

INTRODUCTION À LA BIBLIOMÉTRIE ET SES OUTILS

Abdelghani Maddi, ingénieur de recherche au GEMASS

CONTENU DE LA PRÉSENTATION

- Contexte scientifique, politique et économique
- Bibliométrie : de quoi parle-t-on ?
- Quels outils ?
- Quels enjeux ?
- Exemple d'analyse

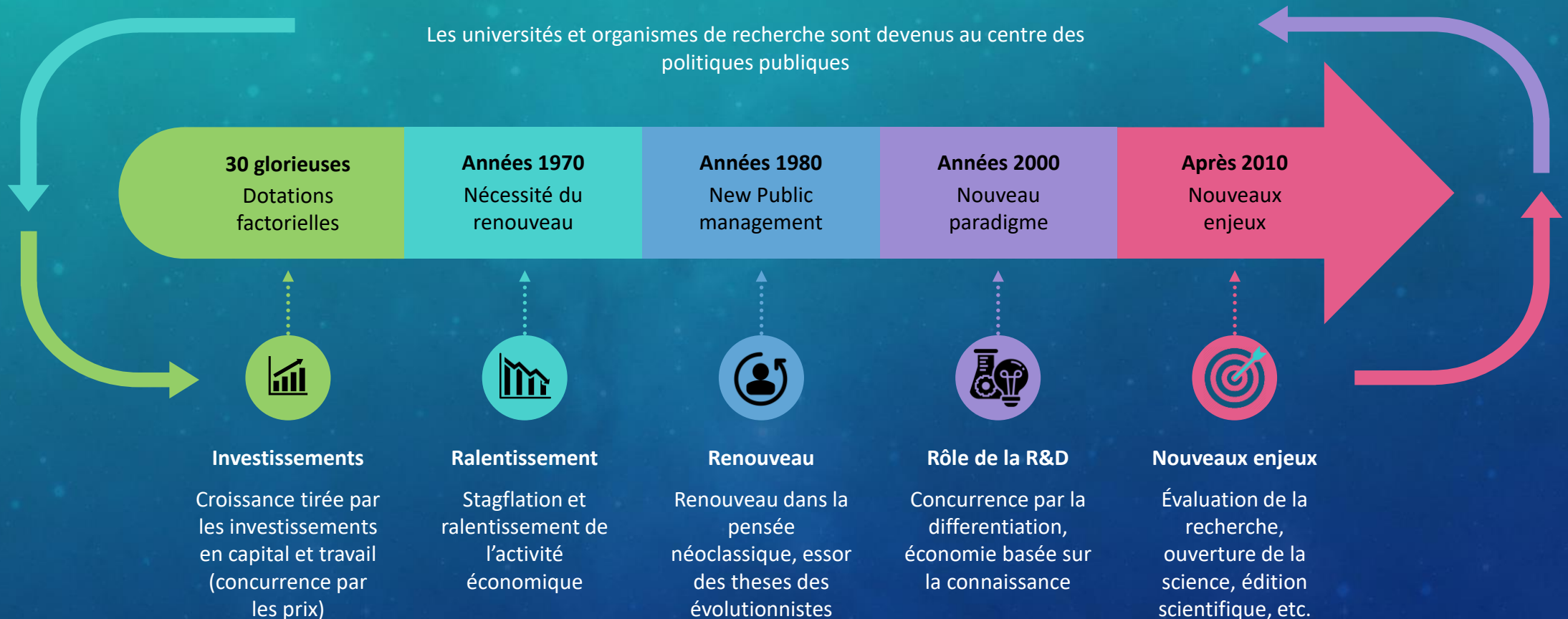
ÉLÉMENTS DE CONTEXTE

CONTEXTE ÉCONOMIQUE ET POLITIQUE

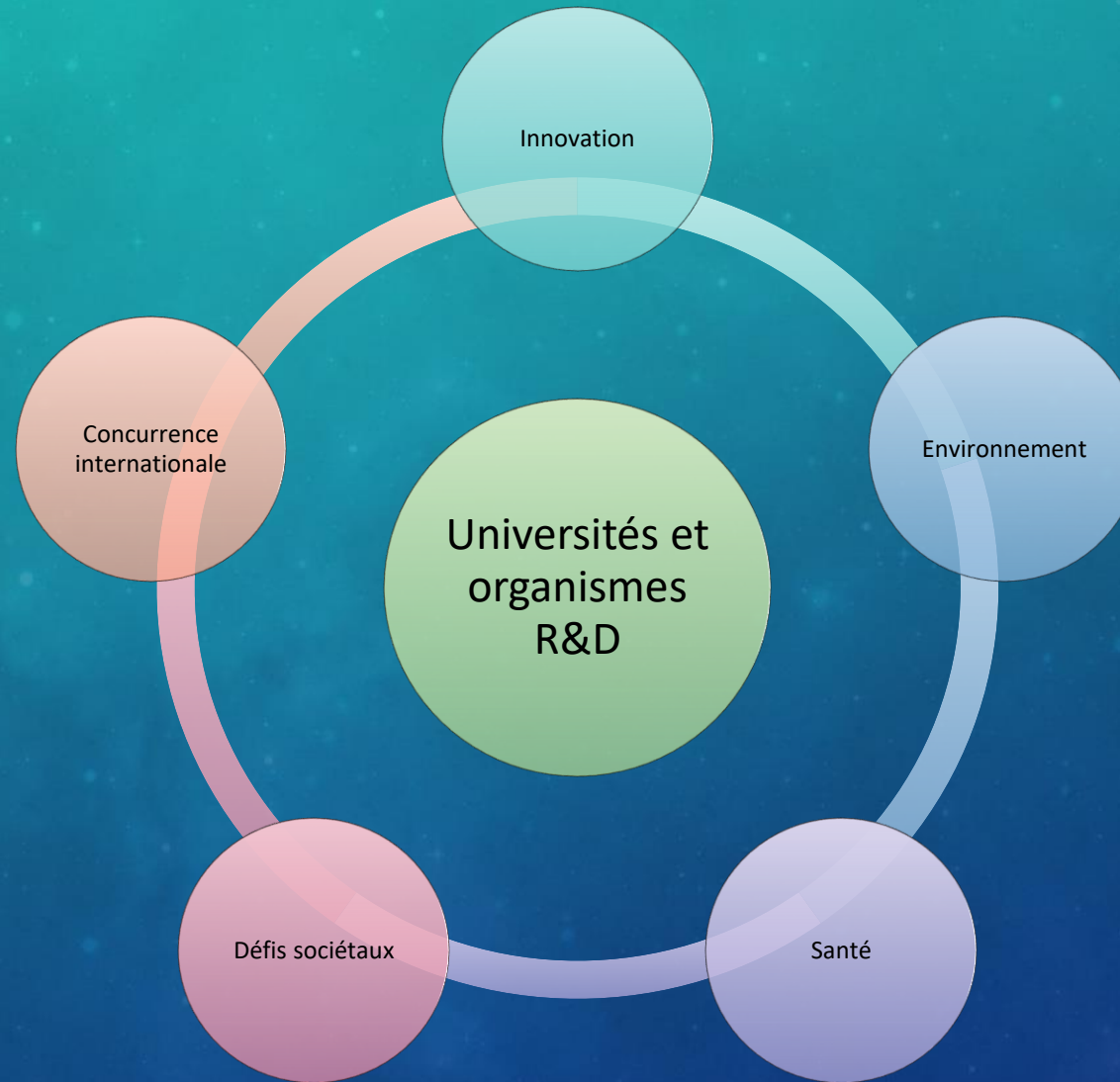
Transformation des politiques économiques des pays à hauts revenus

Vers une économie basée sur la connaissance

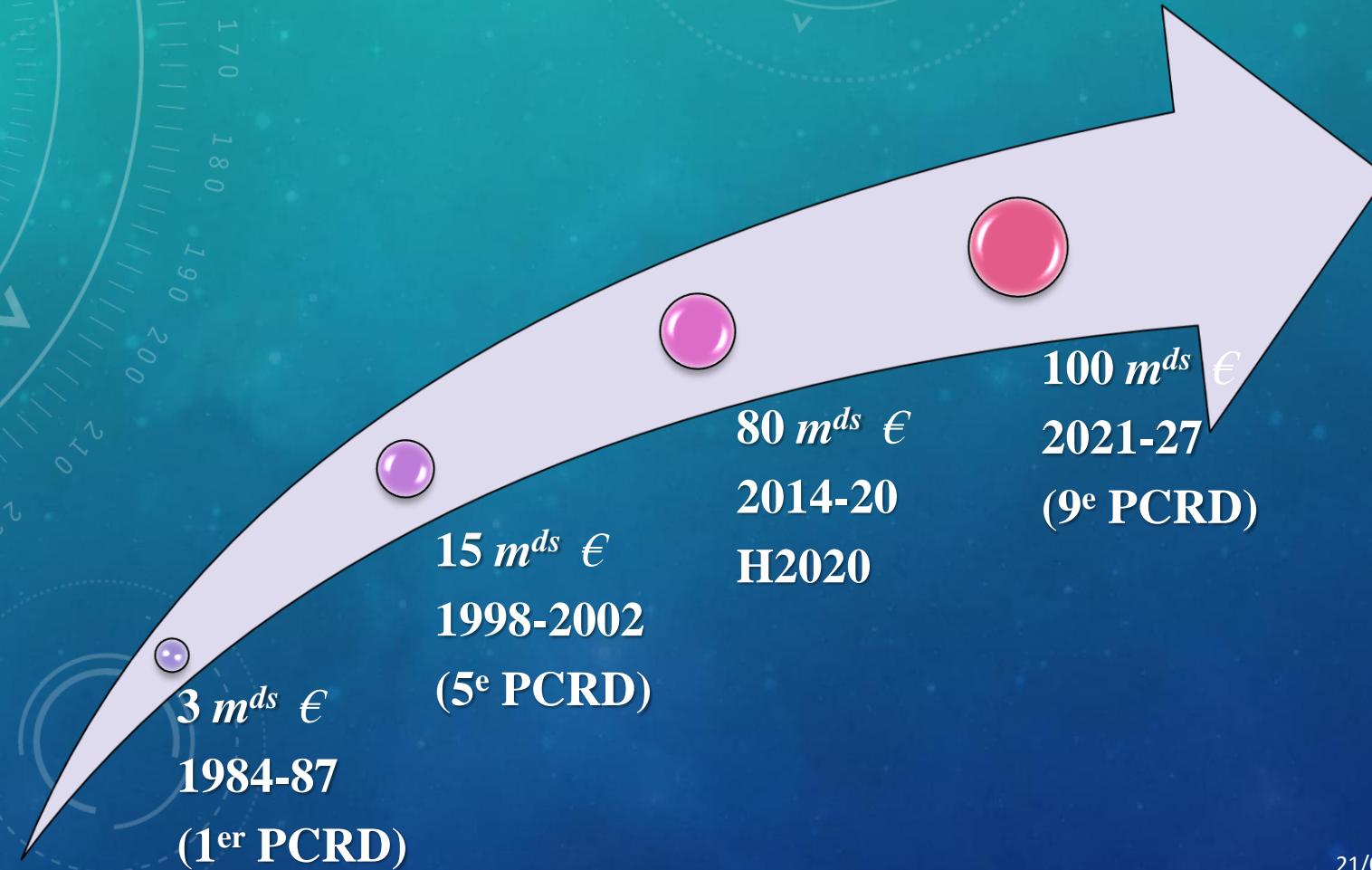
Les universités et organismes de recherche sont devenus au centre des politiques publiques



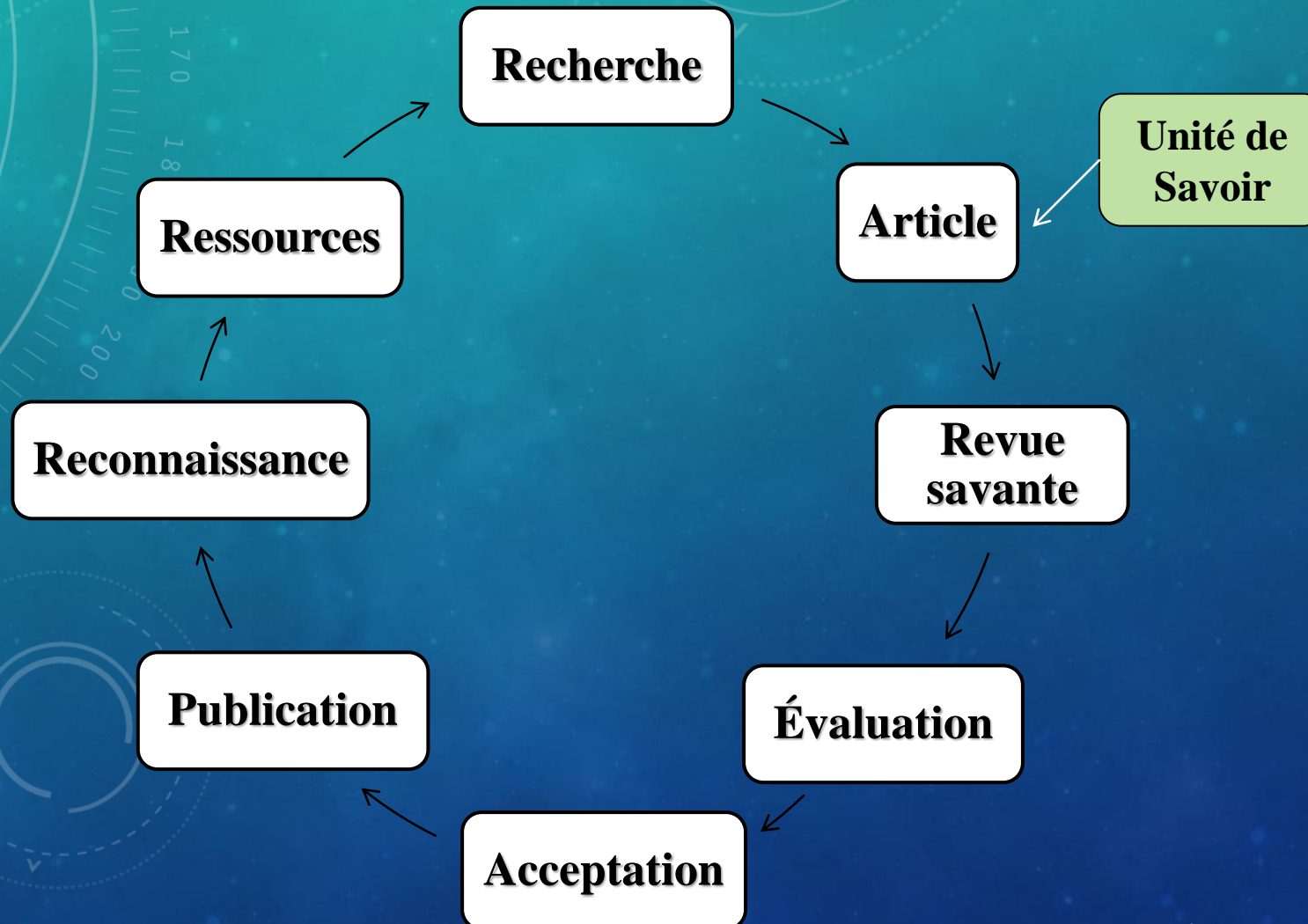
LES UNIVERSITÉS ET ORGANISMES DE RECHERCHE AU CENTRE



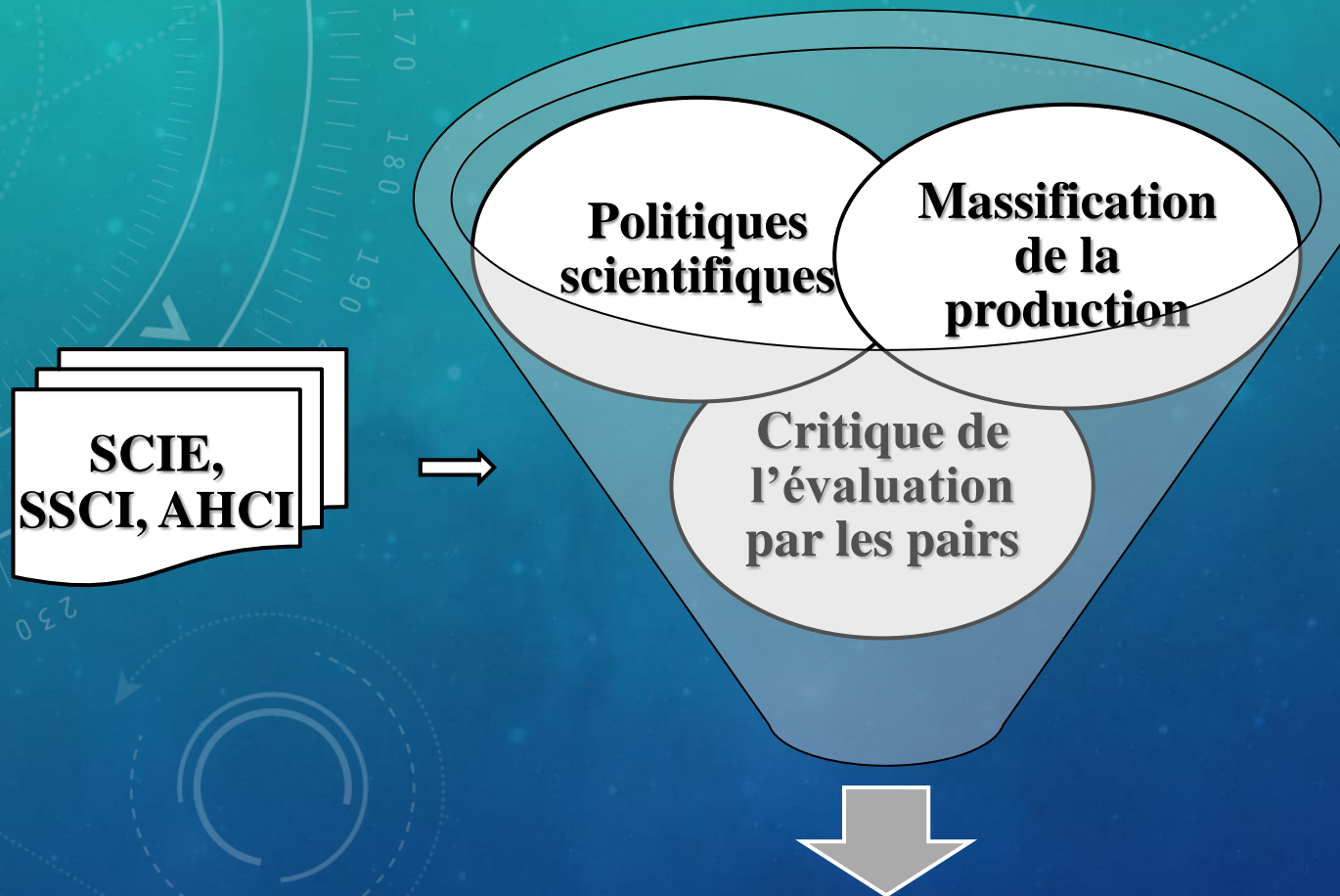
Montants alloués aux programmes cadres R&D européens



Processus « classique » de création du savoir scientifique

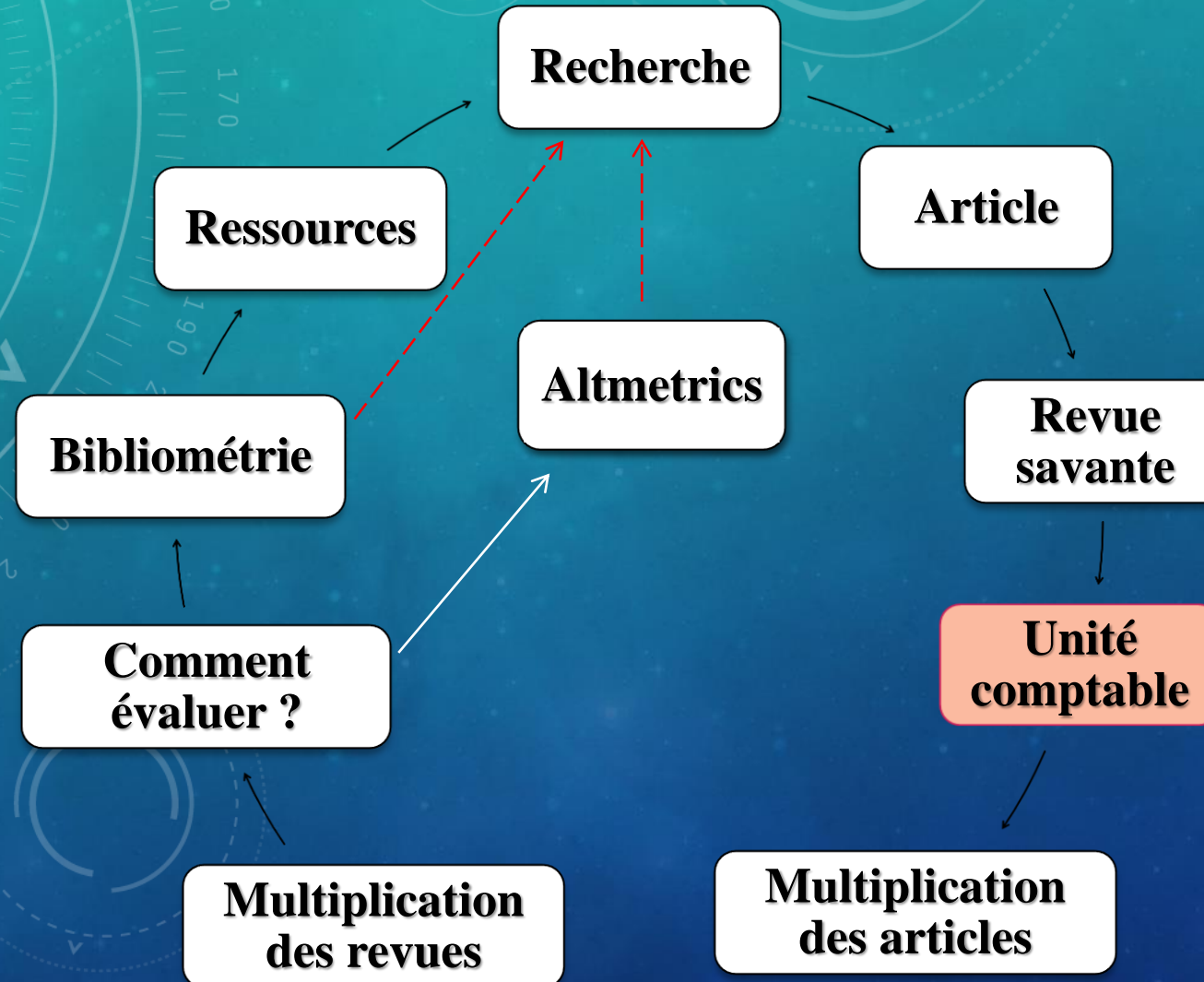


Évolution du processus de création du savoir scientifique



Bibliométrie évaluative

Évolution du processus de création du savoir scientifique

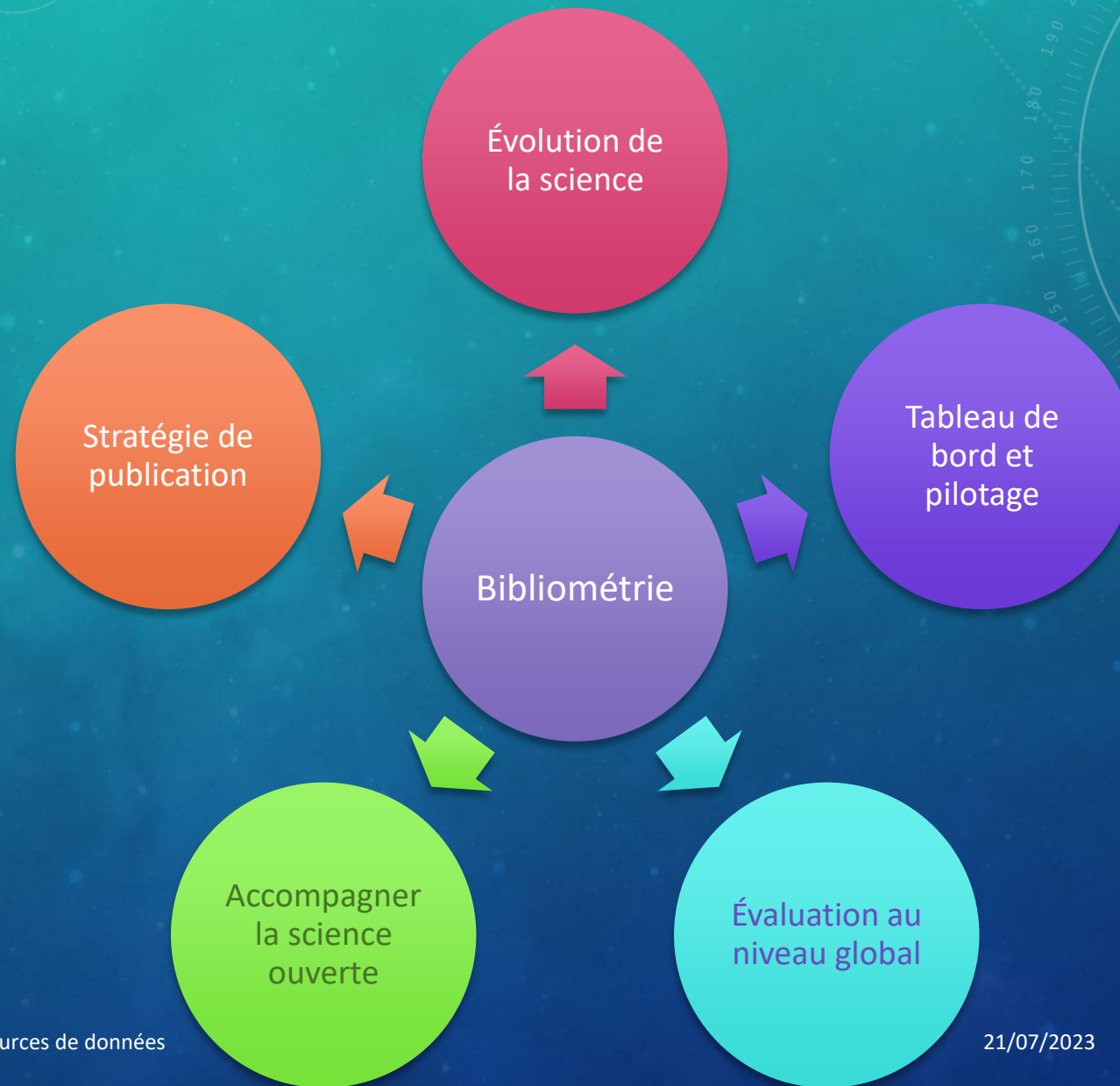


LA BIBLIOMÉTRIE

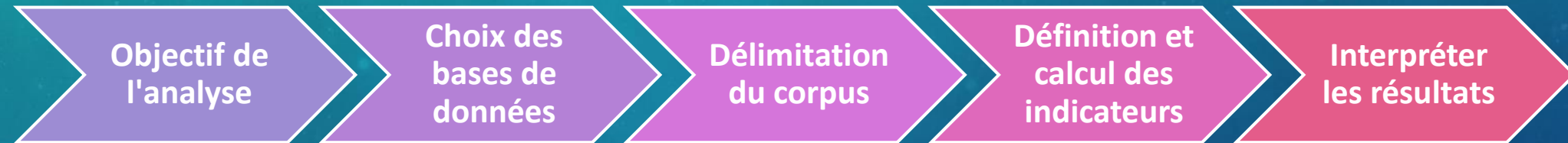
“bibliometrics is the application of mathematical and statistical methods to books and other media of communication”.

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

USAGES DE LA BIBLIOMÉTRIE



MÉTHODE



BOITE À OUTILS

MÉTADONNÉES UTILISÉES EN BIBLIOMÉTRIE



an open access journal



Citation: Pontika, N., Klebel, T., Correia, A., Metzler, H., Knoth, P., & Ross-Hellauer, T. (2022). Indicators of research quality, quantity, openness and responsibility in institutional review, promotion and tenure policies across seven countries. *Quantitative Science Studies*, 3(4), 888–911. https://doi.org/10.1162/qss_a_00224

DOI:
https://doi.org/10.1162/qss_a_00224

Peer Review:
https://publons.com/publon/10.1162/qss_a_00224

Supporting Information:
https://doi.org/10.1162/qss_a_00224

Received: 3 March 2022
Accepted: 25 October 2022

Corresponding Author:
Tony Ross-Hellauer
tross@know-center.at

Auteur de correspondance

Copyright: © 2022 Nancy Pontika, Thomas Klebel, Antonia Correia, Hannah Metzler, Petr Knoth, and Tony Ross-Hellauer. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

RESEARCH ARTICLE

Type de publication

Indicators of research quality, quantity, openness, and responsibility in institutional review, promotion, and tenure policies across seven countries

Nancy Pontika¹, Thomas Klebel², Antonia Correia³, Hannah Metzler^{4,5,6}, Petr Knoth¹, and Tony Ross-Hellauer^{2,7}

¹Knowledge Media Institute, Open University, Milton Keynes, UK

²Open and Reproducible Research Group, Know-Center GmbH, Graz, Austria

³University of Minho, Braga, Portugal

⁴Complexity Science Hub Vienna, Vienna, Austria

⁵Medical University of Vienna, Vienna, Austria

⁶Institute for Globally Distributed Open Research and Education, Gothenburg, Sweden

⁷Open and Reproducible Research Group, Graz University of Technology, Graz, Austria

Keywords: indicators, open science, research assessment, responsible research and innovation, rewards and recognition

ABSTRACT

The need to reform research assessment processes related to career advancement at research institutions has become increasingly recognized in recent years, especially to better foster open and responsible research practices. Current assessment criteria are believed to focus too heavily on inappropriate criteria related to productivity and quantity as opposed to quality, collaborative open research practices, and the socioeconomic impact of research. Evidence of the extent of these issues is urgently needed to inform actions for reform, however. We analyze current practices as revealed by documentation on institutional review, promotion, and tenure (RPT) processes in seven countries (Austria, Germany, India, Portugal, the United Kingdom and the United States). Through systematic coding and analysis of 143 RPT policy documents from 107 institutions for the prevalence of 17 criteria (including those related to qualitative or quantitative assessment of research, service to the institution or profession, and open and responsible research practices), we compare assessment practices across a range of international institutions to significantly broaden this evidence base. Although the prevalence of indicators varies considerably between countries, overall we find that currently open and responsible research practices are minimally rewarded and problematic practices of quantification continue to dominate.

Titre

Auteurs

Adresses

Mots-clés

Résumé

Contributions des auteurs

Financement

Données

Références bibliographiques

Indicators of research quality, quantity, openness, and responsibility

AUTHOR CONTRIBUTIONS

Nancy Pontika: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing—Original draft, Writing—Review & editing. Thomas Klebel: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—Original draft, Writing—Review & editing. Antonia Correia: Data curation, Investigation, Writing—Review & editing. Hannah Metzler: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing—Review & editing. Petr Knoth: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing—Review & editing. Tony Ross-Hellauer: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing—Original draft, Writing—Review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

This work was supported by the project “ON-MERRIT,” funded by the European Commission under the Horizon 2020 program (grant number 824612).

DATA AVAILABILITY

All supporting data and required code are available in Zenodo (Pontika et al., 2022b).

REFERENCES

- Adler, R., Ewing, J., & Taylor, P. (2009). Citation statistics. *Statistical Science*, 24(1), 1–14. <https://doi.org/10.1214/09-STS285>
- Alperin, J. P., Muñoz Nieves, C., Schimanski, L., McKiernan, E. C., & Niles, M. T. (2019). *Terms and concepts found in tenure and promotion guidelines from the US and Canada*. Data set. <https://doi.org/10.7910/DVN/VY4TJE>
- Alperin, J. P., Schimanski, L. A., La, M., Niles, M. T., & McKiernan, E. C. (2020). The value of data and other non-traditional scholarly outputs in academic review, promotion, and tenure in Canada and the United States. In *Open handbook of linguistic data management*. <https://doi.org/10.17613/ye06-n045>
- Blasius, J., & Greenacre, M. (2006). Correspondence analysis and related methods in practice. In J. J. H. van der Aalst (Ed.), *Multiple correspondence analysis* (pp. 4–40). Boca Raton: Chapman & Hall. <https://doi.org/10.1201/9781420011319.ch1>
- Brechelmacher, A., Park, E., & Alperin, J. P. (2020). The rocky road to tenure—Career paths in academia. In T. Fumasoli, G. Coastellic, & B. M. Kehm (Eds.), *Academic work and careers in Europe: Trends, challenges, perspectives* (pp. 13–40). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10720-2_2
- Brown, J. D. (2014). Citation searching for tenure and promotion: An overview of issues and tools. *Reference Services Review*, 42(1), 70–89. <https://doi.org/10.1108/RSR-05-2013-0023>
- Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications. *Journal of the Association of College and Research Libraries*, 91(1), 1–14. <https://doi.org/10.5860/jaalcrl.91.1.1>
- Coonin, B., & Younce, L. (2009). Publishing in open access journals in the social sciences and humanities: Who’s doing it and why. In *Pushing the edge: Proceedings of the Fourteenth National Conference of the Association of College and Research Libraries*.
- Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1(3), 193–203. <https://doi.org/10.1016/j.joi.2007.02.001>
- Curry, S. (2018). Let’s move beyond the rhetoric: It’s time to change how we judge research. *Nature*, 554(7691), 147. <https://doi.org/10.1038/d41586-018-01642-w>, PubMed: 29420505
- European Commission, Directorate General for Research and Innovation. (2020). *Open science: A new paradigm for research and innovation*. Luxembourg: Publications Office of the European Union.
- Gadd, E. (2021). Mis-measuring our universities: Why global university rankings don’t add up. *Frontiers in Research Metrics and Analytics*, 6, 680023. <https://doi.org/10.3389/fрма.2021.680023>, PubMed: 34568739
- Gardner, S. K., & Veliz, D. (2014). Evincing the ratchet: A thematic analysis of the promotion and tenure guidelines at a striving university. *Review of Higher Education*, 38(1), 105–132. <https://doi.org/10.1353/rhe.2014.0045>
- Greenacre, M., & Nenadic, O. (2018). *Ca: Simple, multiple and*

BASES DE DONNÉES

PRINCIPAUX UTILISATEURS DES BASES DE DONNÉES BIBLIOMÉTRIQUES (Leydesdroff et al. 2016)

Producteurs

- Fournisseurs de données (WoS, Scopus, etc.)
- Développeurs d'indicateurs (CWTS, OST, ECOOM, etc.)

Bibliomètres

- Spécialistes dans le domaines qui font les développements théoriques

Scientifiques

- Communauté scientifique au sens large (auto-évaluation, recherche documentaire, etc.)

Gestionnaires

- Universités (RH, administrateurs, etc.)
- Agences de financement (ANR, ERC, etc.)
- Agences d'évaluation (Hcéres, ANVUR, REF, etc.)

QUELLE(S) BASE(S) DE DONNÉES ?

Clarivate™



Scopus®

unpaywall

Google Scholar

ORCID

Connecting Research
and Researchers



Microsoft Academic



Dimensions

1964

2000

2004

2009

2011

2012

2013

2015

2017

2018

2022



PubMed
Central



DataCite
FIND, ACCESS, AND REUSE DATA



LENS.ORG
Solving The Problem Of Problem Solving™



Altmetric



Crossref

I40C



OpenAlex

Base de données	Gratuite	Propriétaire
		✓
	✓	
	✓	
	✓	
	✓	
	✓	
	✓	
	✓	
	✓	
		✓
	✓	
	✓	✓

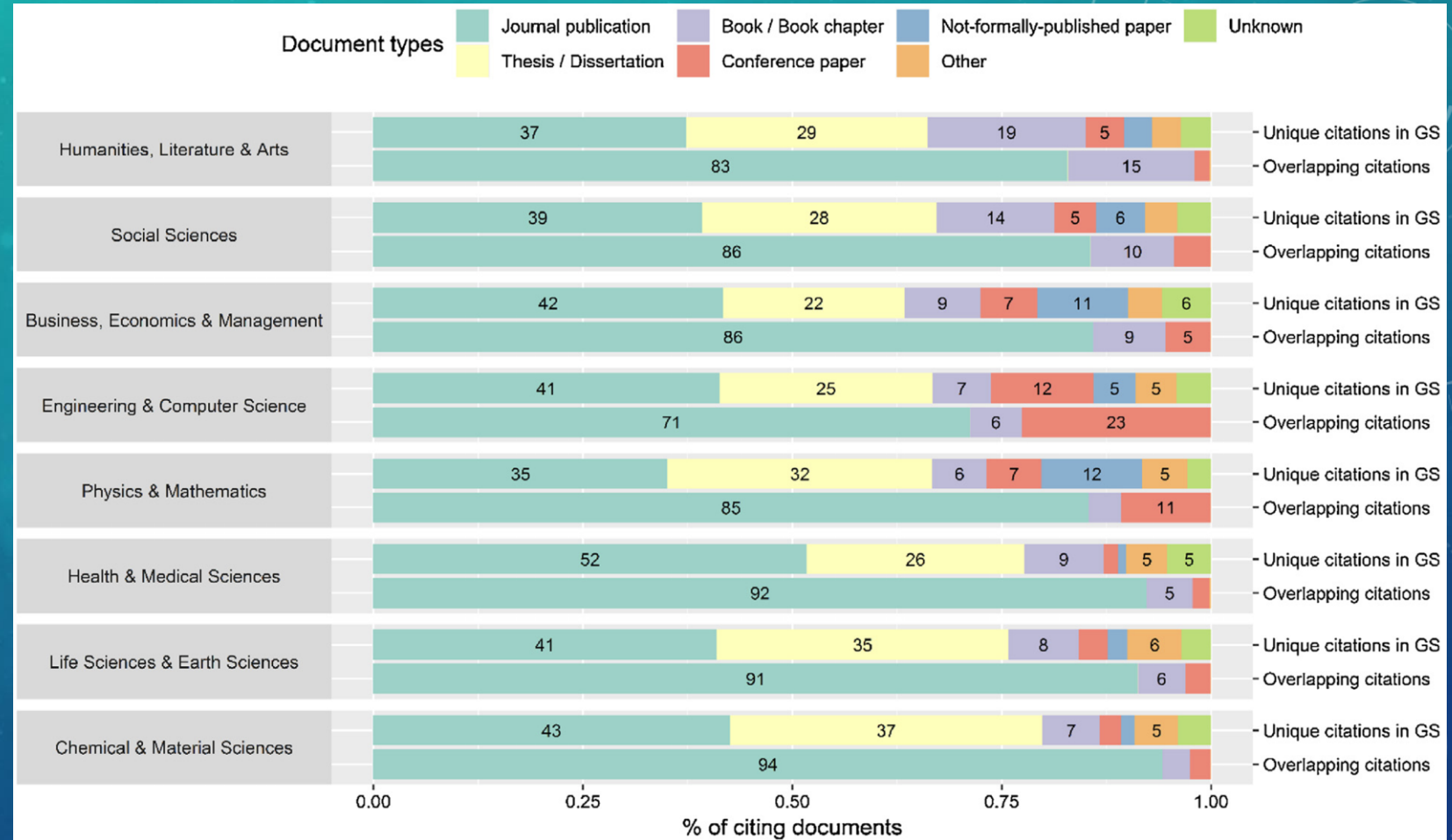
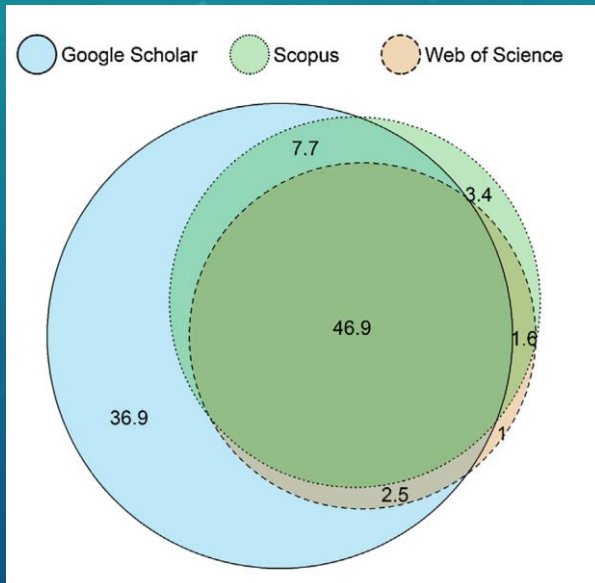
PRINCIPALES BASES DE DONNÉES

Le choix de la base de données dépend de plusieurs facteurs, notamment :

- 1- La qualité de ses données,
- 2- Son niveau de couverture,
- 3- Le sujet de recherche.

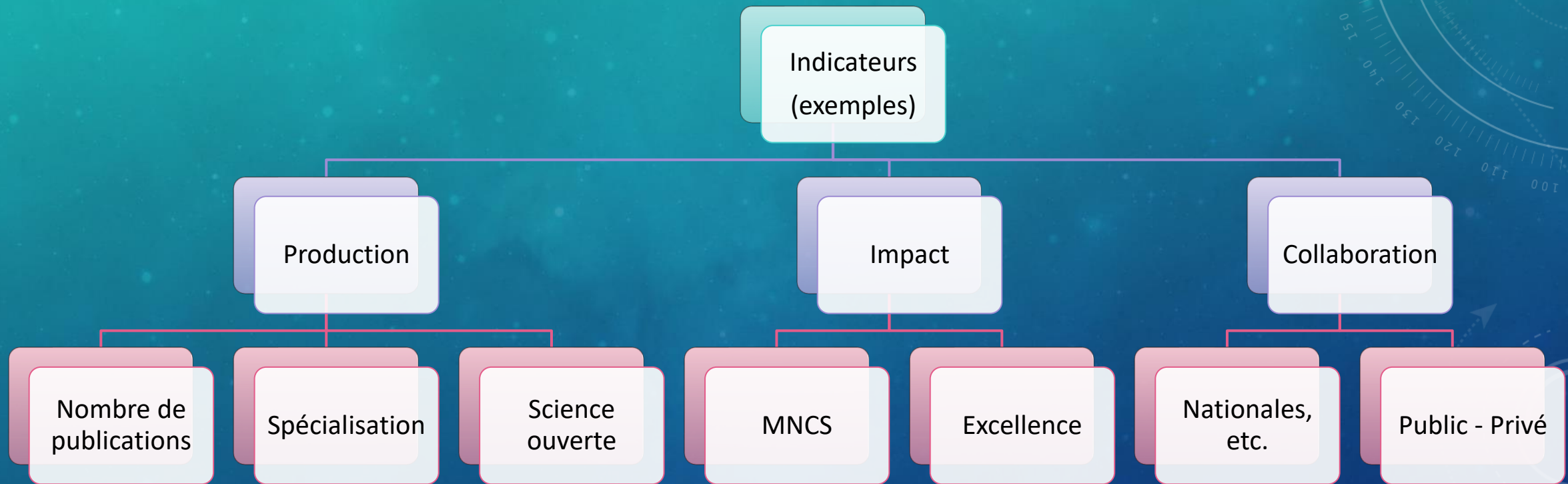
COUVERTURE DES PRINCIPALES BASES DE DONNÉES

- Types de document (articles, livres, conférences, etc.)
- Couverture géographique
- Couverture disciplinaire










Source : <https://www.sciencedirect.com/science/article/abs/pii/S1751157718303249>

QUELS OUTILS À DISPOSITION ?



QUELS OUTILS À DISPOSITION ?

Outils gratuits (liste indicative)	Type d'analyse
 python™	Moissonnage et récupération des données via API, préparation des données, analyse sémantique, désambiguïsation, NLP, ...
 R Studio®	Analyse statistique, topic modeling, visualisation des données, ... (moins performant pour le moissonnage)
 CORTEXT	Traitement et visualisation des données (marche mieux avec données issues des interfaces en ligne, ex. WoS, Scopus)
 vosViewer	Cooccurrences, cocitation, réseaux de collaboration, ... (moins de possibilités que CorTextT). VosViewer : travail en local (les perf. dépendent de la machine). CorTextT : sur des serveurs distants (en 2023, environ 3,5 To de RAM)
 Gargantext	Traitement du texte / text-mining
 Gephi	Visualisation des données (analyse des réseaux)
 Khartis	Visualisation des données géographiques

ENJEUX DE LA BIBLIOMÉTRIE

Quelques problèmes des mesures quantitatives

- Biais ex-ante (caractéristiques des évalués : ex. sexe).
- Bases de données non exhaustives (pas inclusives)
 - Types de produits indexés => biais disciplinaire.
 - Biais linguistiques et géographiques.
- Problèmes de construction de certains indicateurs (indice H, facteur d'impact, MNCS, etc.).

Exemple du Mean Normalized Citation Score (MNCS)

Soit 2 équipes de recherche de même taille :

Équipe de recherche 1

- Discipline : économie de la santé
- Ressources : 200 000 €
- Production :
 - 80 publications dont :
 - 50 => NCS = 2,00
 - 30 => NCS = 0,50
 - $MNCS = ((50*2)+(30*0,5))/80 = 1,44$

Équipe de recherche 2

- Discipline : économie de la santé
- Ressources : 200 000 €
- Production :
 - 120 publications dont :
 - 60 => NCS = 2,00
 - 60 => NCS = 0,50
 - $MNCS = ((60*2)+(60*0,5))/120 = 1,25$

➔ En utilisant le MNCS le group de recherche 1 > 2

- L'équipe 2 a produit 60 publications hautement citées contre 50 pour le groupe 1
- L'équipe 2 a produit 60 publications faiblement citées contre 30 pour le groupe 1
 - ➔ l'équipe 2 est meilleure (voir : Abramo et d'Angelo, 2016)

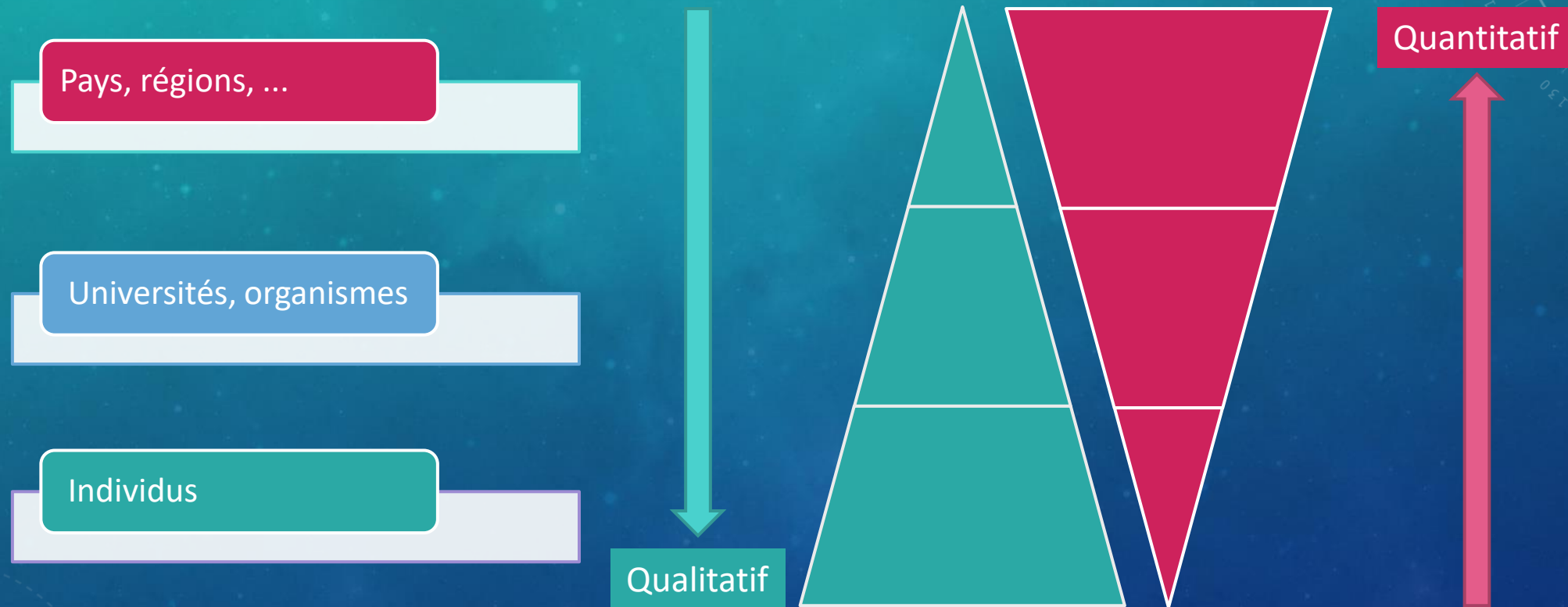
Spécificités des SHS

Orientation nationale et régionale

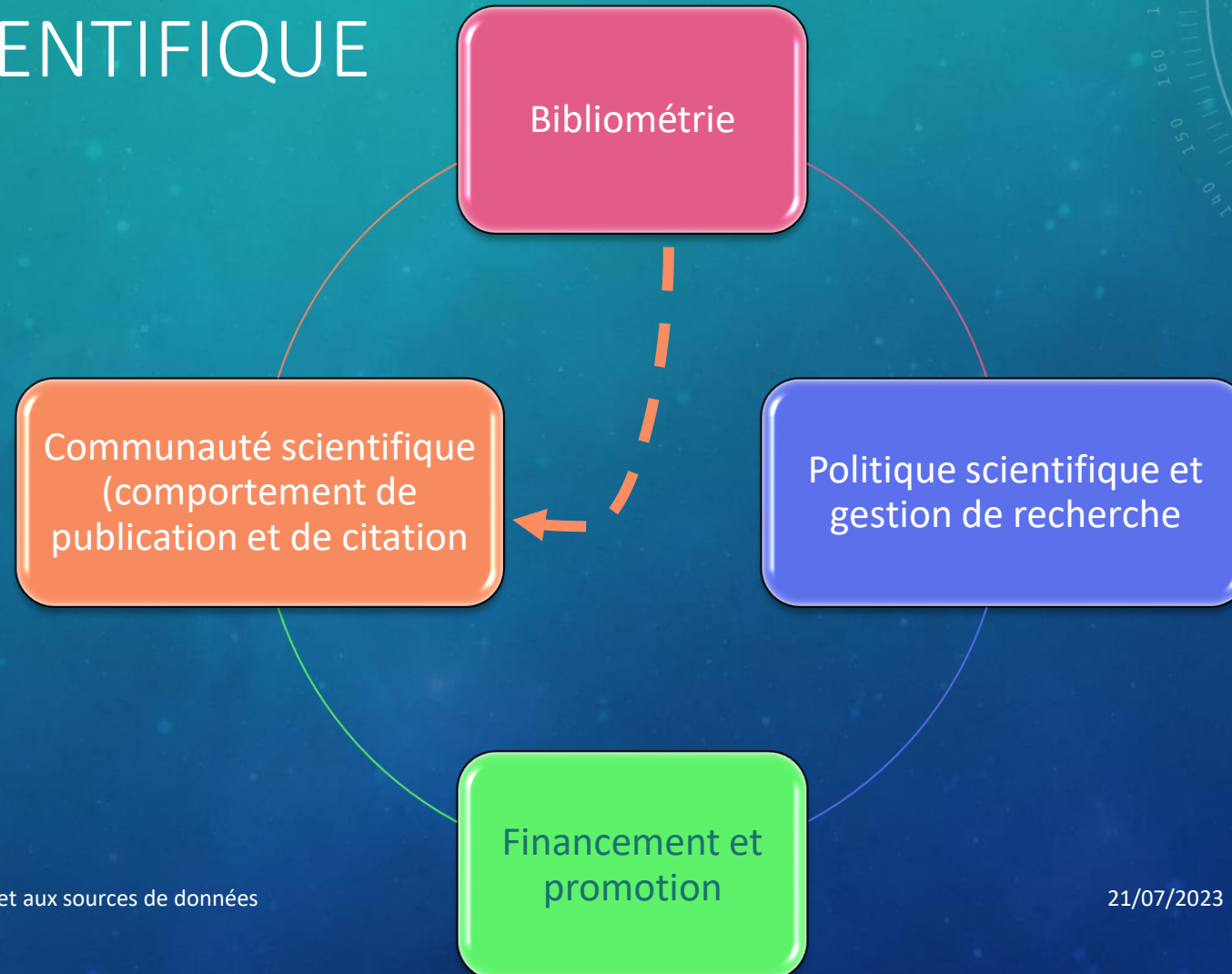
Visent un public plus large (autres produits que les articles)

Pratiques de citation différentes

ADAPTER L'ANALYSE / L'ÉVALUATION EN FONCTION DE LA GRANULARITÉ



IMPACT DES MÉSUSAGES DE LA BIBLIOMÉTRIE SUR LA RECHERCHE SCIENTIFIQUE





RAPPEL SUR LES MÉTHODES DE
« *CLUSTERISATION* »

EXEMPLE SUR L'ANALYSE DE RÉSEAUX

Objectif

- Analyser les publications qui traitent la thématique du Covid-19

Base de données

- Exemple sur la base Web of Science : <https://access.webofknowledge.com/>

Corpus

- Toutes les publications qui contiennent « Covid-19 » dans le titre, résumé ou mots-clés

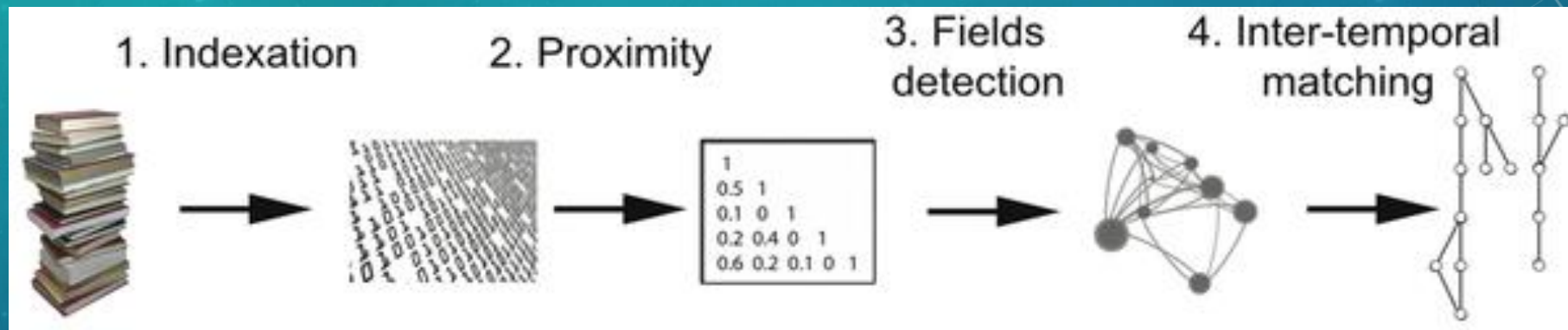
Indicateurs

- Exemple : indicateurs de collaboration, analyse de cooccurrence, etc.

Interprétation

- Analyser et visualiser les données avec Vosviewer : <https://www.vosviewer.com/download>

•Steps contributing towards the reconstruction of a phylomemy.



Chavalarias D, Cointet JP (2013) Phylomemetic Patterns in Science Evolution—The Rise and Fall of Scientific Fields. PLOS ONE 8(2): e54847. <https://doi.org/10.1371/journal.pone.0054847>
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0054847>

Textual Analysis

We believed we could reduce our dependence on foreign oil and protect our planet. And today, America is number one in oil and gas.

Source : Modeling and mapping knowledge dynamics with CorText platform, support de formation, 2017,
<https://docs.cortext.net/training-materials/>

Textual Analysis

Part-Of-Speech Tagging

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil
CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in
NN CC NN
oil and gas.

Extracted noun phrases:

- dependence
- planet
- oil
- gas

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil
CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in
NN CC NN
oil and gas.

Extracted noun phrases:

- *dependence*
- *planet*
- *oil*
- *gas*
- *foreign oil*

Textual Analysis

Chunking

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil
CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in
NN CC NN
oil and gas.

Extracted noun phrases:

- *dependence*
- *planet*
- *oil*
- *gas*
- *foreign oil*
- *dependence on foreign oil*
- *oil and gas*

Textual Analysis

Stemming, Filtering and Standardizing

PRP VBD PRP MD VB PRP NN IN JJ NN
We believed we could reduce our dependence on foreign oil

CC VB PRP NN. CC NN NNP VBZ NN CD IN
and protect our planet. And today, America is number one in

NN CC NN
oil and gas.

Extracted classes:

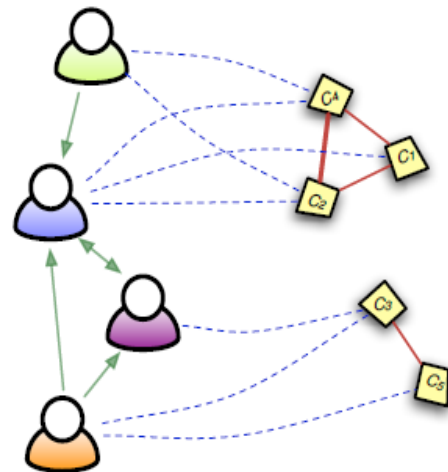
- *dependence on foreign oil: {dependence on foreign oil ; foreign oil dependence}*
- *oil and gas: {oil and gas; gas and oil}*
- *planet: {planet, planets}*
- *etc.*

Source : Modeling and mapping knowledge dynamics with CorText platform, support de formation, 2017,
<https://docs.cortext.net/training-materials/>

Cooccurrences matrices

Cooccurrences matrices construction

- Occurrence matrix O : $O_{ij} = 1$ iff item i is used in document j , 0 otherwise
- The cooccurrence matrix enumerates every joint appearances of two items in the same document. $C = O^t O$

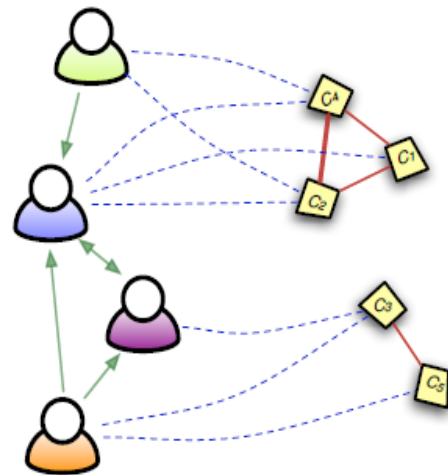


documents \ items	A	B	C	D
C1		1		
C2	1	1		
C3			1	1
C4	1	1		
C5				1

Cooccurrences matrices

Cooccurrences matrices construction

- Occurrence matrix O : $O_{ij} = 1$ iff item i is used in document j , 0 otherwise
- The cooccurrence matrix enumerates every joint appearances of two items in the same document. $C = O^t O$



	C1	C2	C3	C4	C5
C1	x	1		1	
C2	1	x	1	2	
C3			x		1
C4	1	2		x	
C5			1		x

Proximity Measures

Cooccurrences variables

- Co-Occurrence matrix C :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

	C1	C2	C3	C4	C5
C1		1		1	
C2	1		1	2	
C3					1
C4	1	2			
C5			1		

Proximity Measures

Cooccurrences variables

- Co-Occurrence matrix C :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

	C1	C2	C3	C4	C5
C1		1		1	
C2	1		1	2	
C3					1
C4	1	2			
C5			1		

Proximity Measures

Cooccurrences variables

- Co-Occurrence matrix C :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

	C1	C2	C3	C4	C5
C1		1		1	
C2	1		1	2	
C3					1
C4	1	2			
C5			1		

Proximity Measures

Cooccurrences variables

- Co-Occurrence matrix C :
 C_{ij} = number of joint occurrences of i and j in the same document
- total number of cooccurences of i : $s_i = \sum_{j, j \neq i} c_{ij}$
- global number of co-occurences : $N = \sum_i s_i$
- expected number of cooccurences : $e_{ij} = \frac{s_i s_j}{N}$

Direct Measures of Similarity :

- Raw cooccurences : $S_R(i, j) = c_{ij}$
- Association strength : $S_A(i, j) = \frac{c_{ij}}{s_i s_j}$
- Mutual Information : $S_{MI}(i, j) = \log(\frac{c_{ij}}{e_{ij}})$
- Cosine : $S_C(i, j) = \frac{c_{ij}}{\sqrt{s_i s_j}}$
- Inclusion index : $S_I(i, j) = \frac{c_{ij}}{\min(s_i, s_j)}$
- Jaccard index : $S_J(i, j) = \frac{c_{ij}}{s_i + s_j - c_{ij}}$
- χ^2 score : $S_{\chi^2}(i, j) = \frac{c_{ij} - e_{ij}}{\sqrt{e_{ij}}}$
- Cramer index : $S_{\text{cramer}}(i, j) = \frac{c_{ij} - e_{ij}}{e_{ij}}$

