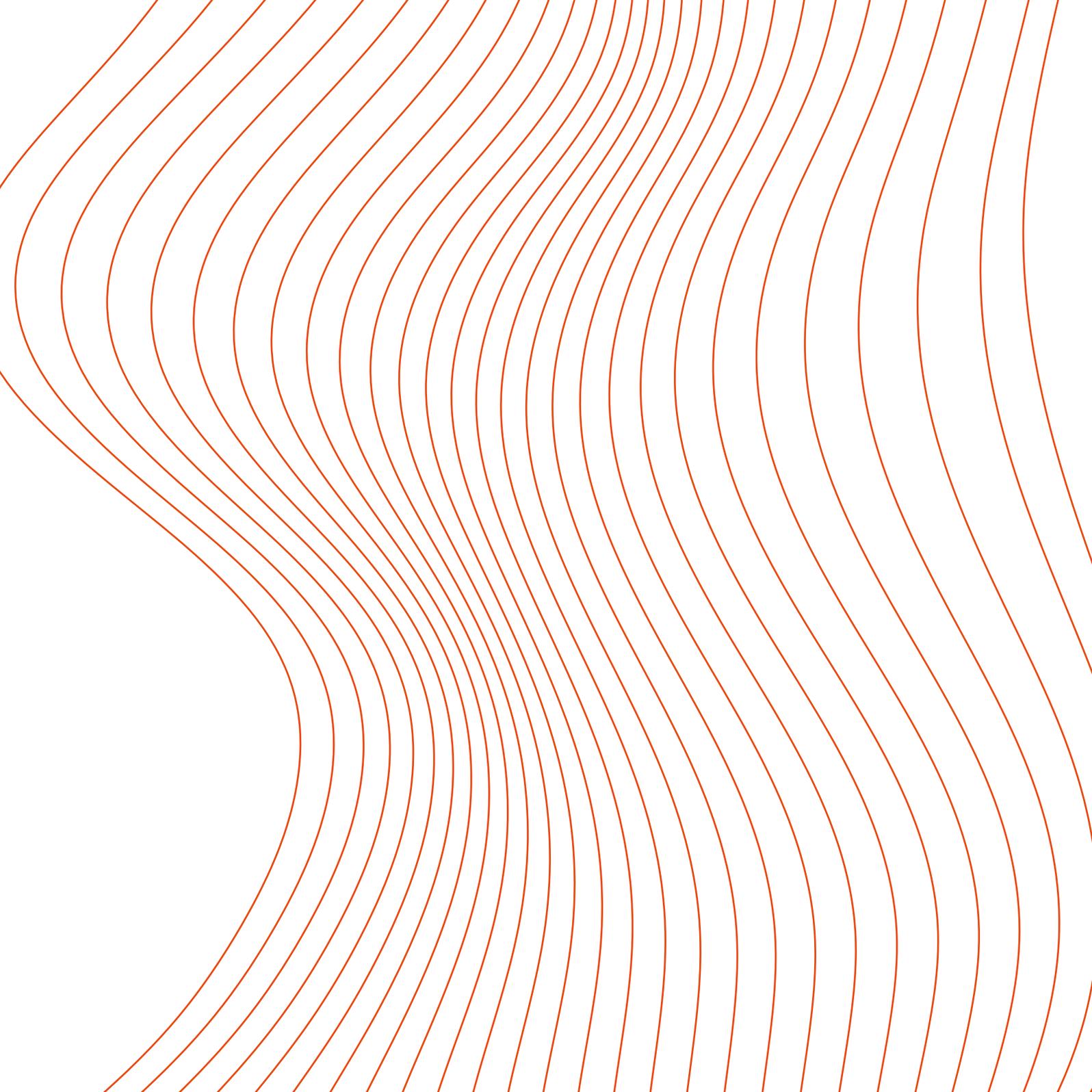


# L'odyssée de HAL

## 2001 – 2021





# Raconter l'aventure HAL

Lorsque HAL a été fondée en 2001, le mouvement du libre accès était balbutiant et un des premiers textes de référence de son histoire, l'initiative de Budapest, n'avait pas encore inscrit l'auto-archivage comme une des stratégies pour le réaliser.

Comment raconter les vingt années de HAL alors que son développement traduit aussi l'essor formidable de la science ouverte, dans les pratiques comme dans les politiques ? Avec près d'un million de documents disponibles

et plus de 130 portails d'institutions actifs, HAL est sans conteste une belle réussite collective grâce au modèle profondément original dessiné au cours de son histoire : dimension nationale, archive multidisciplinaire et collaborative, etc.

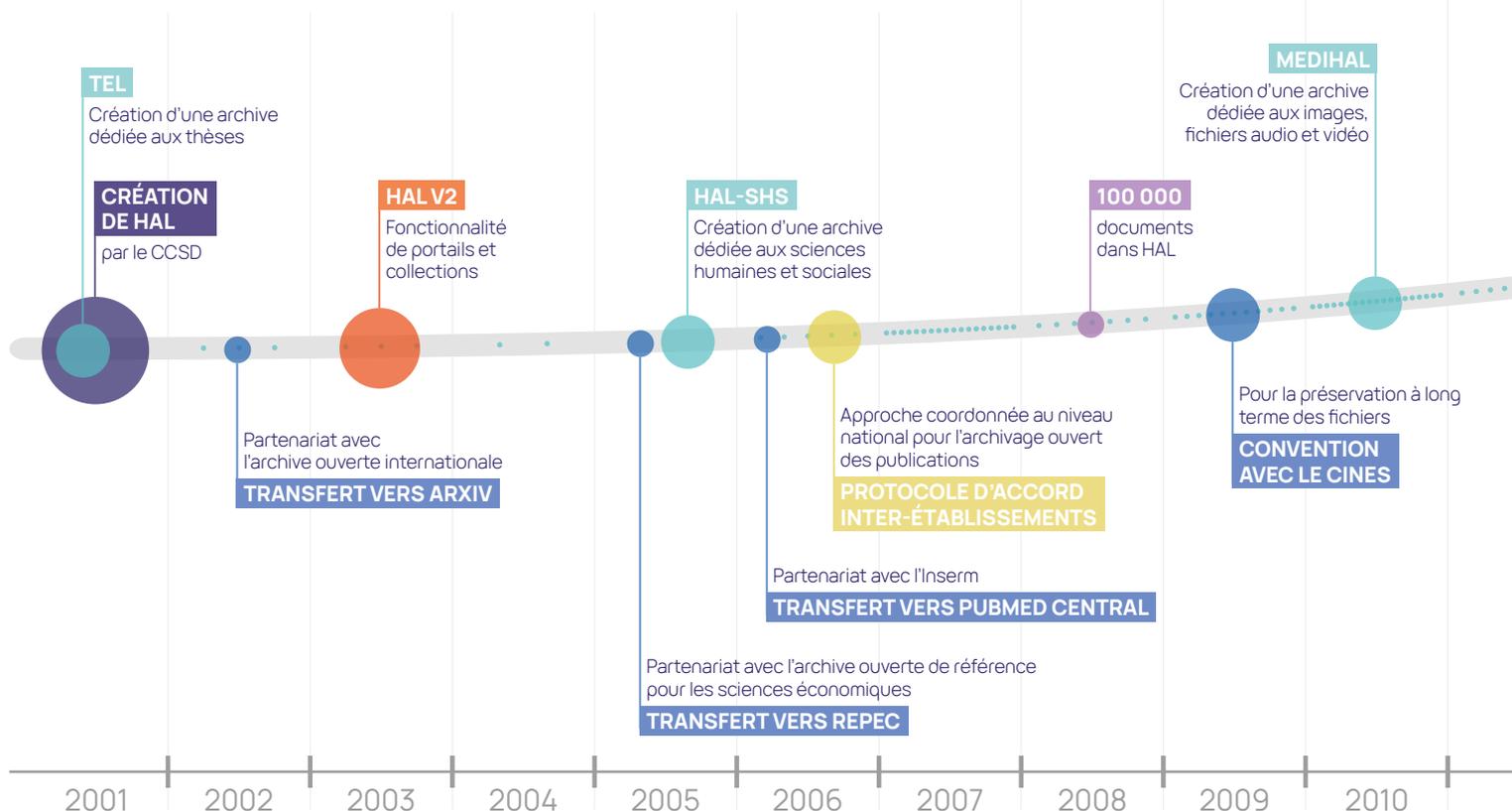
L'histoire de HAL est proposée dans les quelques pages qui suivent via ses multiples facettes permettant ainsi d'appréhender ces vingt années dans toute son originalité.

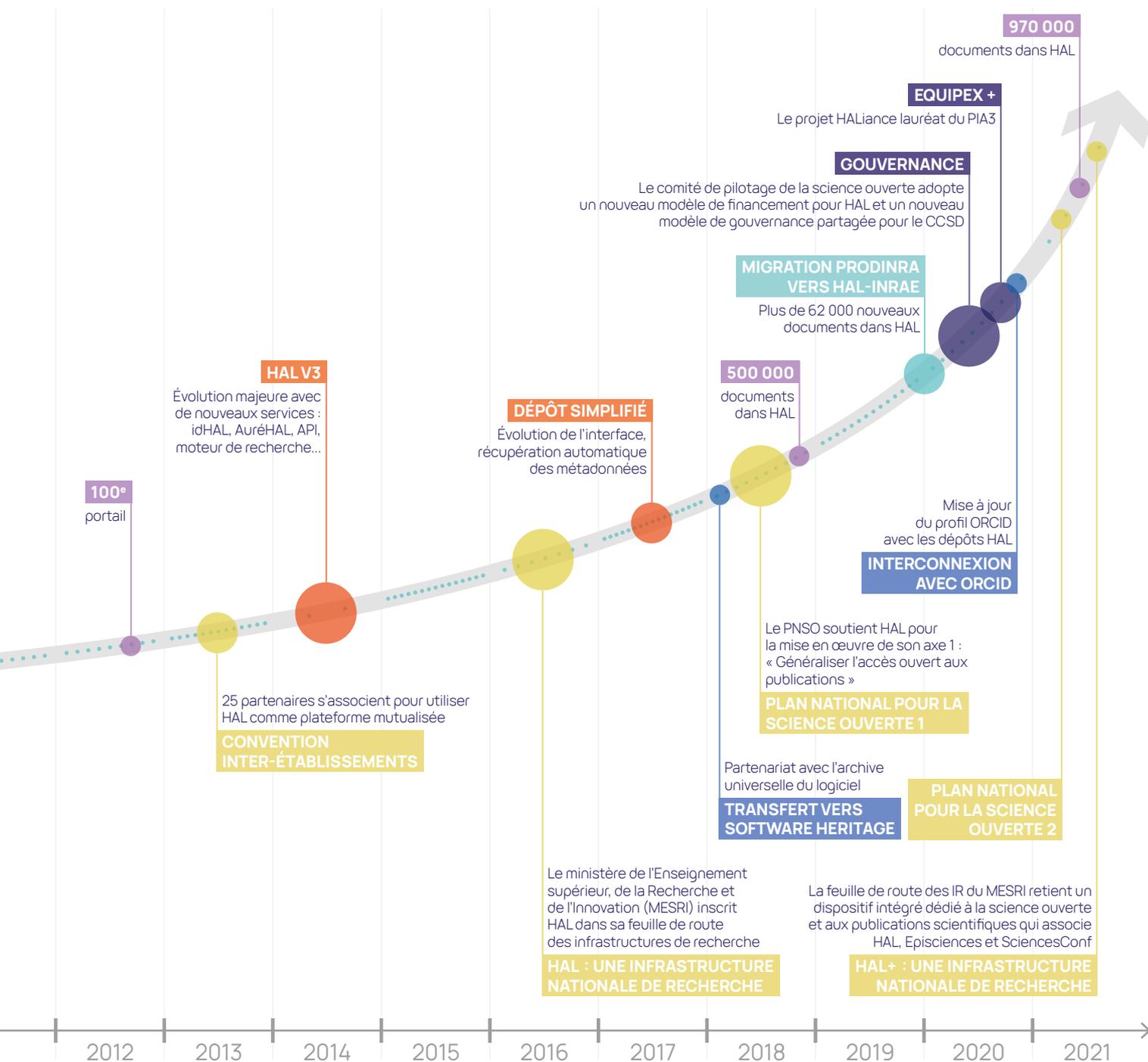
# Les temps forts de la vie de HAL

HAL en quelques événements importants, majeurs voire décisifs qui ont marqué son développement durant ses vingt premières années.

Plus d'informations sur  
[www.ccsd.cnrs.fr/les-temps-forts](http://www.ccsd.cnrs.fr/les-temps-forts)

- Événements structurants
- Développements majeurs
- Événements politiques
- Interopérabilité
- Paliers
- Portails
- Nouveau portail créé





# Au commencement

## Les pionniers les fondateurs

---

**Franck Laloë**

Chercheur en physique, directeur de recherche émérite au CNRS.

**Daniel Charnay**

Ingénieur de physique nucléaire au Centre de calcul de l'IN2P3 CNRS.

**Marco Picco**

Physicien, directeur de recherche au Laboratoire de physique théorique et hautes énergies.

## la première équipe d'informaticien·ne·s

---

**Laurent Capelli**

**Hélène Gaboriaud-Jamet**

**Yannick Barborini**

**Sonia Guérin-Hamdi**

**Estelle Theveniaud**

## Les directions du CCSD

---

**Franck Laloë**

**Daniel Charnay**

**Angel Alastuey**

**Christine Berthaud**

**Nathalie Fargier**

du 1<sup>er</sup> juillet 2000  
au 28 février 2006

du 1<sup>er</sup> mars 2006  
au 31 octobre 2008

du 1<sup>er</sup> novembre 2008  
au 31 mars 2011

du 1<sup>er</sup> avril 2011  
au 31 janvier 2020

depuis le 1<sup>er</sup> février 2020

Comme son modèle arXiv, HAL est née dans le domaine de la physique et a été portée par deux composantes du CNRS : le département des Sciences physiques et mathématiques (SPM) et l'Institut national de physique nucléaire et de physique des particules (IN2P3).

C'est vers ce dernier que s'est tourné Franck Laloë, initiateur du projet, pour sa réalisation. En effet, le Centre de calcul de l'IN2P3 (CC IN2P3) avait, lors du démarrage du projet, l'expérience des plateformes Web : c'est dans ce centre qu'avait été développé, en 1997 – par

Daniel Charnay entre autres – Démocrite, le premier logiciel sur Internet de gestion documentaire et de publications. Cette plateforme intégrait la mise en ligne de publications et ne différait d'une archive ouverte que par son accès privé.

## Les liens avec arXiv

---

Créée en 1991 par Paul Ginsparg, arXiv a fortement inspiré les fondateurs de HAL ; Marco Picco était d'ailleurs à l'époque l'administrateur du site miroir de l'archive américaine en France.

Le modèle, explique Laurent Capelli, est le suivant : « Un chercheur dépose un document, quelle que soit la version, publié ou non, met quelques métadonnées dont la discipline – obligatoire – et ce dépôt est modéré avant sa mise en ligne. Le principe du non retrait après mise en ligne vient aussi d'arXiv. »

Le transfert d'un dépôt HAL vers arXiv a été pensé dès le départ afin d'éviter une double saisie et permet aux auteurs de s'affranchir du système de parrainage (*endorsement*) sur lequel repose arXiv.

## CC-IN2P3 et CCSD

---

L'IN2P3 est étroitement lié à l'histoire du CCSD et de HAL. « Si l'IN2P3 n'avait pas été là pour me donner des machines, il n'y aurait pas eu HAL » témoigne Daniel Charnay<sup>1</sup>. À la naissance du CCSD, l'institut met à disposition des locaux pour l'équipe dans son centre de calcul situé à Villeurbanne, et y héberge les serveurs de HAL.

En mai 2017, l'équipe déménage par besoin de place, mais les serveurs et les métadonnées restent, encore à ce jour, en sécurité au CC IN2P3.

---

<sup>1</sup> Daniel Charnay. « Avec HAL, nous voulions créer un arXiv multidisciplinaire ». *Hermès, La Revue-Cognition, communication, politique*, CNRS-Éditions, 2019, La communication au cœur des connaissances, 2019/3 (85), pp.94-97. {hal-02387841}

# Au fil des développements

## Les versions de HAL

---

Un prestataire externe a écrit les toutes premières lignes de code, d'après un cahier des charges établi par Daniel Charnay, pour une v1 qui ne sera jamais mise en service.

Le développement d'une v2 est assuré ensuite au sein du CCSD. Le tout premier dépôt, hal-00000001, a été effectué le 23 septembre 2002 pour un article de Franck Laloë paru dans *American Journal of Physics*.

Toujours d'actualité en 2021, la v3 a quant à elle été livrée en octobre 2014 : elle est le résultat d'une ré-écriture complète sur un framework PHP avec un moteur de recherche.

Difficile de lister en quelques lignes les nouvelles fonctionnalités, mais on peut toutefois relever une ouverture plus large des données, avec la mise à disposition d'API pour la recherche de métadonnées, ainsi qu'une interface dédiée pour l'administration des référentiels offrant une grande autonomie aux portails pour la gestion de leur structures de recherche.

L'identifiant auteur idHAL, qui s'est imposé rapidement dans les milieux de la recherche, a quant à lui contribué à populariser les questions émergentes liées à l'identité numérique des chercheur-e-s. On en recense plus de 52 400 en novembre 2021.

### Le choix du format PDF

Alors qu'arXiv acceptait principalement des articles avec leurs sources TeX, HAL a dû faire un choix plus large de par sa multidisciplinarité.

Le format PDF s'est imposé ; il est indépendant du logiciel d'édition et les outils de lecture sont

gratuits. La forme normalisée PDF/A a été retenue, pour faciliter l'archivage à long terme.

Le CCSD a donc travaillé très tôt avec le CINES pour la sauvegarde externalisée du contenu de HAL, et pour garantir la lisibilité des PDF.

## De l'affiliation des auteur·trice·s

---

Associer l'auteur·trice à son laboratoire est une étape dans le dépôt qui a toujours été vécue comme compliquée, surtout quand il fallait affilier obligatoirement tous les auteur·trice·s de la publication déposée (un vrai challenge dans certaines disciplines !).

L'ajout des affiliations est définie dès le cahier des charges de HAL, se démarquant ainsi de son modèle arXiv. Une des demandes de Franck Laloë était en effet que la direction du laboratoire puisse voir toutes les publications de son équipe. « Le référentiel des laboratoires était mappé sur le référentiel CNRS », témoigne Laurent Capelli, « on récupérait des infos de Labintel, pour pré-renseigner

les labos CNRS. Dès qu'une institution rejoignait le CCSD pour son archive, la première chose qu'on faisait était de travailler leur référentiel. »

Améliorer l'ergonomie, définir des algorithmes d'affiliation, réduire l'obligation à une seule affiliation par dépôt : des efforts constants jalonnent les développements de HAL pour faciliter cette étape pour le déposant. L'affiliation des auteur·trice·s constitue cependant la clé de voûte pour créer, gérer et mettre à jour les collections des équipes, des laboratoires et les portails institutionnels.

### Refléter la complexité administrative de la recherche publique française

Le référentiel des structures de recherche, disponible dans AuréHAL depuis 2014, a traduit dans son modèle de données la réalité des entités administratives de la recherche publique. Ainsi, un laboratoire peut être organisé en plusieurs équipes et, sur le modèle des unités mixtes du CNRS, dépendre de plusieurs tutelles (organismes de recherche, universités), pouvant elles-mêmes être regroupées sous d'autres entités.

AuréHAL permet ainsi de gérer une hiérarchie de six niveaux. Pas toujours simple à mettre à jour !

# Une archive pour des archives

## TEL

---

TEL, c'est un peu la préhistoire de HAL.

Avant de développer sa propre solution, le CCSD a testé le logiciel ePrint avec la création d'un serveur multidisciplinaire pour la diffusion des thèses, mis en place avec la collaboration de la cellule MathDoc et le soutien de la Société française de physique. Intégrée ensuite dans HAL et élargie aux habilitations à diriger des recherches (HDR), TEL s'impose vite comme le premier diffuseur des thèses en France et, sans doute, comme le seul moyen de diffuser les mémoires HDR.

TEL repose bien sûr sur le principe de l'autoarchivage mais elle s'inscrit aussi

dans le dispositif national de signalement et d'archivage pérenne des thèses géré par l'ABES. Depuis 2009, l'application STAR de ce dispositif importe ainsi le texte intégral si l'auteur-e autorise la diffusion de ses travaux en libre accès et si l'établissement de soutenance a choisi TEL pour la mettre en œuvre.

Déposer sa thèse dans TEL ou dans HAL ? Ce n'est pas toujours évident de s'y repérer. Aussi, le nom de TEL va disparaître pour être remplacé par HAL Thèses afin de faciliter la compréhension des utilisateur-trice-s.

## HAL-SHS

---

Déposer ses articles de recherche en philosophie, histoire ou géographie parmi des publications en mathématiques ou en physique ? Pas si évident au début des années 2000, alors que les principes du libre accès étaient encore peu connus. HAL-SHS est née de ce constat.

Grâce aux efforts concertés de personnes engagées du département des Sciences de l'homme du CNRS, de l'Institut des Sciences de l'Homme à Lyon, de l'université de Rennes 2 et avec le soutien du CCSD, un espace dédié est proposé dès 2005 aux chercheur-e-s en sciences humaines et sociales pour mieux les accompagner sur la voie du libre accès.

Mission accomplie puisqu'en 2021, HAL-SHS représente un quart des dépôts présents dans HAL.

## MEDIHAL

---

Lancée en 2010 en partenariat avec le Centre national pour la numérisation de sources visuelles (CN2SV) et avec le soutien du TGE Adonis, MediHAL est née du constat que les chercheur-e-s disposent d'un grand nombre de photos sur leurs disques durs : clichés de fouilles archéologiques, photographies d'expériences d'astrophysique, reproduction de dessins botaniques, images de synthèses, etc.

Avec ce portail, véritable entrepôt de données, les chercheur-e-s peuvent auto-archiver ces documents iconographiques pour les conserver, les diffuser et les valoriser. Il a paru logique de l'élargir aux données audio-visuelles : c'est chose faite depuis 2014, à la suite de la mise en production de la v3 de HAL.

En 2021, MediHAL compte plus de 50 000 images et 3 000 vidéos scientifiques.



# D'une archive ouverte à l'infrastructure nationale d'aujourd'hui

En 2006, un protocole d'accord fédère autour de HAL la majorité des universités et des organismes de recherche français. Il permet de renforcer et d'accélérer le développement des archives ouvertes en France.

Une convention de partenariat vient actualiser cet accord en 2013 et confirme le rôle national de HAL.

Trois ans plus tard, en 2016, HAL figure parmi les infrastructures de recherche (IR) reconnues par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI). Puis, dans la feuille de route des infrastructures de recherche de 2021, HAL+ obtient à son tour le statut d'IR : le MESRI confirme ainsi l'importance

de HAL, mais aussi celle d'Episciences et de Sciencesconf pour les services complémentaires à l'archive ouverte qu'ils offrent.

En 2018 puis en 2021, avec les deux Plans nationaux pour la Science ouverte (PNSO), la France se dote d'une stratégie pour la diffusion sans entrave des publications et des données de la recherche et consacre la mission centrale de HAL pour la diffusion des publications et l'articulation avec les données et les logiciels de recherche. Dans la foulée de ces PNSO, les établissements de l'enseignement supérieur et de la recherche se dotent progressivement de politiques Science ouverte avec HAL désignée comme levier.

## Des mots pour le dire

Archive ouverte, plateforme mutualisée, plateforme collaborative, infrastructure nationale, archive ouverte nationale, réseau d'archives institutionnelles...

Les mots pour présenter HAL ne manquent pas : ils traduisent la profonde originalité de HAL dans le paysage international des archives ouvertes.

## Pourquoi des notices sans fichier dans une archive ouverte ?

---

Qu'une archive ouverte propose des notices bibliographiques sans texte intégral associé en intrigue ou agace plus d'un. C'est une critique adressée de façon récurrente à HAL, notamment par les défenseurs du libre accès. Mais n'est-ce pas là un facteur qui a contribué à son succès ?

Sur décision du CNRS, HAL a pris le relais de la plateforme PubliCNRS, fermée en 2005

et utilisée par les chercheur-e-s pour la saisie de leurs publications.

Installée dans les pratiques des chercheur-e-s affilié-e-s ou non au CNRS, et des laboratoires pour le recensement des publications, HAL est devenue un outil attractif pour des établissements comme les universités qui ont ce même besoin, d'autant qu'elle permet une mutualisation des efforts. L'accroissement du réseau des portails institutionnels en est le signe le plus évident. Reste à ajouter le texte intégral à toutes ces notices...

### La loi pour une République numérique

Difficile de parler de science ouverte en France sans parler de l'article 30 de la loi pour une République numérique, promulguée le 7 octobre 2016, qui donne de nouveaux droits et de nouvelles libertés aux chercheur-e-s pour mettre leurs écrits en libre accès.

Ce cadre juridique permet de faciliter l'ouverture des publications en étant un véritable bouclier protecteur contre les durées excessives d'embargo imposées par les éditeurs. C'est un argument utilisé massivement par tous les formateur-trice-s au dépôt.



# Les portails

## Un environnement personnalisé

Le besoin de personnaliser un environnement en modifiant les couleurs et/ou d'avoir d'autres métadonnées est apparu dès les débuts de HAL. La fonctionnalité de portail répond à ce besoin ou, selon Laurent Capelli, « l'idée d'un même code qui sert à plusieurs archives ».

TEL, Archivesic, et Jean Nicod, développées pendant la phase expérimentale, intègrent HAL en tant que portails, ainsi que Démocrite, facilitant la

maintenance des différents projets gérés par le CCSD.

EduTice est le tout premier portail thématique créé dans HAL v2 et se présente comme une bibliothèque numérique en éducation et technologies de l'information et de la communication.

Inria ouvre la voie de la mutualisation en 2005 en étant la première institution à lancer son portail et scelle une collaboration sur les développements

informatiques : les outils Bib2HAL et HALtools créés pour faciliter les imports et l'exploitation des données pour les chercheur-e-s de cet institut sont ainsi mis à disposition de toute la communauté des utilisateur-trice-s de HAL.

Le protocole d'accord de 2006 a engendré une réelle dynamique auprès des organismes de recherche dans un premier temps, que les universités et grandes écoles ont très vite rejoint.

## Migrer pour mutualiser L'expérience de l'INRAE

Plus de 62 000 nouveaux documents avec texte intégral au printemps 2020 ! C'est l'aboutissement de deux ans de travail pour préparer la migration de l'archive institutionnelle de l'Institut national de la recherche agronomique (Inra). L'Inra a en effet fait le choix de migrer Prodnra dans HAL, pour contribuer au développement de l'infrastructure

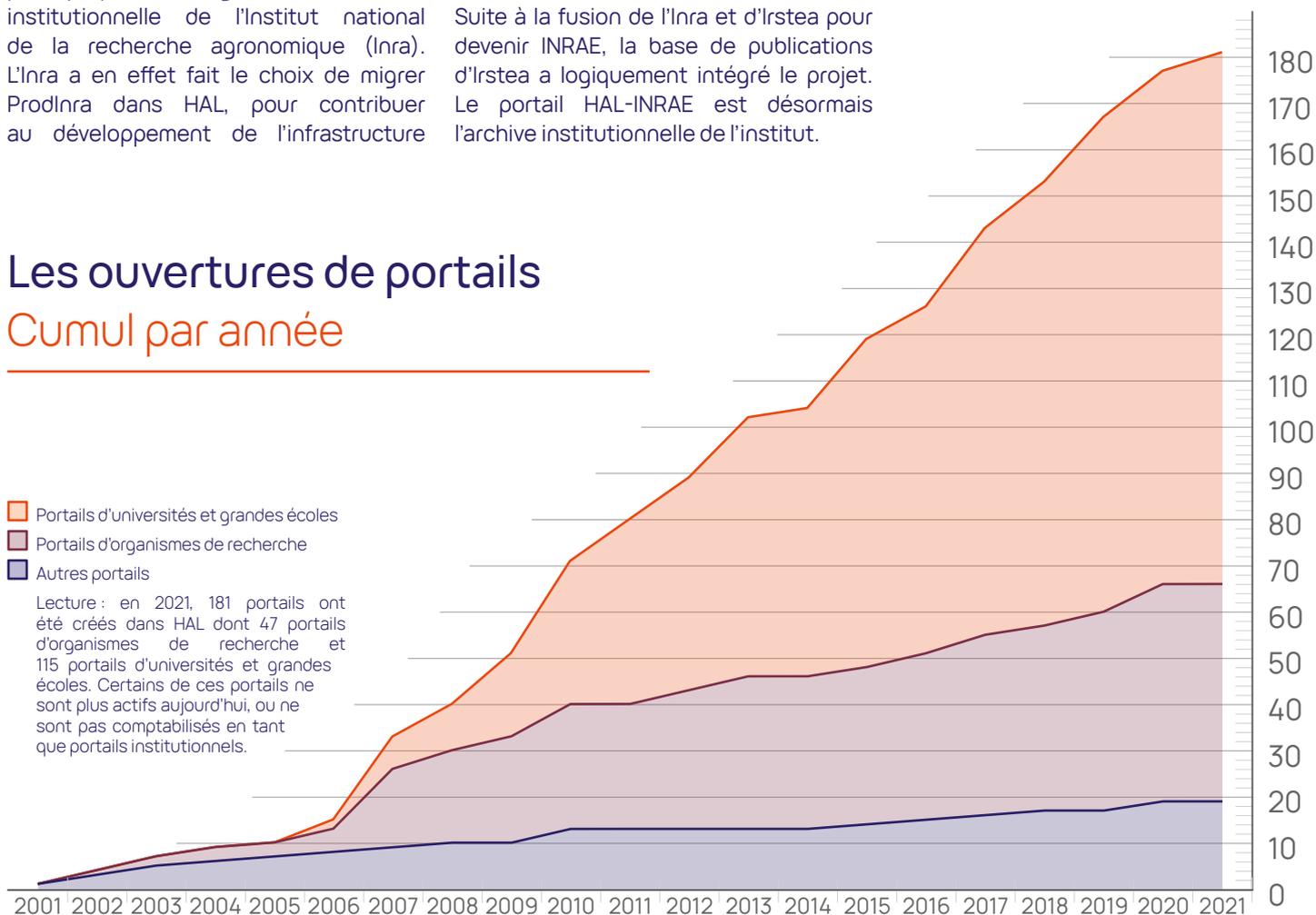
nationale, la mutualisation diminuant les coûts de développement et de maintenance évolutive.

Suite à la fusion de l'Inra et d'Irstea pour devenir INRAE, la base de publications d'Irstea a logiquement intégré le projet. Le portail HAL-INRAE est désormais l'archive institutionnelle de l'institut.

## Les ouvertures de portails Cumul par année

- Portails d'universités et grandes écoles
- Portails d'organismes de recherche
- Autres portails

Lecture : en 2021, 181 portails ont été créés dans HAL dont 47 portails d'organismes de recherche et 115 portails d'universités et grandes écoles. Certains de ces portails ne sont plus actifs aujourd'hui, ou ne sont pas comptabilisés en tant que portails institutionnels.



# Au cœur d'un écosystème

## Pas de success story sans les utilisateur·trice·s !

---

Ce sont eux bien sûr qui font vivre l'archive : les chercheur·e·s au cœur du système, qu'ils soient militant·e·s du libre accès de la première heure, attentif·ve·s au partage de la science au plus grand nombre, invité·e·s, incité·e·s voire obligé·e·s à déposer dans HAL ; les personnels d'accompagnement de la recherche comme les documentalistes

dans les laboratoires et les personnels des bibliothèques universitaires ; les institutions qui ont choisi HAL comme pilier de leur politique de science ouverte. L'appellation plateforme collaborative prend ici pleinement son sens.

En 2021, on recense plus de 83 000 comptes actifs.

### L'assemblée des partenaires

La nouvelle gouvernance du CCSD, mise en place en 2021, vise à mieux associer les utilisateurs au développement des plateformes.

L'assemblée des partenaires est une des instances de cette nouvelle gouvernance : elle réunit les représentants des 132 institutions (universités, grandes écoles, organismes de recherche) qui disposent d'un portail HAL et contribuent ainsi financièrement à son développement.

### Un club utilisateurs : CasuHAL

En 2016, l'association CasuHAL (Communauté des utilisateurs de l'Archive ouverte HAL et services connexes) dépose ses statuts.

Elle rassemble des utilisateurs avancés comme des administrateurs de portail, des gestionnaires de collection et des référents structures qui partagent leurs expériences et pratiques. Elle contribue activement aux évolutions fonctionnelles de l'archive en lien avec le CCSD.

## HAL dans l'écosystème de la science ouverte

---

Les interconnexions avec d'autres systèmes font partie de l'ADN de toute archive ouverte, tel que défini par l'Open Archives Initiative, pour permettre l'échange, le partage et la valorisation des ressources. En vingt ans, les exemples sont nombreux de collaborations qui ont inscrit HAL dans un réseau dense qui constitue un écosystème foisonnant.

Du partenariat avec arXiv dès l'origine de HAL pour le transfert des dépôts en physique à celui avec Software Heritage en 2018 pour les codes sources de logiciels, HAL s'est toujours attachée à répondre aux besoins des chercheur-e-s pour favoriser la visibilité de leurs travaux.

Citons par exemple la collaboration avec l'Inserm dès 2006 pour le transfert du texte intégral de dépôts HAL dans l'archive PubMed Central ou la mise à disposition de métadonnées de dépôts en économie pour la plateforme RePec depuis 2007.

L'identifiant ORCID créé en 2012 s'est peu à peu imposé à l'international comme l'identifiant chercheur qui fait référence : depuis 2020 et grâce à l'adhésion du CNRS, il est possible de compléter son profil ORCID avec ses publications déposées dans HAL. L'identifiant international ROR (Research Organisation Registry) est utilisé depuis 2021 pour enrichir les affiliations institutionnelles.

## Un développement appuyé par un fort soutien institutionnel et politique

---

Créé en 2000 sous forme d'unité propre de services du CNRS, le CCSD devient unité mixte de services en 2014 et unité d'appui et de recherche en 2021. Depuis l'origine, le CNRS soutient HAL et apporte des ressources humaines, matérielles et financières pour assurer son développement. Inria et l'Université de Lyon rejoignent le CNRS comme tutelles contribuant au pilotage de HAL et apportant également des ressources.

Inria a par ailleurs été un soutien des premières heures pour HAL puisque son portail institutionnel a vu le jour en 2005. Dernière tutelle à rejoindre l'aventure, l'Inra, devenu INRAE suite à sa fusion avec Irstea à partir de 2018.

De son côté, le MESRI encourage aussi l'expansion de l'archive ouverte et du CCSD en général par l'attribution de financements réguliers et exceptionnels, mais aussi à

travers la politique en matière de science ouverte qu'il impulse dans le cadre du Comité pour la Science Ouverte. La dynamique est aussi portée par les établissements dotés d'un portail HAL et siégeant au sein de l'Assemblée des partenaires comme par la Conférence des Présidents d'Universités (CPU) qui participe au Comité de pilotage du CCSD.

# Une image d'ouverture

En vingt ans, l'identité graphique de HAL a peu évolué : c'est surtout le personnage créé par le dessinateur Selçuk qui la symbolise, et ce, depuis 2007.



↑ Le personnage dessiné par Selçuk, qui a accompagné HAL de 2007 à aujourd'hui

Le CCSD a fait appel à une agence de graphisme pour un changement de logo en 2014 qui l'identifie depuis la mise en production de la v3.



↑ Logo de 2001 à 2014

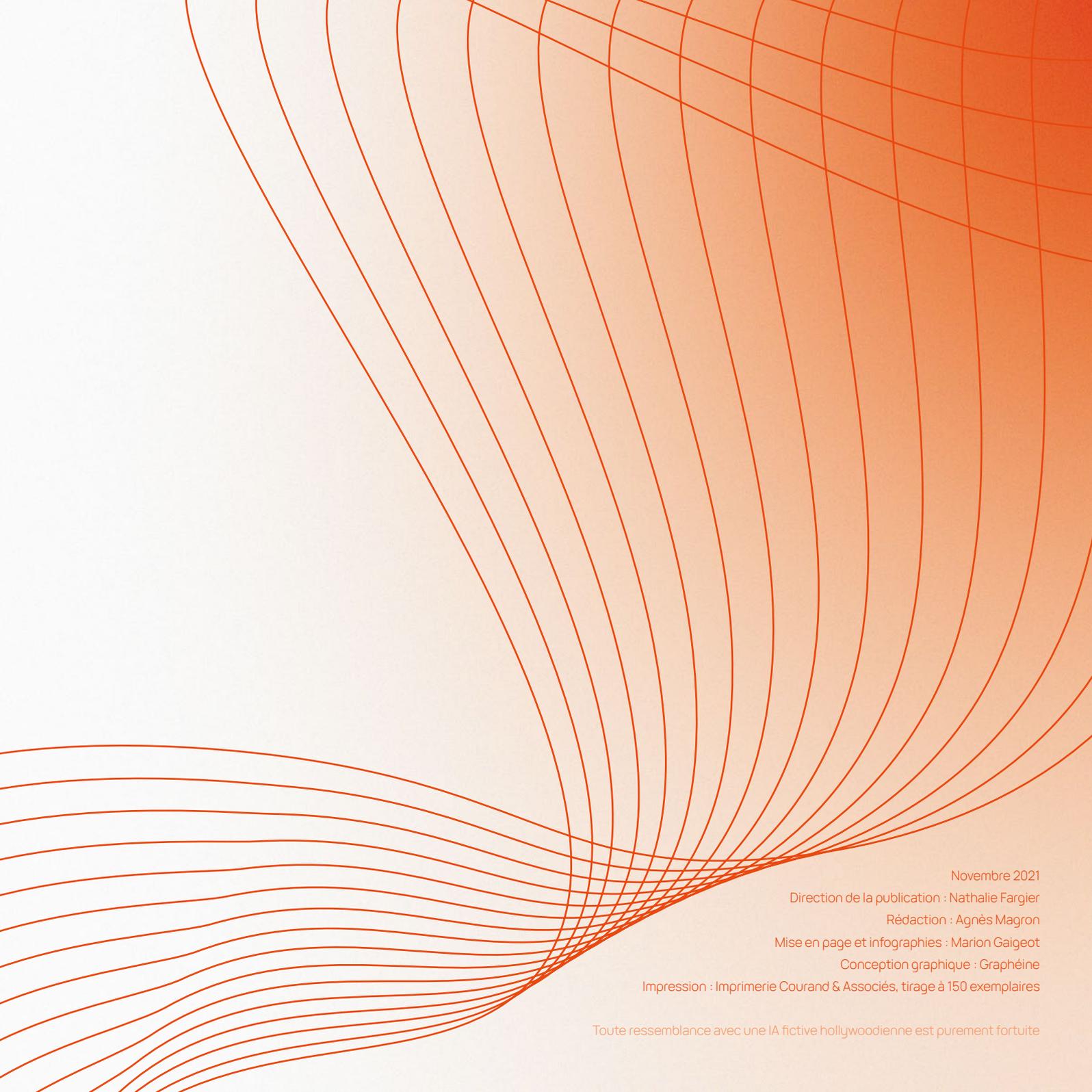


↑ Logo de 2014 à 2021

Fin 2021, à l'occasion de son anniversaire, HAL change de look !



↑ Logo à partir de 2021



Novembre 2021

Direction de la publication : Nathalie Fargier

Rédaction : Agnès Magron

Mise en page et infographies : Marion Gaignot

Conception graphique : Graphéine

Impression : Imprimerie Courand & Associés, tirage à 150 exemplaires

Toute ressemblance avec une IA fictive hollywoodienne est purement fortuite

**CCSD...**  
Centre pour la Communication  
Scientifique Directe



Novembre 2021  
CCSD