



Cortext Manager : un instrument qui impacte la recherche en Science Sociales

P. Breucker¹, G. Enderli¹

(1) INRAE, UMR LISIS, 77454 Marne-la-Vallée

Plan

Introduction

1- L'instrument Cortext

2- Utiliser Cortext pour mesurer son impact

Conclusion



Constituer une plateforme **scientifique et technique** pour soutenir un espace de **recherche** sur les traces et les usages numériques de la science et de l'innovation en société



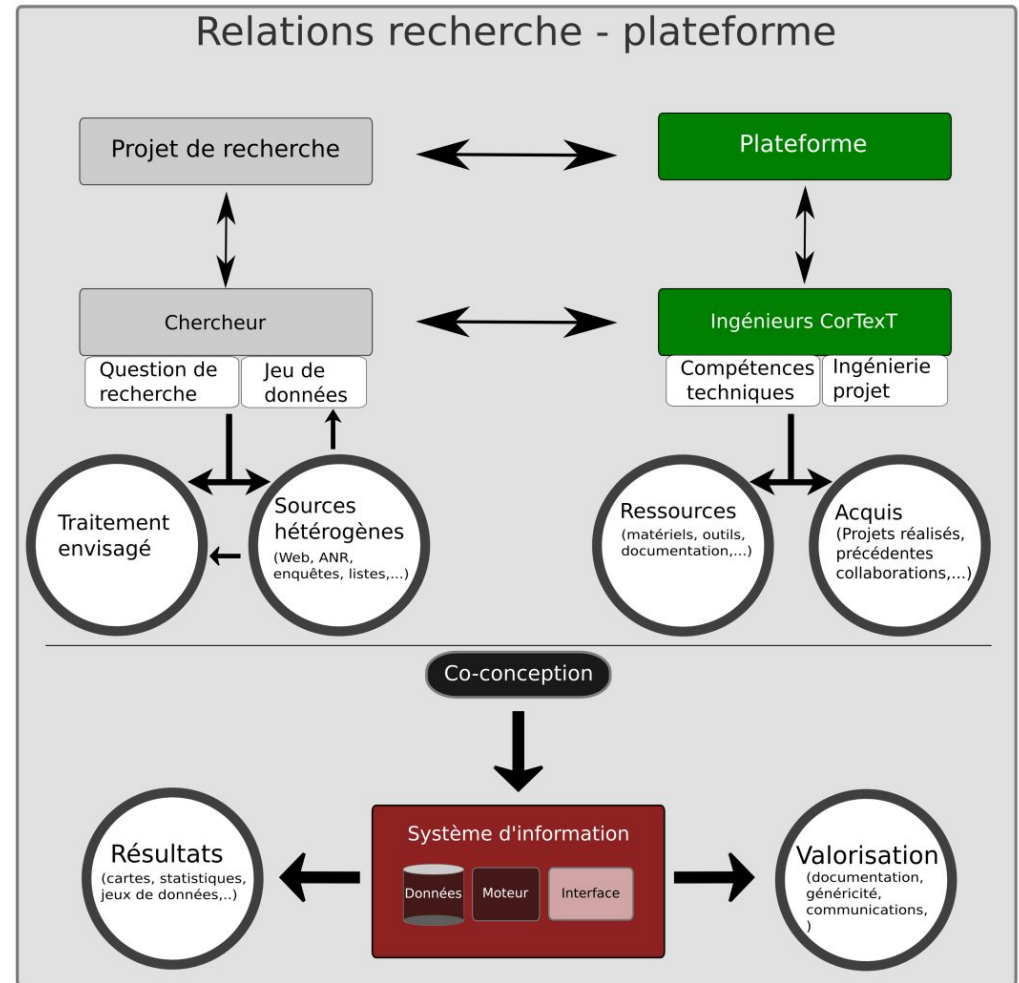
avec les soutiens de



GENÈSE DU PROJET CORTEXT

“L’objectif premier de la plateforme CorText est qu’un chercheur en sciences sociales – ou un autre utilisateur – puisse venir avec une question de recherche et partir en profitant des fruits d’une méthode computationnelle adaptée à sa question.”

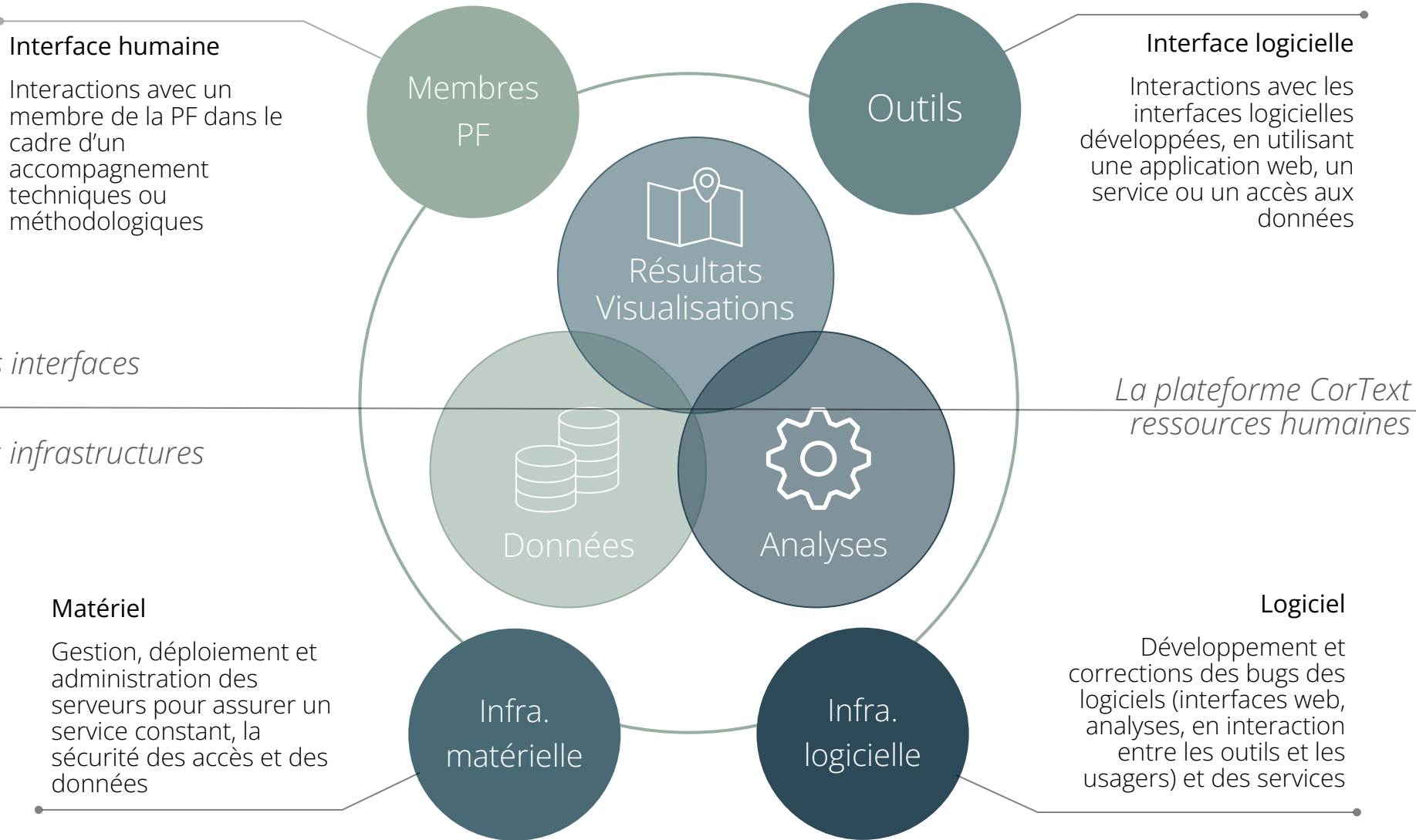
- ❑ Fondé en 2009
- ❑ Financé et soutenu par:
 - INRAE
 - GIS IFRIS et Labex SITES
 - Projet EU RISIS
- ❑ Intégré dans l’UMR LISIS



Principe fondateurs de la plate-forme CorText, 2009

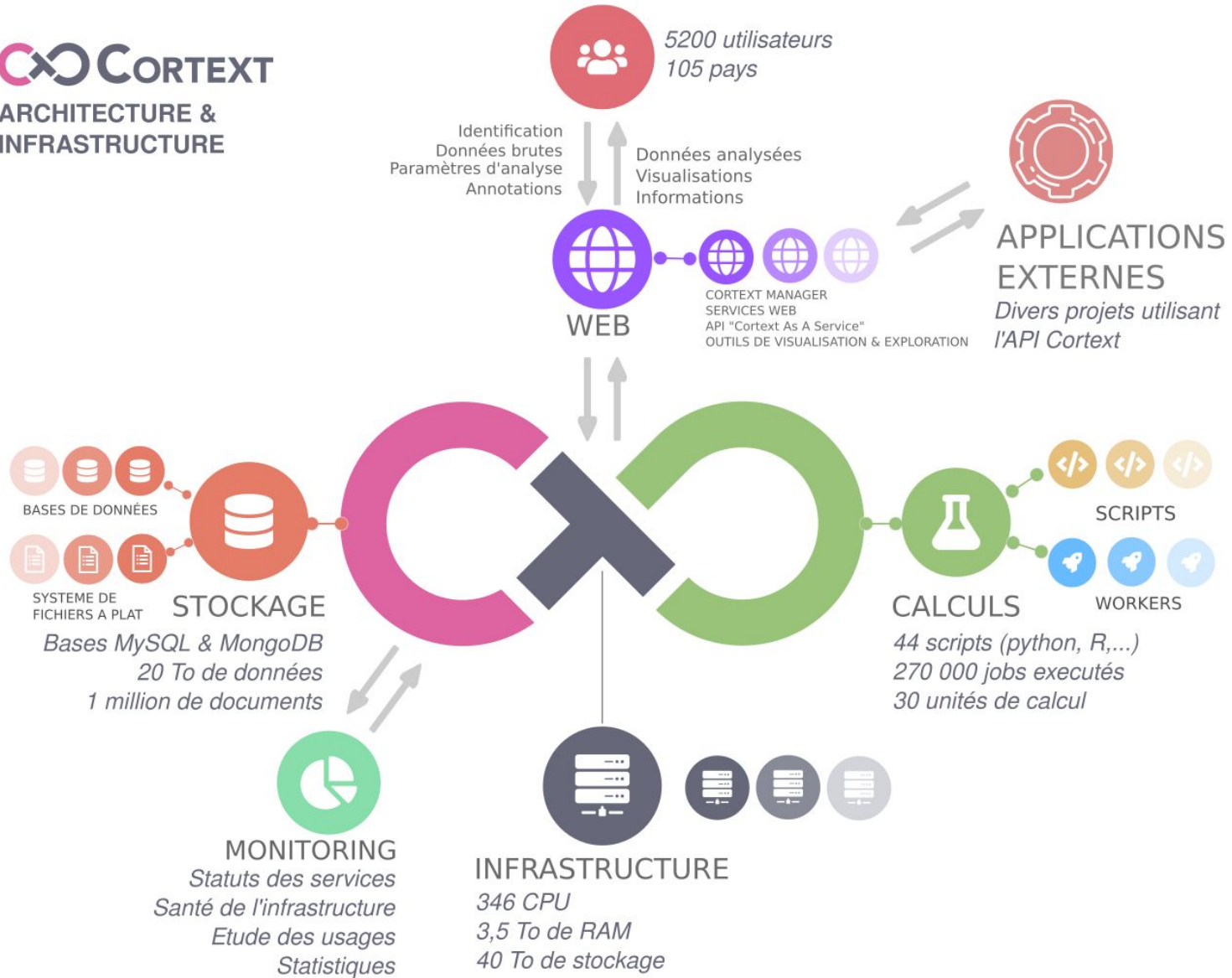
1- L'instrument CorText

L'instrument CorText

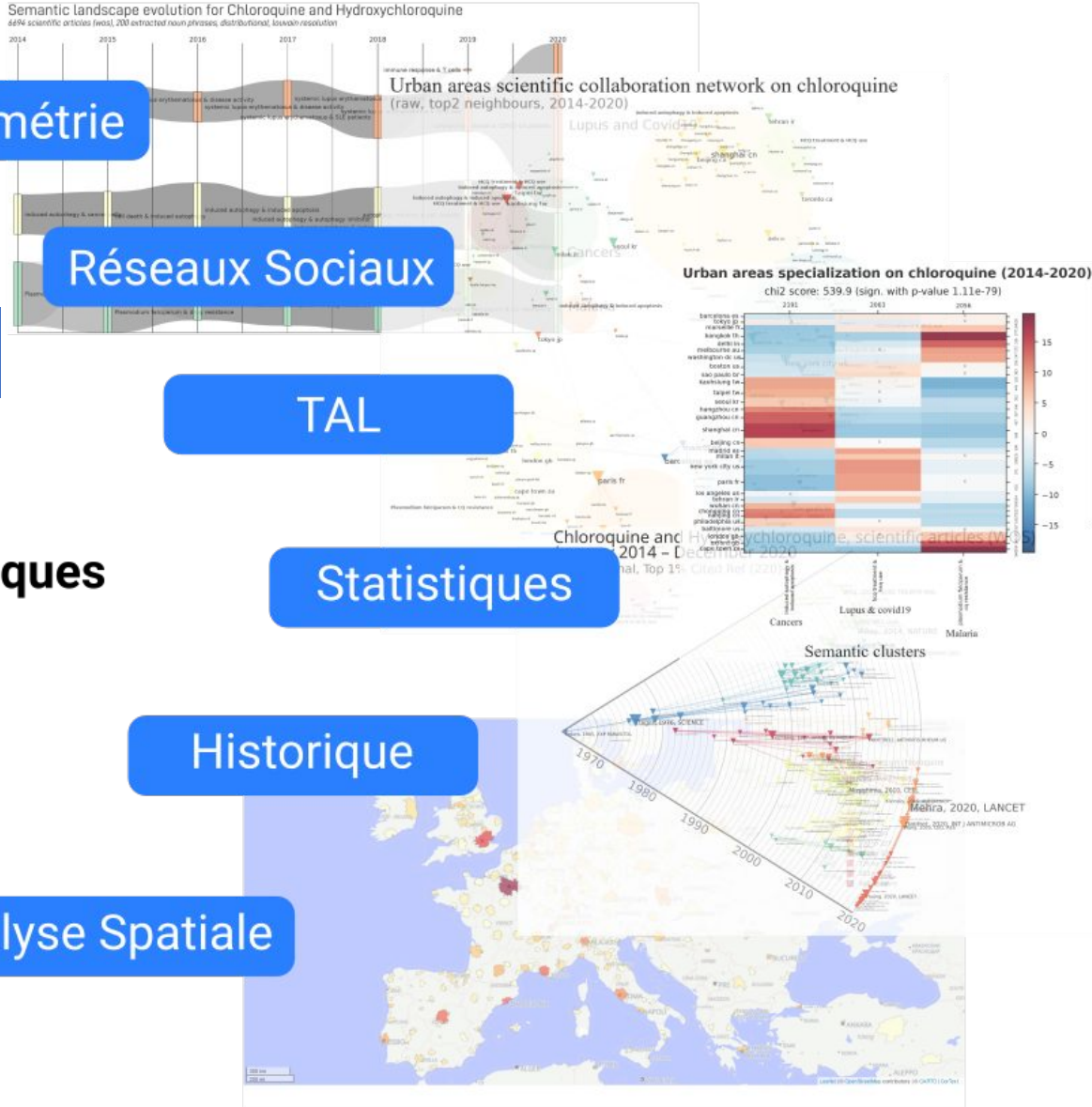


L'instrument CorText

CORTEXT
ARCHITECTURE & INFRASTRUCTURE



Scientométrie



CORTEXT MANAGER

Variété
des champs
méthodologiques

TAL

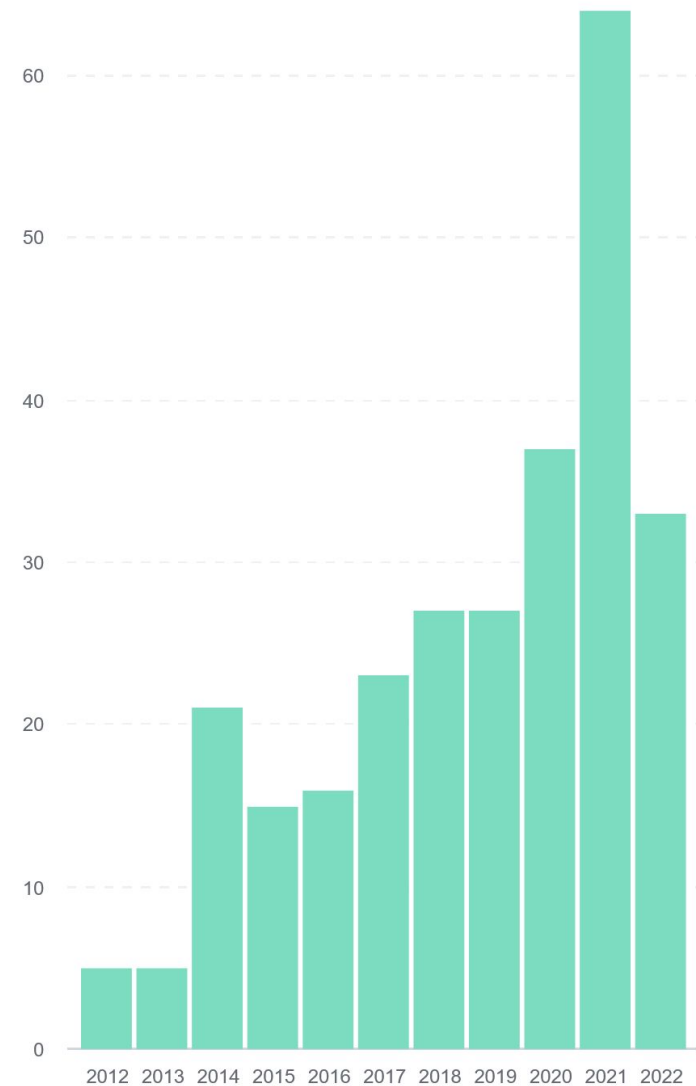
Statistiques

Historique

Analyse Spatiale

Quel impact ?

- ▣ Plus de 6500 utilisateurs en 2022
- ▣ Plus de 100 pays
- ▣ Dans le monde académique:
 - + de **640 auteurs** de publications citant Cortext depuis 2016, soit moins de 10% de la communauté
 - En **dehors de la France** : augmentations importantes en **Chine**, aux **Philippines** et au **Brésil**, avec la structuration de deux communautés à Wuhan et Manille
 - Les enseignements utilisant Cortext se multiplient



<https://www.cortext.net/publications/>

2- Utiliser CorText manager pour mesurer son impact

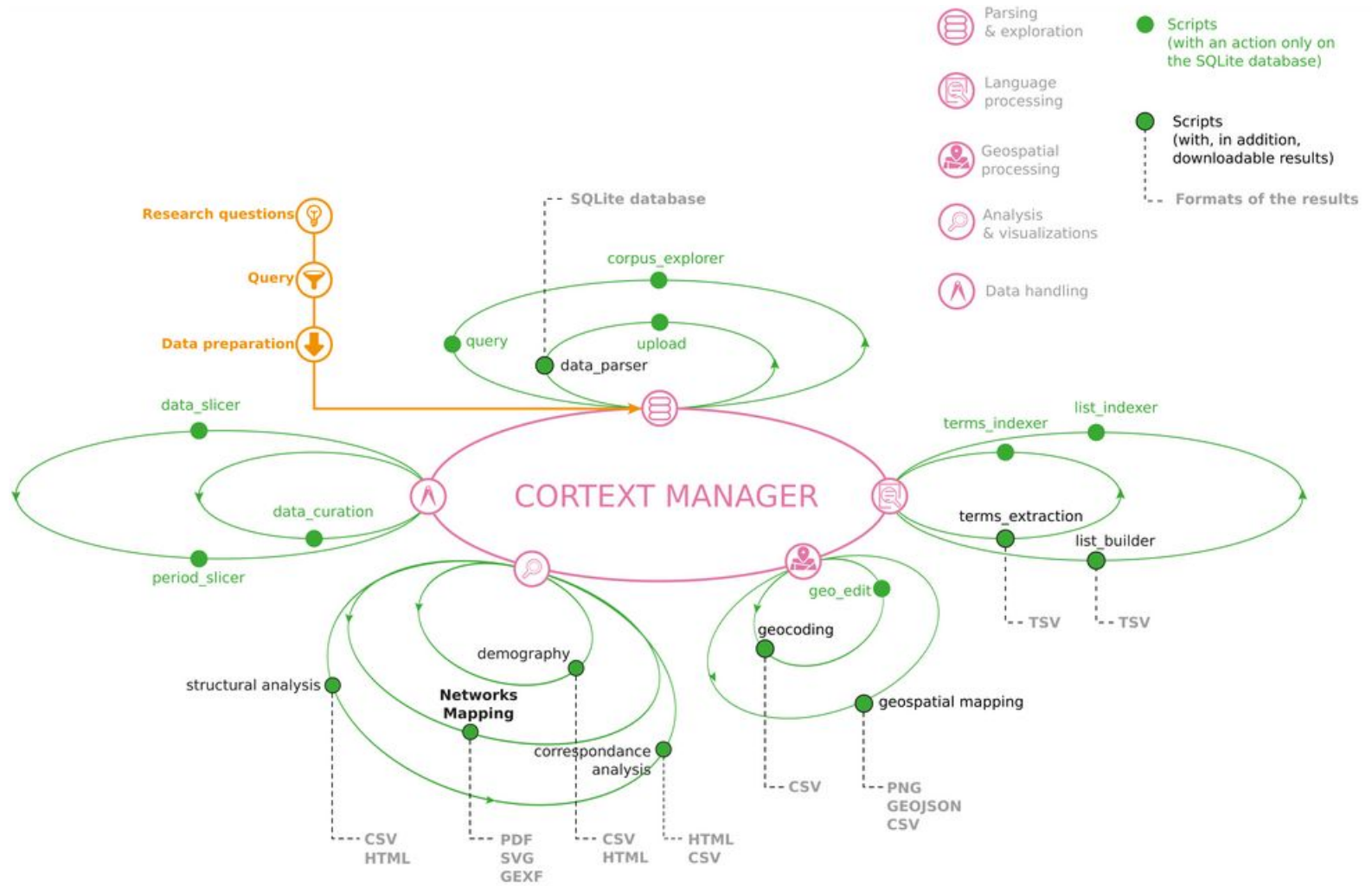
Objectifs

- Comprendre qui publie des travaux en utilisant CorText manager ?
- Quel est l'usage effectué de la plateforme ?
- Quelle est l'étendue des possibles de la plateforme ?
- Comment CorText manager a impacté la recherche ?
- Un exemple concret et « classique » d'analyse pouvant être effectuée

Méthode

- ❑ Collecte des documents mentionnant « CorText Manager », « CorText.net », « CorText plateforme » ou « CorText platform » sur :
 - ✓ Un ensemble de sources de recherche bibliographique (HAL, WOS, Scopus, Science Direct)
 - ✓ Certains sites de conférences spécialisées (STI, EASST, EUSPRI)
 - ✓ Google scholar et google

Schéma des méthodes de CorText manager



Workflow d'analyse

Intégration et parsing des données

Data parsing



Caractérisation du corpus

Demography, List builder



Réseau de collaboration des auteurs

Network mapping



Extraction des termes « importants » des titres et résumés

Term extraction



Extraction des lieux mentionnés

Name entity recognizer



Retravail des termes extraits et classification en catégorie (sujets, lieux, méthodes, données)



Indexation du corpus avec les termes retenus

Term indexer



Réseau de co-occurrence des termes de la catégorie sujets + association avec les termes lieux

Network mapping



Réseau de co-occurrence hétérogène des termes des catégories données et méthodes

Network mapping

Workflow d'analyse

Intégration et parsing des données



Caractérisation du corpus



Réseau de collaboration des auteurs

-> Qui travaille avec CorText Manager ?
Avec qui ? Où ?
Evolution dans le temps des collaborations ?



Extraction des termes « importants » des titres et résumés



Extraction des lieux mentionnés



Retravail des termes extraits et classification en catégorie (sujets, lieux, méthodes, données)



Indexation du corpus avec les termes retenus



Réseau de co-occurrence des termes de la catégorie sujets + association avec les termes lieux



Réseau de co-occurrence hétérogène des termes des catégories données et méthodes

Réseau de collaboration des auteurs

Paramètres
modulables

Question
de recherche

Métriques
modulables

Dimension
temporelle

JOB NAME

Network Mapping ▶ cortex-production-2021-v4-en.db job label (optional)

SCRIPT PARAMETERS

Nodes Selection **Edges** Dynamics Network Analysis and layout

First Field Author

Second Field Author

Number of nodes 150 Define the second field from which nodes of your network will be selected

Nodes advanced settings no

SCRIPT PARAMETERS

Nodes Selection **Edges** Dynamics Network Analysis and layout

Automatically define the Proximity Measure

yes no

Proximity Measure raw Define the proximity measure you wish to use - distributional by default but prefer chi2 if heterogeneous or raw for social networks

Edges filtering advanced settings

yes no

Edges advanced settings

yes no

SCRIPT PARAMETERS

Nodes Selection Edges **Dynamics** Network Analysis and layout

Number of time slices 2

time slices distribution

regular homogeneous regular time slices will create uniform intervals of time ; if you choose homogeneous, time slices will be split in order to preserve the same number of documents at each time step

Overlapping periods

yes no

sequencing

snapshot growing

Réseau de collaboration des auteurs

Enrichissement de
la base de données

Visualisation

Métriques du
réseau

Structure de la Base de Données					Parcourir les données					Éditer les Pragmas					Exécuter le SQL				
Table : PC_Author_Author																			
	file	id	rank	parserank	data														
	Filter	Filter	Filter	Filter	Filter														
1	cortext-production-2021-v4-en	1	0		1	Morrow, D. John & Foley, Aoife M.													
2	cortext-production-2021-v4-en	2	0		1	Penteado, Bruno Elias & Fornazin, ...													
3	cortext-production-2021-v4-en	3	0		1	Cointet, Jean-Philippe & Cambrosio, ...													
4	cortext-production-2021-v4-en	4	0		1	Penteado, Bruno Elias & Fornazin, ...													
5	cortext-production-2021-v4-en	5	0		1	Cointet, Jean-Philippe & Cambrosio, ...													
6	cortext-production-2021-v4-en	6	0		1	Shen, Yuanfei & Li, Xin													
7	cortext-production-2021-v4-en	7	0		1	Liu, Lifan & Zhao, Yi													
8	cortext-production-2021-v4-en	8	0		1	Tancoigne, Elise & Macq, Hadrien													
9	cortext-production-2021-v4-en	9	0		1	Louet, Clémentine & Fabre, Frédéric													
10	cortext-production-2021-v4-en	10	0		1	Rizzo, Davide & Marraccini, Elisa													

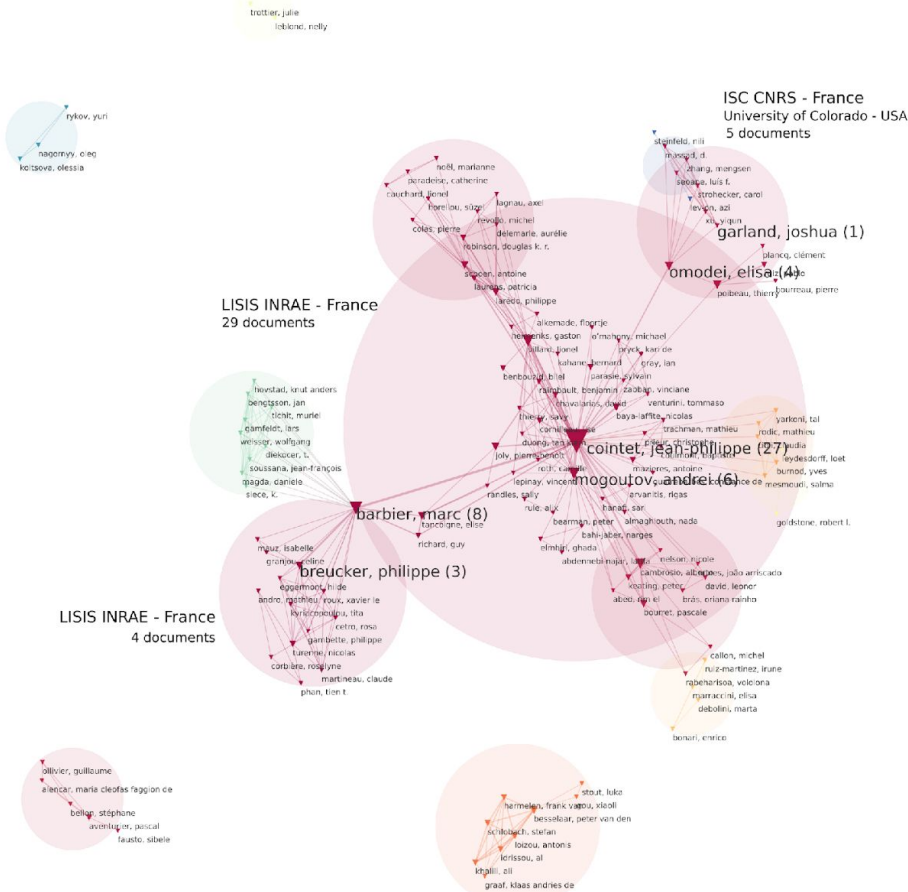
	A	B	C	D	E	F	G	H	I	J
1	id	Label	weight	in-degree	betweenness	out-degree	community	cluster	cluster_label	
2	98	Hu, Jiming	283.4	0.00157939	3.4		52	9	Xu, Xin & Hu, Jiming	
3	176	Barbier, Marc	285.2	0.00768079	5.2		18	48	Perctec, Sophie Le & Réchauchère, Olivier	
4	5	Dedieu, Benoît	202.4	0.0	2.4		55	47	Dedieu, Benoît & Schiavi, Sandra Mara de Alencar	
5	121	Réchauchère, Olivier	2015.2	0.00870947	15.2		18	48	Perctec, Sophie Le & Réchauchère, Olivier	
6	259	Cointet, Jean-Philippe	203.8	0.00120324	3.8		4	28	Cointet, Jean-Philippe & Cambrosio, Alberto	
7	261	Perctec, Sophie Le	2015.2	0.00870947	15.2		18	48	Perctec, Sophie Le & Réchauchère, Olivier	
8	311	Malanski, Priscila Duarte	202.4	0.0	2.4		55	47	Dedieu, Benoît & Schiavi, Sandra Mara de Alencar	
9	371	Rikap, Cecilia	201.0	4.98753117	1.0		11	60	Gagnon, Marc-André & Rikap, Cecilia	
10	392	Schiavi, Sandra Mara de Alencar	202.4	0.0	2.4		55	47	Dedieu, Benoît & Schiavi, Sandra Mara de Alencar	
11	17	Marraccini, Elisa	181.4	6.23441396	1.4		63	25	Rizzo, Davide & Marraccini, Elisa	
12	46	Loconto, Allison	162.2	0.00053615	2.2		5	44	Loconto, Allison & Dorin, Bruno	
13	289	Ubando, Aristotle T.	163.0	0.00023690	3.0		58	2	Ubando, Aristotle T. & Culaba, Alvin B.	
14	73	Cambrosio, Alberto	122.2	0.00010590	2.2		4	28	Cointet, Jean-Philippe & Cambrosio, Alberto	
15	91	Makowski, David	123.6	0.0	3.6		18	48	Perctec, Sophie Le & Réchauchère, Olivier	
16	160	Chen, Wei-Hsin	122.4	1.24688279	2.4		58	2	Ubando, Aristotle T. & Culaba, Alvin B.	
17	170	Akkari, Monia El	124.2	0.0	4.2		18	48	Perctec, Sophie Le & Réchauchère, Olivier	
18	173	Foley, Aoife M.	122.8	0.00054239	2.8		27	36	Morrow, D. John & Foley, Aoife M.	
19	193	Culaba, Alvin B.	122.4	1.24688279	2.4		58	2	Ubando, Aristotle T. & Culaba, Alvin B.	
20	225	Noel Mariannel, Mariann	120.4	0.0	0.4		30	54	Bordignon, Frederique & Ermakova, Liana	
21	236	Bordignon, Frederique	120.4	0.0	0.4		30	54	Bordignon, Frederique & Ermakova, Liana	
22	253	Tancoigne, Elise	120.6	2.49376559	0.6		1	32	Tancoigne, Elise & Macq, Hadrien	
23	389	Rizzo, Davide	121.2	1.24688279	1.2		63	25	Rizzo, Davide & Marraccini, Elisa	
24	10	Villard, Lionel	81.6	4.98753117	1.6		2	55	Villard, Lionel & Larédo, Philippe	
25	12	Deng, Shengli	81.0	0.0	1.0		52	9	Xu, Xin & Hu, Jiming	
26	20	Xia, Sudi	81.0	0.0	1.0		52	9	Xu, Xin & Hu, Jiming	
27	37	Xu, Xin	81.4	8.31255195	1.4		52	9	Xu, Xin & Hu, Jiming	

Documents published between 2012 and 2021 using CorText manager : authors' collaboration networks

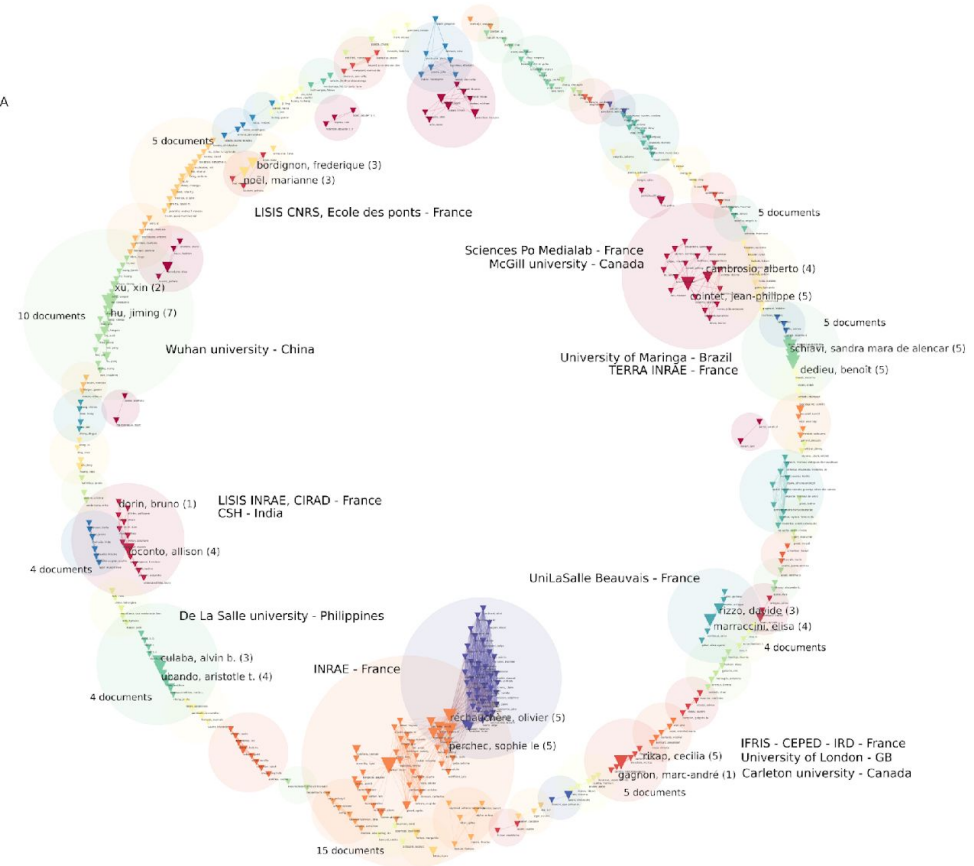
Top 2 central authors from clusters with the largest number of documents ; authors' affiliation

Node size = frequency ; Proximity measure = raw ; Community detection algorithm = Louvain resolution

2012 - 2016 period



2017 - 2021 period



Workflow d'analyse

Intégration et parsing des données



Caractérisation du corpus



Réseau de collaboration des auteurs



Extraction des termes « importants » des titres et résumés

Term extraction



Extraction des lieux mentionnés

Name entity recognizer



Retravail des termes extraits et classification en catégorie (sujets, lieux, méthodes, données)



Indexation du corpus avec les termes retenus

Term indexer



Réseau de co-occurrence des termes de la catégorie sujets + association avec les termes lieux



Réseau de co-occurrence hétérogène des termes des catégories données et méthodes

Workflow d'analyse

Intégration et parsing des données



Caractérisation du corpus



Réseau de collaboration des auteurs



Extraction des termes « importants » des titres et résumés



Extraction des lieux mentionnés



Retravail des termes extraits et classification en catégorie (sujets, lieux, méthodes, données)



Indexation du corpus avec les termes retenus



Réseau de co-occurrence des termes de la catégorie sujets + association avec les termes lieux

-> Sur quels sujets les auteurs ont-ils travaillé ?
Quels sont les terrains d'étude mentionnés dans les documents ?



Réseau de co-occurrence hétérogène des termes des catégories données et méthodes

Réseau de collaboration des auteurs

Question
de recherche

Dimension
géographique

SCRIPT PARAMETERS

Nodes Selection Edges Dynamics **Network Analysis and layout**

Community detection algorithm

louvain louvain resolution infomap clique percolation

Weighted Graph

yes no

Historical map

yes no

Project records onto clusters

yes no

Modify the name of the projected cluster

Assign a unique cluster to each record (best match)

yes no

Penalize large clusters when computing projections

yes no

Add information from a 3rd variable to tag clusters or produce a heatmap

yes no

Label clusters with information coming from other dimensions (! not compatible with historical maps)

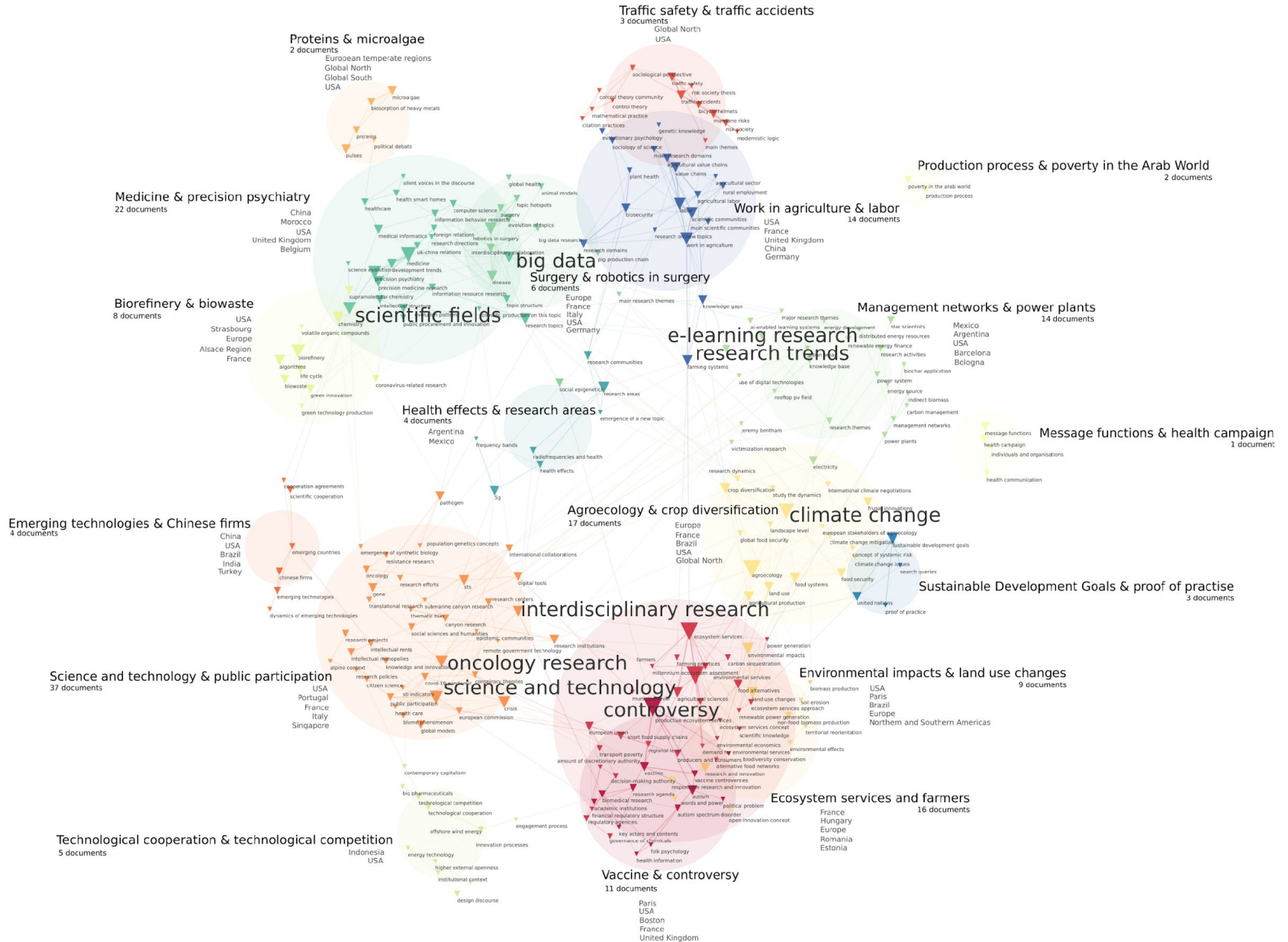
Choose the new field that should be used

Tagging/heatmap Specificity Measure

Documents published between 2012 and 2021 using CoText Manager : co-occurrences of extracted "subjects"

Top 2 central terms and top 5 place names mentioned for each cluster

Node size = frequency ; Proximity measure = raw ; Threshold = top 4 neighbours ; Community detection algorithm = Louvain resolution



Workflow d'analyse

Intégration et parsing des données



Caractérisation du corpus



Réseau de collaboration des auteurs



Extraction des termes « importants » des titres et résumés



Extraction des lieux mentionnés



Retravail des termes extraits et classification en catégorie (sujets, lieux, méthodes, données)



Indexation du corpus avec les termes retenus



Réseau de co-occurrence des termes de la catégorie sujets + association avec les termes lieux



Réseau de co-occurrence hétérogène des termes des > Quels appareils méthodologiques ont été catégories données et méthodes employés ?

Réseau hétérogène de co-occurrence des termes des catégories « données » et « méthodes »

Question
de recherche

Paramètres
modulables

JOB NAME

Network Mapping ▶ cortex-production-2021-v4-en.db job label (optional)

SCRIPT PARAMETERS

Nodes Selection **Edges** Dynamics Network Analysis and layout

First Field

Second Field

Number of nodes

Nodes advanced settings Advanced setting regarding nodes

normalisationtype

No Time dependent

Map exactly the same set of nodes at each time step

yes no

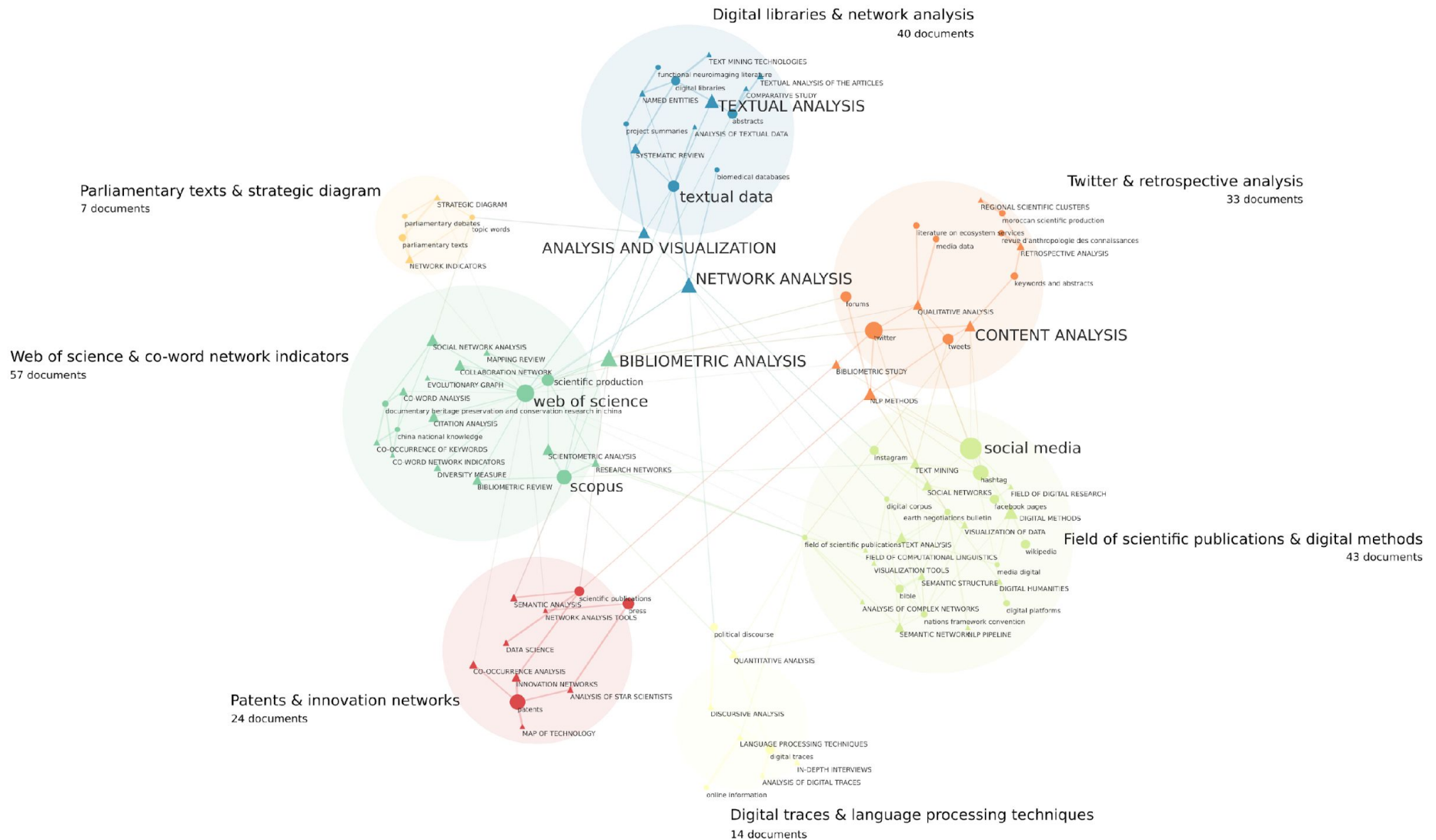
Freeze nodes positions when computing temporal maps

yes no

Documents published between 2012 and 2021 using CorText Manager : co-occurrences of extracted "data" and "methods"

Top 1 central term ("data" & "method") for each cluster

Node size = frequency ; Proximity measure = chi2 ; Threshold = top 8 neighbours ; Optimal proximity threshold = no ; Community detection algorithm = Louvain resolution ; Resolution parameter = 2 ; Assign a unique cluster to each record = no



Conclusion

Conclusion

☐ Axe de travail fréquents

- ✓ Emergence et cartographie de domaine de recherche et d'innovation (par ex. bioénergie, nanotechnologie, biodiversité)
- ✓ Controverses et « hot topics » (par ex. pesticides, biologie de synthèse, sécurité alimentaire mondiale)
- ✓ Analyse du web et des média-sociaux pour caractériser les phénomènes de la Science et de l'Innovation en Société

☐ Modes d'usage

- ✓ Analyse quantitative dans une production académique
 - ✓ cartographie socio-sémantique et relationnelle de productions, notamment celles de la recherche (par ex. publications, brevets, projets)
 - ✓ Etude du positionnement relatif et de la spécialisation d'acteurs (e.g. entreprise, université) et d'espaces géographiques (par ex. ville, région, pays)
- ✓ Appui à la démarche qualitative
 - ✓ Utiliser la carte comme un support efficace pour alimenter le travail d'enquête de terrain (par ex. dialogue avec les acteurs)
 - ✓ Utilisation de CorText Manager sur des données d'enquêtes
 - ✓ Revue de la littérature : certains auteurs utilisent CorText Manager pour circonscrire et situer leurs travaux (par ex.. parcours doctoral, dans un papier).

Conclusion

☐ Modes d'usage sans traces numériques

- ✓ Eclairage politique ou stratégique: note stratégique à destination du politique, de la conduite de la R&D, politique de la recherche, suivi de "trending topics"
- ✓ Rapport d'expertise: contexte de controverse, fabrique de la norme
- ✓ Contextes pédagogiques : traitement automatique des langues et analyse des textes, étude de controverses, étude des dynamiques sociales et politiques sur les média sociaux (par ex. tendances linguistiques, mobilisations sociales, discriminations systématiques), Social Network Analysis, veille technologique et scientifique)

Pour aller plus loin

- ❑  @CorText_team
- ❑ Site web Cortext :
<https://www.cortext.net>
- ❑ Documentation de CorText manager :
<https://docs.cortext.net>
- ❑ Accéder à Cortext Manager :
<https://managerv2.cortext.net>

Merci de votre attention