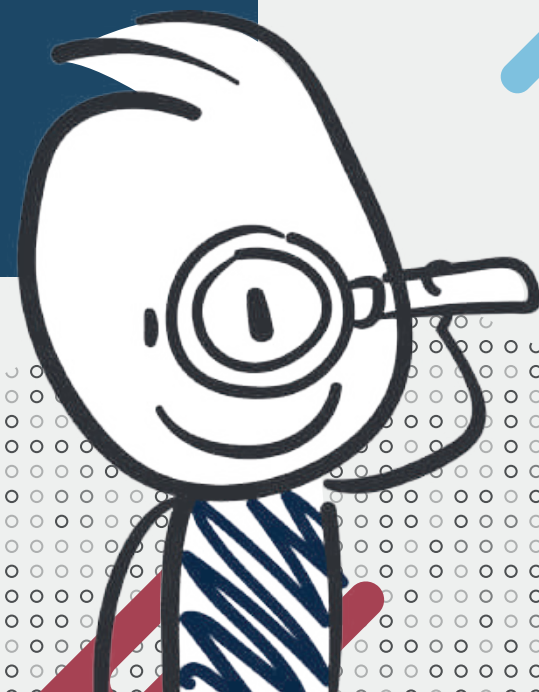




Inist

Rapport d'activité 2019





SOMMAIRE



Édito 4

Les faits marquants... 6

Trois axes forts 8

Accéder à L'information scientifique

BibCnrs, Istex, Panist 10

Négociations, mesure des usages 12

Publications, édition numérique 14

Portail HAL-CNRS 15

Le fonds documentaire 16

Valoriser les données de La recherche

À votre rencontre... OPIDoR Tour et formations 18

OPIDoR 20

DoRANum 22

Exposition de données 24

Analyser et fouiller L'information scientifique

ISTEX pour le TDM 26

ISTEX Web de données 28

Appui au pilotage 30

Traduction 31

Terminologie et Loterre 32

Visa TM 34

Conditor 35

L'Inist, c'est aussi... 36

Systèmes d'information 37

Secrétariat général 37

Communication, Web 38

Veille 39

Pour en savoir plus... 40

ÉDITO



Les services de l'Inist : la science ouverte en pratique

Le CNRS a finalisé en 2019 sa feuille de route pour la science ouverte, dans la lignée du plan national défini en 2018. L'Inist déploie son offre de service dans le cadre de cette feuille de route en liens étroits avec la Direction de l'information scientifique et technique (DIST).

Depuis plusieurs années, l'Inist s'implique dans l'accompagnement des chercheurs pour gérer et **valoriser leurs données de recherche**. Sa plateforme OPIDoR (outil d'aide à la rédaction de plan de gestion de données - DMP et catalogue des services français) est maintenant un outil connu par les chercheurs et porté par les instances du CNRS. L'Inist a travaillé en étroite collaboration avec l'Agence nationale de la recherche (ANR) pour outiller sa directive imposant la rédaction de DMP pour les projets acceptés. Afin d'accompagner cette mesure, les équipes sont allées à la rencontre des chercheurs dans les différentes régions. Organisé avec les délégations régionales du CNRS et les services de proximité des universités, cet OPIDoR Tour a regroupé plus de 1100 participants pour ses cinq premières étapes.

Cette action est le point fort d'une démarche que l'Inist développe, maintenant dans le cadre de son projet d'établissement. Il s'agit de multiplier **les rencontres avec les communautés scientifiques** afin, d'une part, de faire connaître nos services et, d'autre part, de recueillir les avis et doléances des chercheurs dans un objectif d'amélioration continue. Ainsi les formations en présentiel continuent à se développer aussi bien sur la thématique du cycle de vie des données et plan de gestion de données que sur celle de la fouille de textes (création de corpus spécialisés à partir du réservoir ISTEEX et formation complète sur le TDM). Dans le cadre de la mise en œuvre d'une démarche UX (User experience), grâce aux différents échanges avec les utilisateurs, les équipes ont pu repérer les besoins d'évolution du portail BibCnrs, la nouvelle version a été ouverte au premier semestre 2020. Nous remercions tous les collègues qui ont pris le temps de nous donner leurs avis et besoins.

En cohérence avec son objectif d'atteindre **100 % des publications scientifiques accessibles**, le CNRS a ouvert cette année le portail HAL-CNRS. L'Inist est co-administrateur de ce





portail et accompagne les laboratoires dans la réalisation de leurs collections. L'augmentation du taux de publication en accès libre dans la production du CNRS a été mesurée par nos équipes en collaboration avec le Département des outils d'aide à la décision du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI). Cette évolution nous amène à définir une nouvelle politique documentaire que nous mettons en œuvre à travers les négociations que nous menons avec les éditeurs scientifiques et les choix de présentation de la littérature scientifique dans nos portails documentaires.

Avec le transfert de 5810 titres du fonds documentaire papier, 68 bibliothèques françaises ont pu disposer gratuitement d'ouvrages pour compléter leur propre fonds documentaire. Le don de plusieurs centaines d'ouvrages à « Ammareal - le livre solidaire » a généré un bénéfice reversé à une association caritative. La page de la fourniture de documents papiers se referme définitivement avec la double intention de donner une seconde vie aux collections et d'inscrire cette action dans un cadre solidaire et social. Les travaux de manutention ont été

réalisés grâce à un partenariat avec une association locale de réinsertion.

L'année 2019 a débuté par la mise en place d'une nouvelle organisation en phase avec **les axes du projet scientifique qui structurent également ce rapport d'activité**. Si cette réorganisation a impacté chacun d'entre nous, notre activité de service a poursuivi son évolution pour répondre toujours mieux aux besoins exprimés ; ceci grâce à l'engagement des collègues en charge directe de ces services mais également de ceux qui nous concoctent une infrastructure informatique performante et moderne, et qui assurent avec professionnalisme la gestion administrative et financière de notre unité.

Ce rapport d'activité, que nous avons voulu synthétique et illustré, est le reflet de l'implication de tous et de la belle dynamique qui s'instaure avec nos partenaires et utilisateurs.

Bonne lecture.

Claire François, directrice



LES FAITS MARQUANTS

JANVIER

- BibCnrs : sollicitation des utilisateurs via une démarche UX

FÉVRIER

- L'Inist devient co-administrateur du portail HAL-CNRS

MARS

- Mise en ligne des 6 corpus spécialisés pour le TDM ISTEEX

JUILLET

- Projet maDMP 4LS lauréat de ANR Flash
- ACS disponible sur la plateforme ISTEEX

AOÛT

- 2200 titres Elsevier 2018 disponibles dans Panist

SEPTEMBRE

- Don des 10 000 ouvrages du fonds documentaire
- Intégration du modèle DMP ANR dans DMP OPIDoR
- Claire François devient membre du COPIL Science ouverte de l'UL



AVRIL

- BibCnrs : accès simplifié aux ressources (nouvelle extension navigateur)

MAI

- Dépôt du jeu GopenDore (créé par le service Formation) dans Zenodo

JUIN

- Ouverture du service de création en masse de DOI
- Premières formations TDM Ur-fist

OCTOBRE

- Première date de l'OPIDoR Tour à Nancy
- Intervention de Claire François à la journée Science ouverte du CNRS

NOVEMBRE

- Début du transfert des périodiques
- L'Inist intègre le Comité opérationnel données de recherche de l'UL
- Journée de clôture de VisaTM

DÉCEMBRE

- Transfert des données vers HAL-LARA
- Thésaurus MESH bilingue disponible dans Loterre

Trois axes forts

L'Institut de l'information scientifique et technique déploie ses activités autour de 3 axes forts ; un projet qui s'inscrit dans le contexte de politique institutionnelle et nationale de science ouverte.

Faciliter l'accès à la production scientifique internationale

Entre portail d'accès aux ressources courantes et réservoirs d'archives : une combinaison au service des chercheurs.

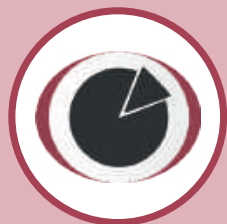
Avec BibCnrs, ISTEEX et Panist, trouvez rapidement et facilement une publication scientifique, qu'elle soit récente ou ancienne, en accès libre ou payant.

Avec ezPAARSE et ezMESURE, mesurez les usages de vos ressources afin de mieux piloter la politique documentaire.



ACCÉDER À
L'INFORMATION
SCIENTIFIQUE





ANALYSER & FOUILLER L'INFORMATION SCIENTIFIQUE

Répondre aux besoins de pilotage scientifique et d'extraction de connaissances

*Web sémantique, terminologie, fouille de textes,
traduction et analyse de contenu numérique :
des services tournés vers la science ouverte.*

Permettre d'extraire, de structurer et de classer
des connaissances à partir de très grands volumes
d'information, la fouille de textes est aujourd'hui un enjeu
majeur pour l'exploration de la littérature scientifique.
L'Inist vous accompagne dans cette démarche à travers des
outils, du conseil et de la formation.



VALORISER LES DONNÉES DE LA RECHERCHE

Faciliter la gestion, le partage, la diffusion ou l'archivage des données

*Collecte, diffusion, valorisation des données,
des outils au service de la recherche.*

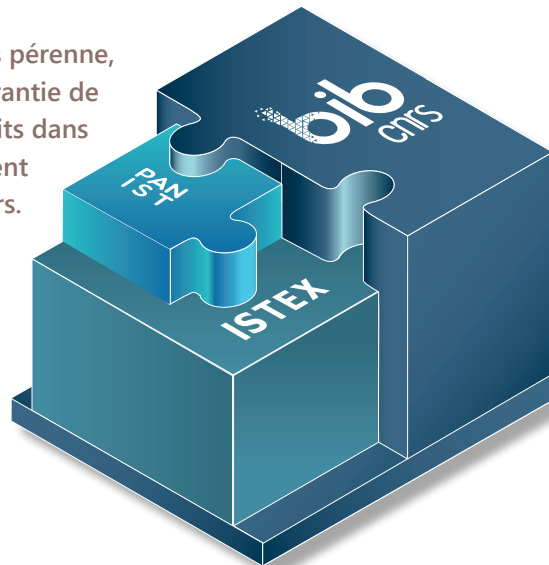
L'Inist propose un accompagnement et des outils pour la
gestion, le partage et la diffusion des données de la recherche
avec l'objectif de faciliter leur découverte, leur accès, leur
interopérabilité et leur réutilisation.

BibCnrs ISTEX - Panist

BibCnrs, portail d'accès aux ressources documentaires électroniques des unités du CNRS, permet la consultation de plusieurs millions d'articles, livres et bases de données scientifiques acquis pour les communautés de recherche du CNRS, ainsi que dans le cadre du projet ISTEX.

ISTEX, réservoir d'archives de plus de 23 millions d'articles et monographies scientifiques accessibles à l'ensemble de l'Enseignement supérieur et de la recherche (ESR), est un complément aux abonnements courants de BibCnrs. La plateforme ISTEX garantit un accès pérenne aux collections rétrospectives quelle que soit l'issue des négociations futures.

PANIST, plateforme d'accès pérenne, offre aux ayants droit la garantie de l'accès aux contenus souscrits dans les contrats indépendamment des plateformes des éditeurs.



Le Fait marquant

L'année 2019 marque l'aboutissement d'une **démarche UX** (User experience) appliquée au portail BibCnrs. Centrée sur les utilisateurs et basée sur des interviews, enquêtes et tests en local, cette démarche a permis de **repérer des besoins d'évolution du portail** tant au niveau ergonomique qu'au niveau fonctionnel. Cela a abouti au lancement d'un marché en vue d'une importante prestation de **refonte du site**, notamment pour simplifier l'accès aux documents dont la réalisation a été programmée en janvier 2020.

BibCnrs : en parallèle de la démarche UX, **une réflexion concernant le positionnement de BibCnrs sur la nécessité de science ouverte** a été menée par un groupe de travail interne qui a consulté de nombreux acteurs CNRS mais aussi d'autres organismes. L'orientation du portail BibCnrs, qui proposait quelques ressources libres parmi l'ensemble des ressources des éditeurs, évoluera en même temps que la refonte du site dès janvier 2020 pour favoriser l'accès aux publications en libre accès.

Panist : le **chargement des données 2018 de l'éditeur Elsevier** complète l'offre d'archivage de Panist qui met à disposition des établissements ayants droit (membres du groupement de commandes 2011-2013 ou de la licence nationale 2014-2018) un total de 7,3 millions d'articles de 2002 à 2018. Ces données prennent ainsi la suite des contenus accessibles sur la plateforme ISTEX.

ISTEX :

- ❑ La négociation liée au **plan de soutien de l'édition scientifique** avec **l'éditeur Cairn** a débuté en octobre.
- ❑ Les échanges sur les formats des premiers contenus du **GIS CollEx-Persée** ont débuté en décembre.

BibCnrs | 1520 unités couvrant 13 391 chercheurs CNRS + 37 294 chercheurs non CNRS = 50 685 ayants droit (+ 4 %)
 5693 abonnés aux listes de diffusion (+ 2 %) 946 abonnés Twitter (+ 29 %) 6 363 901 consultations enregistrées via ezPAARSE/ezMESURE

ISTEX : 24 509 054 consultations enregistrées via ezPAARSE/ezMESURE

Panist : 30 794 consultations enregistrées via ezPAARSE/ezMESURE

Partenariats et collaborations



❑ **Inserm** : le portail d'accès aux ressources documentaires électroniques InsermBiblio a remplacé BiblioInserm (septembre). L'Inist a accompagné le déroulement du projet en apportant son expertise technique et fonctionnelle acquise lors de la refonte du portail BibCnrs.

❑ **GIS CollEx-Persée** : infrastructure de recherche en information scientifique et technique, ce groupement d'intérêt scientifique a pour objectif de faciliter l'accès aux ressources documentaire. L'Inist est opérateur technique avec l'ABES pour le chargement des données du GIS CollEx-Persée dans la plateforme ISTEX. Il met l'expertise acquise par le signalement et l'accès aux ressources ISTEX au service du GIS CollEx-Persée.

❑ **Consortium Couperin.org** : commanditaire de la plateforme Panist, réalisée et maintenue par l'Inist, le consortium est notamment garant des droits d'accès aux contenus souscrits ainsi que de l'évolution de la plateforme.

Actions de communication



❑ **BibCnrs** :

- Présentations à la Journée des nouveaux entrants du CNRS, Dolorès Dardaine, Cécilia Fabry, mars, Paris.
- Forum sur la démarche UX autour de BibCnrs et présentation du poster « Démarche UX & BibCnrs », Laurence Grand, Cécilia Fabry, Journées de l'ABES, mai, Montpellier.

❑ **ISTEX** : poursuite de l'ISTEX Tour dans deux villes supplémentaires (Nice et Nantes) pour présenter la plateforme et ses fonctionnalités.



❑ **Panist** : communiqué aux établissements « Ouverture de la Plateforme d'Archivage National de l'Information Scientifique et Technique », janvier.

Ils parlent de nous...

❑ Publication de l'article « BibCnrs, un réservoir évolutif de publications pour les communautés de recherche du CNRS » dans **La lettre de l'INSHS**, novembre 2019.

❑ « Ouverture de PANIST, la Plateforme d'Archivage National de l'Information Scientifique et Technique » publié sur le **blog Couperin.org**.

Le tweet

  **2156 impressions**
25 oct. 2019

[Ressources en test] Pour tous, jusqu'au 30 novembre 2019 : Découverte de la base de données : ProQuest D&T Global

Négociations Mesure des usages

Chargé de répondre aux besoins exprimés par les communautés scientifiques, le service Négociations & Acquisitions alimente les plateformes d'accès à la production scientifique internationale telles que Panist, ISTEK ainsi que le portail BibCnrs.

Les orientations de la politique documentaire s'appuient sur la mesure des usages afin d'identifier les titres les plus consultés ainsi que les communautés utilisatrices.



Le fait marquant

En 2019, le service Négociations & Acquisitions a mené près de **cinquante négociations** pour assurer le renouvellement des accès aux contenus mis à disposition des chercheurs via BibCnrs.

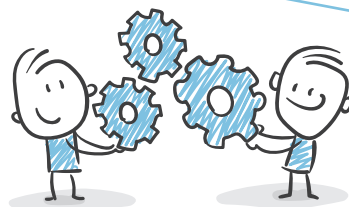
Il s'est également inscrit pleinement dans le paysage de la **science ouverte**.



Dans le contexte national du **Plan pour la science ouverte**, plusieurs actions ont été menées :



- ❑ Les **dépenses en APC** (frais de publication en open access) du CNRS en 2015 et 2017 ont été évaluées. Les données ont été publiées sur le site openAPC, sous l'égide de Couperin.org, avec l'objectif de partager les informations des dépenses des établissements français à l'échelle mondiale.
- ❑ Les données d'usage des ressources électroniques ont été enrichies avec l'**utilisation de l'API UNPAYWALL** afin de déterminer la part des consultations potentiellement en accès libre. Cette expérimentation mise en place pour le consortium du Luxembourg fonctionne également pour d'autres établissements français participant à ezMESURE.
- ❑ La première vague d'acquisition de ressources électroniques dans le cadre du programme mené par le GIS CollEx-Persée a débuté. En partenariat avec l'ABES, l'Inist met en œuvre le **dépôt des données et du texte intégral dans ISTEK** pour une mise à disposition de l'ensemble des établissements de l'ESR.

Partenariats et collaborations



- ❑ **OCLC et Couperin** : suite à l'accord de partenariat signé entre l'Inist, Couperin et l'OCLC (Online Computer Library Center) en février 2019, ezPAARSE est utilisé par l'OCLC pour produire des données d'usages des ressources électroniques de leur solution proposée en mode hosting avec EZproxy. Un nouveau service « Analytics » est proposé aux clients EZproxy d'OCLC reprenant les mêmes socles logiciels qu'ezMESURE.
- ❑ **OCLC** : depuis fin 2019, des collaborateurs d'OCLC participent à l'analyse de plateforme dont les tutoriaux ont été traduits pour l'occasion.
- ❑ **Open Edition** : suite à l'appel d'offre passé par Open Edition pour intégrer ezPAARSE à leurs traitements statistiques de données d'usage, les développements ont commencé. L'équipe ezPAARSE assure un support sous forme de conseil pour réaliser un middleware OpenEdition interne qui permettra l'enrichissement des données d'ezPAARSE pour la production de statistiques au format Counter5.

Le tweet

21 juin 2019

ezPAARSE / ezMESURE

Depuis 2016, les Universités françaises ont déposé **80 millions** de données d'usages des ressources électroniques dans ezMESURE ! Un grand merci à tous les collègues qui nous font confiance !

Actions de communication



- ❑ Au cours d'un séminaire en ligne ezMETRICS en novembre, le principe et les résultats de la **suite logicielle ezPAARSE-ezMESURE** ont été présentés à une vingtaine de membres de l'ICOLC (International Coalition of Library Consortia), une association internationale des consortia de bibliothèques.
- ❑ **Les usages de l'Open Access se mesurent aussi** et **Une mesure des usages « sur mesure » qui dépasse les frontières**, articles publiés sur le site web Inist (lire p. 43 et p. 45).

62 établissements français de l'ESR utilisent ezPAARSE **(+ 5 %)**
648 403 081 consultations réparties sur
 • **441** plateformes **(+ 21 %)** et
 • **3 659 737** titres de revues et d'ebooks différents
28 collaborateurs **(+ 20 %)**
583 abonnés Twitter **(+ 18 %)**
188 tweets **(+ 79 %)**
299 parseurs **(+ 15 %)**

Publications

Édition numérique

L'Inist diversifie et dynamise ses activités d'édition numérique afin de servir les communautés de la recherche en respectant les préconisations de la science ouverte :

- Édition numérique de revues scientifiques pour des partenaires souhaitant les valoriser ;
- Secrétariat de rédaction pour répondre à des besoins de fabrication de revues ou d'ouvrages.

MNHN (*Anthropozoologica*) :

17 articles édités en français ou anglais = **234** pages

ABES : traitement rétrospectif de **28** numéros (n° 64 à 91) de la revue *Arabesques*

OpenEdition : **57** numéros de revues mis en ligne sur OpenEdition Journals (15 titres)

22 numéros de revues mis en ligne sur Cairn ainsi que

37 ouvrages de la Collection Histoire ancienne et médiévale et

37 ouvrages de la Collection Histoire et civilisation

IRevues : **983** documents chargés, **9 179 454** visualisations

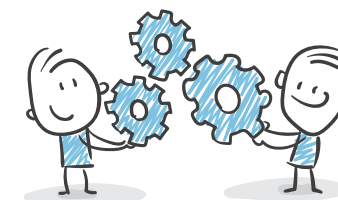


Le Fait marquant

Collaboration avec l'ABES pour la **mise en ligne de la revue Arabesques** : le traitement rétrospectif a été réalisé avec les outils de la chaîne Métopes et le chargement de la revue a été fait sur la plateforme Prairial de l'Université de Lyon 3 qui utilise le logiciel Lodel.

La fermeture de la plateforme éditoriale I-Revue fin 2020 a été annoncée aux éditeurs scientifiques des revues avec l'aide du Pôle Contrats de la Délégation régionale. L'Inist assure une **mission de conseil** aux éditeurs scientifiques dans la recherche d'une plateforme d'accueil et d'accompagnement pour les aspects techniques du transfert des revues.

Partenariats et collaborations



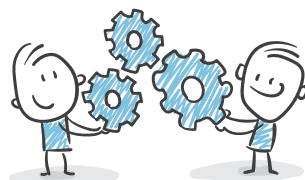
▣ **Muséum national d'Histoire naturelle** : secrétariat d'édition pour la revue *Anthropozoologica* qui est désormais publiée en flux continu. La collaboration sur de nouvelles revues s'est mise en place.

▣ **OpenEdition Center** : des échanges constructifs ont permis de partager nos pratiques respectives en édition numérique ainsi que la prise en main de nouvelles activités telles que la vérification des fichiers avant envoi à Cairn de même que la vérification et le chargement d'ouvrages issus de la numérisation sur OpenEdition Books. (lire p. 41)

Portail HAL-CNRS

Afin de soutenir l'archive ouverte HAL, l'Inist collabore avec le CCSD pour effectuer de la modération des dépôts et en co-administrant avec la DIST le portail HAL-CNRS dédié à la valorisation des productions scientifiques du CNRS.

Partenariats et collaborations



□ **CCSD** : dans le cadre d'un partenariat avec le CCSD, 5 agents assurent la modération des dépôts dans HAL en apportant un soutien (conseil, incitation, aide).

Actions de communication



□ **Inist – HAL : trois missions en appui à la science ouverte**, article publié sur le site web Inist (lire p. 42).

Le fait marquant

L'Inist est co-administrateur du **portail HAL-CNRS** qui a été ouvert en septembre. L'institut apporte son aide à la création de collections pour les unités du CNRS.

Le site Lara (Libre accès aux rapports scientifiques) a été fermé le 1^{er} décembre 2019 ; les 1943 rapports ont été transférés sur **HAL-Lara**, portail HAL des rapports scientifiques et techniques dont l'Inist est administrateur.

Modération dans HAL

61 457 dépôts modérés dont
6 402 affiliations corrigées

46 670 documents mis en ligne

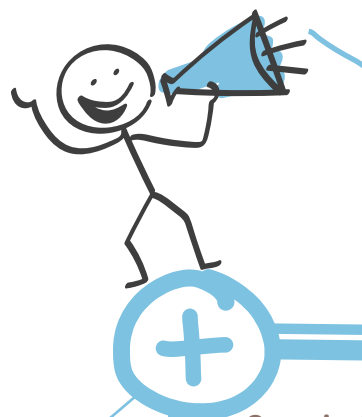
13 626 demandes de modifications ou de conseils apportés aux déposants

Le Fonds documentaire

La fermeture du service Fourniture de documents est effective depuis décembre 2018.

Ce service historique était alimenté par un fonds documentaire de 27 222 titres et de 225 000 monographies, le tout représentant 24 km de collections.

L'année 2019 a vu le transfert d'une partie des collections dans d'autres bibliothèques.



Le Fait marquant

L'Institut a organisé des **transferts d'une partie des périodiques** (5 810 titres) du fonds documentaire vers **68 bibliothèques**.

Organisation du transfert des collections

L'Inist s'est rapproché du CTLes (Centre technique du livre de l'enseignement supérieur), outil de conservation et de gestion coopérative des collections.

Ce dernier a déterminé, avec les plans de conservation partagée thématiques et nationaux (PCP) gérés par le GIS CollEx-Persée (182 bibliothèques), la liste des périodiques que les PCP souhaitent accueillir, les monographies n'étant pas concernées par le PCP. L'arbitrage des demandes en doublon entre établissements a été réalisé par l'ABES.

Ainsi, le transfert d'un quart des collections de périodiques a été réalisé.

Le pilotage en magasin des opérations de transfert a été assuré par l'Inist mais la prise en charge financière et logistique reposait sur les établissements eux-mêmes.

Pour les titres n'ayant pas trouvé preneur, la Délégation Régionale du CNRS gère leur désherbage.



Que sont devenues les monographies ?

Le fonds documentaire de l'Inist était constitué de **225 000 monographies** (ouvrages, thèses, rapports, actes de congrès, etc.), dont 10 000 ouvrages.

Environ 5 000 ouvrages ont été transférés dans des **bibliothèques lorraines** (la bibliothèque Elie Cartan, le jardin botanique Jean-Marie-Pelt, des musées de Nancy...).

Les 5 000 autres ouvrages ont été donnés à **Ammareal**, libraire d'occasions sur Internet. Ammareal reverse un pourcentage de la vente de chaque livre à des organisations caritatives luttant contre l'illettrisme et en faveur de l'éducation. L'Inist a choisi de reverser à **Bibliothèques Sans Frontières** (Paris).

L'Inist a gardé **12 000 thèses** dites AO (archives originales) qui sont principalement des thèses CNRS non détenues par ailleurs et qui peuvent remonter à 1939 pour les plus anciennes.

Le tweet



BU Lyon1
@BULyon1

17 déc. 2019

#Don Reçu ce matin de l'**@INIST_CNRS** : 264 titres de revues de tous les continents (Inde, Brésil, Australie, Sibérie...), essentiellement en **#chimie**.
En chiffres, cela représente 119 mètres linéaires, stockés dans 247 cartons sur 13 palettes! Un peu de travail en perspective...

Partenariats et collaborations



- ❑ **CTLES et ABES** : pour le choix et l'accueil des périodiques.
- ❑ **Ammareal, musées lorrains, bibliothèque Elie Cartan, Jardin botanique Jean-Marie-Pelt** : accueil d'ouvrages.
- ❑ **ULIS** (Unités localisées pour l'inclusion scolaire) : projet de service de Jeunes & Cité (2019-2023) porté par la structure ULIS en faveur de la réinsertion (mise en cartons d'ouvrages et de périodiques).

Périodiques :

68 bibliothèques receveuses

5810 titres transférés

Monographies :

10 000 ouvrages transférés

Thèses :

12 000 thèses AO conservées

À votre rencontre...

OPIDoR Tour Formations

L'Inist a lancé l'OPIDoR Tour avec les délégations régionales du CNRS, et en lien avec les services de proximité des universités, afin de présenter les outils et services OPIDoR pour accompagner la gestion des données. En complément, plusieurs formations en présentiel ont été animées plus particulièrement sur le plan de gestion de données.



Les temps forts :

- ❑ **14 formations en présentiel** : sur la gestion des données, le cycle de vie des données, le plan de gestion des données, les principes FAIR dans le cadre des formations Urfist et de l'Enssib, des JNSO 2019, de l'école thématique e-envir « Les données ouvertes en sciences environnementales », ANR, Journée Archivage numérique des données de recherche Gricad, école inter-organismes sur la qualité en recherche et en enseignement supérieur...
- ❑ Développement du **serious-game GopenDoRe** : outil d'animation de formations mis à disposition de tous, en libre accès, sur la gestion et le partage des données.

Le fait marquant



Les équipes de l'Inist ont organisé avec les délégations régionales du CNRS des déplacements en France métropolitaine afin de présenter les outils et services OPIDoR pour accompagner la gestion des données.

L'OPIDoR Tour s'inscrit dans le cadre de la politique science ouverte du CNRS, notamment sur le volet données de la recherche, et en soutien aux recommandations de l'ANR. En effet, l'Agence demande l'élaboration d'un plan de gestion des données pour tous les projets financés dans les six mois qui suivent le démarrage du projet et ce à partir de l'édition 2019.

Partenariats et collaborations



- ❑ **Délégations régionales CNRS, ANR, BBEES, réseau des Urfist, Enssib, Gricad, Université de Lorraine, Réseau QRRAL, réseau QuaRES, AAU Creneau, SID Université Grenoble Alpes, SCD de Sorbonne Université, de l'Université Poitiers, de l'Université Orléans** : OPIDoR Tour.
- ❑ **Direction de la documentation et de l'édition de l'Université de Lorraine** : co-animation des ateliers sur le plan de gestion des données.
- ❑ **CNRS, GIS Réseau des URFIST, INRAE, EHESS, Universités** : création d'un réseau de formateurs (GT-DMP) dont l'objectif est de mutualiser les ressources pédagogiques sur le Plan de gestion des données, de partager les interventions et les retours d'expérience.

1 174 participants

Initiation aux données de recherche
Urfist - 27 et 28 mars
Yvette Lafosse

DMP et cycle de vie ANR - 9 oct.
Yvette Lafosse
Paolo Lai

Atelier DMP - École thématique e-envir BBES - 28 oct.
Coralie Wysoczynki
Yvette Lafosse

Ateliers DMP
• DR02 - 8 nov.
Yvette Lafosse,
Pierrette Paillassard
• JNSO-DIST - 19 nov.
Yvette Lafosse

Atelier DMP AAU-CRENAU
20 mai
Yvette Lafosse

Nantes

Opidor Tour DR02 - 21 nov.
Pierrette Paillassard
Laurent Rassinoux
Yvette Lafosse

Atelier DMP UL - 24 oct.
Claire Sowinski
Yvette Lafosse

Metz

Opidor Tour Inist - 10 oct.
Paolo Lai

Benjamin Faure
Coralie Wysoczynki
Claire Sowinski
Yvette Lafosse
Pierrette Paillassard

Nancy

Ateliers DMP
• UL - 14 oct.
Pierrette Paillassard
Laurent Rassinoux
Yvette Lafosse
• Institut Jean-Lamour 16 déc.
Claire Sowinski
Coralie Wysoczynki

Orléans

Opidor Tour DR08 - 28 nov.
Benjamin Faure
Coralie Wysoczynki
Yvette Lafosse

Poitiers

Opidor Tour DR08 - 26 nov.
Philippe Boulben
Coralie Wysoczynki
Yvette Lafosse

Ouvrir la science et les données ENSSIB - 27-28 juin
Paolo Lai

Villeurbanne

Lyon

Atelier DMP ENSSIB 27-28 nov.
Jean-Michel Parret

Grenoble

DMP, cycle de vie des données Gricad - 20 nov.
Paolo Lai
Claire Sowinski

Opidor Tour DR11 - 14 nov.
Anne Ciolek
Claire Sowinski
Yvette Lafosse

Montpellier

Atelier DMP QuaRES 11 sept.
Yvette Lafosse

● 5 étapes OPIDoR Tour (2019)

● 14 formations en présentiel (2019)
(+ 129 %)



OPIDoR

2019 a été l'année dédiée à la préparation des projets, au développement d'outils et des savoir-faire en ingénierie pour accompagner les communautés de recherche à l'implémentation progressive des principes FAIR : données Faciles à trouver, toujours Accessibles, Interopérables, Réutilisables.

Ils parlent de nous...

- ❑ **Site de l'ANR**, dans l'article « L'ANR met en place un plan de gestion des données pour les projets financés dès 2019 ».
- ❑ **Antoine Petit** aux JNSO 2019 parle d'OPIDoR (vidéo disponible en ligne : <https://webcast.in2p3.fr/video/feuille-de-route-du-cnrs-pour-la-science-ouverte>)
- ❑ **Journal du CNRS**, décembre 2019, p. 25, dans l'article « Les données scientifiques, un trésor à partager ».

Le fait marquant

Publication du modèle de Plan de gestion de données (PGD) de l'Agence nationale de la recherche dans DMP OPIDoR (disponible en français et anglais) : dans le cadre de sa politique science ouverte, l'ANR demande l'élaboration d'un plan de gestion des données pour les projets financés à partir de 2019. L'Agence a travaillé en concertation avec l'Inist pour que son modèle de PGD soit intégré dans le portail DMP OPIDoR pour une saisie en ligne.

Et aussi...

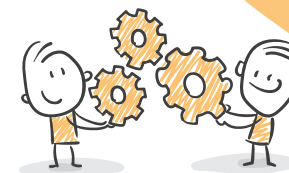
- ❑ **Appel Flash science ouverte** sur les pratiques de recherche et les données ouvertes, lancé par l'ANR : parmi les 25 projets sélectionnés, le projet porté par l'Institut français de bioinformatique, **maDMP4LS** (a machine-actionable DMP for Life Sciences), a été retenu.
- ❑ Accroissement des demandes de **personnalisation de DMP OPIDoR** par les organismes de recherche et établissements d'enseignement supérieur pour mettre en place leur politique de gestion des données (personnalisation de l'interface et adaptation des modèles et recommandations).
- ❑ Mise en ligne d'une **version revisitée de DMP OPIDoR** avec de nouvelles fonctionnalités telles que « Produits de recherche » qui permet de décrire séparément par exemple des jeux de données de nature différente ou d'autres produits de recherche.
- ❑ Mise en ligne de l'outil de gestion des DOI, **PID OPIDoR** qui permet aux utilisateurs de créer, modifier et vérifier leurs DOI, leurs URL et les métadonnées associées.
- ❑ Lancement et mise en œuvre du **projet DMP « Machine actionable »** qui vise à rendre les plans de gestion lisibles et exploitables par les machines pour une exploitation optimale de leur contenu.
- ❑ **Enrichissement de Cat OPIDoR** avec l'identification des services d'appui à la gestion des données pouvant apporter une aide de proximité aux communautés scientifiques pour mettre en œuvre un DMP.

Actions de communication



- ❑ **L'application DMP OPIDoR, un outil au service du chercheur : DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données**, demi-journée « Science ouverte : les données de la recherche », université de Paris 13, 16 avril.
- ❑ **DMP OPIDoR**, atelier DMP, Laurent Rassinoux, université de Lorraine, octobre.
- ❑ **DMP OPIDoR Planifier la gestion des données de recherche**, webinaire ANR « La politique Science Ouverte de l'ANR et son modèle de Plan de Gestion des Données », Laurent Rassinoux, 14 novembre.
- ❑ **Un DMP plus actif et intégré dans le workflow de la recherche**, Benjamin Faure et Florian Mazur, DataBFC2, 22 novembre, Dijon.
- ❑ **DMP OPIDoR - retour d'expériences et évolutions. Outil pour les plans de gestion**, association Aristote - groupe PIN, Marie-Christine Jacquemot, 26 novembre.
- ❑ **DMP OPIDoR. Planifier la gestion des données de recherche**, Anne Busin et Marie-Christine Jacquemot, Plan de gestion des données et traçabilité des activités de recherche – Réseau Qualité Recherche Aquitaine Limousin, 28 novembre.
- ❑ **DMP OPIDoR un outil d'aide à l'élaboration des plans de gestion de données**, Anne Busin et Florian Mazur, Biogenouest, Rennes, 17 décembre.
- ❑ **Mapping de métadonnées et création de DOI en nombre pour l'UNJF**, article publié sur le site web Inist (lire p. 41).

Partenariats et collaborations



- ❑ **Institut français de Bioinformatique (IFB) et Gesco** : l'Inist est partenaire du projet maDMP 4LS (projet ANR Flash).
- ❑ **CINES, cc-IN2P3, GENCI** : partenaires du projet maDMP de l'unité.
- ❑ **ANR, Inra-Irstea (Inrae), INSA, EPFL, HUMA-NUM, Sciences Po, UL, Universités de Paris, Unistra** : publication, adaptation de leurs modèles de plan de gestion de données et des recommandations associées dans DMP OPIDoR.
- ❑ **MITI, RBDD, Gricad** : l'Inist a participé à l'organisation du « Hackathon : comment améliorer le dépôt et le partage de données de recherche ? » (janvier 2020).
- ❑ **Université de Lorraine** : l'Inist a intégré le comité opérationnel des données de recherche.
- ❑ **DataCite** : membre du consortium, l'Inist est agence d'attribution de l'identifiant pérenne DOI en France.
- ❑ **RDA** : l'Institut est membre des working group « exposing data management Plans, Active Data Management Plans IG et FAIR Data Maturity Model: core criteria to assess the implementation level of the FAIR data principles ».
- ❑ **UC3-DCC** : l'Inist participe au développement de code commun DMP Roadmap / active DMP.

DMP OPIDoR : 20 modèles (+ 54%),

2 560 comptes ouverts (+ 54%),

24 établissements de recherche et

plus de 3 000 connexions par mois

Cat OPIDoR : 301 structures (+ 51%), 287 services référencés (+ 44%)

DataCite : 132 adhérents dont 28 nouveaux (+ 23%)

63 100 DOI attribués

DoRANUM

La plateforme DoRANum, réalisée en partenariat avec le Gis Réseau Urfist, met à disposition plus de 80 ressources d'autoformation en libre accès sur la gestion et la conservation, la diffusion et la réutilisation des données scientifiques liées aux activités de recherche. Les ressources pédagogiques accompagnent les chercheurs à chaque étape du cycle de vie des données.



Le Fait marquant

17 nouvelles ressources pédagogiques ont été créées en 2019 :

- ❑ 1 interview de Lionel Maurel (directeur adjoint scientifique de l'Institut sciences humaines et sociales du CNRS) « Questions juridiques liées aux données de recherche » (1 module avec 9 courtes vidéos) ;
- ❑ 1 interview de Véronique Ginouvès (responsable de la phonothèque à la Maison méditerranéenne des sciences de l'Homme d'Aix en Provence) « Les bonnes pratiques dans le cadre juridique et éthique pour la diffusion des données de la recherche » ;
- ❑ 1 module interactif « Les principes FAIR » ;
- ❑ 1 infographie de l'École des Ponts ParisTech « Données de la recherche – contexte juridique : qui a les droits, quelles obligations ? » ;
- ❑ 8 infographies « Les données de la recherche et les codes sources obligatoirement diffusables » ;
- ❑ 1 parcours pédagogique à partir du MOOC de l'université de Bordeaux « Intégrité scientifique dans les métiers de la recherche » ;
- ❑ 2 infographies sur les identifiants pérennes DOI et ORCID ;
- ❑ 1 guide « Le référentiel de gestion des archives de la recherche » ;
- ❑ 1 infographie « Les 3 niveaux de sauvegarde des données de la recherche ».



Démarrage de deux projets d'envergure nationale et internationale :

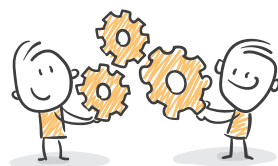
- ❑ **Projet e-learning MIPNES** (Mission de la pédagogie et du numérique pour l'enseignement supérieur) : projet de parcours pédagogique e-learning certifiant sur la gestion, le partage et l'ouverture des données à destination des doctorants, et intégré dans les cursus des écoles doctorales.
- ❑ **Projet OBERRED** (Open Badge Ecosystem for the Recognition of skills in Research Data management and sharing) : projet européen Erasmus+ qui a pour objectif de créer un écosystème d'Open Badges au niveau européen pour la reconnaissance des compétences sur le partage et la gestion des données de la recherche dans le cadre de l'Open Science – Coordination Université Côte d'Azur.

Site web : 26 100 visites (+174 %), 2 899 téléchargements (+23 %), 53 300 pages vues (+103 %),
122 billets postés (+94 %), 83 ressources pédagogiques (+14 %)

Youtube : 32 vidéos disponibles (+23 %), 91 abonnés (+52 %), 8 310 vues (+36 %)

Twitter : 240 tweets (+100 %), 765 abonnés (+62 %)

Partenariats et collaborations



- ❑ **GIS réseau Urfist, Université de Lorraine, Université Côte-d'Azur** : partenariat dans le cadre du projet e-learning MIPNES.
- ❑ **Université Côte-d'Azur** (France), **University of Applied Sciences Lübeck** (Germany), **Beuth University of Applied Sciences** (Germany), **University of Göttingen** (Germany), **World University Service of the Mediterranean** (Spain), **Instytut Badań Literackich Polskiej Akademii Nauk** (Pologne), **Institute of Literary Research of the Polish Academy of Sciences** (Pologne), **Università degli studi di Napoli « Federico II »** (Italie) : partenariat dans le cadre du projet OBERRED.
- ❑ **EHESS, INRAE, Réseau Urfist** : groupe de travail sur les formations aux DMP, échanges de supports et de bonnes pratiques autour des plans de gestion de données de la recherche (Comment y sensibiliser chercheurs, doctorants et personnels IST ? Comment se tenir au courant de l'évolution de la réglementation et des pratiques autour de ces documents ? etc.).

Actions de communication



Trois **vidéos de présentation de DoRANum** ont été réalisées en 2019 :

- ❑ Teaser DoRANum 01 : <https://youtu.be/fMrRj76qUPQ>
- ❑ Teaser DoRANum 02 : <https://youtu.be/5UmG-pIm16o>
- ❑ Teaser DoRANum 03 : <https://youtu.be/UZHrMgYpubs>

Le tweet



Exposition de données

L'Inist apporte son expertise aux laboratoires de recherche qui souhaitent exposer leurs données et qui s'inscrivent ainsi dans une démarche de science ouverte.

Exposer, partager et valoriser les données de la recherche sous forme de bibliothèques numériques, tel est le service proposé par l'équipe chargée de l'exposition des données qui a déjà mis en ligne plusieurs bases de données éditorialisées.



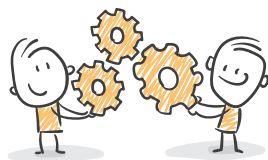
Le fait marquant

Durant l'année 2019, une **dizaine de projets** de valorisation de bases de données bibliographiques et de données de recherche ont été menés par le service exposition de données.

Parmi les projets mis en œuvre en 2019 :

- ❑ **De la Préhistoire à l'Actuel** - Culture, environnement et Anthropologie UMR5199-PACEA : plateforme dédiée à la valorisation des collections du laboratoire PACEA ;
- ❑ **Plateforme Cenhtor** (Centre de ressources numériques des humanités et des territoires) : plateforme dédiée à la conservation et à la diffusion des données de la recherche de la MSH Lorraine USR3261 (CNRS, Université de Lorraine) ;
- ❑ **Histoire et cultures de l'antiquité et du Moyen Âge** - EA 1132 -HISCANT-MA (Université de Lorraine) : plateforme dédiée à la valorisation d'un corpus de statues médiévales ;
- ❑ **Paléorient - Archéologies et Sciences de l'Antiquité** UMR7041-ArScAn : plateforme dédiée à la valorisation des sites archéologiques d'Orient et de bases de données documentaires de la MSH Mondes ;
- ❑ **Base de données en santé publique** (BDSP) en partenariat avec l'École des hautes études en santé publique de Rennes (EHESP), soit environ 530 000 références bibliographiques (articles, ouvrages, thèses, mémoires, travaux académiques, rapports, actes de congrès et retours d'expérience), dans de nombreux domaines de la santé publique ;
- ❑ **AAP ColIEx Persée** - Numérisation enrichie et développement de services à la recherche : l'Inist participe en qualité de partenaire aux candidatures de deux projets :
 - > **PERIODICOL**, coordonné par le SCD d'Aix-Marseille Université (AMU) et l'Institut des mondes africains (IMAF-CNRS) - Enrichissement de la bibliothèque numérique RevColEurop ;
 - > **Datacc'+** (Dispositif d'accompagnement aux données de la recherche), coordonné par l'Université Claude Bernard Lyon 1 (UCBL) et Institut de chimie et de biochimie moléculaires et supramoléculaires (ICBMS).

Partenariats et collaborations



❑ **PACEA** - UMR5199 (CNRS, Université de Bordeaux, ministère de la Culture, École pratique des hautes études, Institut national de recherches archéologiques préventives) : réalisation de la plateforme PACEA.

❑ **Maison des sciences de l'Homme Lorraine** USR3261-MSH Lorraine (CNRS, Université de Lorraine) : réalisation de la plateforme Cenhtor.

❑ **ArScAn** - UMR7041 (CNRS, Université Panthéon-Sorbonne, Université Paris-Nanterre, ministère de la Culture, Université Vincennes - Saint-Denis, Institut national de recherches archéologiques préventives) et **MSH Mondes** : réalisation de la plateforme Paléorient.

❑ **École des hautes études en santé publique de Rennes** (EHESP) : dans le cadre de la reprise et mise en ligne de la base de données en santé publique (BDSP).



L'atelier de numérisation de l'Inist a traité plusieurs collections de données de recherche, le tout représentant plus de 50 000 pages et documents iconographiques, qui sont publiées dans les bibliothèques numériques.

Actions de communication



❑ **CoRea : Corpus & ressources archéologiques**, Bernard Sampité, présentation à l'école thématique « Données Interopérables pour le Patrimoine » (DONIPAT), Aussois, 14 octobre.

18 instances (Omeka & Vibad) réalisées ou en cours de développement

7 000 jeux de données référencées

1,8 million de références bibliographiques (bases Vibad BDSP, Inis CEA, OpenBibArt)

50 000 pages et iconographies numérisées

ISTEX pour le TDM

En plus de permettre l'accès rétrospectif à la littérature scientifique, ISTEX a la particularité d'être une source de données compatible, par sa licence et par la qualité des données, avec la fouille et l'analyse de textes, le Text and Data Mining - TDM.



Le Fait marquant

Depuis le printemps 2019, l'Inist propose des collections de **corpus spécialisés** constituées par l'équipe ISTEX en vue d'une exploitation en traitement automatique des langues et en fouille de textes.

Ces corpus spécialisés sont composés de documents en texte intégral ciblant un domaine scientifique ou une thématique particulière pour répondre à différents cas d'usage :

- > Réaliser des **analyses scientifiques** (analyse syntaxique et terminologique, analyse temporelle de thématiques) ;
- > Mettre au point ou **tester des outils** (permettant la détection de noms d'espèces, de définition de concepts, de termes ou des outils d'exploration thématique) ;
- > Proposer des **corpus Gold** pour évaluer les performances d'un outil (Unitex).

Tous ces corpus spécialisés sont disponibles sur le portail data.istex.fr.



Dans la continuité de l'alimentation de sa plateforme, ISTEX compte de nouveaux chargements de corpus :

- > **4 nouveaux corpus éditeur chargés** : Lavoisier, Brepols e-books, Duke, ACS ;
- > **4 anciens corpus éditeur mis à jour** : Springer E-books, Nature, EEBO, Wiley.

Les documents du fonds ISTEX ont été **enrichis** afin de permettre le TDM.



26 corpus éditeurs chargés (+18 %)

8 corpus spécialisés

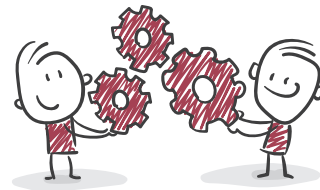
9 279 revues

345 369 monographies

23 027 718 documents (+5 %) dont 22 591 878 documents (98,11 %) avec au moins 1 enrichissement

73 740 142 enrichissements générés sur les documents

Partenariats et collaborations



En collaboration avec le **réseau des URFIST**, l'Inist a proposé des **ateliers pratiques** visant à aller plus loin dans l'usage des services et des outils gravitant autour d'ISTEX. Ces ateliers ont permis de construire un corpus spécialisé à partir du réservoir ISTEX et de le valoriser en utilisant des outils pour la visualisation et la fouille de contenus (Lodex, Gargantext, Cillex).

Cinq sessions ont eu lieu à Nancy, Bordeaux, Rennes, Villeurbanne et Toulouse, et ont réuni 46 stagiaires, parmi lesquels des professionnels de l'IST, des enseignants-chercheurs et des doctorants. Animateurs : Sabine Barreaux, Camille de Salabert, Stéphanie Gregorio, Claude Niederlender, François Parmentier, Pascale Viot.

Actions de communication



- **Présentation du format TEI utilisé dans ISTEX** (schéma ODD-ISTEX, pub2TEI, enrichissements en TEI), ANF TEI, Stéphanie Gregorio et Valérie Bonvallot, 3 décembre, Paris.
- **Se familiariser à la fouille de textes** par l'exemple à travers l'exploitation des corpus Istex afin d'accompagner les chercheurs dans ce nouveau champ d'activité dont ISTEX sera certainement l'infrastructure incontournable, Journées TDM organisées par l'ENSSIB, Laurent Schmitt, Sabine Barreaux, Camille de Salabert, Pascal Cuxac, Nicolas Thouvenin, 25 et 26 septembre, Paris.
- **Réalisation d'un projet d'étude** autour du « Merveilleux scientifique » (exposition BnF) avec 19 étudiants en Master PBD (Politique des Bibliothèques et de la Documentation), TD ENSSIB, Anne-Marie Badolato et Stéphanie Gregorio, 20 novembre, Villeurbanne.
- **ISTEX: A database of twenty million scientific papers with a mining tool which uses named entities**, Denis Maurel (université de Tours), Enza Morale, Nicolas Thouvenin, Patrice Ringot, Angel Turri. *Information*, 10 (5): 178.
- **SKEEFT: indexing method taking into account the structure of the document**, Pascal Cuxac, Nicolas Kieffer, Jean-Charles Lamirel (Loria), 20th Collnet meeting, 5-8 novembre, Dalian, China.

ISTEX

Web de données

Le portail data.istex.fr vise à présenter et permettre un accès à l'archive ISTEX au travers :

- > des référentiels documentaires ;
- > de corpus scientifiques spécialisés ;
- > d'une revue de sommaire (titre de revues).

Toutes les données sont consultables sur le site, mais également au travers d'une base de données sémantique (un triplestore) garantissant l'interopérabilité des données sur Internet.



Le Fait marquant

Après avoir extrait les **entités nommées géographiques** dans les documents ISTEX, les équipes ISTEX ont cherché à aligner chaque lieu avec une entité d'un référentiel géographique (GeoNames). Pour un certain nombre de lieux, l'alignement a nécessité une **désambiguïsation**. Suite à un travail de recherche et développement, **une méthode automatique de désambiguïsation a été mise au point**. Les bons résultats obtenus par cette méthode ont été présentés dans différents colloques, et cela a permis de mettre à disposition, via le triplestore ISTEX, des **données alignées de qualité**.



Et aussi...

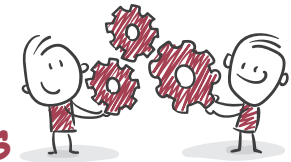
- Nouveau service d'exposition des données ISTEX** avec la mise à disposition des « dumps » (les fichiers RDF du triple store) sur un serveur web.
- Chargement des exports étendus** dans le triplestore, portant le nombre de triplets de 300 millions à presque 500 millions.
- Référencement de l'ontologie ISTEX sur le site **Linked Open Vocabularies**.
- La **navigation dans le portail** a été entièrement repensée pour permettre d'identifier facilement les différents types d'information mais également de naviguer entre chaque information simplement. Le site data.istex.fr est maintenant consultable quel que soit le type d'écran (Interface responsable).

Actions de communication



- ❑ **Des bases de données massives au Web de données : désambiguïsation et alignement d'entités géographiques dans les textes scientifiques**, Pascal Cuxac, Alain Collignon, Stéphanie Gregorio et François Parmentier. In 12^e Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ? Montpellier.
- ❑ **Feature selection and graph representation for an analysis of science fields evolution: an application to the digital library ISTEEX**. Jean-Charles Lamirel (Loria), Pascal Cuxac. BIR 2019, 8th International Workshop on Bibliometric-enhanced Information Retrieval ; 41st European Conference on Information Retrieval, 14-18 april, Cologne, Germany.
- ❑ **From massive databases to the Web of data: disambiguation and alignment of geographical entities in scientific texts**. Pascal Cuxac, 20th Collnet meeting, 5-8 november, Dalian, China.
- ❑ **LODEX : des données structurées au web sémantique**, Stéphanie Gregorio, Alain Collignon, François Parmentier, Nicolas Thouvenin. Atelier Web des Données de la 19^e Conférence sur l'Extraction et la Gestion des Connaissances (EGC 2019), janvier, Metz.
- ❑ **Alignement automatique et désambiguïsation d'entités géographiques dans ISTEEX**, Pascal Cuxac, Journées d'études GIS-URFIST : De P2N à l'IST - analyses brevet, état des lieux et perspectives. IMSIC (Institut méditerranéen des sciences de l'information et de la communication), 4-5 juillet, Toulon.
- ❑ **Désambiguïser les entités géographiques dans les publications de la base ISTEEX**, Pascal Cuxac, Journée d'étude « Données et protocoles dans les Humanités Numériques - Rechercher les lieux, les lieux de la recherche - Identification et indexation des terrains d'étude dans les textes scientifiques : enjeux et pratiques ». Laboratoire Géographie-cités, 18 juin Paris.
- ❑ **Désambiguïser les entités géographiques dans les textes scientifiques**, article publié sur le site web Inist (lire p. 44).

Partenariats et collaborations



- ❑ **IRIS - IRIT** : travail collaboratif lors d'un séminaire sur « Désambiguïsation et alignement d'entités géographiques dans les textes scientifiques ».
- ❑ **Science Miner** : détection et extraction des références citées dans le fonds ISTEEX.

Appui au pilotage

L'équipe Appui au pilotage scientifique propose des études bibliométriques permettant de caractériser l'activité de publication d'un laboratoire, d'un institut ou de toute autre structure ou celle liée à une thématique scientifique : évolution du nombre de publications, mise en évidence des collaborations, principales thématiques, revues les plus utilisées, etc.

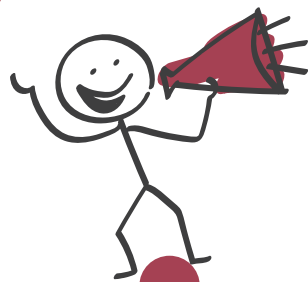
Parallèlement, des prestations de veille documentaire sont proposées pour dresser un panorama thématique, repérer des experts, rechercher des partenaires sur un sujet précis.



L'utilisation de l'outil ezVIS a été abandonnée au profit de l'outil **Lodex** pour la réalisation de rapports web dynamiques avec un thème personnalisé dédié à l'activité.

Le Fait marquant

Une **étude** a été réalisée mettant en évidence le **taux de publication en accès libre dans la production du CNRS**, en collaboration avec le Département des outils d'aide à la décision du MESRI.



Ils parlent de nous...

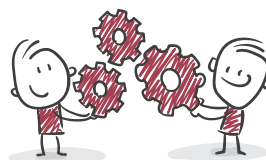
□ **Journal du CNRS**, décembre 2019, dans le dossier « Science ouverte, la révolution nécessaire » (p. 15).

Actions de communication

□ Participation à une **formation à l'outil Lodex** pour les étudiants du Master « Politique des bibliothèques et de la documentation » de l'Enssib, Anne-Marie Badolato et Stéphanie Gregorio.



Partenariats et collaborations



□ **GIS CollEx-Persée** : participation au projet DOPABAT financé par le GIS CollEx-Persée et qui a pour objectifs de connaître l'impact bibliométrique des thèses ainsi que la mise en évidence des coopérations entre laboratoires et entre disciplines (interdisciplinarité) dans les domaines de la Physique et de l'Astronomie.

□ **Service d'appui à la politique et à la prospective scientifiques (SAPPS) du CNRS** : poursuite du partenariat à travers l'hébergement des données, le repérage des laboratoires du CNRS et la réflexion sur de nouveaux indicateurs à produire.

22 études bibliométriques

14 prestations de profils bibliographiques

17 prestations de recherches bibliographiques

Traduction

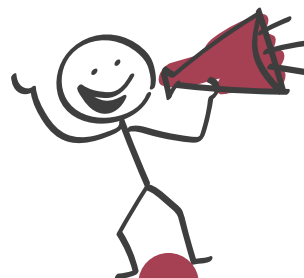
Le service propose des traductions et relectures au profit de communications scientifiques des laboratoires et institutions.

Ils parlent de nous...

Manifestation de satisfaction de Gaëlle BUJAN, déléguée à la protection des données du CNRS, pour la traduction du Guide *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte : guide pour la recherche en anglais*.

103 demandes traitées

720 568 mots (2400 pages environ)



Le fait marquant

L'année 2019 a été marquée par une **sollicitation institutionnelle plus forte** qu'à l'accoutumée en provenance des services centraux du CNRS et en partie du MESRI. On constate également l'apparition de nouveaux demandeurs issus de différents domaines disciplinaires.



Et aussi...

- ❑ L'image du CNRS mettant la **science ouverte** en avant dans sa politique générale, un volume important sur cette thématique a fait l'objet de prestations.
- ❑ Le site « **Ouvrir la Science** » du Comité pour la Science ouverte du MESRI a été alimenté par des traductions produites dans notre unité.
- ❑ Une **campagne de traduction en anglais des sites web des instituts du CNRS** a été lancée, les premiers étant INSU, INS2I et INC.
- ❑ Des didacticiels du **Code COUNTER** (ou Code de bonnes pratiques COUNTER) ont été traduits.

Actions de communication



- ❑ **Des didacticiels du Code COUNTER dans les mains du service Traduction**, article publié sur le site web Inist (lire p. 43).

Terminologie et Loterre

Les activités engagées en 2018 ont été poursuivies, orientées en particulier vers la production et la mise en ligne de nouvelles ressources terminologiques, et vers l'enrichissement des services proposés dans Loterre.



Le Fait marquant

Divers contacts reçus dans le courant de l'année montrent que l'offre de service de l'Inist en matière de terminologie a gagné en **visibilité**, grâce à

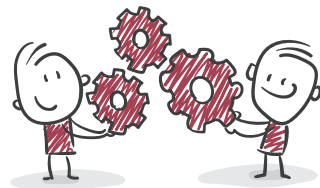
Loterre :

- > Sollicitations de participation à deux consortiums répondant à l'appel à projets ANR « Science ouverte » ouvert au printemps 2019 ;
- > Sollicitations pour exposer dans Loterre les ressources de nouveaux partenaires (INRA, Getty, Inserm, BDSP).

Lancement des travaux d'**évolution de Loterre** : l'objectif était d'enrichir Loterre en proposant des services aux utilisateurs (contrôle et transformations de fichiers terminologiques, alignements de terminologies, annotation de textes avec une terminologie).

L'alimentation de Loterre s'est poursuivie avec l'**ajout de 21 ressources** : 7 terminologies de l'INRA ; 6 terminologies du Getty, en lien avec OpenBibArt ; le thésaurus Santé publique de la BDSP, en lien avec OpenBDSP ; la version bilingue (français-anglais) du MeSH, pour l'Inserm ; 6 ressources produites par l'Inist (Nutrition artificielle, Optique, Tableau périodique des éléments, Théorie et traitement du signal, Chimie, Sciences de la Terre).

Partenariats et collaborations



- ❑ **Frantiq / PACTOLS** : validation des alignements avec GeoNames et enrichissement des branches « Italie », « Espagne », « Maghreb » et « Allemagne » du microthésaurus « Lieux » des PACTOLS.
- ❑ **INRA** : définition en commun d'un modèle de données, et génération de 7 terminologies fournies en SKOS à l'INRA et mis en ligne sur Loterre.
- ❑ **Getty-BHA** : travail d'alignement entre le vocabulaire SH (Subject headings), par rapport au thésaurus AAT (Art & Architecture Thesaurus) du Getty et production d'identifiants pérennes de type ARK pour 6 thésaurus BHA.

Actions de communication



- ❑ Abdelmajid Khayari, Gilles Banzet. **Alignement d'un thésaurus sur GeoNames : retour d'expérience**. 2019. (hal-02159910)
- ❑ Co-rédaction par Frank Arnould (Inist) et Xavier Aimé (Inserm) de l'ouvrage intitulé **Modélisation ontologique & Psychologies : une double influence**, publié en 2020 dans la collection MS2C des Éditions Matériologiques.
- ❑ Élaboration, avec le service Formation, d'une **vidéo de présentation de Loterre** (diffusion en 2020).
- ❑ Signalement des terminologies Inist de Loterre sur le **portail FAIRsharing**.
- ❑ **Des terminologies de l'Inra rejoignent Loterre et Thésaurus des éléments du tableau de Mendeleïev**, articles publiés sur le site web Inist (lire p. 42 et p. 44).

Le tweet



29 août 2019

Découvre **@INIST_Loterre**, ressource qui recense des thésaurus par disciplines scientifiques.. Dont feue la **#BDSP** ! Le tout en **#SKOS** pour un alignement sémantique et booster les usages automatiques. Bravo !

4 902 315 triplets (+407 %)

427 112 concepts (+502 %)

33 terminologies (+175 %)

Visa TM

Le développement de services de fouille de textes pour la recherche a été identifié depuis longtemps comme essentiel pour soutenir l'accès à l'information scientifique. Dans un contexte de science ouverte, la convergence des évolutions juridiques, scientifiques, techniques ouvre de nouvelles perspectives d'organisation autour d'infrastructures de service collectives.

Le projet Visa TM du Comité pour la Science ouverte est financé par le MESRI. Il a pour objectif l'étude d'une e-infrastructure de recherche pour la création d'une offre de service en fouille de textes pour la recherche, basée sur l'analyse sémantique et s'appuyant sur le potentiel de combinaison et d'adaptation offert par la plateforme européenne OpenMinTeD.



Le fait marquant

Une **journée publique** a été organisée afin de clôturer le projet Visa TM. Elle était destinée à dresser un état des lieux et discuter des perspectives concrètes ouvertes par les résultats du projet.

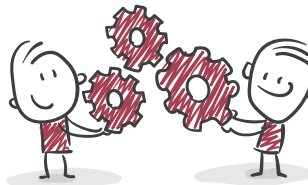


Les **résultats du projet** sont disponibles au travers une série de **8 rapports publics**, consultables et téléchargeables sur le site objectif-tdm.inist.fr. Les principaux thèmes portent sur :

- l'étude de la **création d'une e-infrastructure de fouille de textes** française en exposant ses besoins, ses acteurs, son organisation, ses missions, ses compétences et ses outils ;
- l'étude de **l'intégration logicielle et organisationnelle de services** pour la recherche : ISTEEX, TermSciences, AgroPortal, Alvis/OpenMinTeD ;
- la **communication** sur la fouille de textes vers ses communautés cibles.

Partenariats et collaborations

Inra / LIRMM : le projet Visa TM est le résultat d'une collaboration tripartite entre l'Inra, notamment l'équipe MaIAGE de Claire Nedellec, le LIRMM (Laboratoire d'informatique, de robotique et de microélectronique de Montpellier), et le CNRS avec l'Inist.



Actions de communication

Une **grande enquête nationale** a été menée : elle a reçu plus de 300 réponses.



Conditor

Conditor vise à recenser la production scientifique des unités et établissements de l'ESR. Conditor collecte dans différents entrepôts, agrège et fournit en métadonnées les dispositifs de l'ESR identifiés.

Partenariats et collaborations



- ❑ **AMUE** (Agence de mutualisation des universités et établissements) : intégration de Conditor dans le projet Caplab de l'AMUE afin d'associer les publications aux projets de recherche.
- ❑ **Université de Limoges** : le système de pilotage de la recherche de l'Université de Limoges, dans le cadre de son projet YUZU, a alimenté et enrichi ses rapports d'analyses décisionnelles par des données Conditor.
- ❑ **ANR** : l'ANR a testé l'usage de Conditor pour simplifier la saisie de ses formulaires de montage de projet.
- ❑ **EHESS** (École de hautes études en sciences sociales) : alimentation par Conditor du système de gestion de données de recherche de EHESS.

Le fait marquant



Lancement d'un **travail d'alignement** visant à aligner les notices avec le référentiel national des structures. Pour obtenir les meilleurs résultats possibles, deux chantiers en parallèle ont été lancés, avec deux techniques différentes :

- ❑ Alignements certains par la définition d'un ensemble de règles,
- ❑ Alignement par intelligence artificielle via du « Machine learning ».



Multiplication des sources d'alimentation de la base :
HAL, PubMed, CrossRef, Sudoc.

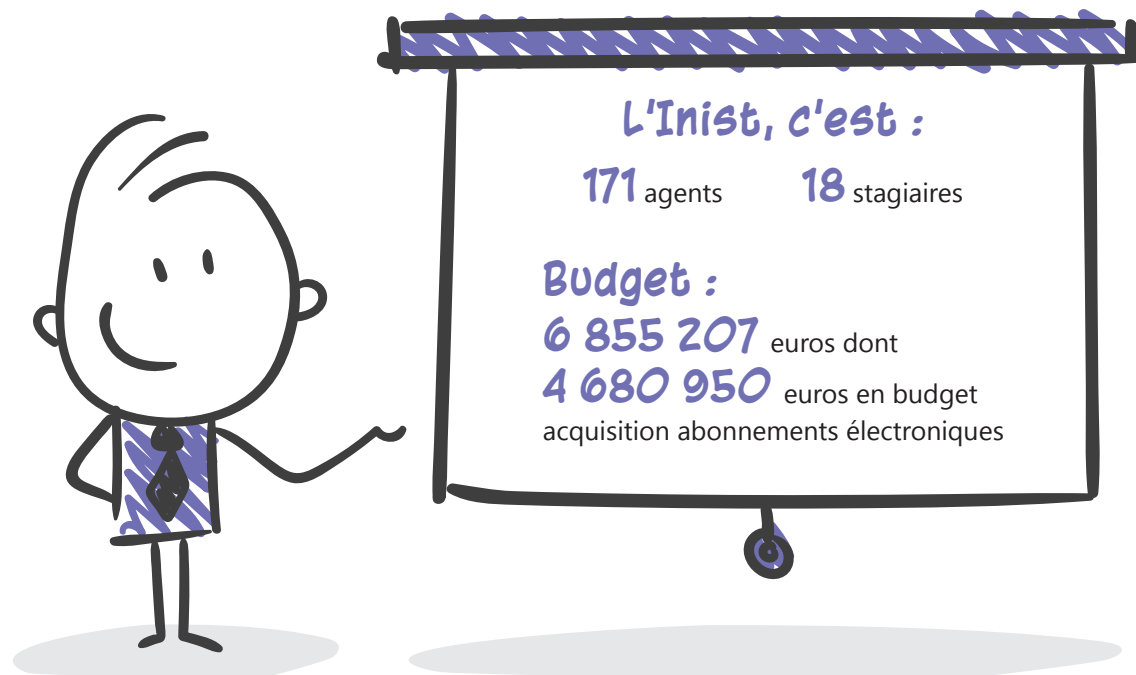
Actions de communication



- ❑ **Présentation de Conditor** à l'Unistra et à Sampra, Valérie Bonvallot et Claude Niederlender, 28 février, université de Strasbourg.
- ❑ **Conditor, un service pilote**, Valérie Bonvallot, présentation au MESRI, 14 mai.
- ❑ **Conditor : vers un référencement national de la production scientifique de l'ESR**, Valérie Bonvallot et Rémy Méja, café biblio aux Jabes, 28-29 mai, Montpellier.
- ❑ **Présentation de Conditor** au Comité National de Coordination de la Recherche (hôpitaux), Valérie Bonvallot, 9 septembre.
- ❑ **Présentation** lors du Comité des utilisateurs du RNSR, Valérie Bonvallot, 30 septembre, Paris.
- ❑ **Atelier sur les alignements avec le RNSR**, Valérie Bonvallot, Pascal Cuxac, Rémy Méja ScanR, OST, 26 novembre, Paris.
- ❑ **L'utilisation de la TEI dans Conditor**, Valérie Bonvallot, ANF du réseau Isore : La TEI pour l'IST, 3 décembre, Paris.

1 430 798 notices 795 940 notices uniques 4 sources

L'Inist, c'est aussi...



Les activités de l'Institut s'articulent autour de trois axes forts structurés par des équipes opérationnelles.

Ces dernières sont épaulées au quotidien dans leurs missions par des équipes dites **fonctions transverses** associées aux systèmes d'information, au secrétariat général, à la communication, au web et à la veille.

Systemes d'information

La stratégie de l'Inist ne pourrait se développer sans un Système d'Information (SI) sécurisé, moderne et performant, mis en œuvre par des professionnels, garantissant la continuité du service informatique fourni aux utilisateurs et implémentant les technologies récentes pour répondre aux besoins de sécurité et de changements à conduire en matière de SI.

Une trentaine d'agents aux **profils divers** se mobilisent chaque jour pour accomplir les larges activités transverses de l'unité. Leur engagement professionnel leur permet ainsi de contribuer au fonctionnement fluide de l'Institut et à l'accomplissement de ses missions générales de services en IST.

Secrétariat général

Le bon fonctionnement de l'Inist doit également s'appuyer sur une gestion rigoureuse, optimisée et prospective sur un ensemble de volets tels que : ressources humaines, financier, réglementaire, logistique et technique.

Le Fait marquant

2019 aura été notamment marquée par une **refonte complète** de notre solution de stockage et de son architecture.

Le Fait marquant

La **réorganisation de l'Institut** a entraîné un effort particulier tant au niveau des déménagements des deux tiers du personnel que des adaptations nécessaires de notre plan comptable et financier.



600 serveurs virtuels 40 serveurs physiques
800 To de stockage 250 bases de données
350 postes utilisateurs 3 robots de sauvegarde

1 500 bons de commande

450 ordres de mission

118 agents ont été formés
(3255 heures de formation)

50 contrats

5 prestations informatiques

Communication Web

Les activités de communication ont pour rôle de **valoriser et de promouvoir les services et les projets de l'Institut** ainsi que les diverses actions dans lesquelles l'Inist est engagé.

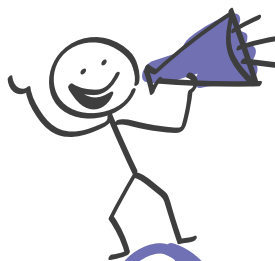
Les réalisations vont de la création de contenu, diffusion des actualités, animation des pages corporate sur les réseaux sociaux (Twitter/Facebook/Youtube/LinkedIn), création graphique pour les supports de communication à la conception de sites web.

Communication

Twitter : 3609 abonnés (+15 %)
252 tweets

Site Inist : 75 824 consultations (+38 %) dont
210 934 pages vues (+83 %)
81 actualités publiées
et 22 articles « Réalisations »

Web : 40 sites en production dont 6 nouveaux



Le fait marquant

Organisation d'un **escape game** pour les agents de l'Inist ; projet lauréat de l'appel à projets Qualité de vie au travail (QVT) 2019 lancé par la Direction des ressources humaines du CNRS.

Parmi les autres **actions de communication** (internes et externes) mises en œuvre :

- ❑ un **film court de présentation de l'Inist** a été réalisé avec l'objectif de présenter les missions de l'Institut (à voir sur le site Inist.fr ou sur la chaîne Youtube de l'Inist) ;
- ❑ deux « **séquences de com'** » ont été organisées dans les locaux de l'Inist afin de permettre aux agents d'élargir leurs connaissances sur certains sujets grâce à la venue d'intervenants extérieurs :
 - > *Des enquêtes orales aux données de la recherche : parcours professionnel d'une archiviste. Enjeux, méthodes, réseaux* par Véronique Ginouvès, responsable de la phonothèque à la Maison méditerranéenne des sciences de l'homme.
 - > *Science ouverte et biodiversité*, par Karim Ramdani, chercheur en mathématiques, Nicolas Fressengeas, chargé de mission « Science Ouverte » de l'Université de Lorraine, Jean-François Lutz, responsable de la bibliothèque numérique de l'Université de Lorraine et membre du CoSO, et Lionel Maurel, directeur adjoint scientifique de l'INSHS.

Actions de communication



- ❑ **Stand Inist** et présentation des trois axes, Congrès de l'ADBU, Cécilia Fabry, septembre, Bordeaux.
- ❑ **Parlons UX : refonte du site Inist**, présentation aux Journées de l'ABES, Cécilia Fabry, mai, Montpellier.

Veille

LaList

Le site **LaList** propose chaque jour des actualités en information scientifique et technique à destination des professionnels du domaine, des chercheurs, des enseignants, des doctorants, voire du grand public.

Ouvrir la science

Le site **Ouvrir la science** piloté par le CoSO est alimenté au fil de l'eau par un travail de veille opéré à l'Inist. Des billets sont notamment rédigés et publiés dans le blog du site « Actualité et veille sur la science ouverte ». À cela s'ajoutent la mise en ligne de ressources signalées et la mise à jour de l'agenda des manifestations.

3017 abonnés à LaList (tous canaux **+12 %**)

67 853 pages vues sur LaList (**+24 %**)

1373 billets publiés sur LaList (**+21 %**)

109 036 pages vues sur le site Ouvrir la science

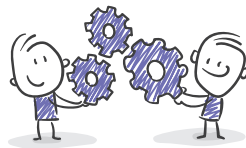
84 références agenda sur le site Ouvrir la science

9 newsletters envoyées pour Ouvrir la science

127 ressources signalées sur le site Ouvrir la science

90 billets publiés sur le site Ouvrir la science

Partenariats et collaborations



□ **CoSO** : pour les Journées nationales de la Science ouverte (JNSO) organisées par le CoSO, le service Communication de l'Inist a pris en charge la réalisation de deux posters et d'un kakémono valorisant les actions mises en œuvre dans le cadre du Plan national pour la science ouverte.

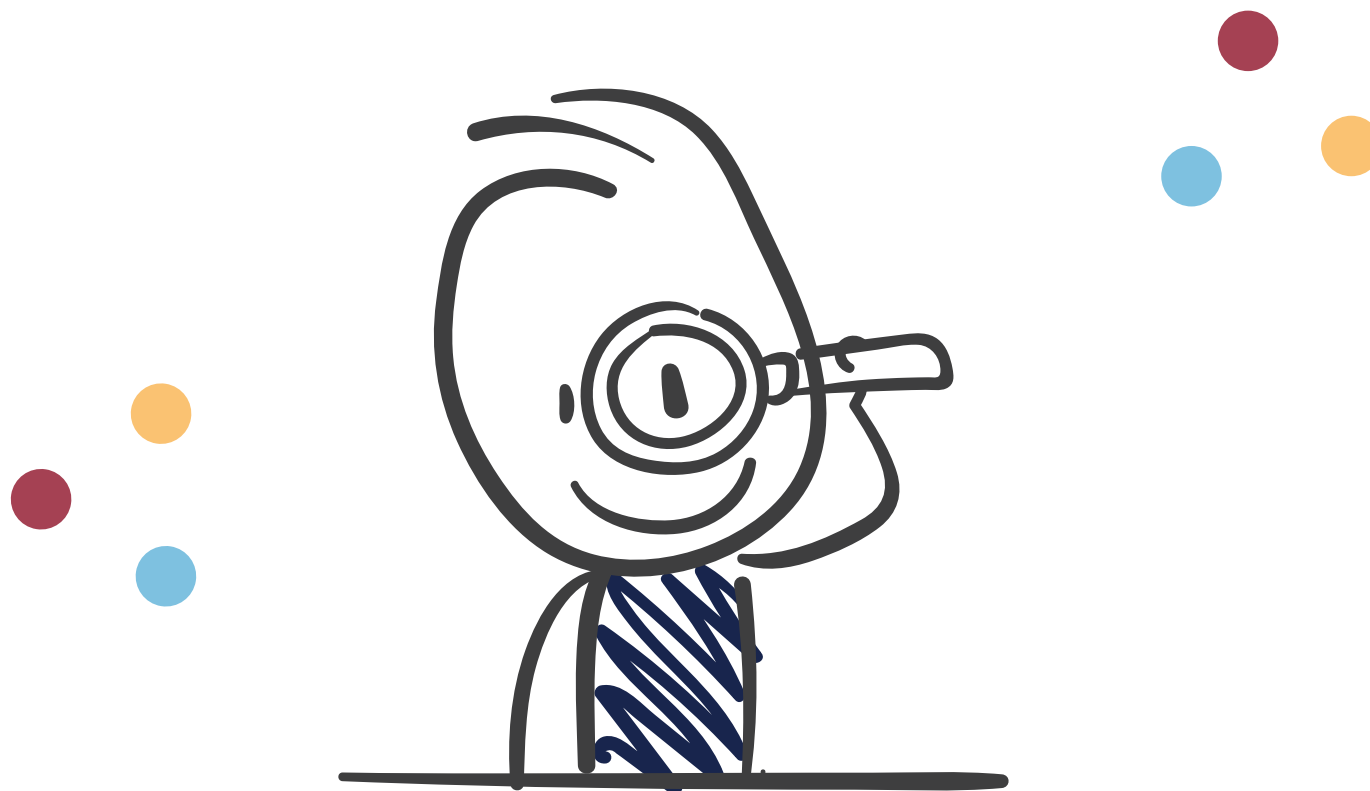
Actions de communication



□ **LaLIST, veille et diffusion de l'information scientifique et technique**, Brigitte Clément, article publié dans *Arabesques* janvier-février-mars 2019, p. 10-11.

Pour en savoir plus...

**Quelques articles
publiés en 2019 sur le site
www.inist.fr**



Le Cléo (devenu OpenEdition Center) et l'Inist partenaires en édition numérique depuis 2015

Publié le 5 février 2019

OpenEdition Center (Cléo) et l'Inist ont signé en 2015 un partenariat en édition numérique visant à développer les services OpenEdition dédiés à la diffusion en libre accès des savoirs produits par la recherche en sciences humaines et sociales.

Depuis 20 ans, le Cléo (devenu OpenEdition Center) développe le portail OpenEdition, un ensemble de plateformes de ressources électroniques en sciences humaines et sociales. : OpenEdition Books, OpenEdition Journals, Calenda et Hypotheses.

La plateforme OpenEdition Journals est dédiée à la diffusion numérique de revues, elle compte actuellement plus de 500 revues. Chaque site de revue bénéficie d'un ensemble fonctionnalités qui permet la structuration des contenus, la citabilité des articles, la pérennité des données, l'interopérabilité avec d'autres systèmes, le référencement ciblé, l'accès à des formats de lecture électronique, etc.

De son côté, l'équipe édition numérique l'Inist a développé un service d'édition électronique utilisant les mêmes techniques et logiciel pour sa plateforme I-Revues.

En 2015, OpenEdition Center (Cléo) et l'Inist se sont rapprochés afin d'apporter un meilleur soutien aux revues adhérentes aux plateformes OpenEdition. À la suite de formations dispensées par le Cléo, l'Inist prend en charge le stylage et la mise en ligne des documents numériques que OpenEdition Center (Cléo) lui transmet.

La volonté des deux partenaires est que ce partenariat s'étende. Ainsi l'équipe édition numérique de l'Inist prend en charge une partie de plus en plus importante du traitement des revues, utilise les outils Métopes[i] et est parfois en contact direct avec l'équipe éditoriale de la revue. Des actions d'amélioration des plateformes OpenEdition et de prise en charge des revues dès acceptation de leur demande de mise en ligne sont prévues à moyen terme.

[i] Métopes : Méthodes et outils pour l'édition structurée

|| Témoignage

L'Inist et OpenEdition sont partenaires pour la préparation éditoriale et la traduction de métadonnées de plusieurs publications diffusées sur OpenEdition.

– OpenEdition

Mapping de métadonnées et création de DOI en nombre pour l'UNJF

Publié le 20 février 2019

Membre fondateur du consortium DataCite, l'Inist est agence d'attribution de l'identifiant DOI en France ; identifiant pérenne favorisant le référencement et la citation des jeux de données.

L'équipe DataCite de l'Inist accompagne et assiste la communauté scientifique dans l'attribution de DOI aux ressources scientifiques.

Cette assistance s'est récemment illustrée dans une collaboration avec l'UNJF (Université numérique juridique francophone). Possesseur d'un compte DataCite à l'Inist, l'UNJF a eu besoin d'aide dans l'attribution de DOI en grand nombre : la demande concernait **127 ressources**.

La particularité de cette demande résidait dans le fait que ces 127 ressources étaient déjà décrites suivant le schéma de métadonnées SupLOMfr (Learning Object Metadata Français) ; schéma dédié aux ressources d'objets pédagogiques.

Afin de convertir ces fichiers au format DataCite, format obligatoire pour la création de DOI, il a été nécessaire dans un premier temps de faire un mapping entre les deux standards. Dans un deuxième temps une

application développée par l'Inist a été utilisée pour créer tous les fichiers de métadonnées selon le standard DataCite.

Une fois le mapping et la conversion réalisés, l'équipe a utilisé une application mise au point par nos informaticiens. Appelée « Pid Opidor », cette application « user friendly » permet aux usagers de créer des DOI en grand nombre, de mettre à jour des métadonnées et des URL ou encore de vérifier des URL associées aux DOI. Elle ne nécessite pas de compétences informatiques.

Grâce à un travail réalisé en étroite collaboration par différentes équipes de l'Inist (DataCite, informatique et web), l'UNJF a ainsi pu octroyer des DOI à ses 127 ressources pédagogiques.

|| Témoignage

Les équipes de l'Inist ont su apporter une aide précieuse et indispensable pour convertir un grand nombre de fichiers de métadonnées des cours de l'UNJF au format SupLOMfr vers le standard Datacite imposé pour la création des DOI.

– Grégory CHRISTIEN, Chef de projets GIS "UNJF"

Inist - HAL : trois missions en appui à la science ouverte

Publié le 22 mai 2019

À l'heure où le CNRS rend obligatoire le dépôt dans l'archive ouverte HAL pour les chercheurs dans le cadre de sa politique en science ouverte, l'Inist a plus que jamais un rôle à jouer aux côtés du CCSD.

Trois missions

L'Inist collabore avec le CCSD (administrateur de HAL) depuis 2015 à travers trois types de missions :

– **la modération** qui consiste à mettre ligne des documents déposés, à garantir la qualité des dépôts (vérification de conformité des métadonnées avec le fichier joint, contrôle du respect du cadre légal) ou encore à refuser des dépôts ne correspondant pas aux critères de l'archive.

De plus, l'équipe de l'Inist vient en aide aux déposants en leur apportant des conseils. L'Inist traite environ 70 % des documents déposés dans HAL, les autres dépôts étant modérés par des institutions (INRA, ArchiveSIC...) ou reversés directement par import.

– **l'administration du portail HAL-LARA** qui propose un libre accès aux rapports scientifiques et techniques.

– **la co-administration du portail HAL-CNRS** (<https://hal-cnrs.archives-ouvertes.fr/>) qui accueille l'ensemble de la production scientifique du CNRS.

Quelques chiffres

En 2018, on compte **62 256 actes de modération** dont **46 631 documents** mis en ligne et **14 637 demandes de modifications**.

3 206 rapports ont été signalés dans HAL-LARA.

Pour accompagner le dépôt

Avec l'objectif d'accompagner le dépôt des publications, l'Inist a collaboré avec le CCSD dans la réalisation de tutoriels vidéo permettant d'aborder rapidement la procédure du dépôt et quelques points précis. L'institut a également apporté des réponses à des questions posées via le service en ligne « Guichet numérique d'assistance-conseil en IST ».

Thésaurus des éléments du tableau de Mendeleïev

Publié le 6 juin 2019

À l'occasion du 150^e anniversaire du tableau de Mendeleïev, l'équipe Loterre de l'Inist a créé un thésaurus comprenant les 118 éléments chimiques reconnus actuellement par l'IUPAC (International Union of Pure and Applied Chemistry).

Sous l'impulsion du ministère de l'Éducation nationale, en association avec le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, l'année 2018-2019 a été déclarée « Année de la chimie de l'école à l'université ». Cette célébration fait écho à celle de l'Unesco qui a déclaré l'année 2019 « Année internationale de la classification périodique des éléments » à l'occasion du 150^e anniversaire du tableau périodique établi par le scientifique russe Dmitri Mendeleïev.

Un thésaurus trilingue

Le CNRS étant partenaire de cet événement anniversaire, l'équipe Loterre a proposé de réaliser un thésaurus des 118 éléments chimiques en trois langues (français, anglais, espagnol). Il est aligné avec l'ontologie ChEBI (Chemical Entities of Biological Interest), axée sur les « petites » entités chimiques dont les atomes, ainsi qu'avec wikidata, wikipédia, DBpedia.

Comme toutes les autres ressources de Loterre, ce thésaurus est libre de droits, consultable et téléchargeable sous différents formats (CSV, SKOS et PDF).

Les usages de l'Open Access se mesurent aussi

Publié le 20 juin 2019

ezPAARSE est devenu un outil de référence dans la mesure des usages. Conçu au départ pour produire des données d'usage sur les ressources payantes, ezPAARSE s'ouvre sur l'Open Access.

Utilisé par un grand nombre d'universités et d'organismes scientifiques français, ezPAARSE produit des indicateurs de consultation des ressources qui peuvent intervenir dans le choix de politiques documentaires. Il s'appuie sur les logs des serveurs mandataires qui gèrent l'authentification pour fournir des données fiables, une méthode qui ne peut pas être appliquée en Open Access.

Des contraintes techniques

Un **travail collaboratif** a été initié en 2018 pour permettre aux équipes de HAL et d'Open Edition d'utiliser ezPAARSE pour générer leurs statistiques d'usage en tenant compte des contraintes propres aux archives ouvertes. En effet, ces plateformes sont souvent accessibles sans authentification (pas de login dans les logs) et elles sont massivement téléchargées par des robots qui « polluent » les consultations réelles des humains. Elles génèrent donc un important volume de consultations dont il faut dissocier les consultations des utilisateurs de celles des robots. ezPAARSE a donc été enrichi d'un middleware (brique logicielle qui crée un réseau d'échange d'informations entre différentes

applications informatiques) adapté, afin de gérer les difficultés techniques liées à l'Open Access.

Un chantier couronné de succès

Grâce au travail des différentes équipes, ezPAARSE prouve désormais son efficacité en permettant d'identifier les consultations des robots et celles des humains dans des temps de traitement courts.

Cette réussite offre désormais la possibilité, pour les plateformes d'archives ouvertes, d'afficher des données de consultations plus fiables et plus réalistes.

Une nouvelle collaboration pour le cas des ressources mixtes

De plus en plus d'éditeurs scientifiques proposent des articles en Open Access au sein des revues traditionnellement payantes (revues mixtes) : comment évaluer la part d'usage de cette partie (gratuite) dans les consultations globales (payantes) ?

Un travail a été initié avec le consortium Luxembourg de la Bibliothèque Nationale du Luxembourg pour, là encore, ajouter un middleware dédié dans ezPAARSE qui permettra d'identifier cette part des consultations.

L'objectif de ce travail est de fournir des métriques destinées aux promoteurs du plan S pour faciliter la migration des publications vers l'Open Access.

Des didacticiels du Code COUNTER dans les mains du service Traduction

Publié le 2 août 2019

Accompagnant la recherche, le service Traduction de l'Inist propose aussi son expertise pour tous types de documents et d'utilisations. Les traducteurs ont récemment travaillé sur des didacticiels du Code COUNTER.

Avant les didacticiels, le code

Le Code COUNTER (ou Code de bonnes pratiques COUNTER) permet aux fournisseurs de contenu tels que les éditeurs de produire des données d'utilisation cohérentes, fiables et comparables, relatives à leur contenu en ligne. Il permet ainsi aux bibliothécaires de comparer les statistiques provenant de différents fournisseurs, des informations précieuses dans des décisions d'acquisitions.

Ce code est régulièrement mis à jour et aujourd'hui, on parle du Code COUNTER 5.

En 2018, le service Traduction de l'Inist avait déjà planché sur le sujet puisqu'il a traduit ce Code COUNTER 5, un document qui compte près d'une centaine de pages. Le suivi de cette traduction avait été coordonné en association avec le groupe Indicateurs de pilotage des politiques documentaires numériques de Couperin, ou GTI.

Des vidéos

Pour poursuivre la diffusion des bonnes pratiques et faciliter l'exploitation de ce code, des didacticiels sont mis à disposition. Cette année, le service Traduction a travaillé sur quatre didacticiels vidéo, aussi bien sur la traduction qu'une partie du doublage. Cette deuxième action a également été réalisée grâce au service Formation de l'Inist qui est intervenu pour apporter son support technique.

Témoignage

Pour traduire les tutoriels, j'ai dû me plonger dans le Code COUNTER lui-même et veiller à respecter la terminologie consacrée pour ne pas déboussoler les utilisateurs en bout de chaîne. J'ai eu l'occasion de participer à la session d'enregistrement audio, une très bonne expérience !

– Alexandre Clément, stagiaire à l'Inist dans le cadre d'un Master Traduction scientifique et technique à l'Université de Haute Alsace, traducteur des didacticiels du Code COUNTER.

Des terminologies de l'Inra rejoignent Loterre

Publié le 22 août 2019

Des ressources terminologiques de l'Inra, publiées il y a plusieurs années sous forme d'ouvrages, viennent d'être intégrées à la plateforme Loterre.

Une collaboration au service de la science ouverte

L'objectif exprimé par la DIST de l'Inra était de convertir les différents référentiels Inra présents dans le portail terminologique multidisciplinaire TermSciences dans un format davantage exploitable dans l'univers du **web sémantique**.

Ce reformatage a permis également de charger les terminologies dans Loterre, la plateforme d'exposition terminologique conçue par l'Inist, qui s'inscrit pleinement dans la science ouverte en offrant notamment différentes possibilités d'interrogation et de téléchargement.

Du format TMF au format SKOS et SKOS-XL

Le challenge de cette conversion depuis le format TMF a consisté à conserver la plus grande richesse possible, notamment les qualificatifs qui décrivent sur le plan linguistique les termes utilisés pour caractériser les concepts.

L'équipe Loterre de l'Inist a donc travaillé à l'optimisation de cette transformation, qui valorise désormais les terminologies en associant les formats SKOS et SKOS-XL. Liste des ressources produites par l'Inra déjà intégrées dans Loterre :

- Biotechnologies de la reproduction chez les Mammifères et l'Homme,
- Glossaire de génétique moléculaire et génie génétique,
- L'ionisation dans l'industrie agroalimentaire,
- Techniques de séparation par membranes,
- Écologie microbienne du tube digestif,
- Vocabulaire de la Politique agricole commune,
- La cuisson-extrusion.

Liste des ressources produites par l'Inra prochainement intégrées dans Loterre :

- Technologies des semences,
- Agriculture et systèmes d'élevage,
- Lexique de pédologie.

À noter également :

- La version 5.1 du thésaurus de la Banque de données en santé publique (BDSP) est intégrée à Loterre,
- Le MeSH produit par la NLM et traduit en français par l'Inserm (avec la participation de l'Inist) va prochainement intégrer Loterre.

Désambiguïser les entités géographiques dans les textes scientifiques

Publié le 21 octobre 2019

L'équipe ISTEEX de l'Inist a travaillé sur la désambiguïstation et l'alignement d'entités géographiques dans les textes scientifiques. Un article relatant ce travail et la méthode utilisée a été publié sur HAL.

Paris... capitale de la France ? Ville du Texas ? Ville de l'Ontario ? Une personne ?

La désambiguïstation de termes peut être très utile, notamment dans les travaux liés au traitement automatique de la langue.

Dans le cadre du projet ISTEEX, des actions d'enrichissement des données ont été menées en interne ou en collaboration avec des laboratoires ; des actions qui ont permis de procéder à une extraction d'entités nommées par le Laboratoire d'informatique fondamentale et appliquée de Tours (Lifat) et l'outil Unitex-Cassys.

Une méthode d'apprentissage automatique

Il restait à relier une chaîne de caractères extraite de son contexte avec une entité géographique précise.

Aussi, pour désambiguïser ces termes et les aligner automatiquement, l'équipe ISTEEX a développé une méthode par appren-

tissage automatique non supervisé (sans ressource) ; une méthodologie utilisant l'algorithme Adagram. L'expérimentation a été réalisée sur 400 000 documents en texte intégral issus d'ISTEEX, tous domaines confondus.

Désambiguïser pour aligner

Avec le projet Data-ISTEEX, il était question de mettre en avant les données produites en les exposant avec les principes du web sémantique (Linked Open Data), donc de les interconnecter avec d'autres réservoirs de données (wikidata, data.bnf, dbpedia, viaf, etc.). Ici, le but était donc d'aligner les entités nommées avec la ressource GeoNames couramment utilisée pour ce type de données géographiques et ensuite de pouvoir faire le lien avec d'autres réservoirs (data.bnf, wikidata par exemple).

L'utilisation de LODEX a permis de publier ces données alignées, le triplestore permettant quant à lui des interrogations complexes du réservoir.

Une évaluation positive et un travail reconnu

La campagne SemEval 2019 (évaluation sémantique) a montré une belle réussite de ce projet puisque l'évaluation s'élève à 0,946 pour l'alignement « géographique » (la meilleure évaluation possible étant 1).

Une mesure des usages « sur mesure » qui dépasse les frontières

Publié le 26 novembre 2019

Le département du Consortium Luxembourg de la Bibliothèque nationale du Luxembourg (BnL) a bénéficié d'une instance ezMESURE, une première ! En effet, cette plateforme était jusque-là réservée aux établissements de l'enseignement supérieur et de la recherche (ESR) français et membres du consortium Couperin.org.

Tout a commencé avec un déménagement : la BnL, où siège le Consortium Luxembourg, a quitté le centre-ville de Luxembourg pour rejoindre le quartier du Kirchberg, en périphérie de la ville. Avec les ouvrages physiques, il a également fallu déplacer toutes les données informatisées, dont les données d'usage.

En effet, depuis 2018, le Consortium Luxembourg utilise ezPAARSE, un logiciel libre et gratuit qui permet l'analyse, l'enrichissement et l'exploitation des logs d'accès aux ressources numériques dont l'Inist est l'opérateur technique, en collaboration avec le consortium français Couperin.org. Une convention a donc été signée entre ces trois acteurs afin que les données d'usage produites avec ezPAARSE soient hébergées à l'Inist jusqu'en juillet 2020.

Lancement du projet ezMETRICS

Ces contraintes de déménagement ont en outre permis de réaliser une expérimentation technique attendue par le Consortium Luxembourg : utiliser ezMESURE pour visualiser les données de consultation au Luxembourg, alors que cet outil n'était utilisé qu'en France jusque-là.

Couronnée de succès, cette expérimentation a initié le lancement d'un nouveau projet, ezMETRICS.

Avec l'objectif d'élargir cette mesure des usages, ezMETRICS vise à réunir ezPAARSE et ezMESURE en un outillage autonome installé en local ou en instance partagée. Ce toolkit repose sur l'évaluation de l'utilisation des contenus numériques sous licence, le monitoring des activités de lecture auprès des éditeurs classiques et des plateformes libres d'accès. Il permet une analyse de données mieux structurée.

En équipant ainsi tous les consortiums, cela permettrait d'obtenir une vue globale des consultations des chercheurs et enseignants et aussi de mesurer le taux de transition vers les ressources en Open Access. Alors qu'ezPAARSE et ezMESURE apportent déjà une aide dans la conduite des politiques documentaires, ezMETRICS pourrait

être un poids supplémentaire dans les accords transformants, un outil qui entrerait donc dans le cadre de la science ouverte et du Plan S au niveau européen.

Les consortiums intéressés

Thomas Porquet de Couperin.org et Patrick Peiffer du Consortium Luxembourg ont présenté le projet ezMETRICS au colloque ICOLC (The International Coalition of Library Consortia) qui s'est tenu à la BnL du 20 au 23 octobre 2019 et qui a accueilli la coalition internationale de consortiums assurant la gestion de ressources électroniques pour bibliothèques, centres de recherche et universités. Un webinar sera organisé rapidement pour les établissements intéressés représentés dans l'ICOLC, qui pourront s'inscrire dans la dynamique proposée et devenir partenaires internationaux.

Le projet ezMETRICS suscite un intérêt certain parmi les participants et les portes qui s'ouvrent pour sa mise en application sont nombreuses et prometteuses.



Le rapport d'activité 2019 de l'Inist est également disponible
en ligne sur son site internet dédié :

<https://ra2019.inist.fr>





Inist - Institut de l'information scientifique et technique

2, rue Jean-Zay - 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  

