

Pourquoi la Science ouverte ?

Un point de vue français

Marin Dacos – Conseiller pour la science ouverte
du Directeur général de la recherche et de l'innovation
Ministère de l'enseignement supérieur, de la recherche et de l'innovation - FRANCE



MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

25/10/2019



**PLAN NATIONAL
POUR LA SCIENCE OUVERTE**

MERCREDI 4 JUILLET 2018



La France s'engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement. "

Pourquoi la science ouverte ?

- Principe : Public money ? Public data !
- Scientifique : une recherche de meilleure qualité.
- Société : une science plus ouverte, plus accessible à la société.

Pourquoi la science ouverte ?



① **IMPACT : CITATIONS ET LECTURES**

② **LE “DERNIER KILOMÈTRE”**

③ **EFFICACITÉ ÉCONOMIQUE**

④ **HEURISTIQUE ! CUMULATIVITÉ DE LA SCIENCE VS REDONDANCE DES EFFORTS**

⑤ **ÉTHIQUE, INTÉGRITÉ, TRANSPARENCE**

Une crise de l'accès

HARVARD – YALE - IIS

“En 2008, la bibliothèque de Harvard était
abonnée à 98 900 périodiques

...contre 73 900 pour l’université Yale...

La bibliothèque de recherche indienne la mieux
financée, celle de l’Indian Institute of Science,
était abonnés à 10 600 périodiques.”

A dark grey world map with numerous red dots of varying sizes scattered across it, representing global locations where pirated papers are downloaded. The dots are most concentrated in North America, Europe, and parts of Asia.

Who's downloading pirated papers?

EVERYONE

In rich and poor countries, researchers turn to the Sci-Hub website.

« These statistics are based on extensive server log data supplied by Alexandra Elbakyan, the neuroscientist who created Sci-Hub in 2011 as a 22-year-old graduate student in Kazakhstan. I asked her for the data because, in spite of the flurry of polarized opinion pieces, blog posts, and tweets about Sci-Hub and what effect it has on research and academic publishing, some of the most basic questions remain unanswered: **Who are Sci-Hub's users, where are they, and what are they reading?** »

Bohannon, John. « Who's downloading pirated papers? Everyone ». *Science*, 28 avril 2016. doi:10.1126/science.aaf5664.

ÉTUDES QUALITATIVE DES USAGES SCI-HUB

■ Etude n°1

- Royaume-Uni, Etats-Unis d'Amérique, Malaisie, Chine, France
- 25% des chercheurs l'utilisent régulièrement

■ Etude n°2

- Six pays d'Amérique latine
- 20% des étudiants l'utilisaient en 2017
- Variations géographiques :
 - Argentine et le Paraguay : seulement 7%
 - Colombie et la Bolivie : supérieur à 34%

ETUDE QUANTITATIVE DES USAGES DE SCI-HUB

■ *The Lancet* étudie les usages de littérature médicale

■ Usages les plus élevés, en valeur absolue :

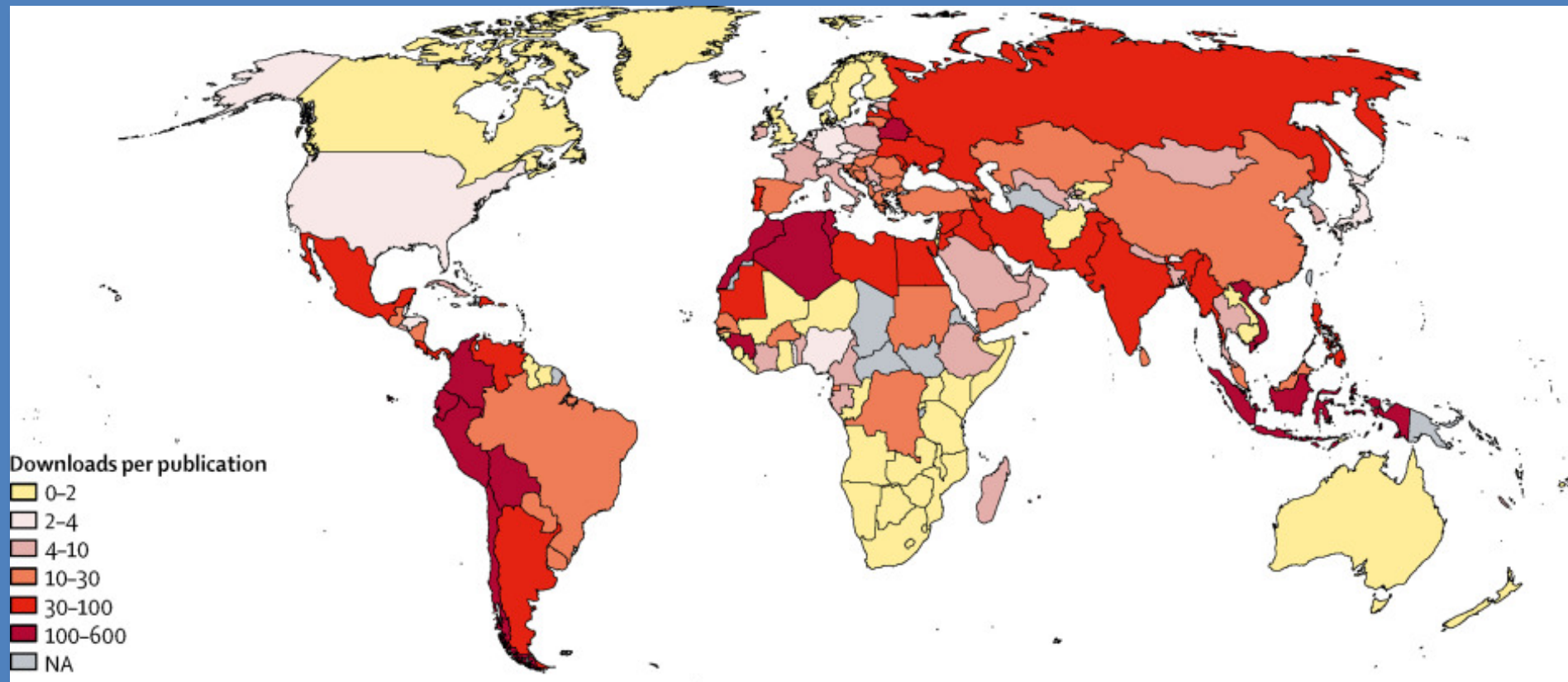
- Inde
- Chine
- Etats-Unis
- Iran

■ En valeur relative à la production de chaque pays :

En téléchargements par publication issue du pays

- Russie
- Algérie
- Brésil
- Turquie
- Mexique
- Inde

Les *lower-middle-income* et *upper-middle-income* utilisent beaucoup plus SCI-Hub, en proportion de leur production



QUELQUES ENSEIGNEMENTS

■ En résumé :

- La crise de l'accès est forte dans les pays riches, mais l'usage massif est lié à l'ampleur démographique de leur recherche
- Les pays les plus pauvres, qui produisent quantitativement beaucoup moins de publications, ont plus de mal à « activer » cette ressource pirate.
- Les pays intermédiaires utilisent beaucoup plus Sci-hub, en proportion de leur production, et s'en servent pour réduire la crise de l'accès.

■ Enseignements

- L'accès ouvert est une condition nécessaire mais pas suffisante.
- Aujourd'hui, les pays intermédiaires ont le plus à y gagner

L'accès
ouvert,
pour qui ?



WELCOME TO CHAOS !

« Il y a **très peu de personnes dans cette salle qui voudraient lire ces publications scientifiques**, et qui en fait en tireraient des conclusions erronées... Je le redis, il faut être très prudent avec l'idée séduisante que tout le monde pourrait lire tout : **cela pourrait nous mener au chaos** »

John Jarvis, Managing Director of Wiley Europe. 2004.

« There are **very few people in this room, who would want to read** some of this scientific information, and in fact draw wrong conclusions from it... I will say again; let us be careful because this rather enticing statement that everybody should be able to see everything **could lead to chaos.** »

John Jarvis, Managing Director of Wiley Europe. 2004.

WELCOME TO « TABLOID-BASED MEDICINE »

2007

Larry Reynolds, directeur de publication du *Journal of Animal Science*, mars 2007 :

« Parce que le public n'a aucune idée de la manière de lire, interpréter ou replacer dans son contexte la science publiée, un accès public immédiat conduirait à un usage sensationnaliste ou impropre de la science. »

R.W. Donnell, médecin, attaque le *New England Journal of Medicine* en 2007 :

Il l'accuse de pratiquer une « **médecine de tabloïd** » pour avoir mis en accès ouvert un éditorial et un article sur Avandia, suite à quoi des millions de lecteurs ont lancé une requête sur ce médicament dans Google

Why open access makes no sense

Robin Osborne

Debating Open Access, published 2013
by the British Academy

*« academic research
publication is a form of
teaching that assumes some
prior knowledge. For those who
wish to have access, there is an
admission cost. »*

*Robin Osborne is professor of ancient history, fellow and senior tutor at [King's College, Cambridge](#)
Cited in Eve, Martin Paul. *Open Access and the Humanities: Contexts, Controversies and the Future*.
University Printing House, Cambridge, United Kingdom: Cambridge University Press, 2014.*

THE PUBLIC IMPACT OF OPEN ACCESS RESEARCH: A SURVEY OF SCIELO USERS

ABOUT SCIELO (SCIENTIFIC ELECTRONIC LIBRARY ONLINE): 14 NATIONAL COLLECTIONS • 1,239 JOURNALS • 544,811 ARTICLES **ALL OPEN ACCESS**

ABOUT THE SURVEYS: SINGLE POP-UP QUESTIONS • 175,757 RESPONDENTS • 49% RESPONSE RATE • BRAZIL, MEXICO, CHILE

MOST PUBLICLY ACCESSED DISCIPLINES



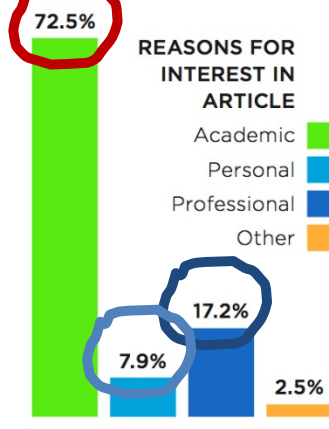
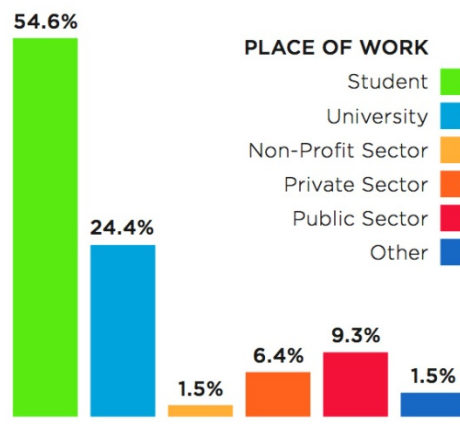
20% of all health science article downloads are for professional practice reasons.

Those interested for personal reasons access humanities and health sciences in equal proportions.

Private sector reads topics most different from academics.

Students and faculty research most similar topics.

OPEN ACCESS IS ALSO PUBLIC ACCESS - John Willinsky

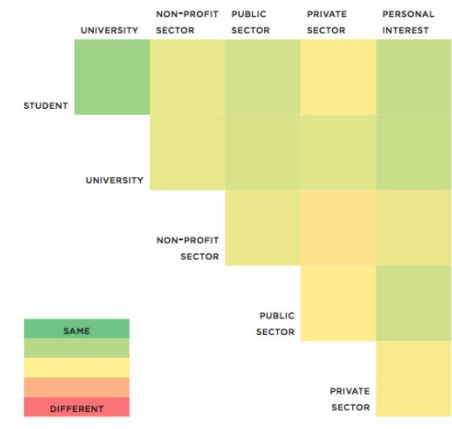


54.6%
STUDENTS

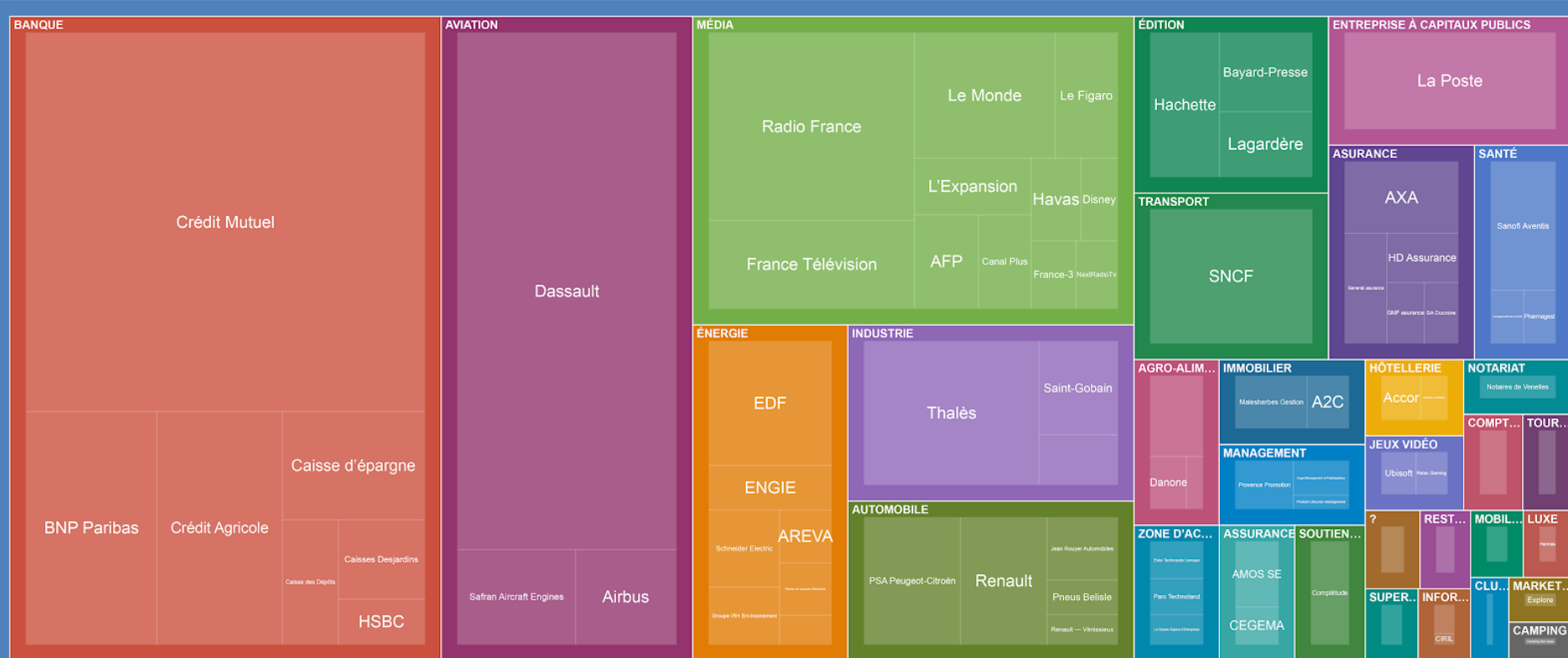
17.2%
RESEARCH FOR PROFESSIONAL REASONS

7.9%
PERSONAL INTEREST

SIMILARITIES IN TOPICS READ BY INTEREST GROUPS



Accès identifiés par des entreprises dans les 5000 IP consultant le plus OpenEdition



Grâce à l'accès ouvert, la plateforme OpenEdition, spécialisée en Sciences humaines et sociales, touche des entreprises dans de nombreux secteurs d'activités. Pour le Crédit Mutuel, par exemple, on identifie 99 003 accès sur la période 2017-mi 2019, portant principalement sur 18000 articles différents. C'est similaire à la consultation de l'université de Lausanne. Source : programme de recherche OpenEdition Lab, Appropriation du savoir ouvert, financé par l'IDEX d'Aix-Marseille Université. Publication à venir.

A la recherche du lecteur inattendu



HISTOIRES D'ACCÈS: SCI-HUB

Open ScreenShot

Nuuk, Groenland



Nuuk

Greenland

 Mostly Sunny · 17°C

6:41 PM

 SAVE

 NEARBY

 SEND TO YOUR PHONE

 SHARE

 Photos



Quick facts

Nuuk, Greenland's capital, is a small city on the country's southwest coast. Its large fjord system is known for waterfalls, humpback whales and icebergs. The waterfront is dotted with brightly colored houses against the backdrop of Sermitsiaq mountain. Greenland National Museum has mummies and Inuit skin boats, while the Nuuk Art Museum displays local works. The Katuaq cultural center offers films, concerts and art.

Population: 17,036 (2015)

Elevation: 5 m

Area: 690 km²

Science a mis en évidence des cas de lecteurs. Ici, un habitant de Nuuk, au Groëland, lit un article sur la meilleure façon d'apporter des traitements du cancer aux populations indigènes.

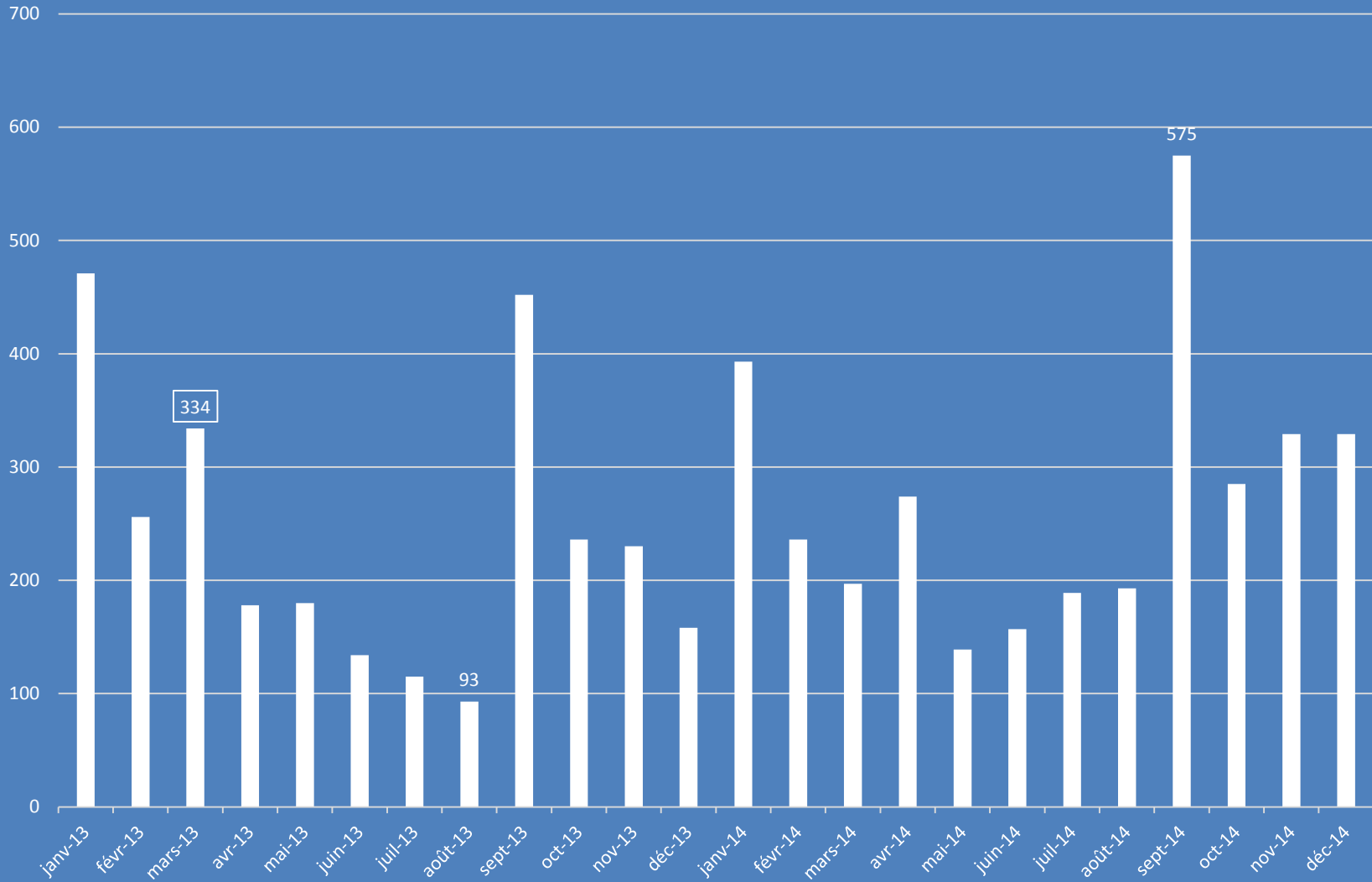
1028 (www.google.com/maps/place/Nuuk,+Greenland/@64.122425,-74.024487,3a/data=!3m1!1e3!3m1!1s0x46205611632723:0x04423545922051023054:1814039145:1564138719v=7m Jul 28 2017 16:41:00 GMT-0400 (EDT)

Map data ©2017 Google, INEGI Canada Terms Send feedback

500 km

LE LECTEUR INATTENDU

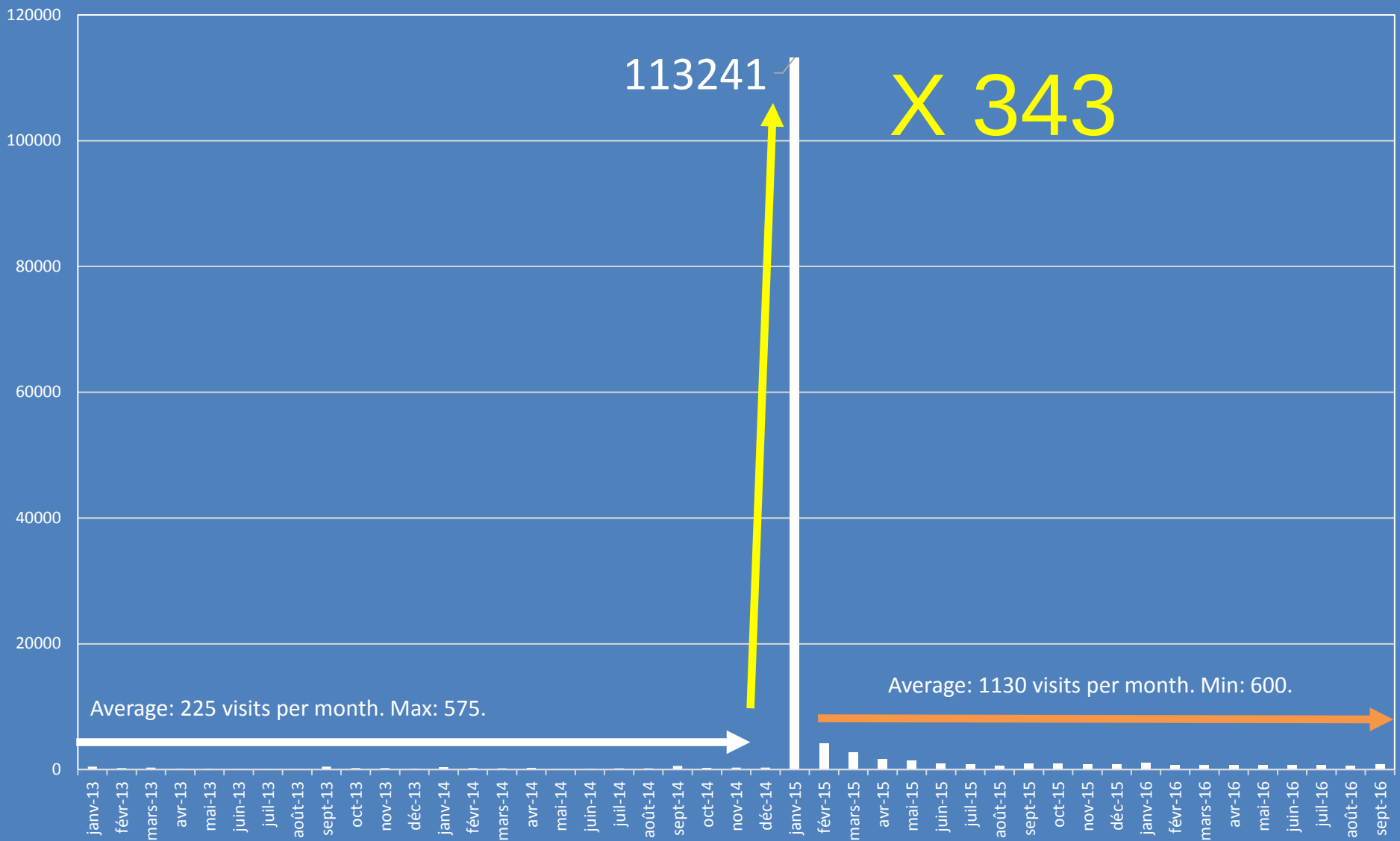
La représentation figurée du prophète Muhammad – Views/downloads



<https://ifpo.hypotheses.org/4445>

LE LECTEUR INATTENDU

La représentation figurée du prophète Muhammad – Views/downloads



ÉTUDES SUR WIKIPÉDIA (2017)

- Wikipédia (anglophone) cite
 - 300.000 articles scientifiques
 - 8.000 ouvrages de recherche
- Un article en accès ouvert a 47% de chances en plus d'être cité

#dataimpact

*Stories about the real-life impact
of Australian research data*

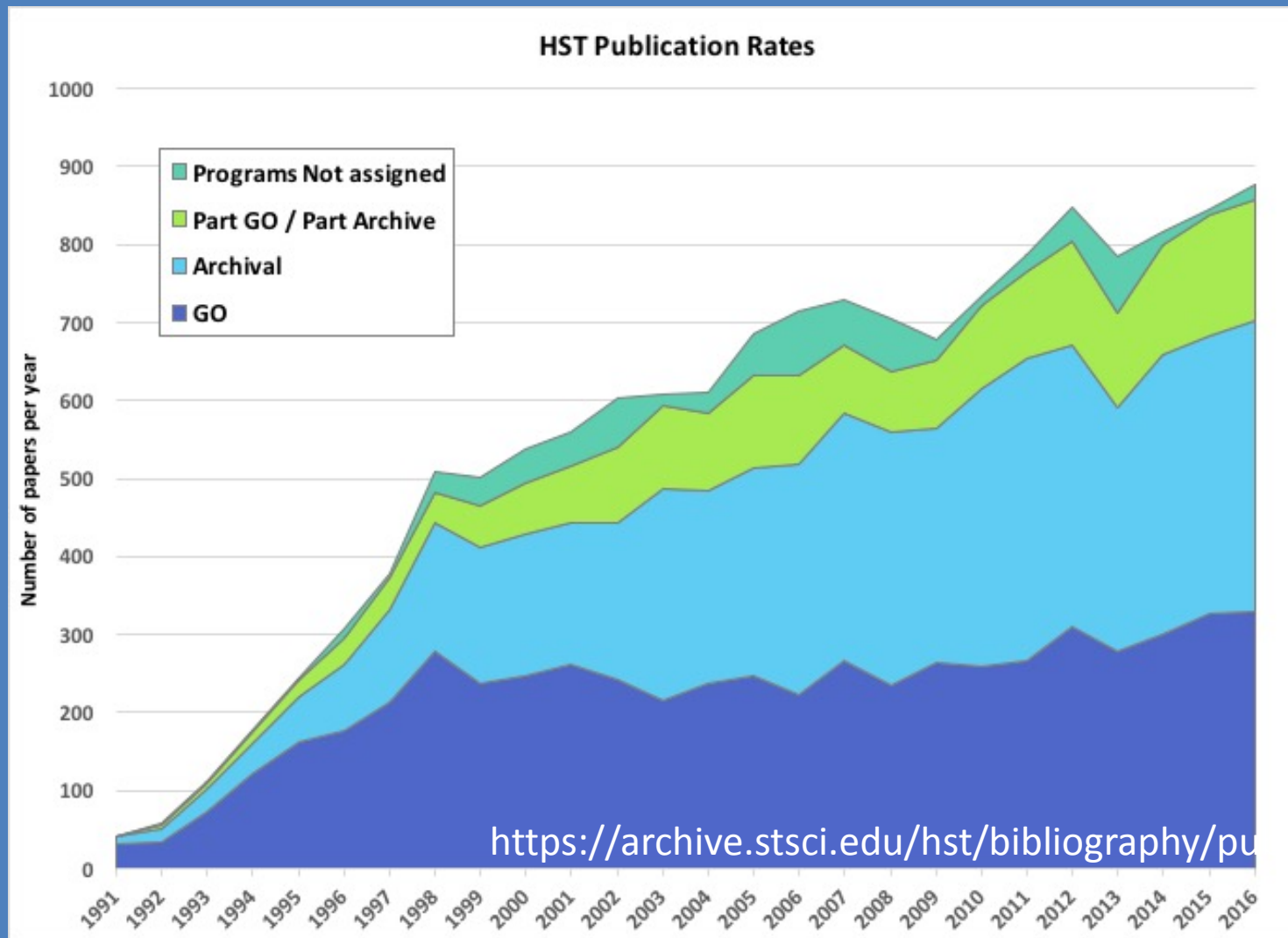
Les données, matière
première de la
connaissance
scientifique

Image : © ANDS

ands.org.au

ands[®]
AUSTRALIAN NATIONAL DATA SERVICE

Cas de l'astronomie : les données « archives » produisent plus de publications que la collecte initiale.



#dataimpact

*Stories about the real-life impact
of Australian research data*

Utilisation et réutilisation des données

Quelques exemples d'impact socio-économique de l'ouverture des données de la recherche.

L'exemple le plus célèbre est celui du génome humain, dont le partage public en temps réel a permis la création de valeur économique et de santé publique de très grande ampleur: coût de 3,6Md\$ pour le décodage, impact économique de plus de 700Md\$.

#dataimpact

*Stories about the real-life impact
of Australian research data*

Utilisation et réutilisation des données

C'est aussi vrai dans le secteur de l'agriculture, par exemple sur les données de phénotypage et génotypage de plantes (céréales par exemple) qui sont diffusées par l'INRA via un connecteur standardisé qui permet d'interconnecter des bases de données utilisées pour la sélection des plantes, par les industriels sélectionneurs et les organismes de recherche. Des hackatons sont organisés pour faciliter la réutilisation.

#dataimpact

*Stories about the real-life impact
of Australian research data*

Utilisation et réutilisation des données

La base de données collaborative d'origine française Open Food Facts, initiative de consommateurs, collabore avec la grande enquête scientifique NutriNet-Santé, et a apporté une contribution essentielle à la définition du Nutriscore, qui est désormais utilisé quotidiennement pour guider les consommateurs en France, en Belgique, en Espagne et en Allemagne.

#dataimpact

*Stories about the real-life impact
of Australian research data*

Utilisation et réutilisation des données

En sismologie, la majeure partie des données sont ouvertes et permettent ainsi une amélioration des connaissances et de la mesure des risques sismiques.

#dataimpact

*Stories about the real-life impact
of Australian research data*



Utilisation et réutilisation des données

Le médaillé de l'innovation Daniel Le Berre a développé un logiciel libre d'aide à la résolution de problèmes complexes, qui est désormais utilisé par de très nombreux industriels.

#dataimpact

*Stories about the real-life impact
of Australian research data*



ands.org.au

ands
AUSTRALIAN NATIONAL DATA SERVICE

Utilisation et réutilisation des données

Les résultats des enquêtes de l'INED sont partiellement publics et utilisables par de nombreux acteurs pour prendre des décisions stratégiques ou pour alimenter le débat public (presse). Même lorsque ces données ne peuvent pas être rendues publiques par respect pour la vie privée ou pour d'autres raisons, une bonne gestion des données de la recherche permet leur communication après arbitrage par un comité d'accès aux données.

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines,^{1,2,*} Arianne Y.K. Albert,³ Rose L. Andrew,¹ Florence Débarre,^{1,4} Dan G. Bock,¹ Michelle T. Franklin,^{1,5} Kimberly J. Gilbert,¹ Jean-Sébastien Moore,^{1,6} Sébastien Renaut,¹ and Diana J. Rennison¹

¹Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

²Molecular Ecology Editorial Office, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

³Women's Health Research Institute, 4500 Oak Street, Vancouver, BC V6H 3N1, Canada

⁴Centre for Ecology & Conservation Biosciences, University of Exeter, Cornwall Campus, Tremough, Penryn TR10 9EZ, UK

⁵Institute for Sustainable Horticulture, Kwantlen Polytechnic University, 12666 72nd Avenue, Surrey, BC V3W 2M8, Canada

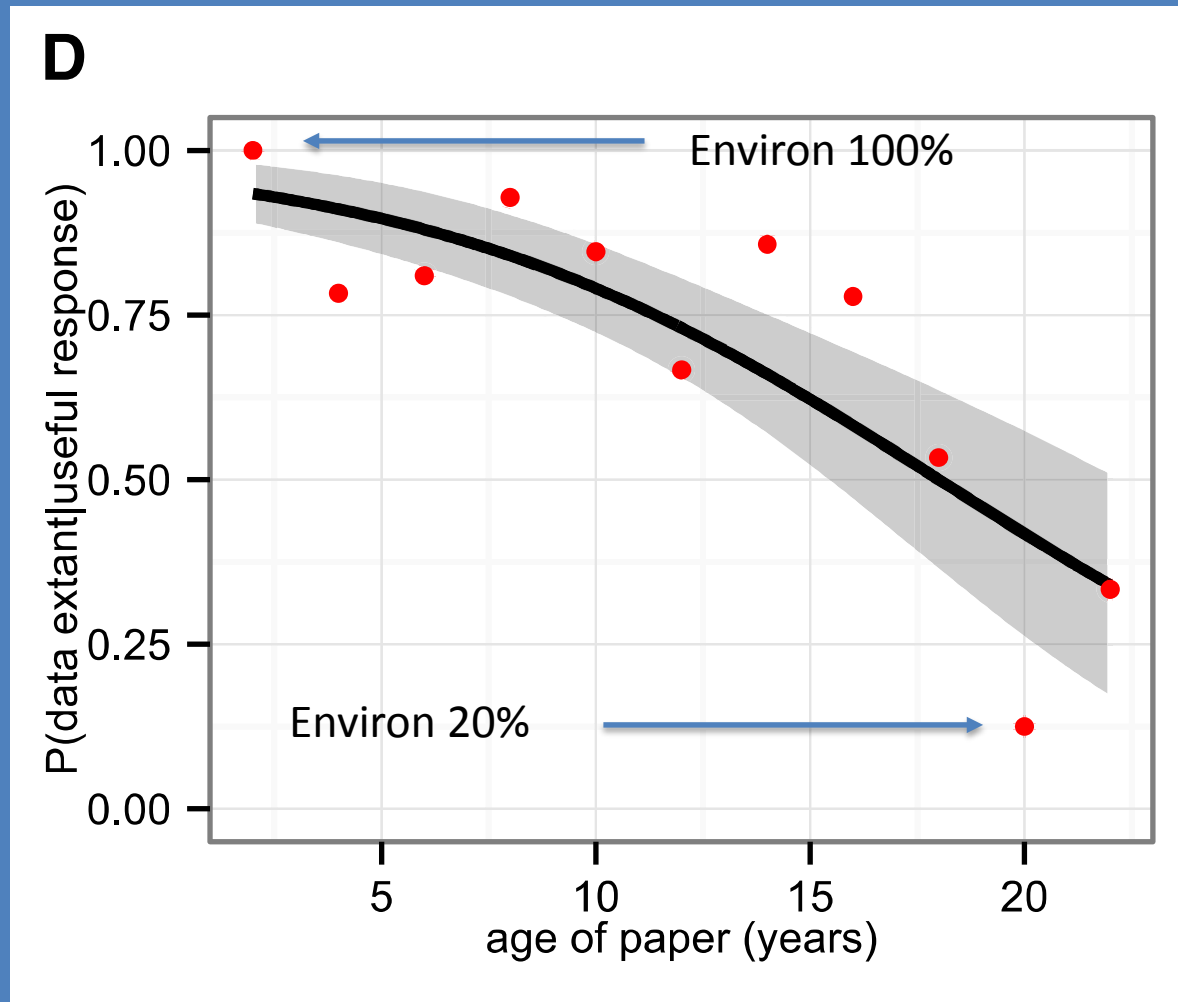
⁶Department of Biology, Université Laval, 1030 Avenue de la Médecine, Laval, QC G1V 0A6, Canada

sets (23%) were confirmed as extant. [Table 1](#) provides a breakdown of the data by year.

We used logistic regression to formally investigate the relationships between the age of the paper and (1) the probability that at least one e-mail appeared to work (i.e., did not generate an error message), (2) the conditional probability of a response given that at least one e-mail appeared to work, (3) the conditional probability of getting a response that indicated the status of the data (data lost, data exist but unwilling to share, or data shared) given that a response was received, and, finally, (4) the conditional probability that the data were extant (either “shared” or “exists but unwilling to share”) given that an informative response was received.

There was a negative relationship between the age of the paper and the probability of finding at least one apparently working e-mail either in the paper or by searching online (odds ratio [OR] = 0.93 [0.90–0.96, 95% confidence interval

LA BIBLIOTHÈQUE DES DONNÉES BRÛLE



(D) Predicted probability that the data were extant (either “shared” or “exist but unwilling to share”) given that we received a useful response. In all panels, the line indicates the predicted probability from the logistic regression, the gray area shows the 95% CI of this estimate, and the red dots indicate the actual proportions from the data.

#dataimpact

*Stories about the real-life impact
of Australian research data*



UK PRIVACY DEBACLES

« Between 24 April 2009 and 29 October 2009 alone, the loss of 878 513 records by 35 organizations is listed. Included in the list is the University of Manchester, after a member of staff emailed an attachment to 469 students with data on 1700 people including information on student disabilities; and Imperial College when six laptops were stolen resulting the loss of medical data containing confidential information on 6000 patients. »

UK Privacy Debacles

https://wiki.openrightsgroup.org/wiki/UK_Privacy_Debacles

#dataimpact

*Stories about the real-life impact
of Australian research data*

ands[®]
AUSTRALIAN NATIONAL DATA SERVICE

ands.org.au

« PUBLIC AVAILABILITY OF PUBLISHED RESEARCH DATA IN HIGH-IMPACT JOURNALS » - 2011

OPEN ACCESS Freely available online

PLOS one

Public Availability of Published Research Data in High-Impact Journals

Atawi A. Alsheikh-Ali^{1,2}, Waqas Qureshi³, Mouaz H. Al-Mallah^{4,5}, John P. A. Ioannidis^{2,6,7,8,9*}

1 Institute of Clinical Sciences, Sheikh Khalifa Medical City, Abu Dhabi, United Arab Emirates, **2** Department of Medicine, Tufts Clinical and Translational Science Institute, Tufts University School of Medicine, Boston, Massachusetts, United States of America, **3** Department of Medicine, Henry Ford Hospital, Detroit, Michigan, United States of America, **4** King Abdul Aziz Cardiac Center, National Guard Health Affairs, Riyadh, Kingdom of Saudi Arabia, **5** Wayne State University, Detroit, Michigan, United States of America, **6** Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, California, United States of America, **7** Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, **8** Biomedical Research Institute, Foundation for Research and Technology Hellas, Ioannina, Greece, **9** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America

Abstract

Background: There is increasing interest to make primary data from published research publicly available. We aimed to assess the current status of making research data available in highly-cited journals across the scientific literature.

Methods and Results: We reviewed the first 10 original research papers of 2009 published in the 50 original research journals with the highest impact factor. For each journal we documented the policies related to public availability and sharing of data. Of the 50 journals, 44 (88%) had a statement in their instructions to authors related to public availability and sharing of data. However, there was wide variation in journal requirements, ranging from requiring the sharing of all primary data related to the research to just including a statement in the published manuscript that data can be available on request. Of the 500 assessed papers, 149 (30%) were not subject to any data availability policy. Of the remaining 351 papers that were covered by some data availability policy, 208 papers (59%) did not fully adhere to the data availability instructions of the journals they were published in, most commonly (73%) by not publicly depositing microarray data. The other 143 papers that adhered to the data availability instructions did so by publicly depositing only the specific data type as required, making a statement of willingness to share, or actually sharing all the primary data. Overall, only 47 papers (9%) deposited full primary raw data online. None of the 149 papers not subject to data availability policies made their full primary data publicly available.

Conclusion: A substantial proportion of original research papers published in high-impact journals are either not subject to any data availability policies, or do not adhere to the data availability instructions in their respective journals. This empiric evaluation highlights opportunities for improvement.

Citation: Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011) Public Availability of Published Research Data in High-Impact Journals. PLoS ONE 6(6): e24367. doi:10.1371/journal.pone.0024367

Editor: Isabelle Boulton, University Paris Descartes, FRANCE

Received: May 17, 2011; **Accepted:** August 6, 2011; **Published:** September 7, 2011

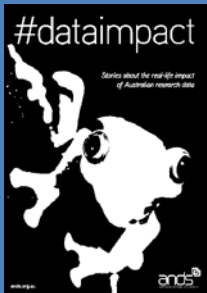
Copyright: © 2011 Alsheikh-Ali et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jioannid@hsph.harvard.edu

These authors contributed equally to this work.



« PUBLIC AVAILABILITY OF PUBLISHED RESEARCH DATA IN HIGH-IMPACT JOURNALS » - 2011

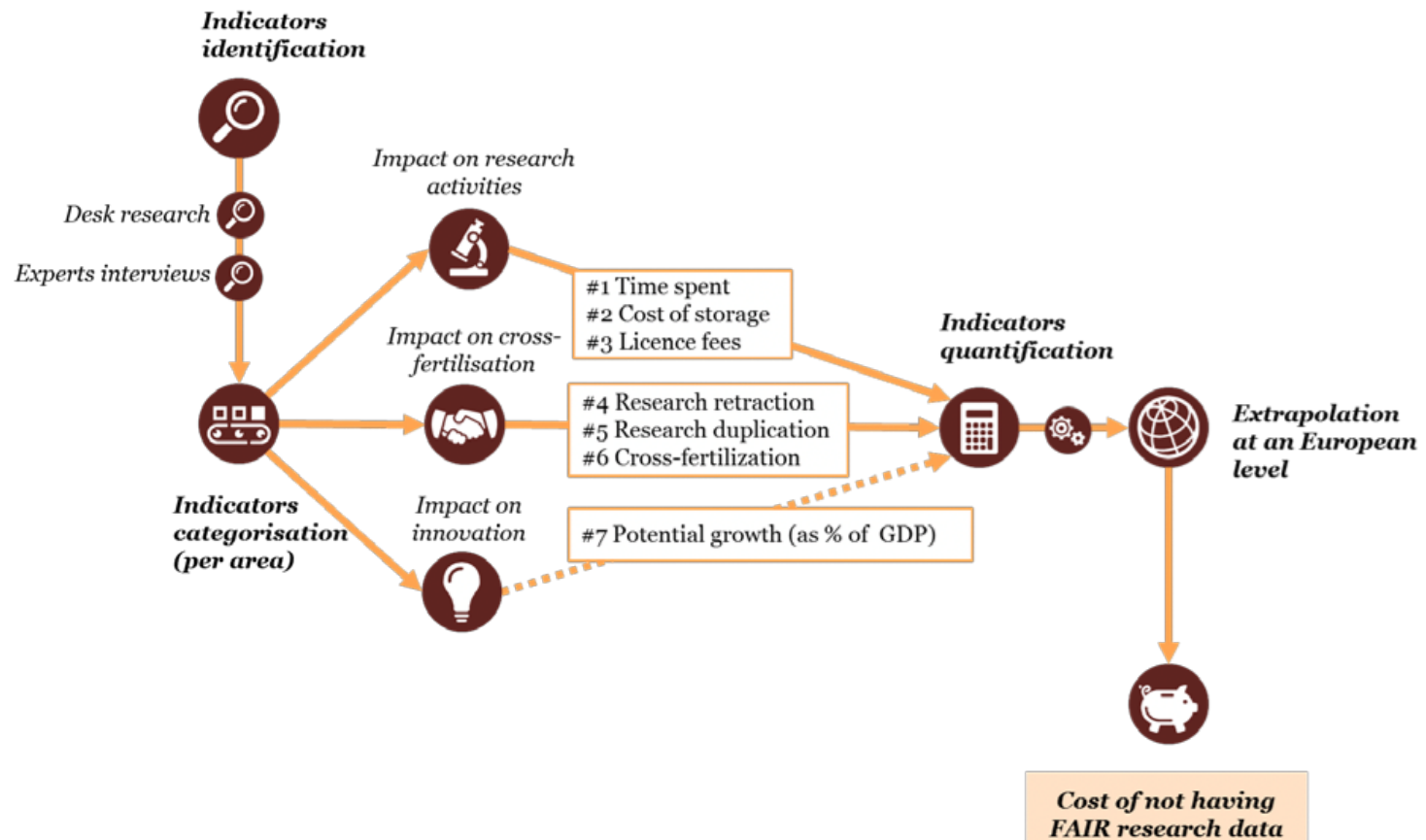
- Echantillon de 50 revues à « fort facteur d'impact »
- 88% des revues demandant l'association de données aux publications
- Sur 500 papiers...
 - ...30% n'ont pas à respecter de politiques concernant les données, et aucun ne partage ses données
- Sur 351 qui ont à respecter des instructions relatives aux données :
 - 59% ne respectent pas les instructions
 - 32% ne partagent que des données intermédiaires
 - 9% partagent les données primaires

« COST OF NOT HAVING FAIR RESEARCH DATA »



« COST OF NOTE HAVING FAIR RESEARCH DATA »

High-level representation of the methodology followed



RÉSULTAT DE L'ÉTUDE

- Dépenses de recherche en Europe : 302MM€ (2016)
- Cost of not having FAIR reseach data :
10,2MM€/an (=3% et 78% du coût annuel de H2020)
- Le coût de « fairisation » serait de 2,5% du coût de la recherche, soit différentiel positif de 0,5% = 2,6MM€ / an
- Perte potentielle pour l'innovation en Europe : 16MM€ (s'ajoutent aux 10,2MM€)

A photograph of a railway track winding through a forest in autumn. The ground is covered in fallen brown leaves. A yellow train is visible in the distance, approaching along the tracks. The trees are mostly bare, with some showing autumn foliage. The sky is overcast and hazy. The text "UNE AMBITION FRANÇAISE" is overlaid in large white letters across the bottom of the image.

UNE AMBITION FRANÇAISE

LES TROIS DÉCENNIES DE LA SCIENCE OUVERTE

- Années 1990: Décennie pratique
- Années 2000 : Décennie politique
- Années 2010 : Décennie politiques publiques

CONTEXTE JURIDIQUE EN FRANCE :

LOI POUR UNE RÉPUBLIQUE NUMÉRIQUE - OCTOBRE 2016

1. Publications : un nouveau droit pour les auteurs

Article 30 : Quand une recherche est financée à au moins 50% par l'Etat, l'auteur dispose de la faculté de déposer dans une archive ouverte le manuscrit auteur accepté, 6 mois (STM) ou 12 mois (SHS) après publication. Si le contrat dit le contraire, il est réputé non écrit.

2. Données : une nouvelle obligation pour les universités et les organismes

Article 6 : principe d'ouverture par défaut pour toutes les données administratives (dans lesquelles on compte les données de la recherche). Cela porte sur des données « achevées ».

PREMIER AXE : **GÉNÉRALISER L'ACCÈS OUVERT AUX PUBLICATIONS**



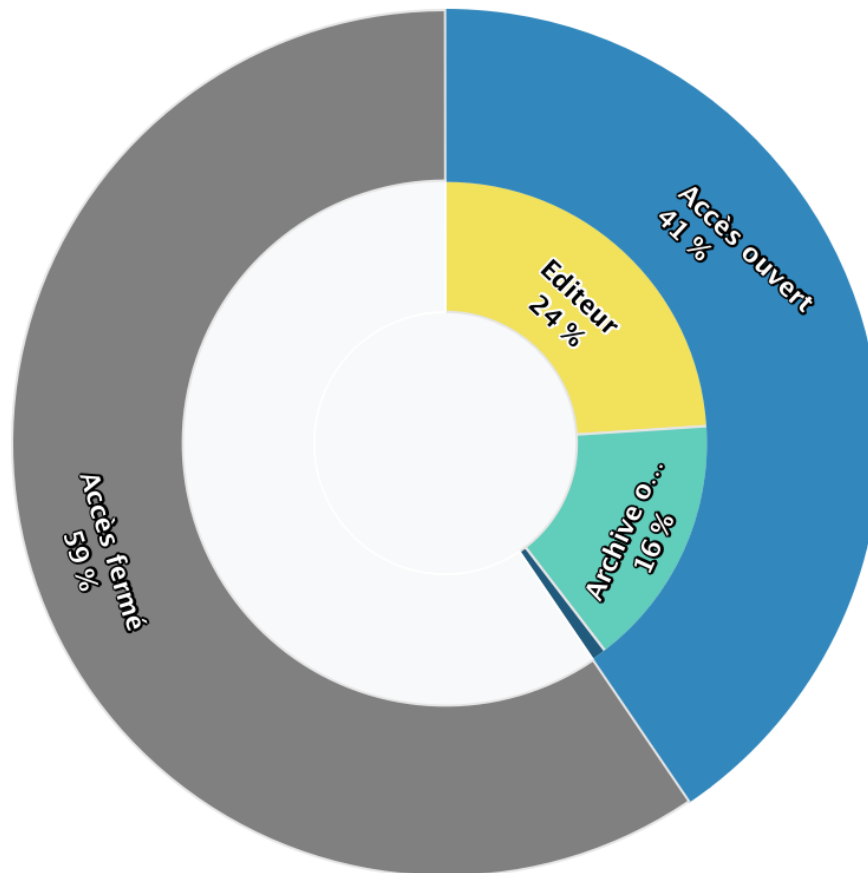
“ *La recherche scientifique est un bien commun
que nous devons partager avec tous.* **”**

BAROMÈTRE DE LA SCIENCE OUVERTE

Proportion des publications 2017 en accès ouvert (mesuré en 2018)

estimé à partir des publications détectées avec une affiliation française

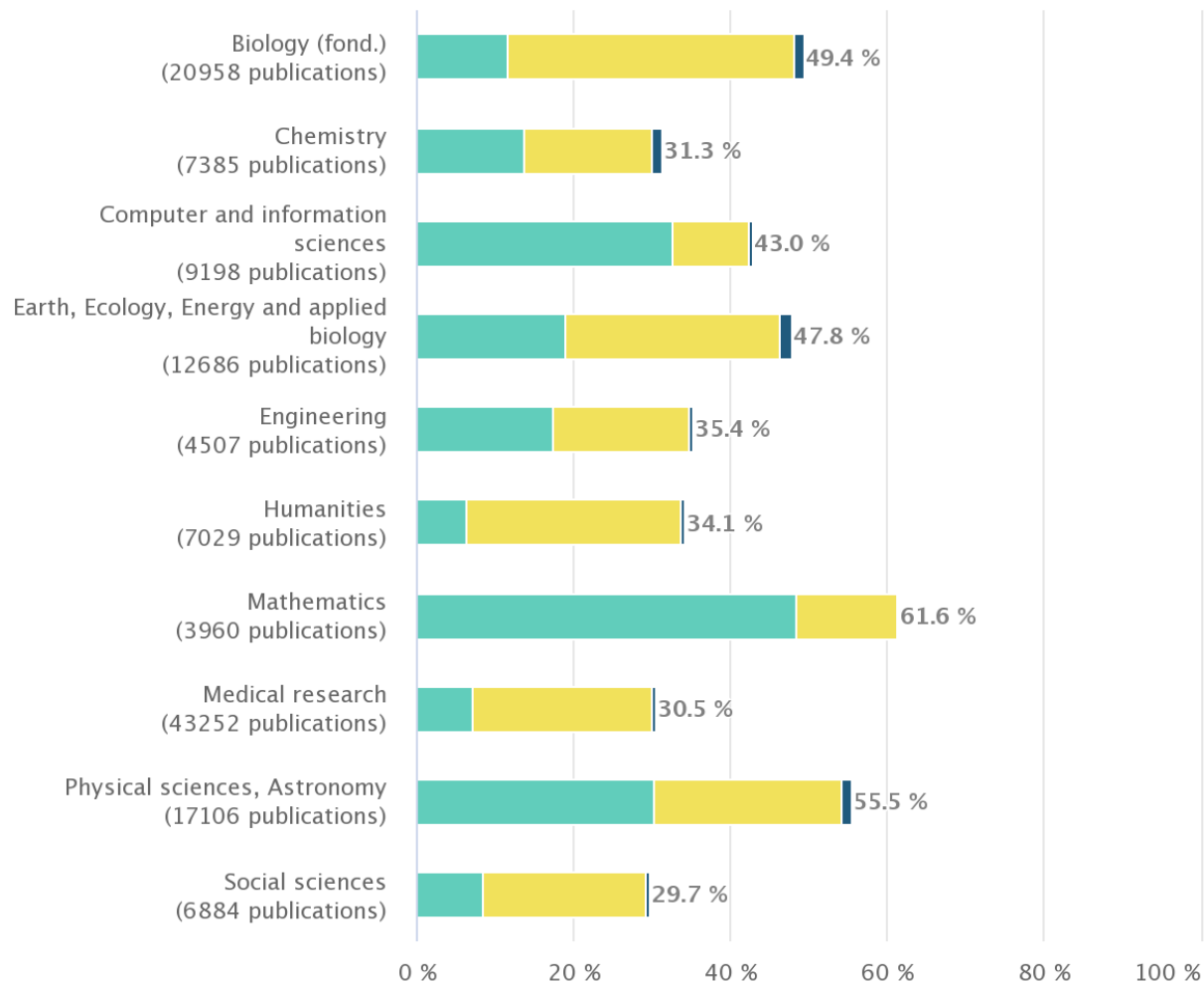
Source : Unpaywall, traitements MESRI



Taux d'accès ouvert aux publications 2017 par discipline (mesuré en 2018)

estimé à partir des publications détectées avec une affiliation française

Source : Unpaywall, traitements MESRI



Type d'hébergement

● Archive ouverte ● Editeur ● Non connu

DEUXIÈME AXE : **STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE**



“ *Les données de la recherche sont la matière première de la connaissance. Les partager, c'est ouvrir de nouvelles perspectives scientifiques.* **”**

Aussi ouvert que possible...

...aussi fermé que nécessaire

Prendre en compte la « longue traîne » *(long tail of data)*

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Mass.: The MIT Press, 2015, pp.9-10.

Retenir une approche
disciplinaire

TROISIÈME AXE : **S'INSCRIRE DANS UNE DYNAMIQUE DURABLE, EUROPÉENNE ET INTERNATIONALE**



La France s'engage pour que la science ouverte devienne la pratique quotidienne par défaut des chercheurs. "

LE COMITÉ POUR LA SCIENCE OUVERTE

OUVRIR LA SCIENCE !

OUVRIR LA SCIENCE !

LE COMITÉ POUR LA SCIENCE OUVERTE

14 personnes

A – Comité de pilotage de la science ouverte

Directeur général de la recherche et de l'innovation + PDG des organismes de recherche + Présidents des conférences des universités / grandes écoles / universités de recherche intensive...

24 personnes

B – Secrétariat permanent de la science ouverte

Représentants des membres du Comité de pilotage et experts issus des groupes

80 personnes

C- Collèges et groupes

Quatre collèges permanents : publications, données, compétences, international, et autant de groupes de travail que nécessaire.

300 personnes

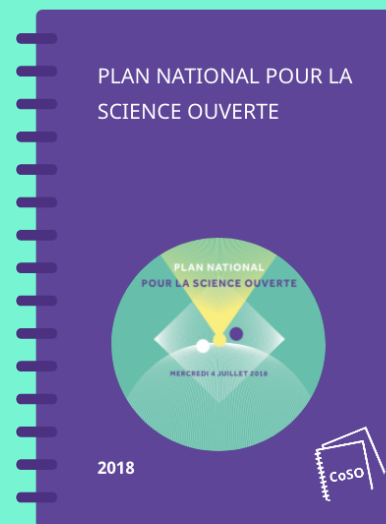
D – Forum de la science ouverte

Appel à candidatures. 50% chercheurs. 41% STM. 48% universités. 55% femmes.



**Que faut-il faire
pour que la science
soit plus ouverte ?**

DÉCOUVRIR LA SCIENCE OUVERTE



Merci !

PLAN NATIONAL POUR LA SCIENCE OUVERTE

MERCREDI 4 JUILLET 2018

“ La France s’engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement. ”

Contact :

Marin Dacos

marin.dacos@recherche.gouv.fr

www.ouvrirlascience.fr

Twitter @marindacos

#scienceouverte

 esr.gouv.fr

