à de nombreux documents numérisés (livres, revues, photos, enluminures, etc.) de la Bibliothèque nationale de France (BNF). La base de données couvre toutes les disciplines, mais particulièrement les sciences humaines et sociales.
- **OpenEdition** (<http://www.openedition.org/>) (contenant : **Revue.org**, **Hypothèses.org**, **Calenda**, etc.) est un portail de publication en SHS créé par le Centre pour l'édition électronique ouverte (Cléo). La majorité des articles sont en « open access freemium », tandis que d'autres sont en « open access simple » et « accès exclusif ». OpenEdition est financé par le CNRS, l'EHESS, l'Université d'Aix-Marseille, l'Université d'Avignon, la Fondation Calouste Gulbenkian et Google.
- **Persée** (<http://www.persee.fr/>) est un portail de revues en sciences humaines et sociales (SHS) dont les documents sont acquis grâce à une numérisation rétrospective de revues françaises. Le texte intégral est accessible en mode image ainsi qu'en mode texte (OCR). Persée est géré par l'UMS Persée, l'Université de Lyon, l'ENS Lyon et le Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche (MENESR).

## Bases de données ouvertes sous condition définies par les entités gestionnaires publiques

- **BibCNRS** (<https://bib.cnrs.fr/>) est un portail documentaire du CNRS, développé par l'INIST, qui centralise dix portails thématiques : biologie, chimie, écologie et environnement, homme et société, ingénierie et systèmes, mathématiques, nucléaire et particules, physique, sciences de l'information, terre et univers. Son développement qui est actuellement au stade bêta est réalisé en open source.
- **ISTEX** (Initiative d'excellence de l'Information Scientifique et Technique) (<http://www.istex.fr/>) est un projet investissement d'avenir. L'objectif principal du projet est d'offrir à l'ensemble de la communauté de l'enseignement supérieur et de la recherche un accès en ligne aux collections rétrospectives de la littérature scientifique dans toutes les disciplines en engageant une politique nationale d'acquisition massive de documentation : archives de revues, bases de données, corpus de textes, etc. ISTEX est géré par l'INIST-CNRS, la DIST-CNRS, l'ABES, le consortium Couperin et la CPU.
- **PASCAL** (<http://www.inist.fr/?PASCAL-73&lang=en>) était une base de données bibliographique en science, technologie et médecine, développée par l'INIST-CNRS. La base de données n'est plus mise à jour depuis 2014.

## Bases de données d'éditeurs

- **CAIRN** (<http://www.cairn.info>) fournit un accès numérique – principalement payant – aux revues de quatre éditeurs francophones en sciences humaines et sociales (Belin, De Boeck, La Découverte et Erès). Seule une partie des ressources est en libre accès, notamment les plus anciennes.

## Catalogue numérique de ressources « physiques »

- **SUDOC** (<http://www.sudoc.abes.fr>) est un catalogue collectif qui permet aux bibliothèques des universités françaises et d'autres établissements d'enseignement supérieur de recenser et d'indiquer la localisation des documents en leur possession. Le SUDOC est développé par l'Agence bibliographique de l'enseignement supérieur (ABES) et est accessible en ligne depuis 2000.

## Moteur de recherche spécialisé dans les documents en libre accès

- **ISIDORE** (<http://www.rechercheisidore.fr/>) est la plateforme de référence pour l'accès aux contenus en libre accès dans les domaines des SHS. Elle moissonne également plusieurs milliers de sources spécialisées (bases de données, réservoirs documentaires, corpus, collections d'ouvrages, de revues, d'images ou de données) et dispose d'une interface mobile (m.rechercheisidore.fr). La plateforme Isidore est portée par la très grande infrastructure de recherche Huma-Num.



## 2. Annuaires de la recherche

### Annuaire généraliste

- **ScanR** (<https://scanr.enseignementsup-recherche.gouv.fr/>) est le nouveau moteur de recherche du MENESR qui recense les orientations et les productions des structures de recherche françaises. Cf. présentation faite par l'ADBU : <http://adbu.fr/scanr-le-nouveau-moteur-de-recherche-du-menesr-qui-identifie-en-quelques-clics-les-orientations-et-les-productions-de-toutes-les-structures-de-recherche-francaises/>

### Annuaire de journaux

- **Mir@bel** (<http://www.reseau-mirabel.info/>) (Mutualisation d'informations sur les revues et leurs accès dans les bases en ligne) est un annuaire qui recense des éditeurs et des revues. La base de données contient principalement des revues francophones en SHS. Mir@bel a été créé par trois établissements d'enseignement supérieur français (Sciences Po Grenoble, Sciences Po Lyon et Ecole Normale Supérieure de Lyon) en 2012.

### Annuaire de plateformes

- **IBISA** (<http://www.ibisa.net/plateformes/>) (Infrastructures en Biologie Santé et Agronomie) contient un annuaire de plateformes issues de la recherche française en biologie, santé et agronomie, qui ont été sélectionnées suite à un appel d'offre. Ce site est géré par le groupement d'intérêt scientifique (GIS) IBISA, créé en mai 2007.
- **La base de données de la bibliothèque de l'Université Paris 8** (<https://www.bu.univ-paris8.fr/ressources/bases-de-donnees#p>) recense des plateformes des ressources bibliographiques (livres numériques, revues, encyclopédies, etc.) ainsi que des plateformes de partage d'IST.

### Annuaire de personnes

- **RENALISTE/CNRS** (<http://renaliste.cnrs.fr/>) est un annuaire des professionnels de l'IST travaillant dans des unités associées au CNRS.

### Prochaines étapes

La version détaillée des bases de données et moteurs de recherche d'IST, à l'échelle mondiale, est accessible en ligne (<http://www.cnrs.fr/dist/acces-ist>).

En complément, le site de la DIST comportera une page présentant les ressources et acteurs impliqués dans la gestion de l'IST (<http://www.cnrs.fr/dist/gestion-ist>)

Par la suite, l'écosystème de l'IST sera analysé et modélisé sous forme de réseaux.

Pour plus d'information : <http://www.cnrs.fr/dist/acces-ist.html>

# Formation, stages

## 6<sup>ème</sup> journée nationale d'étude du Réseau des URFIST

Le 29 septembre prochain aura lieu à Paris (au CNAM) la 6<sup>ème</sup> Journée nationale d'étude du Réseau des URFIST intitulée "Partage des données de la recherche : quels impacts ?". Vous trouverez toutes les informations sur cette journée sur le blog JNE 2016 : <https://urfistjne2016.wordpress.com/>.

Vous pouvez d'ores et déjà vous inscrire sur la plateforme SYGEFOR du Réseau des URFIST : <https://sygefor.reseau-urfist.fr/#!/training/6080/6557/?from=true>

## ANF Renatis «Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données»

Cette ANF organisée par le réseau Renatis soutenue par la Mission pour l'Interdisciplinarité et la DIST-CNRS, en collaboration avec l'INIST-CNRS, le réseau des URFIST, le RIATE et IRSTEA s'est tenue du **6 au 8 juillet 2016 à Paris** dans les locaux de l'**Université Paris Diderot**.

Elle s'inscrivait dans la continuité des FréDoc 2013 dédiés à la "*gestion et valorisation des données de la recherche*" et en complément de celle proposée par le réseau RBDD en 2015 "*conduire et construire un plan de gestion de donnée : de la base de donnée à la pérennisation*".

30 participants ont assisté à la formation organisée autour des axes suivants :

- Enjeux et attendus de la gestion des Données de recherche dans un contexte réglementaire international (H2020 et Open Science) ;
- Aspects pratiques liées à la gestion et partage des données (formats et métadonnées, aspects juridiques, différents étapes du cycle de vie des données) ;
- Ateliers pratiques autour de la mise en œuvre du DMP (plan de gestion des données)
- Plan de gestion des données comme outil de la traçabilité des activités de recherche et gestion des connaissances.

En fin de la formation un temps de réflexion et d'échanges qui s'est avéré très riche, a été organisé autour de trois questions posées aux stagiaires :

- Quels seront les bénéfices de cette formation pour moi et comment pourrai-je exploiter les acquis dans mon unité ?
- De quels moyens dois-je disposer et quelles modalités pratiques faudra-il mettre en œuvre pour réaliser/contribuer au projet autour des Données de recherche ?
- Quel est votre ressenti après la formation : quels sont vos points "faibles", quels thèmes seraient à approfondir ?

De façon unanime les participants ont trouvé la formation très riche et complète. Un gros travail de sensibilisation auprès des chercheurs s'avère nécessaire. Plusieurs questions demeurent : comment « approcher » les chercheurs et les sensibiliser à l'importance de nouvelles pratiques liées à la gestion des données tout au long de la vie d'un projet de recherche, dès le début de la collecte des données, comment les accompagner dans la création de DMP en passant par le travail autour des métadonnées, des identifiants pérennes, de nommage des fichiers.. Il a été clair pour tout le monde que le professionnel de l'IST doit se positionner en tant qu'accompagnateur, la place d'expert revient au chercheur, spécialiste dans sa discipline. Différentes actions de sensibilisation, de communication et d'information, grâce aux dialogues avec les responsables de projets et des équipes peuvent être mise en place dans un délai relativement court dans les unités. Les mises en synergie des compétences diverses (informatiques, juridiques, IST) sont nécessaires. Le positionnement et soutien « institutionnel » est également attendu pour promouvoir les bonnes pratiques en matière de stockage, de diffusion et de partage des données.

Le premier bilan incite les organisateurs à réfléchir et à travailler sur les suites à donner aux journées. La demande et les besoins sont importants. Cette réflexion portera à la fois sur les modalités pratiques de la mise en œuvre des actions de formations autour des données de recherche (par exemple des courts modules thématiques basés sur les vidéos autour des classes virtuelles) et aussi sur le contenu (avec l'accent particulier sur les exercices pratiques et les retours d'expérience). Ce travail pourra se faire en lien avec le projet DoRANum (<http://urfistinfo.hypotheses.org/2980>)

---

## Echos des Organismes de recherche

### [Loi Pour une République numérique](#)

#### **Saluons une mobilisation et une coordination exemplaires de la recherche !**

Le 29 juin au soir, la commission Mixte Paritaire (CMP) trouvait un accord sur la loi Pour une République numérique. Après plusieurs mois d'une mobilisation des milieux de la recherche, nous pouvons nous réjouir du fait que la recherche française dispose désormais d'un cadre juridique plutôt favorable.

Pour rappel, dès la consultation, nos scientifiques, nos établissements et notre association EPRIST, tout comme l'ADBU, Couperin et le CNNum, se sont mobilisés pour défendre nos positions :

- pour limiter au maximum les embargos des éditeurs sur les écrits de la recherche publique (articles scientifiques) – article 17

- pour disposer d'un article complémentaire permettant la pratique de la fouille de texte et de données sans entrave ni juridique ni technique – article 18 bis ajouté et voté par l'Assemblée Nationale grâce à ces contributions.

A l'amont, durant près d'un an, autour du Livre blanc « Une science ouverte dans une République numérique » a pu être initiée par le Conseil scientifique du CNRS, une réflexion sur les pratiques et une expertise sur les textes, qui ont compté pour l'élaboration de la Consultation nationale et, par la suite, pour la défense de nos positions.

Notre engagement a permis de mobiliser les responsables de nos organismes et nos scientifiques pour interpeller le gouvernement et les parlementaires sur ce sujet fondamental pour la place de la France dans le Monde. Il a permis d'écarter la voie contractuelle pour l'usage du TDM qui se dessinait au travers de la version de l'article 18bis proposée par le Sénat.

Jusqu'aux derniers jours avant la CMP, nous avons dû nous mobiliser pour convaincre, face aux pressions exercées.

Mais nos scientifiques ont eu l'occasion de faire la démonstration de la nécessité de disposer des outils actuels à l'ère du Big Data (fouille de texte et de données – FTD ou Text and Data Mining – TDM) pour pratiquer la recherche, afin de pouvoir jouer à armes égales avec leurs concurrents (GB, USA, Japon, etc.) ; ils ont pu montrer qu'attendre la future révision de la directive européenne n'était pas raisonnable étant d'ores et déjà confrontés à cette concurrence. Ils ont pu expliquer que la France, pourtant en avance sur le développement des outils de traitement et d'analyse sémantique (outils de FTD) commence à se faire distancer par des concurrents qui disposent d'outils moins performants mais qui ont en revanche accès à des corpus de textes et de données beaucoup plus importants du fait de l'exception au droit d'auteur dont ils bénéficient pour accéder à toute la littérature scientifique nécessaire. Le préjudice pour la France, à la fois sur le plan scientifique (positionnement sur de nouveaux champs de connaissances pour produire de nouvelles connaissances) et économique (potentiel d'innovation et d'emploi en matière de TDM) aurait été énorme et cela a finalement été bien compris du politique.

La recherche française dispose désormais d'un article 18bis dans la version Assemblée nationale assortie d'une modification du code de la propriété intellectuelle.

Plusieurs questions demeurent cependant en suspens et nous attendons les décrets d'application pour pouvoir œuvrer à l'ajustement des politiques d'établissements en matière de gestion, de traitement et de diffusion des productions de la recherche.

Au niveau européen, cette bonne nouvelle permet d'envisager avec optimisme la révision future de la directive européenne sur le droit d'auteur.

EPRIST



# Documents

## [Final Knowledge Exchange report](#)

A brief [final report](#) about the Dependency of policies on Open Access services has been produced recently by Rob Johnson from Research Consulting on behalf of Knowledge Exchange and published on the [KE website](#). We wish SHERPA and DOAJ all the best with this next step and will continue to support their work where we can.

## [Recherche multidisciplianire : comment évaluer la carrière des chercheurs ?](#)

Un atelier organisé par Science Europe s'est tenu les 1<sup>er</sup> et 2 décembre 2015 à Bruxelles sur ce thème. Le rapport est disponible ici :

[http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/SCsPublicDocs/SE\\_LEGS\\_Careerpaths\\_Workshop\\_Report.PDF](http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/SCsPublicDocs/SE_LEGS_Careerpaths_Workshop_Report.PDF)

**CNRS**

3, rue Michel-Ange  
75794 Paris Cedex 16

T. 01 44 96 40 00

F. 01 44 96 53 90

[www.cnrs.fr](http://www.cnrs.fr)

Couverture Bruno Roulet, secteur de l'imprimé Paris Michel-Ange

© CNRS Photothèque/Bordeaux Imaging Center / Sébastien Marais, Daniel Choquet, Elena Avignone