



INIST-CNRS

**INSTITUT DE L'INFORMATION
SCIENTIFIQUE ET TECHNIQUE**

A modern, multi-story building with a grid-like facade of windows and panels. The building is set against a blue background with a white dot pattern. Several large, stylized speech bubbles with black outlines and blue interiors are scattered around the building, some overlapping it. The overall style is reminiscent of a comic book cover.

**Le rapport
d'activité
2021**

SOMMAIRE



ÉDITO.....	p. 4
LES FAITS MARQUANTS.....	p. 6
ACCÉDER À L'INFORMATION SCIENTIFIQUE.....	p. 9
HAL CNRS et édition numérique.....	p. 10
Portails et plateformes.....	p. 12
Négociations et mesure des usages.....	p. 14
VALORISER LES DONNÉES DE LA RECHERCHE.....	p. 15
Formation - DoRANum.....	p. 16
Partage des données - OPIDoR.....	p. 18
Exposition de données.....	p. 20
ANALYSER ET FOUILLER L'INFORMATION SCIENTIFIQUE.....	p. 21
Istex - Textes et corpus.....	p. 22
Text and Data Mining.....	p. 24
Ingénierie terminologique.....	p. 26
Appui au pilotage scientifique.....	p. 28
Traduction.....	p. 30
L'INIST, C'EST AUSSI.....	p. 31
Département systèmes d'information et secrétariat général... p. 32	
Communication - web - veille.....	p. 34
EN SAVOIR PLUS.....	p. 36



CLAIRE FRANÇOIS
DIRECTRICE

ÉDITO

LES SERVICES DE L'INIST EN ACCORD AVEC LES BESOINS



LAURENT SCHMITT
DIRECTEUR ADJOINT

TOUT AU LONG DE L'ANNÉE 2021, NOUS AVONS COMPOSÉ AVEC LA CRISE SANITAIRE QUI N'EN FINISSAIT PAS DE REVENIR, ET NOUS AVONS INTÉGRÉ DANS NOS PRATIQUES LE DÉVELOPPEMENT DU TÉLÉTRAVAIL.

Soutenu fortement par le CNRS qui porte une politique volontaire de recrutements permettant de faire face aux nombreux départs en retraite, l'Inist poursuit son ancrage dans la **politique de science ouverte**.

La mise en œuvre de l'entrepôt national de données de recherche "**Recherche data gov**" est particulièrement structurante pour déployer un écosystème d'accompagnement des communautés scientifiques dans la gestion et le partage de données de recherche. L'Inist prend toute sa place dans ce projet en pilotant le workpackage "communication, accompagnement, formation". Ce positionnement est en cohérence avec les services déployés dans ce domaine dont le succès ne se dément pas.

Les **formations** organisées à distance ou en présentiel ont permis à 2000 participants d'appréhender les différents aspects de la gestion des données et les principes FAIR qui visent à rendre les données de la recherche faciles à trouver, accessibles, interopérables et réutilisables. Elles complètent le site DoRANum qui reste toujours plébiscité. Le serious game "**Faut pas s'en FAIR**" dévoilé lors de l'ANF Science ouverte résume les compétences qui se déploient au sein des communautés.

La rédaction d'un plan de gestion est centrale pour organiser le traitement des données de recherche mais reste ardue à finaliser. Le déploiement de la version "**DMP OPIDOR machine actionnable**" propose de nouvelles fonctionnalités qui facilitent le travail par l'alimentation automatique de données présentes dans les plateformes partenaires, par exemple les informations liées à un projet ANR. En complément, les équipes accompagnent les acteurs à un rythme toujours plus soutenu (environ 30 interventions d'accompagnement direct par mois).

Le service d'**exposition des données** se structure de plus en plus. L'Inist déploie des bibliothèques numériques dédiées aux données de recherche pour le compte de 15 laboratoires et 4 MSH. De nouveaux projets sont régulièrement à l'étude avant déploiement. Le dynamisme de la communauté développant l'outil Omeka apporte des fonctionnalités de visualisation de plus en plus plébiscitées, ces dernières sont intégrées dans le service apporté à l'Inist aux laboratoires demandeurs.

Afin d'assurer au mieux sa mission historique de donner accès à l'information scientifique pour l'ensemble des communautés de recherche, l'Inist s'est organisé pour être à l'écoute des chercheurs et a déployé cet outil "magique" attendu depuis si longtemps qu'est le bouton **Click & Read**. Il permet en un clic d'accéder au texte intégral d'un article quelle que soit la page web consultée et aussi quel que soit le statut de l'article : en accès libre depuis des sources diverses ou comme ressource électronique acquise par l'établissement. Le succès a été immédiat (plus de 10 000 installations par les utilisateurs) montrant ainsi son adéquation avec les besoins.

En charge de l'administration du portail **HAL CNRS**, et suivant l'objectif fixé par le CNRS d'avoir 100 % des publications scientifiques accessibles, l'Inist poursuit son effort d'accompagnement des laboratoires et de modération pour faciliter le dépôt de leurs articles. C'est en moyenne 500 dépôts par jour qui sont effectués sur la plateforme et, pendant les "campagnes CRAC", jusqu'à 900 par jour.

Dans le domaine de l'**analyse et de la fouille de texte**, troisième axe de la feuille de route pour la science ouverte du CNRS, l'Inist définit ses nouveaux services en s'appuyant sur la plateforme Istex. Afin de faciliter son usage, les équipes vont au-devant des utilisateurs par des actions de formations (ANF TDM et ateliers pratiques) ou lors de la journée Istex de janvier 2021 qui a réuni 480 participants. Elles présentent les corpus thématiques prédéfinis selon des thématiques en lien avec l'actualité et exposés avec l'outil de visualisation Lodex et également l'ensemble du processus de constitution de son propre corpus à partir des ressources disponibles sur la plateforme. Aux côtés des ressources de textes préparés pour la fouille, les ressources terminologiques sont des outils essentiels. La conception et la publication du thésaurus trilingue de la science ouverte ont été plébiscitées.

Faciliter l'analyse et la fouille de texte passe par la mise à disposition d'outils simples d'usage applicables sur différents types de données. En cette année 2021, les équipes ont déployé les premiers **web services** de traitement de textes ou de métadonnées permettant de réaliser des alignements ou des classifications. Ces outils prometteurs faciliteront la réalisation des études bibliométriques déjà très demandées par les entités du CNRS à différents niveaux.

L'ensemble des services de l'Inist a un lien avec le développement de la science ouverte. Le service en charge de la **communication**, de la **veille** et des **supports web** n'est pas en reste en apportant un support actif dans la gestion et l'alimentation du site national "ouvri(l)ascience" qui a fait peau neuve en cours d'année.

Ce rapport d'activité montre d'une part la maturité des services mis en œuvre depuis quelques années dans le cadre renouvelé et impulsé par la lame de fond qu'est l'ouverture de la science et d'autre part notre intégration dans l'écosystème qui se met en place afin d'**accompagner les communautés scientifiques dans cette mutation**.

Bonne lecture...

2021 : LES FAITS MARQUANTS

JANVIER

RENOUVELLEMENT DES
RESSOURCES ÉLECTRONIQUES

INIST, AGENCE D'ATTRIBUTION DE DOI
EN FRANCE, REPRÉSENTÉ AU FESTIVAL
PIDAPALOOZA

FÉVRIER

UN CORPUS D'ENTRAÎNEMENT POUR
LE TDM DISPONIBLE SUR ISTEX

WOW!

MARS

L'INIST PRÉSENT À L'ANF
TDM 2020

UNE NOUVELLE IDENTITÉ
POUR LE SITE LALIST

AVRIL

CLICK & READ...
C'EST PARTI!

OBERRED : UN NOUVEAU
MOOC SUR LES DONNÉES
DE LA RECHERCHE

RÉSEAU :
PASSAGE À
100 GIGAB/S

MAI

L'INIST SE DOTE D'UN
CONSEIL SCIENTIFIQUE ET
TECHNIQUE

JUIN

HALATHON CNRS : L'ÉQUIPE DU
PORTAIL HAL-CNRS SE MOBILISE
POUR LA RECHERCHE

JUILLET

UN THÉSARUS TRILINGUE DE LA
SCIENCE OUVERTE DANS LOTERRE

AOÛT

ÉTUDE TDM SUR LE
TÉLÉTRAVAIL AU CNRS

SEPTEMBRE

L'INIST PARTICIPE À L'ANR TOUR 2022

LANCEMENT DU PROJET DORANUM DISCIPLINAIRE

UTILISATION DES TERMINOLOGIES DE LOTERRE DANS DEEPLPRO

PRODUCTION D'UNE ÉTUDE BIBLIOMÉTRIQUE POUR ANALYSER LA PRODUCTION SCIENTIFIQUE DU LABORATOIRE GÉORESSOURCES

L'INIST INTÈGRE LA TASK FORCE D'EOSC PID POLICY AND IMPLEMENTATION



OCTOBRE

L'INIST DANS LE MAGAZINE "LA RECHERCHE"

POSTER CLICK & READ 1^{ER} PRIX AUX JABES

ANF SCIENCE OUVERTE JEU : FAUT PAS S'EN FAIR SUR LES DONNÉES DE LA RECHERCHE

CRAC : LA MODÉRATION HAL BAT SON PLEIN

NOVEMBRE

EXPOSITION DE DONNÉES : OUVERTURE D'UN SITE AVEC LE LABORATOIRE PACEA

DE L'ONTOLOGIE ONTOTM À TM TOOLS EXPLORER

READMETRICS RETENU POUR LE PROJET EOSC-PILLAR

DMP OPIDOR NOUVELLE VERSION, C'EST PARTI !

DÉCEMBRE

L'INIST DEVIENT LE MIROIR UNPAYWALL POUR LE CNRS

METOSTEM & RAPIDO : L'INIST, PARTENAIRE DANS DEUX PROJETS LAURÉATS

KARGER E-BOOKS : 1^{ER} CORPUS COLLEX-PERSÉE DANS ISTEX

AU 31 DÉCEMBRE 2021,
L'INIST COMPTAIT 154 AGENTS TRAVAILLANT
AUTOUR DE TROIS AXES FORTS :

 ACCÉDER À L'INFORMATION SCIENTIFIQUE

 VALORISER LES DONNÉES DE LA RECHERCHE

 ANALYSER ET FOUILLER L'INFORMATION SCIENTIFIQUE



L'INIST MET À DISPOSITION PLUSIEURS RÉSERVOIRS D'INFORMATION INCONTOURNABLES : BIBCNRs, UNE BIBLIOTHÈQUE NUMÉRIQUE MULTIDISCIPLINAIRE ; ISTEX, UN RÉSERVOIR D'ARCHIVES ; PANIST, UNE ARCHIVE DE PUBLICATIONS RÉCENTES.

LANCÉ EN 2021, CLICK & READ SIMPLIFIE LA RECHERCHE ET L'ACCÈS AU TEXTE INTÉGRAL DEPUIS N'IMPORTE QUEL SITE ET PERMET UN ACCÈS DIRECT À DES MILLIONS D'ARTICLES EN ACCÈS LIBRE ET NOTAMMENT DANS HAL.



ACCÉDER À L'INFORMATION SCIENTIFIQUE



MICHÈLE BONTHOUX
RESPONSABLE DU DÉPARTEMENT

HAL CNRS ET ÉDITION NUMÉRIQUE

L'INIST S'IMPLIQUE FORTEMENT DANS LE SOUTIEN DE L'ARCHIVE OUVERTE HAL EN ADMINISTRANT LE PORTAIL HAL CNRS DÉDIÉ À LA VALORISATION DES PRODUCTIONS SCIENTIFIQUES DU CNRS ET EN ASSURANT LA MODÉRATION DE DÉPÔTS. L'INIST APPORTE SON EXPERTISE POUR ACCOMPAGNER LA MISE EN LIGNE DE DOCUMENTS POUR NOS PARTENAIRES.

MICHÈLE BONTHOUX
RESPONSABLE
DU SERVICE
PAR INTÉRIM

CRAC : LA MODÉRATION HAL BAT SON PLEIN

Ces dernières années, deux règles ont renforcé l'importance du dépôt dans HAL :

- la **Loi pour une République numérique de 2016** qui autorise les chercheurs à mettre en libre accès la version finale acceptée pour publication des manuscrits, quel que soit leur contrat avec l'éditeur ;
- l'**obligation depuis 2019** pour les chercheurs de déposer leurs prépublications dans HAL.

La **campagne CRAC**

s'est étendue

du 5 octobre au

30 novembre 2021.

C'est une période

intense pour les

modérateurs HAL :

en général, c'est en

moyenne 500 dépôts par jour qui sont effectués sur

la plateforme, avec des pics pouvant aller jusqu'à

900 dépôts par jour !



UN ACCOMPAGNEMENT PERSONNALISÉ

L'équipe du portail HAL CNRS a **accompagné** les laboratoires qui le souhaitaient pour :

- le dépôt de leurs publications dans HAL ;
- la création d'une collection pour regrouper leurs publications ;
- l'ajout de texte intégral, la curation (dédoublonnage, enrichissement de métadonnées) ;
- la mise en place de statistiques adaptées.

MOBILISATION POUR LE HALATHON

À l'occasion du **HALathon**

organisé par CasuHAL

du 24 mai au 11 juin 2021,

l'équipe du portail HAL

CNRS de l'inist s'est mobilisée pour déposer

dans l'archive ouverte HAL les manuscrits et

prépublications transmis par les chercheurs

et chercheuses du CNRS.



CÔTÉ SCIENCE OUVERTE...

Les activités du service sont axées à 100 % sur la science ouverte. Parmi elles, le projet **MetoSTEM** a marqué l'année 2021 en devenant lauréat d'un appel à projets du FNSO.

Le projet MetoSTEM a été lauréat de l'appel à projets n° 2 "Publications" du Fonds national pour la science ouverte (FNSO).

Ouvert en janvier 2021, l'appel à projets avait pour objectif d'encourager l'édition, les publications ouvertes et son écosystème ; une démarche qui s'inscrit dans le premier axe "Généraliser l'accès ouvert aux publications" du Plan national pour la science ouverte.

Porté par le Muséum national d'Histoire naturelle (MNHN), dont l'inist est partenaire avec Métopes, MetoSTEM propose une solution technique pour la transformation numérique des monographies STM du MNHN.

Les outils et méthodes développés par ce projet permettront la publication numérique "freemium" des œuvres ainsi que la structuration et l'interopérabilité des données scientifiques sur l'écologie et la biodiversité.



PARTENARIATS ET COLLABORATIONS

Avec le **CCSD** pour la modération des dépôts dans HAL.

Avec le **MNHN** :

- pour le secrétariat de rédaction et d'édition réalisé à l'Inist pour le périodique scientifique *Anthropozoologica* ;
- pour le stylage, avec la chaîne éditoriale *Métopes*, des collections du Museum.

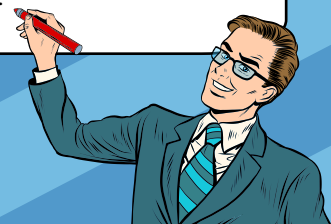
Avec l'**Abes** pour le stylage dédié à la mise en ligne des anciens numéros d'*Arabesques* sur son portail de revues scientifiques *Prairial*.

Collaboration avec **Open Edition Center** dans l'édition de plusieurs revues et ouvrages.

SENSIBILISER LES CHERCHEURS À L'OPEN ACCESS

Les 2 et 3 mars, deux journées d'ateliers ont été organisées à l'Institut Gustave Roussy, animées par les deux administrateurs du portail HAL CNRS.

Les ateliers s'inscrivent dans une **démarche de sensibilisation à l'open access et à l'archive ouverte HAL auprès des chercheurs.**



UN PEU DE COM'

Poster **Arabesques en ligne - 4 partenaires autour d'un même projet**, Journées de l'Abes, 12-13 octobre, Montpellier.

La modération HAL, un gage de qualité ! article publié sur le site web Inist, octobre 2021 (lire p. 40)

Présentation du portail HAL CNRS et retour d'expériences, Colette Orange et Théo Rousselle (administrateurs HAL), Atelier RENATIS C@FÉ RENATIS, 9 novembre.

Témoignage...



"Sans l'engagement de nos collègues de l'Inist, l'installation des différents formats numériques (HTML, PDF et ePub) sur la plateforme OpenEdition Books n'aurait pu se faire."

Albéric GIRARD, MNHN

EN CHIFFRES...

Édition numérique

- 4 numéros d'*Arabesques* traités pour mise en ligne
- 1 ouvrage paru de la collection *Natures en sociétés* (MNHN)
- 1 ouvrage paru de la collection *Archives* (MNHN)
- 470 articles mis en ligne pour OpenEdition
- 18 articles édités d'*Anthropozoologica* (MNHN)

Portail HAL CNRS

- + DE 1 700 000 dépôts affiliés CNRS (dont 183 940 ajoutés en 2021)
- 780 collections d'unités CNRS (88 supplémentaires pour 2021)
- 32 249 d'articles en accès libre (soit 68 % des articles publiés)

Modération dans HAL

- 71 367 dépôts traités par l'Inist
- 55 495 publications mises en ligne
- 15 488 demandes de modifications ou conseils aux déposants



PORTAILS ET PLATEFORMES

LE SERVICE PORTAILS & PLATEFORMES A POUR MISSION D'ACCOMPAGNER LES CHERCHEURS ET PERSONNELS DE RECHERCHE DES UNITÉS DU CNRS DANS LEURS BESOINS DOCUMENTAIRES AU QUOTIDIEN EN LEUR PROPOSANT L'ACCÈS LE PLUS RAPIDE ET LE PLUS EFFICACE POSSIBLE AUX PUBLICATIONS SCIENTIFIQUES. LE SERVICE APPLIQUE UNE DÉMARCHE UX CONSTANTE ET PLACE LES UTILISATEURS AU CŒUR DES PROJETS ET DES RÉALISATIONS (WEBINAIRES D'ÉCHANGES, PHASES DE PROTOTYPAGE AVEC TESTS RÉGULIERS).



DOLORÈS DARDAINÉ
RESPONSABLE DU SERVICE

LANCEMENT DE CLICK & READ

La version 1 de l'extension unifiée **Click & Read** CNRS a vu le jour en mars 2021.



Conçue pour faciliter l'accès aux documents, l'extension **Click & Read** dans sa version 1 remplace et unifie les 3 extensions existantes des plateformes gérées par l'Inist (BibCnrs, ISTEK et Panist) auxquelles s'ajoute l'extension libre Unpaywall.



Lorsqu'un chercheur navigue sur Internet, quel que soit le site, quelle que soit la requête, le bouton **Click & Read** apparaît si le document est disponible.

L'extension a été réalisée dans une logique agile et UX (expérience utilisateurs) : l'équipe projet a consulté en continu les chercheurs et professionnels de l'IST via de nombreux webinaires d'échanges.

ISTEK POURSUIT SON CHEMIN

Poursuite de la **montée en compétences du service** (feuilles de style, chaîne d'ingestion...) et aboutissement des travaux par le chargement du corpus Karger contenant 133 ebooks en médecine humaine, une action réalisée avec le département "Analyser et fouiller l'information scientifique" de l'Inist.

Il s'agit du 1^{er} chargement issu des acquisitions du GIS Collex-Persée mené dans sa totalité par le service !

Cet ajout porte le nombre de corpus disponibles, via la plateforme IsteK, à 31 et accroît le nombre de documents accessibles pour l'ensemble de l'ESR à 23 354 261.



CÔTÉ BIBCNRS

Le marché avec la société **Ebsco Information Services** a été **renouvelé** pour l'usage de l'outil de découverte EDS, de la base de connaissance HLM et du résolveur de lien associé (cœur de fonctionnement du portail BibCnrs pour les recherches d'articles et de revues).

L'équipe a assuré l'**animation du portail** en publiant 75 actualités via le site web et les listes de diffusion de chaque domaine de BibCnrs ainsi que 92 tweets via le compte Twitter. Elle a aussi accompagné les utilisateurs au quotidien en traitant plus de 1800 demandes d'**assistance**.



CLICK & READ, UN OUTIL SCIENCE OUVERTE !



À l'heure de la science ouverte, **Click & Read** signale en priorité les 31 millions d'articles libres de droits du réservoir **Unpaywall** dont les documents de **HAL**, les 23 millions d'articles disponibles dans la plateforme vertueuse **IsteK** puis les 7 millions d'articles disponibles dans la plateforme vertueuse **PANIST**.

L'intégration prioritaire de la source libre **Unpaywall** dans **Click & Read** marque une volonté science ouverte qui se traduit par la signature d'un accord avec la société Our Research pour utiliser le miroir interne Unpaywall et bénéficier d'une mise à jour des données quotidienne.

PRIX DU MEILLEUR POSTER !

Poster **Click & Read** présenté aux Journées de l'Abes, 12-13 octobre, Montpellier

Prix du meilleur poster !



QUELQUES CHIFFRES...

Click & Read

10 748 installations de l'extension via les stores des navigateurs Firefox, Chrome et Safari
29 webinaires de découverte et d'échanges pour préparer la version 2

PANIST

15 260 consultations par 205 établissements

BibCnrs

6 765 443 consultations ezPAARSE en 2021 dont :
 › 2 753 792 articles scientifiques sous licence éditeur
 › 1 964 483 de recherches via les bases de données
 170 243 consultations d'articles issus de ressources libres
 50 898 chercheurs ayants droit : 13 345 chercheurs CNRS et 37 553 chercheurs non CNRS répartis dans 1521 unités CNRS tous types confondus
 7 625 abonnés aux listes de diffusion
 1 200 abonnés au compte Twitter

ISTEX

1 081 822 consultations documentaires

PARMI LES TWEETS QUE NOUS AVONS VUS...



OSUR-Observatoire des Sciences de l'Univers Rennes @OSURennes

Alors ça c'est vraiment bien ! Merci @INIST_CNRS 🙌
Une extension unifiée pour faciliter l'accès aux publications scientifiques : Click And Read
youtu.be/EIT4QfEOHd0 via @YouTube



youtube.com
Click And Read
L'Inist lance son nouveau service Click & Read : une extension unifiée pour faciliter l'accès aux publications...



Iris @laboIRIS

Le bouton magique attendu depuis si longtemps, merci @BibCnrs ! 🥰

BibCNRS @BibCnrs - 12 oct. 2021
[Le saviez-vous ?] [#ClickAndRead]
Besoin de gain de temps ?
Besoin de simplifier vos recherches documentaires ?
Pensez à Click & Read !
Une extension unifiée pour le CNRS mais pas que...
clickandread.inist.fr

NÉGOCIATIONS ET MESURE DES USAGES



CATHERINE FOURNIER
RESPONSABLE DU SERVICE

LE SERVICE NÉGOCIATIONS MET EN PLACE ET GÈRE LES ABONNEMENTS QUI ALIMENTENT LE PORTAIL BIBCNRS. IL INTERVIENT DANS TOUTES LES ÉTAPES, DE LA NÉGOCIATION AU SIGNALLEMENT DES RESSOURCES.

L'ACTIVITÉ "MESURE DES USAGES" DÉVELOPPE ET MET EN PLACE DES OUTILS RELATIFS À LA QUANTIFICATION DE L'UTILISATION DES RESSOURCES ÉLECTRONIQUES, À DESTINATION DU CNRS, DES ÉTABLISSEMENTS DE L'ESR ET DE CONSORTIA ÉTRANGERS.

LES FAITS MARQUANTS

Pour un meilleur signalement des bouquets dans les outils de gestion, l'Inist a développé **ezHLM** afin de visualiser l'ensemble des titres souscrits pour les 10 espaces de BibCnrs sur un même tableau de bord.

Dans ezMESURE, le module **ezCOUNTER** a été mis en production : il permet la saisie des identifiants SUSHI des établissements participants.

LES CONTENUS BIBCNRS ÉVOLUENT

Suite à la **campagne annuelle de découverte de nouvelles ressources**, BibCnrs s'est enrichi d'une nouvelle base de données, Sociology Database, et de 9 titres en chimie des polymères (Wiley). De plus, les accès à la revue *Science* ont été étendus aux instituts INP et INC. Près de 20 nouvelles ressources ont été testées par les chercheurs : 4 d'entre elles ont intégré BibCnrs.

CÔTÉ SCIENCE OUVERTE...

Pour identifier l'évolution des pratiques lors des consultations effectuées depuis BibCnrs, un **nouveau tableau de bord** a été constitué.

Ainsi, en 2021, sur les 2,758 millions d'articles consultés via BibCnrs, près de 150 000 consultations ont été effectuées sur des sources en open access.

Les **analyses des dépenses** réalisées par le CNRS pour publier en accès ouvert ont repris, avec le dépouillement des fichiers 2018 à 2020.

Près de 10 000 lignes ont été passées au crible afin d'identifier les montants mis en œuvre dans les laboratoires.

L'analyse couvre également l'évaluation des coûts des frais annexes (couleur, pages, figures) souvent associés, quels que soient les modes de publication.

Le **miroir Unpaywall** a été implémenté dans ezMETA et l'API développée. Ce service est utilisé dans les requêtes lancées pour identifier les sources en open access via Click & Read. Depuis sa mise en production en octobre, plus de 4 millions de requêtes ont été effectuées, correspondant à 89 millions de DOI recherchés.



LES PARTENARIATS ET COLLABORATIONS

Consortium Couperin : partenaire historique des développements autour de la mesure des usages - ezMESURE.

OCLC : renouvellement de la convention Couperin/OCLC. Ce partenariat a permis d'enrichir la base de connaissances d'ezPARSE avec de nouveaux parseurs.

EN CHIFFRES

45 négociations
Plus de 2000 titres de périodiques

ezMESURE

80 établissements français contributeurs

652 plateformes déclarées dans analogIST

468 parseurs actifs

Twitter

594 abonnés

209 tweets

LA COM'

Mise en place de **deux sessions mensuelles** de supports mutualisés de formation des contributeurs d'ezMESURE aux outils et notamment Kibana.

Les ressources documentaires, ça se négocie ! article publié sur le site web Inist, avril 2021 (lire p. 37)

À L'HEURE OÙ LA RECHERCHE TEND À RENDRE ACCESSIBLES LES TRAVAUX SCIENTIFIQUES DANS UNE DÉMARCHE DE "SCIENCE OUVERTE", L'INIST PROPOSE UN ACCOMPAGNEMENT ET DES OUTILS POUR LA GESTION, LE PARTAGE ET LA DIFFUSION DES DONNÉES DE LA RECHERCHE.

CES SERVICES ONT POUR OBJECTIF DE FACILITER L'ACCÈS AUX DONNÉES, LEUR INTEROPÉRABILITÉ AINSI QUE LEUR RÉUTILISATION.



"La valeur des données réside dans leur usage"
National Research Council, Bits of Power

VALORISER LES DONNÉES DE LA RECHERCHE



PAOLO LAI
RESPONSABLE DU DÉPARTEMENT

FORMATION - DORANUM

LE SERVICE PROPOSE UNE OFFRE DE FORMATION HYBRIDE SUR LA GESTION ET LA CONSERVATION, LA DIFFUSION ET LA RÉUTILISATION DES DONNÉES SCIENTIFIQUES LIÉES AUX ACTIVITÉS DE RECHERCHE ET PLUS LARGEMENT SUR LA SCIENCE OUVERTE. LA PLATEFORME DORANUM, RÉALISÉE EN PARTENARIAT AVEC LE GIS RÉSEAU URFIST, MET À DISPOSITION PRÈS DE 100 RESSOURCES D'AUTOFORMATION QUI ACCOMPAGNENT LES CHERCHEURS À CHAQUE ÉTAPE DU CYCLE DE VIE DES DONNÉES. CE DISPOSITIF EST COMPLÉTÉ PAR UNE OFFRE DE FORMATIONS À DISTANCE (WEBINAIRES ET CLASSES VIRTUELLES) ET EN PRÉSENTIEL (ATELIERS ET FORMATIONS) ANIMÉES PAR LES ÉQUIPES.



CLAIRE SOWINSKI
RESPONSABLE DU SERVICE

DORANUM DISCIPLINAIRE, C'EST PARTI !

Porté par l'Inist et le GIS Réseau Urfist, soutenu et financé par le MESRI, le **projet DoRANum disciplinaire (2021-2022)** a démarré.

Il vise à enrichir la plateforme DoRANum grâce à la réalisation de ressources pédagogiques dédiées aux éléments qui structurent les principes FAIR dans les disciplines scientifiques (ex. : les identifiants pérennes, les schémas de métadonnées, les entrepôts de données).

Un comité scientifique composé de 15 chercheurs a été constitué afin de créer les ressources avec les équipes pédagogiques.



UN MOOC POUR OBERRED

Dans le cadre du projet Oberred, l'équipe a participé à la réalisation d'un MOOC (cours en ligne) en anglais sur la gestion et le partage des données de recherche intitulé "**Basics of managing and sharing data**" lesson 2 "**Concepts and processes of RDM**" (99 inscrits).



Oberred est un projet Erasmus+, financé par la Commission européenne, dont l'Inist est partenaire.

Il a pour vocation de concevoir un écosystème pour reconnaître des compétences liées à la gestion et au partage des données de la recherche.

La reconnaissance des compétences se fait au travers d'open badges.

CÔTÉ SCIENCE OUVERTE...

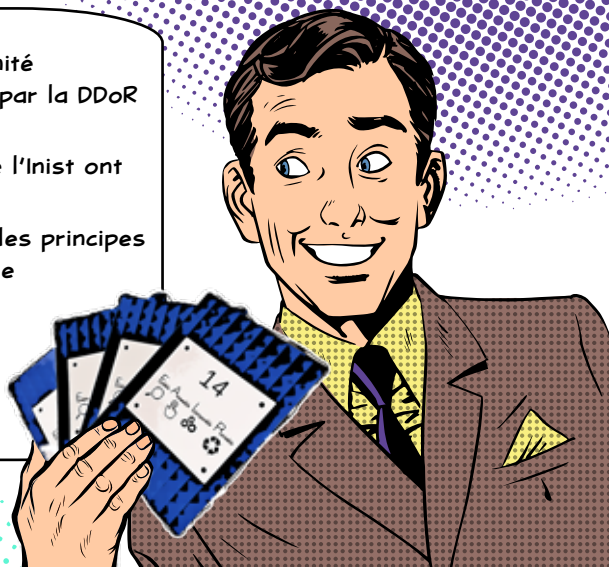
FAUT PAS S'EN FAIR !

Le service formation-DoRANum a participé au comité d'organisation de l'**ANF science ouverte** organisé par la DDoR les 19 et 20 octobre 2021.

À cette occasion, les ingénieurs pédagogiques de l'Inist ont créé le serious game "**Faut pas s'en FAIR**".

Il s'agit d'une extension de GopenDoRe axée sur les principes FAIR ; principes qui visent à rendre les données de la recherche faciles à trouver, accessibles, interoperables et réutilisables.

Accessible pour tous, le jeu est téléchargeable sur la plateforme Zenodo.



LES PARTENARIATS ET COLLABORATIONS

MESRI, INRAE, ESR : dans le cadre de Recherche Data Gov.

GIS réseau Urfist : partenariat dans le cadre de DoRANum et du projet DoRANum disciplinaire.

GIS réseau Urfist, CRFCB : partenariat dans le cadre de la plateforme Callisto (plateforme de formation à la documentation, aux métiers des bibliothèques et à l'information scientifique).

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, Gis 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.

Comité pour la science ouverte, Université de Lorraine, Université Toulouse Jean-Jaurès, GRICAD Grenoble, IRD, Consortium CAHIER, MATé SHS, MSH Dijon, MSH Lyon St Etienne, PUD Université de Poitiers, Université Aix-Marseille : partenariats dans le cadre de la mise en place de formations.

QUELQUES CHIFFRES

Site web DoRANum

81 883 visites

5094 téléchargements de ressources

138 199 pages vues

151 billets postés

+ DE 100 ressources pédagogiques

Webinaires, formations en présentiel et à distance

+ DE 1920 participants

Twitter

1351 abonnés

840 tweets



ON EN PARLE

À LIRE

Faut pas s'en FAIR avec le service Formation
article publié sur le site web Inist, octobre 2021 (lire p. 39)

EN TWEET

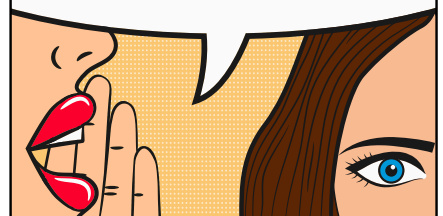
Urfist de Bordeaux @urfiatbordeaux
[On a testé pour vous] FAIR ? GPD ? DPM OPIDoR ? Pour y voir plus clair, on a suivi le parcours dédié de @DoRANum . C'est simple, clair et rapide, et on a eu tout juste à la fin ! On ne peut que vous le conseiller : à vous de jouer ! doranum.fr/enjeux-benefic...



4-08-PRÉ - 05 Juin 2021 - Twitter Web App

ON NOUS L'A DIT...

DoRANum est indispensable, c'est un site de référence sur la gestion des données, les ressources sont très claires et simples à comprendre.



PARTAGE DES DONNÉES - OPIDOR

LE PORTAIL OPIDOR MET À DISPOSITION DU CNRS ET DE L'ENSEMBLE DE L'ESR DES SERVICES ET ACCOMPAGNEMENTS SUR LES PLANS DE GESTION DE DONNÉES (PGD), LES IDENTIFIANTS DOI ET LES RÉFÉRENTIELS ET STANDARDS À CONNAÎTRE. L'OBJECTIF EST DE FAVORISER LA VALORISATION DES DONNÉES ET DE RÉPONDRE AUX PRINCIPES FAIR QUI VISENT À RENDRE LES DONNÉES FACILES À TROUVER, ACCESSIBLES, INTEROPÉRABLES ET PLUS FACILEMENT RÉUTILISABLES.

INSCRITS DANS LE PLAN D' ACTIONS DU PLAN DONNÉES DE LA RECHERCHE DU CNRS ET DANS LE PROJET RECHERCHE DATA GOUV (RDG), LES SERVICES OPIDOR ACCOMPAGNENT LE MOUVEMENT NATIONAL SUR L'OUVERTURE DES DONNÉES.

**JEAN-MICHEL
PARRET**

RESPONSABLE
DU SERVICE

DMP OPIDOR, NOUVELLE VERSION !

Mise en production de la **version DMP OPIDoR "machine actionnable"** : l'outil DMP OPIDoR a évolué pour permettre la production de plans de gestion de données non seulement lisibles par les humains, mais aussi exploitables par les machines.



POUR LA SCIENCE OUVERTE

Le service a rejoint le projet **Recherche Data Gov** (plateforme nationale fédérée des données de la recherche) au travers de groupes de travail (GT) nommés "**Work Package**" (WP) :

- Coordination du WP5 - Accompagnement, formation et communication ;
- Participation au WP6 : définition de la structure cible (structure, gouvernance, modèle économique, organisation RH et financière).

D'AUTRES TEMPS FORTS ONT PONCTUÉ L'ANNÉE...

Lancement du **projet METADoRe** pour un meilleur suivi des données de recherche et une meilleure qualité de la description de leurs DOI. Un pas de plus vers FAIR !

L'inist devient **membre de la Task force EOSC PID Policy and Implementation**. Le GT a pour objectifs de formuler des recommandations à la CE pour l'intégration des services PID dans l'écosystème de l'EOSC (PID actionnables par machine, méta résolveur, standardisation du graphe PID, intégration des PID dans la gestion des données FAIR, nouvelles technologies PID...).

Accompagnement à la rédaction de PGD dans différentes disciplines en : archéologie, urbanisme, architecture, géographie, géopolitique, économie, sociologie, mécanique des fluides, recherche participative, rénovation urbaine, géomatique, environnement, pollution marine, droit de la mer, nano matériaux, nano physique...

Création d'un **modèle Science Europe** "Machine actionnable".

Partenariat sur cinq cas d'usage CNRS : Microscopie électronique, Données nucléaires, Chimie analytique, Dynamique de fluide, Séquençage du COVID 19. Travail mené avec des laboratoires et les Instituts du CNRS afin de **développer l'adoption et l'implémentation des principes FAIR**.



LES PARTENARIATS ET COLLABORATIONS

Institut français de bioinformatique (IFB) et GenOvest : partenariat dans le cadre du projet maDMP4LS (projet ANR Flash).

GENCI, Meso@LR : collaborations pour la mise en œuvre du DMP machine actionnable (demande d'espace de stockage/calcul).

ANR, ANRT : actions de sensibilisation au modèle de DMP structuré.

Université de Lille, Le Mans Université, Université de Montpellier, Université de Paris Saclay, Université de Paris Est Créteil, Anses, INERIS, Normandie Université, Université de Paris 8 Vincennes-Saint Denis, ACTA Les Instituts Techniques Agricole (ITA), IRD, Université Picardie Jules Verne, Université Gustave Eiffel, ONERA, CEA : administrateurs de DMP OPIDoR, personnalisation de l'interface, recommandations, modèles de DMP.

Université de Paris 8 Vincennes-Saint Denis : intervention dans le cadre de l'Open Access Week organisée sur la gestion des données de recherche, les plans de gestion de données.

Université de Lorraine : l'Inist participe au comité opérationnel des données de recherche, notamment sur la mise en place de l'entrepôt Dorel au sein de l'université.

DataCite : membre du consortium, l'Inist est agence d'attribution de l'identifiant pérenne DOI en France.

RDA : membre des groupes Exposing Data Management Plans WG et Active Data Management Plans IG.

UC3-DCC : contribution au développement du code commun DMP Roadmap.

ARTICLE

Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M.-C., Cardoso, J., Kvamme, T., Praetzelis, M., Suchánek, M., Hooft, R., Faure, B., Moa, H., Hasan, A. and Jones, S., 2021.

Application Profile for Machine-Actionable Data Management Plans. *Data Science Journal*, 20(1), p. 32.

DOI: <http://doi.org/10.5334/dsj-2021-032>

CONFÉRENCES

DOIs Bulk update (1^{er} juillet 2021), Mohamed Yahia.

DataCite Member Meeting (septembre 2021) Mohamed Yahia.

Présentation de DMP OPIDoR et de ses évolutions à la **18^e plénière de RDA**, Benjamin Faure et Konogan Bourhy (IFB).

TÉMOIGNAGE



Je tenais tout d'abord à vous remercier pour la clarté de votre présentation d'hier, mais également pour l'utilité de l'outil que vous nous proposez.

Clémence LEROY
(Le Mans Université)

EN CHIFFRES...

DMP OPIDoR

- 43 modèles de DMP disponibles
- 9 200 comptes utilisateur
- 107 comptes administrateur
- 50 établissements de recherche
- Plus de 5000 connexions par mois
- 30 interventions d'accompagnement direct à DMP OPIDoR par mois avec des délais de réponse de moins de 24 heures



Cat OPIDoR

1135 structures et/ou services nationaux dédiés aux données de recherche ajoutés pour 436 pages créées dans le wiki.

471 services de données des infrastructures de recherche, centres et entrepôts de données thématiques.

DataCite

18 nouveaux adhérents
50 367 DOI créés

EXPOSITION DE DONNÉES

LE SERVICE VALORISE LES DONNÉES DE RECHERCHE SOUS FORME DE BIBLIOTHÈQUES NUMÉRIQUES OUVERTES SUR LE WEB DANS LE RESPECT DES FORMATS STANDARDS, DES PRINCIPES FAIR ET DES PRÉCONISATIONS DU PLAN NATIONAL POUR LA SCIENCE OUVERTE (PNSO). DEUX TYPES DE SERVICE SONT DISPONIBLES : L'UN POUR LA MISE EN LIGNE DE GRANDES BASES DE DONNÉES DOCUMENTAIRES, L'AUTRE DESTINÉ AUX LABORATOIRES QUI SOUHAITENT METTRE EN LIGNE DES DONNÉES DE RECHERCHE VIA LA CRÉATION DE BIBLIOTHÈQUES NUMÉRIQUES AVEC LE LOGICIEL OMEKA.

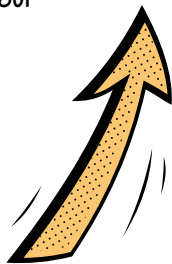


ALAIN ZASADZINSKI
RESPONSABLE DU SERVICE

UNE ACTIVITÉ EN HAUSSE EN 2021 !

Doté d'un scanner ZEUSCHEL et d'une plateforme de logiciels spécialisés pour le traitement numérique de documents, le service a réalisé **72 000 numérisations et traitements d'images** dont 60 000 pour l'alimentation de quatre plateformes Omeka.

Avec 53 000 réalisations en 2020, le service a augmenté de **40 %** son activité de numérisation.



LES PLATEFORMES OMEKA ÉVOLUENT

Mise en place d'une nouvelle plateforme avec OMEKA-S en remplacement d'Omeka-Classic : de nouveaux services vont pouvoir être proposés tels que la technologie IIIF pour la visualisation et la manipulation d'images en haute définition, l'intégration de vidéos 3D, une nouvelle version de cartographies ou encore la possibilité d'exporter les métadonnées sous des formats variés, transformant les plateformes d'exposition de données de recherche en outils de travail collaboratifs au service des laboratoires et des projets de recherche.



CÔTÉ SCIENCE OUVERTE

VALORISATION D'ARCHIVES DOCUMENTAIRES

Mise en production de la ressource "OpenBibArt", base de données bibliographique en histoire de l'art en libre accès réalisée en partenariat avec l'INHA (Institut national en histoire de l'art) et le Getty Research Institut (US).

Avec cette 4^e réalisation de mise en ligne d'archives documentaires, s'inscrivant dans la dynamique d'ouverture des données scientifiques telle que le recommande le PNSO, le service de l'Inist a ainsi mis en ligne plus de **27 millions de références bibliographiques** dont une grande

partie propose un **accès au texte intégral** en étant connecté à de nombreuses ressources telles qu'Istex, HAL, PANIST, BibCnrs, PubMed, Unpaywall, ArXiv, Base, WorldCat et GoogleScholar.



ON FAIT LES COMPTES



À la fin de l'année 2021...

le service gère 22 plateformes déployées avec Omeka-Classic dont 12 en production et 10 en cours de développement.

4 nouveaux projets de plateformes étaient également à l'étude fin décembre 2021.

15 laboratoires et 4 MSH ont ainsi déjà confié à l'Inist la mise en ligne et la valorisation de leurs jeux de données de recherche.

POUR RÉPONDRE À DES BESOINS DE TRAITEMENTS OU D'ANALYSE DE TEXTES, DE CORPUS OU DE DONNÉES SCIENTIFIQUES, FORMULÉS PAR DES CHERCHEURS, DES LABORATOIRES OU DES STRUCTURES INSTITUTIONNELLES, L'INIST PROPOSE SELON LA DEMANDE, DES RESSOURCES QUALIFIÉES ET PRÊTES À L'EMPLOI, DES OUTILS ET SERVICES D'EXTRACTION DE LA CONNAISSANCE, OU DE VISUALISATION DE DONNÉES.



ANALYSER ET FOUILLER L'INFORMATION SCIENTIFIQUE



NICOLAS THOUVENIN
RESPONSABLE DU DÉPARTEMENT

ISTEX - TEXTES ET CORPUS

LE SERVICE MET À DISPOSITION DE LA COMMUNAUTÉ SCIENTIFIQUE FRANÇAISE DES COLLECTIONS RÉTROSPECTIVES DE LA LITTÉRATURE SCIENTIFIQUE MONDIALE. IL DÉVELOPPE DES SERVICES DE RECHERCHE DE DOCUMENTS, DE TÉLÉCHARGEMENT ET DE VISUALISATION DE DONNÉES POUR PERMETTRE L'EXPLOITATION DE SES RESSOURCES POUR UN USAGE AVANCÉ EN FOUILLE DE TEXTES. IL MET EN PLACE DES ACTIONS DE FORMATION ET DE COMMUNICATION POUR ACCOMPAGNER LES CHERCHEURS ET LES PROFESSIONNELS DE L'IST DANS LA CONSTITUTION DE CORPUS DE TEXTES INTÉGRAUX ADAPTÉS À DES TRAVAUX D'ANALYSE DE CONTENU.



SABINE BARREAUX
RESPONSABLE DU SERVICE

19 JANVIER 2021 : JOURNÉE ISTEX

La **journée Istex** a réuni plus de 480 participants (partenaires, utilisateurs, décideurs) en visio pour faire un bilan de ce service proposé à la communauté scientifique française depuis 2019. Ce moment de rencontre a permis de réaliser un état des lieux sur les usages et les avancées de la plateforme et de ses outils associés, et d'échanger sur l'évolution des contenus et des services.

ISTEX

UN PREMIER CORPUS COLLEX-PERSÉE

De nouveaux corpus éditeurs sont régulièrement chargés pour enrichir la couverture scientifique des ressources mises à disposition. Cette année a vu l'arrivée d'un **premier corpus** acquis dans le cadre du groupement d'intérêt scientifique **CollEx-Persée** : Karger (133 e-books en médecine, publiés entre 2015 et 2019).

D'AUTRES TEMPS FORTS ONT MARQUÉ L'ANNÉE...

Les services associés à Istex s'enrichissent pour faire découvrir les ressources Istex et faciliter leur utilisation.

L'outil de visualisation **LODEX** se dote d'un **niveau de navigation supplémentaire** dans les données décrivant les publications et permet ainsi de mettre en valeur les enrichissements produits sur ces données.

Le site **Revue de Sommaire**, service phare de **data.istex**, propose deux nouvelles fonctionnalités de navigation :

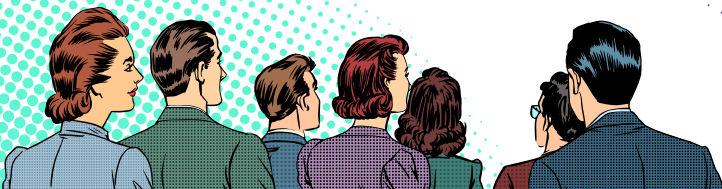
- la recherche sur le regroupement en collections proposé par l'éditeur, pour les ouvrages ;
- l'affichage des articles citant un ouvrage ou une revue dans l'archive Istex.

L'interface de téléchargement **ISTEX-DL** est maintenant accompagnée d'un **nouveau tutoriel** accessible en ligne sur le site Inist, sur **data.istex**, mais aussi sur la plateforme Callisto.

Science ouverte : accompagner la mise en place de nouveaux projets

En 2021, le service Istex était à nouveau présent lors de l'**ANF TDM** organisée conjointement par la DDoR du CNRS et l'INRAE sur le thème de "l'exploration documentaire et l'extraction d'information".

La **présentation de la constitution de corpus à partir du réservoir Istex** a permis de faire découvrir aux participants (chercheurs, des doctorants et des ingénieurs d'appui à la recherche) la richesse et la pertinence des données et des services Istex pour mettre en œuvre des projets de fouille de textes en toute indépendance.



Après l'ANF, les stagiaires ont pu bénéficier d'un **accompagnement personnalisé** dans le cadre de leur cas d'usage spécifique en architecture, en histoire et philosophie des sciences, et en astrophysique.

DES NOUVEAUX JEUX DE DONNÉES !

Le site data.istex.com compte cinq nouveaux jeux de données pour valoriser les ressources Istex :

- **EN-ISTEX** : Un corpus gold de 200 documents disposant d'entités nommées annotées manuellement pour permettre l'évaluation d'outils de détection automatique d'entités nommées ;
- **e-Éducation et Réfugiés - demandeurs d'asile** : deux corpus dans la collection SHS pour éclairer les problématiques actuelles à l'aide de publications d'archives ;
- **Publications sur Istex** : un jeu de données réunissant 48 publications provenant de HAL, Google Scholar, ArXiv sur l'utilisation des données Istex ;
- **Métriques sur les enrichissements Istex** : un jeu de données présentant l'évolution chronologique des différents types d'enrichissement des données Istex.

LES PARTENARIATS ET COLLABORATIONS

ARK Alliance : dans le groupe de travail "NAAN Registry Working Group" composé d'une équipe de conservateurs bénévoles et internationaux, il s'agit de gérer et superviser le registre des institutions qui demandent un numéro d'attribution de nom (NAAN) afin qu'elles puissent commencer à attribuer des ARK.

CÔTÉ COM'

Vidéo sur le **corpus Gold EN-ISTEX** présenté à TALN 2021.

Pastille vidéo sur **Istex pour le TDM** réalisée dans le cadre de l'ANF-TDM 2020.



À VOTRE RENCONTRE

Atelier pratique : Constitution de corpus à partir des ressources Istex

Medial-UL, Villers-lès-Nancy, mars 2021.

Journée sur la fouille de textes par l'exemple à travers l'exploitation des corpus Istex

ENSSIB - Inist (visio), avril 2021.

Journée sur l'usage des données ISTEX

URFIST Occitanie (visio), mai 2021.

Conférence : TALN 2021

Enza Morale, Denis Maurel, Jeanne Villaneau, Jean-Yves Antoine.

Corpus EN-Istex : un corpus d'articles scientifiques annoté manuellement en entités nommées.

Traitement Automatique des Langues Naturelles, 2021, Lille, France. pp.6-7. hal-03265916

EN CHIFFRES...

Istex, c'est :

31 corpus éditeurs chargés

13 corpus spécialisés pour la fouille de textes

3 corpus d'actualité

9 318 revues / 348 769 monographies

23 584 204 documents dont :

- 23 005 776 avec au moins 1 enrichissement (97,5 %)

- 86 796 568 enrichissements générés sur les documents

UN RETOUR QUI FAIT PLAISIR !

Journée Istex
19/01/21

Merci pour cette excellente journée, la haute tenue des échanges, la qualité des présentations et le déroulé de la journée ne méritent à mon avis que des éloges. L'intérêt du public est une reconnaissance de tout le travail accompli depuis toutes ces années, et les perspectives ouvertes par la table ronde donnent envie de continuer !

David Aymonin
(Abes)

EN TWEET


#istex2021 Bravo @ISTEX_Platform pour la démo #Beethoven en live de constitution d'un corpus thématique avec les différentes étapes pour affiner, corriger, etc. Excellent et très clair !

@ISTEX_Platform @ISTEX_Platform · 10 jan. 2021
#istex2021 Concrètement le corpus #Beethoven avec trois outils : 1) app #ISTEX pour l'interrogation, #ISTEX pour l'extraction, #Abes pour la visualisation.
#TDM #Abes #redim #abes #istex
#Istex cette discussion



10 jan. 2021 · Twitter Web App

TEXT AND DATA MINING



PASCAL CUXAC
RESPONSABLE DU SERVICE

LE SERVICE A POUR BUT DE METTRE À DISPOSITION DES WEB SERVICES DE FOUILLE DE TEXTE PERMETTANT L'ENRICHISSEMENT DES DONNÉES ET LEUR ALIGNEMENT AVEC DES RESSOURCES DIVERSES (RÉFÉRENTIELS, TERMINOLOGIES...). DES MÉTHODES DE TRAITEMENT AUTOMATIQUE DE LA LANGUE (TAL), D'APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING) OU ENCORE D'ANALYSE SÉMANTIQUE SONT MISES EN ŒUVRE AFIN D'EXTRAIRE DES TEXTES LES INFORMATIONS ESSENTIELLES.

Vers une offre de web services TDM pour une analyse assistée par ordinateur...

LE FAIT MARQUANT

Lancement des premiers web services sur la classification Pascal/Francis, l'attribution de RNSR, la détection de langue, Teeft (indexation), Unpaywall, et des services de mapping (halAuthorId/IDref/ORCID).

Enjeu : Faciliter l'usage de traitements de text & data mining pour :

- extrapoler de nouvelles données à partir des données d'origine,
- mettre en évidence des données existantes, mais cachées par le volume de données,
- réduire le volume de données pour des représentations graphiques.

Ambition : interconnecter les compétences des différents services :

- Bibliométrie et TDM,
- Corpus et TDM,
- Terminologie/Traduction et TDM.

ET AUSSI...

Le service a travaillé sur l'import de plus de 800 thèses d'exercice en santé de l'Université de Lorraine (UL) dans le portail HAL-UL à travers un dépôt massif et automatisé. Il s'est appuyé sur la base CorHAL.

Dans un contexte de science ouverte, le service TDM a réalisé une première étude comparative du WoS avec les archives ouvertes HAL et Pubmed pour analyser la complétude des réservoirs ouverts en vue d'études bibliométriques, mais également pour faire une analyse de l'Open Access, principalement pour les publications du CNRS.

L'ACTION SCIENCE OUVERTE LA PLUS IMPORTANTE

CorHAL : Alimentation de HAL par des bases externes internationales

Afin de proposer aux chercheurs des publications à déposer dans HAL, l'outillage développé doit permettre de collecter les métadonnées de publications et d'importer automatiquement le texte intégral.

Pour cela, l'Inist moissonne les données, les homogénéise, les dédouble et les enrichit via différents traitements TDM (classification thématique, attribution de codes RNSR, etc.), et met ensuite les corpus enrichis à disposition du CCSD.



LES PARTENARIATS ET COLLABORATIONS

Université de Lorraine (UL) : travail à la demande de l'UL sur l'extraction des thèses d'exercice de l'UL dans Conditor et le reversement automatique dans HAL-UL.

IRD : Utilisation de Lodex et des web services de TDM pour la mise en valeur des données de la base Horizon, également avec les données Conditor.

Abes : participation à l'enrichissement et à l'alimentation de Conditor.

CCSD : collaboration dans le projet CorHAL.

EN CHIFFRES

CorHAL :

2 174 406 notices

1 648 821 notices uniques

4 sources

Poster présenté aux Journées de l'Abes (JABES),
12-13 octobre 2021, Montpellier

CÔTÉ COM'

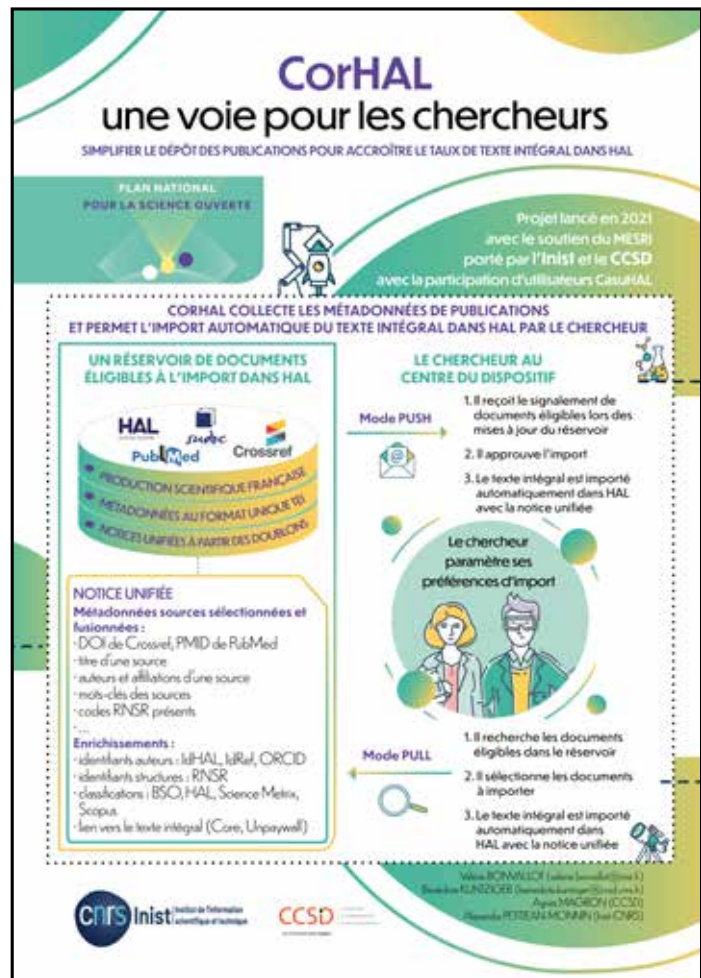
Atelier TextMine - Fouille de textes : Pascal Cuxac, Vincent Lemaire, Cédric Lopez, EGC 2021, Montpellier, Janvier 2021.

Cuxac P., Lamirel J.C., Biffet A. : **Workshop on IncrLearn "Incremental classification and clustering, concept drift, novelty detection in big/fast data context"**, ICDM 2021, 7 december 2021, Auckland, New-Zealand.

Lamirel, J.C., Gueddari, Y., Wang, Y. et al. **Analysis of the dynamics and influence of the research work of Prof. Liu Zeyuan in China featuring a new hybrid approach combining community detection with topic tracking.** *Scientometrics* 126, 6273-6300 (2021).

Intervention en **licence 3 Information et Communication** à l'UL pour une sensibilisation à la bibliométrie.

De l'**ontologie OntoTM à TM Tools Explorer** article publié sur le site web Inist, novembre 2021 (lire p. 42).



IL TÉMOIGNE...



"Les travaux de l'équipe TDM permettent de mettre en valeur le corpus constitué ainsi que de tester, valider et proposer à l'ensemble de l'ESR des nouveaux web-services qui interagissent entre les nombreux et différents corpus internationaux en faveur des problématiques actuelles telles que la Science Ouverte."

Nicolas Barts (IRD)

INGÉNIERIE TERMINOLOGIQUE



NATHALIE VEDOVOTTO
RESPONSABLE DU SERVICE

L'ANNÉE 2021 A ÉTÉ PLACÉE SOUS LE SIGNE DE LA SCIENCE OUVERTE AVEC LA POURSUITE DE LA MISE À DISPOSITION DE RESSOURCES TERMINOLOGIQUES SOUS LICENCE OUVERTE DANS LOTERRE, ET NOTAMMENT UN "THÉSAURUS DE LA SCIENCE OUVERTE" TRILINGUE : FRANÇAIS, ANGLAIS ET ESPAGNOL. DÉDIÉE À L'EXPOSITION, L'INTERROGATION ET LE TÉLÉCHARGEMENT DE RESSOURCES TERMINOLOGIQUES SCIENTIFIQUES, LA PLATEFORME LOTERRE S'INSCRIT DANS L'ÉCOSYSTÈME DU WEB DE DONNÉES OUVERTES ET FAIR.

Publié sur Loterre en juillet 2021, le thésaurus de la science ouverte a récolté un gros succès !

Réalisation et publication, sous licence ouverte CC-by 4.0, d'un **thésaurus trilingue** (français, anglais, espagnol) qui définit et structure plus de **400 concepts relatifs au domaine de la science ouverte** (<https://www.loterre.fr/skosmos/TSO/>)



À noter : la hiérarchie du thésaurus sera prise en compte par le projet "Cartographie des acteurs mondiaux de la science ouverte", porté par le collège Europe et International du CoSO.

LE FAIT MARQUANT



"Thank you to Inist-CNRS for publishing the Open Science Thesaurus (TSO)! I work with an organization based in the United States that conducts research about funding of open science infrastructure. We were very happy to find the TSO. We especially love its multilinguality and polyhierarchy."
Anne Britton (Research Fellow - Invest in Open Infrastructure <https://investinopen.org>)

UN AUTRE TEMPS FORT DANS L'ANNÉE...

RÉALISATION D'UN TUTO POUR LES TRADUCTEURS

Comment transformer une ressource terminologique hébergée dans Loterre en un glossaire multilingue pour **DeepL Pro**, un programme de traduction automatique ?



Réponse sur la chaîne YouTube "Inist-CNRS"



LES PARTENARIATS ET COLLABORATIONS

LSCE (Laboratoire des sciences du climat et de l'environnement) et DibiSO (Direction des bibliothèques, de l'information et de la science ouverte) de l'Université Paris-Saclay : projet de création d'un thésaurus de paléoclimatologie.

Inserm : finalisation de la version XML bilingue fre/eng du thésaurus MeSH pour les millésimes 2020 et 2021.

DibiSO de l'Université Paris-Saclay, DipSO, MalAGE et projet TransformOn de l'INRAE : projet TyDI qui consiste en la création d'un outil de sélection et organisation de candidats termes extraits par un extracteur de termes.

Participation aux Groupement de recherche Mémoire (GDR 2013) et SémanDiv (GDR 2011) : appui en ingénierie terminologique.



CÔTÉ COM'

EN VIDÉO

Vidéo de **présentation de Loterre** (à voir sur la chaîne YouTube "Inist-CNRS")



EN CONFÉRENCE

Présentation de Loterre et de la gestion FAIR des ressources terminologiques à l'Inist, atelier du groupe de travail "Catalogue, vocabulaires" de l'infrastructure de recherche Data Terra, mai 2021.

EN ARTICLES

Majid Khayari, Véronique Reszetko, Dominique Vachez, Nathalie Vedovotto, Jérémy Yon, et al.. **De TermSciences à Loterre : comment l'Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR.** 2021. hal-03176063

Dominique Vachez. **Étude comparative de thésaurus en Sciences de l'Environnement - Bonnes pratiques de conception et FAIRisation de thésaurus.** 2021. hal-03264803

Patricia Fener, Claude Dahdouh. **Repérage automatique d'énoncés définitoires avec Unitex pour l'aide à l'enrichissement de ressources terminologiques : retour d'expérience.** 2021. hal-03390661



EN CHIFFRES...

Loterre

Ajout de 2 ressources terminologiques :
- Taxonomie des poissons (en partenariat avec l'équipe Rameau de la BnF),
- Thésaurus de la science ouverte.

42 terminologies (+ 5 %)

464 280 concepts (+ 2,08 %)

5 295 211 triplets (+ 1 %)

APPUI AU PILOTAGE SCIENTIFIQUE



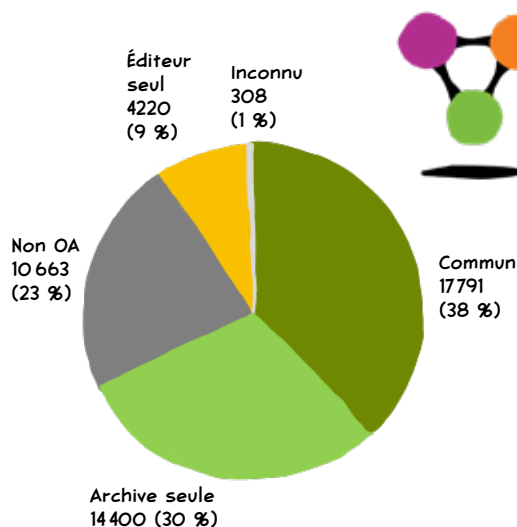
PATRICIA RANOARISON
RESPONSABLE DU SERVICE

LE SERVICE RÉALISE DES ÉTUDES BIBLIOMÉTRIQUES PERMETTANT DE CARACTÉRISER LA PRODUCTION SCIENTIFIQUE D'UN LABORATOIRE, D'UN INSTITUT OU DE TOUTE AUTRE STRUCTURE, OU CELLE LIÉE À UNE THÉMATIQUE SCIENTIFIQUE. PARALLÈLEMENT, DES PRESTATIONS DE VEILLE DOCUMENTAIRE SONT PROPOSÉES POUR DRESSER UN PANORAMA THÉMATIQUE, REPÉRER DES EXPERTS OU ENCORE POUR RECHERCHER DES PARTENAIRES SUR UN SUJET PRÉCIS.

Le fait marquant : l'étude des publications CNRS en accès ouvert (OA)

Le service a **quantifié les publications scientifiques en accès libre produites par le CNRS**. Ce chiffre clé correspond à l'indicateur 18 du Contrat d'objectifs et de performance 2019-2023 du CNRS "Pourcentage des publications des unités, dont le CNRS est tutelle, publiées en accès ouvert".

Réalisé avec l'outil **LODEX**, ce graphique présente la part des publications en accès ouvert sur 47 382 publications produites en 2020 par des unités du CNRS. Il propose la répartition selon le moyen d'accès : site de l'éditeur, archive ouverte, ou site et archive (Commun). (données WoS et Unpaywall).



LES AUTRES TEMPS FORTS DU SERVICE...

Participation au **développement de LODEX**, outil libre permettant de mettre en ligne et présenter des données structurées, notamment via des représentations graphiques.

Mise en place d'**actions de valorisation de la bibliométrie** : analyse des retours de nos usagers à propos des études bibliométriques et de l'outil LODEX.

LES PARTENARIATS ET COLLABORATIONS

SAP2S : poursuite du partenariat pour produire les indicateurs bibliométriques du CNRS (hébergement des données utilisées, repérage des unités de recherche du CNRS dans les affiliations des publications françaises, réflexion sur de nouveaux indicateurs).

IN2P3 : démarrage d'un partenariat pour créer un dispositif de génération d'indicateurs bibliométriques, à partir de la base INSPIRE - HEP et l'outil LODEX.

EN CHIFFRES

19 études bibliométriques
9 recherches bibliographiques
19 profils bibliographiques

FORMER...

Formation à l'IST, la bibliométrie et l'outil LODEX, **Accès expert à l'information**, pour les étudiants de licence 3 "Information et Communication" de l'Université de Lorraine, par Anne-Marie Badolato, Valérie Bonvallot, et Philippe Houdry.

... ET INFORMER

Bibliométrie : une collaboration étroite avec GeoRessources article publié sur le site web Inist, septembre 2021 (lire p. 38)

QUELQUES EXEMPLES D'ÉTUDES BIBLIOMÉTRIQUES

Suivi de la production scientifique des **unités de recherche de plusieurs instituts du CNRS** (INEE, INP, INSB, INSIS et INSU).

Suivi de la production scientifique française en **oncopédiatrie entre 2015 et 2020** à la demande de l'INSB.

Suivi de la production scientifique de **plusieurs unités de recherche du CNRS** (Géorressources, GANIL, IRIT, CRIOBE, OSU OTELO).

Suivi de la production scientifique française dans le **domaine de l'énergie entre 2014 et 2020** à la demande de la Cellule énergie du CNRS (INSIS).

Suivi des publications liées aux **projets scientifiques** que l'IPEV (Institut polaire français Paul-Émile-Victor) déploie en régions polaires et subpolaires.

Élaboration de stratégies de recherche pour obtenir plusieurs corpus et accompagnement pour l'utilisation des résultats à la demande d'équipes de chercheurs de l'INEE et l'INSMI qui ont répondu à l'appel à projets ANR "Programmes et équipements prioritaires de recherche".



ILS TÉMOIGNENT...

Merci à toute l'équipe pour ce travail conséquent de haute couture.

GUILLAUME CABANAC,
enseignant-chercheur,
chargé de mission
Scientométrie à l'IRIT (Institut
de recherche en informatique
de Toulouse)

Je vous remercie très sincèrement pour ce travail de recherche de publications, extrêmement utile pour notre Institut.

JÉRÔME CHAPPELLAZ, directeur de l'IPEV
(Institut polaire français Paul-Émile
Victor)

Je vous remercie pour votre étude que je trouve excellente.

FANNY FARGET, directrice adjointe du
GANIL (Grand accélérateur national
d'ions lourds)

TRADUCTION

LE SERVICE PROPOSE DES TRADUCTIONS ET RELECTURES ANGLAIS-FRANÇAIS ET FRANÇAIS-ANGLAIS DE COMMUNICATIONS SCIENTIFIQUES POUR LES LABORATOIRES ET INSTITUTIONS DU CNRS.

NICOLAS THOUVENIN
RESPONSABLE
DU SERVICE

AU SERVICE DU CNRS ET DES INSTITUTS

THE CNRS ACTIVITY REPORT

À la demande de la Direction de la communication du CNRS, le service s'est chargé de la **traduction du rapport d'activité 2020 du CNRS**.

AU PLUS PRÈS DE L'ACTUALITÉ

Traduction du **rapport de recherche Pandémie de Covid-19 : ce qu'en disent les SHS** coordonné par Marie Gaille, directrice de l'InSHS, et Philippe Terral, sociologue et professeur à l'Université Toulouse Paul Sabatier. Le document réunit les contributions de plus de 70 professionnels en SHS. La version anglaise a été publiée en livre sous le titre **The Covid-19 pandemic : What the HSS say?** par Alliance Athéna.

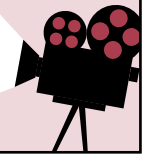


AU NIVEAU EUROPE

Traduction pour Knowledge Exchange du rapport **Openness Profile: Modelling research evaluation for open scholarship** publié en mars 2021. Ce rapport a ensuite été présenté au SpSO, à l'Institut du Cancer en décembre et discuté lors des journées Science ouverte du CNRS.

DES DOCUMENTAIRES

Trois textes à propos des éco-documentaires pour la revue numérique *Traverses* du site Internet film-documentaire.fr demandé par François-Xavier Destors, responsable éditorial, historien et auteur-réalisateur de documentaires.



L'ACTION SCIENCE OUVERTE LA PLUS IMPORTANTE



Les préprints : guide pratique

À la demande du CoSO (Comité pour la science ouverte), le service a traduit ce guide d'information et d'utilisation montrant aux chercheurs comment publier et employer les préprints, et à un public plus large (journalistes, patients, personnel de santé, etc.) et comment interpréter les travaux de recherche qui y sont présentés.

LES PARTENARIATS ET COLLABORATIONS

EOSC (European Open Science Cloud) : traduction (anglais/allemand/français) de textes significatifs sur la constitution de cette infrastructure européenne de la recherche destinés à une éditorialisation pour un support de sensibilisation auprès des communautés de recherche françaises.

EN TWEET

Des questions sur les préprints ? Les réponses sont dans ce guide pratique dispo sur hal.fr
co-rédigé notamment par
@JFric67 de l'INIST_CNRS
et Serge Bauin du @CNRS
#Preprint #OpenScience #ScienceOuverte
istat.fr/KATc2BX3



À LIRE

Traduction d'un rapport de recherche sur la pandémie de Covid-19 article publié sur le site web Inist, octobre 2021 (lire p. 41)

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 AU SECRÉTARIAT GÉNÉRAL, AUX SYSTÈMES D'INFORMATION, À LA COMMUNICATION, AU WEB ET À LA VEILLE.



**L'INIST,
C'EST AUSSI...**

DÉPARTEMENT SYSTÈMES D'INFORMATION ET SECRÉTARIAT GÉNÉRAL



SECRÉTARIAT GÉNÉRAL

LE FONCTIONNEMENT DE L'INSTITUT S'APPUIE SUR UNE GESTION RIGOUREUSE ET PROSPECTIVE NÉCESSAIRE AU SUIVI DE L'ACTIVITÉ. LE SECRÉTARIAT GÉNÉRAL ENGLOBE LES VOLETS RH (RESSOURCES HUMAINES), FINANCIER, RÉGLEMENTAIRE, LOGISTIQUE ET TECHNIQUE.

**STÉPHANIE
SIMON**

RESPONSABLE
DU SERVICE

GPEC

L'année a été marquée par la relance d'une **démarche GPEC** (gestion prévisionnelle des emplois et des compétences) afin de faire face à de nombreux départs en retraite pour les années à venir, et à de nombreuses arrivées.

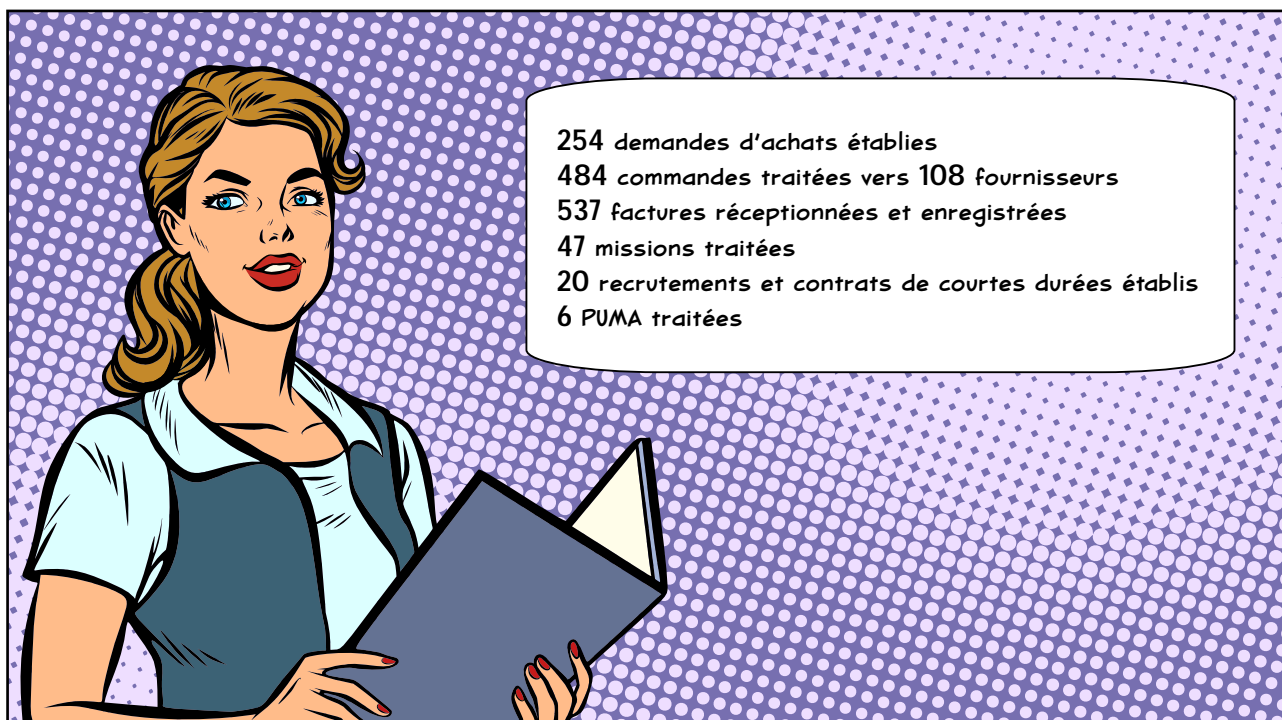
PARMI LES AUTRES FAITS MARQUANTS

Mise à jour d'une centaine de **fiches de postes**.

Don de mobiliers à des laboratoires rattachés à la Délégation Régionale du CNRS.

Optimisation de la procédure de suivi des recettes et premiers essais d'**Étamine Missions**.

QUELQUES CHIFFRES...





ANGEL TURRI
RESPONSABLE DU SERVICE

SYSTÈMES D'INFORMATION (SI)

LE SERVICE GARANTIT LA CONTINUITÉ DU SERVICE INFORMATIQUE FOURNI AUX UTILISATEURS.

MODERNE ET PERFORMANT, IL RÉPOND NOTAMMENT AUX BESOINS DE SÉCURITÉ.

2021 A ÉTÉ MARQUÉE PAR...

L'augmentation de la puissance de notre cœur de réseau passant de 10 GigaB/s à 100 GigaB/s.

La séparation du réseau des usagers de celui de notre datacenter afin d'anticiper la future migration de celui-ci vers le DataCenter Mutualisé Lorrain (DCML).

D'autres actions ont été réalisées telles que :

- ✓ Mise à disposition des environnements et accompagnement des projets : Click & Read, Unpaywall, CorHAL, Metadore, TDM services, Omeka, Lodex.
- ✓ Sécurisation des postes de travail : durcissement des règles des mots de passe.
- ✓ Nouveau logiciel de gestion du parc informatique : mise en œuvre de GLPI (Gestionnaire libre de parc informatique).
- ✓ Migration des applications et des données sous Ubuntu18.
- ✓ Salles de réunions adaptées pour permettre des réunions hybrides (présentiel/distanciel).
- ✓ Acquisition et Installation de commutateur 25 Gb/s et d'un cœur de réseau à 100 GB/s.
- ✓ Acquisition de 6 nouveaux serveurs Dell de type R640 pour un accroissement de 30 % de la puissance de calcul.

LES SYSTÈMES D'INFORMATION EN CHIFFRES...

- 28 serveurs physiques regroupés en 6 clusters (Environnement de virtualisation VMware)
- 660 machines virtuelles allouées
- 850 To de capacité brute de stockage
- 280 bases de données
- 35 structures d'indexation ElasticSearch
- 4 millions de requêtes https en entrée par jour
- plus de 98 % de disponibilité des applications
- Moyens de sauvegardes pour une capacité de 3 Po :
 - 7 serveurs physiques
 - 3 appliances disques VTL
 - 2 robots bandes LTO



COMMUNICATION - WEB - VEILLE



CÉCILIA FABRY
RESPONSABLE DU SERVICE

COMMUNICATION-WEB

LE SERVICE A POUR MISSION DE VALORISER ET DE PROMOUVOIR LES SERVICES ET PROJETS DE L'INIST. À TRAVERS UNE STRATÉGIE DE COMMUNICATION, IL MET EN PLACE DES ACTIONS DE COMMUNICATION EN ÉTROITE COLLABORATION AVEC LES ÉQUIPES DE L'INSTITUT.

IL CONÇOIT AUSSI DES SITES WEB ET DES SUPPORTS DE COMMUNICATION.

LE TEMPS FORT

Refonte du site *Ouvrir la science*, du Comité pour la science ouverte, hébergé à l'Inist. Elle s'est appuyée sur la méthodologie UX : ateliers de conception, ateliers de web design, architecture du site et développement.

EN VIDÉO

Mise en ligne de **9 nouvelles vidéos** de présentation des services...



... à voir sur la chaîne YouTube de l'Inist !

CÔTÉ PRESSE

Plusieurs articles ont été publiés dans la **presse spécialisée** :

- La science ouverte pour mieux gérer l'information scientifique, p. 114, n° 567 - Octobre-Novembre-Décembre 2021, *La Recherche*.
- Données de la recherche : "L'automatisation de service correspond vraiment à une attente", interview de la directrice de l'Inist Claire François, par Cyril Duchamp, *AEF Info*, dépêche n° 662398.
- Arnaud Moign, Cécilia Fabry, Anne-Marie Badolato et Claire François, "Du papier à la science ouverte : évolutions des services d'accès à l'information", *Histoire de la recherche contemporaine* [En ligne], Tome X - n° 2 | 2021, mis en ligne le 31 décembre 2021.

ET AUSSI...

3 posters ont été réalisés pour les journées de l'Abes : le prix du meilleur poster a été remporté avec Click & read.

Réalisation de campagnes de com' (Click & Read, DMP OPIDOR V3).

Reprise des stands lors des journées de l'Abes et de l'ADBU

LES PARTENARIATS ET COLLABORATIONS

Sciences sur la Place : Comme chaque année, l'Inist a collaboré avec la Délégation régionale Centre-Est du CNRS à l'organisation de l'événement multipartenarial "Les Sciences sur la Place" dans le cadre du Livre sur la Place à Nancy.

EN CHIFFRES

Web	38 sites en production dont 5 nouveaux
Site Inist	
65 479 visites	Twitter
239 106 pages vues	4 337 sites
21 308 visiteurs	363 nouveaux abonnés
70 actualités publiées	167 tweets
7 articles "Réalisation" publiés	22 186 visites du profil

EN TWEET

La **#Scienceouverte** en français, anglais, espagnol !
Le **#thésaurus** de la **#Scienceouverte** élaboré par le service Ingénierie terminologique de l'**@INIST_CNRS** en ligne sur la plateforme **@INIST_Loterie** sous licence **@creativecommons** CC-by 4.0 **#SemanticWeb**



Info :
Un thésaurus trilingue de la science ouverte dans Loterie - Inist
Nos actualités Thésaurus de la science ouverte (français, anglais et espagnol) actuellement riche de plus de 600 concepts science ouverte

2021-2022 | 2 juil. 2021 | Twitter Web App

LE SERVICE VEILLE

IL ACCOMPAGNE LE MONDE DE LA RECHERCHE ET LES PROFESSIONNELS DU DOMAINE EN METTANT À DISPOSITION EN CONTINU UNE MINE D'INFORMATIONS ACTUALISÉES SUR LE DOMAINE DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE VIA LES SITES LALIST ET OUVRIR LA SCIENCE.

**CÉCILIA
FABRY**
RESPONSABLE
DU SERVICE

UNE NOUVELLE IDENTITÉ POUR LALIST

Site de veille en IST (information scientifique et technique), LaLIST a fait peau neuve avec une **nouvelle identité visuelle** et un nouveau logo alliant sobriété et dynamisme.



VEILLE DE LA CPU

Les informations IST hebdomadaires collectées dans LaLIST contribuent à l'alimentation de la **veille d'information de France Universités**.

PARTENARIATS ET COLLABORATIONS

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, et Comité pour la science ouverte : collaboration pour la mise en ligne de contenus sur le site Ouvrir la science.

Cartographie internationale des acteurs de la science ouverte (CartoSO) : mise en œuvre de la base de données et de son outil de visualisation associé (Lodex)

Le projet CartoSO a été lancé par le collège Europe et international afin de permettre aux acteurs français de **se repérer dans le paysage mondial de la science ouverte**.

La science ouverte prend une place centrale dans l'écosystème de la recherche scientifique et le nombre d'acteurs qui intervient sur le sujet évolue de manière exponentielle. La CartoSO met à disposition un point d'entrée unique permettant de repérer les acteurs européens et internationaux, d'avoir accès à des informations descriptives et stratégiques sur ces acteurs.

L'Inist est impliqué dans la gestion du projet, dans son suivi et dans la création de l'application web.



LaLIST

2057 abonnés Twitter

1048 abonnés newsletter

Site web :

1328 billets publiés

52 317 visites

92 227 pages vues

23 606 visiteurs

Ouvrir la science

3876 abonnés Twitter

1372 abonnés newsletter

Site web :

250 articles publiés

98 570 visites

204 607 pages vues

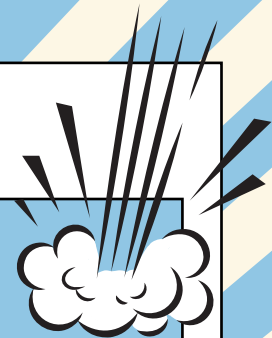
40 606 visiteurs



POUR EN SAVOIR PLUS...

- p. 37 : Les ressources documentaires, ça se négocie !
- p. 38 : Bibliométrie : une collaboration étroite avec GeoRessources
- p. 39 : Faut pas s'en FAIR avec le service Formation
- p. 40 : La modération HAL, un gage de qualité !
- p. 41 : Traduction d'un rapport de recherche sur la pandémie de Covid-19
- p. 42 : De l'ontologie OntoTM à TM Tools Explorer

LES RESSOURCES DOCUMENTAIRES, ÇA SE NÉGOCIE



Toutes les acquisitions de ressources thématiques pour le portail BibCnrs sont gérées à l'Inist par six négociateurs. Ils négocient chaque année avec une cinquantaine d'éditeurs et diffuseurs internationaux pour le renouvellement ou la mise en place de nouveaux abonnements. Les discussions portent sur les tarifs, les licences et les conditions d'accès. Focus sur un métier peu commun, mais ô combien utile à la communauté scientifique.

À l'Inist, l'activité de négociation existe depuis la création des portails du CNRS. En effet, depuis le début des années 2000, l'Institut gère BibCnrs, un portail multidisciplinaire qui permet d'accéder à un réservoir central de plusieurs millions d'articles, à des livres électroniques et à des bases de données scientifiques.

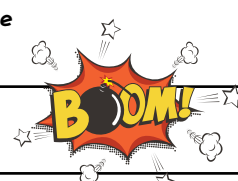
BibCnrs, c'est en tout 10 domaines scientifiques, à l'image des 10 grands instituts du CNRS.

Pour l'alimenter et permettre aux ayants droit d'accéder aux articles, cela passe par des négociations auprès des éditeurs.

QU'EST-CE QU'ON NÉGOCIE ?

Les ressources sont réparties dans des bouquets thématiques que les agents de l'Inist négocient avec les éditeurs. Il y a deux types de négociation : pour les renouvellements de licences, dont l'enjeu principal est de limiter la hausse des tarifs, et pour l'acquisition de nouvelles ressources.

Outre la négociation, les chargés de ressources documentaires ont aussi pour mission la gestion de ces dernières. Cela se traduit notamment par des analyses des consultations, en collaboration avec le service Mesure des usages-ezMESURE de l'Inist, à partir des statistiques recueillies auprès des éditeurs. Ces données sont précieuses, car elles peuvent servir de leviers dans les négociations de renouvellement, mais aussi dans les négociations pour l'ouverture de nouvelles ressources généralement testées lors des campagnes de découvertes annuelles.



UNE CAMPAGNE DE TEST POUR IDENTIFIER LES BESOINS

Afin de servir la communauté scientifique au plus près de ses besoins, une campagne de découverte est organisée chaque année. Gérée en lien avec les Cor'IST (correspondants en IST de chaque institut) qui interviennent dans la validation des ressources, elle permet aux chercheurs de tester en "conditions réelles", des revues, des bases de données, des livres électroniques.

Les ressources testées sont issues des suggestions des utilisateurs et des propositions des éditeurs. Leur acquisition dépend ensuite de l'avis des Cor'IST associés au portail ainsi que des budgets alloués par le CNRS.

LES NÉGOCIATIONS, UN BRAS DE FER ENTRE PUBLIC ET PRIVÉ

Les champs des négociations tournent autour du tarif, du modèle économique de l'éditeur, de la qualité de la revue, de la taille de la communauté et des conditions d'accès. Les discussions se font majoritairement à distance, par mail ou par téléphone, et très souvent en anglais.

Il s'agit d'échanges argumentés. Chaque négociateur a sa façon de faire, mais tous sont armés d'une grande patience. Ils sont ouverts à la discussion, mais ils savent être fermes quand cela est nécessaire et parfois il faut l'être, dans l'intérêt de l'institution, car il s'agit d'argent public.

Ils portent aussi une vigilance accrue aux licences. Ils vérifient scrupuleusement les termes du contrat tels que la pérennité de l'accès aux ressources souscrites même en cas de désabonnement.

Les négociateurs font parfois appel au service Traduction de l'Inist, notamment pour les parties plus "techniques" liées aux contrats.

Pour certains bouquets, l'Inist négocie avec et pour les membres de Couperin.org. En effet, certains négociateurs sont par ailleurs volontaires auprès de Couperin. Ils sont mandatés par le consortium et les établissements pour négocier des offres commerciales dans l'intérêt de l'ensemble de la communauté de l'Enseignement supérieur et de la recherche.

ET APRÈS...

À l'issue des négociations et des achats des ressources, l'équipe procède ensuite au signalement de chaque titre acquis et des autorisations d'accès associés dans l'outil de gestion de la bibliothèque : une action indispensable pour les rendre visibles et accessibles sur les portails.

BIBLIOMÉTRIE : UNE COLLABORATION ÉTROITE AVEC GEORESSOURCES

La bibliométrie est une activité qui, à partir d'indicateurs définis, permet de produire un bilan exhaustif de la production scientifique d'un laboratoire ou d'une thématique donnée. Installé à Nancy, le laboratoire de recherche en géologie GeoRessources s'est tourné vers l'Inist pour ses analyses bibliométriques. Une collaboration qui dure depuis plusieurs années.

En effet, la collaboration a commencé en 2018 lors de la venue dans les locaux de l'Institut du directeur de GeoRessources. Son objectif principal était de qualifier les interactions avec les autres structures et d'avoir une vue sur la production du laboratoire.

UNE ÉTUDE BIBLIOMÉTRIQUE PERSONNALISÉE

Au commencement, GeoRessources a transmis ses données à savoir que le laboratoire compte entre 100 et 150 publications par an pour quelque 60 chercheurs et enseignants-chercheurs.

À partir du fichier Excel reçu, les agents de l'Inist ont effectué des recherches dans les bases de données afin d'enrichir les publications du laboratoire par des informations de type thématique par exemple, trouvant aussi parfois des publications supplémentaires.

La première réalisation a été faite en juin 2018 sur les données 2017. S'appuyant sur l'outil ezVIS, l'équipe de l'Inist a établi une vue sur la production du laboratoire par année, par chercheur ou encore par équipe thématique.



VERS UN SUIVI DE LA PRODUCTION

Le travail s'est poursuivi par une mise à jour en 2019 avec les données 2018, puis en 2021 avec les données 2019. Ces mises à jour ont été importées dans Lodex, un logiciel open source dédié à la valorisation de données structurées produites par l'Inist. Cet outil offre une meilleure visualisation de l'analyse bibliométrique via des graphiques dynamiques. Il met en évidence l'évolution des publications, les thèmes et aussi les liens avec les laboratoires nationaux et internationaux.

Ainsi, grâce à une collaboration de plusieurs années, GeoRessources bénéficie d'un suivi de sa production. Il est à noter qu'à l'heure de la science ouverte, un indicateur de production en libre accès a été intégré à l'analyse bibliométrique du laboratoire permettant ainsi de montrer son implication dans cette démarche fortement soutenue par le CNRS.

LA BIBLIOMÉTRIE, POUR QUOI ?

La bibliométrie est un outil d'appui au pilotage scientifique. Elle permet de dresser un panorama et d'identifier des collaborations potentielles avec d'autres chercheurs et enseignants-chercheurs, d'autres laboratoires, d'autres institutions.

Le suivi de la production qu'elle fournit peut être utilisé pour appuyer les éléments qualitatifs lors des campagnes du HCERES (Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur).

Et bien sûr, la bibliométrie peut être mise en avant dans la communication officielle afin de valoriser la production du laboratoire.

AU SERVICE DES LABORATOIRES

Les membres de l'équipe Appui au pilotage scientifique mettent leur expertise au service des laboratoires CNRS en proposant un ensemble d'indicateurs et de représentations graphiques à partir d'un corpus de références bibliographiques.

Avec l'intervention de GeoRessources, le service a également partagé son expérience avec l'Université de Lorraine lorsque cette dernière a créé en 2020 sa nouvelle offre de services autour de la bibliométrie.

TÉMOIGNAGE

Le travail réalisé en collaboration avec les collègues de l'Inist fournit à GeoRessources un outil d'analyse et de visualisation des publications scientifiques du laboratoire. La force de l'outil Lodex est (1) d'être un outil pérenne, consolidé par des professionnels de l'information scientifique, (2) de permettre une communication sereine et exhaustive sur nos activités scientifiques, et (3) d'avoir un outil de pilotage scientifique efficace et "clef en main", le tout à la fois pour la direction, mais également pour tous les acteurs de GeoRessources.

Anne-Sylvie ANDRÉ-MAYER

Directrice de GeoRessources UMR 7359

FAUT PAS S'EN FAIR AVEC LE SERVICE FORMATION

Après le succès du jeu de plateau sur les données de la recherche GopenDoRe, le service Formation de l'Inist frappe à nouveau avec "Faut pas s'en FAIR", un jeu de cartes sur les principes FAIR.

En effet, les ingénieurs pédagogiques de l'Inist ne sont pas à court d'idées pour rendre ludique le thème des données de la recherche. Leur objectif premier est de briser l'appréhension liée aux principes FAIR. C'est réussi ! Et dans la bonne humeur !



RETOUR SUR GOPENDORE

C'est en 2018 que l'idée d'un jeu pédagogique a germé au service Formation. Le but était d'organiser des ateliers sur les bonnes pratiques de gestion des données de la recherche basés sur un format d'apprentissage ludique.

Après de longues semaines de travail et de réflexion, le jeu de cartes GopenDoRe a vu le jour début 2019. Découpé en quatre grandes étapes, le jeu suit le cycle de vie des données :

- Créer, traiter et analyser les données
- Préparer les données pour les rendre accessibles
- Diffuser les données (partage et réutilisation)
- Pérenniser les données (archivage)

Les cartes ont même été catégorisées par des couleurs à l'instar des thématiques du site DoRaNum permettant de s'auto-former sur la gestion et le partage des données de recherche.

PRÈS DE 700 TÉLÉCHARGEMENTS

S'adressant aux chercheurs comme aux personnels administratifs, le jeu a pour objectif principal de favoriser les bonnes pratiques de gestion des données de la recherche et le partage d'expériences, tout en suscitant des échanges, notamment avec les personnes soutien en local.

GopenDoRe a été joué aux quatre coins de la France lors d'ateliers de formation et aussi à l'occasion de l'OPIDoR Tour.

Fait pour être réutilisé sans modération, le jeu est imprimable depuis Zenodo. Il comptabilise d'ailleurs plus de 600 téléchargements à ce jour !

Suite aux demandes des utilisateurs, une adaptation numérique est en cours de réalisation.

UNE EXTENSION SUR LES PRINCIPES FAIR

Suite à ce succès, le service Formation est resté sur sa lancée en créant Faut pas s'en FAIR, en collaboration avec la DDOR (Direction des données ouvertes de la recherche). Il s'agit d'une extension de GopenDoRe axée sur les principes FAIR ; principes qui visent à rendre les données de la recherche faciles à trouver, accessibles, interopérables et réutilisables.

Le jeu se compose de 20 questions (5 par principe FAIR). Sa durée est de 1h30, et il peut être utilisé en complément de GopenDoRe.

Faut pas s'en FAIR sera présenté pour la première fois à l'occasion de l'ANF Science ouverte qui se tiendra les 19 et 20 octobre 2021 à Meudon. À l'issue de cet événement, le jeu sera téléchargeable sur la plateforme Zenodo.



LA MODÉRATION HAL, UN GAGE DE QUALITÉ !

La modération est un passage obligatoire pour le dépôt des publications dans HAL. Elle garantit la qualité du dépôt. Pour ce faire, le CCSD qui a créé, développe et administre l'archive ouverte HAL, s'accompagne d'une équipe de modérateurs à l'Inist.

En effet, en renfort des modérateurs du CCSD, plusieurs agents de l'Inist sont modérateurs HAL depuis 2015. Aujourd'hui, ils sont trois à inspecter quotidiennement les documents déposés sur la plateforme.



DES RÈGLES À RESPECTER

Un fichier éditeur est un fichier créé et produit par l'éditeur : mise en page, pagination, logo de l'éditeur, mention du copyright.

Suivant les consignes du CCSD et la procédure de modération, les agents comparent donc la notice au fichier PDF. Ils vérifient que toutes les métadonnées liées à la publication sont correctes (année, date, etc.) et ils vérifient aussi qu'au moins une affiliation française soit présente.

Si l'éditeur impose une durée d'embargo, l'auteur peut déposer son fichier, mais il devra fixer la durée dans HAL : la référence bibliographique sera visible en consultation, mais le fichier ne le sera qu'à l'expiration du délai fixé par l'éditeur.

Les chercheurs déposent sur la plateforme HAL leurs travaux contenant diverses métadonnées. Des principes généraux de contrôle s'appliquent alors à tous les types de documents, parmi eux : la vérification des fichiers PDF joints (le fichier principal doit contenir le texte intégral lisible) et la correspondance entre le type de publication et le document déposé. Les modérateurs vérifient également le statut juridique, qu'il s'agisse d'un fichier auteur ou d'un fichier éditeur.

Un fichier auteur est un fichier créé et produit par le ou les auteurs. On le distingue de la version publiée notamment parce que la mise en page comme la pagination ne reproduisent pas celles effectuées par l'éditeur.

UN VRAI TRAVAIL DE RECHERCHE

Les modérateurs ne se contentent pas de vérifier. Au besoin, ils corrigent ou complètent certaines données de la notice. Par exemple, pour les communications de congrès, ils vont jusqu'à rechercher et ajouter la date du congrès si elle est manquante. Ils peuvent aussi compléter les affiliations des auteurs.

Leur expérience respective dans le milieu de la documentation est un réel atout dans ces recherches documentaires.

PUIS VIENT LA MISE EN LIGNE

Si tout est correct, le modérateur valide le dépôt pour la mise en ligne sur HAL de la notice et du texte intégral. Mais le cas échéant, le modérateur peut demander des modifications au déposant ou refuser la publication si elle ne répond pas aux critères. Il peut aussi refuser le dépôt en cas d'article "douteux". Effectivement, il arrive que des personnes mal intentionnées tentent de faire passer des documents qui vont à l'encontre de l'éthique, mais heureusement les modérateurs veillent !

DES PÉRIODES PLUS INTENSES QUE D'AUTRES...

Les chercheurs étant tenus de fournir un compte rendu de leur activité (CRAC) pour leur évaluation, une campagne est organisée chaque année pour les inciter à déposer dans HAL ; une période très chargée pour les modérateurs

En 2020, durant la campagne CRAC, le stock quotidien des dépôts à traiter était de 3000 demandes. Bien que l'équipe de l'Inist traite en moyenne 300 dépôts par jour, les modérateurs ont eu besoin d'un renfort assuré notamment par d'autres agents de l'Inist.

LE CHIFFRE

44601 dépôts sur HAL ont été mis en ligne par les modérateurs HAL de l'Inist entre décembre 2019 et septembre 2020.

TÉMOIGNAGE

Depuis 2015, l'équipe de l'Inist contribue amplement à la modération des dépôts dans HAL, au bénéfice de toute la communauté scientifique française. Le CCSD est très heureux de cette collaboration.

Agnès MAGRON

Responsable communication et coordinatrice des activités de modération au CCSD

TRADUCTION D'UN RAPPORT DE RECHERCHE SUR LA PANDÉMIE DE COVID-19

Fort de son travail continu au sein du CNRS, le service Traduction de l'Inist a été sollicité pour traduire en anglais un rapport de recherche en SHS sur la pandémie.

Les sciences humaines et sociales face à la première vague de la pandémie de Covid-19 —Enjeux et formes de la recherche est un rapport de recherche publié par l'InSHS le 20 novembre 2020 (version 1).

Ce document réunit les contributions de plus de 70 collègues de différentes disciplines SHS (sciences humaines et sociales). Il a été coordonné par Marie Gaille, directrice de l'InSHS, et Philippe Terral, sociologue et professeur à l'Université Toulouse Paul Sabatier.

Sollicité par ces coordinateurs scientifiques, l'Inist a géré la traduction de la version 2 qui vient d'être publiée par Alliance Athéna, une alliance nationale de la recherche publique française dédiée aux sciences humaines et sociales.

57 000 MOTS À TRADUIRE

En février 2021, Marie Gaille s'est adressée à l'Inist pour traduire en anglais ce rapport comportant 57 000 mots ; une quantité bien supérieure aux travaux habituellement traités à l'Institut.

Au total, 36 jours ouvrables de travail ont été estimés. En raison de cette importante charge de travail, Richard Dickinson, traducteur à l'Inist en charge de ce projet, a proposé de diviser le rapport en trois parties, l'une étant traduite à l'Inist par lui-même (15 000 mots !) et les autres par deux traducteurs freelance.

UN TRAVAIL D'HARMONISATION

Après un travail de traduction qui a duré plusieurs semaines, Richard a rassemblé les trois parties du rapport pour harmoniser les textes. En effet, il fallait appliquer le même format à certains éléments du texte (ex. : les dates, les fonctions, etc.). Il a aussi veillé à homogénéiser les termes du glossaire et leur signification.

Outre les bons échanges avec les coordinateurs scientifiques, le traducteur n'a pas manqué de souligner l'intérêt de ce rapport qui met en lumière les travaux des différentes disciplines SHS à travers un regard mondial sur la pandémie.



Le livre *The Covid-19 pandemic : What the HSS say?* a été publié en version imprimée par Alliance Athéna.



DE L'ONTOLOGIE ONTOTM À TM TOOLS EXPLORER

Fruit d'une collaboration entre les services "Text et Data Mining" et "Ingénierie terminologique" de l'Inist, TM Tools Explorer est la version bêta d'une application d'exploration des outils de fouille de textes construite à partir d'OntoTM et de Lodex.

Au cours du projet Visa TM dont l'objectif principal était de décrire une infrastructure capable de rendre la fouille de texte accessible à tous, une liste de trois cents outils spécialisés dans le traitement automatique du langage et la fouille de textes avait été établie. Elle présentait un premier panorama des applications disponibles.

Cependant, l'équipe chargée du projet a décidé d'aller plus loin en transformant cette liste en ontologie avec l'objectif de proposer une représentation enrichie, formelle, explicite et manipulable par une machine de ces outils.

LA NAISSANCE D'ONTOTM

Pour construire OntoTM (Ontology of Text Mining), la liste d'outils a été transformée en fichier OWL (Web Ontology Language). Le langage OWL permet de représenter les connaissances dans les ontologies informatiques. Dans un deuxième temps, un ensemble de classes et de propriétés provenant d'ontologies existantes ont été sélectionnées afin d'aider à la modélisation du domaine.

La liste d'outils OWL et les ontologies externes ont ensuite été fusionnées au sein d'une même ontologie : OntoTM.

PASSAGE PAR LODEX

Produit par l'Inist, l'outil Lodex (Linked Open Data EXperiment) permet de publier des jeux de données (.csv, .tsv, .xml, .json, etc.) dans des formats du web sémantique (JSON-LD, N-Quads) et d'exposer ces données dans un rapport web dynamique sous la forme d'un tableau de bord avec des graphiques.

Aussi, un jeu de données a été extrait d'OntoTM en .csv. Avant son versement dans Lodex, il a bénéficié d'enrichissements :

- un alignement des outils avec ceux répertoriés dans Wikidata afin d'enrichir le fichier avec les liens vers les notices Wikidata trouvées ;
- une recherche automatisée de références bibliographiques liées aux outils à la fois dans la base Istex et dans la base Conditor.

UNE VUE DES FACETTES DANS TM TOOLS EXPLORER

Ensuite, le chargement du fichier .csv dans Lodex en vue de son exploitation a mis en avant un certain nombre de facettes dans TM Tools Explorer : tâche(s) effectuée(s) par l'outil, pays de production, langue(s) traitée(s), langage(s) de programmation, format(s) d'entrée, système(s) d'exploitation, interface(s) utilisateurs.

Les croisements de différentes facettes sont susceptibles de constituer une aide à la recherche pour des outils spécifiques.



UNE OPTIQUE COLLABORATIVE

Les données étant renseignées majoritairement en anglais, et afin de proposer plus rapidement une version utilisable, l'équipe a fait le choix de produire cette version bêta en anglais. Une version en français devrait suivre assez rapidement néanmoins.

Afin d'enrichir la base de données, l'équipe compte sur une appropriation de l'outil par les utilisateurs et les communautés de recherche impliquées dans des travaux sur le Traitement Automatique des Langues ainsi que la fouille de textes.

C'est donc clairement dans une optique collaborative et ouverte que cette première proposition a été construite au service d'une facilitation d'appropriation par tous des outils de fouille de textes, la plus simple possible.

Toutes les propositions d'amélioration sont les bienvenues.



À bientôt !

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

www.inist.fr

