

# Inist

## RAPPORT D'ACTIVITÉ 2017



cnrs

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

Inist  
Information et services numériques pour la recherche



# Édito

## « Ingénierie des connaissances 2018-2022 » : acte 2 du repositionnement de l'Inist

2017 fut une année de transition (encore une, pourrions-nous dire !) pour l'Inist. En effet, l'unité est en évolution depuis un certain nombre d'années. Le projet « Ingénierie des connaissances 2014-2018 » porté par mon prédécesseur représente un tournant majeur pour l'unité, avec une offre de services renouvelée en direction de la recherche. Ce projet s'est accompagné d'une réflexion importante sur la partie RH, ayant conduit au repositionnement de 50 % des personnels et s'étant poursuivie dans le cadre d'une démarche GPEC.

« Ingénierie des connaissances 2018-2022 » entérine le repositionnement de l'unité en tournant définitivement la page de l'Inist tel qu'il avait été conçu dans les années quatre-vingt.

La rédaction du projet a permis l'instauration d'un dialogue permanent avec la Direction générale déléguée à la science (DGDS). Les activités sont recentrées sur l'analyse et la fouille de données, la valorisation des données de la recherche, l'accès à la documentation scientifique et, en partenariat, l'édition numérique et la modération dans HAL. Le projet implique notamment l'arrêt du service Fourniture de documents acté au 31 décembre 2018 et s'accompagne d'une optimisation des fonctions support.

Ce rapport d'activité se veut l'expression du travail des équipes – que je remercie ici sincèrement pour leur implication – mené malgré les soubresauts que continue de connaître l'Inist. Le portail OPIDoR a été ouvert, la plateforme ISTEEX s'est enrichie de milliers de documents ; nous contribuons activement à de nombreux projets nationaux destinés à favoriser la diffusion de la connaissance et à valoriser la recherche, comme Conditor ou Visa TM. En interne, nous tentons d'améliorer encore nos modes de travail, favorisant agilité et transversalité.

Cet éditto est le dernier que je signe en tant que directrice de l'Inist. Je souhaite vivement que l'Institut puisse consolider sa place d'opérateur de la stratégie IST du CNRS et plus largement de l'ESR (enseignement supérieur et recherche), au service de la science ouverte qui fait désormais l'objet d'un plan national et dont nous pouvons – devons, en tant qu'unité propre de service du CNRS – être un acteur majeur.

Dominique WOLF  
Directrice de l'Inist



@INIST\_CNRS

# Somr

**Édito** 3

**Faits marquants** 6

## **AXE 1**

### **Analyse et fouille de l'information scientifique**

Contribuer au pilotage scientifique 8

Construire des terminologies et fouiller des textes 10

## **AXE 2**

### **Valorisation des données de la recherche**

Valoriser les bases de données des laboratoires 12

Optimiser le partage et l'interopérabilité des données de recherche 14

Se former à distance sur les données de la recherche 16

# maire

## **AXE 3** **Accès à l'information**

Consulter les ressources documentaires électroniques du CNRS	18
Accéder aux collections rétrospectives de la littérature scientifique	20
Mesurer les usages des ressources électroniques	22

## **Accompagner la recherche**

Publier les résultats de la recherche scientifique	24
Former à l'information scientifique	27

<b>L'Institut en chiffres</b>	<b>28</b>
-------------------------------	-----------

# Faits marquants



## Janvier

Lancement du projet **LOTERRÉ** (Linked Open Terminology Resources) plateforme d'exposition de ressources terminologiques.

Création du **GDR SémanDiv** (Sémantique de la biodiversité), Inist est membre de ce GDR.



## Juin

Refonte des sites **ezPAARSE** et **ezMESURE**.

Premiers chargements dans **Conditor**.



## Mars

**Carrefour de l'IST** « Les innovations de rupture : quelles missions & compétences à réinventer pour les professionnels de l'IST ? »

**Forum ouvert** sur le thème : « Soyons accélérateur de la recherche ; osons les métiers de demain. »



## Juillet

Démarrage du projet **Visa TM** : instance nationale de services de Text and Data Mining.

Création du **GIS « Réseau des Urfist »**, Inist est membre du GIS.



### Août

Ouverture du service dl.istex.fr qui permet le téléchargement de **corpus ISTEX**.



### Octobre

Validation du projet « **Ingénierie des connaissances 2018-2022** » par la DGDS du CNRS.

Ouverture du service permettant le dépôt des statistiques d'usage dans **ezMESURE**.

**Journée d'étude Open Science** : partage des publications et des données.



### Septembre

Ouverture de **Cat OPIDoR**, cartographie des services français de gestion et valorisation des données de la recherche.

Ouverture du **portail OPIDoR** : outils et services pour optimiser le partage et l'interopérabilité des données de la recherche.



### Décembre

**DoRANum** : mise en accès libre de plus de 60 ressources pédagogiques sur les données de la recherche.

Cap des 20 millions d'objets documentaires disponibles dans **ISTEX**.

## Contribuer au pilotage scientifique

---

L'offre de services de l'Institut dédiée à l'appui au pilotage scientifique propose une veille documentaire, afin de dresser un panorama, repérer des experts, rechercher des partenaires sur un sujet précis et mener des études bibliométriques permettant de caractériser l'activité scientifique d'un laboratoire, d'un institut ou celle liée à une thématique particulière.



### Collaboration avec la Direction de l'innovation et des relations avec les entreprises (Dire)

L'Inist collabore avec la Dire du CNRS dans le cadre de la réalisation d'études thématiques utilisant les publications et brevets pour la mise en évidence des acteurs CNRS concernés, de leurs collaborations, de leur position internationale, en coopération avec « CNRS Innovation » dans le cadre des Focus transfert CNRS. Les études réalisées en 2017 portent sur les térahertz, les diagnostics et traitements non médicamenteux de la maladie d'Alzheimer, les lasers à cascade quantique (QCL) et enfin le VIH. D'autres études bibliométriques, portant sur la valorisation du CO ou l'épigénétique en oncologie, ont été conduites.

Enfin, des réponses à des questions concernant les collaborations industrielles du CNRS, notamment avec les groupes Solvay et PSA, ont par ailleurs été apportées.



### Collaboration avec le Service d'appui à la politique et à la prospective scientifiques du CNRS (SAP2S)

L'Inist participe au repérage des adresses du CNRS dans les données du Web of Science pour les besoins de l'OST (Observatoire des sciences et des techniques), pour le calcul des indicateurs qui accompagnent le rapport annuel d'activité du CNRS ou la réalisation d'études bibliométriques commandées par les instituts. L'implication dans ce marquage a été renforcée, les données du SAP2S étant désormais traitées selon une méthodologie proposée par l'unité.

L'hébergement des données acquises par le SAP2S auprès de Thomson Reuters/Clarivate Analytics entre également dans ce champ d'activité à travers un travail de consolidation et de reformatage des données livrées.



### Collaboration avec Envirhônalp (Groupement d'intérêt scientifique de recherche dans le domaine de l'environnement)

L'Inist a réalisé un rapport ezVIS (outil libre développé par l'Inist dédié à la constitution de rapports web consultables en ligne) pour les données de publication du GIS de l'année 2016 suite à l'extension du GIS à la région Auvergne-Rhône-Alpes.

### Publications de l'INSU (Institut national des sciences de l'Univers)

Les études réalisées pour cet institut portent sur les recherches dans le domaine des surfaces et interfaces continentales, sur les différents programmes du projet Mistrals et enfin sur la production scientifique liée aux TGR auxquelles participe l'Institut : CFHT (Télescope Canada-France-Hawaï) et IRAM (Institut de radioastronomie millimétrique).

#### Chiffres clés

**11**  
études  
bibliométriques

**15**  
recherches  
bibliographiques

**23**  
profils



#### Formations

Participation à l'enseignement dans l'UE 906 EC1 Services documentaires électroniques du Master 2 Information-Communication de l'Université de Lorraine.



## Construire des terminologies et fouiller des textes

À l'heure des nouveaux usages numériques et particulièrement de la massification des données, les enjeux liés au traitement automatique de la langue sont désormais bien identifiés : faire progresser l'intelligence artificielle, mais aussi développer des services et outils en mesure de procéder à la fouille et l'analyse de grandes quantités d'informations, notamment par des traitements relevant du Text and Data Mining (TDM).

Acteur clairement identifié dans ce domaine, l'Inist propose des services centrés sur la mise à disposition de ressources terminologiques fortement orientés vers le web de données et l'utilisation d'outils de fouille de texte.



### Visa TM : vers une infrastructure de services avancés pour le « Text-Mining »

Le projet « Visa TM » associe l'Inra, l'Inist et l'Université de Montpellier. Il a été validé par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI), sous l'égide de la BSN 10, au printemps 2017, et ce pour une durée de deux ans. Son objectif est d'étudier les conditions de production de services TDM à haute valeur ajoutée fondés sur l'analyse sémantique destinés aux chercheurs, à l'aide d'infrastructures qui facilitent le partage et la réutilisation par des non-spécialistes.

Le projet est organisé en trois volets :

- le volet « étude » a pour objectif de dresser un état des lieux des acteurs (recherche et service) à l'échelle nationale et de consolider l'analyse des besoins ;

- le volet « conception » vise à démontrer, *via* des développements informatiques, la faisabilité et l'intérêt de l'interconnexion des trois types d'infrastructure, TDM (OpenMinTed), bibliothèque numérique (HAL ou ISTEEX) et ressources sémantiques (Agro-Portal) ;
- le volet « application pilote » doit démontrer la facilité de déploiement d'un nouveau service TDM à base de composants des trois infrastructures.

L'Inist s'appuie sur une équipe pluridisciplinaire composée d'informaticiens et de membres de l'équipe Terminologie-TAL pour mettre en œuvre ce projet.

<https://admin.core-cloud.net/ou/UPS76/visatm/Site-Pages/Accueil.aspx>



### LOTERRÉ

Tout au long de l'année 2017 se sont déroulés les travaux de réalisation de la plateforme de mise à disposition de ressources terminologiques scientifiques multidisciplinaires LOTERRÉ (Linked Open Terminology Resources).

Cette plateforme est fondée sur les technologies du web de données. Pour faciliter les échanges et l'interopérabilité, elle s'appuie sur un triplestore (base de données de triplets RDF) et permet la consultation des ressources, leur interrogation *via* une API et leur téléchargement.

Les travaux réalisés s'organisent en deux volets :

- conception et développement du site : architecture ; test, choix et paramétrage des applicatifs ; charte graphique ; contenu éditorial ;
- travaux sur les données à exposer : définition du modèle de données ; détermination des métadonnées ; mise en forme des ressources ; choix d'identifiants pérennes.

L'objectif visé est une mise en ligne au premier trimestre 2018.

<https://www.loterre.fr>



@INIST\_Loterre



## Construire et partager des terminologies

Au cours de l'année, l'Inist a mis les compétences de ses ingénieurs au service des chercheurs, des laboratoires, des groupements ou institutions de recherche, pour répondre à leurs besoins de ressources terminologiques normalisées, validées et interopérables.

Ainsi les réalisations de 2017 portent sur :

- Enrichissement de la traduction française du MeSH (Medical Subject Headings) réalisé avec l'Inserm ;
- Conversion en ontologie du thésaurus de la mémoire conçu à l'Inist, mise à jour de la ressource et traduction anglaise des définitions des concepts de ce thésaurus. Participation à la création du GDR « Mémoire » validée en décembre 2017 ;
- Publication sur le site Pôle lorrain de compétences

en éducation thérapeutique du patient (ETP) du glossaire réalisé en collaboration avec nos partenaires du CHRU de Nancy, de l'École de santé publique de l'université de Lorraine, le Pôle lorrain de compétences en ETP et du réseau régional LORDIAMN) ;

- Alignement du micro thésaurus « Lieux » de PAC-TOLS produit par Frantiq avec la base de données géographiques GeoNames ;
- Partage d'une terminologie relative à l'écotoxicologie avec l'Université de Saclay ;
- Aide méthodologique à la structuration des mots-clés d'auteur issus d'un corpus de la revue *Rééducation orthophonique* (1997-2014), du projet Orthocorpus (Atilf-CNRS).



## GdR SémanDiv / Thésaurus de la biodiversité

L'Inist est membre du GDR SémanDiv (Sémantique de la biodiversité), créé en janvier 2017, pour une durée initiale de quatre ans (2017-2020).

Membre du comité de pilotage, l'Inist :

- co-anime l'axe 3 (recensement des standards et référentiels terminologiques disponibles) ;
- contribue à l'axe 2 (développement de thésaurus et d'une ontologie de la biodiversité) par la réalisation d'un thésaurus des concepts les plus gé-

néraux de la biodiversité et son alignement avec des ressources plus spécifiques.

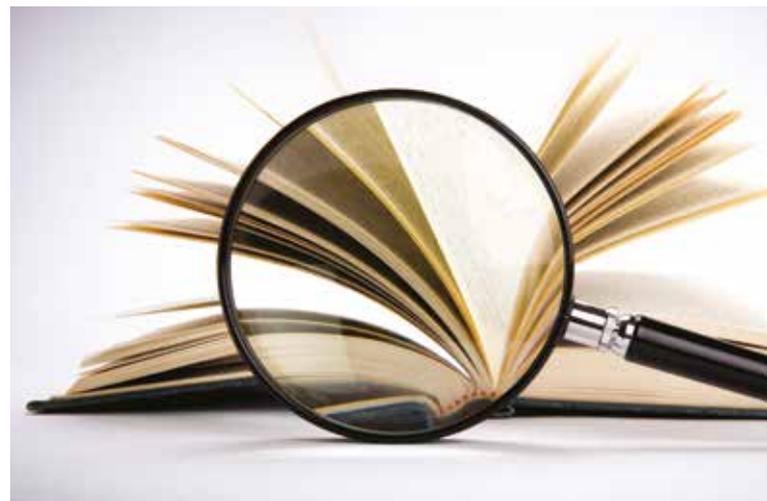
Le thésaurus de la biodiversité, qui est réalisé à l'Inist, sert de socle à la réalisation des travaux terminologiques/ontologiques du GDR ; il est mis à jour et structuré en continu (définitions de 82 concepts en 2017). L'Inist a également rédigé un document technique définissant les différents concepts cœur en matière de thésaurus et d'ontologies.

## Communications

Le Pôle lorrain de compétences en ETP publie le glossaire « L'ETP au fil des mots » - Poster au Séminaire annuel d'ETP, Metz, le 10 octobre 2017.

« Enrichissement et exposition de terminologies : une expérience collaborative Inist/CNRS » - 8<sup>e</sup> journée nationale Rameau, BNF - Paris, le 15 novembre 2017

« Données et évaluation de la biodiversité : quels enjeux, quels défis ? » Séminaire ECOSCOPE 2017, Paris, 12 et 13 décembre 2017.



## Présence internationale

Depuis novembre 2017, l'Inist participe au groupe d'intérêt « Controlled Vocabularies for Repository Assets » de COAR. Ce groupe d'intérêt a pour objectif de concevoir et promouvoir des vocabulaires contrôlés multilingues destinés à l'indexation des entrepôts de données.

## Valoriser les bases de données des laboratoires

La valorisation des bases de données se situe au cœur des recommandations exprimées par la Direction de l'information scientifique et technique (Dist) du CNRS en matière d'ouverture des données. À travers ce type de service, l'Institut apporte son soutien aux laboratoires, groupements ou institutions de recherche qui souhaitent diffuser ou mettre en commun sur les réseaux leurs fonds ou bases documentaires multimédias et les rendre interrogeables de manière structurée, interopérables et moissonnables.

Le service rendu repose sur un traitement documentaire des données incluant numérisation ou océrisation des données, reformatage, normalisation, alignement, enrichissement des métadonnées, et mise en ligne dans des applications web dédiées.



### CoReA : Corpus & ressources archéologiques

Valorisation de fonds d'archives multimédias et carnets de terrain en archéologie

La plateforme « CoReA » offre l'exposition de stèles gallo-romaines de la Narbonnaise à travers un vaste ensemble de fiches archéologiques consultables par anthroponymes, chronologie, créateur, lieu de conservation, peuple, sujet, support et ville de découverte.

Réalisée en partenariat avec le Centre Camille Jullian (UMR CNRS/Aix Marseille Université) et le groupement de service Fédération et ressources sur l'antiquité (Frantiq), cette bibliothèque numérique est construite autour du logiciel libre « OMEKA » et de ses plug-ins. Elle propose un ensemble de fonctionnalités telles que la géolocalisation, l'import de données, le moissonnage OAI-PMH, l'interrogation en langage naturel ou encore la recherche par facette.

<http://ccj-corea.cnrs.fr/>

### Réalisations

Mise en ligne de 636 carnets de terrains d'archéologues, ce qui implique la numérisation des documents, le reformatage des métadonnées, leur enrichissement avec les thésaurus PACTOLS et les données de géolocalisation.

### Communication

« CoReA : plateforme collaborative OMEKA pour la valorisation de corpus et de ressources archéologiques », Centre Camille Jullian - MMSH - Aix-en-Provence - 12 décembre 2017.



## Valorisation de la base INIS « International Nuclear Information System »

Opérée par l'Agence internationale de l'énergie atomique (AIEA), la base INIS héberge l'une des plus importantes collections d'informations publiées au monde concernant les usages pacifiques des sciences et technologies nucléaires. Cette base couvre également les aspects économiques et environnementaux relatifs aux énergies non nucléaires (fossiles, renouvelables, hydrogène), à leur développement (ressources, exploitation, production) ainsi qu'à leur utilisation (distribution, stockage, conversion, etc.).

La plateforme Vibad (Valorisation et indexation de base de données), développée par l'Inist, permet d'interroger la base INIS grâce à un moteur de recherche performant et des fonctionnalités variées

(navigation, rebond, recherche par facettes, export de données dans différents formats, stockage des requêtes, etc.).

La base INIS compte plus de quatre millions d'instances et s'enrichit chaque mois de plusieurs dizaines de milliers de documents intégrés par des processus manuels (contrôle qualité des données) et automatiques.

<https://inis-cea.inist.fr/>

### Réalisations

Mises à jour mensuelles de la base et automatisation des chargements associés.



## Conditor

### Référencement de la production scientifique française

Projet national multipartenaires, Conditor a pour objectif de recenser l'ensemble de la production scientifique (articles, ouvrages, congrès, thèses, rapports...) de la communauté de l'enseignement supérieur et de la recherche. À terme, Conditor fournira aux différents organismes de l'ESR des métadonnées validées concernant leur production scientifique. Après une phase projet de deux ans, le service opérationnel Conditor sera progressivement mis en place courant 2019. Le réservoir sera alimenté à partir de sources multiples et géré par une équipe organisée en réseau, en s'appuyant sur les outils fournis par la plateforme Conditor.

### Construire un référentiel mutualisé, dédoublonné et enrichi

- Conditor s'alimentera à partir des archives et autres réservoirs de métadonnées sur la produc-

tion scientifique et servira également de source pour ces réservoirs ;

- La valorisation de la production scientifique référencée dans Conditor sera facilitée, au travers d'applications qui pourront exploiter ses données ;
- Les partenaires du projet pourront bénéficier de l'apport de chacun en matière de métadonnées ;
- L'administration du référentiel sera assurée par des professionnels de l'IST.

### Réalisations

En 2017, la plateforme Conditor a été créée sur la base du socle ISTEEX. Les deux premiers corpus ont été chargés avec les données du WoS 2014 et celles de HAL 2014, soit environ 200 000 références. En fin d'année, le module de dédoublonnage automatique à base de règles a été mis au point permettant un repérage d'environ 48 000 doublons sur les 200 000 notices de la base.



## Optimiser le partage et l'interopérabilité des données de recherche

Seul le ralliement aux bonnes pratiques de gestion des données de la recherche, conjuguant les technologies de l'information et de la communication, permet de garantir véritablement leur valorisation. En effet, enclencher ce processus vertueux correspond à faciliter la découverte, l'accessibilité et la réutilisabilité des données pour leur permettre de jouer pleinement leur rôle moteur dans l'ouverture vers des nouvelles avancées scientifiques et dans la stimulation de la capacité d'innovation. Au travers de ses réalisations, l'unité accompagne cette mutation inéluctable en matière de gestion des données de la recherche.



### Portail OPIDoR (Optimisation du partage et de l'interopérabilité des données de la recherche)

Ouvert à l'automne 2017, le portail OPIDoR met à disposition des communautés scientifiques un ensemble d'outils et de services pour faciliter la gestion et la valorisation des données de la recherche. Ces services permettent de répondre aux principes FAIR qui visent à rendre les données Faciles à trouver, Accessibles, Interopérables et Réutilisables.

Actuellement trois services sont proposés :

- attribution de l'identifiant pérenne DOI en France.

L'Inist opère ce service pour le compte du CNRS, membre du Consortium DataCite ;

- Cat OPIDoR cartographie les services français qui accompagnent la communauté scientifique dans les processus de production, gestion et valorisation des données de recherche ;
- DMP OPIDoR est un outil d'aide à la création en ligne d'un plan de gestion de données, élément clé d'une bonne gestion des produits issus de la recherche.



### DOI Datacite

L'Inist offre à ses utilisateurs un accompagnement personnalisé à l'attribution de DOI aux données de la recherche (fourniture des préfixes DOI, assistance à la création et à la conversion de métadonnées) et à l'utilisation des différents services proposés par DataCite.

#### Quelques chiffres clés :

- **79** centres de données servis, dont **24** nouveaux en 2017 ;
- **27 000** DOI attribués en 2017.



### Cat OPIDoR

Financé par le MESRI, et co-construit avec l'Université de Strasbourg, ce catalogue permet de repérer, dès la planification d'un projet, les services utiles à chaque phase du cycle de vie des données produites :

- information ou accompagnement ;
- plateforme d'acquisition de données, de calcul, d'archivage ;
- entrepôts et annuaires de données ;
- outils de gestion.

Ouvert en septembre 2017, l'alimentation est conti-

nue et collaborative, à la manière de Wikipédia. Elle se fait à l'aide d'un formulaire qu'un fournisseur ou utilisateur d'un service non encore signalé peut renseigner. L'administration et la maintenance de l'appliquatif (wiki sémantique) sont réalisées par l'Inist, qui assure également une légère modération pour éviter des inconsistances ou des indécidables.

#### Quelques chiffres clés :

- **100** services répertoriés ;
- **30** collaborateurs actifs.



## DMP OPIDoR

Cet outil, fondé sur une réalisation du Digital Curation Center, permet au chercheur et à ses partenaires de rédiger un plan de gestion de données, en proposant différents modèles préconisés par les institutions et financeurs, dont la Commission européenne, ainsi que des guides et des exemples personnalisés.

L'Inist assure un accompagnement à l'élaboration de plans de gestion de données selon un des modèles mis à disposition et à la construction de modèles personnalisés.

L'année 2017 a vu, d'une part, la stabilisation du service et, d'autre part, la poursuite d'une collaboration internationale en vue de la migration de DMP OPIDoR vers un code commun, DMP Roadmap, aligné sur la convergence de deux outils historiques développés respectivement par le Digital Curation Center et l'University of California Curation Center.

### Quelques chiffres clés :

- **800** comptes ouverts ;
- **10** modèles disponibles.



### Formations

Gestion et partage des données de recherche - École doctorale RP2E de l'Université de Lorraine, Nancy, 23 janvier 2017.

DMP OPIDoR-ANF DERCI - Paris - 19 juin 2017.

Présentation de Cat OPIDoR et DMP OPIDoR et Atelier DMP OPIDoR-ANF Renatis - Inist-CNRS - Vandœuvre - 3 juillet 2017.

### Communications

« La gestion et le partage des données de recherche »  
Laboratoire d'imagerie médicale (LIB) – Institut Gustave Roussy Villejuif, 1<sup>er</sup> décembre 2017.

« Offre de service de l'Inist : le portail OPIDoR » -  
Journée Open Science Inist-CNRS - Vandœuvre -  
16 novembre 2017.



### Présence internationale

- Research Data Alliance (RDA) : participation au groupe de travail DMP Common Standards et aux groupes d'intérêt Sharing Rewards and Credit (SHARC) et Mapping of the Landscape (MoL) ;
- Assemblée générale DataCite, 3-4 avril 2017, Barcelone : participation et représentation du CNRS, membre du Consortium DataCite.
- Inserm Workshop 246 - Management and reuse of health data: methodological issues, 15-17 mai 2017, Bordeaux (coorganisation).

## Se former à distance sur les données de la recherche

À l'image de la politique de la Commission européenne inscrite dans le programme-cadre de recherche H2020, les politiques institutionnelles mises en place dans le contexte du développement de la Science ouverte incitent fortement les communautés scientifiques à gérer, partager et diffuser les résultats de leur recherche.

### *L'ouverture des données de la recherche (principes FAIR) au cœur du dispositif d'autoformation DoRANum* <https://doranum.fr>

Pour accompagner la communauté scientifique, l'Inist est engagé avec le réseau des Urfist dans le projet DoRANum (Données de recherche : apprentissage numérique à la gestion et au partage) afin de mettre en place un dispositif de formation à distance ouverte qui intègre différentes ressources d'autoformation sur la thématique de la gestion et du partage des données de la recherche.

Plusieurs modalités et ressources d'apprentissage complémentaires sont ainsi proposées aux utilisateurs afin de répondre aux attentes et usages des publics visés : chercheurs, enseignants chercheurs, doctorants et professionnels de l'information des établissements d'enseignement supérieur et de la recherche. Elles sont entièrement accessibles en ligne en libre accès et réutilisables.



### Principales réalisations 2017 des équipes pédagogiques de l'Urfist de Nice et de l'Inist

#### Ressources pédagogiques

Plus de 60 nouveaux modules de formation (fiches synthétiques, vidéos pédagogiques, animations et tutoriels multimédias, infographies, quiz, cours e-learning, vidéos d'interventions d'experts) ont été conçus, réalisés et mis en ligne sur neuf thématiques prioritaires de formation.

#### Neuf thématiques de formation



Enjeux et bénéfices de la gestion et du partage des données de recherche  
Pourquoi partager les données ? Qu'est-ce que l'Open Science ?



Les aspects juridiques et éthiques  
Que puis-je partager ? Que puis-je réutiliser ?



Le plan de gestion de données  
Pourquoi et comment rédiger un plan de gestion de données ?



Les métadonnées et formats  
Comment décrire mes données ?



Les identifiants pérennes  
Comment citer un jeu de données ?



Le stockage et l'archivage des données  
Quelles données conserver à long terme et comment ?



Le dépôt et les entrepôts  
Où et comment déposer mes données ?



Data journals, data papers  
Pourquoi et comment publier mes données comme un article scientifique ?



Accès, visualisation  
Où et comment extraire les données qui m'intéressent ?

### Interviews filmées

- Panorama des activités et métiers d'ingénierie pédagogique dans DoRANum ;
- Retours d'expérience de la MSH Ange-Guépin ;
- Retours d'expérience et évolution des métiers à l'Inist ;
- Interviews d'experts de l'Inra, Couperin, LIRMM, Université de Saclay.

### Site DoRANum - CMS

- Conception et mise en ligne du site V2. DoRANum devient un site d'accès aux ressources pédagogiques ;
- Mise en production du moteur de recherche à facettes ;
- Audit du site.



### Formations

Atelier « DoRANum : une plateforme de sensibilisation et de formation pour les chercheurs et les bibliothécaires » - Formation « Gérer, valoriser et préserver les données de la recherche », Ensisib, Villeurbanne, 27 juin 2017.

Atelier de formation DoRANum, Journée Open Science, Inist-CNRS, Vandœuvre, 16 novembre 2017.

### Communications

« DoRANum : service de formation à distance à la gestion de données de recherche », séminaire « Données scientifiques : outils et services pour la gestion et l'ouverture des données scientifiques », Inra, Paris, 16 février 2017.

« Présentation de DoRANum », GTA0 de Couperin, Paris, 30 mai 2017.

« Présentation et atelier DoRANum », ANF Renatis Medici « Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données », Vandœuvre, 3-6 juillet 2017.

« Présentation DoRANum », ateliers du numérique LabexMed, MMSH, Aix-en-Provence, 29 septembre 2017.

« Présentation de DoRANum », atelier des Journées RNBM « Accès ouvert : rêve ou réalité » CIRM, Marseille, 19 octobre 2017.

Webinaire sur DoRANum pour l'Urfist de Nice, Inist-CNRS, 17 octobre 2017 et 19 décembre 2017.



### DoRANum

- Projet placé sous l'égide de la Bibliothèque scientifique numérique ;
- Le comité de pilotage réunit les instances directement impliquées (Dist, Inist, Réseau des Urfist, BSN 9, BSN 10) ;
- Près de 226 000 euros sur deux ans. Financement MESRI (BSN9) ;
- Un partenariat CNRS, Réseau des Urfist, BSN 9 et 10 du MESRI ;
- Partenaires associés : Inra, Couperin, MSH Ange-Guépin, GIS « Réseau des Urfist ».

# Consulter les ressources documentaires électroniques du CNRS

### **BibCnrs : une plateforme pour répondre aux besoins des chercheurs**

Ce portail d'accès aux ressources documentaires électroniques met à disposition des chercheurs un ensemble négocié de revues, d'ouvrages et de bases de données, ainsi que des ressources libres qualifiées.

Le portail BibCnrs, *via* ses 10 domaines scientifiques (INSB, INEE, INSHS, INC, INSU, INS2I, INSIS, INP, IN2P3, INSMI), offre l'accès à un contenu scientifique de qualité :

- 19000 revues ;
- 18000 ouvrages ;
- 40 bases de données dont les bases de données citationnelles Web of Science et Scopus ;
- 60 plateformes éditeurs dont ScienceDirect, Wiley Online Library, Cairn.info, Nature.com, JStor, SpringerLink, ACM Digital Library...

L'Inist a élaboré une stratégie d'évolution de la couverture des portails du CNRS permettant de mettre les budgets alloués par le CNRS en adéquation avec l'évolution des besoins documentaires des communautés. La sélection des nouvelles ressources a été effectuée sur la base des usages obtenus lors de quatre campagnes de découverte après validation par les correspondants IST des instituts du CNRS.



#### **Les nouvelles ressources 2017**

##### **Revues**

- 9 titres IOP (Institute of Physics) ;
- Science ;
- 10 titres RSC (Royal Society of Chemistry) : extension à la collection Gold 2016.

##### **Ouvrages**

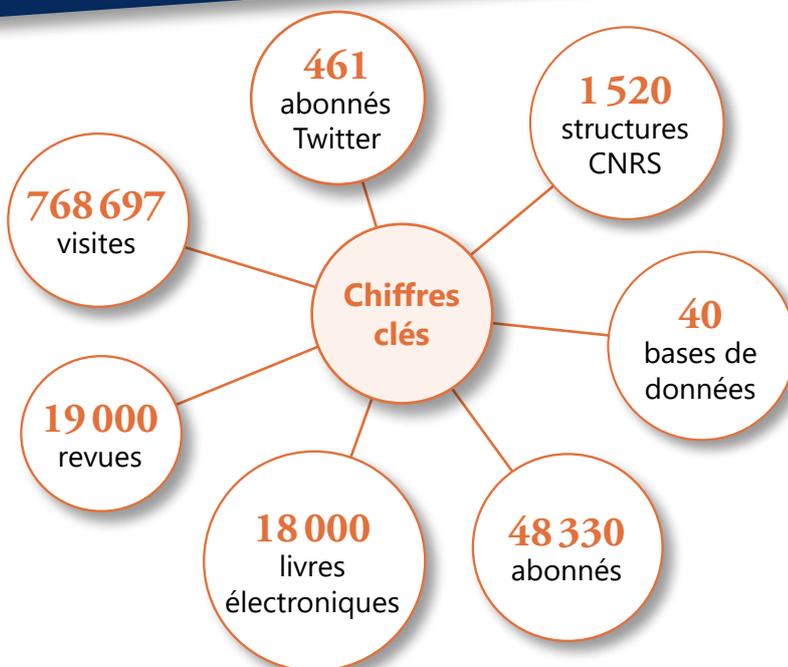
- 2 collections ebooks (Springer) : Biomedical et Physics and Astronomy : 1 084 titres ;
- 10 ebooks IOP (offerts).



#### **Négociations nationales**

La préparation du renouvellement du contrat Springer Nature s'est appuyée sur la mesure de la part croissante des articles en accès ouvert dans les revues Springer. Trois protocoles de soutien à l'édition scientifique française ont été signés sous l'égide du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, avec CAIRN, OpenEdition et EDP Sciences. Les objectifs sont de maintenir et faire évoluer une édition scientifique française de qualité, en s'appuyant notamment sur les acquis de la loi pour une République numérique.

- **Accéder aux articles directement via Google Scholar.**  
Il suffit de paramétrer votre moteur de recherche. Vous repérez ainsi dans la liste des résultats de Google Scholar de façon instantanée les articles BibCnrs disponibles en texte intégral dans votre domaine.
- **Accéder aux articles directement via PubMed.**  
Il suffit de paramétrer votre compte personnel sur PubMed. Vous repérez dans la liste des résultats de façon instantanée les articles disponibles en texte intégral grâce au logo BibCnrs INSB ou BibCnrs INEE.
- **Être reconnu ayant droit directement sur un site éditeur.**  
Les signets des navigateurs permettent de s'identifier comme ayant droit lorsque vous naviguez sur un site éditeur et d'obtenir ainsi le texte intégral d'un article sans passer par une recherche via le portail BibCnrs.



### Vidéos pédagogiques

Cinq nouveaux modules multimédias ont été créés en anglais et cinq modules ont été mis à jour :

- S'authentifier dans BibCnrs ;
- La recherche dans BibCnrs ;
- Comment paramétrer Google Scholar ;
- Comment être reconnu comme ayant droit via un signet ;
- Paramétrer PubMed pour obtenir des articles accessibles via BibCnrs ;
- Adjust the settings on Google Scholar ;
- Adjust the settings on PubMed ;
- How to be recognized as a rights holder ;
- Running searches in BibCnrs ;
- Signing in on BibCnrs ;

Deux webinaires ont été animés sur BibCnrs.

# Accéder aux collections rétrospectives de la littérature scientifique

Le projet **ISTEX** est un vaste programme d'acquisition de ressources scientifiques visant à créer une bibliothèque numérique aux meilleurs standards internationaux, accessible à distance par tous les membres des établissements de l'ESR *via* des services multiples pouvant facilement s'intégrer dans les outils et/ou portail déjà existants.

Ce réservoir de plus de 20 millions d'articles scientifiques ou de monographies est un complément aux abonnements courants et permet de s'affranchir des liens avec les éditeurs commerciaux.



### ISTEX : les ressources viennent à vous

La plateforme ISTEX offre fin 2017 un accès multiple et simple à des ressources multilingues et pluridisciplinaires couvrant une période de 1401 à 2015. ISTEX est actuellement disponible dans plusieurs grandes universités françaises et organismes de recherche soit à travers les outils de découverte, les espaces numériques de travail ou les portails d'accès à la documentation. Une extension web (<http://addons.istex.fr>) permet à toutes et tous d'accéder à ISTEX depuis n'importe quel site qui signale des ressources *via* un DOI (Digital Object Identifier).

### Ils utilisent ISTEX

La base de données LISSA (Littérature scientifique santé) propose environ 1,2 million d'articles scientifiques en français dans le domaine de la santé. Une coopération entre les équipes techniques du LISSA et d'ISTEX a permis en quelques jours d'y ajouter un bouton intégré ISTEX à l'application web <https://www.lissa.fr> permettant l'accès direct au plein texte d'ISTEX depuis le site tiers.

L'Institut des systèmes complexes de Paris et d'Île-de-France (ISCIPIF) développe GarganText, une application web qui combine des outils de text mining avancés et vous permet de produire une cartographie des connaissances en quelques minutes. Depuis 2017, les ressources ISTEX sont directement et automatiquement accessibles comme nouvelle source pour un traitement dans GarganText.



### Le saviez-vous ?

L'extension navigateur est disponible pour Firefox et Chrome et s'installe en trois clics. Elle permet d'accéder aux ressources ISTEEX depuis tous vos outils de recherche documentaire habituels (Google Scholar, PubMed, WoS, Scopus...) Essayez-la ! <https://addons.istex.fr/>

### Web sémantique

En 2017, ISTEEX commence à s'afficher sur le web sémantique à travers le développement d'une plateforme de publication sémantique : LODEX, et un portail linked open data ([data.istex.fr](http://data.istex.fr)).

### Communications

ISTEX était présent sur le stand du MESRI au Salon du livre de Paris - Paris, 24-27 mars 2017.

« ISTEEX : des enrichissements au web de données » - *I2D – Information, données & documents* - avril 2017 (Volume 54).

« ISTEEX, un projet national d'archives documentaires : au-delà de l'accès au texte intégral, l'enrichissement des données par méthodes de fouille de textes » - 85<sup>e</sup> congrès annuel de l'Association francophone pour le savoir (Acfas) 8-12 mai 2017.

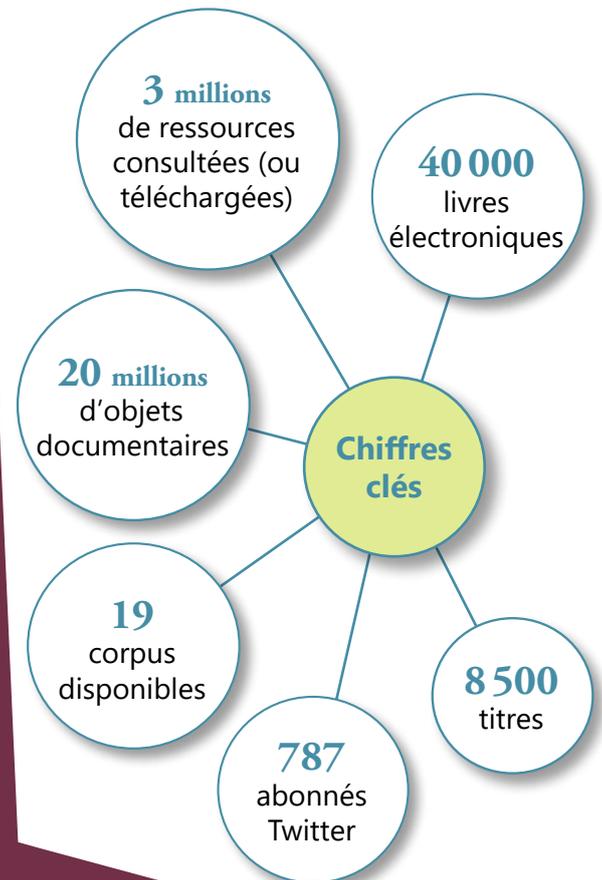
« ISTEEX : signalement et accès aux ressources » - Journées de l'ABES - Montpellier, 10-11 mai 2017.

Séminaire sur des chantiers d'usages et les services à valeur ajoutée développés autour d'ISTEX - Nancy, 6-7 juin 2017.

« Apprendre à utiliser l'API ISTEEX, depuis l'interrogation jusqu'à l'extraction de documents » - Journées nationales du DEveloppement logiciel - Marseille, 4-7 juillet 2017.

« Découvrir ISTEEX, plateforme d'accès aux collections rétrospectives de la littérature scientifique » - Biennale du numérique 2017 - Lyon, 13-14 novembre 2017.

« L'Étincelle Triplex : un triple store pour ISTEEX » - SemWebPro-Paris, 22 novembre 2017.



### Formations

Atelier « Interroger et exploiter l'API ISTEEX » - Carrefour de l'IST - Palais des Congrès Nancy - 20 au 21 mars 2017.



### Vidéos pédagogiques

« Interrogation de l'API ISTEEX » : trois nouveaux modules multimédias ont été créés, trois autres ont été mis à jour en 2017.



### Présence internationale

Présentation d'ISTEX aux directeurs de l'Office européen des brevets - Delft (NL), 25 janvier 2017.



@ISTEX\_Platform

# Mesurer les usages des ressources électroniques

Dans le cadre d'une politique documentaire nationale, de plus en plus contrainte budgétairement, la plateforme **ezMESURE** (financée en partie par le MESRI) permet à chaque organisme de recherche ou d'enseignement de déposer ses statistiques générées avec le logiciel libre **ezPAARSE**.



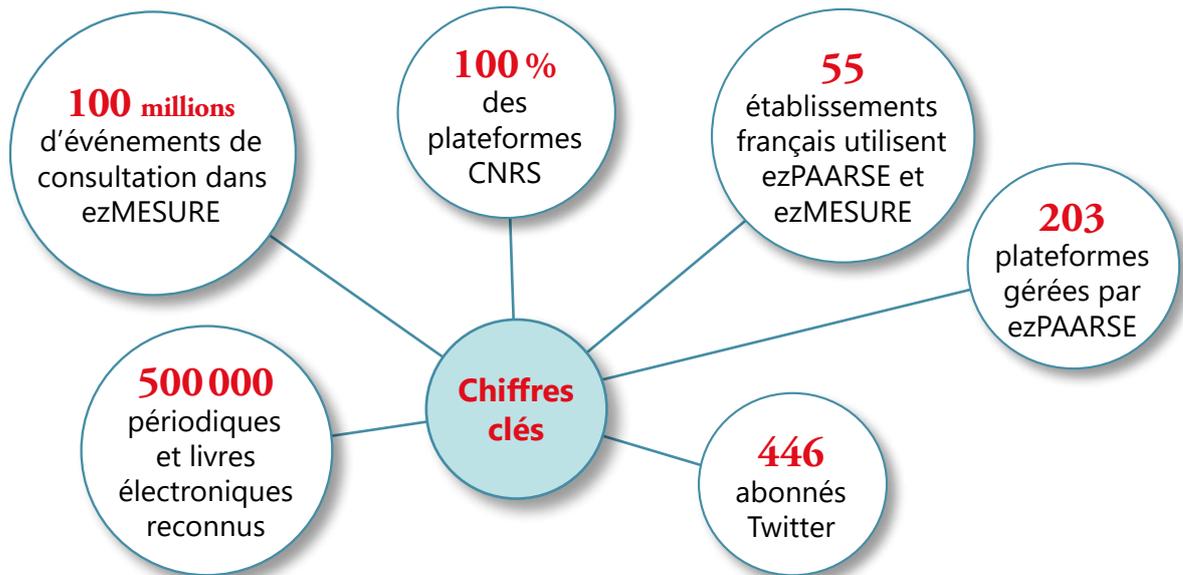
### ezMESURE / ezPAARSE

ezMESURE propose une interface en ligne de visualisation dynamique, consolidée et comparative des données de consultation qui est utilisée par plus de 50 établissements de l'enseignement supérieur et de la recherche. Elle agrège fin 2017 plus de 100 millions d'événements de consultations (accès à un PDF, une page HTML, une revue de sommaire...).

ezMESURE offre un entrepôt national pour collecter des données, homogènes et indépendantes, sur les accès aux ressources numériques payantes par les établissements de l'ESR. Cet entrepôt offre une visualisation, de type cartographie dynamique, des indicateurs nationaux, régionaux et par établissement des ressources documentaires consultées, par grandes catégories d'utilisateurs, ainsi que les archives de ces données pour des analyses rétrospectives. En 2017, l'alimentation d'ezMESURE a considérablement augmenté grâce à la nouvelle fonctionnalité qui permet aux usagers de réaliser eux-mêmes les imports.

Cette plateforme vient en complément du logiciel libre ezPAARSE qui permet l'analyse, l'enrichissement et l'exploitation des logs d'accès aux ressources électroniques.

Des outils simples, efficaces et indissociables, une équipe disponible et réactive, une communauté d'usagers grandissante.



### Vidéos pédagogiques

- Installation d'ezLOGGER dans Chrome ;
- Utilisation d'ezLOGGER ;
- Paramétrage d'ezLOGGER.



### Formations

Atelier « Interroger et exploiter l'API ISTEK » - Carrefour de l'IST - Palais des Congrès Nancy - 20 au 21 mars 2017.

ezMESURE - Création de tableaux de bord - Sciences Po Paris - 25 avril 2017.

Atelier ezPAARSE / ezMESURE - Lyon - 30 juin 2017.

### Communications

« Présentation d'ezMESURE » - Journées annuelles DIRBUIST - Paris, 5 avril 2017.

« ezPAARSE, outil d'analyse, d'enrichissement et d'exploitation des logs » - Biennale du numérique 2017 - Lyon, 13-14 novembre 2017.



### Présence internationale

« ezMESURE » - International Coalition of Library Consortia (ICOLC) - Prague (CZ), 15-18 octobre 2017.



@ezparse

## Publier les résultats de la recherche scientifique

### Édition numérique

Enjeu de taille pour l'édition scientifique, l'intégration de nouveaux modèles de publication électronique offre aujourd'hui un réel potentiel en termes de visibilité et de valorisation des contenus éditoriaux.

Le savoir-faire de l'Institut en matière de publication numérique permet de proposer la consultation de documents scientifiques, la création de sites pour la diffusion de publications numériques, la numérisation d'archives, ainsi que des conseils pour optimiser la production éditoriale.

#### La plateforme I-Revues

La plateforme I-Revues permet la consultation de documents scientifiques, la création de sites et la numérisation d'archives pour la diffusion de publications scientifiques. En 2017, une nouvelle publication a été diffusée sur I-Revues : le 23<sup>e</sup> Congrès français de mécanique. Et l'antériorité du *Bulletin de l'Académie vétérinaire de France* (1948 à 2002) a été numérisée et mise en ligne.

Un plan de curation des métadonnées des revues dans Dspace a été réalisé afin que les articles soient mieux référencés, notamment dans Google Scholar. Suite à la loi sur la République numérique, nos conseils ont permis à des revues d'abaisser, voire d'annuler leur barrière mobile.

#### La mise en ligne de revues scientifiques sur I-Revues

2 363 documents ont été traités et chargés sur la plateforme I-Revues et plus de 3 874 000 documents ont été vus à ce jour.

#### L'Inist participe au soutien de l'édition électronique aux côtés de ses partenaires :

**Cléo** : chargement de 744 documents sur Revues.org et OpenEdition Books ;

**Inserm** : conseil en ingénierie documentaire et mise en ligne des articles des revues *Médecine Sciences* et *Sciences et Santé* sur la plateforme de production éditoriale iPubli (606 documents mis en ligne) ;

**Muséum national d'Histoire naturelle** : fabrication des deux numéros du volume 52 de la revue *Anthropozoologica*, activité de secrétariat d'édition. L'antériorité de cette revue a également été mise en ligne sur le site du MnHn et maintenant les versions papier et numérique sont publiées simultanément ;

**Académie européenne interdisciplinaire des sciences (AEIS)** : maquettage de l'ouvrage *Ondes, matière et Univers* pour une version numérique et une version papier ; l'ouvrage est mis en ligne et imprimé par EDP Sciences.

## Modération de l'archive ouverte HAL

La collaboration avec le Centre pour la communication scientifique directe (CCSD) a été initiée en 2015.

En 2017, l'Inist a réalisé 50 242 actes de modération, dont 11 708 demandes de modifications et conseils et 37 808 mises en ligne de documents.

L'équipe a également réalisé le dépôt en masse de plus de 500 documents dans HAL pour le compte de l'European Synchrotron Research Facility (TGIR de l'INSU).



### Vidéos pédagogiques

- Comment distinguer un fichier auteur d'un fichier éditeur ? ;
- Comment importer dans HAL ? ;
- Conception d'un parcours pédagogique « Hypothèses » (mise en ligne 2018).

## Numérisation et maquetage

L'Inist a diversifié ses activités de numérisation. En effet, des projets ayant des objets et objectifs différents ont été réalisés :

- mise en ligne de textes océrisés : monographies pour Prodig ;
- mise en ligne de textes océrisés et réalisation d'une notice en XML par article : revue *Bulletin de l'Académie vétérinaire de France* mise en ligne dans I-Revues ;
- mise en ligne de données de la recherche : photos pour le projet CoReA, photogravures pour l'Université de Lorraine.

Le tout représente un équivalent de 24 197 pages.

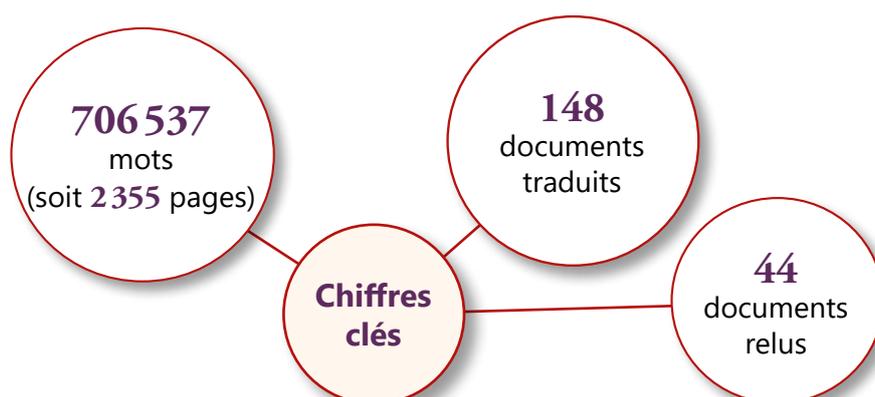


## Traduction

L'activité de l'équipe de traduction est caractérisée par la prise en charge de documents d'une importante complexité, plus particulièrement en raison de leur nature scientifique et savante. Une large gamme de thématiques en sciences humaines et sociales, sciences et technologies, sciences de la vie ont été couvertes dans les articles, rapports et divers documents confiés en 2017 à l'équipe de traduction par les chercheurs, enseignants-chercheurs, d'autres acteurs de la recherche dans son ensemble. Il en va de même dans les domaines bibliothéconomiques et juridiques pour le compte des autres équipes de l'Inist, les acteurs IST du CNRS, voire notre ministère de tutelle.

Au cours de l'année, l'équipe a accentué son relationnel avec les utilisateurs pour affiner la pertinence et la précision de ses prestations. Cela s'est traduit par de nombreux dialogues en cours de travaux pour en soigner la qualité, les adapter au mieux aux contenus et aux usages rédactionnels attendus. L'équipe compte poursuivre ce mode dialogué de réalisation de ses prestations en raison de la variété croissante des thématiques de contenus.

Le service rendu engendre un regain d'intérêt renouvelé au travers de la demande qui augmente. Il a fallu à l'occasion rechercher des solutions à proposer en dehors d'une intervention directe face à des plannings chargés en raison d'une demande croissante.



### Formations

Communication scientifique en anglais : « Mieux rédiger en anglais un article scientifique » & « Optimiser sa communication orale en anglais » - Nantes - 10 au 12 mai 2017.

# Former à l'information scientifique

L'avènement d'une science ouverte s'accompagne de profondes transformations dans la production et la diffusion de la science.

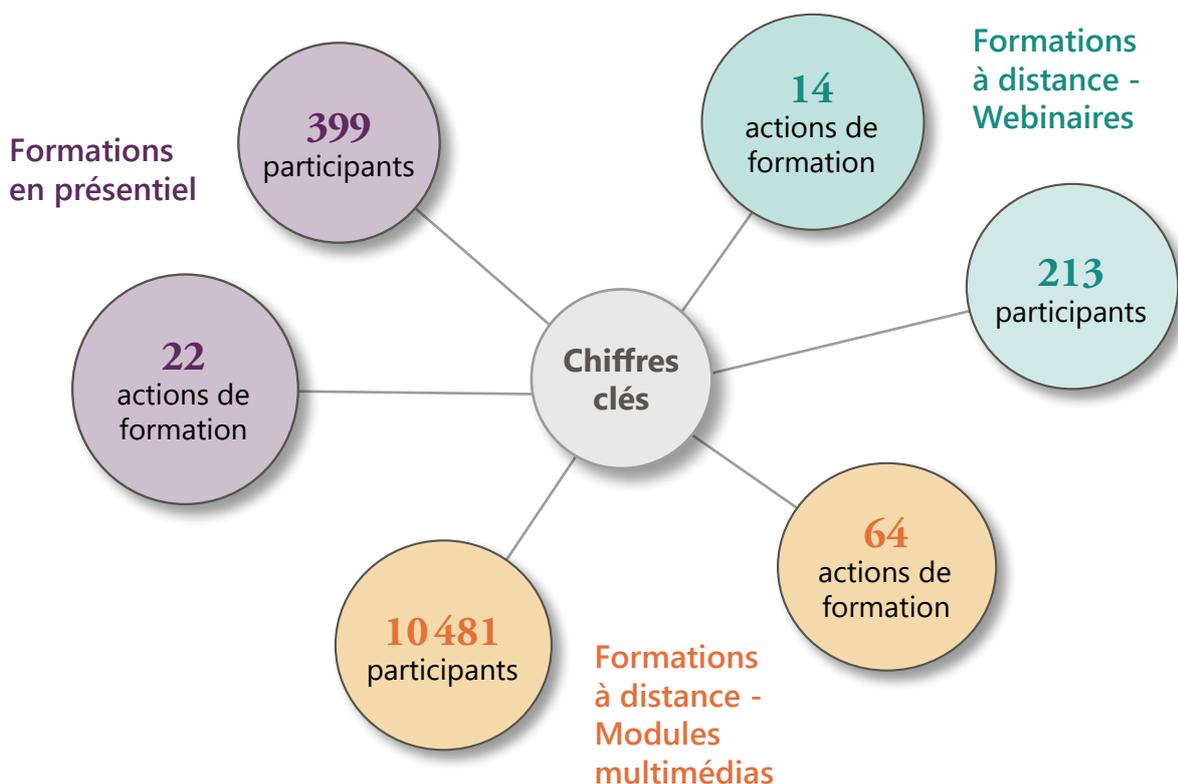
Afin de répondre à de tels enjeux, l'offre de formation de l'Institut poursuit son développement : de la formation classique à l'autoformation, l'Inist propose des solutions multiples et adaptées aux besoins de la communauté scientifique.

Cette offre s'appuie sur une équipe d'ingénieurs pédagogiques et un réseau d'experts internes et externes.



## Réalisations

Les principales interventions du dispositif de formation en présentiel et/ou à distance (webinaires et vidéos - animations pédagogiques multimédias) ont porté sur le libre accès aux publications et aux données de recherche, sur les services et plateformes IST ainsi que sur les réseaux sociaux scientifiques. Des formations sur les nouvelles pratiques d'apprentissage ont été dispensées.





---

# L'Institut en chiffres



Ressources  
humaines

**198**  
agents



Budget d'acquisition  
de ressources  
électroniques

**5 847 319 €**



Twitter  
Inist

**2 700**  
abonnés

Formation IST



**11 093**

participants

**100**

actions de  
formation



Data center

**40** serveurs  
physiques

**400** serveurs  
virtuels

**800 To**  
de données  
(articles scientifiques)

**219**  
bases de  
données



**1 520**  
structures CNRS

**48 330**  
abonnés

**768 697** visites

**19 000** revues

**18 000** livres



**60**

ressources  
pédagogiques  
multimédias en  
accès libre



**20 millions**  
d'objets  
documentaires

**8 500** titres

**19**  
corpus  
disponibles

**40 000** livres électroniques

**3 millions**  
de ressources consultées  
(ou téléchargées)



**55**

établissements français versent  
leurs données statistiques



**500 000** périodiques et livres  
électroniques reconnus

**203**

plateformes gérées par  
ezPAARSE



**100**  
services  
référéncés



**800**  
comptes  
ouverts

**10** modèles  
disponibles



**27 000**  
DOI attribués

**79** centres de  
données  
servis





**Inist - Institut de l'information scientifique et technique**

2, allée du Parc de Brabois - CS 10310  
54519 Vandœuvre-lès-Nancy

Tél. : +33 (0)3 83 50 46 00

Fax : +33 (0)3 83 50 46 50

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

