

L'importance de l'information scientifique (IS)¹ s'est accrue dans le monde de la recherche à partir des années 2000, du fait notamment de la numérisation accélérée des publications scientifiques et du développement des plateformes internationales de diffusion des publications qui ont facilité la création d'outils scientométriques, bouleversé les équilibres socio-économiques existants et les pratiques scientifiques elles-mêmes.

En défendant un « accès ouvert » à la science et en élaborant des plans successifs pour la « science ouverte ²», la recherche publique s'est dotée de moyens pour maîtriser, au moins partiellement, les évolutions en cours. Dans de nombreux pays, la question des équipements en information scientifique a ainsi été inscrite sur l'agenda des politiques de recherche.

Les roadmaps ESFRI 2006 – 2021

Afin de faciliter une politique des grands équipements de la recherche coordonnée à l'échelle européenne, est créé en 2002 le « Forum stratégique européen sur les infrastructures de recherche » (*European Strategy Forum on Research Infrastructures* - ESFRI).

Dans la première feuille de route européenne « ESFRI » (2006), l'information scientifique (IS) a été identifiée comme un sujet d'intérêt pour l'ensemble des sciences. Cette roadmap mentionne le besoin de disposer d'outils de traitement des connaissances scientifiques sans pour autant consacrer une rubrique spécifique à l'IS ni même labelliser une infrastructure dans ce domaine.

Il faudra attendre la feuille de route ESFRI de 2021 pour qu'une infrastructure d'IS soit reconnue sous le nom d'OPERAS, l'acronyme signifiant en français : « Communication scientifique ouverte dans l'Espace européen de la recherche en sciences humaines et sociales ».

Cette reconnaissance comble une lacune dans le paysage européen. OPERAS est une infrastructure dédiée aux questions de diffusion et d'accès aux résultats de la

recherche qui soutient les cultures de communication scientifique multilingues en sciences humaines et sociales. Pilotée par l'infrastructure de recherche (IR) française OpenEdition en partenariat avec la très grande infrastructure (IR*) Huma-Num, OPERAS témoigne de l'avancée des équipes françaises dans le domaine de l'IS depuis les années 2010 et confirme l'intérêt, reconnu dès la première feuille de route française, de conférer à certaines réalisations relevant de l'information scientifique, le label d'infrastructure.

Les feuilles de routes françaises 2008-2021

À la différence d'ESFRI, l'information scientifique a toujours été présente dans les feuilles de route françaises des infrastructures de recherche. Toutefois, les modalités de sa prise en compte ont évolué au fil du temps.

L'IS ne dispose pas d'une rubrique dédiée en propre dans la première feuille de route, en revanche, dès 2008 (Cf. Fig 1) est labellisé dans la section « sciences humaines et sociales » le projet de « bibliothèque scientifique numérique » (BSN) qui entend construire un projet global d'information scientifique revisitée à l'aune du numérique. La section « Infrastructures documentaires » a été introduite dans la feuille de route de 2012. Cette rubrique a été renommée « Infrastructures en information scientifique et technique » (IST) en 2016 et a conservé cette dénomination en 2018. En 2021, la section s'intitule « Infrastructures en information scientifique », elle est intégrée à la rubrique « Services et infrastructures numériques ». Dans le monde de la recherche où les typologies sont rarement laissées au hasard, ces changements d'appellation traduisent effectivement des intentions particulières.

Dans un premier temps, avec le terme d'infrastructures documentaires, il s'agit, pour les responsables politiques, d'affirmer la volonté d'impliquer le monde des bibliothèques (très présent dans les universités), longtemps resté à la marge des bouleversements subis par la communication scientifique.

¹ L'information scientifique (ou information scientifique et technique) désigne l'ensemble des informations produites par la recherche et nécessaires à l'activité scientifique. Elle intervient en amont et en aval dans tout le cycle de production de nouveaux contenus scientifiques quelle que soit leur forme : articles, données, ouvrages, archives ouvertes, etc.

² « La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Elle s'appuie sur

l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche. » <https://www.ouvrirlascience.fr/initiez-vous-a-la-science-ouverte/>

En dédiant ensuite une rubrique spécifiquement à l'IST, le projet est alors de souligner que ce volet du travail scientifique est devenu, avec l'informatisation, un sujet aux implications importantes pour la pratique et le pilotage de la recherche qui va bien au-delà de simples choix techniques. L'expression « information scientifique » (IS), débarrassée du mot « technique » traduit l'importance de ces enjeux qui sont moins techniques que scientifiques.

FIG 1. Les infrastructures d'information scientifique
Feuilles de route françaises : 2008-2021

Feuille de route	Infras IS présentes dans la rubrique SHS	Rubrique spécifique IS Intitulé de section
2008	BSN	–
2012		Infrastructures documentaires
		BSN CollEX
2016	OpenEdition	Infrastructures en information scientifique et technique » (IST)
		CollEx-Persée HAL, CCSD NUMEDIF OpenEdition
2018	OpenEdition	Infrastructures en information scientifique et technique » (IST)
		Huma-Num (dominante infra SHS) CollEx-Persée HAL, CCSD METOPES OpenEdition
2021	OpenEdition	Infrastructures en information scientifique » (IS) (rubrique « Services et infrastructures numériques »)
		CollEx-Persée HAL+ METOPES OpenEdition ISTEX Software heritage

D'une feuille de route à l'autre : continuité et clarification

L'examen du tableau présenté ci-dessus permet de constater que, dans le prolongement de plusieurs rapports consacrés à l'IS et aux bibliothèques universitaires³, un dispositif national de coordination et de coopération est créé au moment de l'élaboration de la première feuille de route française des infrastructures de recherche : la Bibliothèque Scientifique Numérique (BSN).

- ♦ BSN (Bibliothèque scientifique numérique) est pensée comme la forme renouvelée d'une partie du secteur de l'information scientifique et technique. BSN est construite autour de 9 segments :
- BSN 1 : L'acquisition des archives de l'édition scientifique et l'abonnement aux revues scientifiques
 - BSN 2 : Les dispositifs d'accès et d'hébergement
 - BSN 3 : Le signalement (bases de données bibliographiques)
 - BSN 4 : Les archives ouvertes
 - BSN 5 : La numérisation du patrimoine scientifique de l'ESR national (revues et ouvrages)
 - BSN 6 : L'archivage pérenne
 - BSN 7 : L'édition scientifique : coordination de l'aide publique, soutien à l'innovation de nouvelles formes d'édition, etc.
 - BSN 8 : La fourniture de documents/prêt entre bibliothèque
 - BSN 9 : Formations, Compétences, Usages

Les quatre infrastructures pivots de l'IS que sont OpenEdition, Persée, HAL, METOPES pilotent dès les débuts de BSN chacune un des segments de cette *Bibliothèque Scientifique Numérique*. En 2016, chacune de ces infrastructures ayant atteint un degré de maturité suffisant, le choix est fait qu'elles figurent en propre sur la feuille de route, Persée étant associée à CollEx.

Partageant une volonté d'ouverture à l'international et assurant une indépendance technologique à la recherche française, ces plateformes ont continué à interagir entre elles et avec de nombreuses infrastructures internationales (métadonnées partagées, accords de partenariat, dispositifs de réversion, etc.). Ensemble, elles assurent ou accompagnent aujourd'hui tout ou partie des fonctions de l'information scientifique dans la recherche académique : collecter et produire, certifier, structurer et relier, améliorer la diffusion et inscrire dans la durée. Complémentairement, elles respectent au plus près les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable). Inscrites dans la durée, leurs actions conjointes constituent un écosystème associant méthodes, données, codes et publications.

³ Communication scientifique, enjeux et propositions (rapport de la mission IST, dite « mission Pau »), CNRS, juillet 2004 ; Mission sur le devenir du CINES, Centre informatique national de l'enseignement supérieur, Yann Boaretto, rapport du 22 janvier 2007, notamment partie II « Propositions pour l'organisation de l'information scientifique et

technique en France », p. 19-43 ; Contribution de Couperin (Consortium universitaire de publications numériques) à la réflexion sur l'IST, 2 pages, 21 février 2008 ; Rapport du Comité IST information scientifique et technique, sous la direction de Jean Salençon, mai 2008.

♦ **OpenEdition** est une infrastructure d'édition électronique au service plus particulièrement de la communication scientifique en sciences humaines et sociales. Elle rassemble quatre plateformes complémentaires dédiées respectivement aux collections de livres avec OpenEdition Books, aux revues avec Revues.org – qui devient OpenEdition Journals en 2017–, aux blogs de recherche avec Hypothèses et aux annonces scientifiques avec Calenda. La mission d'OpenEdition est de favoriser le développement d'une édition électronique en SHS qui dépasse la simple mise en ligne de documents scientifiques, en misant sur la formation à de nouveaux usages éditoriaux, sur l'internationalisation, sur un modèle économique durable et sur des innovations technologiques régulières. **Les tutelles** d'OpenEdition sont le CNRS, AMU, EHESS, et l'Université d'Avignon.

En 2016, le vote de la loi pour une République numérique qui permet l'adoption par la France de mesures en faveur du libre accès aux publications scientifiques, dans la lignée du mouvement mondial de la Science ouverte, et parallèlement à l'ouverture des données publiques (Open Data) va permettre d'amplifier le mouvement en faveur de la science ouverte.

Sous l'égide du Comité pour la Science Ouverte (CoSO), qui prend désormais le relais de la BSN, les quatre infrastructures voient leurs missions consolidées, et réinscrites sur la feuille de route 2018, avec pour objectif d'accélérer la diffusion des résultats de la recherche en facilitant le dépôt pour les chercheurs, le travail d'édition et de publication des résultats et le développement de services pour un accès et une exploitation facilités. Le Fonds national pour la science ouverte (FNSO) permet de financer de nombreuses actions et constitue une garantie de cohérence de l'ensemble.

♦ **CollEx-Persée** est une infrastructure de recherche en information scientifique et technique mise en place en 2017 qui réunit, sous la forme d'un groupement d'intérêt scientifique 21 établissements de recherche (membres ou associés), 4 opérateurs nationaux (Persée pour la numérisation ; l'Abes pour les données bibliographiques ; le CTLES pour la conservation des documents ; l'INIST-CNRS pour la fouille de données), la Bibliothèque nationale de France et une centaine de structures labellisées au titre de leurs collections d'excellence. Son objectif est de faciliter l'accès et de favoriser l'usage des collections de bibliothèques par les chercheurs dans un contexte où les relations entre ces deux communautés sont mises à l'épreuve du fait des évolutions techniques et de la disponibilité massive de ressources documentaires en ligne qui ont modifié le processus de recherche. La vision de CollEx-Persée est celle d'une bibliothèque qui développe des collections hybrides (ressources numériques, imprimées, matériaux de la recherche comme les archives et les ressources iconographiques) adossées à des services qui répondent aux nouveaux besoins des chercheurs en tant qu'utilisateurs de sources d'information de toute sorte et producteurs de données et de résultats de recherche. **Établissement porteur** : Campus Condorcet.

♦ **HAL+** est une infrastructure intégrée de diffusion et d'archivage de la production scientifique, couvrant tous les domaines scientifiques, les différents types de documents produits par la recherche (articles publiés ou non dans des revues à comité de lecture, communications scientifiques, thèses...) et leur articulation avec les données. Sa stratégie s'inscrit dans le mouvement international en faveur de la science ouverte. Elle repose sur trois plateformes complémentaires : HAL, l'archive ouverte pluridisciplinaire de la recherche française (et tous ses portails dérivés) qui en est le moteur premier ; Episciences pour l'innovation éditoriale et la publication en OA diamant de revues et SciencesConf pour l'organisation d'événements scientifiques. L'ensemble de ces outils et services permet aux chercheurs et aux établissements de recherche de rendre leur production visible, découvrable, utilisable et accessible à long terme et de se réappropriier la maîtrise du processus de communication scientifique. Le Centre pour la Communication Scientifique Directe en assure la mise en œuvre. **Tutelles/Partenaires** : 124 universités, organismes de recherche et écoles de l'ESR français.

De NUMEDIF à METOPES

Initialement attachée à la mise au point et à la dissémination d'un écosystème normé de production numérique de contenus scientifiques structurés interopérables, et à la construction d'un appareil spécifique de diffusion-distribution, NUMEDIF résulte de l'association de deux services : le service de diffusion des presses universitaires confié à la FMSH, et le service de l'Université de Caen à l'origine de la plateforme METOPES. Cette dernière repose sur la prescription de normes et standards dans le domaine de l'édition et de la diffusion numérique ainsi que sur l'animation d'un réseau (formation, partenariats...). Les difficultés rencontrées à la FMSH à partir de 2017 ont conduit à ne retenir que METOPES à partir de la feuille de route de 2018.

♦ **MÉTOPES** est conçue à l'usage des éditeurs et au service de l'activité éditoriale de l'ensemble des établissements publics d'enseignements supérieur et de recherche. L'infrastructure fournit un ensemble de méthodes et d'outils propres à permettre la constitution de fonds éditoriaux numériques structurés et à assurer, dans le contexte du passage d'une économie de l'imprimé diffusé par abonnement à celle de la diffusion de contenus numériques en accès ouvert, la mutation des fonctions d'édition et de diffusion au plus près des missions et des objectifs de diffusion des résultats et des données de la recherche. Métopes est une infrastructure technique de veille, de développement et de formation au service des producteurs de données édités. Elle leur fournit des méthodes et des outils permettant la mise en œuvre, à partir du vocabulaire de la Text Encoding Initiative, d'un standard de représentation des contenus édités pour la production de fonds numériques normés, reconfigurables, pérennes, dotés de métadonnées riches, à fort potentiel d'interopérabilité. **Établissement porteur** : Université de Caen-Normandie.

L'élargissement : ISTEK et Software heritage

La feuille de route de 2021 est l'occasion d'introduire deux nouvelles infrastructures : ISTEK qui est soutenu par un solide partenariat d'établissements et *Software heritage* qui introduit de nouveaux objets d'archives (notamment les codes sources des logiciels) indispensables au déploiement de la science ouverte.

◆ ISTEK (création 2012 - exploitation 2015)

Information scientifique et technique d'excellence

ISTEK abrite une bibliothèque scientifique numérique donnant accès à un corpus de 23 millions de documents (articles, ebooks...) couvrant toutes les disciplines. Ces ressources sont pérennes, accessibles et exploitables par la communauté de l'ESR. Complément aux abonnements courants, la plateforme répond à deux besoins : i) la recherche de documents ; ii) la fouille de contenus (les documents sont prétraités et enrichis afin d'en faciliter l'exploitation ; des fonctionnalités d'extraction API permettent de générer des corpus à la demande, etc.). L'association entre un moteur de recherche puissant et une intégration dans les environnements numériques locaux permet une navigation simple entre les ressources courantes et les collections rétrospectives.

La plateforme ISTEK résulte d'un partenariat entre le CNRS, l'Abes, le consortium Couperin, FU et l'Université de Lorraine, elle s'appuie sur un système d'adhésion des établissements pour financer sa maintenance logicielle et matérielle

Tutelles/Partenaires : ABES, Consortium Couperin, INIST-CNRS, France Université, Université de Lorraine

Software Heritage (création 2015 - exploitation 2018) est une archive universelle ayant pour mission de collecter, préserver et rendre facilement accessible le code source de tous les logiciels disponibles publiquement, avec tout l'historique de leur développement. Software Heritage contribue à construire le pilier logiciel de la Science ouverte, en répondant à deux besoins clés : l'archivage pérenne et public des codes sources des logiciels produits ou utilisés par la recherche, et le référencement exact, des versions utilisés dans les expérimentations, ce qui en facilite la reproductibilité. Le code source publiquement disponible est collecté de plusieurs façons : i) automatiquement et régulièrement depuis des plateformes d'hébergement de code, comme GitHub, GitLab.com ou Bitbucket, et des archives de paquets, comme Npm ou Pypi ; ii) à la suite d'un signalement sur l'interface « Save Code Now » ; iii) par le dépôt depuis la plateforme multidisciplinaire d'archives ouvertes HAL. L'ensemble du code source, avec son historique de développement, est intégré dans une structure de données spécifique, un graphe dirigé acyclique de Merkle, et à tous les artefacts logiciels dans l'archive sont associés des identifiants cryptographiques intrinsèques et pérennes, appelés SWHID, qui en garantissent l'intégrité.

Tutelles/Partenaires : CEA, CNRS, DINUM, INRIA, MESRI, Université de Paris, Sorbonne U.

Un horizon à moyen terme

Les infrastructures d'information scientifique ont atteint une taille critique. Par exemple, Software Heritage héberge 9,6 milliards de fichiers sources provenant de 151 millions de projets. HAL+ diffuse en accès ouvert 865000 documents scientifiques en texte intégral, dont plus de 100 000 thèses couvrant toutes les disciplines, et héberge près de 9 000 conférences. OpenEdition regroupe sur ses différentes plateformes plus d'un million de documents scientifiques publiés, dont 96 % sont disponibles en accès ouvert, répartis en 550 revues scientifiques, 10 000 livres électroniques et 6 000 carnets de recherche. Enfin, ISTEK donne accès à 23 millions de documents.

En termes d'utilisation, le succès est encore plus massif. En 2020, HAL a comptabilisé 11,8 millions de visites, et l'ensemble des services d'OpenEdition en a enregistré 96 millions. La population des usagers ne se limite pas à la communauté nationale : 51,8 % des visites de l'infrastructure HAL proviennent d'autres pays que la France, et 47,9 % des visites d'OpenEdition sont en provenance de l'international. Cette fréquentation impose aux établissements tutelles, avec l'appui du ministère en charge de la recherche, un investissement à la hauteur des enjeux pour garantir à la recherche française indépendance et respect de la propriété des données qu'elle produit.

Au-delà des objets traditionnels de l'information scientifique, il est nécessaire de prendre en compte de nouvelles formes ou celles historiquement considérées comme mineures. Ainsi, le code et les logiciels, les archives scientifiques et la vidéo sont des objets qui doivent recevoir une attention particulière à l'avenir, soulevant des questions essentielles d'information scientifique.

Cet ensemble d'infrastructures constitue un paysage dynamique, complémentaire et structuré. Pour le Ministère, la perspective, à présent, serait d'aller plus avant en rassemblant les porteurs de ces infrastructures pour qu'ils construisent collectivement l'écosystème français des infrastructures de la science ouverte.

À paraître en 2025 aux éditions A. Athéna, un ouvrage sur PROGEDO et un ouvrage sur HUMA-NUM (genèse, conditions de développement, positionnement européen et international).

Collection Repères Athéna

Sciences humaines et sociales, les structures de recherche *Repères* n°1, juin 2021

Sciences humaines et sociales, les réseaux de recherche *Repères* n°2, juin 2022

La publication scientifique à l'ère numérique, comprendre le *Closed Access* et l'*Open Access*, *Repères* n°3, décembre 2024

Les infrastructures de recherche en SHS, *Repères* n°4, février 2025