



Initiation à la bibliométrie :

un outil pour piloter la recherche et la documentation

Solenn Bihan

Université de Lille, Direction Valorisation de la Recherche
Plateforme Lillometrics

Objectifs du stage

- Comprendre à quoi sert la bibliométrie et ses limites
- Savoir interpréter les indicateurs courants
- Produire des analyses simples
- Améliorer la qualité des données
- Transposer ces compétences dans son établissement pour proposer des services aux chercheurs

Se présenter

Présenter son établissement

- **Définition, usages et limites**

Définition

L'application des **mathématiques** et des méthodes **statistiques** aux **livres**, **articles** et autres moyens de communication.*

**Source : Pritchard A. Statistical Bibliography or Bibliometrics. Journal of Documentation. 1969;25(4):348-9.*

Cartographier
Décrire **Analyser** **Classer** **Valoriser**
Trier

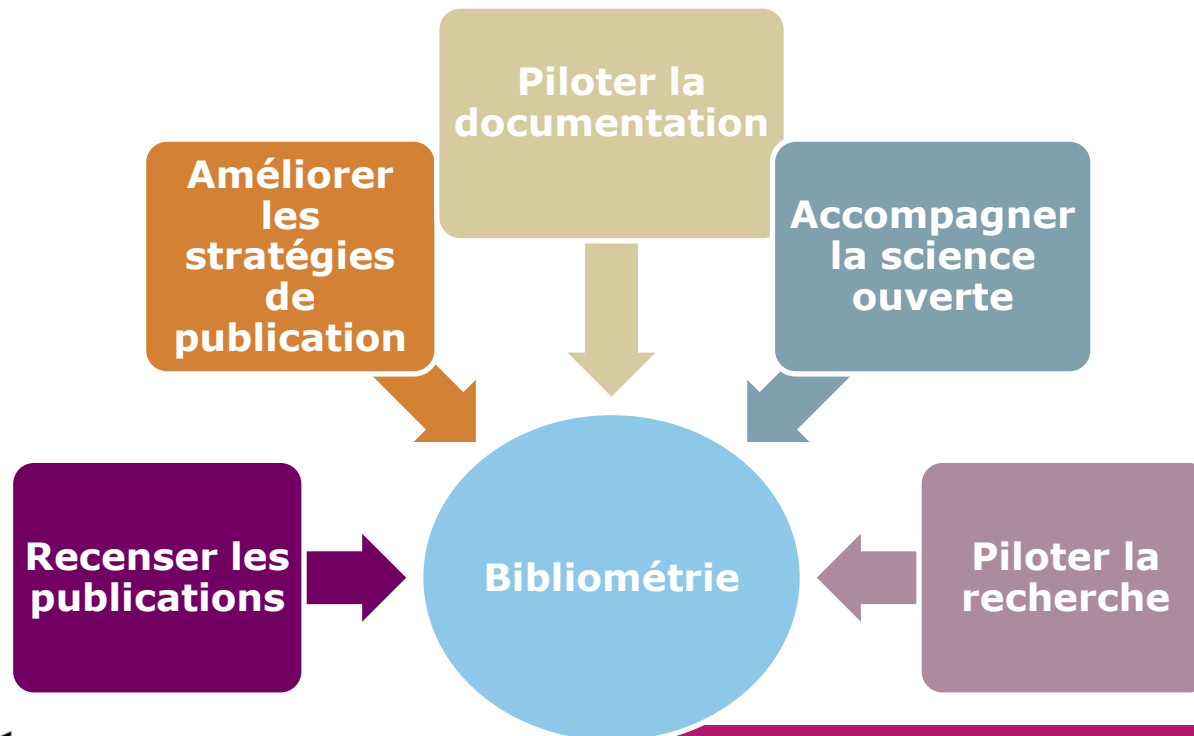
Dr Jekyll ou Mr Hyde ? Parmi ces usages, lesquels vous semblent légitimes ?

1. Mesurer le taux de publications déposées dans une archive ouverte pour évaluer la stratégie open access de l'établissement
2. Attribuer des financements aux chercheurs en fonction de leur h-index
3. Evaluer la progression des thèses en accès libre sur Thèses.fr
4. Dessiner des communautés d'intérêt grâce aux réseaux de citations
5. Choisir les articles les plus cités sur un sujet pour les lire en premier
6. Classer les universités en fonction de leur nombre de publications et de citations

Bibliométrie et évaluation de la recherche

Le problème, c'est l'évaluation de la recherche basée uniquement sur la bibliométrie.

L'évaluation de la recherche n'est qu'un usage parmi d'autres de la bibliométrie

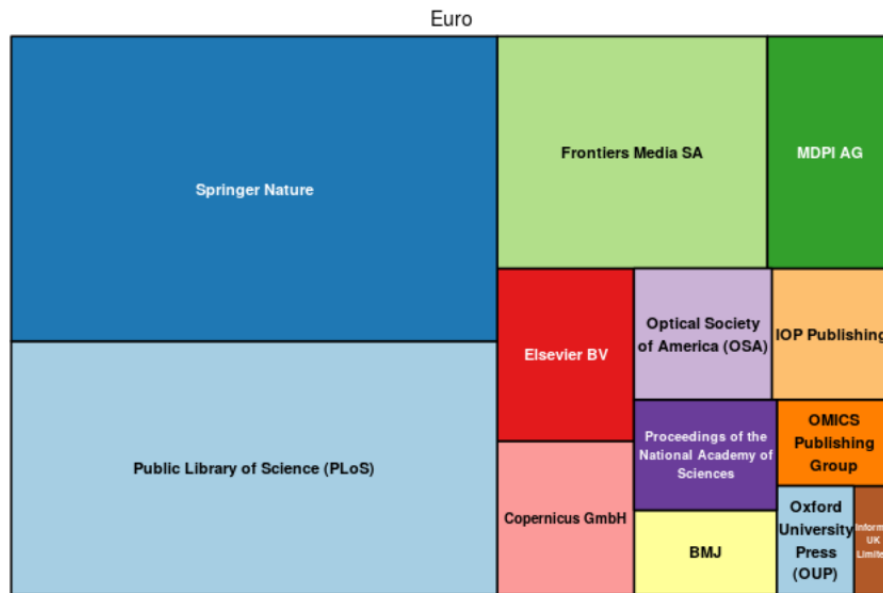


Examples

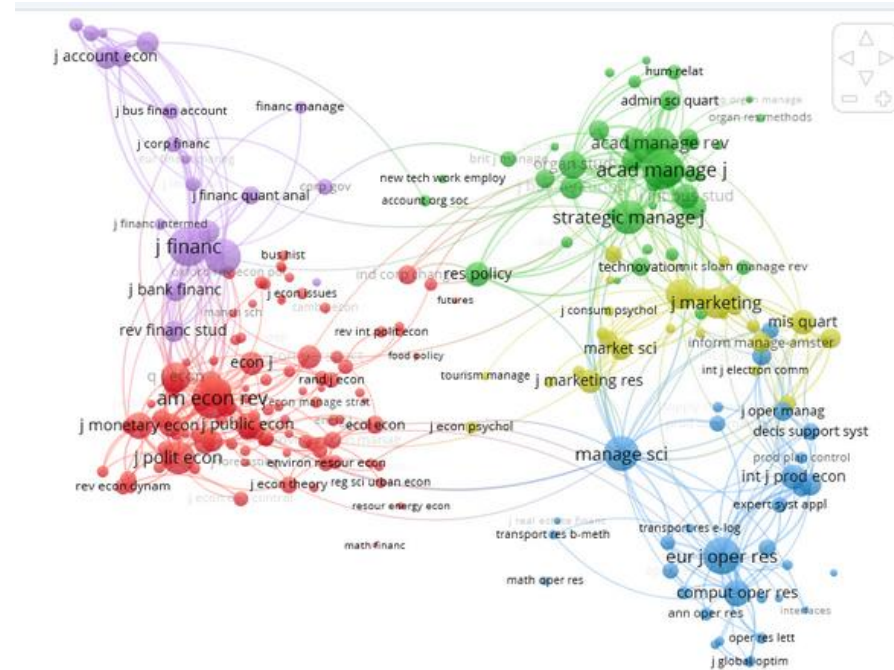
Etude des frais de publication en open access payés aux éditeurs

Analyse des revues en économie par les réseaux de citations

Fees paid per publisher (in EURO)



Source : OpenAPC. Données Couperin
<https://www.intact-project.org/blog/>



Source : Vosviewer
<http://www.vosviewer.com/features/examples>

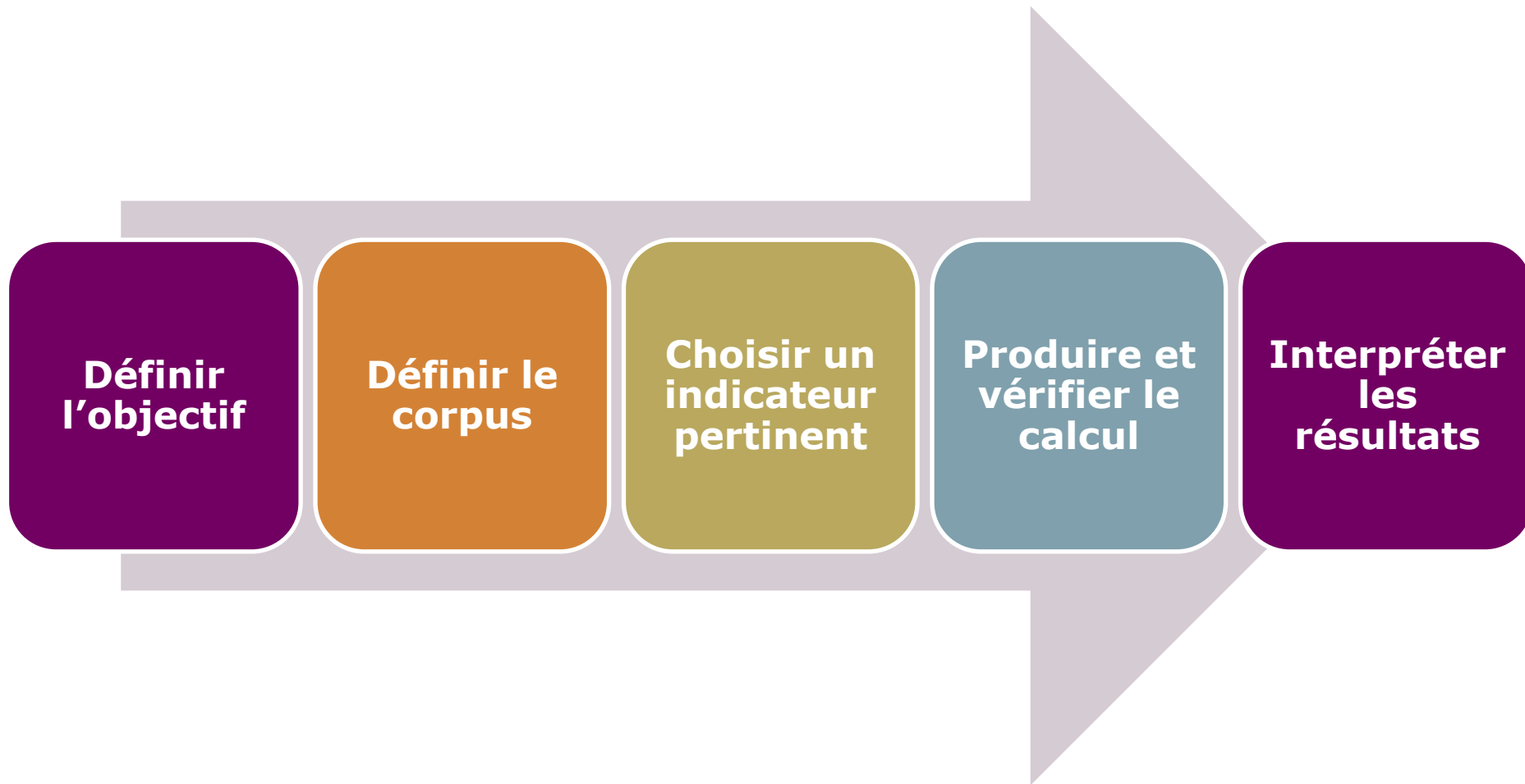
La bibliométrie scientifique, selon Yves Gingras

- Outil indispensable d'analyse des sciences et de leurs évolutions
- Qui suppose une démarche rigoureuse et validée
- Et une connaissance fine des modes de production

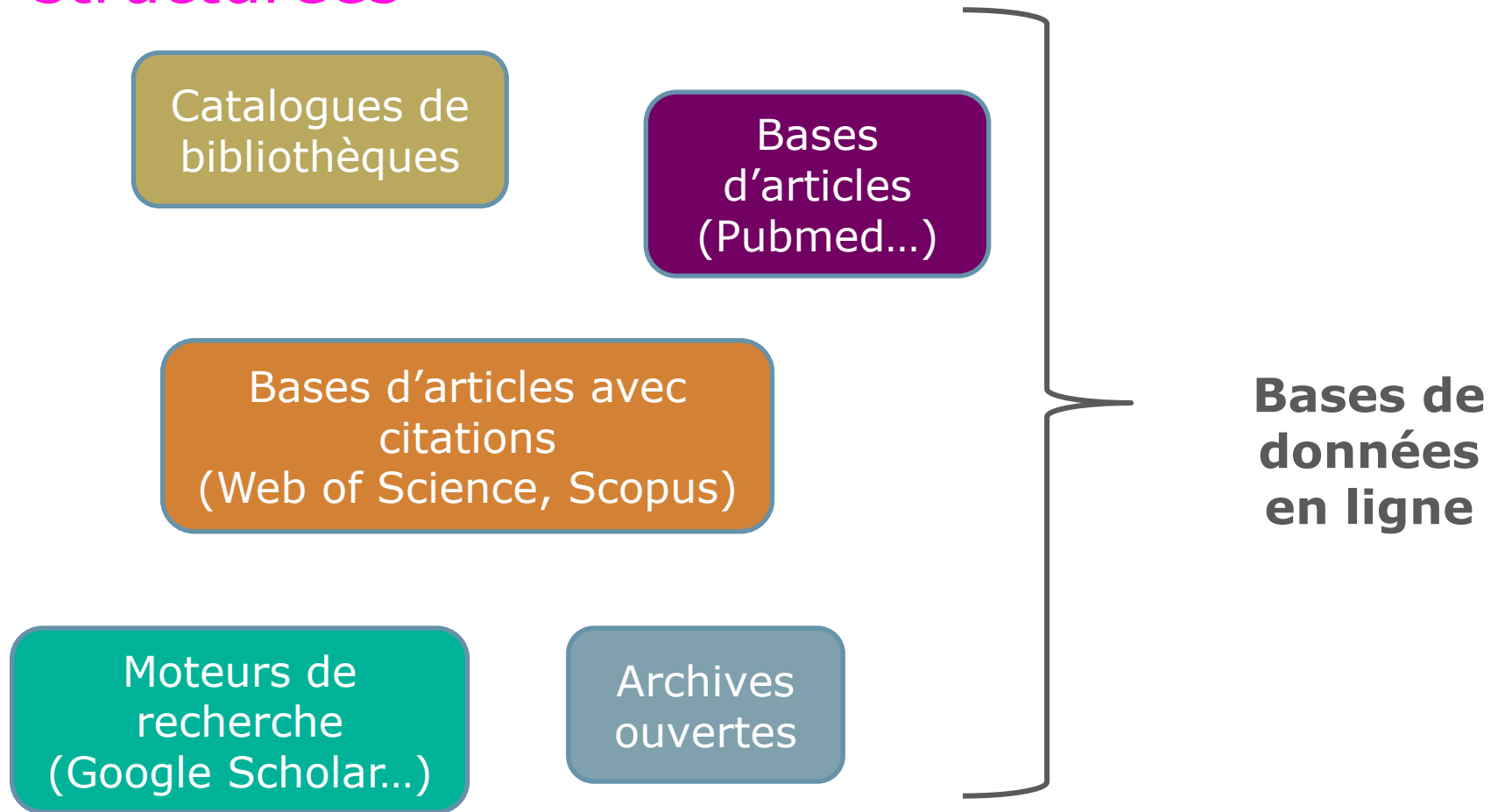


- **Méthode de la bibliométrie scientifique**

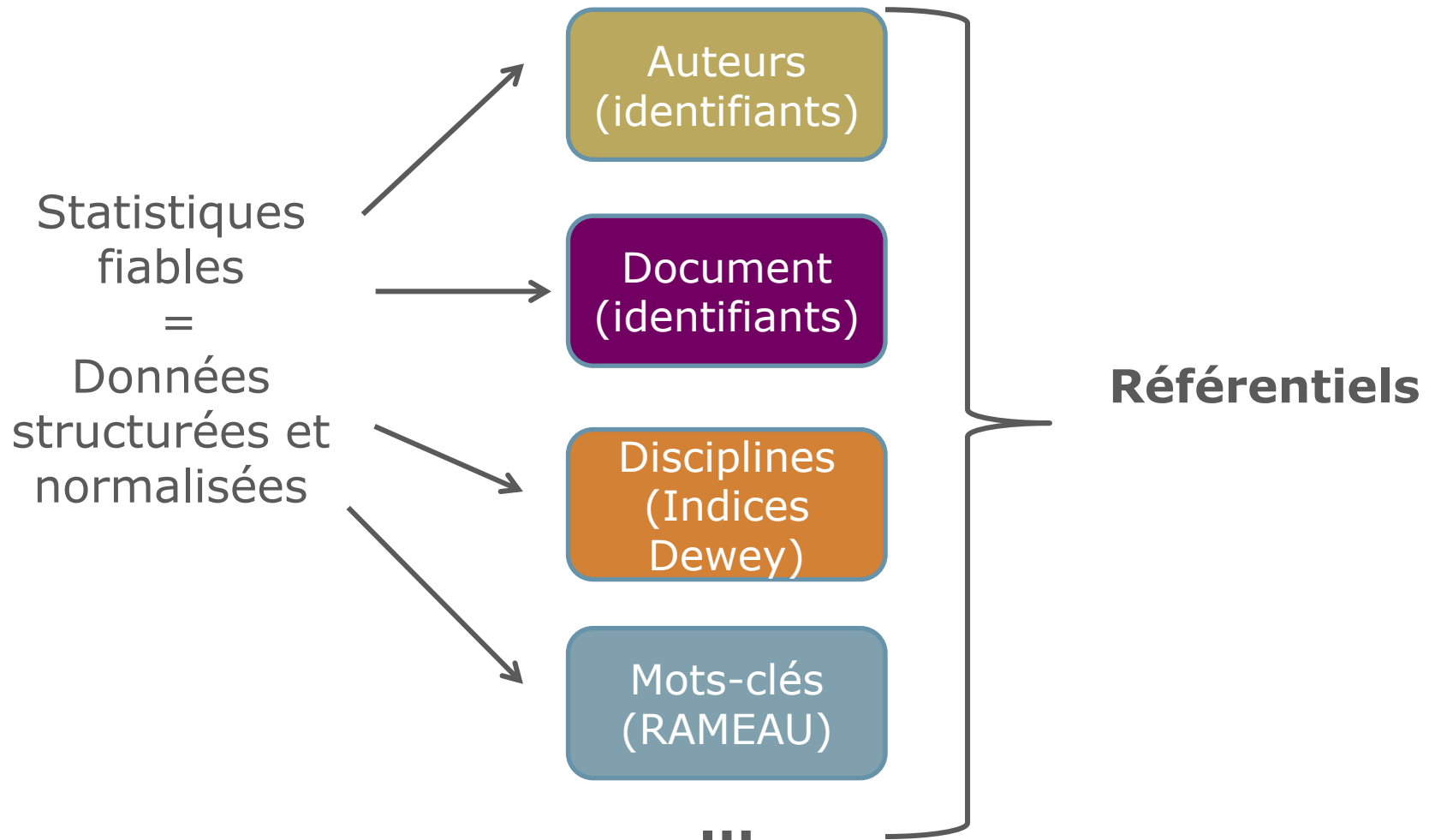
Méthode



Condition 1 : des données accessibles et structurées



Condition 2 : des données de qualité



Données de qualité = Référentiels

Un **référentiel**, c'est un ensemble de bases de données contenant les « références » d'un système d'information. Un référentiel clair, logique et précis est un des gages de bonne **interopérabilité** d'un système d'information.*

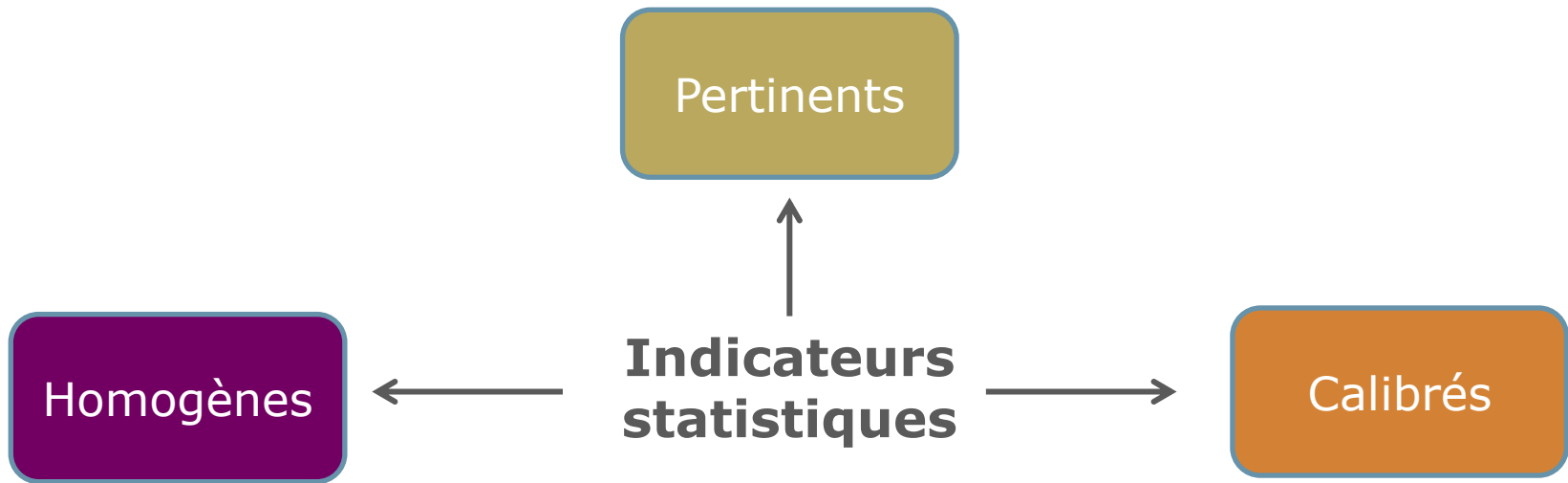
**Source : Wikipedia.*

Un référentiel est **partagé** entre plusieurs applications. C'est le garant de la **qualité** des données.

Un référentiel inclut des **références** (=un fichier d'autorités) qui sont administrées (=suivi, dédoublonnage...) + des **services** (=applications pour produire/réutiliser)

Ex : [IdRef](#) est le Référentiel des autorités Sudoc, Calames, Theses.fr, ISTEEX, Persée...

Condition 3 : des indicateurs fiables



L'indicateur n'est pas le **concept** lui-même mais une façon **approchée** de mesurer [...] le concept visé.*

**Source : Gingras Y, UQAM. La fièvre de l'évaluation de la recherche : du mauvais usage de faux indicateurs. 2008. Disponible sur: [http:// cpham.perso.univ-pau.fr/DOC/Gingras_Bibliometrie.pdf](http://cpham.perso.univ-pau.fr/DOC/Gingras_Bibliometrie.pdf).*

- **Les compétences nécessaires à la bibliométrie**

Etude de cas 1 : les thèses

Vous êtes responsable de la diffusion des thèses. Suite à des formations ou un accompagnement des doctorants, vous voulez savoir si la part de thèses en accès libre évolue favorablement.

- **Quelle proportion de thèses de votre établissement est en accès libre ? Comment ce nombre évolue-t-il depuis 3 ans ?**
- **Y a-t-il plus ou moins de thèses en accès libre dans votre établissement que dans les autres comparables ?**

Etude de cas 2 : l'informatique et les SHS

Vous avez rendez-vous dans un laboratoire pour promouvoir l'archive ouverte, et vous voulez savoir s'il est déjà impliqué dans l'open access : un peu ? beaucoup ? Quels sont les chercheurs moteurs ?

- **Combien de documents du laboratoire XXX sont déposés dans HAL ?**
- **Quel % en texte intégral ?**
- **Quelle évolution des dépôts selon les années ?**
- **Quels sont les champions de l'OA dans ce labo ?**

Bilan des études de cas

- Une méthode commune
- Applicable à toutes les disciplines et tous les types de documents
- Des contraintes:
 - Des sources de données accessibles et de qualité
 - Des valeurs de référence pour interpréter les résultats

Bibliométrie = fouille de textes et de données (TDM) sur des données bibliographiques

ISTEX

Améliorer la qualité des données bibliographiques

Priorités actuelles:

- Profils d'établissements et laboratoires (RNSR, IDHAL, IdRef, WoS & Scopus)
- Identifiants chercheurs (IdRef, IDHAL, ORCID, WoS & Scopus ...)

Mais aussi :

- Identifiants documents (DOI, ISSN, ISBN...)
- Thésaurus (MeSH, RAMEAU), classifications (Dewey, OCDE)

Projets en cours :



Repérage
pour
IPERU

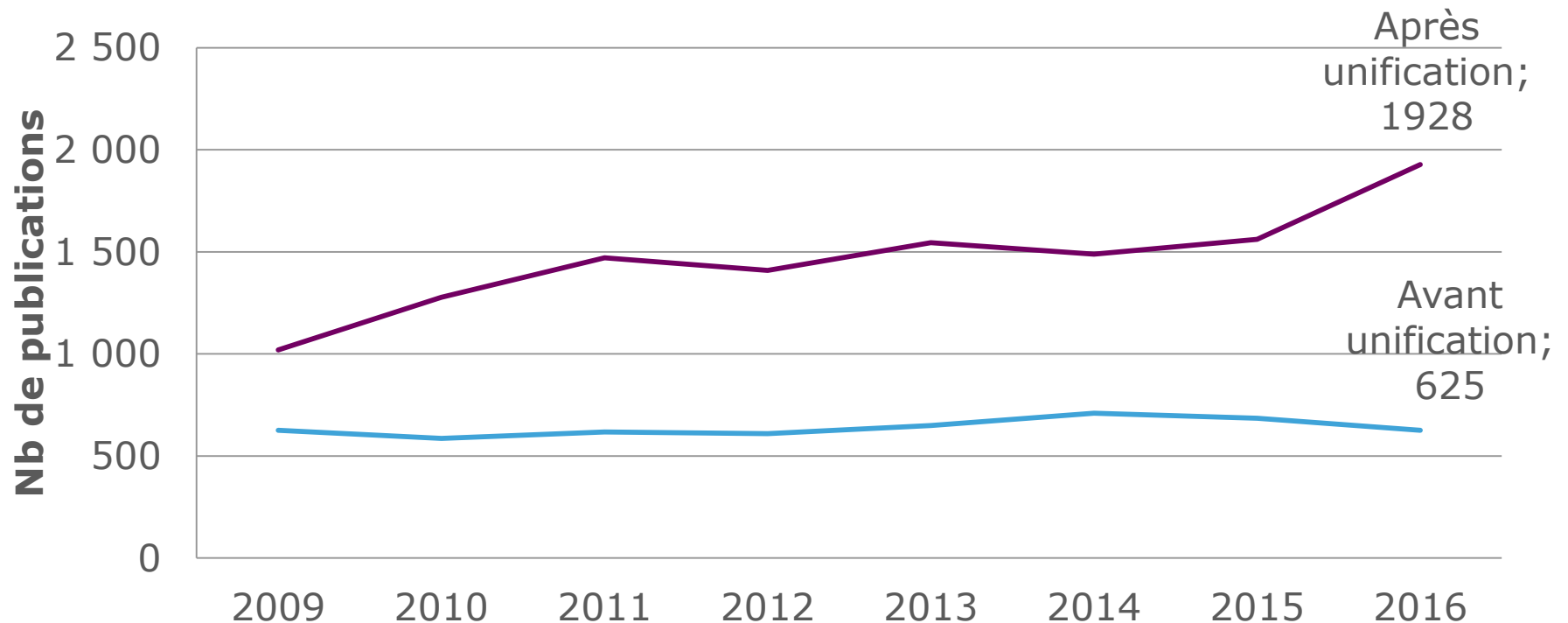


Observatoire des Sciences
et Techniques



Exemple d'amélioration des affiliations dans le Web of Science

Nb publications l'université Lille 2 dans le WoS X3 après l'unification des affiliations en 2017



Les compétences liées à la bibliométrie

Bibliographie et paysage éditorial

- Connaissance générale de l'écosystème éditorial, des différentes publications selon les communautés
- Maîtrise des bases bibliographiques et de leur structuration
- Habitude des requêtes avancées dans les bases bibliographiques

Administration de la recherche

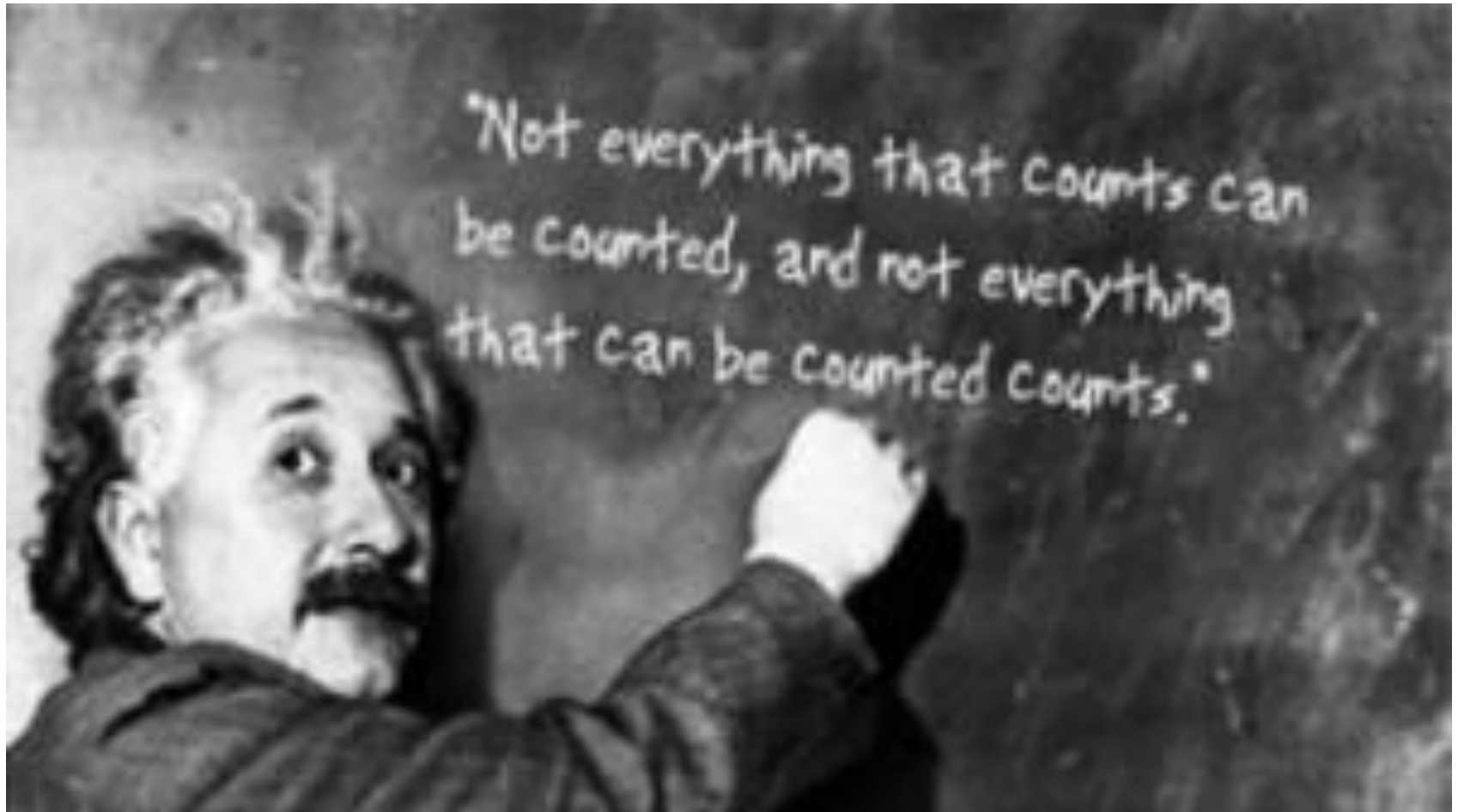
- Connaissance de l'administration de la recherche (locale, nationale) et de son pilotage (lien avec la gouvernance)
- Connaissance des modes de financement et d'évaluation de la recherche
- Compréhension fine des besoins des chercheurs selon leur communauté

Analyse, statistique et qualité

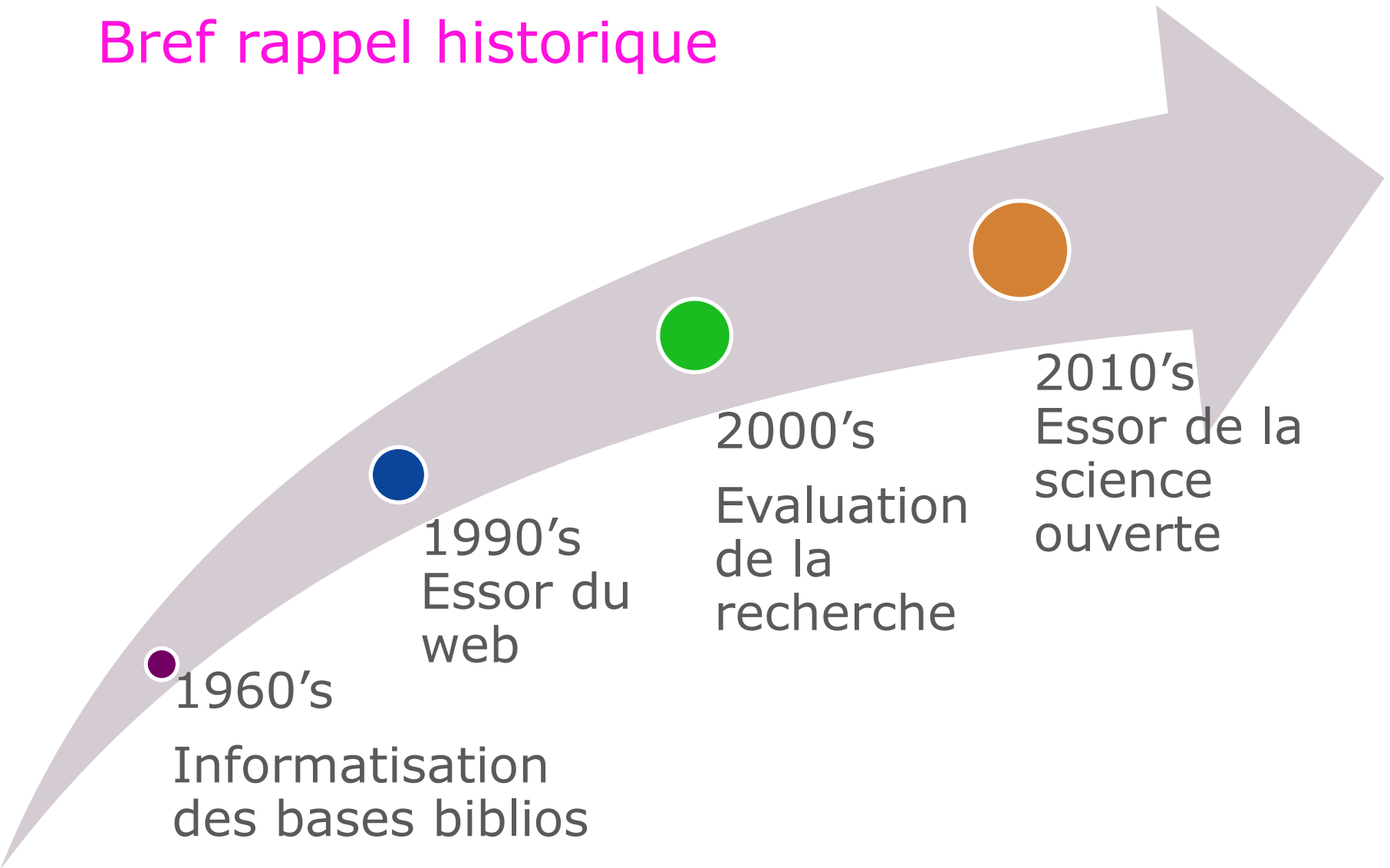
- Compétences en statistiques descriptives, en démarche qualité
- Gestion de données (tableur, base de données, interrogation API)
- Infographie (mise en forme et visualisation de données)

- **Connaître les bases citationnelles**

La bibliométrie au service de l'évaluation de la recherche



Bref rappel historique



Les bases citationnelles

Web of Science

Base d'articles & citations.
Couverture sélective, surtout sciences dures
Assez bonne normalisation des données.

Payant

Scopus[®]

Base d'articles & citations.
Couverture plus large que le Wos (sciences sociales).
Normalisation des données en amélioration.

Payant



Moteur de documents toutes disciplines (articles, thèses, livres) + citations. Aucune transparence des sources, ni métadonnées

Gratuit

- **Interpréter les principaux indicateurs**

Les principaux indicateurs

- Le facteur d'impact (pour une revue)
- Le nb de publications & de citations (auteur, labo, université, thématique)
- Le h-index (pour un auteur)
- Les top 1%, 10% (pour les articles)

Le facteur d'impact d'une revue

- **Définition :**
- Nb moyen de fois où un article a été cité dans les 2 ans qui suivent sa parution.
 - Ex : un article du BMJ est cité en moyenne 20 fois en 2 ans
- L'IF mesure l'audience d'une revue, donc son prestige.
- L'IF est la propriété de Clarivate Analytics, et toutes les revues n'ont pas d'IF

Le facteur d'impact d'une revue

NEW ENGLAND JOURNAL OF MEDICINE

How is Journal Impact Factor Calculated?

$$\text{JIF} = \frac{\text{Citations in 2017 to items published in 2015 (32544) + 2016 (20560)}}{\text{Number of citable items in 2015 (342) + 2016 (328)}} = \frac{53,104}{670}$$

Journal Impact Factor Calculation

$$\text{2017 Journal Impact Factor} = \frac{53,104}{670} = 79.260$$


Source : extrait du JCR Science Edition 2017

- **Accès :**
- Il faut être abonné au Journal Citation Reports (JCR)

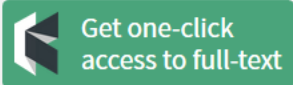
Le facteur d'impact d'une revue

- <https://webofknowledge.com/>
- Web of Science > Journal Citation Reports

Web of Science InCites **Journal Citation Reports** Essential Science Indicators EndNote Publons Kopernio Sign In ▼ Help ▼ English ▼

Web of Science 

Tools ▼ Searches and alerts ▼ Search History Marked List

Select a database Web of Science Core Collection ▼ 

Basic Search Cited Reference Search Advanced Search + More

Example: oil spill* mediterranean × Topic ▼ Search Search tips

+ Add row

Le facteur d'impact d'une revue

• Commentaires

- Pas de comparaison possible entre spécialités
- Pas pertinent pour évaluer les citations d'un article (moyenne pour la revue)
- Cercle vicieux de renforcement
- Influence de paramètres matériels (langue, nb pages, nb articles/fascicule)
- Pas d'erreur de calcul possible (garanti par Clarivate)



Le facteur d'impact est de plus en plus critiqué, et les institutions s'engagent à ne plus l'utiliser ([pétition DORA](#))

Quelles alternatives au facteur d'impact ?

- Le rang de la revue au sein sa discipline
 - quartiles des revues (Q1 journals)
 - ou percentiles de revues (JIF)

| | Full Journal Title | JCR Abbreviated Title | ISSN | Journal Impact Factor ▼ | Average JIF Percentile |
|---|--------------------------------------|-----------------------|-----------|-------------------------|------------------------|
| 1 | Annual Review of Applied Linguistics | ANNU REV APPL LINGUI | 0267-1905 | 4.880 | 99.724 |
| 2 | JOURNAL OF SECOND LANGUAGE WRITING | J SECOND LANG WRIT | 1060-3743 | 3.324 | 99.171 |
| 3 | APPLIED LINGUISTICS | APPL LINGUIST | 0142-6001 | 3.225 | 98.619 |
| 4 | BRAIN AND LANGUAGE | BRAIN LANG | 0093-934X | 2.851 | 78.919 |
| 5 | JOURNAL OF MEMORY AND LANGUAGE | J MEM LANG | 0749-596X | 2.829 | 81.089 |
| | MODERN LANGUAGE | | | | |

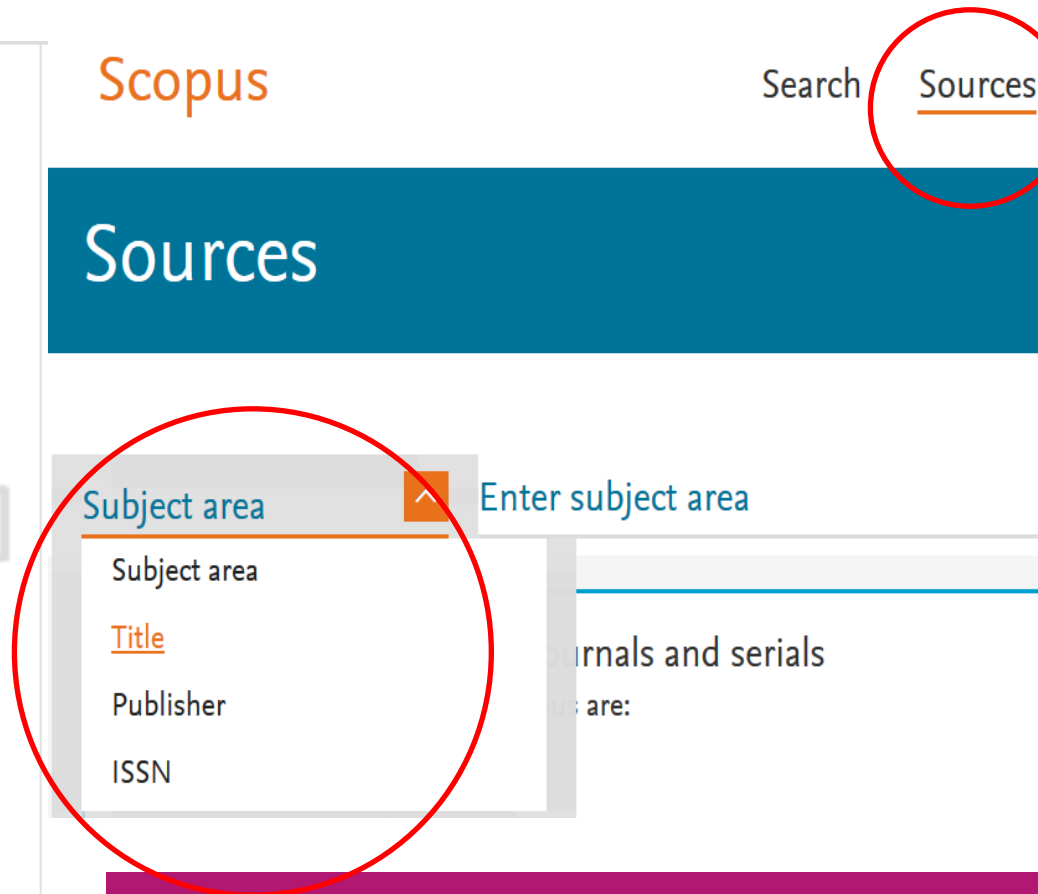
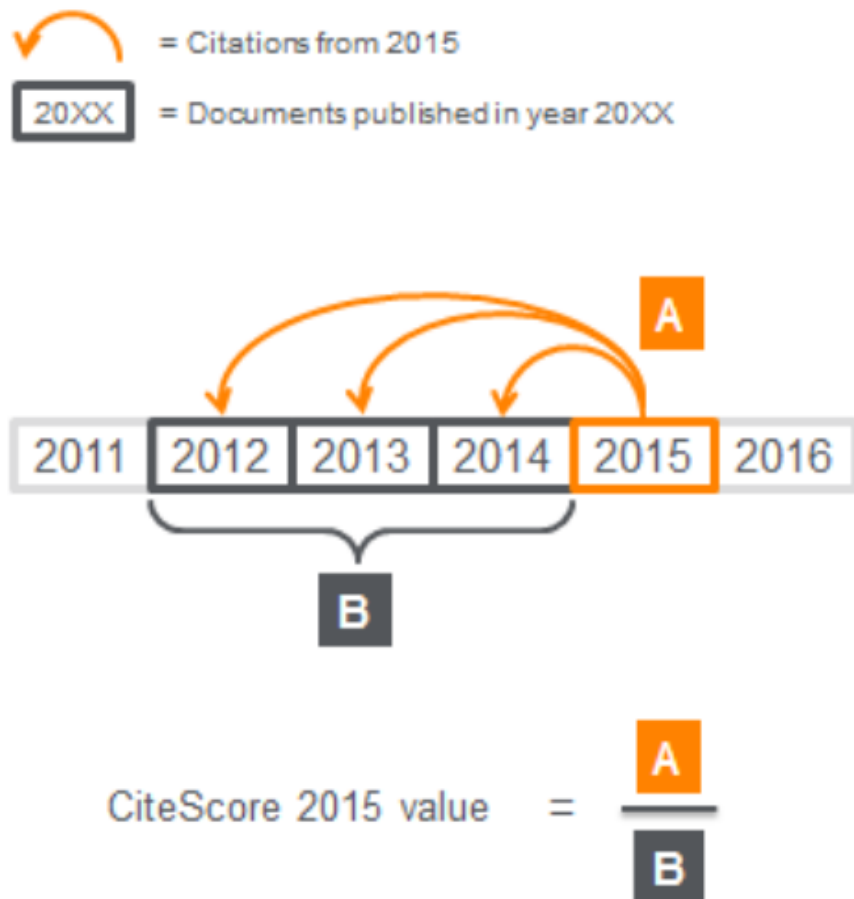
- Les citations au niveau des articles plutôt que des revues

CiteScore metrics for journals and serials

Alternative au Facteur d'impact

Les citations sont calculées sur 3 ans au lieu de 2

Accessible via Scopus > Sources



Quelles alternatives au facteur d'impact ?

- **Le Scimago Journal Ranking**
- Les citations sont pondérées selon la source
- Il est accessible gratuitement :
- <https://www.scimagojr.com/journalrank.php>
- Lancé en 2016, encore peu connu des chercheurs

The screenshot shows the Scimago Journal & Country Rank website. At the top, there's a header with 'SJR' and 'Scimago Journal & Country Rank'. Below this is a navigation bar with links: Home, Journal Rankings (selected), Country Rankings, Viz Tools, Help, and About Us. A search bar is on the right. Below the navigation bar are several filter dropdowns: 'All subject areas', 'All subject categories', 'All regions / countries', 'All types', and '2017'. There are also checkboxes for 'Only Open Access Journals', 'Only SciELO Journals', and 'Only WoS Journals'. A 'Display journals with at least' field is set to 0, and a 'Citable Docs. (3years)' dropdown is set to 0. An 'Apply' button is present. Below the filters is a 'Download data' button. At the bottom, a table displays journal rankings. The first row shows 'CA - A Cancer Journal for Clinicians' with an SJR of 61.786 (Q1), 137 H index, 43 Total Docs. (2017), 130 Total Docs. (3years), 3160 Total Refs., 16834 Total Cites (3years), 109 Citable Docs. (3years), 198.90 Cites / Doc. (2years), and 73.49 Ref. / Doc. The table also includes a US flag icon.

| Title | Type | ↓ SJR | H index | Total Docs. (2017) | Total Docs. (3years) | Total Refs. | Total Cites (3years) | Citable Docs. (3years) | Cites / Doc. (2years) | Ref. / Doc. |
|--|---------|--------------|---------|--------------------|----------------------|-------------|----------------------|------------------------|-----------------------|-------------|
| 1 CA - A Cancer Journal for Clinicians | journal | 61.786 Q1 | 137 | 43 | 130 | 3160 | 16834 | 109 | 198.90 | 73.49 |

Le nombre de publications & de citations

- Quelle source choisir ?
 - Critère de recherche : auteur, labo, université ? (qualité des données)
 - Pour les auteurs : quel niveau de contribution ?
 - Filtres : années, types de documents
 - Citations : exclure les auto-citations ?
-
- Ex : Devos, Patrick
 - Pubmed = 153 publis
 - Web of Science = 167 publis / 2670 citations
 - Scopus = 155 publis / 3262 citations

Le nombre de publications & de citations

- Comment interpréter les résultats?
- Quelle valeur de référence pour le nb de publications ?
- Quelle valeur de référence pour le nb de citations ?

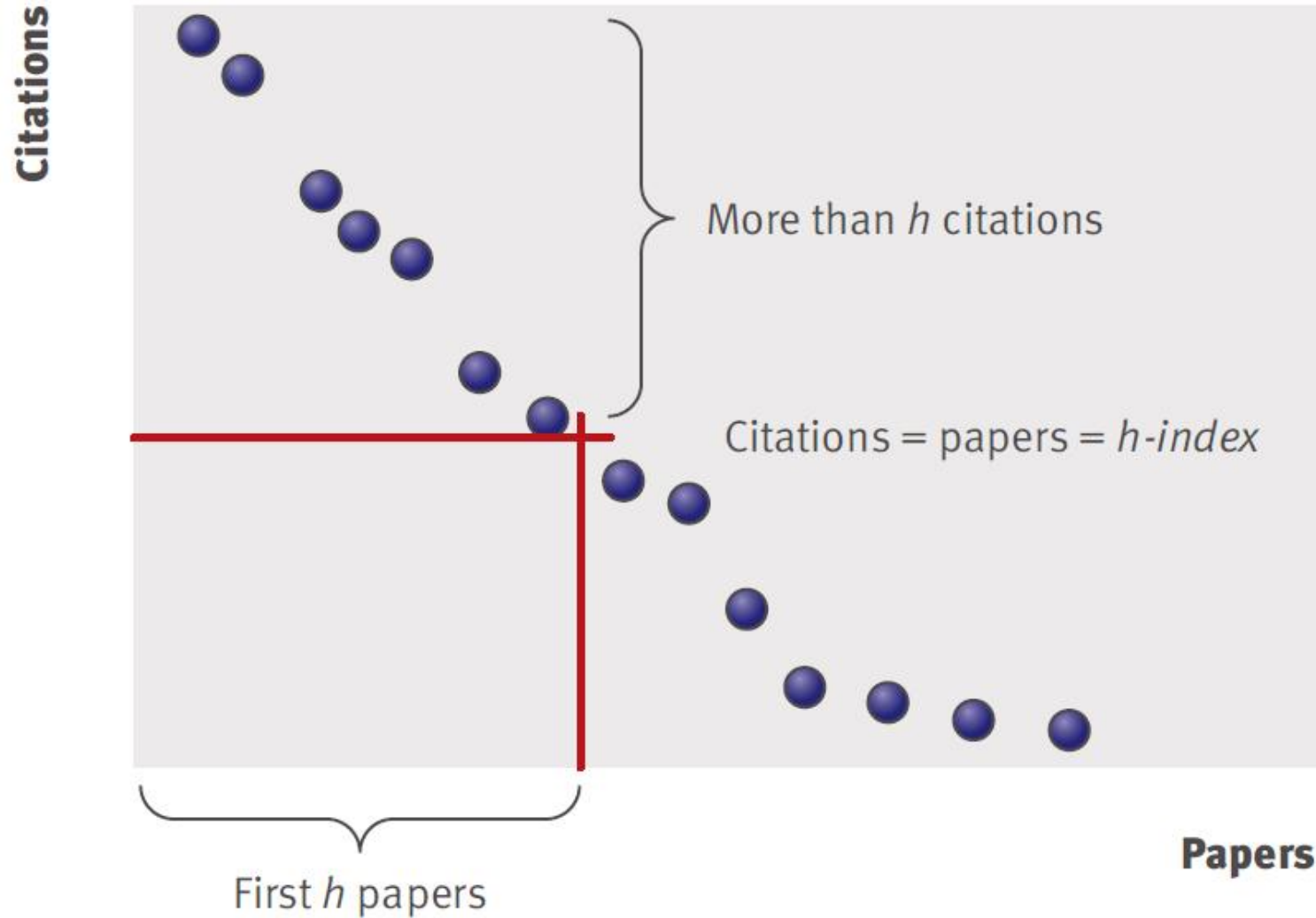
Le H-index d'un auteur

- Le H index est construit à partir de 2 indicateurs simples :
 - Nb de publications d'un auteur
 - Nb de citations
- Ex : Devos, Patrick
 - Pubmed = 153 publis
 - Web of Science = 167 publis / 2670 citations
 - Scopus = 155 publis / 3262 citations

Le H-index d'un auteur

- **Définition :**
- Indice composite de la production scientifique et de la notoriété d'un chercheur.
- Si $H = 20$, cela signifie que cet auteur a publié 20 articles qui ont été cités au moins 20 fois.
- Cette simplicité explique son succès
- Contrairement à l'IF, cet indicateur peut être librement calculé par n'importe qui.

Le H-index d'un auteur



Source : Horne R, Petrie KJ, Wessely S. H-index pathology: implications for medical researchers and practitioners. *BMJ*. 2009;339:b5356.

Le H-index dans le Web of Science

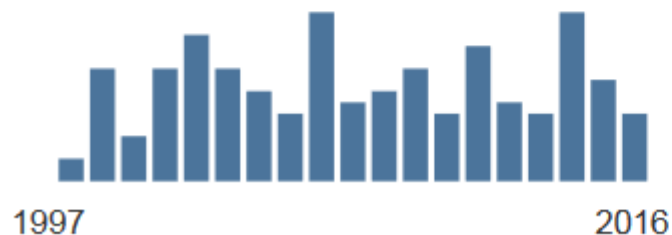
Citation report for 167 results from Web of Science Core Collection b

You searched for: **Article Group for: DEVOS P ...**[More](#)

This report reflects citations to source items indexed within Web of Science Core Collection. Perform a Cited Re

Total Publications

167



h-index

28

Average citations per item


15.99

Le H-index dans Scopus

Devos, Patrick


Sante publique : epidemiologie et qualite des soins, Lille,
France

Author ID: 21639126700

 <http://orcid.org/0000-0001-7803-9552>

Documents: 155

Citations: 3262 total citations by 3003 documents

h-index: 31 

Co-authors: 150 (maximum 150 co-authors can be displayed)

Subject area: Medicine , Biochemistry, Genetics and Molecular Biology

Le H-index dans Google Scholar

Facile si l'auteur a créé son profil :



The screenshot shows the Google Scholar interface. At the top, the Google Scholar logo is on the left, and a search bar contains the text "tessier Frédéric". Below the search bar, it says "Articles" and "Environ 6 500 résultats (0,04 s)". On the left side, there are filters: "Date indifférente" (with sub-options: Depuis 2018, Depuis 2017, Depuis 2014, Période spécifique...), "Trier par pertinence" (with sub-option: Trier par date), and "Toutes les langues" (with sub-option: Rechercher les pages en Français). The main content area shows a "Conseil : Recherchez des résultats uniquement en Français. Vous pouvez indiquer votre langue de recherche." Below this, it says "Profils utilisateurs correspondant à 'tessier Frédéric'". A profile for "Frederic J Tessier" is shown, with details: "Full Professor - Lille University", "Adresse e-mail validée de univ-lille2.fr", and "Cité 2007 fois". Below the profile, a list of publications is shown, starting with "[HTML] The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance ...", followed by the authors "E Stein, J Jeong, AP McMahon, M Tessier-Lavigne" and the journal "Cell, 2003 - Elsevier". The snippet continues: "Developing axons are guided to their targets by attractive and repulsive guidance cues. In the embryonic spinal cord, the floor plate chemoattractant Netrin-1 is required to guide commissural neuron axons to the midline. However, genetic evidence suggests that other ...". At the bottom of the snippet, there are icons for a star, a document, and a link, followed by the text "Cité 615 fois", "Autres articles", and "Les 14 versions".

Sinon, on peut utiliser l'outil [Publish or Perish](#), qui permet de récupérer les informations de Google Scholar sous forme de tableau.

Le H-index d'un auteur

- **Commentaires :**

- Pas de base de calcul unique
- Problèmes de référencement des auteurs
- Le rang de l'auteur ne compte pas
- Les disparités entre disciplines ne sont pas prises en comptes
- Pas de comparaison possible des chercheurs d'âge ou de spécialités différentes

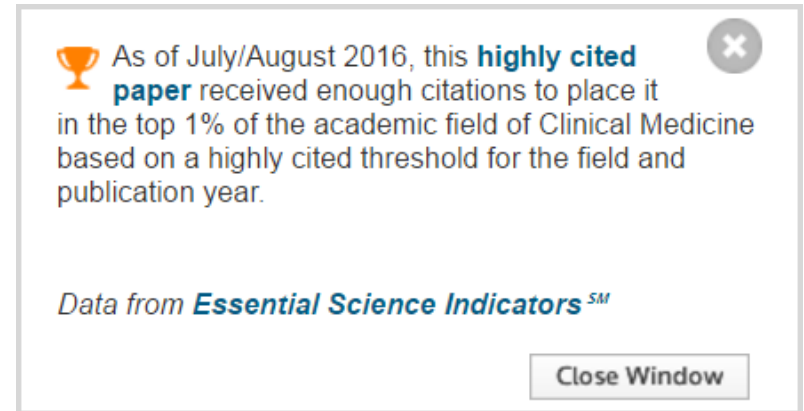
Les percentiles de citations (Top 1%, 10%...)

- **Définition :**
- Indicateur qui se réfère à la distribution des citations, et non pas à la moyenne : Classer tous les documents du plus cité au moins cité, puis déterminer ceux qui font partie du top 1%, 10%, 20%... les plus cités.
- Indicateur normalisé par type de document, discipline et année.
- Très utilisé par l'OST et l'INSERM

Les percentiles de citations (Top 1%, 10%...)

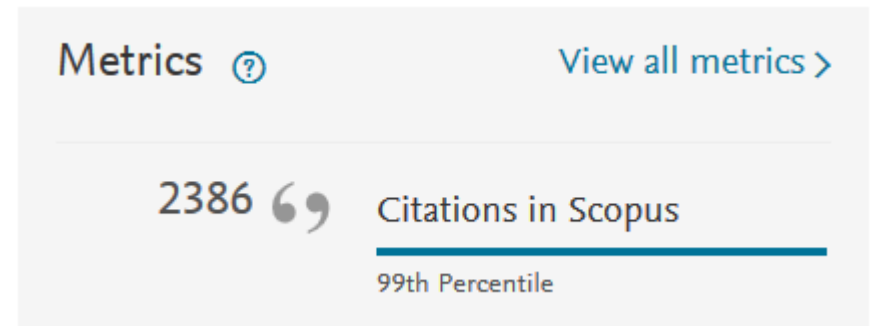
Dans le Web of Science :

- Highly cited : top 1%



Dans Scopus :

- 99th percentile



Top 1% a la même signification statistique que 99^e percentile

Les percentiles de citations (Top 1%, 10%...)

Commentaire :

- Indicateur au niveau de l'article
- Normalisé par année, type de document et discipline
- Facilité d'interprétation
- Respecte la distribution asymétrique des citations
- Mais variable selon la source de données
- le découpage disciplinaire est +/- fin
- Pas de pondération du rôle d'auteur
- Pas pertinent au niveau individuel

Synthèse sur les indicateurs

| Impact à mesurer | | | | |
|---|---|---------------------------------|------------------------------------|-----------------------------------|
| Mesurer l'impact d'une revue | Facteur d'impact | CiteScore, Scimago Journal Rank | Quartiles des revues Q1 Journals | |
| Mesurer l'impact d'un article | Nombre de citations / valeur de référence | Percentile de citation (Top 1%) | | |
| Mesurer l'impact d'un auteur/labo /université | Part de la production / France / Europe / Monde | H-index | % d'articles dans le top 1% ou 10% | % d'articles dans les Q1 journals |

Etude de cas : analyser le profil de publications

Un futur doctorant qui travaille sur le diabète vient vous voir. Il hésite entre 3 directeurs de thèse possibles : Bart Staels, François Pattou et Philippe Froguel. Pouvez-vous lui préciser le profil bibliométrique de chacun pour l'aider à faire son choix ?

- **Quelles bases de données utiliser ?**
- **Explorer les fonctions « analyse des résultats »**
- **Comparer les profils des 3 chercheurs pour faire apparaître les éléments communs ou spécifiques**

- **Proposer des services dans son établissement**

Quels services proposer dans votre établissement ?

- Adaptée à l'établissement (et aux utilisateurs)
- Adaptée aux moyens disponibles
- Discuter ensemble des services possibles
- Puis restitution collective

Quels services proposer dans votre établissement ?

Aider les chercheurs

Quels services proposer dans votre établissement ?

Améliorer la qualité des données

Piloter la documentation et l'open access

Pistes de lecture

1. Larivière V, Sugimoto CR (2018). *Mesurer la science*. Montréal, Canada: Les Presses de l'Université de Montréal.
<https://pum.umontreal.ca/catalogue/mesurer-la-science>
2. Gingras, Y. (2014). *Les dérives de l'évaluation de la recherche : Du bon usage de la bibliométrie*. Paris: Liber.
3. Gingras, Y. (2008). La fièvre de l'évaluation de la recherche : du mauvais usage de faux indicateurs. CIRST, UQAM. http://cpham.perso.univ-pau.fr/DOC/Gingras_Bibliometrie.pdf
4. Vidaillet, B. (2013). *Évaluez-moi ! : Évaluation au travail : les ressorts d'une fascination*. Paris: Seuil.
5. Dassa, M., Deniau, A., Durand-Barthez, M., Girard, F., Pothier, N., & Sene, A. (2014). Le millefeuille des affiliations françaises dans les publications scientifiques. *Documentaliste-Sciences de l'Information*, 51(4), 12-16.
http://archivesic.ccsd.cnrs.fr/sic_01097580
6. Académie des Sciences. (2011). *Du bon usage de la bibliométrie pour l'évaluation individuelle des chercheurs*. <http://www.academie-sciences.fr/activite/rapport/avis170111.pdf>
7. Observatoire des Sciences et des Techniques (OST). (2015). Le manifeste de Leiden pour la mesure de la recherche. <http://www.obs-ost.fr/manifesto> .



lillometrics@isite-ulne.fr



Solenn Bihan
Laurence Crohem
David Bélart

Patrick Devos
Hélène Lefranc
Sylvain Vandamme

