

# NOTE D'ÉTONNEMENT

Baromètre de la science ouverte



Justine BOUWY-OUNNOUGH  
DTR/D2MC2/SEARCH

## Qu'est-ce que le Baromètre de la Science Ouverte ?

- Le Baromètre de la Science Ouverte (BSO) en France est un **outil** mis en place depuis 2018 pour mesurer l'avancement de l'**ouverture des publications scientifiques**. Inscrit dans le **Plan national pour la science ouverte**, il est développé par le **ministère de l'Enseignement supérieur et de la Recherche**, en partenariat avec **Inria** et l'**Université de Lorraine**.
- Il est construit à partir de **données ouvertes** et propose un site internet offrant des **dizaines d'indicateurs** regroupés en **thématiques**, accompagnés de **visualisations interactives**. Les données sous-jacentes au baromètre sont mises à disposition sous **licence ouverte**, son **code est ouvert** et sa méthodologie est présentée en détail dans une publication elle-même en accès ouvert.

Source : [Baromètre de la Science Ouverte](#)

## En résumé

### Objectifs

- **Suivre et analyser** l'évolution de l'accès ouvert aux publications scientifiques produites par les chercheurs en France.
- **Encourager les pratiques de science ouverte** en offrant une vision globale de la proportion d'articles publiés en libre accès.
- **Identifier les disciplines et les institutions** qui progressent ou doivent encore améliorer l'accès ouvert.

### Critères analysés

- **Types d'accès** : L'outil distingue différents types d'accès ouvert, comme la voie dorée (Gold open access) et la voie verte (Green Open Access).
- **Disciplines** : Les résultats sont présentés par discipline, ce qui permet de voir les variations d'une spécialité à l'autre.
- **Institutions** : Le baromètre montre comment différentes institutions françaises s'en sortent en termes de science ouverte.

### Mesures

- Le baromètre **examine les publications scientifiques**, en particulier les articles de revues, et détermine le pourcentage de ceux qui sont disponibles en accès ouvert (via archives ouvertes, revues en libre accès, etc.).
- Il se base sur des **bases de données et des outils de référence** comme Crossref, Unpaywall, ou HAL pour collecter les données.



## Crossref VS Datacite – pourquoi le choix du BSO ?

- Il n'y a **pas d'explication clairement donnée** par le site gouvernemental du BSO quant à savoir pourquoi les DOI pris en compte ne sont que les DOI Crossref. Mais en étudiant les sites des 2 entreprises, on a **une piste** qui pourrait expliquer ce choix.



- **Couverture** : Principalement axée sur les publications académiques traditionnelles telles que les articles de revues, les chapitres de livres et les actes de conférences.
- **Services** : Offre des services tels que le [Reference Linking](#), qui permet aux chercheurs de suivre des liens depuis les listes de références vers d'autres documents en texte intégral, facilitant ainsi la découverte de nouvelles informations.



- **Couverture** : Cible une variété de ressources de recherche, notamment les jeux de données, les logiciels, les images et d'autres types de contenus non textuels.
- **Services** : Propose des outils comme le [DOI Fabrica](#), une plateforme de gestion de DOI adaptée à la création manuelle ou à la curation humaine de DOI générés automatiquement.

- A ses débuts en 2018, le Baromètre de la Science Ouverte se concentrait **essentiellement** sur les **articles de revues scientifiques**. Il est donc cohérent que le choix se soit porté sur **Crossref**, qui couvre les **publications académiques textuelles**.

Sources : [Crossref](#) & [Baromètre de la Science Ouverte](#)

## Analyse SWOT – Forces au BSO institutionnel IRSN

- **Engagement institutionnel et communautaire** : L'équipe HAL-IRSN (Audrey Legendre, Karen Payrar, Bruno Cosenza) et, en tant que chargée de bibliométrie, Irène Sorokine-Durm sont des référents importants pour la mise en place du BSO. Il en va de même pour les 105 établissements ayant déjà fait leur BSO ou partageant les mêmes problématiques que nous\*.
- **HAL** : Le BSO se base – entre autres – sur l'archive ouverte HAL pour calculer les taux d'accès ouvert. Un travail de complétude est fait depuis plusieurs années à l'IRSN pour alimenter HAL IRSN.
- **Production scientifique** : La production scientifique de l'IRSN permettra d'alimenter un baromètre évolutif chaque année.
- **Expertise scientifique** : L'IRSN bénéficie d'une expertise reconnue dans les domaines de la radioprotection et de la sûreté nucléaire. Le BSO sera un gage supplémentaire de transparence, d'ouverture et de confiance pour le domaine sensible qu'est le nucléaire .

\* Source : [Ministère chargé de l'Enseignement Supérieur et de la Recherche](#)

STRENGTH



## Analyse SWOT – Freins au BSO institutionnel IRSN

- **Variété des supports** : Les supports variés (publications, chapitres d'ouvrages, communiqué dans un congrès, etc.) peuvent complexifier l'élaboration des indicateurs car les références à intégrer ne sont pas forcément les mêmes. A noter que ces typologies de documents n'ont pas toujours de DOI associés, rendant leur repérage compliqué.
- **Nouveauté des outils** : Les outils que l'on a choisi d'utiliser ne sont pas spécifiques à toutes les institutions et nécessitent une appropriation (langage python, Google Colab, Openalex).
- **Manque d'uniformité** : Il n'y a pas d'uniformité dans la méthodologie. Chaque établissement suit sa propre méthode et a ses propres outils, rendant compliqué la comparaison entre les BSO.
- **Questions encore en suspens** : Des questions restent encore sans réponses, faisant que la création du BSO se fera sur cette première année avec des erreurs possibles et/ou des références manquantes (question du DOI et des collections).
  - Question sur les DOI qui ne sont pas des DOI Crossref mais qui ont un identifiant HAL. Est-ce qu'ils remontent dans le BSO quand on active le filtre HAL ?
  - Si on envoi un seul fichier global avec toutes les structures, est-ce qu'on aura des statistiques pour chaque structure ?
  - Faut-il envoyer autant de fichier qu'il y a de structures ?

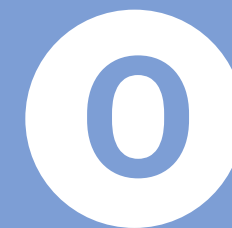
WEAKNESS



## Analyse SWOT – Opportunités au BSO institutionnel IRSN

- | **Développement de nouveaux outils** : Utilisation de nouveaux outils pour le suivi qui pourront être réutilisés et qui permettent une montée en compétence (pour mes collègues et moi) : requêtes Python ; base de données SQL
- | **Valorisation des résultats** : Créer un BSO pour renforcer la bibliométrie et la visibilité des publications IRSN en exploitant les outils HTML du MESRI en interne (intranet) et en externe (site ASNR, magazine institutionnel).
- | **Indicateurs bibliométriques** : Les indicateurs du BSO pourraient devenir de nouveaux outils bibliométriques pour l'évaluation par la recherche et être intégrés au Contrat d'Objectifs et de Performance (COP) de l'ASNR.
- | **Conformité aux directives nationales et européennes** : Répondre aux attentes des organismes financeurs qui soutiennent le mouvement de la science ouverte.
- | **Usage éducatif et de communication** : Nouveau support de sensibilisation des chercheurs et étudiants sur les enjeux de la science ouverte.

OPPORTUNITY



## Analyse SWOT – Menaces au BSO institutionnel IRSN

- **Fusion avec l'ASN** : Aucune politique n'a été (re)définie quant à l'Open Science, posant un voile d'incertitude sur la continuité du projet et pour le domaine de la recherche en général.
- **Evolution des structures** : Avec la fusion, sous quelle structure publieront les chercheurs ? Uniquement ASNR ou avec une identification IRSN ? Faut-il créer un nouveau BSO englobant ASNR et/ou le corpus IRSN ?
- **Résistance au changement** : L'introduction de nouvelles pratiques peuvent rencontrer des résistances au sein du SEARCH (nouvelle charge de travail), mais aussi une résistance quant au changement d'indicateurs de bibliométrie au niveau de l'institution.
- **Évolution technologique rapide** : Des outils ou méthodologies utilisés pour le baromètre pourraient rapidement évoluer et l'IRSN n'aura pas forcément la ressource humaine pour suivre.

THREAT



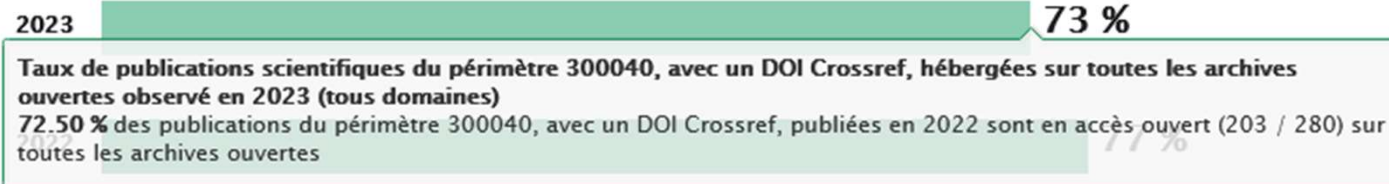


## Prospection du baromètre IRSN

Graphiques trouvables sur le [site du Baromètre de la Science Ouverte](#) avec l'identifiant IRSN **300040**

### [ RÉFLEXIONS

- Les graphiques proposés ne sont pas forcément parlant. Pour une population néophyte, les titres gagneraient à être simplifiés pour plus de compréhension.
- N'ayant pas eu la main sur les données prises en compte pour le BSO généré « automatiquement » pour l'IRSN, on a aucun visu sur les références qu'il englobe. On a un pourcentage et des nombres, mais pas de moyen de savoir de quelles publications il s'agit → Cela ne nous servira pas pour la constitution du fichier final.

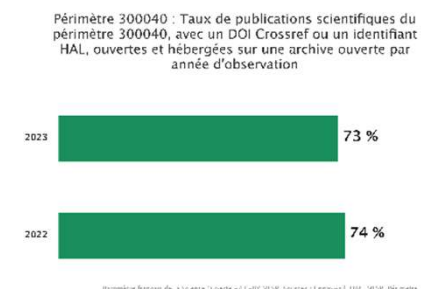
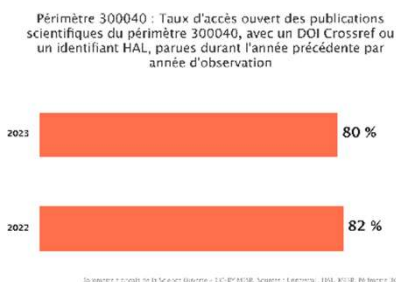
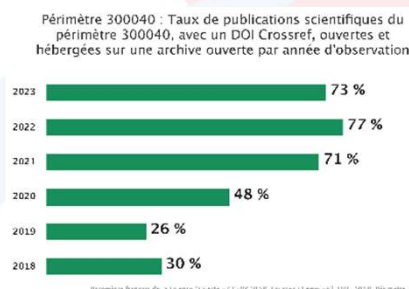
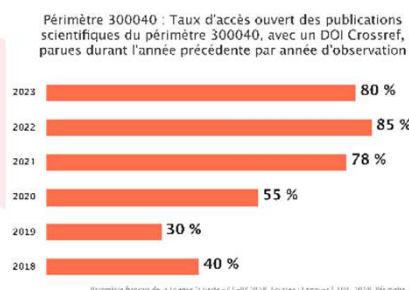


*Exemple des détails d'un graphique*

# Prospection du baromètre IRSN

J'ai pris le parti de focaliser mon attention sur 4 graphiques :

- **Taux d'accès ouvert des publications scientifiques**
  - avec un DOI Crossref
  - Avec un DOI Crossref ou un identifiant HAL
- **Taux de publications scientifiques ouvertes et hébergées sur une archive ouverte**
  - avec un DOI Crossref
  - Avec un DOI Crossref ou un identifiant HAL



Le choix de ces graphiques a été fait car les données peuvent être recoupées facilement. Il y a de nombreuses autres possibilités mais se concentrer sur l'indicateur des publications scientifiques reste le mieux pour appréhender le BSO. Ce seront aussi les principaux graphiques concernés par mon travail.

## Prospection du baromètre IRSN – Comparatif graphique 1/3



### CONSTATATIONS

- Bien que la **somme globale des publications soit identique**, il y a un **certain écart** entre le nombre d'accès ouvert des publications et le nombre de publications hébergées sur une archive ouverte.
- Le nombre de publications scientifiques ouvertes et hébergées sur une archive ouverte regroupe l'**addition de 25 archives ouvertes** françaises, européennes et internationales) + les sources **Unpaywall, HAL, MESR, Périmètre 300040**
- Le taux d'accès ouvert des publications scientifiques a pour sources **Unpaywall, HAL, MESR, Périmètre 300040**
- **Question** : Pourquoi le taux hébergé sur une archive ouverte est plus bas alors qu'il sollicite plus de sources ?
  - **Réponse** : Les publications sont déposées directement sur le site de l'éditeur et pas nécessairement une archive ouverte.

Année d'observation (année de publication)	Nombre d'accès ouvert des publications scientifiques (Crossref)	Nombre de publications scientifiques ouvertes et hébergées sur une archive ouverte (Crossref)	Nombre total de publications
2018 (pour 2017)	106	80	<b>267</b>
2019 (pour 2018)	96	83	<b>320</b>
2020 (pour 2019)	141	122	<b>256</b>
2021 (pour 2020)	211	192	<b>271</b>
2022 (pour 2021)	246	224	<b>291</b>
2023 (pour 2022)	225	203	<b>280</b>

## Prospection du baromètre IRSN – Comparatif graphique 2/3



### CONSTATATIONS

- Cette fois-ci les **identifiants de HAL** ont été **inclus**.
- En sélectionnant le filtre « inclure les identifiants de HAL », on a une **différence minime** par rapport au précédent tableau.
- En calculant la **différence** entre les 2 sources avec le tableau précédent, on trouve un **écart** de 6 et 12 documents « non référencés ».
- **Question** : Pourquoi l'antériorité ne concerne que 2021 et 2022 ?
  - **Réponse** : Depuis mars 2023, les publications disponibles dans HAL mais ne disposant pas de DOI peuvent être traitées dans le Baromètre.

Année d'observation (année de publication)	Nombre d'accès ouvert des publications scientifiques (Crossref + id HAL)	Nombre de publications scientifiques ouvertes et hébergées sur une archive ouverte (Crossref + id HAL)	Nombre total de publications
2022 (pour 2021)	252	230	<b>309</b>
2023 (pour 2022)	243	221	<b>304</b>

## Prospection du baromètre IRSN – Comparatif graphique 3/3



### CAS DU FILTRE “SÉLECTIONNER UNE ARCHIVE OUVERTE”

- Par curiosité, j’ai recensé le nombre des publications scientifiques hébergées sur une archive ouverte par année et pour chacune des 25 archives ouvertes disponibles dans le filtre « sélectionner une archive ouverte ».
- La dernière ligne en rouge correspond au nombre de publications avec le filtre par défaut « toutes les archives ouvertes ».
- On constate un écart entre les sommes totales qui ne devrait pas être et n’est pas explicable sans accès aux données.
- La présence de publications IRSN sur des archives ouvertes étrangères et/ou inconnue mériteraient une vérification de leurs métadonnées.

Date d'observation	2018	2019	2020	2021	2022	2023
HAL (France)	34	24	91	168	195	169
Pubmed Central (USA)	24	29	24	33	45	55
ArXiv (USA)	7	1	4	2	10	9
Archimer (France)		2	5	5	2	4
LIRIAS (Belgique)	1				2	3
LilloA (France)				1	1	3
US OSTP (USA)			4	3	1	3
dora		1		3	4	3
nora.nerc.ac.uk		2	1	5		3
Pure (Pays-Bas)		1				3
usir.salford.ac.uk				1		3
Research Square (Springer)				1	1	2
Zenodo (Europe)			1	4	2	2
bioRxiv (USA)				2	3	2
ora.ox.ac.uk				2		2
repositorio.ul.pt		1				2
upcommons.upc.edu			1	3	3	2
DiVA (Suède)	2		1	1	1	1
UCL Discovery (USA)	2	2	3		3	1
archive.ugent.be	1		2			1
cris.unibo.it						1
ddd.uab.cat						1
diposite.ub.edu	2					1
dspace.stir.ac.uk			1	1		1
egusphere.copernicus.org						1
Total/ différentes archives	73	63	138	235	273	278

Total / toutes les archives ouvertes	80	83	122	192	224	203
--------------------------------------	----	----	-----	-----	-----	-----

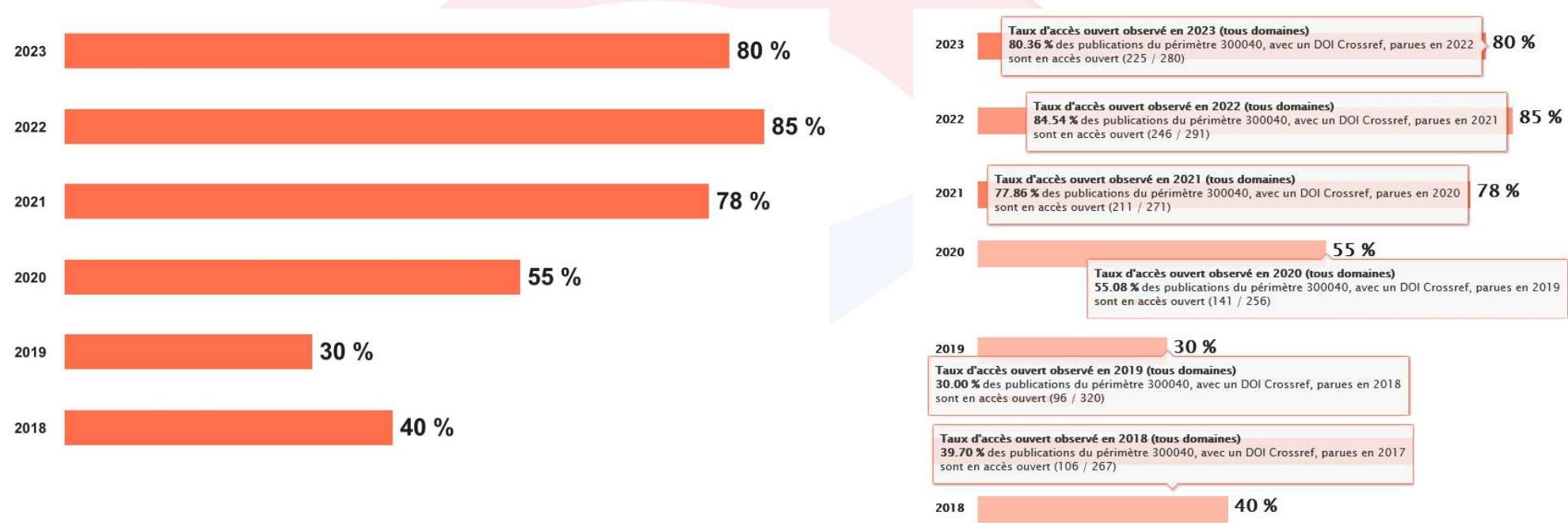
## Prospection du baromètre IRSN – Conclusions

*Ces conclusions sont à placer sous le spectre de la note d'étonnement. Il faudra les reprendre à la fin de ma mission pour voir si elles sont toujours d'actualité.*

- Peu importe les sélections, **les sources** données au bas de chaque graphique **restent les mêmes**. Il aurait été intéressant qu'on puisse savoir qu'elle(s) source(s) précisément le baromètre va chercher selon les **différents graphiques** et encore plus selon les **filtres** utilisés.
  - ❑ Le baromètre semble avant tout conçu pour offrir une vue d'ensemble statique, plutôt que pour fournir des analyses approfondies et interactives adaptées aux filtres ou sélections.
- Les chiffres présentés par le BSO IRSN **au moment de la rédaction** de la note d'étonnement **ne permettent pas** de les prendre en compte pour la mise en place du fichier d'intégration :
  - ❑ Impossibilité de savoir quels documents ils concernent → pas d'extraction de base de données possible
  - ❑ Les chiffres donnés ne sont pas les mêmes selon les filtres → manque de fiabilité

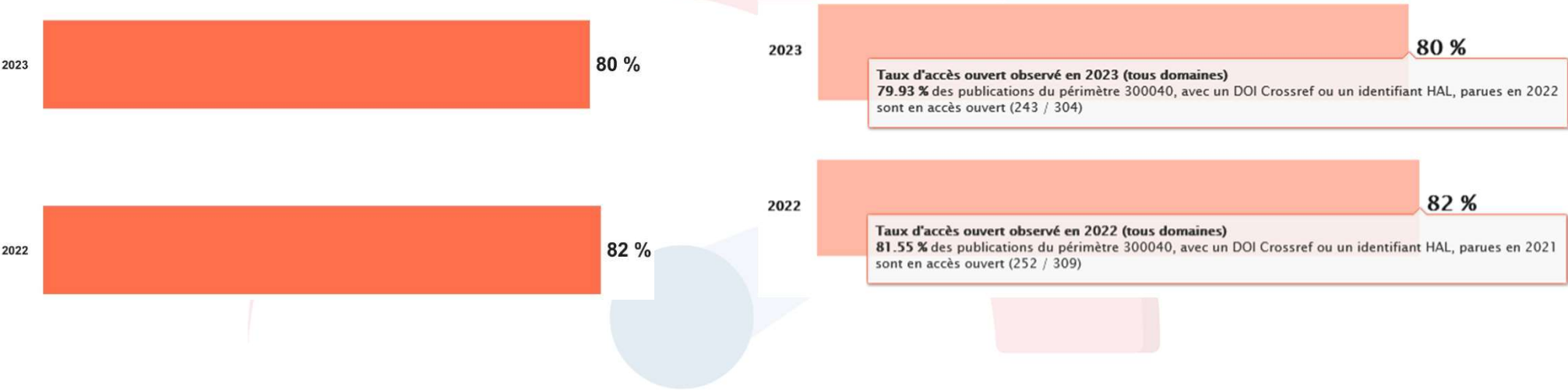
# État des lieux du baromètre IRSN – Novembre 2024

Taux d'accès ouvert des publications scientifiques du périmètre 300040, avec un DOI crossref, parues durant l'année précédente par année d'observation



# État des lieux du baromètre IRSN – Novembre 2024

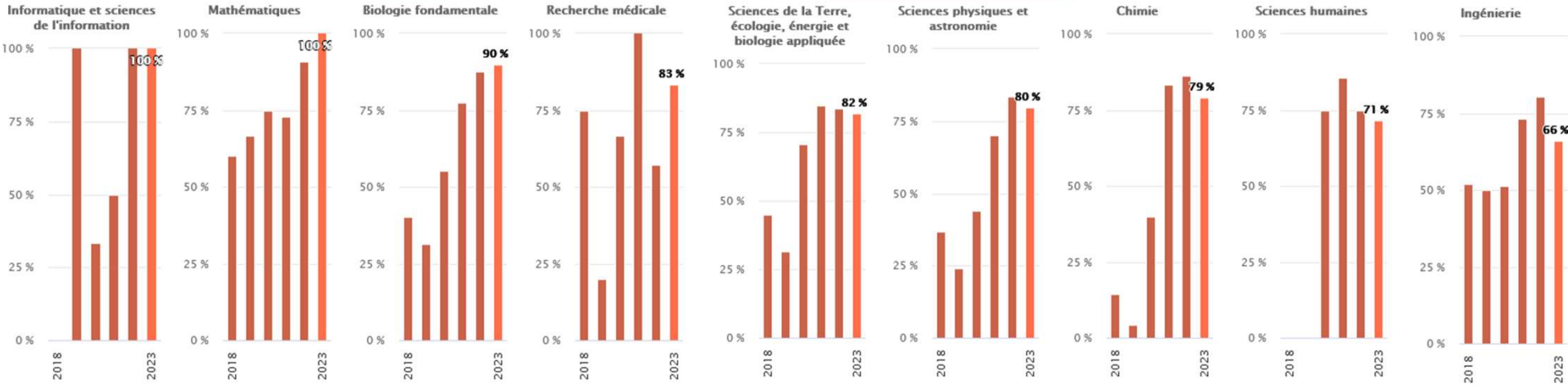
Taux d'accès ouvert des publications scientifiques du périmètre 300040, avec un DOI Crossref ou un identifiant HAL, parues durant l'année précédente par année d'observation





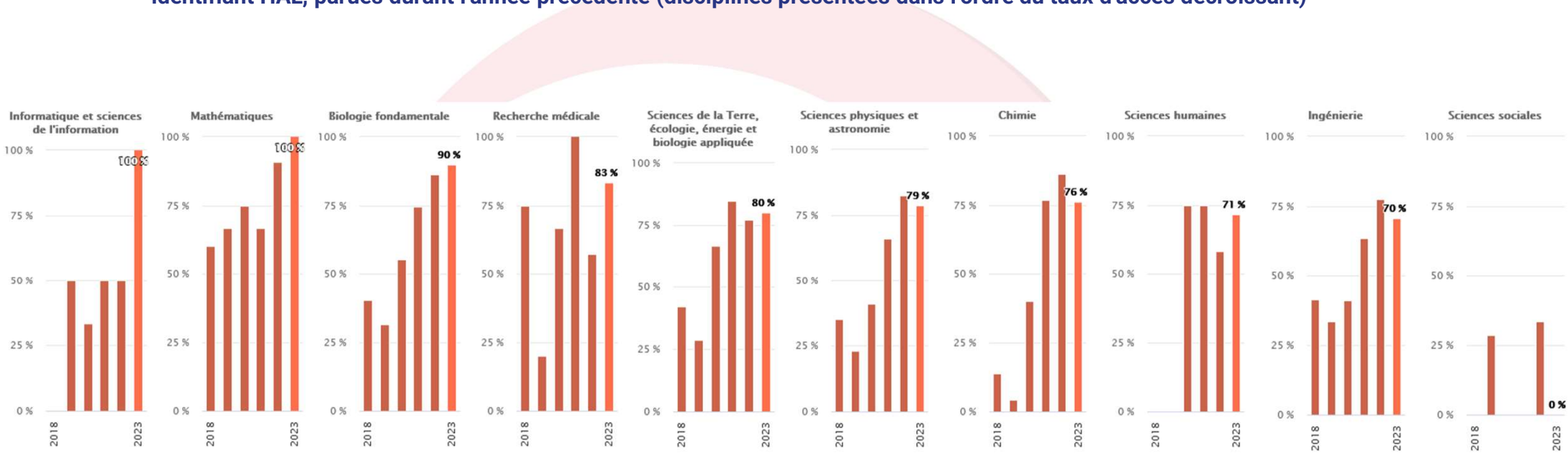
# État des lieux du baromètre IRSN – Novembre 2024

Taux d'accès ouvert par discipline et par année d'observation, pour les publications du périmètre 300040, avec un DOI Crossref, parues durant l'année précédente (disciplines présentées dans l'ordre du taux d'accès décroissant)



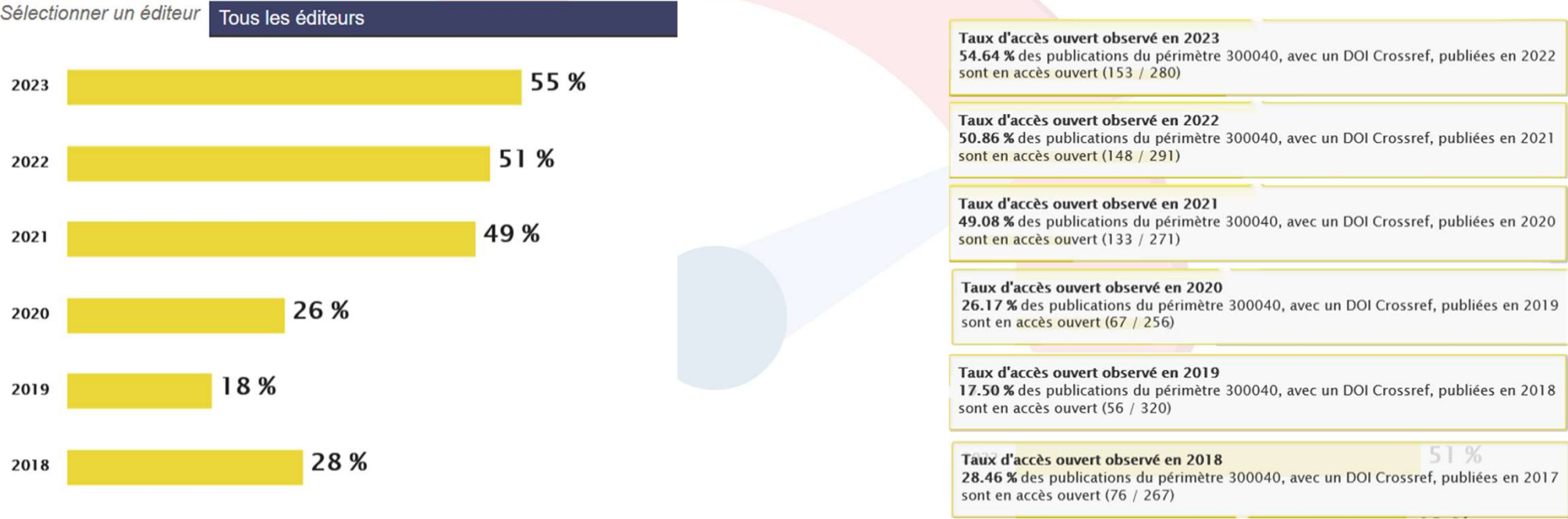
# État des lieux du baromètre IRSN – Novembre 2024

Taux d'accès ouvert par discipline et par année d'observation, pour les publications du périmètre 300040, avec un DOI Crossref ou un identifiant HAL, parues durant l'année précédente (disciplines présentées dans l'ordre du taux d'accès décroissant)



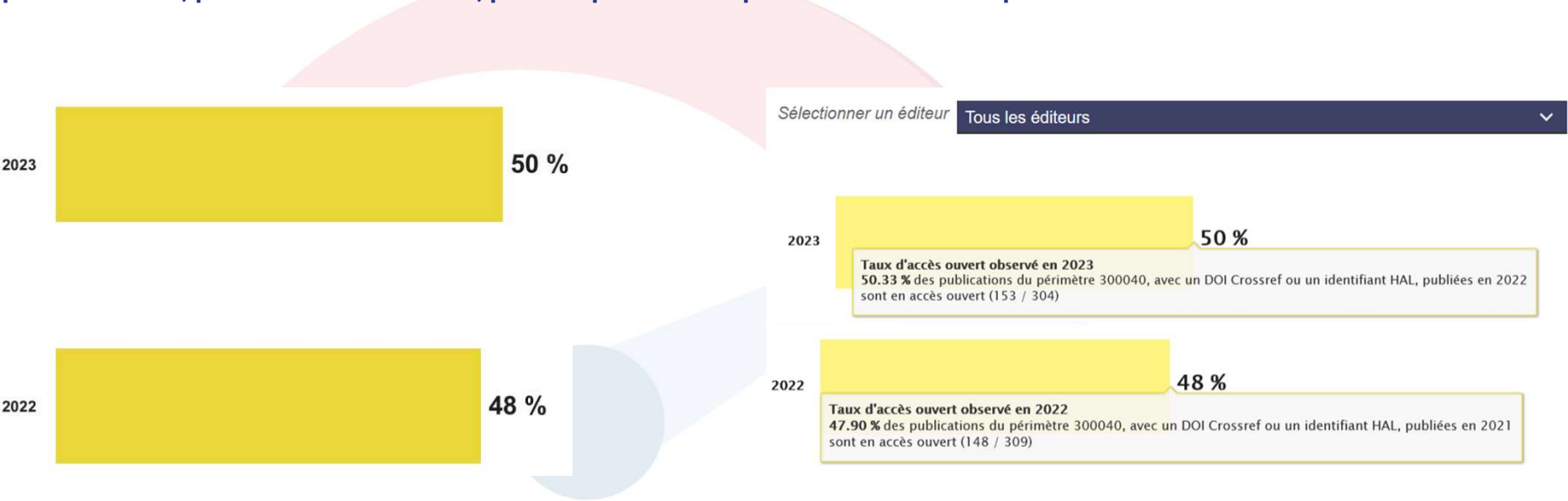
# État des lieux du baromètre IRSN – Novembre 2024

Part des publications scientifiques du périmètre 300040, avec un DOI Crossref, mises à disposition en accès ouvert par leur éditeur, par année d'observation, pour les publications parues durant l'année précédente



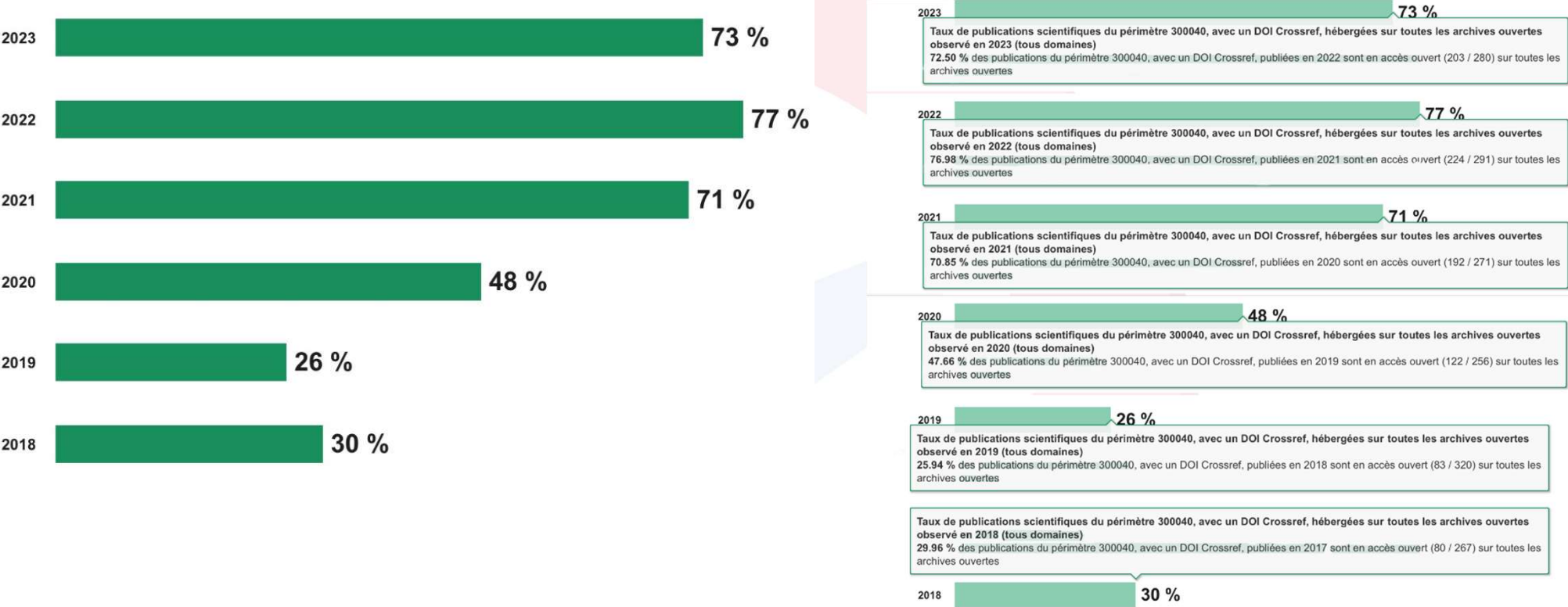
# État des lieux du baromètre IRSN – Novembre 2024

Part des publications scientifiques du périmètre 300040, avec un DOI Crossref ou un identifiant HAL, mises à disposition en accès ouvert par leur éditeur, par année d'observation, pour les publications parues durant l'année précédente



# État des lieux du baromètre IRSN – Novembre 2024

Taux de publications scientifiques du périmètre 300040, avec un DOI Crossref, ouvertes et hébergées sur une archive ouverte par année d'observation



# État des lieux du baromètre IRSN – Novembre 2024

Taux de publications scientifiques du périmètre 300040, avec un DOI Crossref ou un identifiant HAL, ouvertes et hébergées sur une archive ouverte par année d'observation

