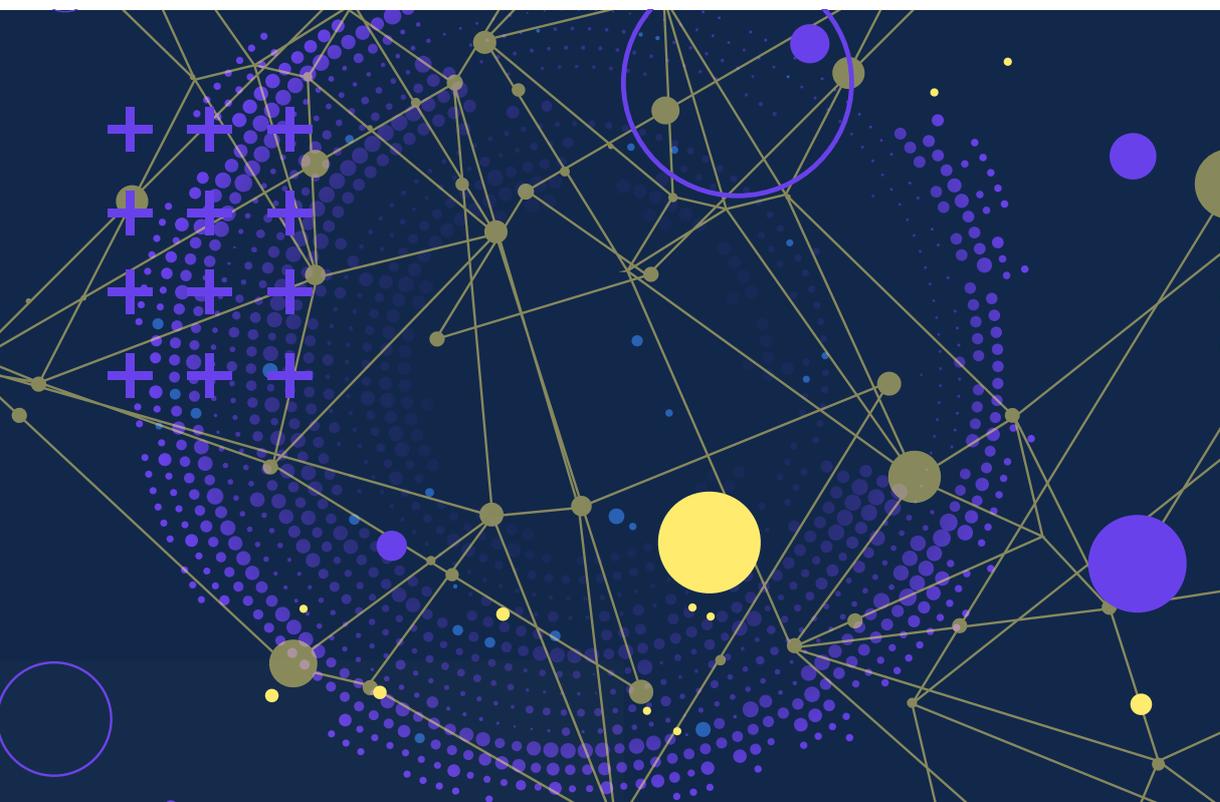


INIST

Institut de
l'information
scientifique et
technique



RAPPORT D'ACTIVITÉ
2023

cnrs



Sommaire

Édito	4
Portails & Plateformes	8
Négociations & Acquisitions	10
Mesure des usages	11
HAL-CNRS	12
Édition numérique	13
OPIDoR	16
DoRANum - Formation	18
Exposition des données	20
Textes & Corpus - Istex	24
Ressources langagières multilingues	26
Text & Data Mining (TDM)	28
Analyses bibliométriques	29
Communication	31
Web	32
Veille	32
Systèmes d'information	33
Secrétariat général	33
À votre rencontre	34
Un peu de détente...	36



Édito

UN PROJET RENOUVELÉ

Au premier janvier 2024, l'Inist a été renouvelé avec un projet sur 5 ans prévoyant d'une part de déployer son accompagnement vers la science ouverte pour les 3 types de productions scientifiques que sont les publications, les données et enfin les codes et logiciels, et d'autre part de faire de la fouille de texte une réalité pour les chercheurs. Ce projet s'appuie sur nos réalisations les plus récentes décrites dans ce document.

L'ACCÈS À TOUT TYPE DE PRODUCTIONS SCIENTIFIQUES FACILITÉ

L'outil **Click & Read** qui permet un accès rapide aux textes des publications poursuit son extension aux ressources électroniques des partenaires ; aujourd'hui 67 réservoirs documentaires sont accessibles montrant ainsi une bonne appropriation de la part de nos collègues des différents établissements. Tant par cet outil que par le portail documentaire **BibCnrs**, l'Inist favorise l'accès aux publications disponibles en Open Access et expérimente également l'accès aux données de recherche.

Initié en 2019, le portail **HAL-CNRS** a atteint en 2023 le cap des 2 millions de références. L'équipe en charge du portail s'est organisée pour en assurer la modération et la curation tout au long de l'année, y compris pendant les périodes de dépôts intensifiées par les campagnes d'évaluation des chercheurs.

PREMIÈRE ANNÉE DE FONCTIONNEMENT DE L'ÉCOSYSTÈME RECHERCHE DATA GOUV

Les **centres de ressources OPIDoR et DoRANum** ont trouvé un mode de fonctionnement collaboratif avec les ateliers de la donnée permettant de mutualiser sur nos plateformes les outils et ressources pédagogiques indispensables à la mise en œuvre de leurs missions d'accompagnement des communautés scientifiques.

Dans la même logique, l'Inist est fortement impliqué dans l'administration et la curation de l'espace générique national ainsi que de l'espace institutionnel CNRS Research Data. Ces entrepôts, qui sont complémentaires des entrepôts thématiques nationaux ou internationaux disponibles, permettent de disposer d'une solution de confiance répondant à chaque besoin. Mis en place en 2022, le **consortium DataCite France** organise son activité autour de 6 groupes de travail portant aussi bien sur la qualité des métadonnées associées aux identifiants pérennes que sur le modèle économique de ce service nécessaire pour assurer l'accès aux jeux de données.

DES ATTENTES FORTES VIS-À-VIS DE LA FOUILLE DE TEXTES

Les équipes poursuivent l'alimentation de la plateforme **Istex** pour arriver bientôt à 28 millions de documents qui sont prétraités et enrichis pour un usage en fouille de texte. Afin de faciliter ce dernier usage, les équipes packagent les outils dans des **web-services**, 38 sont d'ores et déjà accessibles sur la plateforme et peuvent être utilisés en autonomie soit sur un corpus extrait des ressources Istex soit sur un corpus préparé par l'utilisateur.

Une bonne façon de mettre en valeur les ressources documentaires disponibles sur la plateforme est d'en extraire un corpus, de l'enrichir en appliquant des méthodes de fouille de texte et de le visualiser avec l'outil **Lodex** qui est maintenant arrivé à maturité. Les équipes déploient un ensemble de formations afin d'accompagner les utilisateurs dans les différentes phases d'interrogation de la plateforme, de traitement et de visualisation des corpus ainsi constitués. L'Inist est le premier utilisateur de ces avancées en fouille de texte. Ainsi pour réaliser le travail de titan que demandent le recensement et l'analyse des frais engagés par le CNRS pour publier les travaux scientifiques en Open Access selon le mode gold, l'équipe fait appel aux collègues pour réaliser une première passe automatisée dans le dépouillement des quelque 3 000 dépenses.

Ce rapport d'activité décrit précisément l'ensemble de ces développements, qui pour la plupart sont réalisés en collaboration avec nos partenaires nationaux et s'inscrivent dans la feuille de route science ouverte du CNRS pilotée par la Direction des données ouvertes de recherche (DDOR). Elles ont également leur prolongement dans une implication dans le groupe de travail de l'initiative **EOSC** « PID Policy&implementation » et celui d'« EOSC FAIR impact », ou dans le projet **Skills4EOSC**. Cette implication ainsi que la participation aux projets portés par EOSC seront à renforcer.

Bonne lecture.

Claire FRANÇOIS
Directrice de l'Inist



++++
**ACCÉDER
AUX PRODUITS
DE LA RECHERCHE**

L'accès aux produits de la recherche est important pour garantir la transparence et l'intégrité de la recherche.

Il permet aux chercheurs, aux étudiants et aux professionnels de rester informés des avancées récentes dans leur domaine, facilitant ainsi l'innovation.

De plus, un accès ouvert favorise l'équité dans la diffusion des connaissances, permettant à tous, indépendamment de leur affiliation institutionnelle ou de leur situation géographique, de bénéficier des avancées scientifiques et de contribuer au développement de la science.

Pour cela, l'Inist met à votre disposition plusieurs réservoirs d'information incontournables ainsi que des outils pour être au plus près des attentes et des besoins.

Portails & Plateformes

Le service dédié aux Portails & Plateformes gère plusieurs sources d'information proposées aux usagers telles que BibCnrs, une bibliothèque numérique couvrant plusieurs disciplines, et Panist, une archive de publications récentes.

Lancée en 2021, l'extension Click & Read complète cette offre en facilitant la recherche et l'accès au texte intégral à partir de n'importe quel site, offrant un accès direct à des millions d'articles en libre accès.



L'enquête « Vos pratiques documentaires et BibCnrs dans un contexte de science ouverte » a été publiée du 9 juin au 15 septembre auprès de plus de 41 000 usagers. Elle a recueilli 4 911 réponses (11,8 %), ce qui représente une participation active (merci aux utilisateurs !).

Il est apparu que 75 % des répondants utilisent régulièrement BibCnrs et que 62 % accèdent aux documents via l'onglet « Article », 37 % via « Base de données » et 30 % via « Revue/Ouvrage ».

Malgré une note globale de 8 sur 10 qui confirme l'évolution positive de BibCnrs durant ces dernières années, quelques souhaits d'évolutions ont émergé : ergonomie plus moderne, accès direct aux ressources phares en moins de clics, ajout d'une recherche avancée...

L'équipe BibCnrs a étudié des solutions pour répondre au mieux à ces demandes dans les prestations de développements 2024-2025.

CLICK & READ

Tous les établissements sont éligibles quel que soit leur système de gestion de bibliothèque.

» Ajout de 25 nouvelles sources consultables depuis Click & Read issues des établissements de l'ESR suivants :

Université de Lille, Université de Lorraine, Université Paris-Est Créteil, Université Grenoble Alpes, Université Paris Cité, Université de Toulon, La Rochelle Université, Université de Strasbourg, Université Paris 8 Vincennes Saint-Denis, École Centrale de Lyon, Université Lumière Lyon 2, Université de Technologie de Compiègne, Arts et Métiers Sciences et Technologies, IFP Énergies nouvelles, INSA Toulouse, INRAE, IMT MINES ALBI, École Nationale de l'Aviation Civile, Avignon Université, Humathèque Condorcet, INP Toulouse, Université Rennes 2, VetAgro Sup, Université de Franche-Comté, Université Paris Dauphine-PSL.

Ces sources complètent celles des partenaires (universités, organismes de recherche dont le CNRS, grandes écoles...) déjà disponibles. Soit un total de **67 réservoirs documentaires** paramétrables par les chercheurs selon leurs droits d'accès.

L'extension repère ensuite, pour toute navigation sur Internet, les articles scientifiques accessibles et affiche le bouton cliquable  qui permet d'ouvrir le document.

» Internationalisation des pages de Click & Read qui s'affichent en français ou anglais selon la langue du navigateur.

» Ajout d'un connecteur pour les établissements clients du résolveur de liens ExLibris SFX.

» Mise en place d'un mode d'utilisation des API optimisé pour un service **plus éco-responsable**.

» Ajout de la compatibilité avec le **navigateur web Opéra**.

+ de 24 000 installations depuis mars 2021

+ de 15 000 utilisateurs quotidiens

+ de 3 000 consultations d'articles par jour

BIBCNRs

BibCnrs a évolué avec l'objectif de favoriser la science ouverte :

» Utilisation du **miroir interne** du réservoir libre **Unpaywall** pour proposer plus de ressources libres dans les résultats de recherche.

» Mise en avant des revues accessibles via BibCnrs publiées selon le **modèle ouvert Diamant** (publication gratuite pour les auteurs et consultation gratuite pour les lecteurs) par l'ajout d'un logo bleu en forme de Diamant. 

» Ajout d'un **onglet « Données de recherche »** permettant de consulter des jeux de données issus de Datacite (organisation mondiale à but non lucratif qui attribue notamment des DOI aux jeux de données internationaux). Ce service apporte un complément à la consultation des publications scientifiques et permet d'identifier de nouvelles connaissances.

» Réalisation de **20 webinaires** de présentations de ces nouveautés.



PANIST

Nouvelle dynamique pour PANIST

(Plateforme d'archivage national de l'information scientifique et technique).



Mis en place à l'initiative du ministère de l'Enseignement supérieur et de la Recherche, du consortium Couperin.org, et géré par l'Inist, PANIST est un service d'accès pérenne aux archives. Il offre aux établissements de l'ESR français ayants droit la garantie d'accéder aux contenus souscrits au terme d'un accord consortial avec un éditeur, et ce indépendamment des accès proposés en ligne par l'éditeur.

Cette plateforme propose l'accès aux collections Freedom, Bibliothèque Médicale Française et Cell Press de l'éditeur Elsevier (2002 à 2022) ainsi qu'à la collection SpringerLink de l'éditeur Springer (2015-2017).

Les processus de traitements des données et des droits d'accès ont été améliorés en 2023.

Courant 2023, 2,6 millions d'articles de 2019 à 2022 de l'éditeur Elsevier sont venus compléter les données existantes et portent le nombre de documents scientifiques disponibles via PANIST à plus de 10,1 millions.

Les ressources contenues dans PANIST peuvent être consultées, selon les droits d'accès des établissements, via l'extension web gratuite Click & Read qui facilite et centralise l'accès en un clic aux articles.



OPENBIBART

Plateforme de recherche ouverte conçue par l'Inist, OpenBibArt réunit le contenu de plusieurs ensembles de données : **le Répertoire d'art et d'archéologie (RAA)**, **le Répertoire international de la littérature d'art (RILA)** et la **Bibliography of the History of Art (BHA)**.

» **23 160** vues en 2023 par **3 451** visiteurs

PASCAL & FRANCIS

Le site OpenPascalFrancis est une archive des bases de données bibliographiques PASCAL et FRANCIS en **sciences exactes, humaines et sociales**, produites par l'Inist entre 1972 et 2015.

» **2 184 749** vues en 2023 par **1,4** million de visiteurs

INIS

L'Inist héberge et actualise, pour le CEA, la base de données bibliographiques consacrées aux publications mondiales en **sciences nucléaires** : l'INIS (International Nuclear Information System). Elle compte plus de 4 millions de références et s'enrichit chaque mois.

» **5 369** vues en 2023 par **1 132** visiteurs



— / Négociations & Acquisitions / —

Le service Négociations & Acquisitions gère toutes les acquisitions de ressources thématiques pour le portail BibCnrs. Annuellement, les négociateurs collaborent avec une cinquantaine d'éditeurs et de distributeurs internationaux pour le renouvellement des abonnements existants et la mise en œuvre de nouvelles souscriptions.



Des techniques de text and data mining (TDM) pour l'Enquête APC du CNRS

Mandaté par la DDOR depuis 2019 pour effectuer l'Enquête APC (Article Processing Charges) du CNRS, à savoir le recensement des frais engagés pour des prestations externes d'édition scientifique par les structures de recherche sur du budget CNRS, le service fait désormais appel aux compétences du service TDM de l'Inist pour automatiser une partie des opérations.

Cette collaboration permet d'optimiser l'étape de dépouillement, qui porte chaque année sur un panel d'environ 3000 transactions.

Une fois normalisées, ces données viennent ensuite alimenter la base de données internationale OpenAPC, par l'intermédiaire du consortium Couperin.org, et servent d'outil d'aide à la décision pour la DDOR.

Ce qu'il s'est passé en documentation électronique...

- » Acquisition pour INSHS de l'**accès à un bouquet d'une cinquantaine de revues de l'éditeur Brill** consacrées principalement aux études moyen-orientales et orientales, dont les 41 revues de la collection *Middle-East and Islamic Studies*.
- » Acquisition pour INSB et INEE de l'**accès aux bases de données BioCyc** (sur la thématique des voies métaboliques) et **RepBase** (sur la thématique des éléments d'ADN répétitif).
- » Participation active à **plusieurs négociations** au sein du consortium Couperin.org.
- » **11 ressources proposées en essai gratuit** au cours de l'année dont *RetroNews*, les ebooks *GeoScienceWorld*, *JoVE* et *IEEE*. L'enquête menée auprès des utilisateurs suite à cette campagne de découverte a suscité près de 400 réponses.



Soutien à la science ouverte

L'Inist soutient les initiatives SPARC Europe, DOAJ et Mir@bel à hauteur d'environ 11 000 € au total pour 3 ans via un dispositif de soutien à des infrastructures œuvrant pour la science ouverte mis en place par le consortium Couperin.org.

Mesure des usages

L'équipe « ezTEAM - mesure des usages » accompagne les établissements de l'ESR et des consortia étrangers dans leur politique documentaire en développant et mettant en place des outils relatifs à la quantification de l'utilisation des ressources électroniques.



L'équipe a vu une montée en charge d'ezCOUNTER avec la mise en œuvre, via la partie ressources numériques, de l'enquête ESGBU du ministère.

Au total, 60 structures documentaires de l'ESR sur 137 actives en 2022 ont répondu à l'enquête. Les répondants sont à 75 % en charge des ressources numériques, les autres responsables d'enquête ESGBU. On note une progression de 11 à 13 points du taux de réponses à cette enquête par rapport à l'année précédente.

Recommandé par le ministère !

Pour la première fois, l'outil ezCOUNTER figurait, via ezMESURE, dans les recommandations de l'enquête statistique sur les bibliothèques universitaires et les services documentaires de l'enseignement supérieur et de la recherche (ESGBU).

Cette dernière a pour objectif de collecter les données d'activité des bibliothèques des établissements ESR favorisant ainsi un suivi régulier par l'État de l'activité des bibliothèques de l'enseignement supérieur.

Le ministère de l'Enseignement supérieur et de la Recherche (MESR) a indiqué que les établissements utilisant le dispositif ezMESURE pouvaient accéder en février 2023 au tableau de bord ESGBU dédié au groupe de questions ressources numériques Counter. Ce module intégré à ezMESURE, développé en partenariat avec l'Inist, le Consortium Couperin.org et le MESR, permet de consulter et d'agréger les données annuelles COUNTER5 des éditeurs pour générer des tableaux de bord personnalisés mettant en avant les 7 indicateurs du volet « usages numériques » de l'enquête.

Parmi les actions réalisées en 2023...

- » **Moissonnage courant** (2022 et 2023) des données d'usage éditeurs de plus de 60 établissements de l'ESR grâce au module ezCOUNTER.
- » **Moissonnage rétrospectif** (2019-) des données d'usage éditeurs de plus de 60 établissements de l'ESR pour les négociateurs des licences nationales (Elsevier, Springer-Nature, Wiley...).
- » Mise à disposition d'une **version en préproduction d'ezMESURE** à des établissements bêta-testeurs pour valider le passage à la gestion des établissements complexes (regroupements universitaires par exemple).
- » Maintenance du **miroir ezUnpaywall** (mises à jour quotidiennes) et préparation d'une version incluant l'historique de la disponibilité en Open Access.



Des **rendez-vous** ont été organisés tout au long de l'année avec les utilisateurs des outils et les négociateurs, en ligne, par sessions zoom, animés par Frédéric Truong et Thomas Porquet (Couperin) :

- 16 supports mutualisés
- 14 webinaires en ligne
- 7 revues de sprint



L'équipe a annoncé en décembre le renommage de la suite logicielle ezPAARSE, ezMESURE, ezCOUNTER en **ReadMETRICS** avec un lancement prévu en 2024.

HAL-CNRS

Le portail HAL-CNRS recense toute la production scientifique du CNRS. Cet outil institutionnel sert de plateforme de mise en avant des publications scientifiques des chercheurs du CNRS. Au sein de l'Inist, l'équipe HAL-CNRS est chargée de l'administration du portail, de la modération des dépôts, de l'animation et de la valorisation du portail, ainsi que de l'accompagnement des utilisateurs.



La nouvelle organisation de l'équipe autour des missions clés du portail HAL-CNRS a porté ses fruits : une meilleure visibilité, plus de dépôts en open access (OA), qualité des dépôts optimale grâce à des actions de modération/curation et d'accompagnement renforcées. 2023 marque aussi un autre cap : celui des 2 millions de références dans le portail HAL-CNRS !

Poursuite du soutien à l'objectif 100 % OA du CNRS via le portail HAL-CNRS :

- » **+3 points de taux d'OA** pour les articles 2022 et **+8 points** pour les articles 2023 grâce aux différentes actions de modération/curation (ajout de DOI, d'affiliation, etc.) et intervention dans les référentiels d'AuréHAL.
- » Modération et mise en ligne de **57 308 dépôts**, dont ceux en lien avec le CRAC et le RIBAC.
- » **+ 60 % de trafic** sur le portail HAL-CNRS (changement d'URL et meilleure visibilité).
- » **Ouverture de collections tests** pour des instituts du CNRS avec d'importantes actions de curation.

Des activités en hausse :

- » 70 heures consacrées à **l'assistance/accompagnement** des chercheurs CNRS via le nouveau service mis en place : hal-cnrs@inist.fr
- » **Construction d'ateliers de formation** à la demande d'unités CNRS : 13 ateliers organisés et animés pour 2 UMR soit 20 heures de formation et 65 chercheur(ses) sensibilisés.
- » **Interventions** auprès de personnels en IST/communauté HAL : partage de bonnes pratiques, retours d'expérience, assistance sur la gestion de collection, etc.

HAL CNRS
accueille l'ensemble de la
production scientifique du CNRS

LE PORTAIL INSTITUTIONNEL HAL-CNRS, C'EST :

- la vitrine de la production scientifique du CNRS.
- l'archive ouverte pour les dépôts des publications scientifiques du CNRS.

Une ÉQUIPE pour :

- ✓ la gestion/administration de la plateforme.
- ✓ la modération des dépôts.
- ✓ l'animation, la valorisation du portail, et l'accompagnement des utilisateurs.

Des SERVICES pour :

- ✓ l'ouverture de collections et la formation à leur gestion.
- ✓ l'accompagnement et l'assistance au dépôt.
- ✓ des ateliers à la carte.

INFORMATIONS ET CONTACT :

inist CNRS HAL

cnrs.hal.science
hal-cnrs@inist.fr

Poster présenté lors de la
6^e édition des **Journées CasuHAL**
à Toulouse du 14 au 16 juin 2023.

Édition numérique

L'équipe de l'Inist réalise un travail de secrétariat d'édition en partenariat avec le Muséum national d'histoire naturelle (MNHN).



En lien avec le MNHN et l'équipe Métopes, l'Inist a opéré des tests sur trois collections pilotes (Faune et flore tropicales, Stratotypes et Patrimoines naturels) dans le cadre du projet MetoSTeM.

L'objectif de **MetoSTeM**, projet porté par le MNHN, est d'adapter la chaîne Métopes et l'outil Metotaxa aux monographies dédiées aux STM (Science, Techniques et Médecine). Cette évolution permettra de structurer finement et de disséminer des contenus et des données issus de la recherche en biodiversité et écologie au sens large.

Metostem participera activement à la science ouverte, en conformité avec les principes FAIR : données en open access (interopérables), identifiées et liées, traçabilité des publications (DOI).

L'adaptation de la chaîne de production des périodiques scientifiques est en cours de réflexion (typographies française et anglaise, annotation des noms vernaculaires, structuration du matériel examiné, liaison de plusieurs fichiers xml les uns aux autres, etc.), ainsi que le passage des périodiques scientifiques en CC-BY.



MetoSTeM est un
projet lauréat FNSO 2021

PARMI LES ACTIONS RÉALISÉES EN 2023...

Pôle monographies

- » 3 parutions papier et 5 parutions numériques d'ouvrages, pour les collections *Natures en sociétés* et *Archives* du MNHN sur SciencePress et OpenEdition Books.

Exemple de réalisation :

Le Roux Géraldine 2022 — *L'art des ghostnets: approche anthropologique et esthétique des filets-fantômes*. Paris: Muséum national d'Histoire naturelle, 447 p. (*Natures en Sociétés*; 6).



Pôle périodiques

- » 13 articles *Anthropozoologica*, 7 articles *Comptes rendus Palevol* et 1 article *Naturae* publiés en flux continu.
- » Mise en ligne des archives *Comptes rendus Palevol* 2002-2019 sur le site Sciencepress (122 volumes disponibles intégralement en open access).
- » 23 numéros d'*Arabesques* (ABES) publiés sur Prairial dans le cadre du chantier de rétroconversion de la revue.
- » Poursuite de la collaboration Inist/Calenda : calendrier multilingue en sciences humaines et sociales en ligne et en libre accès.



**GÉRER ET VALORISER
LES DONNÉES
DE LA RECHERCHE**



La gestion et la valorisation des données de la recherche sont fondamentales pour la science ouverte.

Elles assurent l'intégrité, la qualité et la pérennité des données de la recherche tout en favorisant leur partage et leur réutilisation.

L'inist propose des services et des outils spécialisés pour vous aider dans la gestion, le partage et la diffusion de vos données de recherche.

Notre mission est de rendre l'accès aux données plus facile, d'assurer leur interopérabilité et de promouvoir leur réutilisation.

Outils et services pour accompagner la gestion des données de la recherche



OPIDoR

Le portail OPIDoR, initié par l'Inist et géré par l'équipe dédiée, offre à la communauté de l'Enseignement supérieur et de la recherche une gamme d'outils et de services pour faciliter l'application des principes FAIR, rendant ainsi les données Faciles à trouver, Accessibles, Interopérables et Réutilisables. Ce portail s'inscrit dans le cadre du Plan national pour la science ouverte orchestré par le Comité pour la science ouverte (CoSO).



Lancement officiel de CNRS Données de la Recherche : le Catalogue des entrepôts et des services dédiés aux données de la recherche.

Les services français qui contribuent à la gestion, au partage et à la réutilisation des données de la recherche tout au long de leur cycle de vie sont recensés dans Cat OPIDoR qui affiche désormais le catalogue CNRS des entrepôts et des services.

DMP OPIDoR



DMP OPIDoR est un **outil d'aide à la création en ligne de DMP** qui permet d'anticiper la gestion des données dans le cadre de projets de recherche. Parmi les actions réalisées en 2023 :

- » Élaboration et publication du modèle « **PGD CNRS** » en français (20/11/2023) et « **DMP CNRS** » en anglais (27/11/2023).
- » Développements informatiques pour la mise en production d'une **nouvelle version de DMP OPIDoR (V4)** prévue en 2024.
- » Collaboration avec les **ateliers de la donnée** pour l'élaboration d'un modèle de plan de gestion de logiciel.
- » **Constitution d'un corpus** de PGD disciplinaires dans le cadre d'un groupe de travail RDA France.
- » Participation au groupe de travail du **Comité pour la science ouverte** sur les plans de gestion de données.

CAT OPIDoR



Cat OPIDoR est un **catalogue** qui recense et décrit les services français contribuant à une bonne gestion, conservation et réutilisation des données scientifiques. Parmi les actions réalisées en 2023 :

- » **Étude de faisabilité** Cat OPIDoR comme base de catalogue de services à intégrer dans EOSC.
- » Participation à un **groupe de travail INRAE** sur le signalement des services numériques : ce groupe est en charge de décrire les services numériques mutualisés disponibles à INRAE pour pouvoir y faire mention par un lien unique dans les plans de gestion de données. Des métadonnées de Cat OPIDoR ont été adaptées et un formulaire spécifique a été mis en place pour répondre à ce besoin (en phase d'expérimentation).

PID OPIDOR - CONSORTIUM DATA CITE FRANCE

Dirigé par l'Inist, le Consortium DataCite France a tenu son assemblée générale le 8 décembre 2023, marquant une étape clé dans la définition de ses objectifs pour l'année à venir.

L'événement a permis de présenter la composition des cinq groupes de travail formés suite à un appel à participation. Ces groupes se concentrent sur les métadonnées, les interactions avec ORCID et d'autres identifiants persistants, le modèle économique, la sensibilisation et la formation, ainsi que la communication.

Chacun de ces groupes a pour mission spécifique de contribuer à la promotion et à la ges-

tion des données de la recherche. Ils proposeront prochainement une feuille de route et un calendrier détaillé pour leurs actions à venir lors de réunions prévues en 2024.

Le Consortium DataCite France s'engage ainsi à renforcer son rôle en tant qu'acteur majeur dans la gestion des données de la recherche.

La première réunion du comité exécutif s'est tenue le 14 avril 2023 au cours de laquelle **Geneviève Michaud** et **Pascal l'Aventurier** ont été nommés par consensus co-présidents.

L'Inist est impliqué dans des groupes de travail de l'initiative **European Open Science Cloud (EOSC)** de la Commission européenne.

Le groupe **EOSC-A Task Force PID Policy & Implementation** a formulé les recommandations suivantes :

- Cartographie des activités actuelles liées au PID dans le contexte de l'EOSC,
- Analyse des critères d'évaluation de la conformité à la politique PID de l'EOSC des services d'identifiants pérennes,
- Harmonisation des vocabulaires liés au PID.

Le groupe **EOSC FAIR Impact** a animé des workshops sur les identifiants pérennes et les bonnes pratiques :

- 2^e atelier sur la force de synchronisation - novembre 2023.
- Implémentations PID conformes à l'EOSC - Lignes directrices pratiques pour la mise en œuvre des meilleures pratiques, novembre 2023.
- Meilleures pratiques PID pour la citation de données complexes, les artefacts sémantiques et les services associés - mai 2023.



Centre de ressources
outils communs aux ateliers
 Recherche Data Gouv

Porteur du **Centre de ressources OPIDoR de Recherche Data Gouv**, l'Inist a poursuivi ses actions :

- » **Accompagnement** (conseils, curation, amélioration des métadonnées) au dépôt de jeux de données dans l'entrepôt générique et dans l'entrepôt institutionnel CNRS Research Data.
- » Participation au **second séminaire** des dispositifs d'accompagnement de Recherche Data Gouv organisé par l'Université de Reims et l'Atelier de la Donnée DAT@URCA le 14 décembre. L'équipe OPIDoR y a animé deux sessions afin de présenter DMP OPIDoR, d'échanger et de co-construire avec les ateliers de la donnée. Le poster « **DMP OPIDoR, un compagnon pour FAIR** » a été présenté à cette occasion.



DONNÉES DE LA RECHERCHE APPRENTISSAGE NUMÉRIQUE

DES RESSOURCES POUR ACCOMPAGNER LA COMMUNAUTÉ SCIENTIFIQUE DANS LA GESTION ET LE PARTAGE DE LEURS DONNÉES

DoRANum - Formation

Plateforme de formation à distance, DoRANum permet de se former sur la gestion et le partage des données de la recherche.



L'offre de formations disciplinaires a été développée pour accompagner les communautés scientifiques à la gestion des données dans leurs disciplines, notamment grâce aux résultats du projet DoRANum disciplinaire, soutenu par le Comité pour la science ouverte, et mené en partenariat avec le GIS Réseau Urfist. Des ressources pédagogiques ont été réalisées, entre autres, en Bio-informatique, Science Humaines et sociales, Économie-gestion, Humanités numériques et Linguistique.

Parmi les actions mises en œuvre en 2023 par l'équipe DoRANum :

- » Alimentation importante de DoRANum grâce à la réalisation de **21 nouvelles ressources pédagogiques génériques et 5 nouvelles ressources disciplinaires**.
- » Important travail de **mise à jour des ressources** pédagogiques DoRANum (43 ressources).
- » **27 formations en présentiel ou à distance** dispensées sur les sujets de l'ouverture et du partage des données et des réseaux sociaux scientifiques, pour un total de **+ de 910 participants**.
- » Démarrage d'un **programme de formation multimodal** sur la gestion des données de recherche pour les ingénieurs plateforme de CNRS Ingénierie.

SKILLS4EOSC

Le service formation DoRANum est impliqué dans le projet SKILLS4EOSC (2022-2025) avec la DDOR pour le partenaire CNRS.

Financé par la Commission Européenne, ce projet a pour objectif de **faire progresser les compétences en science ouverte** en unifiant le paysage européen actuel de la formation (définition de profils métier en science ouverte, mutualisation de formations, mise en place de réseaux de centres de compétences).



PARMI LES WEBINAIRES PROPOSÉS...

- » **Les exigences de la Commission européenne sur l'ouverture des données** pour le réseau des financeurs de la recherche, organisé par l'ANR (21 mars).
- » **Présentation des enjeux de l'ouverture de données de la recherche** pour les lauréats de Winning Normandy, organisé par la DR CNRS Paris-Normandie et le service enseignement supérieur et recherche de la région Normandie (26 juin).
- » **Plan de gestion de données, bonnes pratiques de rédaction** organisé par IFSEM CNRS pour les chercheurs et membres de SPV de la région parisienne (21 septembre).
- » **PGD bonnes pratiques de rédaction et outils** lors des journées RDA France (10 octobre).
- » **Présentation du Centre de ressources DoRANum Recherche Data Gouv** aux ateliers de la donnée (21 novembre).



Centre de ressources
supports pédagogiques, e-formations
Recherche Data Gouv

Porteur du **Centre de ressources DoRANum de Recherche Data Gouv**, l'Inist a contribué à son développement et à ses interactions avec les membres de l'écosystème Recherche Data Gouv :

- » Définition et mise en place des **workflows de travail** pour l'ingénierie pédagogique.
- » Participation aux **groupes de travail** des Ateliers de la donnée.
- » Étude pour la **mise en place d'un entrepôt** pour le dépôt et le partage de ressources pédagogiques.
- » **Ateliers collaboratifs** sur l'évolution de la plateforme DoRANum durant le séminaire des dispositifs d'accompagnement de l'écosystème Recherche Data Gouv, en décembre 2023.
- » Réalisation de **tutoriels interactifs sur l'entrepôt Recherche Data Gouv** en collaboration avec le Centre de ressources Entrepôt-catalogue. Ces tutoriels viennent compléter le dispositif d'accompagnement (guides, classes virtuelles) déjà mis en place par le Centre de ressources Entrepôt-catalogue. Ils ont pour objectif d'accompagner pas à pas, et à leur rythme, les déposants lors du dépôt de données dans l'entrepôt.
- » Réalisation d'une **ressource sur les ZRR** (zones à régime restrictif) en collaboration avec l'atelier de la donnée ADOC Lorraine.
- » **Participation au second séminaire** des dispositifs d'accompagnement de Recherche Data Gouv organisé par l'Université de Reims et l'Atelier de la Donnée DAT@URCA le 14 décembre. L'équipe Formation-DoRANum a pu échanger et répondre aux questions relatives à la réalisation de ressources pédagogiques avec les ateliers de la donnée et les centres de ressources. Deux posters y ont été présentés : « **Processus de création d'une ressource pédagogique** » et « **Processus d'adaptation d'une ressource pédagogique** ».



Exposition des données

Ce service de l'Inist a pour mission de valoriser et partager les données de la recherche par la création de bibliothèques numériques avec le logiciel libre OMEKA.



Produit par l'Inist en 2022, le site Chronospedia (<https://chronospedia.com>) a bénéficié d'une refonte en 2023 avec la mise en oeuvre d'une nouvelle plateforme de valorisation de données patrimoniales dans le domaine de l'horlogerie. Ce site est le fruit d'une collaboration avec une grande diversité d'acteurs issus du milieu de l'enseignement supérieur et de la recherche (CNRS, Universités), de l'industrie (Dassault Systèmes), de l'artisanat d'art, du monde associatif et de la culture.

Parmi les actions réalisées en 2023...

- » Création, ou évolution avec OMEKA-S, de **19 nouvelles plateformes** pour la valorisation de 34 projets de recherche.
- » Collaboration avec **22 laboratoires de recherche, 4 MSH et un institut à l'étranger**, tous étant des UMR ou UAR CNRS rattachées à 8 universités, 2 grandes écoles et au Muséum national d'histoire naturelle.
- » Accompagnement des laboratoires de recherche pour la **création et la valorisation de jeux de métadonnées** dédiés à la mise en ligne de contenus sous OMEKA ou concernant des données de recherche déposées dans des entrepôts, en application des principes FAIR.
- » **Dématérialisation de 62 000 pages ou images** de documents patrimoniaux pour alimenter les bibliothèques numériques et déposées dans des entrepôts de données tels que Nakala ou des serveurs au sein des organismes de recherche.
- » **Déploiement d'un serveur et de connecteurs OMEKA basés sur la technologie IIIF** (International Image Interoperability Framework) afin de mettre en place des technologies innovantes pour les documents numériques nécessitant une haute définition permettant d'éviter la duplication des données. L'empreinte numérique de données très volumineuses est ainsi fortement réduite, tout en conservant une qualité optimale des images.

- » Participation aux **groupes de travail** des consortiums d'HumaNum : Mémoires des Archéologues et des sites archéologiques plus (MASA+) et 3D SHS (CST3D).



Poster présenté aux **JABES**,
23-24 mai 2023, Montpellier.

EXEMPLES DE SITES MIS EN PRODUCTION EN 2023...



CORPUS VISUELS

Le service Exposition des données de l'Inist a travaillé avec le laboratoire InVisu sur la réalisation d'un site dédié aux expositions virtuelles du laboratoire.

<https://corpus-visuels-invisu.inist.fr>

DHAF

Dante d'hier à aujourd'hui en France

Le projet DHAF a pour objet la réception en France de l'œuvre de Dante (1265-1321), systématiquement explorée sous toutes ses facettes depuis le Moyen Âge jusqu'à nos jours. Il est porté par un consortium de quatre universités : l'Université Sorbonne Nouvelle, l'Université Rennes 2, l'Université Côte d'Azur, l'Université Grenoble Alpes.

<https://dhaf.inist.fr>



ÉTUDES NORDIQUES

Ce portail rassemble des informations, des ressources documentaires et pédagogiques et des outils de recherche utiles aux études nordiques. Il est le fruit d'une collaboration entre le réseau Collex Europe du Nord et les chercheurs francophones de l'Association pour les études nordiques (APEN).

<https://etudes-nordiques.cnrs.fr>



Titouan Boudart, Konstantin Protassov.
La 3D au secours du patrimoine horloger. CHRONOSPEDIA : Encyclopédie virtuelle du savoir horloger. *JC3DSHS 2023, Les Journées du Consortium 3D SHS*, Nov 2023, Lyon, France.
(hal-04484358)

Anne-Lise Christmann, Pierre Willaime, Alain Zasadzinski. **Cenhtor, une plateforme pour les humanités numériques en Lorraine.** *Journée de lancement de Cenhtor*, MSH-Lorraine, Jan 2023, Nancy & Metz, France.
(hal-03983823)

++++
++++

ANALYSER & FOUILLER L'INFORMATION SCIENTIFIQUE

++++
++++
++++

++++
++++

Les outils d'analyse et de fouille de texte sont essentiels en recherche scientifique pour extraire des informations pertinentes, synthétiser des données, classifier des textes, faciliter la traduction multilingue et s'intégrer avec d'autres méthodes d'analyse de données.

Ils permettent aux chercheurs d'explorer efficacement de grandes quantités de données textuelles, favorisant ainsi la découverte de nouvelles connaissances et l'avancement de la science.

Pour répondre aux demandes de gestion scientifique et d'extraction de connaissances émises par les institutions, les entités fonctionnelles ou les acteurs de la recherche, l'Inist offre des services dédiés à l'analyse et à la fouille de corpus textuels scientifiques.

Textes & Corpus - Istex

Avec plus de 27 millions de publications scientifiques s'étalant sur 700 ans et couvrant toutes les disciplines, Istex est le plus vaste réservoir d'archives scientifiques normalisées au service de la recherche française. Sa particularité unique est d'être compatible techniquement et juridiquement avec la fouille et l'analyse de texte.



L'écosystème Istex a poursuivi son évolution afin de proposer des services à la recherche à partir des données Istex mais aussi à partir d'autres données. En effet, plus qu'une base documentaire, l'infrastructure Istex offre à la communauté de l'enseignement supérieur et de la recherche des outils dédiés à l'exploration et l'exploitation de documents à travers des services en ligne.

PARMI LES NOUVEAUTÉS 2023...

Corpus d'actualité publiés sur le site data.istex.fr :

- Nuit des étoiles : l'espace scruté par l'IA,
- 2023 Année internationale des millets.

Corpus spécialisés publiés sur le site data.istex.fr :

- 2 nouveaux corpus sur la mémoire,
- 2 corpus sur la traduction automatique,
- 2 corpus en paléoclimatologie.

Ressources numériques acquises auprès des éditeurs et chargées en 2023 :

- 1 nouveau bouquet chargé dans la base Open Edition - Journals, grâce au Plan de soutien à l'édition scientifique française,
- 1 année supplémentaire pour 2 bouquets déjà présents dans Istex :
 - Elsevier - Année 2013, grâce aux acquisitions pérennes des Licences Nationales ;
 - EDP Sciences - Année 2022, grâce au Plan de soutien à l'édition scientifique française.

LE CORPUS MACHINE TRANSLATION

Une exploration diachronique des (méta)données Istex

Mathilde Hugulin & Sabine Barreaux
(1) Inist-CNRS (UAR 76), 54 519 Vandœuvre-lès-Nancy, France - mathilde.hugulin@inist.fr, sabine.barreaux@inist.fr

Atelier sur l'Analyse et la Recherche de Textes Scientifiques, CORIA-TALN 2023

Contexte

Objetif : Explorer l'histoire de la traduction automatique

Ressource utilisée : Istex (APR, Istex OL)

Outils : Lodes, web services Inist

Motivations sociale, scientifique et politique : dynamique nationale et européenne autour de la traduction automatique (ANR 22-Ma705-0003 ; Fiorini et al., 2020 ; OPERAS ; Helsinki Initiative, 2019)

Constitution de corpus

Procédure itérative (de Sabatier & Barreaux, 2020)

1. Interrogation d'Istex (Syntaxe Lucene)
2. Détection du bruit dans Lodes
3. Révision de la requête

Le corpus Machine Translation contient **7 160 documents** en anglais et en français (54 millions de mots)

ISTEX

Initiative d'excellence en Information Scientifique et Technique

En chiffres	Non contenu
27 millions de publications scientifiques	Textes intégraux aux formats d'origine et standardisés (XML TEI)
51 langues	Textes nettoyés
41 corpus éditeurs	Métadonnées
700 ans de publications	Enrichissements : entités nommées (Unitex), termes (TextE), structuration du pdf (Gobid), domaines scientifiques (Nb, Multicat)

Linked Open Data Experiment

Lodes est un logiciel open source créé pour les besoins du projet Istex afin de valoriser ses données structurées (Gregoire et al., 2018). Il permet de concevoir des sites web offrant des interfaces pour explorer visuellement un jeu de données (CSV, JSON, etc.) à travers de tableaux de bord dynamiques présentant des indicateurs bibliométriques. Lodes offre également la possibilité d'utiliser des web services afin d'extraire les documents à l'aide de programmes d'analyse, de curation, d'annotation et d'indexation.

Résultats de l'exploration

Indicateurs bibliométriques

Exploitation des métadonnées

CoSib permet de détecter les références bibliographiques et de les structurer dans un format interprété par Lodes.

Utilisation de deux web services :

1. Le web service de découpage d'adresses retourne une adresse au format texte en tableau de champs
2. Le web service Pays et subdivisions normalise les graphiques des pays en exploitant le thésaurus Lodes

Diachronie de la traduction automatique

Exploitation du texte intégral

- Création d'une ressource terminologique permettant d'annoter le corpus
- Annotation du corpus grâce à une feuille de style XSL

Poster présenté à **CORIA-TALN 2023**
5 - 9 juin 2023 à Paris.

EN CHIFFRES...

Au 31/12/2023, la base Istex représente :

- 27 917 892 documents au total,
- 42 bouquets éditeurs acquis,
- 10 249 revues / 437 306 monographies,
- 539 975 documents chargés,
- 3,9 millions de documents disponibles par ailleurs en Open Access (14,2 %),
- 14,3 millions de textes intégraux nettoyés (51,5 %),
- 27,2 millions avec au moins un enrichissement (97,6 %).

Ces ressources sont valorisées dans data.istex au travers de :

- 20 corpus spécialisés pour la fouille de textes,
- 11 corpus d'actualité.

Quelques données d'usage de la plateforme en 2023 :

3 729 461 accès à la plateforme (hors Inist et hors robots) :

- parmi lesquels 62 % pour un usage TDM et 38 % pour un usage documentaire,
- par 286 établissements,
- pour 9,8 millions de documents téléchargés (via istex-DL, revue de sommaire, corpus data.istex, Google Scholar, etc.).

FORMATIONS

L'équipe Istex est intervenue dans des formations et a animé des ateliers sur la **constitution d'un corpus spécialisé** à partir des ressources Istex pour un usage en TDM :

- 10 février : à Villeurbanne, Enssib, dans le cadre du Diplôme de conservateur des bibliothèques. Intervenantes : Sabine Barreaux, Mathilde Huguin & Justine Revol.
- 10 mai : en distanciel, Université Paris 1 — Panthéon Sorbonne, dans le cadre du programme Cycl@doc destiné aux doctorants. Intervenantes : Mathilde Huguin et Fabienne Schmittheisler.



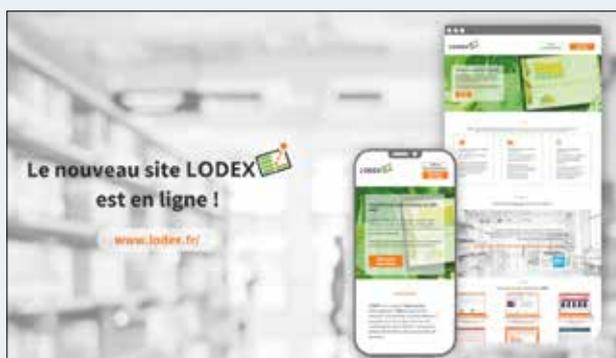
Mathilde Huguin, Sabine Barreaux.
**Le corpus « Machine Translation » :
une exploration diachronique des
(méta)données Istex.** *18^e Conférence en
Recherche d'Information et Applications -
16^e Rencontres Jeunes Chercheurs en RI -
30^e Conférence sur le Traitement Automatique
des Langues Naturelles - 25^e Rencontre des
Étudiants Chercheurs en Informatique pour le
Traitement Automatique des Langues, 2023,
Paris, France. pp. 54-59.
(hal-04131599)*

LODEX



Développé par l'Inist, Lodex est un logiciel open source de traitement et de valorisation des données à travers différents graphiques et pages entièrement configurables pour réaliser des bases de données ou tableaux de bord personnalisés.

- » Le **site vitrine de Lodex** a fait peau neuve avec une interface plus moderne et ergonomique. Quelques témoignages d'utilisateurs y sont accessibles, illustrant comment Lodex a contribué au partage de données pour des organisations telles que le CIRAD et les Hospices civils de Lyon.
- » **9 nouvelles instances** Lodex ont été ouvertes.





— Ressources langagières multilingues —

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. L'appui linguistique se concrétise également avec les compétences de l'équipe de traducteurs en matière de traductions, de relectures de documents scientifiques ou de supports de présentation.



7 nouveaux thésaurus ont été publiés dont :

- Un **thésaurus Mathématiques** : ce thésaurus bilingue (français et anglais) concerne le domaine des mathématiques. Il renferme actuellement plus de 2 600 concepts enrichis de synonymes et pourvus de définitions issues en grande majorité de Wikipédia.
- Un **thésaurus Astronomie** : réunissant un vocabulaire de 2 493 concepts, il reflète le contenu des articles indexés, incluant tous les aspects de la recherche en astronomie y compris les éléments mathématiques, la chimie interstellaire et la physique théorique présents dans les articles.

TERMINOLOGIE



CORPUS ET THÉSAURUS...

- » Mise à jour du **corpus mémoire psychologie — neurosciences** et création d'un **nouveau corpus Mémoire et SHS** dans le cadre de l'utilisation des terminologies dans l'exposition de corpus Istex.
- » Mise à jour du **thésaurus Mémoire** avec ajout du marquage des bibliographies sélectives (type d'étude et type d'accès au document).
- » Création d'une nouvelle ressource, le thésaurus bilingue (français-anglais) « **Composés inorganiques** ».

DU CÔTÉ DES PROJETS...

- » Poursuite des travaux de constitution et d'enrichissement du **thésaurus en paléoclimatologie** (Paleosaurus), en collaboration avec la

DiBiso et des chercheurs de l'Université Paris-Saclay, du LSCE et de Géosciences Montpellier, avec notamment l'attribution de définitions (en français et en anglais) à chacun des termes.

- » **Projet TyDI - Phase 2** : définition des spécifications (IHM essentiellement) ; tests des développements réalisés, révision de la documentation utilisateur.

Loterre

Nombre de terminologies chargées
au 1^{er} janvier 2024 : **72**

TRADUCTION

QUELQUES EXEMPLES DE TRADUCTIONS...

De l'anglais vers le français :

- » Traductions pour le **CNRS** (Direction de la communication, Direction des données ouvertes de la recherche - DDOR, mission pilotage et relations avec les délégations régionales et les instituts - MPR, Direction Europe de la recherche et coopération internationale - DERCI) tout au long de l'année : **61 110 mots**.
- » Traductions pour le **Comité pour la science ouverte** : **12 685 mots**.

Du français vers l'anglais :

- » Traduction du **rapport d'activité annuel du CNRS** à la demande de la Direction de la communication : **20 799 mots**.

- » Traduction du rapport « **L'égalité femmes-hommes dans le financement de la recherche. Étude de 11 pays européens, Israël et le Canada et de son annexe** » dans le cadre du projet Gender-Net Plus pour la Mission pour la place des femmes au CNRS : **83 500 mots**.
- » Traductions de rapports et documents pour **EOSC** : total de **69 406 mots**.

Volumes pour les activités de traduction et relecture :

Français vers anglais : 121 documents - **378 134 mots**.

Anglais vers français : 29 documents - **261 093 mots**.

ET AUSSI...

- » Participation au **Groupe de travail COAR** (Confederation of Open Access Repositories) sur la prise en compte technique des langues non anglaises dans les archives ouvertes.

- » **Travaux d'extractions et de dépouillements terminologiques** et coopération avec le service Textes & Corpus - Istex dans le cadre du projet ANR MaTOS consacré à l'emploi et l'évaluation de la traduction automatique de textes scientifiques.

LANCEMENT DU PROJET MATOS

L'Inist participe au projet **MaTOS (projet ANR)** dont le lancement officiel a eu lieu en février 2023 à Paris.

Le projet **MaTOS, Machine Translation for Open Science** (Traduction automatique pour la science ouverte) a pour objectif de fluidifier la circulation et la diffusion des connaissances scientifiques par une **traduction automatique améliorée**. Il vise à créer un outil expérimental de traduction automatique autour de la science ; un domaine où cette technologie se heurte parfois à des difficultés.

L'Inist y apporte des ressources qui ressortent de ses activités usuelles, comme les **ressources textuelles Istex & terminologiques Loterre**, ainsi que ses **multiples compétences** dans le domaine du traitement des vocabulaires.



Soumission d'un billet de blog sur OPUS-CAT-MT pour publication sur Hypothèses.org : **Traductions et science ouverte - Explorer le rôle de la traduction dans la science ouverte et multilingue**. Jean-François Nominé.

Andrea Chiarelli, Lucia Loffreda, Rob Johnson, Jean-François Nominé, Manon Parrinello, et al. **Publication et Données Reproductibles de la Recherche : Que Faire ? : Soutenir et Encourager de Nouvelles Pratiques**. Knowledge Exchange. 2023, 82 p. (hal-04170637)

Text & Data Mining (TDM)

L'Inist propose différents services dédiés au Text & Data Mining en développant et adaptant des méthodes de fouille de textes pour proposer des enrichissements adaptés aux textes, selon les besoins.



L'offre de service en TDM a été intégrée comme un service à la recherche dans l'infrastructure Istex. Pour les projets de fouille de texte, les web-services ont été mis à disposition sur le site Istex.

PARMI LES RÉALISATIONS 2023...

- » Poursuite de la création et de la mise à disposition de web-services de TDM : **38 web-services accessibles.**
- » **Premier web-service de traitement de corpus** (LDA ou Topic modelling) : modèle probabiliste permettant de déterminer des sujets ou thèmes dans un corpus de documents.
- » **Premiers développements du service « IA factory »** : interface simple de chargement de corpus utilisateurs et d'exécution de programmes d'Intelligence Artificielle (IA).
- » Animation de 2 sessions du **webinaire « Découverte du TDM »** (120 inscrits).



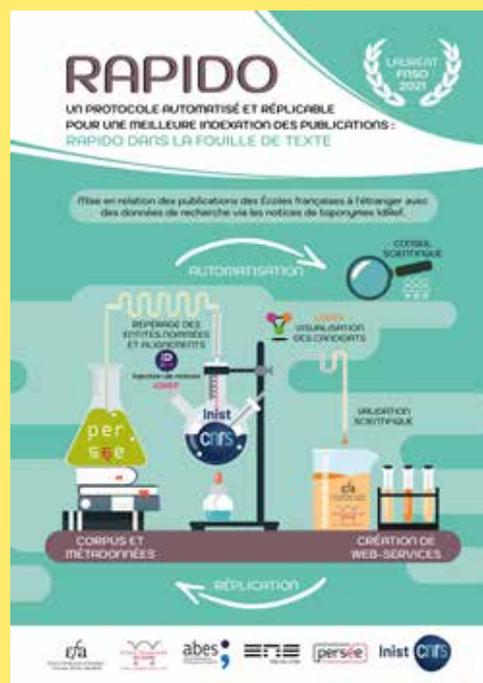
Kévin Cousot, Cédric Lopez, Pascal Cuxac, Vincent Lemaire. **Défi Text-Mine'23 - Reconnaissance d'entités d'intérêts dans les signatures d'e-mails.** Atelier TextMine'23, Groupe de travail sur la fouille de textes de l'association EGC, Jan 2023, Lyon, France. (hal-03951933)

Pascal Cuxac, Cédric Lopez, Vincent Lemaire. **TextMine'23 : Atelier sur la fouille de textes.** 2023. (hal-04188281)

Projet RAPIDO

Rapido (Rendre Accessibles des Publications scientifiques Indexées et liées à des DONnées certifiées) est porté par l'ENS de Lyon pour Persée, en partenariat avec l'Inist, l'École française de Rome, l'École française d'Athènes et l'Abes. Il a pour but de mettre en place, d'abord sur un corpus délimité de revues des Écoles françaises à l'étranger, un **protocole automatisé** permettant la mise en relation de ces publications avec des données de recherche via les notices de toponymes Idref.

La première réunion du Conseil Scientifique s'est déroulée le 13 septembre.



Poster présenté aux **JABES**, 23-24 mai 2023, Montpellier.

Analyses bibliométriques

Les analyses bibliométriques offrent la possibilité de surveiller la production scientifique d'une unité de recherche, d'un institut ou de toute autre entité, ainsi que celle associée à un domaine scientifique spécifique, permettant d'établir une vue d'ensemble sur un sujet donné.



Le service a réalisé dix études bibliométriques, à partir de la base Conditor, pour suivre la production scientifique des unités de recherche des instituts du CNRS pour la période 2017-2022 : CNRS Biologie, CNRS Chimie, CNRS Écologie & environnement, CNRS Ingénierie, CNRS Mathématiques, CNRS Nucléaires & Particules, CNRS Physique, CNRS Sciences humaines & sociales, CNRS Sciences informatiques et CNRS Terre & Univers.

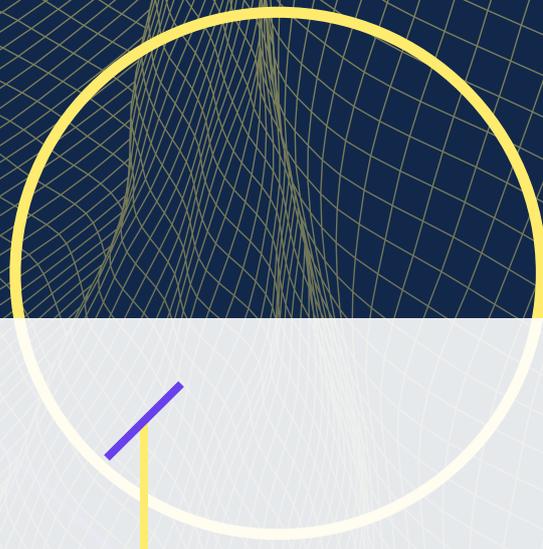
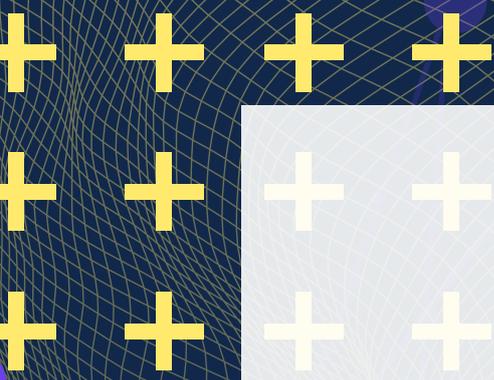
Conditor est une base de signalement des publications scientifiques françaises qui recueille les métadonnées des bases en accès libre HAL, PubMed et Crossref.

PARMI LES ACTIONS RÉALISÉES EN 2023...

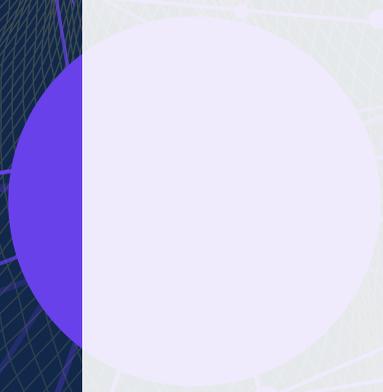
- » Participation à deux Expertises Scientifiques Collectives (ESCo) : **ESCo sur les impacts des parcs éoliens en mer, biodiversité et socio-écosystèmes marins & côtiers** (co-piloté par le CNRS et l'IFREMER) et **ESCo sur les terres rares** (piloté par le CNRS), en appui au service Textes & Corpus - Istex.
- » **Quantification des publications scientifiques en accès libre produites par le CNRS** pour l'année 2021 à partir des bases du Web of Science. 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 ».
- » Réalisation d'une **étude pour identifier les pratiques en fouille de texte** des chercheurs français non-experts en fouille de texte.
- » Réalisation d'une **étude pour établir le poids des APC** (Article Processing Charges) dans la production 2018 du CNRS afin de comparer sur une même année les coûts réels et théoriques.
- » Mise en oeuvre de **6 recherches bibliographiques et 46 études bibliométriques**.

Quelques exemples d'études bibliométriques :

- Suivi de la production scientifique des unités de recherche CNRS implantées à **La Réunion** (OSU-REUNION, LACy, PIMIT, ENTROPIE) pour la période 2012-2021 ;
- Suivi de la production scientifique de plusieurs **unités de recherche** du CNRS : I2BC Chrono-environnement, GANIL, B3OA, Georesources ;
- Suivi de la production scientifique des **entités de recherche** de plusieurs instituts du CNRS en lien avec des TGIRs-Très Grandes Infrastructures de Recherche : CNRS Physique, CNRS Biologie, CNRS Terre & Univers, CNRS Écologie & Environnement, CNRS Chimie ;
- Suivi de la production scientifique française dans les domaines de la **paléoclimatologie** et du **paléoenvironnement** pour la période 2000-2020 (LSCE et IGE) ;
- Suivi de la production scientifique française dans le domaine de **l'énergie** pour la période 2014-2021 (Cellule Énergie de CNRS Ingénierie) ;
- Suivi de la production scientifique des unités de recherche CNRS implantées en **Outre-Mer** pour la période 2015-2021 ;
- Suivi de la production scientifique de l'**Université Bordeaux Montaigne** pour la période 2013-2023 ;
- Suivi de la production scientifique française dans le domaine de **l'autonomie** (vieillesse et situation de handicap) pour la période 2000-2021 (Programme Prioritaire de Recherche Autonomie) ;
- Suivi de la production scientifique française sur **la fin de vie et les soins palliatifs** pour la période 2010-2021 (Plateforme nationale pour la recherche sur la fin de vie).



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



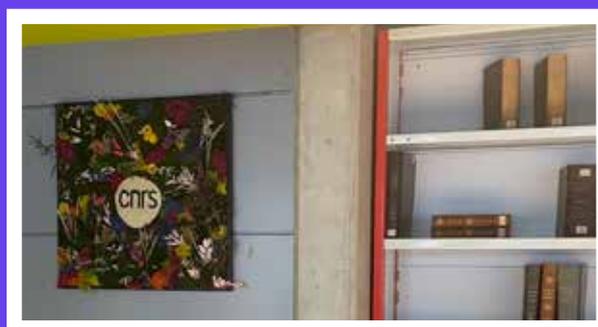
Communication



L'Inist a fêté ses 35 ans en 2023 !

Afin de célébrer cet anniversaire comme il se doit, et toujours avec l'objectif de créer des moments de cohésion pour les agents de l'unité, le service Communication a organisé la constitution de 6 fresques végétales collectives.

Chacun a pu laisser aller sa créativité pour réaliser ces fresques destinées à venir colorer plusieurs salles de réunion de l'Inist.



ACTIONS DE COMMUNICATION

» Réalisation de **plusieurs posters** :

- Exposer ses données de recherche : une offre de service de l'Inist-CNRS (JABES, mai 2023, Montpellier)
- Rapido (JABES, mai 2023, Montpellier)
- HAL-CNRS (Journées CasuHAL, juin 2023, Toulouse)
- Le corpus Machine Translation, une exploration diachronique des (méta)données Istex (CORIA-TALN, juin 2023, Paris)
- DMP OPIDoR, un compagnon pour FAIR (séminaire Recherche Data Gouv, décembre 2023, Reims)
- Processus d'adaptation d'une ressource pédagogique et Processus de création d'une ressource pédagogique (séminaire Recherche Data Gouv, décembre 2023, Reims)

» Participation à des **manifestations** :

- **ADBU**, 27-29 septembre, Versailles
- **Journées ABES**, 23-24 mai, Montpellier
- **Journée des nouveaux entrants** (DRO6), 12 mai, Vandœuvre-lès-Nancy

» Réalisation d'une **vidéo métier** diffusée sur CANAL U :

Lucile, chargée d'études bibliométriques



RÉSEAUX SOCIAUX

L'Inist est présent sur **3 nouveaux réseaux sociaux** :

» Instagram

» Mastodon

» BlueSky



LinkedIn

2 800 abonnés

(+ 370 abonnés en 2023 !)

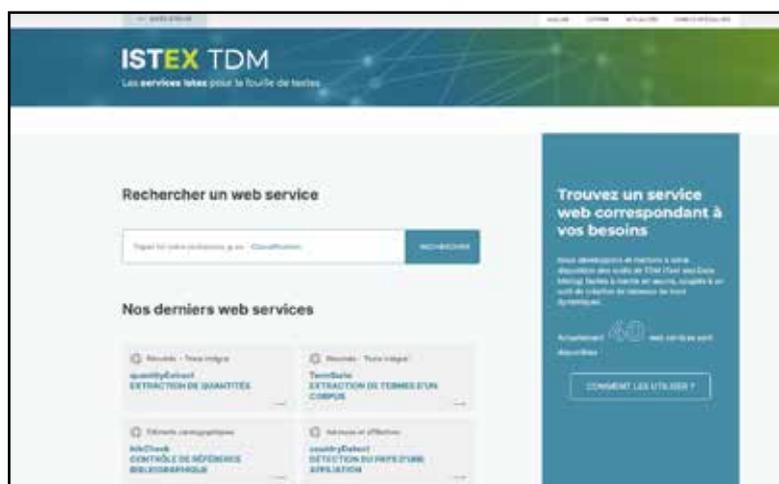


Twitter

4 650 abonnés

Web

Réalisation du site **Istex TDM** qui vient compléter l'offre de services Istex en matière de Text & Data Mining.



Veille

LALIST

» À travers son **service LaLIST**, véritable base de connaissances, l'Inist accompagne le monde de la recherche et les professionnels du domaine en mettant à disposition une mine d'informations actualisées.

1 071 billets publiés

1 188 abonnés

OUVRIR LA SCIENCE

» L'Inist poursuit son travail de veille pour le **site Ouvrir la science**, piloté par le Comité pour la science ouverte, en collaboration avec le ministère de l'Enseignement supérieur et de la Recherche pour la mise en ligne de contenus.



— / Systèmes d'information / —

Lancement de la rénovation complète de notre infrastructure informatique : le service se renforce

Deux actions principales ont été mises en œuvre par le service dédié :

- » **Opérations préparatoires au déménagement vers le DCML** (lire ci-contre) : participation à la rénovation du câblage, adaptation des infrastructures actuelles.
- » **Définition et premiers travaux sur le système d'information (SI) cible 2028** : une usine logicielle adaptée aux nouveaux besoins (fouille de textes, sécurité, haute disponibilité).

2023 a aussi été marquée par la fin des travaux du **Datacenter mutualisé lorrain (DCML)**, projet dans lequel sont engagés l'Université de Lorraine, le CHRU de Nancy-Brabois et la Métropole du Grand Nancy.

Le CNRS a intégré le projet en tant que partenaire opérationnel.

L'Inist y transférera son infrastructure informatique courant 2024.

— / Secrétariat général / —

Parmi les actions majeures réalisées par le service de gestion administrative et logistique :

- » **ESR Missions** : compréhension des flux, debuggage des premières missions, accompagnement des agents et enfin conception d'un mode opératoire incluant les « Trucs, astuces et points d'attention » à destination des missionnaires.
- » **Gestion budgétaire** : refonte de la gestion et rédaction de fiches de synthèse adaptées à nos achats spécifiques.

Effectif au 31/12/2023 : **149 agents**

Budget exécuté : **6,1 millions d'euros**, dont 4,7 millions pour les acquisitions (BibCnrs).

Mouvements RH :

- 15 départs de l'unité
- 7 lauréats au concours
- 10 recrutements (CDI et contrats de courte durée)
- 6 stagiaires accueillis

À votre rencontre

10 janvier

Séminaire des URFIST

Paris

Présentation : Intelligence artificielle et fouille de textes

Pascal CUXAC

17 janvier

EGC'23

Lyon

Animation de l'Atelier

TextMine

Pascal CUXAC

8 février

Après-midi Science Ouverte

Organisé par l'Université de Paris Sorbonne,

Université de Villetaneuse

DMP OPIDoR Un outil au service du chercheur pour l'élaboration de PGD

Laurent RASSINOX

23 mai > 24 mai

JABES

Montpellier

Présentation du poster « Un protocole automatisé et répliquable pour une meilleure indexation des publications : RAPIDO dans la fouille de texte »

Justine REVOL

Présentation du poster « Exposer ses données de recherche : une offre de service de l'Inist-CNRS »

Alain ZASADZINSKI

5 juin > 9 juin

CORIA-TALN 2023

Paris

Présentation du poster sur « Le corpus Machine Translation : Une exploration diachronique des (méta) données Istex »

Mathilde HUGUIN

Sabine BARREAUX

6 juin > 8 juin

Rencontre entre les Ateliers de la donnée de l'écosystème Recherche Data Gouv

Les Centres de référence thématiques et le Centre de ressource entrepôt catalogue Lyon

Animation d'ateliers

Marie-Christine JACQUEMOT

7 juin

ANF

La protection des données personnelles dans les projets de recherche dans le contexte de la science ouverte, Lyon

Plan de gestion de données

DMP OPIDoR : un fil conducteur pour la gestion des données personnelles

Anne BUSIN

Françoise COSSERAT

8 juin > 9 juin

Atelier « Enrichissement du thesaurus biodiversité de l'Inist et usage du thesaurus »

Organisé par le Conseil scientifique du Pôle National de Données de Biodiversité, Paris

Animation de l'atelier

Dominique VACHEZ

9 juin

Journée EBSCO

Paris

Présentation de l'extension Click & Read

Laurence GRAND

Florence JOIN

Dolorès DARDAINE

14 juin > 16 juin

Journées CasuHAL

Toulouse

Présentation du poster

« HAL-CNRS : une nouvelle

équipe valorise la production scientifique du CNRS »

Théo ROUSSELLE

Séverine KOEHL

22 juin

Journée SOFT

Tous formés à la science ouverte en SHS

Bordeaux

DMP OPIDoR — Plan de gestion de données

Lauriane LOCATELLI

28 juin

Une journée de formation pour les lauréats de Winning Normandy

Webinaire : Présentation des enjeux de l'ouverture des données de la recherche

Gabriela TORRES RAMOS

Erwan GUILLOU

5 juillet

ANF « Pratiquer la science ouverte » (DDoR)

Paris

Animation de 2 ateliers ludiques « Dép'osez »

Coralie WYSOCZYNSKI

Lynda KEHLI

11 juillet

Journée des Directeurs d'Unités de la DRO6

Vandœuvre-lès-Nancy

Les APC en contexte(s) : de la vision laboratoire à la vision Direction

Xavier LAUNOIS

28 septembre

ADBU

Versailles

Présentation de l'extension Click & Read

Dolorès DARDAINE

28 septembre

Ateliers à la demande de l'IFSEM CNRS

Paris

Analyse de DMP

Coralie WYSOCZYNSKI
Claire SOWINSKI

28 septembre

Journées DOREL

UL-Nancy

Présentation de Loterre

Stéphane SCHNEIDER

3 octobre > 6 octobre

2^e congrès du GDR Mémoire

Carry-le-Rouet

Communication « Mémoire et justice : la terminologie au service de la recherche et du travail de terrain »

Frank ARNOULD

3 octobre

Open Science Month

Paris

Le Paléosaurus, vocabulaire normé en paléoclimatologie

Ludovic HAMIAUX

9 octobre > 13 octobre

Journées RDA France

Webinaire « Plan de Gestion des Données : Bonnes pratiques et outils »

Pierrette PAILLASSARD
Coralie WYSOCZYNSKI

Webinaire « Les nouveautés de DMP OPIDoR »

Anne BUSIN
Françoise COSSERAT

12 octobre > 13 octobre

ANF « TDM 2023, Exploration documentaire et extraction d'information »

La constitution et l'exposition d'un corpus documentaire à partir des ressources Istex
Mathilde HUGUIN

23 octobre > 26 octobre

Journées 2023 du GDR EMILI

Étude des Milieux Ionisés :

Plasmas froids créés par décharge et laser
Vandœuvre-lès-Nancy

Animation de l'atelier participatif « Pratiques de publication »

Xavier LAUNOIS

10 novembre

20 ans de Persée

Lyon

Atelier « Enrichissement, diffusion et visualisation des données : collaborations pour une offre de service à destination de la recherche »

Justine REVOL

22 novembre > 24 novembre

Journées 2023 du Consortium « 3D pour les Sciences Humaines et Sociales »

Lyon

La 3D au secours du patrimoine horloger. CHRONOSPEDIA : Encyclopédie virtuelle du savoir horloger

Alain ZASADZINSKI
Titouan BOUDART

24 novembre

Journées MAGIS 2023

Atelier Humanités Numériques
Spatialisées
Bordeaux

Présentation du projet Rapido

Justine REVOL

27 novembre

Atelier GopenDoRe

Toulouse

Atelier pour les ingénieurs plateforme de CNRS Ingénierie

Coralie WYSOCZYNSKI
Lynda KEHLI

1^{er} décembre

ICDM'23

Shangai

Animation de l'atelier IncrLearn

Pascal CUXAC

12 décembre > 13 décembre

Journées OMEKA

Paris

Travailler en équipe et en communautés avec Omeka : Accompagner les équipes de recherche jusqu'à l'administration autonome de leurs plateformes

Edwige PIEROT
Louis MAILLARD

13 décembre

Semaine Data SHS

Dijon

Les plans de gestion de données en SHS : intérêt du DMP et l'outil DMP OPIDoR

Lauriane LOCATELLI

14 décembre > 15 décembre

Séminaire des dispositifs d'accompagnement de l'écosystème Recherche Data Gouv

Reims

Atelier « DMP OPIDoR du modèle narratif au modèle structuré »

Anne BUSIN
Françoise COSSERAT
Erwan GUILLOU

Workshops « Évolution de la plateforme DoRANUM »

Gabriela TORRES RAMOS
Claire SOWINSKI

Bar camp « Les processus d'ingénierie pédagogique DoRANUM »

Coralie WYSOCZYNSKI
Pierrette PAILLASSARD



Un peu de détente...

MOTS CACHÉS

N	C	H	E	R	C	H	E	U	R	A	K	F	P	P
I	N	F	O	R	M	A	T	I	O	N	V	K	O	J
K	A	C	C	O	M	P	A	G	N	E	M	E	N	T
V	I	K	T	S	C	I	E	N	C	E	F	X	B	D
N	H	P	O	R	T	A	I	L	L	B	P	Y	T	Q
J	Q	V	A	L	O	R	I	S	A	T	I	O	N	A
G	T	E	R	E	C	H	E	R	C	H	E	P	K	G
S	Y	T	C	J	A	S	E	R	V	I	C	E	R	T
F	O	G	D	F	C	C	Z	J	H	W	C	I	V	B
V	N	I	O	J	F	O	U	I	L	L	E	B	Z	Y
V	X	W	M	I	S	S	I	O	N	R	E	H	B	V
N	N	T	A	N	A	L	Y	S	E	C	U	C	N	G
G	E	S	T	I	O	N	C	K	B	O	U	T	I	L
X	T	B	D	O	N	N	É	E	K	X	Z	J	Y	D
Z	Z	I	A	C	C	È	S	Q	A	P	P	U	I	R

ACCOMPAGNEMENT

ACCÈS

ANALYSE

APPUJ

CHERCHEUR

DONNÉE

FOUILLE

GESTION

INFORMATION

MISSION

OUTIL

PORTAIL

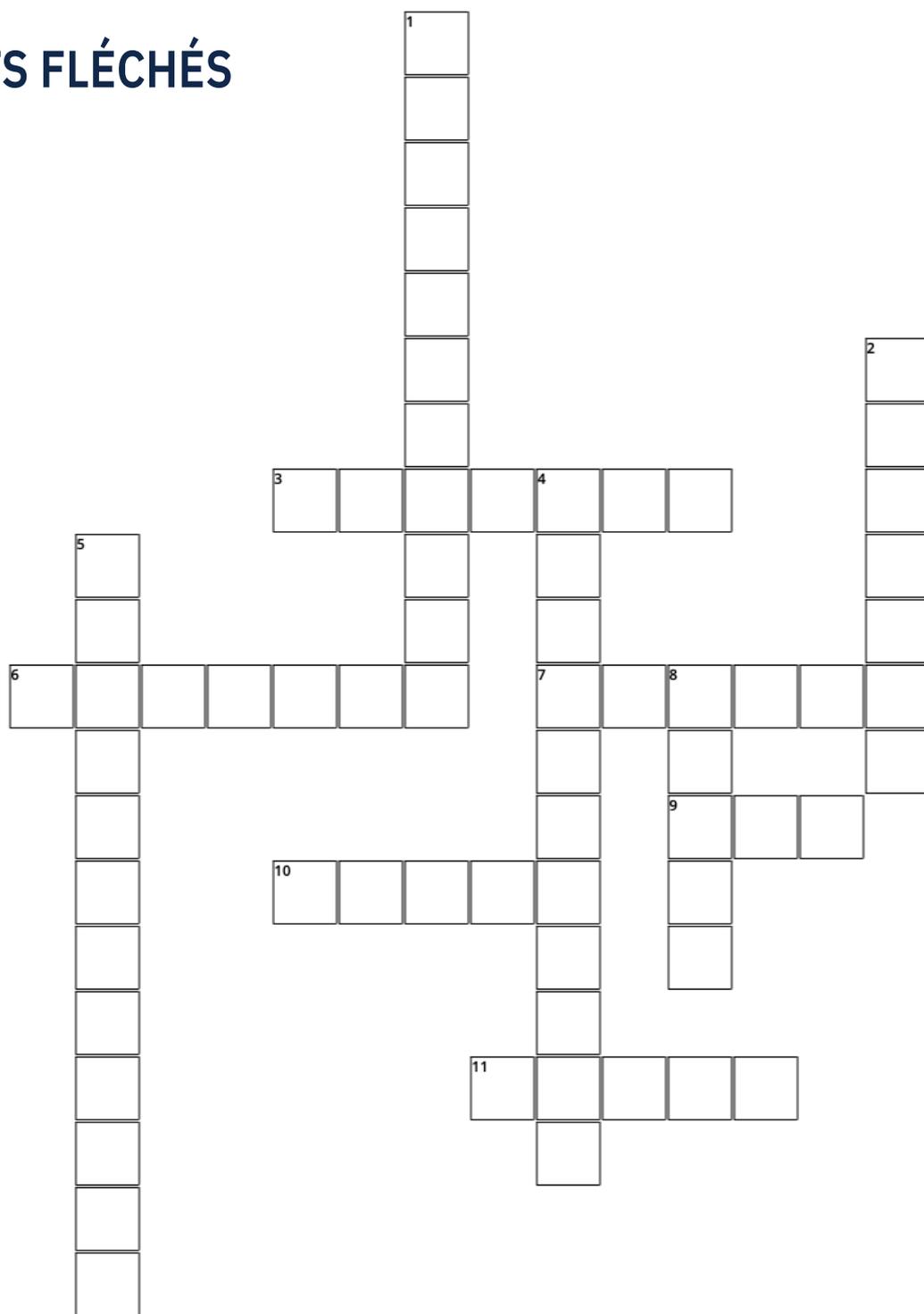
RECHERCHE

SCIENCE

SERVICE

VALORISATION

MOTS FLÉCHÉS



HORIZONTAL

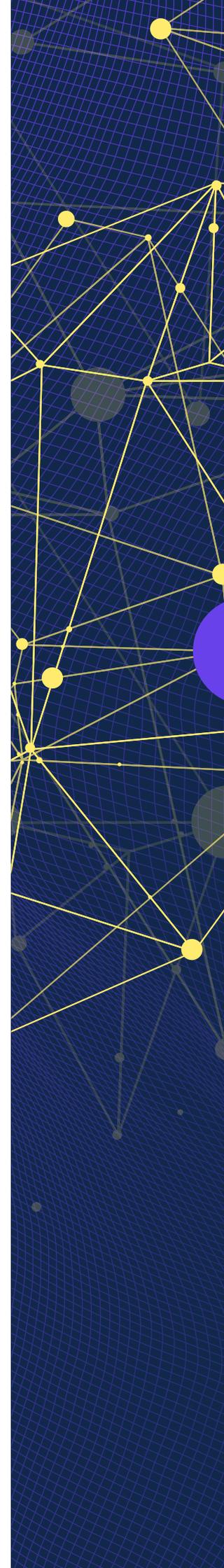
3. Plateforme de formation
6. Portail multidisciplinaire
7. Outils et services pour accompagner la gestion des données de la recherche
9. Text & Data Mining
10. Logiciel de création de bibliothèque numérique
11. Logiciel de traitement et de valorisation des données

VERTICAL

1. Boîte à outils pour la mesure des usages
2. Portail de la production scientifique du CNRS
4. Opération qui permet de renouveler ou d'acquérir de nouvelles revues en limitant la hausse des tarifs
5. Extension web pour l'accès aux documents
8. Plateforme accueillant plus de 27 millions de publications scientifiques

Ce rapport d'activité a été réalisé par le service Communication de l'Inist, en lien avec l'ensemble des services de l'Institut.

Conception graphique : Alexandra Petitjean-Monnin





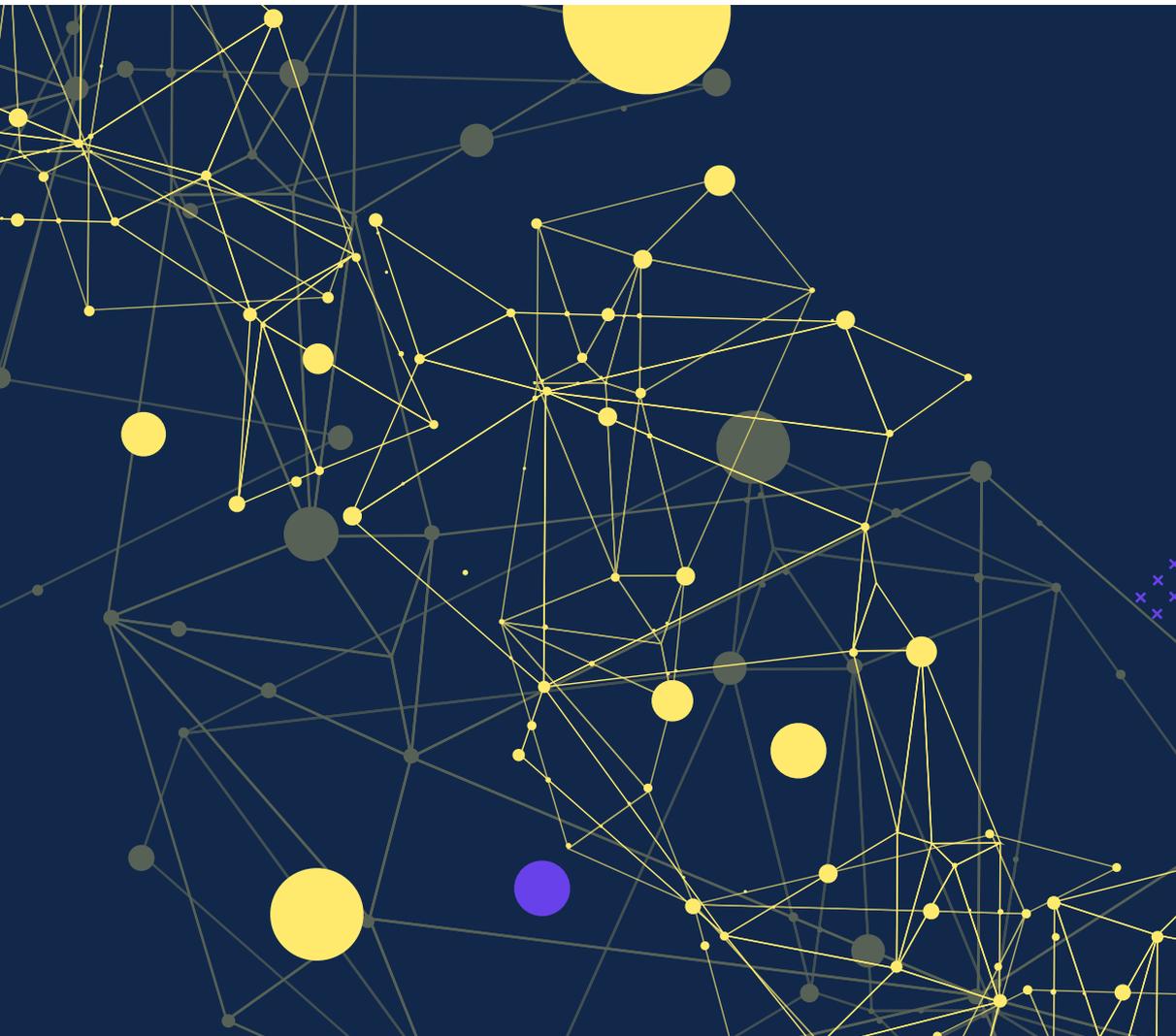
+++
+++
+++
+++

+++
+++
+++
+++

+++
+++
+++

+++
+++
+++

+++



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

