

# LA LETTRE **CorlST**

Information bimestrielle de la DIST  
et des correspondants IST des instituts

N° 15 | HIVER 2017

*Analyse systémique de la loi  
« Pour une République numé-  
rique ». Un regard novateur  
pour les décrets d'application*

**cnrs**

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

Direction de l'information scientifique et technique

# Editorial

Bonjour,

Construire l'application de la loi numérique est la grande priorité de ce début 2017 : un projet de décret, discuté par tous les acteurs et par les départements ministériels concernés (MCC et MENESR) est le trait d'union entre toutes les exigences posées par la loi. Soumis pour avis au Conseil d'Etat, ce texte deviendra le premier jalon d'une application de la loi, préalablement à toutes les dispositions, réglementaires et éventuellement législatives (droit des données) qui ponctueront les étapes d'un renouvellement en profondeur du travail de la science à l'heure numérique en France.

L'ampleur des transformations ainsi initiées par le législateur est loin d'être aujourd'hui perceptible : bon nombre des pratiques nouvelles résultant de ces choix législatifs, ne sont pas encore rendus au stade des applications pratiques. La DIST conduira en 2017 un ensemble d'actions stratégiques prioritaires qui viseront toutes, à travers les outils d'IST, à valoriser le travail législatif. La formation comme les plans de gestion des données seront des moyens privilégiés de réalisation de ces orientations. Le COPIST, de son côté, appréhendera ces évolutions à travers le renouvellement de l'offre nationale de services partagés proposés par l'Inist, Persée et le CCSD.

Plusieurs manifestations nationales (rendez-vous de Meudon en novembre, journées innoétriques) viendront ponctuer la réflexion partagée accompagnant ces actions.

Voilà donc une année au programme bien rempli que je souhaite heureuse et féconde à tous les lecteurs de la lettre CorlST.

Renaud FABRE  
Directeur de la DIST  
[renaud.fabre@cnrs-dir.fr](mailto:renaud.fabre@cnrs-dir.fr) / 01 44 96 46 57



# Actualités IST

## [DMP OPIDoR : Data Management Plan pour une Optimisation du Partage et de l'Interopérabilité des Données de la recherche](#)

Planifier la gestion des données de recherche constitue un élément clé pour leur conservation et leur réutilisation par la communauté scientifique.

Dans le cadre de la stratégie IST du CNRS, l'Inist propose [DMP OPIDoR](#) un nouveau service à destination de l'Enseignement Supérieur et de la Recherche.

Cet outil intuitif et collaboratif permet au chercheur et à ses partenaires de rédiger un plan de gestion de données<sup>1</sup>, en proposant différents modèles préconisés par les institutions et financeurs (Commission européenne...), ainsi que des guides et des exemples personnalisés.

**DMP OPIDoR** s'inscrit ainsi dans une démarche de collaboration et d'échange et intégrera les réflexions et évolutions menées au niveau national et international.

Pour toute information complémentaire: [info-opidor@inist.fr](mailto:info-opidor@inist.fr)

## [ScanR, moteur de la recherche et de l'innovation, un produit au service des communautés de l'enseignement supérieur et de la recherche](#)

Le ministère en charge de l'Enseignement supérieur et de la Recherche a mis en ligne en juillet 2016 scanR, moteur de la recherche et de l'innovation. scanR concentre, pour plus de 6 000 laboratoires publics et institutions publiques et près de 25 000 entreprises, les informations ouvertes sur leurs activités de recherche et d'innovation. scanR est donc un outil au service de la visibilité de l'ensemble des acteurs français de la recherche et de l'innovation.

scanR constitue une réponse, parmi d'autres, pour dépasser les difficultés du paysage de la donnée sur la recherche et l'innovation : multiplicité de sources dépourvues faiblement interopérables, faiblesse de l'offre ouverte de données. Une trentaine de sources majeures (HAL notamment) ont été alignées, enrichies et fiabilisées, parfois en s'appuyant sur des acteurs de terrain comme les principaux organismes de recherche nationaux ou des acteurs locaux. Pour rendre compte de la recherche en cours, l'équipe exploite les informations non structurées diffusées par les entreprises et les laboratoires publics eux-mêmes sur leur site web.

Pour réaliser cette extraction des informations du web et pour concevoir une application adaptée aux besoins des publics, l'équipe s'est enrichie des compétences de c-Radar, sélectionné au terme d'un appel d'offre. Le projet a été conduit en 6 mois.

Lancé le 6 juillet 2016, scanR a suscité l'intérêt du public et des professionnels (déjà plus de 1 100 tweets et retweets) et a rapidement trouvé ses utilisateurs (plus de 170 000 pages vues). On compte plus de 5 000 téléchargements des jeux de données proposés autour des sources majeures ([tinyurl.com/OPENDATAscanR](http://tinyurl.com/OPENDATAscanR)). L'API « Dis-moi tout sur la

---

<sup>1</sup> Un Plan de gestion des données ou Data Management Plan (DMP) est un document évolutif décrivant la collecte, le traitement, la documentation, le stockage, l'archivage, la conservation et le partage des données tout au long de leur cycle de vie dans le cadre d'un projet de recherche.

recherche de cette entreprise/ce laboratoire » ([tinyurl.com/APIscanR](http://tinyurl.com/APIscanR)) est exposée de manière ouverte.

L'équipe d'administration s'attache à développer des liens avec les différentes communautés intéressées. scanR est encore un projet en devenir qui doit constituer une gouvernance en accord avec sa philosophie « ouverte ». Il doit intégrer de nouvelles sources, internationaliser l'approche pour caractériser avec autant de pertinence la recherche de nos partenaires étrangers, travailler avec ses utilisateurs pour en adapter les fonctionnalités et améliorer la qualité de l'information. L'équipe souhaite ainsi organiser en 2017 un évènement de type data camp pour développer les usages des données de scanR par une communauté élargie.

scanR <https://scanr.enseignementsup-recherche.gouv.fr/>

Accès aux données ouvertes liées à scanR <https://scanr.enseignementsup-recherche.gouv.fr/>

### Persée ouvre son triplestore : data.persee.fr et renforce son inscription dans l'environnement des données liées

*Afin de répondre aux usages actuels des communautés scientifiques et d'aller encore plus loin dans l'exploitation et la réutilisation des données, l'UMS Persée expose ses données au format RDF<sup>2</sup> au travers de son triplestore data.persee.fr.*

L'UMS Persée est une Unité Mixte de Service réunissant l'Université de Lyon, le CNRS et l'ENS de Lyon et bénéficiant du soutien du MENESR. L'UMS, réalise depuis plus de 10 ans le portail Persée<sup>3</sup> regroupant des milliers de publications scientifiques accessibles librement et gratuitement et comptabilisant 30 millions de visites annuelles. Le portail permet une recherche dans plus de 650 000 documents au travers d'outils de navigation exploitant trois index (documents, illustrations et auteurs) et proposant des fonctionnalités de tri par facettes.

Le triplestore rassemble les données décrivant les ressources produites par Persée de manière structurée selon les principes du web sémantique. Ce nouveau point d'entrée à l'ensemble des métadonnées offre aux chercheurs plus de liberté. Il est ainsi possible à chacun de définir des corpus correspondant à son propre questionnement, ou de récupérer des jeux de données. Les utilisateurs du triplestore pourront réutiliser, retravailler et à terme enrichir les données sélectionnées librement et gratuitement.

Pour les chercheurs, les alignements avec des référentiels nationaux et internationaux permettent de questionner non plus seulement les données de Persée, mais également celles proposées par IdRef, DBpedia, data.bnf.fr, le Cairo Gazetteer, GBIF, ... en une seule et même requête. Plusieurs modes d'exploration sont disponibles : le sparql endpoint pour les initiés et l'outil Sparklis qui permettra au plus grand nombre de formuler ses requêtes en langage naturel, sans connaissance *a priori* des contenus et de leurs modèles. Différents outils de visualisation permettent d'afficher les données et jeux de données sous une forme

---

<sup>2</sup> RDF : Resource Description Framework

<sup>3</sup> [www.persee.fr](http://www.persee.fr)

graphique. Enfin, une gamme de services est également proposée pour aider à s'approprier data.persee.fr et à en exploiter toute la richesse :

- un ensemble de tutoriels afin de prendre en main les interfaces de recherche ;
- des lieux d'interaction avec Persée permettant la co-construction de jeux de données et d'outils pour les exploiter.

Des formations seront également mises en place afin de répondre aux besoins des chercheurs.

Avec l'ouverture de son triplestore, Persée entend :

- affirmer sa position d'acteur de la dynamique nationale d'ouverture des données ;
- répondre aux besoins de constitution de corpus complexes, de fouille et d'exploitation des données de la communauté scientifique.

Le site : <http://data.persee.fr>

Contacts :

Nathalie Fargier | [nathalie.fargier@persee.fr](mailto:nathalie.fargier@persee.fr) | 04 26 73 14 80

Viviane Boulétreau | [viviane.bouletreau@persee.fr](mailto:viviane.bouletreau@persee.fr) | 04 26 73 14 82

### **Conditor : Signalement mutualisé de la production scientifique de l'ESR**

#### **Un projet multi-partenaires soutenu par le MENESR dans le cadre de la BSN et porté par la DIST du CNRS**

Malgré la variété des dispositifs de référencement existants au sein des établissements de l'ESR (Hal et les archives institutionnelles, les systèmes d'information recherche tels que Ribac,...), il reste impossible de savoir ce qui est produit par la recherche publique française dans son ensemble. Les bases internationales telles que WoS ou Scopus n'en répertorient qu'une partie, certes importante mais qu'il faut réussir à repérer et dans tous les cas leur objectif n'est pas l'exhaustivité en la matière : Conditor est né de ce constat dans le cadre de la BSN.

Il vise non seulement à identifier la production de l'ESR mais à rationaliser progressivement les façons de faire pour éviter les saisies et repérages multiples d'une même production dans le WoS ou Scopus par exemple.

Il s'agit concrètement de mettre en place un service mutualisé pérenne :

- un « pot » commun de métadonnées articulé à Hal alimenté au fil de l'eau par des sources existantes ESR ou internationales (libres et gratuites ou commerciales), signalant l'ensemble de la production de la recherche publique française, en s'appuyant sur des référentiels communs dont le Répertoire National des Structures de Recherche (RNSR),
- un « outillage » adapté pour le construire (conversion de format, détection des doublons, alignement avec les référentiels,...) et fournir en métadonnées les applicatifs existants ou en développement chez les partenaires de l'ESR pour en améliorer la complétude et la qualité,
- une organisation multipartenaires pour le faire vivre : réseau métier réparti dans les établissements (fournisseurs, administrateurs, utilisateurs des données produites), équipe informatique (administration, maintenance), structure de pilotage du service (évolutions du produit ...).

Dès la genèse du projet, le pragmatisme, la complémentarité, la collaboration inter-établissement et la transparence ont été la règle. Après une expérimentation réussie, **le lancement officiel a eu lieu le 5 décembre 2016** et a réuni les représentants du ministère de l'Education nationale et de l'enseignement supérieur (MENESR), les partenaires de l'expérimentation et de la phase projet (voir ci-dessous) ainsi que les principaux soutiens.

Les travaux à mener pour aboutir en 2 ans à un premier corpus, un outillage aussi industrialisé que possible et une organisation multipartenaires du service ont été découpés en six ensembles pris en charge par une équipe multi-partenaires autour d'un ou deux partenaires pilote ou copilotes :

1. Gestion du projet (Dist du CNRS)
2. Négociation des sources de données et formalisation des usages des données collectées et produites par Conditon (Istea)
3. Elaboration de la stratégie de construction d'un corpus de notices (Abes et Inist CNRS)
4. Conception/développement itératif/déploiement de l'applicatif (Inist CNRS et CCSD)
5. Mise en place du service Conditon (Université Paris Diderot)
6. Communication (Université de Bordeaux)

Les acteurs impliqués à différents titres dans la phase projet sont l'Abes, l'Amue, le CCSD, l'Inist, la Dist, l'InSHS et le SAP2S du CNRS, Huma-Num, l'Inra, Inria, l'IRD, l'Irstea, le MENESR (RNSR et ScanR), l'OST du HCERES, ainsi que les universités de Bordeaux, Paris-Dauphine, Paris-Diderot et l'UPMC.

### Indicateurs Couperin : premiers enseignements

Le département des négociations documentaires Couperin.org propose des indicateurs présentés à la dernière Journée des négociateurs qui a eu lieu en décembre 2016. Ils sont établis à partir :

- des données 2015 du Journal Citation Report de Clarivate Analytics
- des données de l'enquête nationale sur les ressources électroniques (ERE 2015)
- des statistiques d'utilisation 2015 des revues électroniques (chiffres partiels bruts et/ou consolidés).

Ces indicateurs permettront aux négociateurs de comparer 13 ressources entre elles que ce soit au niveau du coût ou des usages.

On peut notamment en tirer quelques éléments globaux parmi les 13 éditeurs étudiés (Elsevier, Springer, Wiley, OUP, IOP, APS, AIP, OSA, ACS, RSC, ACM, EDP Sciences et Taylor & Francis).

- Les éditeurs de revues scientifiques qui totalisent le plus d'usage au sein de Couperin.org sont Elsevier/Wiley/Springer mais aussi American Chemical Society (ACS)

Rapportés au nombre de revues proposées, ce sont les titres des sociétés savantes (dont 2 en physique non proposées sur BibCNRS) qui sont les plus utilisés et en tête American

Physical Society (APS), ACS et American Institute of Physics (AIP). Ils sont suivis par les titres de la Royal Society of Chemistry et Elsevier.

- Les éditeurs les plus cités par article : ACS/APS/Wiley et EDP Sciences/Elsevier/ACM/Springer.
- Les éditeurs les plus coûteux par article (moyenne rapportée à l'ensemble des abonnés Couperin.org) : EDP Sciences/Wiley/ACM.
- Les éditeurs les plus coûteux par article (médiane calculée sur l'ensemble des abonnés Couperin.org) : Wiley/Springer/APS.

Le seul éditeur français de la liste est EDP Sciences dont les articles sont bien cités (3ème position avec Wiley) mais qui se situe parmi les plus chers (Coût moyen/ téléchargement ; sont pris en compte dans ce calcul les établissements qui bénéficient d'un tarif négocié Couperin.org, les statistiques d'utilisation disponibles et les coûts déclarés dans l'ERE, coûts vérifiés, modifiés ou complétés le cas échéant grâce aux documents de négociation).

### Séminaire données scientifiques - 16 février 2017

La BSN10 et EPRIST ont organisé le 16 février 2017 une journée intitulée : « **outils et services pour la gestion et l'ouverture des données de la recherche : retour d'expériences** » <http://seminaire.inra.fr/data>

Ce séminaire, qui s'est tenu à l'Inra ( rue de l'université, Paris), a été l'occasion :

- de présenter les outils et services (d'ores et déjà opérationnels ou en gestation) et d'en étudier la pertinence à la lueur de cas d'usage
- de discuter les orientations institutionnelles visant à instaurer de bonnes pratiques de gestion **et d'ouverture** des données scientifiques parmi les chercheurs.

Cette journée , très appliquée, était destinée à un large public d'acteurs ESR dans la chaine de gestion/diffusion/valorisation des données de la recherche : professionnels IST, bibliothécaires, ingénieurs/chercheurs gestionnaires de données, ingénieurs de centre de traitement de données,... consultez le programme en ligne

<https://seminaire.inra.fr/data/Pre-programme>

[Michael Witt](#), directeur du D2C2 (Distributed Data Curation Center) de l'Université de Purdue, Indiana est intervenu en keynote speaker.

# Dossier

## L'analyse systémique de la loi «Pour une République numérique» - Un regard novateur pour les décrets d'application

Les articles 30 et 38, soit l'inscription d'un droit d'exploitation secondaire des publications scientifiques pour les chercheurs et l'autorisation de la fouille de textes et de données, de la Loi pour une République numérique vont être précisés dans des décrets d'application. Ces objets juridiques, rédigés par le Ministère de l'Education nationale et de l'Enseignement supérieur, vont restreindre ou au contraire élargir la portée juridique des articles législatifs. Ils sont donc au cœur d'un vaste enjeu d'écriture.

Afin de bien comprendre les objectifs de la légistique, soit de l'écriture de la législation, la DIST du CNRS a réalisé une analyse systémique de la « Loi pour une République numérique ». L'idée est d'identifier les dispositions qui seront ou pourront être appliquées aux trois notions clés de la science ouverte : les plateformes, les écrits scientifiques, les données de la recherche.

Ces trois concepts existent légalement sous plusieurs angles : les lois (Code de la recherche, Code de la propriété intellectuelle...) ou encore les décrets (Décret constitutif du CNRS par exemple). L'objectif de cette analyse systémique est donc de clarifier leur utilisation, de cerner les limites associées à cette utilisation (critère d'exhaustivité) et de mettre en lumière les points de divergence dans l'écriture des décrets d'application (critère de vérification), afin que la loi facilite des pratiques concrètes et simples pour les chercheurs.

### **Qu'est-ce qu'une analyse systémique ? Un point théorique**

L'analyse systémique est une étude interdisciplinaire qui cherche à comprendre les interactions entre les différents éléments d'un système complexe. En l'occurrence, le droit de la science est ce système complexe dans lequel les différents éléments sont écrits à plusieurs endroits, de façon différente, selon des définitions parfois contradictoires.

Outil théorique, d'inspiration philosophique, l'analyse systémique trouve aujourd'hui des applications pratiques dans certains domaines comme l'éducation, le droit ou la science.

Pour plus d'informations, voir *La systémique* par Daniel Durand (1979).

### **Comment l'analyse a-t-elle été faite ?**

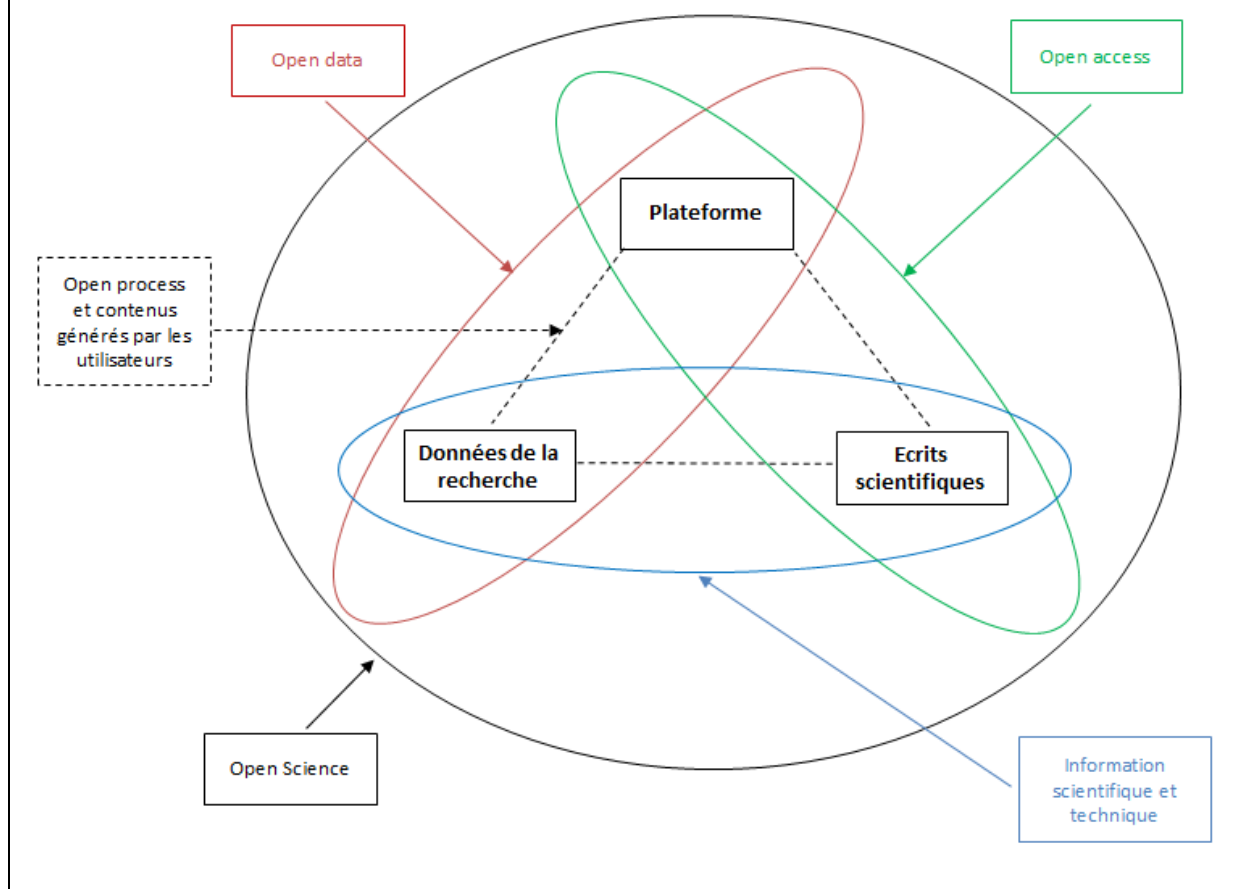
L'analyse a mis en écho différents éléments pour comprendre leurs similarités, leurs contradictions, afin de les résoudre. Elle a procédé par analyse quantitative (représentativité forte ou faible d'une expression selon sa présence dans l'ensemble des textes par exemple), par analyse qualitative, par analyse sémantique (quelle différence entre « exploitation de données » et « fouille de données » par exemple) et par analyse lexicographique.

A partir de ces analyses, on aboutit à des croisements entre les concepts centraux (plateformes, données de la recherche, écrits scientifiques) et la manière dont ils



s'articulent entre eux. Sont ainsi mis en lumière les grands morceaux du système de la loi : l'Open Data, l'Open Access, l'Information Scientifique et Technique.

Le schéma ci-après résume ces mises en relation.



Forte de cette analyse, la DIST a pu dessiner les relations entre les concepts clés. Nous résumons ici ces relations ainsi que les recommandations finales de l'analyse systémique de la Loi pour une République numérique.

## **Les relations entre «Plateformes», «Données de la recherche» et «Ecrits scientifiques» dans l'analyse systémique**

Comme le précise le texte de l'analyse, « lorsque ces concepts sont pris deux à deux, ils forment une nouvelle notion juridique encadrée par un régime qui lui est propre. Ce régime défini par la loi pour une République numérique peut se heurter à des régimes juridiques existants. » (p.49)

Du point de vue de l'Open Data, l'ouverture des données de la recherche se distingue du cadre général sur deux points : c'est un acte de mise à disposition (i) obligatoire et (ii) volontaire. L'objectif devra donc être de préciser ce qui peut ou ne pas être mis en ligne selon la réglementation particulière, ce qui peut ou ne peut pas pour des raisons de valorisation (interrogeant ainsi le statut *volontaire* du dépôt) et ce qui doit être mis en ligne à des fins de réutilisation par des tiers. Une chartre éthique sur ces pratiques de dépôt est souhaitable, ainsi qu'une procédure interne de mise à disposition. Une licence d'utilisation des données de la recherche devrait être homologuée afin d'uniformiser les pratiques.

Règles et bonnes pratiques de mise à disposition des écrits scientifiques devront être précisés dans un guide d'application. Afin d'aiguiller encore plus l'activité des chercheurs, il conviendrait de créer un contrat type de publication scientifique, par arrêté, afin de limiter les interprétations divergentes entre dispositions contradictoires. Du point de vue de l'Open Access, il convient aussi de préciser par décret l'application dans le temps de la loi, afin de bien souligner que les dispositions s'appliquent aux contrats en cours.

Dans ce travail d'écriture des décrets, des recouvrements sémantiques sont à mettre en avant, afin de clarifier au maximum l'interprétation possible de la loi. C'est le cas notamment de la fouille de textes et de données. Surtout, pour celle-ci, d'autres précisions sont à porter, comme par exemple : le caractère d'ordre public de l'exception de TDM, la fixation des règles relatives aux limitations techniques, les règles de communications des données de recherche issues de la fouille. Les décrets ne pourront de toute façon que tenir compte du projet de directive Droit d'auteur dans le marché unique numérique de la Commission européenne.

## **Quelles recommandations à la suite de l'analyse systémique ?**

L'analyse systémique explique ainsi que la dynamique du système pris dans son ensemble et des concepts et relations impliquent :

- une sécurisation des textes, expressions et conditions d'application par décret ;
- une sécurisation des pratiques par :
  - o un guide d'application ;
  - o un guide des bonnes pratiques ;
  - o une charte éthique ;
  - o une sensibilisation ;
- un encadrement des pratiques par un contrat type de publication scientifique ;
- une supervision par une Agence de la Science ouverte.

Concernant la sécurisation des pratiques et l'encadrement, il est possible d'apporter ici quelques précisions supplémentaires. Le guide d'application viserait à définir en des termes clairs les droits, points de vigilance et interdictions des chercheurs dans l'application du droit d'exploitation secondaire, dans la publication des données de la recherche et dans l'utilisation des outils de *text and data mining*. Formation et session de sensibilisation seront de bonnes modalités d'apprentissage juridique pour l'ensemble des chercheurs. Dans le même sens, une charte éthique doit définir les valeurs d'accès et de partage de la Science, encadrer les pratiques et alerter les chercheurs sur le respect des droits fondamentaux que sont le respect des secrets, le droit à la vie privée ou encore le droit des données à caractère personnel, le droit de la propriété intellectuelle.

L'Agence de la Science ouverte serait quant à elle le lieu de convergence des points de vue des acteurs publics et privés de la Science ouverte, de sécurisation et de contrôle des pratiques, de définition des bonnes pratiques et de respect de la charte éthique et pourrait, entre autre, être en charge de la rédaction du rapport sur l'impact du principe de libre accès aux données scientifiques sur le marché de l'édition scientifique et sur la circulation des idées et des données scientifiques.

Pour en savoir plus :

[http://www.cnrs.fr/dist/z-outils/documents/20170203\\_analyse%20syst%C3%A9mique\\_vf.pdf](http://www.cnrs.fr/dist/z-outils/documents/20170203_analyse%20syst%C3%A9mique_vf.pdf)

## **Les articles 30 et 38 de la loi « Pour une République numérique »**

### **Article 30**

Le chapitre III du titre III du livre V du code de la recherche est complété par un article L. 533-4 ainsi rédigé : « Art. L. 533-4.-I.-Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un périodique paraissant au moins une fois par an, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, sous réserve de l'accord des éventuels coauteurs, la version finale de son manuscrit acceptée pour publication, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique ou, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est au maximum de six mois pour une publication dans le domaine des sciences, de la technique et de la médecine et de douze mois dans celui des sciences humaines et sociales.

« La version mise à disposition en application du premier alinéa ne peut faire l'objet d'une exploitation dans le cadre d'une activité d'édition à caractère commercial.

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

« III.- L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.

« IV.- Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. »

### **Article 38**

Le code de la propriété intellectuelle est ainsi modifié :

1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé :

« 10° 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. 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 ; »

2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé :

« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui

y a licitement accès, en vue de fouilles de textes et de données incluses ou associées aux écrits scientifiques dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites. »



# Formation, stages

## ANF des réseaux Médiçi, rBDD et Renatis

En 2017 les réseaux Médiçi, rBDD et Renatis seront porteurs des ANF suivantes, soutenues par la Mission pour l'Interdisciplinarité :

### Médiçi

#### **« De la réception du manuscrit à la relecture d'épreuve - organisation et suivi éditorial »**

L'objectif de la formation est de préparer les personnels de l'édition à la bonne organisation et gestion des différentes étapes de production et les différents intervenants pour tenir les délais proposés. L'enjeu est de maintenir un outil de production de qualité au sein des établissements publics. Le souhait est aussi de transmettre une base solide de la connaissance des bonnes pratiques de l'édition scientifique publique qui en l'absence de réglementation précise doit pouvoir être contrôlée par le réseau professionnel qui est garant de bonne transmission des savoir-faire professionnels.

La délégation régionale de Normandie assurera le soutien logistique de cette formation qui se tiendra à Caen en septembre pendant 4 jours en résidentiel.

### Renatis et Médiçi

#### **« Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données - deuxième session »**

Ayant suscité un grand intérêt l'année dernière cette formation fait suite à l'ANF 2016 du même nom. Le comité de pilotage du réseau Renatis a voulu reconduire cette formation, cette fois-ci en collaboration avec le réseau Médiçi et cocommanditée par la MI et la DIST.

Renatis souhaite continuer à travailler tout au long de l'année 2017 avec les comités de pilotage des réseaux RBDD et Médiçi sur la thématique des données, de façon à croiser les demandes des différents publics, et les diverses façons d'y répondre et ce en tenant compte des spécificités disciplinaires et des structurations par les TGIR et les Infrastructures de Recherche.

Cette formation a pour objectif de fournir des éléments pédagogiques, méthodologiques, techniques, juridiques, aux professionnels de l'IST afin qu'ils puissent sensibiliser les chercheurs à la gestion des données et à leur cycle de vie. Cette formation intéressera également les chercheurs, enseignants-chercheurs et doctorants souhaitant s'impliquer dans la gestion et la gouvernance des données de la recherche. La participation de binômes chercheurs / ingénieurs est encouragée. Cette ANF s'inscrit dans un cadre global, complémentaire au programme des formations à distance DoRANum soutenu par le MENESR dans le cadre des segments BSN9 et BSN10 de la Bibliothèque Scientifique Numérique.

La délégation régionale de Bretagne et Pays de la Loire assurera le soutien logistique de cette formation qui se tiendra à Nancy (INIST), la première semaine de juillet.

### rBDD

#### **« Utilisation et maîtrise d'un ETL : intégrations de données avec Talend Open Studio »**

Cette formation fait suite à l'atelier **« Découvrir Talend un TP pour manipuler les briques de base »** des journées RBDD intitulées **« Conduire et construire un plan de gestion des données : de la base de données à la pérennisation »** tenues en octobre 2015.

Il a mis en évidence un besoin important de la communauté du réseau rBDD, chercheurs et ingénieurs, concernant l'organisation de l'intégration et du traitement des données. Talend Open Studio est l'outil ETL plébiscité par la communauté. Il permet la conception de processus d'intégration de données sous un environnement graphique de développement simple à utiliser. Avec son organisation de connecteurs intégrés, il autorise des déploiements rapides de traitement et supporte tous les types d'opérations d'intégration, de migration et de synchronisation de données.

La délégation régionale Languedoc-Rousillon assurera le soutien logistique de cette formation de deux jours.

#### « PostgreSQL Administration »

Partant du constat que de plus en plus d'ingénieurs en bases de données souhaitent se former à PostgreSQL, le réseau a fait le choix de former des formateurs internes en PostgreSQL en 2016. Avant de la déployer en région par l'intermédiaire des formations permanentes, le réseau souhaite la rôder dans le cadre d'une ANF.

La DR Languedoc-Rousillon assurera le soutien logistique de cette formation de trois jours en résidentiel.

#### « Autour du langage SQL ... »

Le langage SQL est le langage normalisé servant à exploiter les bases de données relationnelles. Sa connaissance est indispensable pour mener à bien un projet de bases de données relationnelles (de sa création à son exploitation). A l'issue de la formation les stagiaires seront capables de : maîtriser les 3 domaines du langage (manipulation des données (LMD), définition des données (LDD), contrôle d'accès aux données (LCD)), avoir un aperçu du SGBD PostGreSQL et savoir utiliser les principaux clients (PSQL, PGAdmin, QGIS, R, ...).

La délégation régionale Languedoc-Rousillon assurera le soutien logistique de cette formation de deux jours en résidentiel.

---

## Echos des Organismes de recherche

### [Bienvenue aux Journées Abes 2017 !](#)

Les Journées Abes auront lieu les 10 et 11 mai au Corum, Palais des Congrès de Montpellier. Comme chaque année, les professionnels de l'IST du CNRS et des organismes de recherche - y sont cordialement invités. Remarque importante : les tarifs d'inscription à leur destination ont été revus à la baisse et alignés sur ceux des établissements membres des réseaux (Sudoc, Thèses, Calames) gérés par l'Abes.

David Aymonin, directeur de l'Abes depuis septembre 2016 ouvrira les Journées 2017 qui proposeront un panorama des projets à l'œuvre dans le monde de la documentation en cette période de transition –numérique, bibliographique ...et des pratiques. Ainsi, lors de différentes tables rondes, il sera question de la mise en œuvre de l'infrastructure Collex-Persée, des premières vagues de déploiement des Systèmes de Gestion de Bibliothèques « nouvelle génération » (SGBm) ou des chantiers menés avec les grands partenaires tels

l'INIST-CNRS - pour le dispositif ISTEEX - ou la BnF - pour le projet « Fichier National des Entités »...

Un programme chargé donc, mais qui, comme chaque année, fera la part belle à la convivialité !

L'équipe organisatrice vous donne rendez-vous sur le site de l'Abes (<http://www.abes.fr/>) à partir du 15 mars, date de la publication du programme et de l'ouverture des inscriptions.

## Documents

### Fouille de textes et de données (TDM)

Si la situation du TDM sur le plan national est en passe d'être éclaircie avec la loi numérique et la prochaine parution du décret d'application correspondant, il n'en est pas de même à l'échelle européenne. En effet, le processus de révision de la directive européenne sur les droits d'auteurs et droits voisins (2001/29/CE) à peine enclenché, le débat autour du TDM commence. A la suite du rapport du parlement porté par la députée européenne Julia Reda, un projet de modification de la directive, soutenant une exception en faveur du TDM a été voté par le parlement :

<https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market>

La commission CULT (Culture et Education) du Parlement Européen vient de fournir un avis concernant ce projet de directive, qui demande explicitement que les éditeurs soient rémunérés pour la mise à disposition des contenus pour des activités de TDM (amendement 5 page 9), niant par là même l'idée portée par le monde de la recherche : « the right to read is the right to mine ».

<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=%2f%2fEP%2f%2fNONSGML%2bCOMPARL%2bPE-595.591%2b01%2bDOC%2bPDF%2bV0%2f%2fEN>

### TDM et European Open science Cloud

Le développement du Cloud Recherche européen est l'occasion d'aborder également le sujet de la fouille de textes et de données à l'échelle européenne. A la suite de la communication de la Commission présentant le projet de Cloud européen () les instances destinataires (Parlement, Conseil, Comité économique et social européen et comité des régions) sont amenées à commenter ce texte. La Commission parlementaire ITRE (Industrie, Recherche et Energie) en charge de ce dossier du Cloud pour le Parlement est la première à réagir à travers ce rapport qui propose une motion au vote du Parlement, soutenant notamment le TDM et plaidant pour un environnement juridique des droits d'auteurs modernisé (§ 86 et suiv.)

<http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&mode=XML&reference=A8-2017-0006&language=EN>

### Pour mémoire et en savoir plus sur le TDM....

La DIST et l'Inist coopèrent à une veille IST en continu sur le TDM, mise à votre disposition à partir du site : <http://www.donneesdelarecherche.fr/spip.php?article332>

## 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 et impression secteur de l'imprimé du service mutualisé d'Ile de France

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