

INFORMATION NUMÉRIQUE

Enjeux et Pratiques

Licence 3 / UE602 – Accès expert à l'information – 2025-2026

Initiation à la bibliométrie

valerie.bonvallot@inist.fr

Sommaire : initiation à la bibliométrie

26 mars 2025

- 1. Exemples, définition et usages**
- 2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte**
- 3. Indicateurs bibliométriques**
- 4. Méthodologie de réalisation d'études bibliométriques**

1. Exemples, définition et usages

1. Exemples, définition et usages

Avez-vous déjà entendu parler de **bibliométrie** ?

A quoi vous fait penser ce terme ?

➡ **Biblio-métrie** = livres (« papier ») + mesure

1. Exemples, définition et usages

Exemple avec OpenAlex : données sur le réchauffement climatique chez les jeunes
Sur les 10 dernières années

("réchauffement climatique" OR "global warming") AND (jeune* OR adolescen* OR young* OR youth* OR student* OR étudiant*)

- ▶ Utiliser les facettes (filtres) ou le requêtage pour répondre aux questions :
 - ❖ que pouvez-vous dire de l'évolution du nombre de publications ?
 - année avec le plus grand nombre de publications
 - année avec le moins de publications : pourquoi ?
 - ...
 - ❖ quelles sont les sources dans lesquelles les auteurs publient le plus ?
 - ❖ quel est le principal topic d'OpenAlex représenté par les articles ?
 - ❖ combien d'auteurs ont écrit strictement plus de 10 articles ?
 - ❖ quelle place occupe la France ? Avec combien de publications ? La première publication a été citée combien de fois ?
- ▶ Quelles autres questions pourrait-on poser ? Que pouvez-vous dire de plus sur ce corpus ?

1. Exemples, définition et usages

Déf. : Bibliométrie :

- Sous-discipline des sciences de l'information
- "La bibliométrie est l'application des mathématiques et des méthodes statistiques aux livres, articles et autres moyens de communication" (A. Pritchard, « Statistical Bibliography or Bibliometrics ? » , in *Journal of Documentation*, n°25(4), 1969, p.348-349)

Déf. : Production scientifique (rappel)

Ensemble des publications scientifiques produites par des chercheur.se.s d'un laboratoire ou d'une structure de recherche, permettant la diffusion des résultats de recherche. Ces publications peuvent être : des articles, des ouvrages ou des chapitres d'ouvrages, des actes de congrès ou des communications, des thèses, des rapports, des fiches techniques, des cartes, etc.

- ➡ Analyse **statistique et quantitative** sous forme d'étude bibliométrique
- ➡ Sur la **production scientifique ou littérature scientifique** (liste des publications ou articles sous forme de notices / références bibliographiques) = grande quantité de données

1. Exemples, définition et usages

A quoi sert la bibliométrie (usages) ?

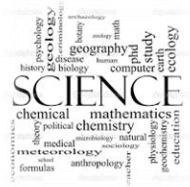
Qui ?

Auteur(s)
Organisme(s)
Pays
Continent(s)



Fait quoi ?

Mots-clés
Classification
Texte intégral
Résumé



Où ?

Organisme(s)
Pays
Continent(s)



Quand ?

Date de publication



Avec qui ?

Auteur(s)
Organisme(s)
Pays
Continent(s)



Comment ?

Type(s) de document
Langue(s)
Revue(s)
Editeur(s)
Financement
OA (open access)



1. Exemples, définition et usages

A quoi sert la bibliométrie (usages) ?

Une étude bibliométrique peut répondre à de multiples objectifs, qui permettent de :

- rechercher et recenser de l'information (ex : sur un domaine)
- caractériser la science, mieux connaître les sujets actuels, et les grandes tendances
- identifier les acteurs de la recherche (ex : pour établir de nouveaux partenariats)
- répondre à l'évaluation de la recherche scientifique au niveau d'un laboratoire, d'un institut de recherche, ou d'une structure de recherche
- répondre à un appel à projets spécifique
- préparer un rapport annuel, une rétrospective ...

Une étude bibliométrique peut devenir un réel

→ outil de pilotage et d'aide à la décision

pour les laboratoires, les instituts, les structures de recherche, les chercheurs et chercheuses.

1. Définition, usages, exemples

A quoi sert la bibliométrie (usages) ?

Quelques problématiques rencontrées, quelques questions auxquelles répond une étude bibliométrique :

- Quelle est la production scientifique de mon laboratoire ces 5 dernières années ?
- Comment est représenté le CNRS sur la thématique de l'énergie ?
- Quelle part des publications scientifiques de mon université est en accès ouvert ?
- Quelles sont les collaborations à l'international pour mon institut de recherche ?
- Quels sont les experts européens reconnus dans le domaine de l'énergie ?
- Quel est l'impact des publications publiées par les chercheurs français ?



Ces exemples vont permettre de définir les indicateurs bibliométriques à réaliser dans l'étude.

Déf. : Indicateur bibliométrique :

Un indicateur bibliométrique donne une appréciation quantitative des articles scientifiques, des revues, ou des chercheurs. Ils ne peuvent jamais qu'être complémentaires à l'évaluation qualitative (qualité du contenu des articles, revues...).

1. Définitions, usages, exemples

Quelques exemples d'études bibliométriques réalisées (résultat final) :

- Etude sur la production scientifique de l'IN2P3 (CNRS Nucléaire et Particules, Institut national de physique nucléaire et de physique des hautes énergies)
- Etude sur la bibliographie de volumes du 6^e rapport du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat)
- Plusieurs exemples sur le site [Istex](#)

2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

Institutionnalisation française et internationale de la bibliométrie

- Evaluation des établissements de recherche et d'enseignement français par le **HCERES** (Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur)
- Création d'organismes et production de rapports bibliométriques français et internationaux :
 - **France : OST** (Observatoire des sciences et techniques)
<https://www.hceres.fr/fr/observatoire-sciences-techniques>
<https://www.hceres.fr/fr/publications-ost>
 - **France : Baromètre français de la Science Ouverte**
<https://barometredelascienceouverte.esr.gouv.fr/>
 - **Pays Bas : CWTS** (Centre for Science and Technology Studies)
<https://www.cwts.nl/>
 - **Europe : OCDE** (Organisation de coopération et de développement économiques) :
http://www.oecd-ilibrary.org/science-and-technology/main-science-and-technology-indicators_2304277x
 - **Etats-Unis : NSF** (National Science Foundation)
<https://nces.nsf.gov/indicators>

2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

Institutionnalisation française et internationale de la bibliométrie

- **Classements internationaux des universités**
 - Shanghai (ARWU-Academic Ranking of World Universities) : <http://www.shanghairanking.com>
 - CWTs Leiden : <http://www.leidenranking.com/>
<https://open.leidenranking.com/>
 - **Pour la prochaine séance : à quelle place se positionne l'université de Lorraine ? Vos impressions ?**
- **Les services dans les bibliothèques universitaires**
 - UL : <https://www.univ-lorraine.fr/recherche-et-innovation/bibliometrie/>
<https://scienceouverte.univ-lorraine.fr/bibliometrie/barometre-lorrain-de-la-science-ouverte/>
- **Les formations :**
 - URFIST (Unité Régionale de Formation à l'Information Scientifique et Technique) ou Universités
 - European summer school for scientometrics <https://esss.info/programme/>

2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

Prise de conscience collective des limites de l'évaluation quantitative de la recherche et de la mainmise des éditeurs scientifiques

Quelques liens :

- *Les dérives de l'évaluation de la recherche. Du bon usage de la bibliométrie*, Yves Gingras, 2014, Editions Raisons d'agir
- *Du bon usage de la bibliométrie pour l'évaluation individuelle du chercheur* (Avis Académie des sciences 01/2011) <http://www.academie-sciences.fr/pdf/rapport/avis170111.pdf>
- Blog intéressant sur la bibliométrie, ses usages, ses dérives : <https://openscience.pasteur.fr/tag/bibliometrie/>
- BD pédagogique sur la bibliométrie et l'évaluation de la recherche : <https://zenodo.org/record/3820169#.YDNuG3njJaT>

2. De la diffusion de la bibliométrie vers le mouvement de la Science Ouverte

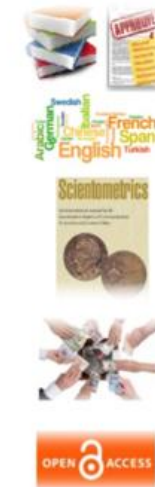
Le mouvement de la Science ouverte instaure la création de nouveaux indicateurs

- Budapest Open Initiative - 2001 : <https://www.budapestopenaccessinitiative.org/> (problématique du libre accès aux publications scientifiques et incitation à l'utilisation des archives ouvertes ou des revues en libre accès)
- La déclaration de San Francisco - 2012 : <https://anr.fr/fr/lanr/engagements/la-declaration-de-san-francisco/> (critique de l'indicateur « facteur d'impact »)
- Le manifeste de Leiden - 2015 : <http://www.leidenmanifesto.org/> (critique de l'indicateur « h-index »)
- Position du CNRS : feuille de route du CNRS pour la science ouverte (novembre 2019)
https://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaquette_ScienceOuverte.pdf
 - ➡ L'évaluation des publications (articles) doit désormais porter sur les résultats scientifiques et non leur environnement de publication. Les chercheur.ses sont amenés à choisir leurs meilleures productions et en expliquer l'impact. Tous les types de production peuvent servir à l'évaluation, et toutes les productions citées doivent être accessibles en libre accès.
- Leiden Ranking Open Edition 2023 (janvier 2024) : création d'un nouveau classement international des universités à partir d'OpenAlex (base de données en libre accès) <https://open.leidenranking.com/>
 - ➡ Commencement d'une nouvelle ère pour l'évaluation de la recherche.

3. Indicateurs bibliométriques

Comment ?

Type(s) de document
Langue(s)
Revue(s)
Editeur(s)
Financement
OA (open access)



Stratégie d'édition

3. Indicateurs bibliométriques

3.1. Indicateurs sur la production scientifique (Qui ?)

Calcul du nombre de publications au niveau de :

- Chercheur.se → repérage d'experts
- Organisme de recherche (laboratoire, institut, université)
- Pays
- Continent

⇒ **Interrogation des bases de données** : champs *Auteurs*, *Affiliations* (=adresse d'un organisme de recherche, signature de l'auteur), *Pays...* : penser à toutes les formes d'écriture + historique et évolution du nom + identifiants (ex. nom de jeune fille, un laboratoire, université de Lorraine, France et outremer, liste de pays pour un continent ...)

⇒ **Traitement des données** : il peut exister plusieurs formes d'écriture pour un même auteur, une même affiliation. Il est nécessaire d'homogénéiser les données : une forme préférentielle, un identifiant.

3. Indicateurs bibliométriques

3.1. Indicateurs sur la production scientifique (Qui ?)

Zoom sur les identifiants auteurs-organismes : code alphanumérique unique pérenne (homonymie, changement de nom ...)

International

- researcherId (WoS)
- ORCID (Open Researcher and Contributor ID)
- id Ror (Research Organization Registry)

France

- idRef auteur (ABES, liens entre identifiants auteurs : orcid, idRef, idHal)
- idHal (HAL)
- RNSR (Répertoire National des Structures de Recherche) → RNeST

3. Indicateurs bibliométriques

3.2. Indicateurs de contenu thématique (A propos de quoi ?)



Calcul du nombre de publications au niveau du contenu thématique :

- Indexation : Mots-clés auteurs / Descripteurs - Index. manuelle / automatique
- Classifications : au niveau de l'article ou du périodique ?
- A partir du résumé ou du texte intégral (full text) ?
- ➔ Interdisciplinarité : cooccurrence de termes

3.3. Indicateurs sur la répartition géographique (Où ?)

Calcul de la répartition géographique des publications à partir de l'affiliation de la publication (le champ « Adresses »).

- Organisme de recherche (laboratoire, institut, université)
- Niveau local, régional, national
- Pays
- Continent

Même logique d'interrogation et de traitement que la question « QUI ? »

3. Indicateurs bibliométriques

3.4. Indicateurs sur la périodicité (QUAND ?)

Répartition de l'évolution temporelle des publications sur une périodicité donnée (sur plusieurs années par exemple). Utile pour montrer les grandes tendances d'évolution.

⇒ Traitement possible des données au préalable : bien homogénéiser les dates de publication (ex: « 2 023 », « 2023 » => écritures différentes pour une même année).

!! Changement de politique de la base : modification de la couverture, de la structure

3.5. Indicateurs sur les coopérations et relations scientifiques (AVEC QUI ?)

Calcul des coopérations des auteurs, des organismes et des régions, des pays. Mise en évidence des co-publications, des réseaux

Même logique d'interrogation et de traitement que la question « QUI ? »

Enrichir ces données : tag sur les organismes (privés/publics), tag sur les localisations (régions)

3. Indicateurs bibliométriques

3.6. Indicateurs sur les stratégies d'édition (COMMENT ?)

Répartition par stratégie d'édition :

- Type de document (article, ouvrage, chapitre d'ouvrage, acte de conférence...)
- Éditeur (Elsevier, Springer...)
- Source (revue, ouvrage, conférence...)
- Langue de publication (anglais, français, allemand...)
- Sources de financement (projet ANR, subvention européenne...)

3.7 Indicateurs sur la Science ouverte (l'accès ouvert - AO ou Open access - OA)

Répartition des publications en accès ouvert (oui, non...), par type d'accès ouvert (bronze, gold, green, hybrid, closed...)

- ⇒ Travail d'enrichissement des données à partir d'*Unpaywall* (base de données qui recense les publications en libre accès sur le web à partir de leur numéro DOI, Digital Object Identifier).
- ⇒ Ces indicateurs sont de plus en plus demandés et deviennent obligatoires dans les études du CNRS.



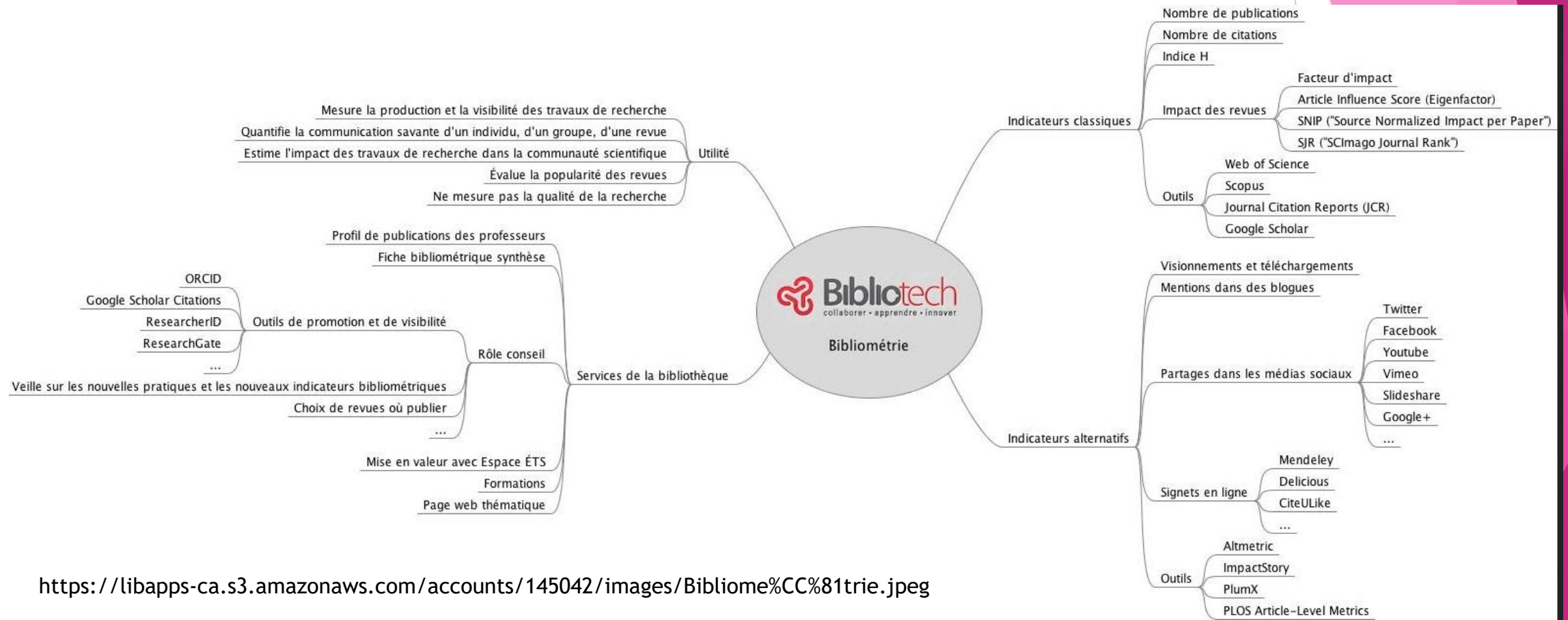
3. Indicateurs bibliométriques

3.8. Indicateurs de notoriété ou de visibilité = impact / citations

- **Le facteur d'impact (FI ou impact factor, IF)** : indicateur créé par Clarivate Analytics (WoS), qui estime directement la visibilité d'une revue scientifique, calculé au niveau d'une revue pour une année donnée (le FI d'une revue est égal à la moyenne des nombres de citations des articles de cette revue publiés durant les 2 années précédentes).
 - Une revue avec un facteur d'impact élevé serait considérée comme plus importante qu'une revue avec un FI faible.
- **Le « h index »** : un.e chercheur.se obtient un indice de h quand il a publié h articles, et que chacun a été cité au moins h fois.
 - L'indice ne sera jamais plus élevé que le nombre de publications du chercheur.se (pénalisation des jeunes chercheur.ses). Il assimile productivité et impact.

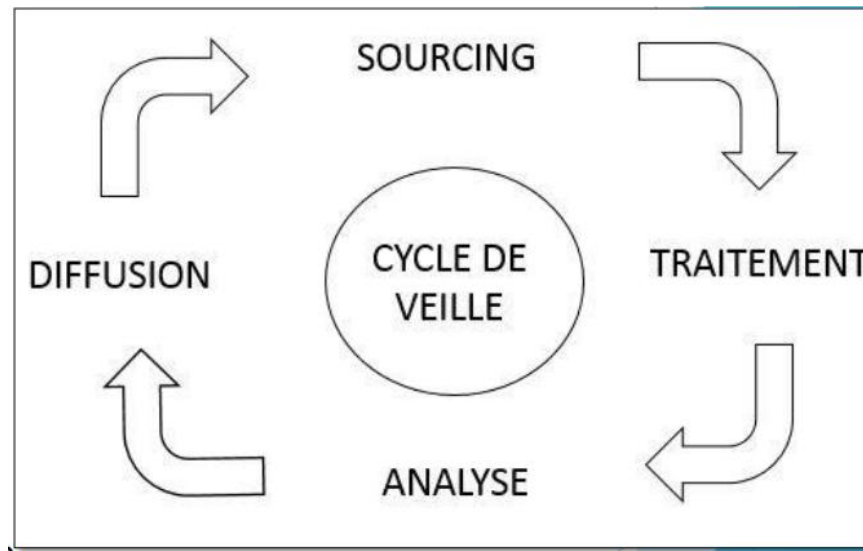


Ces indicateurs de notoriété ou visibilité sont de moins en moins utilisés car décriés (autocitations, qualité ?, différences selon les domaines ...) ²³
 Internet → statistiques d'usage et métrique alternative (altmetrics)



<https://libapps-ca.s3.amazonaws.com/accounts/145042/images/Bibliome%CC%81trie.jpeg>

4. Méthodologie de réalisation d'une étude bibliométrique



4. Méthodologie de réalisation d'une étude bibliométrique

Méthodologie de réalisation d'une étude bibliométrique ?

➡ 2 grands types d'études les plus couramment réalisés :

- Etude sur la production scientifique d'une entité de recherche (ex : laboratoire, département universitaire, université, institut, organisme de recherche...)
- Etude sur la production scientifique sur une thématique (ex : énergie, blob...)

4. Méthodologie de réalisation d'une étude bibliométrique

4.1. Identification des besoins du commanditaire

- **Périmètre et objet de l'étude** : étude thématique (ex : énergie, blob...) ou étude sur la production scientifique d'un acteur de la recherche (ex : laboratoire, structure de recherche, université...)
- **Période souhaitée** : année ou fourchette d'années
- **Indicateurs souhaités**
- **Bases de données bibliographiques à interroger** : pluridisciplinaire, thématique, propriétaire, en open access... (cf: voir cours sur les bases de données)
- **Définition du choix des mots-clés** (// équation de recherche pour interroger les bases de données bibliographiques)

4. Méthodologie de réalisation d'une étude bibliométrique

4.2. Constitution du corpus de notices bibliographiques

- Interrogation de(s) base(s) de données bibliographiques (ou sources d'information)
- Collecte des données plus ou moins « structurées »
- Traitement des données pour préparer un corpus de notices bibliographiques (fichier de données)
- Envoi du corpus au commanditaire pour validation



4. Méthodologie de réalisation d'une étude bibliométrique

4.3. Traitement des données

- Opérations de dédoublonnage
- Opérations d'homogénéisation (ex : nom des revues différents, auteurs écrits sous plusieurs formes d'écriture...)
- Opérations d'enrichissement (ex : ajout des informations liées à l'Open access, indexation et classifications automatiques → Istex TDM - Text & Data mining : <https://services.istex.fr/>)

➡ Objectif : rendre les données les plus propres possibles pour garantir leur qualité et améliorer leur utilisation dans l'analyse bibliométrique ou statistique.

Etape la plus longue et technique. Traitement des données à partir d'outils de curation de données comme OpenRefine, ou avec des traitements automatisés sous forme de scripts (sous Python par exemple).



Déf. Dédoublonnage : action de suppression des doublons ou données redondantes d'une liste, d'une base de données ou d'un fichier

Déf. Homogénéisation : action de nettoyage des données afin d'obtenir une valeur unique et cohérente pour des entrées de même type (ex : adresses, auteurs, mots-clés...)

Déf. Enrichissement : action d'ajouter des informations complémentaires à des données existantes à partir d'autres fichiers (ex : ajout du taux d'open access, de catégories thématiques spécifiques...)

4. Méthodologie de réalisation d'une étude bibliométrique (5)

4.4. Valorisation et visualisation des données sous forme d'indicateurs bibliométriques

- Utilisation d'outils de visualisation de données structurées comme VOSviewer ou Lodex (développé par le CNRS-INIST)
- Envoi du site web au commanditaire. L'étude peut être en mode public ou confidentiel (avec usage d'un identifiant / mot de passe).

➔ Création d'un outil web clé en main de type tableau de bord, qui permet l'exploration du corpus, et qui met en valeur les données sous forme d'indicateurs, à l'aide de graphiques.



5. Conclusions

Indicateurs **quantitatifs** pour la recherche scientifique

qui fait quoi où quand avec qui comment ?

**Productivité
scientifique**

**Contenu
thématique**

**Evolution
Tendances**

**Coopération
Relation scientifique**

Stratégie d'édition

Reconnaissance

Visibilité scientifique (citations)

Connaissance des sources

Données structurées, homogènes, enrichies

**Résultats accompagnés d'une méthodologie :
sources, modes de calcul, outils utilisés**

**Ne pas être utilisés seuls + environnement de la science ouverte,
multiplication des documents**

Pour aller plus loin

Calculs et représentations graphiques

Pour aller plus loin : le mode de comptage

Nombre de publications

- ✓ comptage d'occurrence
un acteur / une publication
→ Ne pas additionner
- ✓ comptage fractionnaire
un acteur / une part de publication
→ On peut additionner

Pour aller plus loin : le mode de comptage

Exemple : une notice bibliographique

Titre : Links between particle surface hardening and rehydration impairment ...

Source : FOOD HYDROCOLLOIDS 2016, Vol 61, P 277-285

5 Auteurs : Burgain, J (Burgain, Jennifer)^[1,2] ; Scher, J (Scher, Joel)^[2] ; Petit, J (Petit, Jeremy)^[2] ; Francius, G (Francius, Gregory)^[3] ; Gaiani, C (Gaiani, Claire)^[2,4]

4 Adresses :

- ▶ [1] **CNIEL**, 42 Rue Chateaudun, F-75314 Paris 9, **France**
- ▶ [2] **Univ Lorraine**, LIBio Lab Ingn Biomol, 2 Ave Foret de Haye, TSA 40602, F-54518 Vandoeuvre Les Nancy, **France**
- ▶ [3] **Univ Lorraine**, LCPME, UMR 7564, F-54600 Villers Les Nancy, **France**
- ▶ [4] **Univ Queensland**, Sch Agr & Food Sci, St Lucia, Qld 4072, **Australia**

Pour aller plus loin : le mode de comptage

5 coauteurs

	Occurrence	Fractionnaire
Burgain, Jennifer	1	0,2
Scher, Joel	1	0,2
Petit, Jeremy	1	0,2
Francius, Gregory	1	0,2
Gaiani, Claire	1	0,2
Total	5	1

Pour aller plus loin : le mode de comptage

3 institutions - 2 pays

		Occurrence	Fractionnaire
Organismes	Univ Lorraine	1	0,33
	Univ Queensland	1	0,33
	CNIEL	1	0,33
	Total	3	1
Pays	France	1	0,5
	Australie	1	0,5
	Total	2	1

Pour aller plus loin : le mode de comptage

Part de publications / Pourcentage pour rapporter,
pour avoir un élément de comparaison

Type de document	Nb de publications	Part
ARTICLES	94 223	75,8%
REVIEW ARTICLES	9 103	7,3%
MEETING ABSTRACTS	6 670	5,4%
EARLY ACCES	6 131	4,9%
Autres	15 692	0,17%
Total	131 819 (124 316)	105,5% (100)

Pays coopérant	Nb de copublications avec	Part
USA	21 730	32,7%
GERMANY	14 975	22,58%
ENGLAND	14 729	22,21%
ITALY	13 247	19,97%
...		
Total	(66 316)	(100)

Une publication peut
avoir plusieurs types
de documents

cu=France and py=2021 (WoS)
124 316 publications

Une publication peut
avoir plusieurs pays
coopérants

Pour aller plus loin : les représentations graphiques

Tableaux

Radars

Bâtons

Courbes

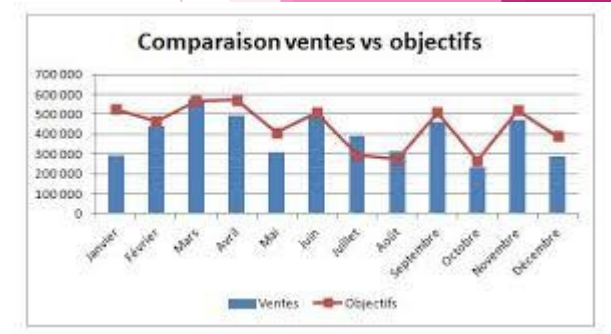
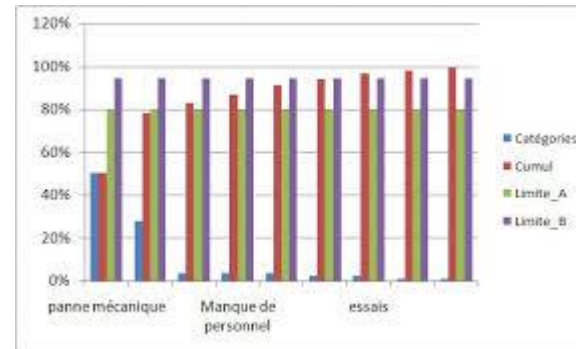
Bulles

Cartographie

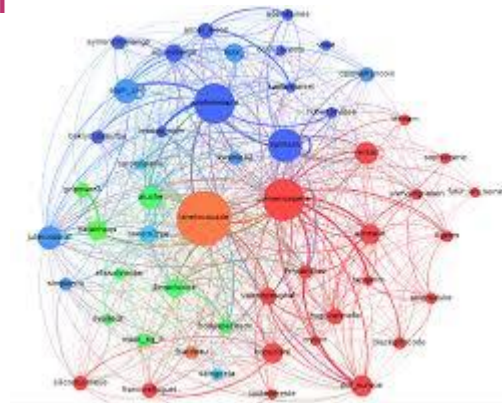
Réseaux

....

Excel



Gephi



Pour aller plus loin : les représentations graphiques

Tableaux

Radars

Bâtons

Courbes

Bulles

Cartographie

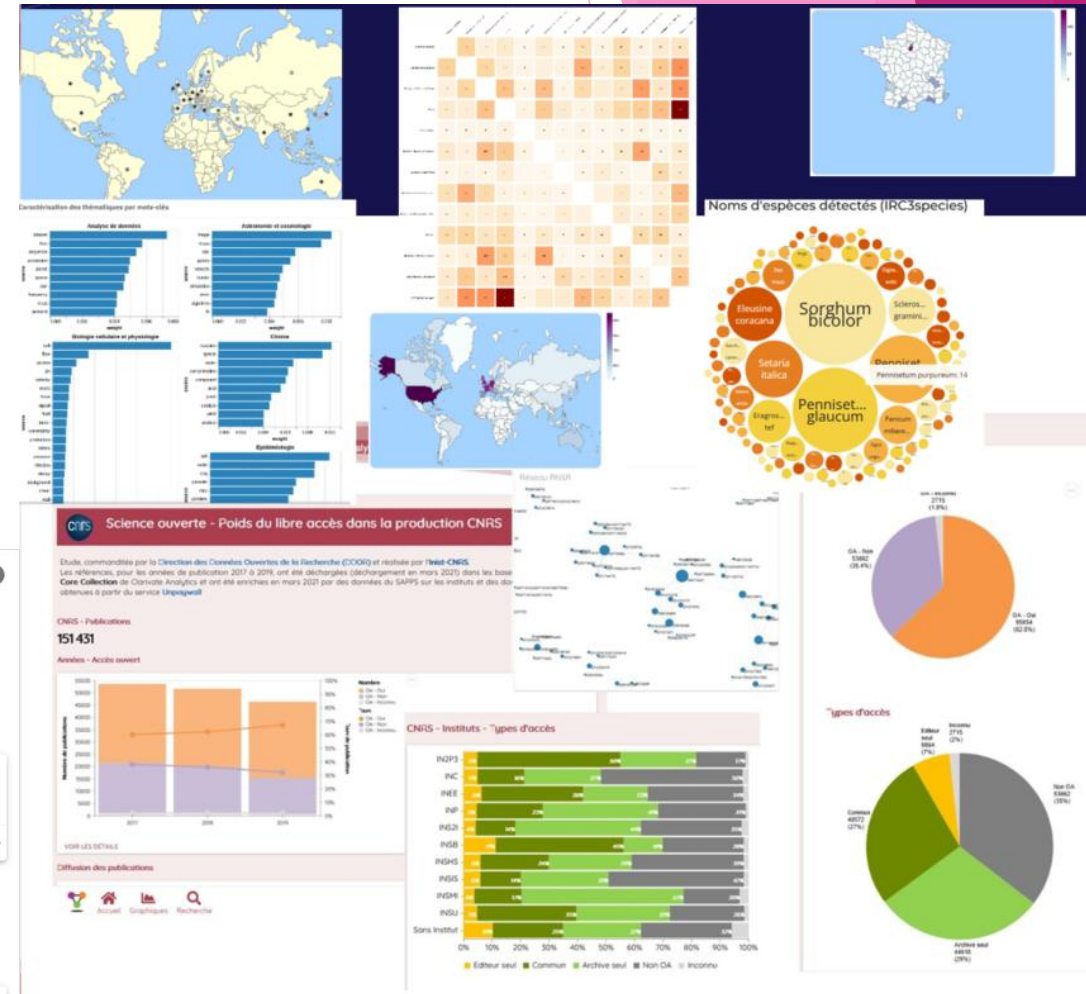
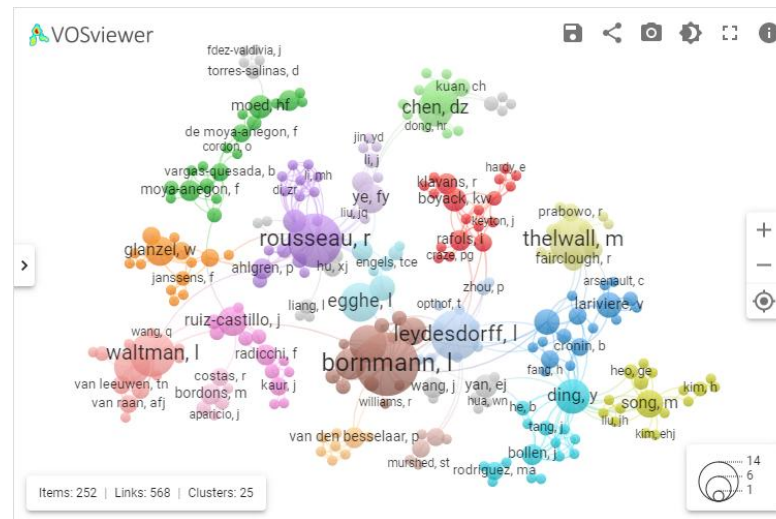
Réseaux

....

Lodex



VOSviewer



Prochaine séance : découverte du logiciel de visualisation de données Lodex et les web services de fouille de textes



Merci pour votre attention