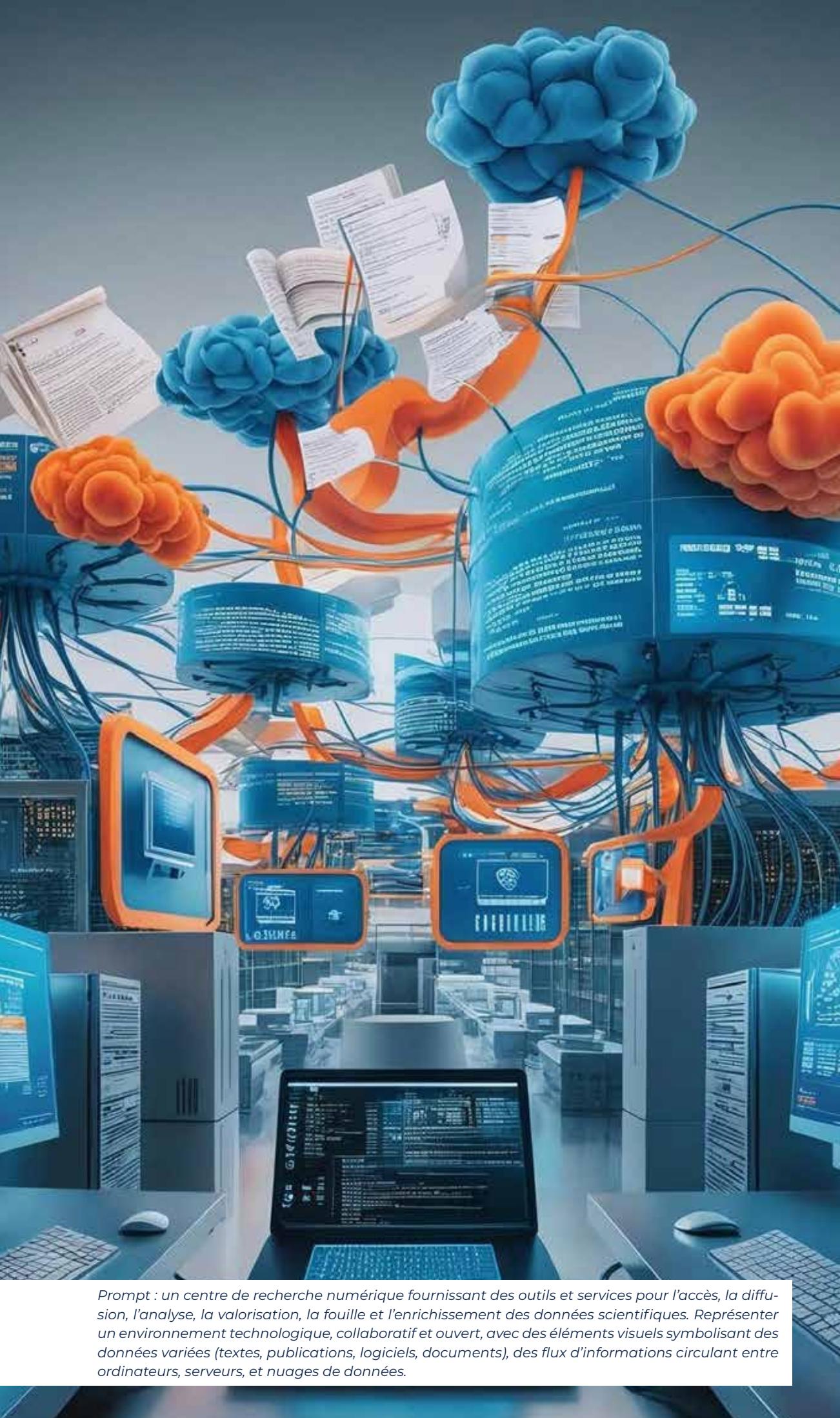
The background of the entire page features a large, modern glass building with a curved facade, identified as the INIST building. Superimposed over the building is a complex digital interface. This interface includes several floating computer monitors displaying code and data, glowing blue and orange light trails, and several open books with glowing pages, suggesting a fusion of traditional knowledge and modern digital information. The overall color palette is dominated by blues, oranges, and yellows.

# INIST

## Institut de l'information scientifique et technique

*Rapport d'activité 2024*



Prompt : un centre de recherche numérique fournissant des outils et services pour l'accès, la diffusion, l'analyse, la valorisation, la fouille et l'enrichissement des données scientifiques. Représenter un environnement technologique, collaboratif et ouvert, avec des éléments visuels symbolisant des données variées (textes, publications, logiciels, documents), des flux d'informations circulant entre ordinateurs, serveurs, et nuages de données.

# SOMMAIRE

<b>Édito</b>	<b>5</b>
<b>Accéder aux produits de la recherche</b>	
<b>Portails et Plateformes</b>	<b>8</b>
<b>Négociations et Acquisitions</b>	<b>9</b>
<b>Mesure des usages</b>	<b>10</b>
<b>HAL-CNRS</b>	<b>12</b>
<b>Édition numérique</b>	<b>13</b>
<b>Gérer et valoriser les données de la recherche</b>	
<b>OPIDoR</b>	<b>16</b>
<b>Formation - DoRANum</b>	<b>18</b>
<b>ExpoDoR</b>	<b>20</b>
<b>Analyser et fouiller l'information scientifique</b>	
<b>Textes &amp; Corpus - Istex</b>	<b>24</b>
<b>Ressources langagières multilingues</b>	<b>26</b>
<b>Text &amp; Data Mining (TDM)</b>	<b>28</b>
<b>Appui au pilotage scientifique</b>	<b>29</b>
<b>Et aussi...</b>	
<b>Communication, Web, Veille</b>	<b>31</b>
<b>Secrétariat général</b>	<b>32</b>
<b>Ingénierie des systèmes d'information</b>	<b>33</b>
<b>À votre rencontre...</b>	<b>35</b>





# ÉDITO

## Des services à votre écoute

L'ensemble des services de l'Inist s'organisent autour d'applications numériques que nous faisons évoluer en fonction des demandes et des usages. Cette année 2024, nombre de ces applications ont profité d'**évolutions importantes**. Pour les définir, nous tenons compte de l'ensemble de vos retours qu'ils soient directs auprès des équipes ou le résultat d'enquêtes ou de tests, nous vous en remercions chaleureusement.

Afin de permettre un accès toujours plus fluide aux productions scientifiques internationales, l'interface du portail **BibCnrs** a été modernisée permettant des modes de navigation adaptés à la diversité de besoins exprimés. Tandis que l'outil **Click & Read** a été adapté pour faciliter le paramétrage par les utilisateurs et également pour répondre aux exigences des navigateurs sur lesquels il s'appuie (ex. FireFox). Développée en collaboration avec le Consortium Couperin, la suite **ReadMetrics**, pour la mesure des usages documentaires dans les différents établissements, évolue afin d'améliorer la fiabilité des indicateurs produits et faciliter la personnalisation des tableaux de bord générés.

De même, le déploiement d'une version plus interactive et mieux connectée aux différentes sources d'information de l'outil de rédaction de plan de gestion de données **DMP OPIDoR** a reçu un vif succès. Afin d'accompagner l'exposition des matériaux de recherche, l'Inist s'est doté d'un **atelier de photographie numérique** en très haute définition qui a été inauguré avec la numérisation de 10 000 planches de l'herbier du Jardin botanique du Lautaret, dans le cadre d'un partenariat avec l'Université Grenoble-Alpes.

L'Inist poursuit régulièrement l'alimentation des ressources documentaires de l'infrastructure **Istex**, en complément la nouvelle fonctionnalité **Istex Search** propose un parcours complet incluant l'interrogation du réservoir de publications, l'exploration des résultats et le téléchargement du corpus ainsi réalisé. Afin d'effectuer une analyse avancée de ces corpus, ou de leurs corpus personnels, les utilisateurs disposent de nouveaux **web-services** asynchrones permettant de réaliser, par exemple, l'extraction de thématiques, la détection de données bruitées, l'extraction de terminologies ou l'identification de références phares. Cette infrastructure s'est également enrichie d'un catalogue de **ressources terminologiques** mobilisables pour l'analyse.

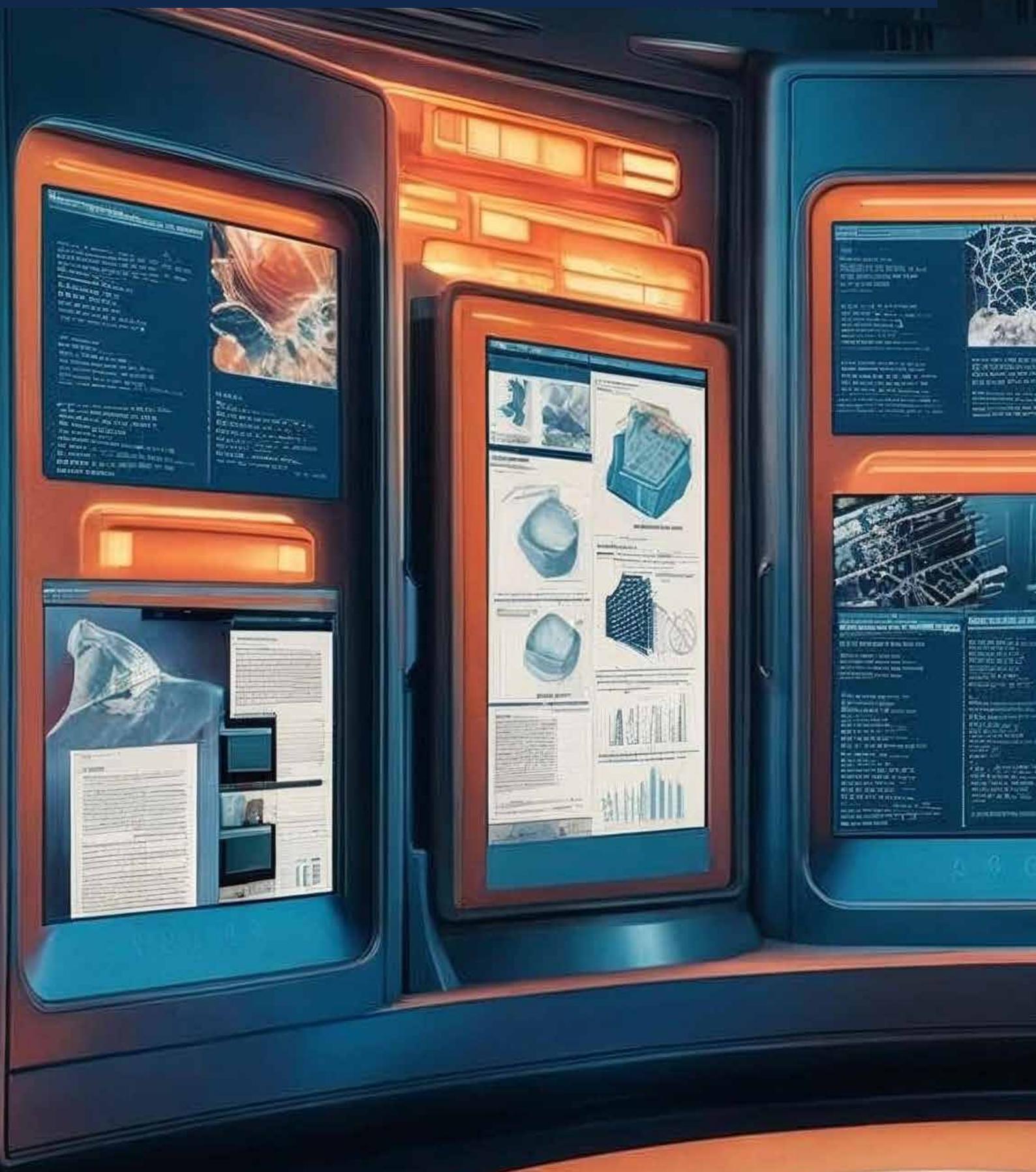
L'Inist s'organise également pour vous rencontrer lors des **rendez-vous nationaux** réguliers comme le congrès de l'ADBU ou les JABES. En cette année 2024, nous vous avons également accueillis dans nos locaux de Vandœuvre-lès-Nancy d'une part lors des premières rencontres organisées par le nouveau **Consortium DataCite France** les 18 et 19 juin. Elles ont permis aux membres du comité exécutif et des groupes de travail de se retrouver et d'avancer les différentes thématiques portées par le consortium. D'autre part, nous avons accueilli le **quatrième séminaire Recherche Data Gouv** du 20 au 22 novembre qui a été l'occasion d'échanges très riches rendus possibles par la place importante réservée aux ateliers.

Le déploiement de l'**intelligence artificielle** est une réalité qui impactera l'ensemble des métiers, il est donc important de nous y préparer sans précipitation. L'illustration de ce rapport d'activité porte sur cette thématique. Les images qui occupent une place suffisamment importante sont accompagnées du « prompt » utilisé pour les générer.

Bonne lecture.

**Claire FRANÇOIS**  
Directrice de l'Inist

# Accéder aux produits de la recherche





Prompt : Illustration de l'accès à l'information scientifique et technique avec des écrans futuristes affichant des publications scientifiques et des bases de données.

# PORTAILS ET PLATEFORMES

L'équipe Portails et plateformes facilite l'accès aux publications scientifiques pour les chercheurs en leur proposant des outils performants et adaptés à leurs besoins. Elle gère et améliore des plateformes et outils comme BibCnrs, PANIST et Click & Read qui permettent de consulter facilement des ressources en ligne, qu'elles soient négociées ou en accès ouvert.

## Une année de refonte pour BibCnrs !

S'appuyant sur les retours de l'enquête réalisée en 2023 auprès des utilisateurs du portail BibCnrs, l'année 2024 a été consacrée à l'idéation et la mise en œuvre d'améliorations pour **optimiser les parcours d'utilisation** des ressources accessibles via le portail BibCnrs.

La nouvelle version a été lancée le 8 janvier 2025.

## BibCnrs

Épaulée par une prestation de développements informatiques et guidée par les avis des usagers testeurs, l'équipe BibCnrs a préparé une **version repensée** alliant de nouvelles fonctionnalités, une ergonomie plus moderne, le tout dans un environnement paramétrable pour laisser à chacun le choix de son mode de navigation idéal. En parallèle, le contenu de BibCnrs a été enrichi de nombreuses ressources en accès ouvert de qualité.

Les équipes « Portails et Plateformes » et « Négociations et Acquisitions » ont œuvré main dans la main pour proposer une offre de service d'accès aux publications scientifiques reconnue et répondant au mieux aux besoins des communautés de recherche du CNRS.

En 2024, BibCnrs c'est :

- **5 172 857 consultations** mesurées par ezMeasure dont 2,7 millions d'articles scientifiques ;
- **55 792 ayants droit** (15 452 chercheurs CNRS et 40 340 chercheurs non CNRS) présents dans l'ensemble des unités du CNRS ;
- **8 891 abonnés** aux listes de diffusion ;
- **1 464 abonnés** via le compte X.



## PANIST, plateforme nationale pour l'accès pérenne à la documentation scientifique

De **nouveaux contenus** ont été ajoutés, notamment les articles Elsevier de 2023 et Springer de 2018 à 2022. Ces ressources sont accessibles via l'extension Click & Read.

Les droits d'accès ont été mis à jour pour s'adapter aux évolutions des établissements membres des groupements de commandes.

Par ailleurs, la page d'accueil du site a été modernisée pour une navigation plus fluide.

Enfin, une analyse des données IOP de 2013 à 2021 a été réalisée en vue d'un futur chargement.

PANIST en chiffres :

- **11,6 millions d'articles scientifiques disponibles** : 10,5 millions de l'éditeur Elsevier 2002-2023 et 1,1 million de l'éditeur Springer 2015-2022 ;
- **67 381 consultations** (hausse de 185 % des consultations de 2023 à 2024) ;
- **96 % des consultations** de PANIST effectuées via Click & Read.

## Concernant les autres bases gérées par l'Inist...

- **INIS** : la convention de partenariat entre le CEA, l'Inist et le CNRS pour l'hébergement de la base de données INIS a été renouvelée jusqu'en juin 2027. **107 061 nouvelles notices bibliographiques** ont été ajoutées, portant le total des références disponibles à **4 899 036. 3 619 vues** réalisées par 971 visiteurs ;

- **OpenPascalFrancis** : **1,7 million de vues** réalisées par 1,2 million de visiteurs ;
- **OpenBibArt** : **67 670 vues** réalisées par 28 633 visiteurs.

## Click & Read, un accès facilité aux ressources scientifiques



En 2024, Click & Read a continué de simplifier l'accès aux ressources électroniques pour les chercheurs des unités mixtes, en intégrant **13 nouveaux établissements partenaires** : Dauphine PSL, Université de Caen Normandie, INSA Lyon, Université de La Réunion, Université de la Nouvelle-Calédonie, Université Sorbonne Paris Nord, Université Bretagne Sud, Université d'Angers, Nantes Université, Université Clermont Auvergne, INRIA, Université Bourgogne Europe, Université Polytechnique Hauts-De-France.

Afin d'améliorer l'expérience des utilisateurs, l'**interface de paramétrage** a été repensée pour une utilisation plus intuitive, en collaboration avec les établissements partenaires. Une page de consentement a également été ajoutée afin de respecter les exigences des navigateurs comme Firefox.

De plus, un **système de notifications** sur Chrome et Edge a été mis en place pour informer en temps réel des mises à jour et nouvelles sources disponibles.

Enfin, un **kit de communication** a été créé pour mieux accompagner les utilisateurs.

- Plus de **40 000 installations** depuis son lancement en mars 2021 ;
- Plus de **17 000 utilisateurs** quotidiens ;
- **767 800 consultations** d'articles en 2024 ;
- **3 000 consultations** quotidiennes d'articles ;
- **65 établissements** partenaires (48 universités, 7 organismes de recherche et 10 grandes écoles).

## NÉGOCIATIONS ET ACQUISITIONS

L'équipe Négociations et Acquisitions gère l'achat des ressources thématiques pour le portail BibCnrs. Chaque année, le service travaille avec une cinquantaine d'éditeurs et de distributeurs pour renouveler les abonnements et en souscrire de nouveaux.

### Une année de transition

En 2024, une grande partie de son activité a été consacrée à la collaboration avec le service Portails et Plateformes pour **préparer le lancement du nouveau site BibCnrs**. Il a également accompagné la revue Acta Acustica dans son changement de modèle d'édition.

### L'Inist aux côtés de la DDOR pour accompagner l'évolution de la revue Acta Acustica.

Acta Acustica, revue de référence en acoustique éditée par EDP Sciences pour l'European Acoustics Association (EAA) a adopté le **modèle d'édition en accès ouvert diamant**, effectif au 1<sup>er</sup> janvier 2025. Grâce au soutien de plusieurs partenaires académiques, dont le CNRS, la revue reste **accessible gratuitement sans imposer de frais de publication** (APC, article processing charges) aux auteurs.

Le CNRS, par l'action conjointe de la DDOR et de l'Inist, a œuvré dès mi-2023 et tout au long de l'année 2024 à la transition économique de cette revue majeure dans le domaine de l'acoustique, d'une part en participant aux réflexions pour établir le nouveau modèle économique, d'autre part en participant financièrement à la transition, avec une contribution à hauteur de **60 000 € sur trois ans** (2025-2027).

Ce financement s'inscrit dans la politique de **science ouverte** du CNRS, qui promeut le développement de l'accès ouvert diamant et dénonce la généralisation des frais de publication en accès ouvert.



# MESURE DES USAGES

L'équipe ezTEAM – mesure des usages accompagne les établissements de l'Enseignement supérieur et de la recherche (ESR), ainsi que des consortia étrangers, dans l'optimisation de leur politique documentaire. Elle développe et met en place des outils permettant d'analyser et de quantifier l'utilisation des ressources électroniques. Grâce à ces solutions, les établissements disposent de données précises pour évaluer l'usage de leurs abonnements et ajuster leurs stratégies d'acquisition en fonction des besoins réels des utilisateurs.

## ReadMetrics : une nouvelle identité pour la suite logicielle de mesure des usages

En 2024, la coopération entre l'Inist et le consortium Couperin a franchi une nouvelle étape avec la refonte et le changement de nom de sa suite logicielle dédiée à l'analyse des usages des ressources numériques. Désormais baptisée ReadMetrics, cette **boîte à outils** offre aux établissements de l'ESR une solution complète pour collecter, analyser et visualiser les données d'usage des ressources documentaires.

## La boîte à outils dédiée à la mesure des usages

**ReadMetrics** est une suite d'outils regroupant trois modules complémentaires :

- **ezPAARSE**, un outil de traitement des logs de connexion permettant de générer des données d'usage détaillées (qui a fêté ses 10 ans en 2024 !) ;
- **ezMESURE**, une plateforme d'exploration et de visualisation des données pour analyser les usages en temps réel ;
- **ezCOUNTER**, un service de collecte et de normalisation des statistiques conformes à la norme COUNTER.

Cette nouvelle identité s'accompagne d'une **évolution des fonctionnalités** visant à améliorer la fiabilité des indicateurs, à personnaliser les tableaux de bord et à optimiser l'exploitation des données pour éclairer les décisions stratégiques.

ReadMetrics constitue ainsi un **outil clé** pour gérer les abonnements, mesurer l'impact des ressources mises à disposition des chercheurs et enseignants-chercheurs, et adapter l'offre documentaire aux besoins des communautés scientifiques.

Par ailleurs, l'équipe ezTEAM, en collaboration avec Couperin, a poursuivi plusieurs développements pour affiner la collecte et l'analyse des données d'usage :

- **Collecte automatique avec ezCOUNTER** : moissonnage courant des statistiques d'usage des éditeurs scientifiques pour le CNRS et les établissements de l'ESR ;
- **Moissonnages rétrospectifs** : collecte des statistiques antérieures à 2023 avec un maximum d'historique à destination des négociateurs Couperin ;
- **Modernisation des outils** : transition progressive de l'ancienne interface de reporting vers ezReport, offrant une mise à jour plus réactive des versions de la suite Elastic ;
- **Amélioration d'ezUnpaywall** : intégration d'un historique des données pour garantir un enrichissement fiable des informations d'accès aux ressources en Open Access, sans perte de pertinence.



## Theses.fr désormais sur Bibliomap

Une nouvelle collaboration entre l'Inist, Couperin et l'ABES a permis d'améliorer le suivi des usages des thèses en ligne. Désormais, les données d'usage de Theses.fr sont hébergées sur la plateforme ezMESURE, mises à jour quotidiennement et accessibles via un nouvel outil de visualisation : <https://bibliomap.theses.fr>. Bibliomap affiche en direct les événements de consultation générés par l'activité bibliographique de la communauté de recherche CNRS.

Ce partenariat favorise des échanges à la fois techniques et institutionnels, renforçant ainsi l'analyse des usages des ressources documentaires.

### Au plus près des utilisateurs

Les membres de l'équipe ont animé des rendez-vous tout au long de l'année avec les utilisateurs des outils et les négociateurs, en ligne, par sessions Zoom :

- 26 supports mutualisés ;
- 21 webinaires en ligne ;
- 6 revues de sprint.

Retrouvez ces rendez-vous sur <https://blog.readmetrics.org/category/rendez-vous/>



Dans le cadre de la série vidéo « **Métiers d'Inist** », Frédéric Truong explique son métier de chargé de la mesure des usages et des relations utilisateurs pour Readmetrics.

La vidéo a été mise en ligne le 6 novembre 2024 sur la chaîne Canal U de l'Inist :

<https://www.canal-u.tv/chaines/inist-cnrs>



# HAL-CNRS

HAL-CNRS est le portail institutionnel dédié à la production scientifique des unités du CNRS. Opéré par l'Inist, il s'appuie sur l'archive ouverte HAL, gérée par le CCSD, pour offrir un accès centralisé aux publications des chercheurs.

L'équipe HAL-CNRS assure l'administration du portail, la vérification des dépôts et l'animation de la plateforme. Elle accompagne les chercheurs, les instituts et les unités du CNRS dans la diffusion de leurs travaux, contribuant ainsi à la politique de science ouverte du CNRS.

## HAL-CNRS : au service des unités et chercheurs

En 2024, en complément des services proposés par le CCSD, le service HAL-CNRS a renforcé son **offre d'accompagnement pour la communauté scientifique du CNRS** :

- **Pour les unités et instituts du CNRS**, l'équipe a procédé à la création, la curation et la mise en ligne de leurs collections associées à la publication d'indicateurs. Elle a également accompagné les unités dans le recensement de leurs publications, notamment dans le cadre de leur évaluation HCERES.
- **Pour les chercheuses et chercheurs du CNRS**, l'accompagnement s'est renforcé autour du dépôt des publications dans HAL-CNRS. L'équipe a également assuré la modération des publications affichées dans leur CV-HAL et leurs rapports d'activité (CRAC, RIBAC), facilitant ainsi la valorisation des travaux scientifiques.

## Un accompagnement renforcé

À la fin de l'année 2024, l'archive institutionnelle HAL-CNRS comptait **2,2 millions de documents** déposés par les personnels des unités du CNRS.

Le **taux d'accès ouvert** des publications postérieures à 2019, date officielle de création du portail, atteignait **51,5 %**, une progression constante favorisée par la politique du CNRS, l'obligation de dépôt, la levée progressive des embargos éditeurs et le travail d'enrichissement des dépôts mené par l'équipe HAL-CNRS.

**L'accompagnement des chercheurs et des unités** est resté au cœur des missions du service. Dans le cadre de la modération, l'équipe

a vérifié et validé plus de **40 000 dépôts**, avec une activité renforcée lors des périodes de soumission des rapports d'activité CRAC et RIBAC. Elle a également soutenu les instituts et unités dans la création et la gestion de leurs collections, notamment avec l'ouverture des collections **CNRS Chimie** et **CNRS Mathématiques**, ainsi que celles de **14 unités**.

Par ailleurs, l'équipe a accompagné les unités dans la préparation de leur **dossier HCERES** et a adapté le portail HAL-CNRS ainsi que ses collections à la **nouvelle charte graphique** du CNRS, assurant ainsi une meilleure cohérence visuelle et institutionnelle.



# ÉDITION NUMÉRIQUE

L'équipe Édition numérique assure un travail de secrétariat d'édition en collaboration avec le Muséum national d'Histoire naturelle (MNHN).

Son rôle consiste à accompagner la publication de contenus scientifiques en garantissant leur qualité éditoriale.

## MetoSTEM, pour la publication numérique des monographies STM

Le projet MetoSTEM, porté par le Muséum national d'Histoire naturelle et lauréat du Fonds national pour la science ouverte (FNSO), s'est achevé en mai 2024.

Ce projet visait à proposer une solution technique innovante pour la transformation numérique des monographies en Sciences, Techniques et Médecine (STM) du MNHN.

## Une expertise en structuration de contenus et export de fichiers

L'Inist s'est impliqué dans le projet MetoSTEM en participant notamment aux tests des outils développés, grâce à ses compétences techniques en matière de structuration de contenus éditoriaux et d'export de fichiers pour publication. Cette contribution a permis d'expérimenter la solution sur trois collections pilotes du MNHN (Faune et flore tropicales, Stratotypes et Patrimoines naturels), optimisant ainsi la diffusion et l'exploitation des contenus scientifiques en biodiversité et écologie.

Plus largement, les outils et méthodes de MetoSTEM facilitent la publication numérique « **freemium** » et garantissent l'interopérabilité des données, dans une démarche de science ouverte alignée sur les principes FAIR.

## Collaboration avec le MNHN

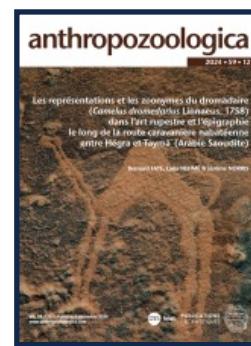
### Monographies :

Dans le cadre de la coopération qui lie l'Inist et le MNHN, l'équipe Édition a contribué à 2 parutions papier et 2 parutions numériques d'ouvrages, pour les collections *Natures en sociétés* et *Archives* du MNHN, accessibles sur le site des Publications scientifiques du Muséum national d'Histoire naturelle et OpenEdition Books.



### Périodiques :

- Édition et publication en flux continu de 13 articles pour la revue *Anthropozooologica* ;
- Éditions ponctuelles pour les revues *Zoosystema* et *Naturae*.



## Collaboration avec Calenda – OpenEdition :

Le service poursuit son engagement au sein de l'équipe de rédaction du calendrier Calenda, une plateforme multilingue en libre accès dédiée aux sciences humaines et sociales (SHS).

Cette collaboration vise à assurer la publication et la diffusion d'annonces scientifiques (colloques, séminaires, appels à contributions, etc.), renforçant ainsi l'accessibilité et la visibilité des événements académiques dans le domaine des SHS.

# Gérer et valoriser les données de la recherche





Prompt : Illustration moderne de la science ouverte montrant la gestion, le partage et la valorisation des données de la recherche : dossiers connectés, symboles d'accès libre et outils numériques, pour évoquer qualité, pérennité et réutilisation des données.

# OPIDOR

L'équipe OPIDoR accompagne les chercheurs dans la gestion et l'ouverture des données de la recherche. À travers la plateforme OPIDoR, elle met à disposition des outils et des services permettant notamment d'élaborer des plans de gestion des données (PGD), de partager, citer et préserver les données scientifiques.

## Lancement de la nouvelle version de DMP OPIDoR

L'année 2024 a été marquée par le lancement d'une **nouvelle version de DMP OPIDoR**, l'outil dédié à la création de plans de gestion de données. Cette mise à jour apporte une interface plus ergonomique, une navigation améliorée et des recommandations personnalisées pour **mieux accompagner les chercheurs**. Elle introduit également un modèle spécifique pour les entités de recherche et simplifie l'importation des informations de projet. Par ailleurs, la gestion des produits de recherche a été optimisée et l'intégration des identifiants pérennes, comme ORCID et ROR, facilitée.

## Les autres actions autour de DMP OPIDoR...

- Enrichissement de l'outil DMP OPIDoR en proposant un **nouveau modèle de plan de gestion des données spécifiquement dédié aux logiciels**. Ce modèle a été conçu pour répondre aux besoins des chercheurs et des ingénieurs développant des logiciels scientifiques, en leur offrant un cadre structuré pour documenter et organiser les différentes étapes de gestion de leurs codes et programmes. Il prend en compte des aspects clés tels que la documentation, la conservation, l'attribution d'identifiants pérennes, les licences d'utilisation et la reproductibilité des logiciels.
- Constitution d'un recueil de plans de gestion disciplinaires de référence**, un projet mené avec RDA France. Ce recueil rassemble

des exemples de plans de gestion de données adaptés à des disciplines spécifiques, afin de guider les chercheurs dans la création de leurs propres PGD. L'objectif est de fournir des modèles concrets et adaptés aux particularités de chaque domaine de recherche, facilitant ainsi l'adoption des bonnes pratiques en matière de gestion, de partage et de conservation des données scientifiques.

- Harmonisation des vocabulaires liés à la science ouverte**, un projet réalisé en collaboration avec les services DoRANum et Loterre de l'Inist. L'objectif était de standardiser les termes utilisés dans les plans de gestion de données afin de faciliter leur compréhension et leur utilisation à travers différents projets et disciplines.



## Autour de PID OPIDoR...

- Accueil et animation du **séminaire DataCite France**, qui s'est tenu les 18 et 19 juin, en collaboration avec le service Communication de l'Inist. Organisé dans le cadre du consortium DataCite France, cet événement a rassemblé les acteurs de la gestion des identifiants pérennes et de la science ouverte afin d'échanger sur les enjeux et les évolutions liés à l'attribution des DOI aux données de recherche.
- Participation à l'**assemblée générale du Consortium DataCite France** le 6 décembre au cours de laquelle le rôle de DataCite a été mis en avant dans le développement d'un cadre mondial pour des données de recherche ouvertes, connectées et réutilisables grâce aux métadonnées et aux identifiants persistants.
- **EOSC Opportunity Area (OA) Expert Groups** : participation à la mise à jour de la

politique EOSC sur les identifiants pérennes « EOSC Pid policy ».

- **EOSC FAIR Impact** : participation à la rédaction d'un guide utilisateur sur les identifiants pérennes « User guidelines on EOSC PID implementation ».

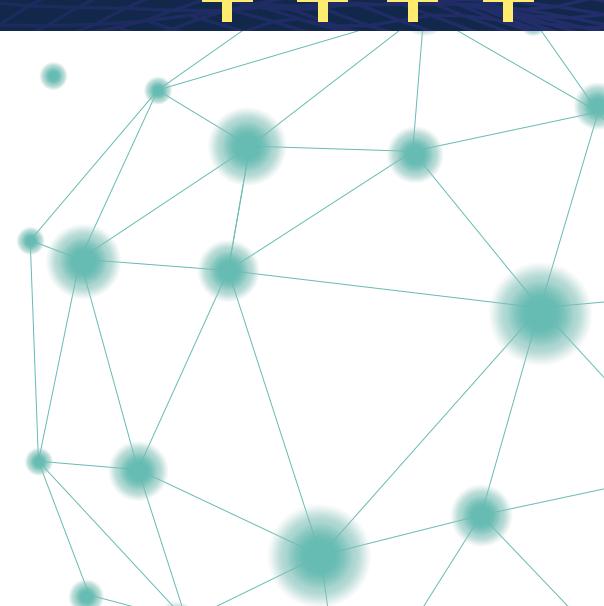
## Autour de CAT OPIDoR...

- Amélioration du **modèle de description des services dans Cat OPIDoR**, outil destiné à recenser les entrepôts de données. Cette amélioration vise à offrir une description plus claire et détaillée des différents entrepôts recommandés par le Comité pour la science ouverte (CosO). De plus, des filtres spécifiques ont été mis en place pour faciliter la recherche et la sélection des entrepôts en fonction de critères précis, comme le type de données ou les exigences de gestion.



Centre de ressources  
outils communs aux ateliers  
FR Recherche Data Gouv

- Participation à l'organisation du **séminaire Recherche Data Gouv**, un événement majeur dédié aux enjeux de la gestion et du partage des données de recherche en France, qui s'est tenu du 20 au 22 novembre 2024.
- Participation à l'administration, la curation et à l'accompagnement au dépôt dans l'**entrepôt CNRS Research Data**, plateforme développée par le CNRS pour l'archivage, la préservation et le partage des données de recherche.



## Et aussi...

Le service OPIDoR a accompagné le laboratoire Imopa dans le transfert des données de la plate-forme EsCoLa vers une nouvelle solution logicielle dédiée à la gestion des données scientifiques en laboratoire, LabCollector. Cet accompagnement a consisté à garantir la bonne migration des données scientifiques tout en assurant leur intégrité et leur accessibilité dans la nouvelle plate-forme.



## PUBLICATION

Miksa, T., Chodacki, J., Suchánek, M., Praetzellis, M., Papadopoulou, E., Jacquemot, M.-C., & Ashley, K. (2024). **Salzburg Manifesto on machine actionable Data Management Plans**.  
<https://doi.org/10.5281/zenodo.10658521>

# FORMATION - DORANUM

L'équipe Formation-DoRANum accompagne les chercheurs, ingénieurs et professionnels de l'information dans l'appropriation des bonnes pratiques en gestion et partage des données de la recherche. Elle propose des ressources pédagogiques variées (autoformations, webinaires, jeux sérieux, ateliers) accessibles via la plateforme DoRANum, ainsi que des formations adaptées aux besoins des communautés scientifiques.

## Un accompagnement sur mesure

En 2024, le service Formation-DoRANum de l'Inist a accompagné les ingénieurs plateforme de l'Institut CNRS Ingénierie dans leur montée en compétences sur la gestion des données de la recherche. Pour cela, un **parcours de formation en e-learning** a été mis à leur disposition sur la plateforme Callisto.

### Un parcours adapté aux besoins et à la demande

Le parcours de formation dédié à **CNRS Ingénierie** comprenait :

- Une quinzaine de ressources en autoformation ;
- Un jeu pédagogique dédié à la gestion des données ;
- Un webinar d'introduction à la gestion et au partage des données de la recherche ;
- Un test en début et en fin de parcours pour mesurer la progression des connaissances.

En complément, l'équipe a animé deux ateliers ludiques lors de l'Action nationale de formation (ANF) sur la gestion des plateformes de CNRS Ingénierie, qui s'est tenue à Rouen du 10 au 12 décembre. Ces ateliers, basés sur les jeux « **Gopen DoRe** » et « **Faut pas s'en FAIR** », ont permis aux participants d'aborder de manière interactive les enjeux du partage et de la gestion des données.



### Une plateforme e-learning enrichie et mise à jour

La plateforme e-learning DoRANum a continué de s'étoffer pour accompagner la communauté scientifique dans la gestion et le partage des données de recherche :

- **13 nouvelles ressources pédagogiques** ont été publiées, couvrant des thématiques essentielles pour mieux comprendre et appliquer les bonnes pratiques en gestion des données ;
- **2 ressources disciplinaires spécifiques** ont été ajoutées, permettant aux chercheurs de découvrir des recommandations adaptées à leur domaine : « **Principaux data journals en sciences de l'environnement** » et « **Une culture de la donnée en Sciences humaines et sociales est-elle possible ?** » (réalisée par l'Université de Lorraine) ;
- **40 ressources existantes** ont été **mises à jour** afin de garantir des contenus toujours pertinents.

### Des ressources pédagogiques en anglais pour un public international

Le service a élargi l'accessibilité de ses contenus en proposant des ressources pédagogiques en anglais sur la plateforme DoRANum. Cette initiative répond aux attentes des utilisateurs étrangers, notamment des doctorants. Pour cela, deux actions majeures ont été menées :

- **la traduction de 9 fiches synthétiques**, permettant d'aborder de manière claire et concise les principes clés de la gestion et du partage des données de recherche ;
- **l'ajout de sous-titres en anglais** sur les 9 vidéos « Minutes », courtes vidéos pédagogiques qui apportent des explications essentielles sur les bonnes pratiques en matière de gestion des données.

## Et aussi...

- **Organisation sur demande de webinaires** sur la gestion des données de la recherche pour IFSEM CNRS et pour les membres du programme Traces de l'UMR AUSser : « **Plan de gestion de données : bonnes pratiques de rédaction** » .
- Mise en place d'un nouveau webinaire « **Principes FAIR et cycle de vie des données** » et poursuite du webinaire « **Plan de gestion de données : bonnes pratiques de rédaction** ».
- **Dépôt sur Zenodo du jeu « Dép'Osez, le jeu sur les entrepôts de données** » co-conçu par le service formation-DoRANum et la DDOR : <https://zenodo.org/records/14234220>

## EN CHIFFRES...

**33 sessions de formations** organisées pour un total de **1139 participants**



## PUBLICATION

Antoine, D. (2024). **FAIR-by-Design Methodology Implementation Success Story (1.0.0)**. Zenodo.

<https://doi.org/10.5281/zenodo.13286232>

## Engagement continu dans le projet européen SKILLS4EOSC

L'Inist a poursuivi son implication dans le projet européen **SKILLS4EOSC** (2022-2025) aux côtés de la DDOR. Ce projet vise à renforcer les compétences en science ouverte à l'échelle européenne en développant des ressources et des cadres de formation adaptés. Plusieurs actions ont été menées :

- **Mise à jour du catalogue des profils de carrière en science ouverte** afin de mieux définir les rôles et compétences en science ouverte, avec un alignement en cours sur les ontologies science ouverte ;
- **Développement du cadre de reconnaissance pour les ressources d'apprentissage** afin d'établir des critères d'évaluation et de valorisation des ressources de formation en science ouverte ;
- **Développement du curriculum en science ouverte pour les data stewards** afin de leur fournir les compétences nécessaires pour accompagner les chercheurs dans la gestion et le partage des données de recherche ;
- **Conception de parcours d'apprentissage** sur la reproductibilité pour les data librarians et sur les licences pour les data stewards, avec des programmes spécifiques développés par l'Inist pour garantir la fiabilité des résultats scientifiques et permettre une meilleure compréhension et application des licences de diffusion des données de recherche ;
- **Application de la méthodologie « FAIR-by-design »** développée par Skills4EOSC pour accompagner la création de supports pédagogiques conformes aux principes FAIR : mise en œuvre dans la conception de plusieurs ressources d'apprentissage, puis élaboration d'un retour d'expérience structuré, contribuant ainsi à son amélioration par les auteurs ;
- **Développement d'un « advocacy kit »** (kit de plaidoyer) destiné à promouvoir la science ouverte auprès des communautés scientifiques ;
- **Reconnaissance de Recherche Data Gouv** en tant que « **Competence Centre** » de Skills4EOSC.

## Une première ressource co-conçue pour accompagner la gestion des données



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

Dans le cadre de Recherche Data Gouv, le service Formation-DoRANum de l'Inist a publié « **Le minimum à FAIR** », la première ressource issue d'une collaboration entre le groupe de travail 3 des Ateliers de la donnée et le Centre de ressources DoRANum.

Cette ressource a été conçue pour **sensibiliser et accompagner les chercheurs dans l'application des principes FAIR** (Facile à trouver, Accessible, Interopérable, Réutilisable) à leurs données de recherche. Elle propose des conseils pratiques et des recommandations essentielles pour rendre les données plus ouvertes et réutilisables, en accord avec les exigences de la science ouverte.

# EXPODOR

L'équipe *Exposition des données - ExpoDoR* accompagne les chercheurs et les institutions dans la structuration, la visualisation et la diffusion de leurs données scientifiques. Son rôle est de mettre en valeur ces données à travers des outils et des méthodes adaptés, facilitant ainsi leur réutilisation et leur compréhension.

## Un nouvel atelier pour numériser le patrimoine scientifique

ExpoDoR a élargi son activité en mettant en place un **atelier de photographie numérique en très haute définition**, spécialement conçu pour capturer des objets patrimoniaux et scientifiques. Ce nouveau service, lancé en juillet, a débuté avec la numérisation de 10 000 planches de l'herbier du Jardin botanique du Lautaret. Réalisé en partenariat avec l'Université Grenoble-Alpes, ce projet contribue à la valorisation des données de recherche sur la botanique alpine.

## 335 000 ressources numériques valorisées !

En 2024, le service ExpoDoR a créé **22 plateformes OMEKA-S** hébergeant **33 sites, 107 collections** valorisant **335 000 ressources numériques** correspondant à des données de recherche ou patrimoniales.

Ces prestations ont été réalisées pour le compte de 32 laboratoires de recherche, plusieurs MSH, OSU, Bibliothèques universitaires, écoles et consortium d'HumaNum.

Dans le cadre de la série vidéo « **Métiers d'Inist** », Edwige Pierot explique son métier de chargée de projet de valorisation des données de la recherche.

La vidéo a été mise en ligne le 27 septembre 2024 sur la chaîne Canal U de l'Inist :

<https://www.canal-u.tv/chaines/inist-cnrs>

Co-organisation des Journées du consortium 3D d'huma-Num (JC3DHN2024) à Nancy (Inist, Université de Lorraine) 26 au 28 novembre 2024



## QUELQUES EXEMPLES DE RÉALISATIONS EN 2024 :

### Théophile

Le projet Théophile, nommé d'après le moine allemand du XII<sup>e</sup> siècle, auteur d'un traité sur l'art et l'artisanat, vise à documenter et exposer trois collections de photographies mettant en valeur l'architecture et les arts monumetaux, en particulier ceux du Moyen Âge.

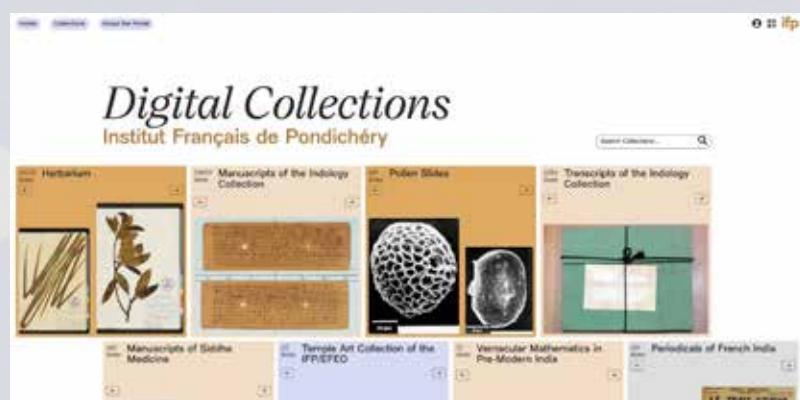
<https://theophile-centrechastel.inist.fr/>



### Digital Collections - Institut Français de Pondichéry

Ce portail regroupe et met à disposition les collections numérisées de l'Institut français de Pondichéry (IFP), issues de ses recherches depuis 1955. Il donne accès à un large corpus de ressources, allant des manuscrits en langues indiennes aux archives écologiques, tout en valorisant les projets de numérisation menés dans le cadre du programme Endangered Archives de la British Library.

<https://digitalcollections.ifpindia.org/>



### Documents de Recherche sur l'Architecture Antique – DoRAA

La plateforme DoRAA met en valeur les archives de l'Institut de recherche en Architecture antique (IRAA) en offrant un accès aux fonds de fouilles, études et publications sur l'architecture antique en France. Elle expose des documents numérisés issus des archives papier du laboratoire, permettant aux chercheurs et au public de consulter ces ressources.

<https://doraa-iraa.inist.fr/>



# Analyser et fouiller l'information scientifique





Prompt : Scène illustrant des outils d'analyse et de fouille de texte sur des écrans : visualisation de données textuelles, graphiques, nuages de mots, et traduction automatique. Avec des piles de documents numériques et algorithmes en action.

# TEXTES & CORPUS - ISTEX

L'équipe Textes & Corpus - Istex accompagne les chercheurs dans l'exploitation des ressources Istex pour leurs projets de fouille de textes. Elle facilite l'interrogation et la constitution de corpus grâce à des services adaptés et propose des formations pour maîtriser ces outils. Elle aide également à concevoir des corpus sur mesure, en fonction des besoins spécifiques de chaque projet.

## Lancement d'ISTEX Search pour simplifier l'accès aux ressources

L'Inist a enrichi l'offre de services Istex avec le lancement d'**Istex Search**, un nouvel outil visant à **simplifier l'accès aux ressources Istex** : il propose un parcours complet incluant l'interrogation du réservoir de données, l'exploration des résultats et le téléchargement de corpus.

Développé avec l'appui d'une société spécialisée en UX et web design, puis intégré par une entreprise de développement informatique, Istex Search améliore l'expérience utilisateur et facilite la constitution de corpus pour la recherche. Un dispositif d'accompagnement à son utilisation a également été mis en place.

## Réservoir Istex : 301 659 ressources numériques chargées en 2024

**Ressources numériques acquises auprès des éditeurs dans le cadre du programme CollEx-Persée :**

- Springer : 4 nouveaux bouquets d'ebooks et 1 complément d'ebooks en Sciences humaines et sociales (Europe centrale, orientale et balkanique, lot allemand – études germaniques – philosophie – droit et criminologie - sciences politiques) ;
- Taylor & Francis : 1 nouveau bouquet d'ebooks en Sciences humaines et sociales (Europe centrale, orientale et balkanique, lot anglais) ;

- EMS : 1 nouveau bouquet d'ebooks en Mathématiques ;
- Casalini : 1 complément d'ebooks en Sciences humaines et sociales ;
- Bentham : 1 nouveau bouquet de revues en Chimie.

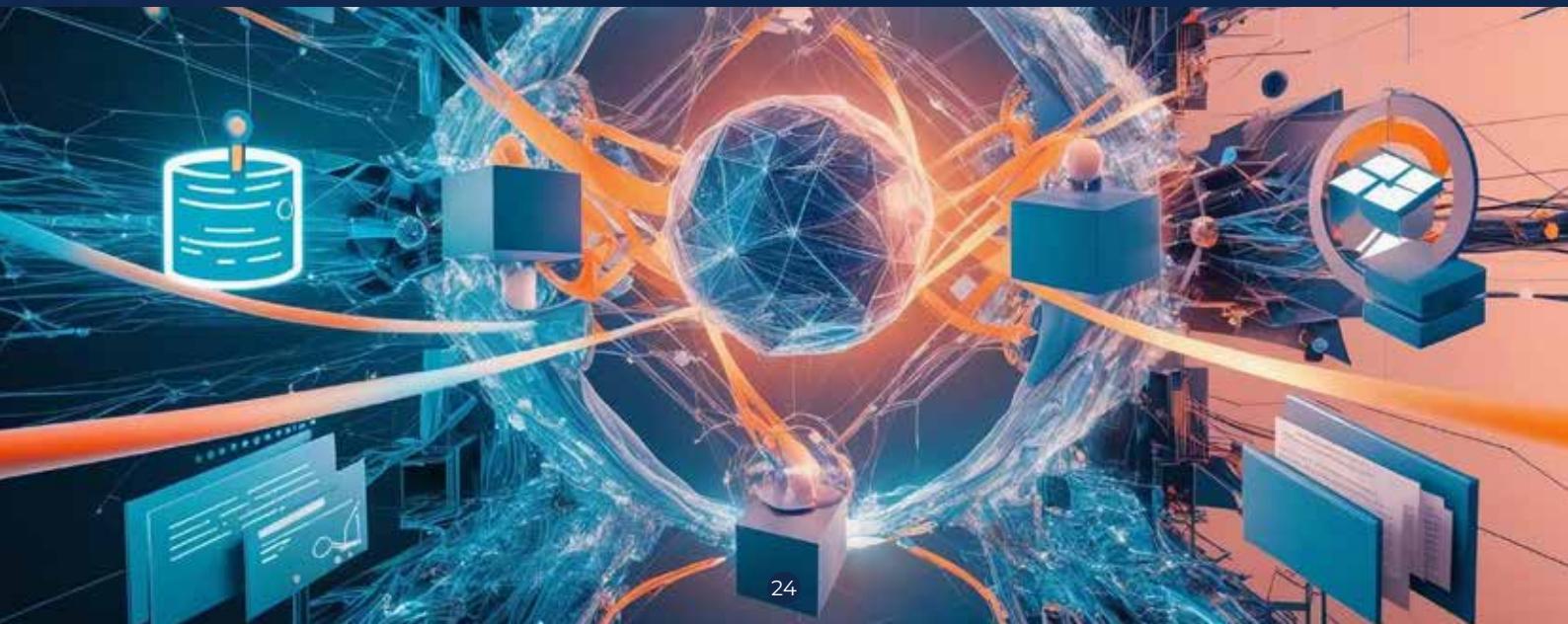
## Ressources numériques en Open Access moissonnées dans le cadre de la collaboration avec la société Science Miner :

PloS : 1 nouveau bouquet de revues en Sciences de la santé, Sciences de la vie et Sciences du développement durable.

## De nouveaux corpus scientifiques

Istex met à disposition des collections de corpus illustrant diverses exploitations de sa base documentaire dans des contextes scientifiques variés. Ils permettent d'analyser, d'extraire et de structurer des informations spécifiques afin de répondre à des problématiques précises.

**5 nouveaux corpus** ont été publiés sur le site [data.istex.fr](http://data.istex.fr) : Langues signées ; Biodiversité ; Supra-conducteurs ; Scientifiques célèbres ; Prix Nobel et médaillés Fields.



## Formations

**8 ateliers et interventions** sur la constitution d'un corpus spécialisé à partir des ressources Istex pour un usage en fouille de texte ont été organisés pour un total de **130 personnes formées** :

- **Enssib** : formation des étudiants du Diplôme de conservateur des bibliothèques à Villeurbanne ;
- **Écoles doctorales** : une nouvelle offre de formation proposée aux doctorants des universités de Haute Alsace, Lorraine, Paris 1 – Panthéon Sorbonne, Strasbourg, Lille ;
- **ANF-TDM** : présentation d'Istex lors de la 5<sup>e</sup> édition de l'ANF consacrée à l'exploration documentaire et l'extraction d'informations à Villejuif ;
- **Média Normandie** : formation des chercheurs et personnels d'appui à la recherche à Caen.

## Présentations

**28 séances** ont été animées à distance avec l'objectif de présenter l'écosystème Istex (son contenu, son usage, son accès, ses services associés pour l'interrogation, le téléchargement, la visualisation des données et la fouille de textes) :

- **16 séances à la demande des établissements ayant droit** (IFP Énergies nouvelles, Réseau des écoles françaises à l'étranger, Université de Montpellier, CY Cergy Paris Université, ArScAn, Anses, LAM, École Pratique des Hautes Études PSL, Université Paris Saclay, vetAgroSup, ICMCB, CHU de Caen, MPES) ou à la demande des services « Négociations et acquisitions » et « Portails et plateformes – BibCnrs » de l'Inist. Au total, **46 personnes** ont bénéficié de ces présentations.
- **12 séances sous forme de webinaires** ouverts à tous et sans inscription.

## Istex en chiffres

Au 31 décembre 2024, la **base Istex** comptait :

- **28 219 752 documents** au total ;
- **50 bouquets** de ressources numériques publiées entre 1455 et 2023 ;
- **9 541 revues / 439 061 monographies** ;
- **4,2 millions de documents** disponibles par ailleurs en accès libre (14,9 %).

Des ressources valorisées à travers **36 corpus scientifiques** prêts à l'emploi.

Les **données d'usage** de la plateforme en 2024 :

- **10 millions de documents** téléchargés (via Istex-Search, Istex-DL, Harvester, Click & Read, Google Scholar, outils de découverte, etc.) ;
- **61 % des utilisateurs** accèdent à Istex pour un usage TDM et 39 % pour un usage documentaire.

## LODEX

Bien plus qu'un simple outil de visualisation, Lodex permet le traitement et l'analyse de données, leur enrichissement ainsi que leur visualisation et leur publication sous forme de représentations graphiques.

### Création d'une liste de diffusion

Une liste de diffusion dédiée à la communauté Lodex a été créée afin de **favoriser les échanges** et de **diffuser les informations** relatives aux évolutions de l'outil. Elle permet aux utilisateurs d'être informés des mises à jour et nouveautés, de poser leurs questions et de partager leurs retours d'expérience.

Pour rejoindre la communauté Lodex : <https://groupes.renater.fr/sympa/info/lodex>

# RESSOURCES LANGAGIÈRES MULTILINGUES

L'équipe dédiée accompagne la communauté scientifique en facilitant l'accès à des ressources terminologiques et linguistiques adaptées à leurs besoins. À travers le site Loterre, le service met à disposition des bases de données terminologiques conformes aux principes FAIR. En complément, l'équipe apporte un appui linguistique en assurant la traduction et la relecture de documents scientifiques et aussi de supports de présentation.

## Intégration de Loterre dans Istex

Le site Loterre (Linked open terminology resources) a rejoint l'infrastructure Istex, **renforçant ainsi l'accès aux ressources numériques** pour les chercheurs. Loterre, spécialisée dans les vocabulaires scientifiques et les bases de connaissances, complète les services d'Istex dédiés à la fouille de textes et à l'analyse de documents.

Cette intégration facilite la recherche d'informations et améliore l'exploration des contenus scientifiques.

## TERMINOLOGIE

**Deux nouveaux services** ont été mis en place :

- **Vocab Explorer** : un catalogue de ressources terminologiques accessibles en ligne (entrepôts ou répertoires de terminologies, vocabulaires contrôlés, glossaires, lexiques, dictionnaires, ontologies, thésaurus...) à explorer et à filtrer par domaines, thématiques, type de ressources, langues...
- **Loterre Explorer** : un service qui propose une vision analytique des ressources terminologiques exposées dans Loterre. Il permet d'explorer et de filtrer les ressources de Loterre par domaines, thématiques, langues, producteurs.

## Mise en chantier de quatre nouvelles ressources

Dans le cadre du développement de nouvelles ressources terminologiques, quatre projets de théaurus ont été engagés en 2024, en collaboration avec plusieurs partenaires, afin de structurer et valoriser des vocabulaires spécialisés dans différents domaines de recherche :

- **Thésaurus IA** en collaboration avec le GDR RADIA : collecte de termes à partir de divers lexiques et glossaires, et de références de revues fournies par le GDR, pour la réalisation d'une ressource terminologique structurée sur l'intelligence artificielle ;
- **Thésaurus Math-Code-Log** en collaboration avec le GT Maths : constitution d'un théaurus adapté aux données, codes et logiciels pour pouvoir appliquer des mots-clés pertinents à des codes et logiciels afin de faciliter les recherches dans Recherche Data Gouv ou dans toute autre plateforme de dépôt de jeux de données, codes et logiciels ;
- **Thésaurus de primatologie** en collaboration avec les primatologues de la Revue de Primatologie : extraction terminologique à partir de la Revue de Primatologie, curation de la liste de termes (environ 700 après curation), et première structuration à partir de cette collecte de termes ;
- **Thésaurus TAL** en collaboration avec le CLILLAC-ARP, l'ISIR et l'Inria dans le cadre du Projet ANR MaTOS : préparation d'une version en preuve de concept d'un théaurus bilingue en traitement automatique des langues (TAL), destiné à améliorer un système de traduction automatique spécialisé.

## Loterre en chiffres...

**72 terminologies ont été mises en ligne et 7 théaurus ont été mis à jour** (Pays et subdivisions ; Science ouverte ; Psychologie cognitive de la mémoire humaine ; ThésoTM ; Mathématiques ; Communes de France ; Paléoclimatologie)

## TRADUCTION

Les traductions les plus marquantes de l'année :

**Anglais vers Français** : à la demande de la DDOR, le service a traduit un rapport de 25 082 mots rédigé par l'entité Knowledge Exchange, intitulé *Approaches to scaling up reproducibility in research organisations*. Ce rapport explore les méthodes qui pourraient permettre de favoriser la reproductibilité au sein des organismes de recherche dans un contexte de science ouverte.

**Français vers Anglais** : à la demande de la Direction de la communication du CNRS, le service a traduit 77 documents dont le Rapport d'activité annuel du CNRS. Il a également traduit le document *Positionnement stratégique de CNRS Sciences Humaines & Sociales* qui explicite la stratégie de positionnement de CNRS Sciences humaines & sociales pour accompagner la recherche fondamentale sur les individus et les sociétés humaines et renforcer sa capacité à traiter des enjeux qui les traversent.

## Volumes pour les activités de traduction et de relecture :

**Français -> Anglais** (175 documents) : traduction = 250 219 mots ; relecture = 138 662 mots

**Anglais -> Français** (14 documents) : traduction = 105 734 mots



## PUBLICATIONS

Gilles Banzet, Franck Bassinot, Henri Bretel, Valérie Daux, Ludovic Hamiaux, et al. *Le théâtre de paléoclimatologie : création et perspectives d'utilisation*. 2024. hal-04555209

Frank Arnould : Co-édition du numéro spécial « Words of memory » de la revue Lexis.



# TEXT & DATA MINING (TDM)

L'équipe TDM accompagne les chercheurs dans l'exploitation automatisée des textes et des données scientifiques. Elle développe des outils et des méthodologies sur mesure pour faciliter l'extraction d'informations, l'analyse de corpus et la fouille de textes. Ce service s'appuie sur des technologies avancées pour optimiser l'exploitation des ressources documentaires.

## De nouveaux web-services pour l'analyse avancée de corpus

Le service TDM a développé de nouveaux web-services asynchrones, permettant de traiter des corpus dans leur ensemble. Ces services facilitent des tâches telles que le clustering, l'extraction de thématiques, la détection de données bruitées, l'extraction de terminologie et l'identification de références phares.

Fin 2024, **39 web services** étaient en ligne (dont 11 nouveaux).

## Ça s'est passé en 2024...

- **Intégration des services TDM dans Istex** : les outils de traitement des données textuelles (TDM) ont été intégrés dans l'infrastructure Istex, facilitant ainsi leur utilisation et leur gestion par les utilisateurs.
- **IA-Factory** : Des web-services de traitement de corpus ont été développés et rendus accessibles via IA-Factory, une interface de chargement de corpus et d'exécution d'outils de TDM disponible sur Istex, permettant ainsi une utilisation simplifiée des outils d'intelligence artificielle.
- **Données de validation de web-services** : des corpus de données annotées ont été réalisés afin de tester et évaluer la qualité des web-services de manière méthodique.
- **Projet RAPIDO** : dans le cadre de ce projet qui vise à développer des outils automatisés pour l'extraction et l'alignement de données dans des corpus scientifiques, les pro-

grammes d'extraction et d'alignement ont été finalisés. Ils ont ensuite été déployés pour des tests sur la plateforme Persée, permettant d'évaluer leur efficacité et leur intégration aux flux de traitement documentaire.

- **Aide à l'utilisation de web-services** : le site Istex-TDM a été enrichi avec des fiches de description détaillées et des guides d'utilisation, facilitant l'accès aux services de traitement de données textuelles.

## Et aussi...

- **Co-Organisation du Workshop InCr-Learn'24 @ ICDM 2024** (International Conference on Data Mining), Abu Dhabi, 9 décembre 2024 (Pascal Cuxac) ;
- **Participation à 10 conférences/rencontres** qui ont donné lieu à des publications (voir p. 35-37).

## Parmi les nouveaux web-services :

### TopRefExtract, pour analyser l'influence des publications

Intégré à l'infrastructure Istex, TopRefExtract facilite l'analyse des citations dans un corpus scientifique. En entrant une liste de DOI, l'outil interroge l'API OpenAlex afin d'extraire et classer les références citées. Il met en avant les publications les plus citées et identifie les articles qui les mentionnent.

### BibCheck, pour détecter les références bibliographiques invalides

Ce web-service a été conçu pour aider à identifier les références bibliographiques invalides dans un corpus scientifique. Que les références soient rétractées ou générées par des intelligences artificielles (hallucinations), BibCheck permet de les repérer facilement.



# APPUI AU PILOTAGE SCIENTIFIQUE

L'équipe dédiée accompagne les établissements de l'ESR dans l'analyse et la valorisation de leurs données stratégiques. Il propose des outils et des expertises pour exploiter efficacement les données bibliométriques, scientifiques et institutionnelles. Grâce à ces analyses, les décideurs disposent d'indicateurs fiables pour orienter leurs stratégies et évaluer l'impact de leurs actions.

## Exploration de nouvelles bases de données bibliographiques

Le service a testé de nouvelles bases bibliographiques ouvertes. Il a évalué l'usage de Conditor pour une étude bibliométrique sur la production scientifique française en accès libre et exploré OpenAlex pour analyser la production d'un laboratoire, d'un institut du CNRS ou d'une thématique. Ces travaux ont abouti à une méthode partagée, incluant **un modèle et un mode opératoire simplifié** pour exploiter OpenAlex avec Lodex.

## Parmi les autres temps forts du service...

- **Contribution aux Expertises Scientifiques Collectives (ESCo)** : participation à l'ESCo sur les impacts des parcs éoliens en mer (co-pilotée par le CNRS et l'IFREMER) et à celle sur les terres rares (pilotée par le CNRS), en appui au service Textes & Corpus - Istex ;
- **Quantification des publications scientifiques en accès libre** produites par le CNRS en 2022 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 » ;

- **Création d'un tableau de bord dynamique avec Lodex** pour visualiser les résultats de l'enquête APC coordonnée par le consortium Couperin.

## Quelques exemples d'études bibliométriques :

- Suivi du **poids du libre accès dans la production scientifique du CNRS pour l'année 2022** à partir des bases du Web of Science (DDOR - Direction des données ouvertes de la recherche) ;
- Suivi de la **production scientifique des unités de recherche de plusieurs instituts du CNRS** à partir des bases du Web of Science : CNRS Biologie, CNRS Écologie & Environnement, CNRS Ingénierie ;
- Suivi de la **production scientifique du CNRS pour la période 2017-2023** à partir de la base Conditor (DDOR - Direction des données ouvertes de la recherche) ;
- Suivi de la **production scientifique des unités de recherche** des instituts du CNRS à partir de la base Conditor ;
- Suivi de la **production scientifique de plusieurs unités de recherche du CNRS : GANIL** (Grand accélérateur national d'ions lourds), **IRIT** (Institut de recherche en informatique de Toulouse) ;
- Suivi de la **production scientifique de plusieurs infrastructures de recherche** : Infrastructure de recherche littorale et côtière (ILICO), Analyses et expérimentations sur les écosystèmes continentaux (AnaEE France) ;
- Suivi de la **publications scientifiques françaises**, ou produites par les chercheurs du GDR Sport, **dans le domaine du sport et de l'activité physique** pour la période 2013-2024 (Groupe de recherche sports et activités physiques) ;
- Étude approfondie de près de **9 000 publications sur la paléoclimatologie et le paléoenvironnement** (2000-2020) issues de la base WoS, à la demande du LSCE et de l'IGE, contribuant ainsi à l'élaboration d'un livre blanc dédié.

# Et aussi...



Prompt : Illustration représentant l'activité de communication, veille et web dans un univers d'information scientifique et technique.

# COMMUNICATION, WEB, VEILLE

L'équipe conçoit et met en œuvre les actions de communication interne et externe, en accompagnement des projets et des services de l'unité. Elle veille à valoriser les activités de l'Inist, à diffuser l'information de manière cohérente et à produire des supports adaptés aux publics cibles.

## Fin de l'utilisation de Twitter

L'Inist a décidé de mettre fin à l'utilisation de son compte Twitter fin décembre afin de rester en cohérence avec ses **valeurs institutionnelles**, notamment en matière d'éthique et de responsabilité numérique. Ce choix s'inscrit dans une démarche plus large de vigilance à l'égard des plateformes dont les pratiques ne garantissent pas un cadre de confiance et de modération adapté.

## Parmi les événements de l'année...

Le service a participé à plusieurs événements en assurant une présence sur stand, notamment lors de la **Journée des nouveaux entrants organisée** par la Délégation Régionale Centre-Est (avril), des **Journées ABES** (mai), et du **congrès de l'ADBU** (octobre).

Il a également organisé et accueilli dans les locaux de l'Inist deux manifestations nationales : les **Journées du consortium DataCite France** (juin) et le **séminaire Recherche Data Gouv** (novembre).

Par ailleurs, une **visite des étudiants de Médial** (centre de formation aux métiers des bibliothèques Grand Est) a été organisée en février, dans le cadre de leur diplôme universitaire « Préparation aux métiers et concours des bibliothèques ».

## Présence sur les réseaux sociaux :



## Création de deux sites web

### Kit de design pour Istex

<https://design.istex.fr/>

Éléments de conception préconçus et regroupés pour aider à l'application de la charte graphique Istex.

## En vidéo

Deux nouvelles vidéos de la série « **Les métiers de l'Inist** » ont été réalisées pour valoriser les métiers et projets portés par l'Inist :

- PROFESSION #6 : Edwige, chargée de projet de valorisation des données de la recherche ;
- PROFESSION #7 : Fred, chargé de la mesure des usages et des relations utilisateurs pour Readmetrics.

De plus, des vidéos courtes et didactiques, ont été produites dans la série « **Comment faire avec Lodex ?** » afin de faire découvrir pas à pas les fonctionnalités de Lodex.

Les vidéos sont visibles sur la [chaîne Canal U de l'Inist](#).

## L'Inist veille...

**LaLIST**, site de veille scientifique et technique animé par l'Inist depuis 2012, a franchi le cap des 15 000 billets publiés, confirmant son rôle de ressource de référence pour la communauté scientifique.

Dans le cadre du site **Ouvrir la science**, piloté par le Comité pour la science ouverte, l'Inist poursuit ses activités de veille en collaboration avec le ministère de l'Enseignement supérieur et de la Recherche pour l'alimentation en contenus.

### Consortium DataCite France

<https://datacite-france.fr/>

Valorisation du consortium, présentation de ses objectifs et missions, diffusion des travaux et activités des groupes de travail.

## Quelques chiffres...

- LinkedIn Inist : 3400 abonnés (+ 600 abonnés en 2024 !) ;
- 899 billets publiés en 2024 sur LaLIST ;
- 1069 abonnés aux newsletters LaLIST.

# SECRÉTARIAT GÉNÉRAL

L'équipe dédiée assure la gestion administrative, financière et humaine de la structure, en supervisant le budget, les dépenses, les ressources humaines et les archives. Elle veille également à l'entretien des bâtiments pour garantir un environnement de travail fonctionnel et sécurisé.

## Évolution des effectifs

En 2024, 13 personnels permanents ont quitté l'Inist, dont 9 pour un départ à la retraite. Dans le même temps, 12 personnels permanents ont rejoint l'unité, parmi lesquels 7 recrutés par concours externe. L'appui des personnels contractuels demeure significatif pour la réalisation des missions, avec 9 agents en poste au 31 décembre.

À cette date, l'Inist comptait **150 agents** pour 141 équivalents temps plein travaillé (ETPT).

## Répartition des agents par domaine d'activité

Production, diffusion  
des savoirs

59,3 %

Informatique et  
développements

24,7 %

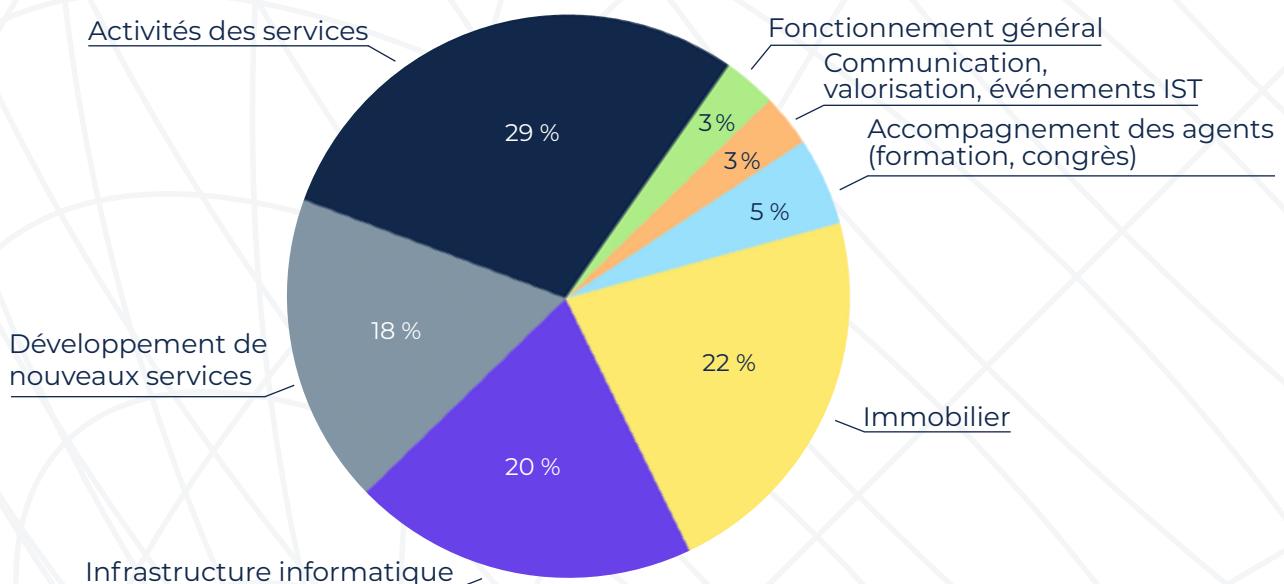
Direction, pilotage,  
communication

16 %

## Données budgétaires

Le budget exécuté en 2024 s'élève à **5,9 millions d'euros**, dont 4,5 millions dédiés aux acquisitions pour le réseau BibCnrs. Hors acquisitions, les dépenses, d'un montant de 1,4 million d'euros, se répartissent principalement entre quatre grands postes : le développement des services (18 %), l'accompagnement de l'activité (29 %), les infrastructures informatiques (20 %) et les coûts liés à l'hébergement (22 %). Par ailleurs, 5 % du budget sont consacrés à l'accompagnement et à la montée en compétences des agents, hors coûts liés aux missions.

## Répartition des dépenses hors acquisitions



# INGÉNIERIE DES SYSTÈMES D'INFORMATION

L'équipe veille à la mise en œuvre, au déploiement et à la continuité opérationnelle des infrastructures et outils numériques de l'Inist, en s'appuyant sur des compétences complémentaires allant du développement applicatif à la gestion des systèmes et de la sécurité, jusqu'au support aux utilisateurs.

## Préparation de la migration

L'année a été marquée par la préparation de la migration du système d'information de l'Inist vers le **Datacenter mutualisé lorrain** (DCML), une étape stratégique pour la modernisation et la sécurisation de l'infrastructure informatique. Ce projet a mobilisé l'ensemble de l'équipe autour de la planification du déménagement des équipements, de la transition vers une solution de stockage en location, et de la coordination avec les partenaires techniques du DCML.

## Lancement du projet GitLab-K3S

Le service a lancé un nouveau projet pour **moderniser ses outils informatiques** utilisés au quotidien par les équipes techniques. Il s'agit d'installer GitLab, une plateforme de travail collaboratif pour les développeurs, sur une infrastructure plus légère et plus facile à gérer appelée K3S. Ce changement permettra de mieux automatiser certaines tâches et de rendre les services numériques plus fiables et plus simples à faire évoluer. Ce projet s'inscrit dans une démarche d'amélioration continue des outils mis à disposition par l'Inist.

## Évolution continue des machines et services d'infrastructure

L'année a été rythmée par l'évolution constante des machines et des services d'infrastructure, dans un souci d'adaptation aux besoins et de maintien en conditions opérationnelles. Plusieurs chantiers ont été menés, notamment l'optimisation du service SMTP mail, la mise en place ou la mise à jour d'accélérateurs pour améliorer les performances, ainsi que l'automatisation de la gestion des certificats de sécurité.





Prompt : Illustration montrant des conférences et stands autour des données scientifiques.

# À VOTRE RENCONTRE...

## 23 JANVIER

Dijon

**Conférence Extraction et Gestion des Connaissances (EGC)**

Atelier TextMine

Pascal Cuxac

hal-04419844

Détection d'entités nommées géographiques par réseau de neurones récurrents

Lucas Anki, Léo Gaillard, Justine Revol

hal-04461354

**Défi TextMine'24 : Reconnaissance d'entités géographiques dans un corpus des Instructions nautiques**

Pascal Cuxac

hal-04434981

**Extraction d'acronymes torturés dans la littérature scientifique**

Pascal Cuxac

hal-04426448

## 19 FÉVRIER

Édimbourg

**IDCC workshop on interoperability**

Elevating data management planning : interoperability of RDM services through machine-actionability

Marie-Christine Jacquemot

## 18 MARS

Paris

**Journée des négociateurs Couperin**

**En savoir plus sur PANIST**

David Castelain, Dolorès Dardaine

## 27 - 29 MARS

Aix-en-Provence

**Rencontres annuelles Frantiq organisées par la**

## Fédération et ressources sur l'Antiquité

Table ronde « Les outils Frantiq dans l'environnement numérique de l'Enseignement supérieur et de la Recherche »

Alain Zasadzinski

## 28 MARS

Visioconférence

**Séminaire des correspondants Huma-Num**

Présentation de la ressource pédagogique DoRANum « Exemples d'entrepôts de données en SHS »

Lynda Kehli

## 4 AVRIL

Visioconférence

**Journée « Cycle de vie / Gestion des données de la biologie »**

DMP OPIDoR : sur le chemin de FAIR

Marie-Christine Jacquemot

## 14 MAI

Visioconférence

**Webinaire pour Média Normandie**

Présentation de Lodex

Mahafaka Ranoarisoa

## 16 MAI

Visioconférence

**Webinaire « Recherche Data Gouv et EOSC »**

Présentation du projet SKILLS4EOSC

Gabriela Torres Ramos, Diane Antoine, Claire Sowinski

## 28-29 MAI

Montpellier

**Journées ABES**

**Stand**

Cécilia Fabry, Léa Antoine, Pascale Viot

## 28-30 MAI

La Mure

**Séminaire CHRONOSPEDIA**

Chronospedia : une plateforme pour partager le savoir horloger : mise en ligne et archivage

Alain Zasadzinski, Titouan Boudart

## 28-31 MAI

Nancy

**Congrès INFORSID**

La fouille de textes en IST : les outils Istex-TDM

Pascal Cuxac  
hal-04597734

## 2-5 JUIN

Athènes (Grèce)

**World Conference on Research Integrity (WCRI 2024)**

Mining tortured acronyms from the scientific literature

Pascal Cuxac  
hal-04311600v1

## 4-6 JUIN

Poitiers

**3<sup>e</sup> séminaire de Recherche Data Gouv**

Ateliers pour l'élaboration de profils disciplinaires

Anne Busin, Françoise Cosserat, Marie-Christine Jacquemot, Paolo Laï, Laurent Rassinoux

Atelier « Quelle implication dans EOSC pour Recherche Data Gouv ? »

Gabriela Torres-Ramos, Diane Antoine

Atelier de présentation des étapes d'ingénierie pédagogiques du Centre de ressources DoRANum

Gabriela Torres-Ramos, Pierrette Paillassard, Diane Antoine

## 6 JUIN

Visioconférence  
**Centre National Universitaire de Documentation Scientifique et Technique de Tunisie**

Présentation du Consortium DataCite France pour le consortium tunisien  
Isabelle Cao

## 12-24 JUIN

Université Paris-Saclay  
**Rencontres annuelles CasuHAL**  
Poster de présentation HAL-CNRS  
Théo Rousselle, Séverine Koehl



## 13 JUIN

Visioconférence  
**Webinaire pour Média Normandie**

À la découverte du TDM  
Valérie Bonvallot

## 18-19 JUIN

Vandœuvre-lès-Nancy  
**Journées du consortium Datacite France**  
Animation des ateliers des groupes de travail DOI et autres identifiants, Communication, Métadonnées, Formation, Modèle économique  
Valérie Warth, Paolo Laï, Marie-Christine Jacquemot, Isabelle Cao, Françoise Cosserat, Claire Sowinski

## 26-28 JUIN

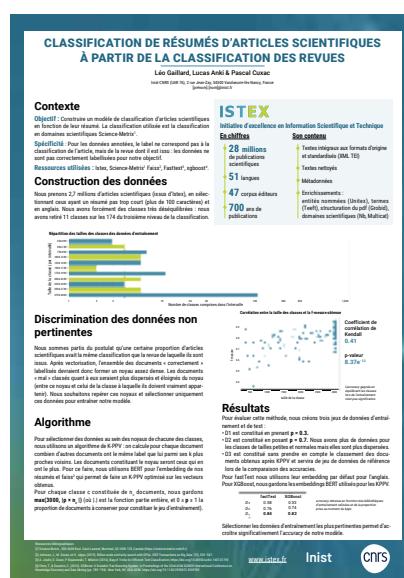
Lyon  
**12<sup>e</sup> journées du réseau Médici**

L'IA pour contrer l'IA : détection de citations peu fiables dans les articles scientifiques  
Pascal Cuxac  
hal-04884535v1

## 11-13 SEPTEMBRE

Marseille  
**29<sup>e</sup> Rencontres de la Société Francophone de Classification (SFC)**

Classification de résumés d'articles scientifiques à partir de la classification des revues - poster  
Léo Gaillard, Lucas Anki, Pascal Cuxac  
hal-04702632



## 3-4 OCTOBRE

Villejuif  
**ANF TDM 2024, Exploration documentaire et extraction d'informations**

La constitution et l'exploration d'un corpus documentaire à partir des ressources Istanx  
Mathilde Huguin

Des web services dédiés à la fouille de textes  
Valérie Bonvallot, Justine Revol

## 4 OCTOBRE

Mulhouse  
**Journée Mondiale de la Traduction**

Approches et outils de la terminologie scientifique  
Jean-François Nominé, Manon Delorme

Atelier « Terminologie et outils »

Jean-François Nominé, Manon Delorme

## 4-11 OCTOBRE

Visioconférence  
**Journées RDA France**

Atelier « OPIDoR au service des données de la recherche : focus sur les nouveautés de DMP OPIDoR »

Anne Busin, Françoise Cosserat, Laurent Rassinoux

Atelier « OPIDoR au service des données de la recherche : focus sur les entrepôts dans Cat OPIDoR »  
Anne Ciolek, Sidonie Raffy, Valérie Warth

## 24-25 OCTOBRE

Paris Nanterre  
**Rencontre PROFIL 1 – Les images-preuves de crime de guerre comme « données de recherches »**

Exposer et valoriser des données de recherche avec Omeka S

Alain Zasadzinski, Louis Maillard

Atelier-formation à l'outil OMEKA-S pour la plateforme en construction PROFIL  
Équipe Omeka

## 20-22 NOVEMBRE

Nancy  
**4<sup>e</sup> séminaire de Recherche Data Gouv**

Ateliers pour l'élaboration de profils disciplinaires

Anne Busin, Françoise Cosserat, Marie-Christine Jacquemot, Paolo Laï, Laurent Rassinoux

**Atelier pour la description des Centres de référence thématiques dans Cat OPIDoR**

Anne Ciolek, Sidonie Raffy

**Atelier Curation dans l'espace générique**

Françoise Tisserant

**Ateliers ludiques « Faut pas s'en FAIR »**

Coralie Wysoczynski, Léa Poirot, Lynda Kehli, Claire Sowinski

**Atelier « Stratégie internationale de Recherche Data Gouv : actions concrètes »**

Gabriela Torres-Ramos, Claire Sowinski

**Atelier « Présentation du logigramme interactif » avec le Centre de ressource établissement INRAE**

Pierrette Paillassard

**Présentation des « Outils ludiques réutilisables pour accompagner à la gestion des données » dans le cadre d'un barcamp**

Équipe DoRANum

**20-22 NOVEMBRE**

Lyon

**Journées Omeka & NumaHOP**

**Reprise de données patrimoniales : de Dspace à Omeka S associant IIIF et Nakala**

Louis Maillard, Bernard Sampité

**26-28 NOVEMBRE**

Nancy

**Journées du Consortium 3DHN**

**Accompagnement à la gestion des données de la recherche par l'Inist : DoRANum et OPIDoR**

Claire Sowinski, Marie-Christine Jacquemot

**Atelier « Plan de gestion de données pour la 3D »**

Alain Zasadzinski, Lauriane Locatelli

**27-28 NOVEMBRE**

Aubervilliers

**Journées annuelles 2024 du Cluster 1 Biblissima+ : « Des outils IIIF pour valoriser et explorer les collections en ligne »**

**Omeka S et IIIF**

Titouan Boudard, Louis Maillard

**29 NOVEMBRE**

Vandœuvre-lès-Nancy

**Journée scientifique « Concepts et taxonomies de la mémoire : de la philosophie à l'ingénierie des connaissances »**

**CogMemo : un thésaurus des concepts de la psychologie cognitive de la mémoire humaine**

Frank Arnould

**L'ontologie formelle comme cadre taxinomique de la mémoire : l'exemple de BFO (Basic Formal Ontology)**

Frank Arnould

**3-4 DÉCEMBRE**

Vandœuvre-lès-Nancy

**ANF B@O Renatis - La boîte à outils du réseau Renatis**

**Présentation de l'ANF**

Sylvie Gresillaud

**Atelier API Rest**

Philippe Houdry

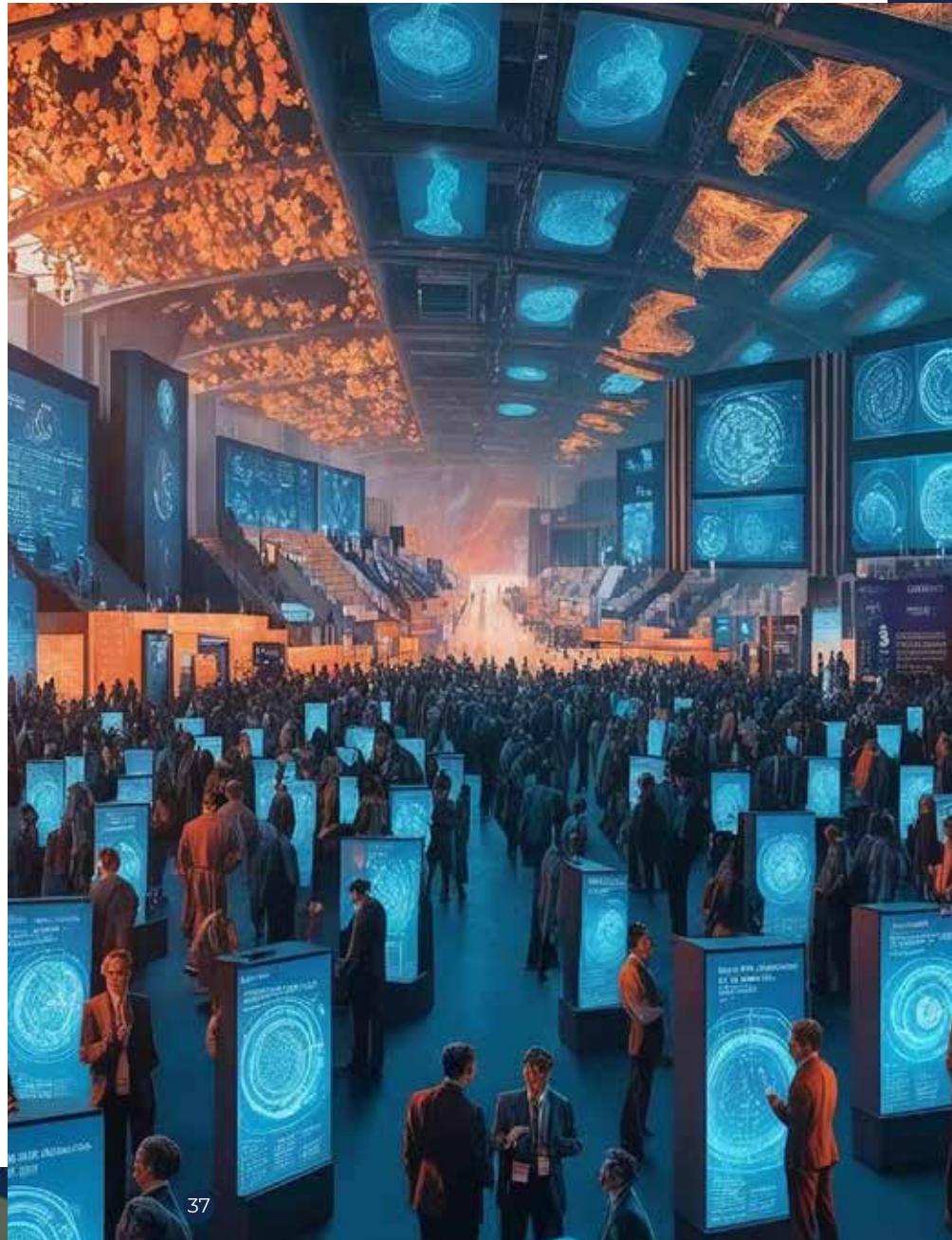
**10 DÉCEMBRE**

Rouen

**ANF gestion des données de CNRS Ingénierie**

**Ateliers ludiques « Gopen DoRe et Faut pas s'en FAIR »**

Coralie Wysoczynski, Lynda Kehli



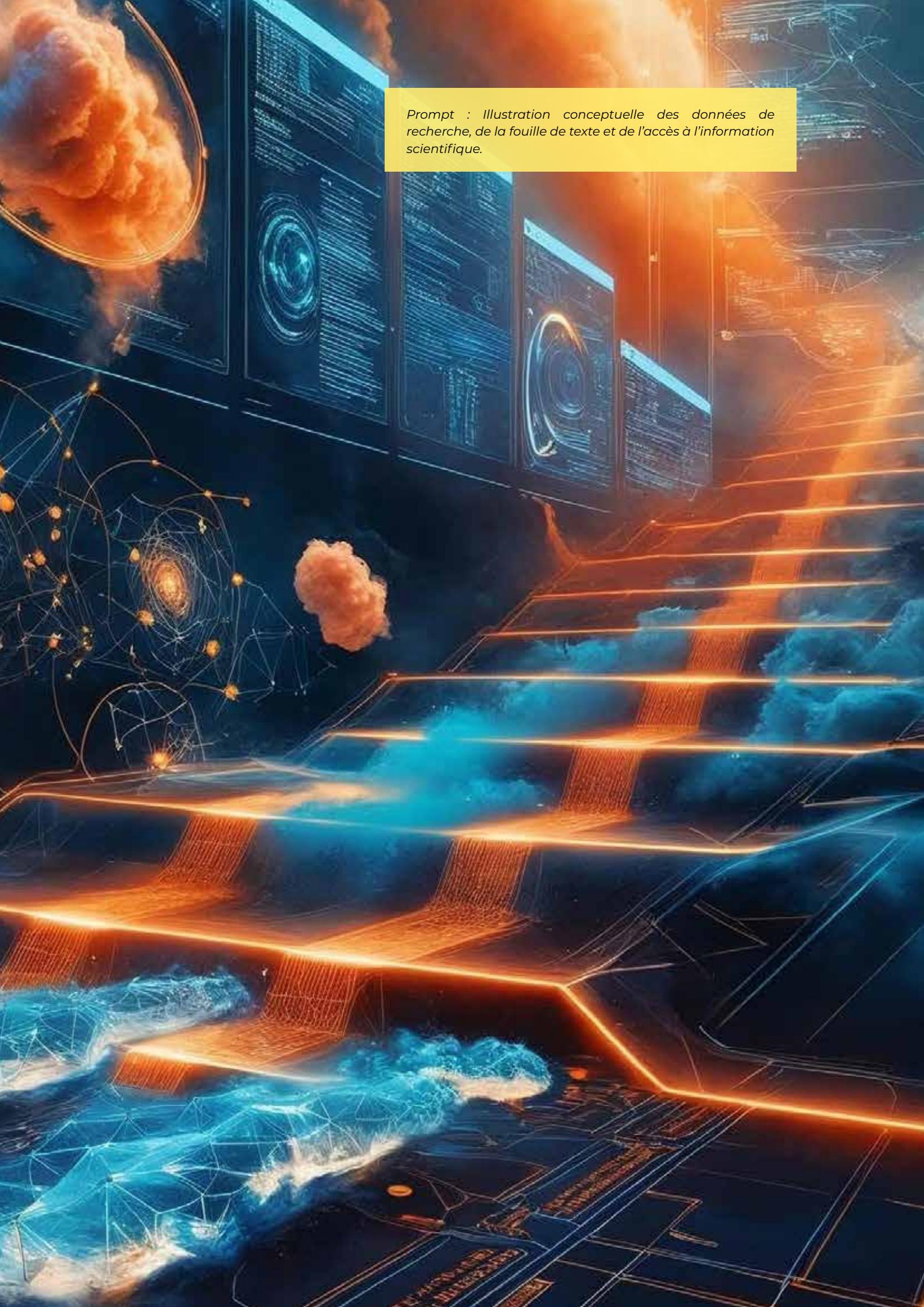
Ce rapport d'activité a été réalisé par le service Communication de l’Inist, en lien avec l’ensemble des services de l’Institut.

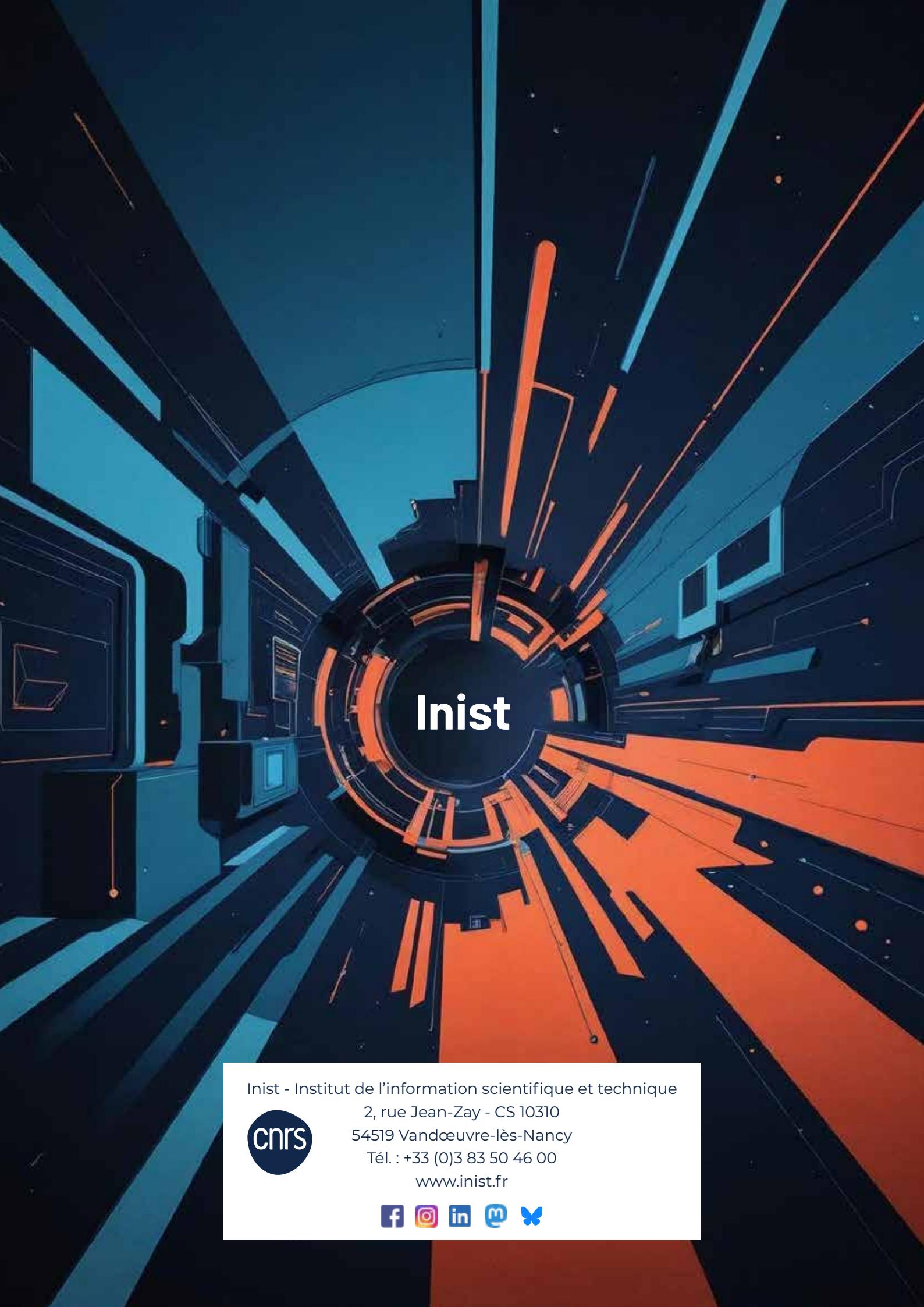
La plupart des images ont été générées avec IA Canva.

Conception graphique : Alexandra Petitjean-Monnin



*Prompt : Illustration conceptuelle des données de recherche, de la fouille de texte et de l'accès à l'information scientifique.*





# Inist

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

