



Étude Scientométrique de la Recherche en Agriculture Numérique

Cas des financements liés à **DigitAg**



Mémoire de Master 2

Modélisation et Analyse Numérique (MANU)

Présenté par : **Rabah DJEBRA**

Stage réalisé à : **INRAE Montpellier, UMR MoISA**

Encadrant : **Jongheon KIM**

Année universitaire : **2024–2025**

Date de soutenance : **09 juillet 2025**

Dédicace

C'est avec une profonde gratitude et un grand respect que je
dédie ce travail :

À mon défunt père,
dont l'absence renforce chaque jour mon courage et ma
détermination.

À ma chère mère,
pour son amour inconditionnel, ses sacrifices silencieux
et son soutien indéfectible tout au long de mon parcours.

À mes amis, sans exception,
pour leur présence, leur bienveillance et leurs encouragements
constants.

Rabah

Remerciements

Avant tout, je rends grâce à Dieu pour m'avoir donné la force, la volonté et la persévérance nécessaires à la réalisation de ce travail.

Je remercie sincèrement mes encadrants, Monsieur Jongheom Kim, Madame Karine Gauche, Madame Véronique Bellon-Maurel ainsi que l'équipe DigitAg et mon enseignante référente Madame Hélène Mathis pour leur encadrement de qualité, leur disponibilité et leur confiance, qui ont grandement contribué à la réussite de ce stage.

Je remercie également les membres du jury pour l'honneur qu'ils me font en acceptant d'évaluer ce travail et de contribuer à son enrichissement par leurs remarques et suggestions.

Enfin, je remercie chaleureusement mes amis ainsi que toutes les personnes qui, de près ou de loin, m'ont soutenu, encouragé ou aidé dans l'élaboration de ce travail.

Rabah

Table des matières

1	Collecte et traitement des données	2
1.1	Définitions et objectifs	2
1.2	Revue des sources de données	3
1.3	Sources de données et formats - WOS & SCOPUS	3
1.4	Traitement et nettoyage des données	4
1.5	Intégration des publications HAL	5
2	Modélisation des données scientométriques	7
2.1	Objectifs de la modélisation	7
2.2	Présentation des sources et stratégie d'intégration	7
2.3	Modélisation relationnelle sous MySQL	8
2.3.1	Structuration des sources WOS & SCOPUS	8
2.3.2	Gestion des doublons inter-bases	9
2.3.3	Intégration des publications HAL	9
2.3.4	Table centrale de fusion	9
2.3.5	Vue finale pour export Cortext	9

2.4	Représentation par un graphe orienté pondéré	11
2.4.1	Définitions : Graphes orientés et pondérés	11
2.5	Construction des matrices d’adjacence	12
2.5.1	Exemples de tables d’adjacence	12
3	Analyse et interprétation des résultats	14
3.1	Objectif du chapitre	14
3.2	Évolution temporelle des publications	14
3.3	Typologie des documents et des sources	15
3.4	Analyse des financements – Focus sur DigitAg	17
3.5	Analyse géographique de la répartition des publications	19
3.6	Analyse avec Cortext Manager	21
3.6.1	Exportation des données vers Cortext	21
3.6.2	Analyse lexicale et temporelle avec Epic Epoch	21
3.6.3	Cartographie des réseaux termes–auteurs	23
3.6.4	Analyse croisée des mots-clés par pays	24
3.7	Analyse des publications issues de HAL	26
3.7.1	Contexte et objectifs de l’intégration de HAL	26
3.7.2	Typologie des publications HAL financées par DigitAg	27
3.7.3	Évolution temporelle des publications HAL financées par DigitAg	28
3.7.4	Répartition institutionnelle des publications HAL financées par DigitAg	30

3.8 Synthèse des résultats	31
--------------------------------------	----

Table des figures

2.1	Schéma relationnelle de la base de données	8
2.2	Exemple de graphe orienté et pondéré	11
3.1	Evolution des publications scientifique wos et scopus	15
3.2	Top 3 types de documents et sources de publications par années	16
3.3	Top 5 financements et le financement DigitAg	17
3.4	Répartition géographique des publications scientifiques (Power BI)	20
3.5	Visualisation des thématiques dominantes avec Epic Epoch dans Cortext Manager	22
3.6	Réseau de co-occurrence entre termes scientifiques et auteurs (Cortext Manager)	24
3.7	Corrélation entre mots-clés dominants et pays contributeurs	26
3.8	Répartition des publications HAL financées par DigitAg selon le type de document	28
3.9	Évolution annuelle des publications HAL financées par DigitAg (en valeur absolue et en pourcentage)	29
3.10	Top 10 des UMR et Top 5 institutions pour les publications HAL financées par DigitAg	30

Introduction

L'agriculture numérique s'impose aujourd'hui comme un domaine stratégique pour répondre aux défis agricoles du XXI^e siècle. Ce champ de recherche interdisciplinaire, à l'intersection de l'agronomie, de l'informatique et des sciences sociales, connaît un développement rapide à l'échelle mondiale. Dans ce contexte, la France a lancé en 2017 l'Institut Convergences DigitAg, un programme ambitieux visant à structurer la recherche en agriculture numérique autour du pôle montpelliérain.

Ce mémoire propose une évaluation scientométrique de l'impact de DigitAg sur la structuration de ce champ scientifique émergent. L'étude repose sur l'analyse systématique des publications scientifiques issues des bases Scopus, Web of Science et HAL, couvrant depuis 1960 jusqu'à 2025. Nous mobilisons des méthodes quantitatives avancées pour répondre à des questions centrales : DigitAg a-t-il favorisé l'émergence d'un réseau de recherche cohérent ? Comment a-t-il influencé les thématiques de recherche ? Quelle est sa place dans le paysage international ?

Notre approche méthodologique combine plusieurs innovations :

- Un pipeline de traitement des données utilisant Python et Power BI pour le nettoyage et la normalisation des métadonnées scientifiques.
- Une modélisation relationnelle des données dans Power BI et MySQL permettant des analyses multidimensionnelles.
- L'application de méthodes scientométriques inspirées des travaux de Callon et al. (1986) et Zupic & Čater (2015), combinant analyse de réseaux, co-mots et performance bibliographique.

Les résultats obtenus éclairent d'un jour nouveau les dynamiques de construction d'un champ scientifique interdisciplinaire. Ils montrent notamment l'émergence d'une signature montpelliéraine distinctive dans l'agriculture numérique, tout en révélant les limites des approches purement quantitatives pour appréhender ce phénomène complexe.

Cette recherche apporte une contribution originale à la compréhension des mécanismes de l'innovation scientifique, avec des implications potentielles pour les politiques de recherche. Elle démontre la pertinence des analyses scientométriques pour évaluer l'impact des investissements ciblés dans l'enseignement supérieur et la recherche.

Chapitre 1

Collecte et traitement des données

1.1 Définitions et objectifs

La scientométrie est une technique d'analyse quantitative de la production scientifique, mobilisée pour mesurer, cartographier et interpréter les dynamiques de recherche. Au-delà de simples indicateurs de performance, elle constitue un outil d'intelligence scientifique permettant de révéler les réseaux de collaboration, les évolutions thématiques ou encore l'impact institutionnel.

Comme le rappellent Callon et al. [2], la scientométrie permet de visualiser les processus de structuration d'un champ scientifique à partir d'analyses de co-mots et de co-auteurs. Elle participe à identifier les controverses, les courants dominants et les reconfigurations épistémiques.

Les principaux objectifs sont :

- **Évaluation de la recherche** : mesurer la productivité et l'impact des chercheurs et institutions via des métriques bibliométriques (publications, citations, h-index, etc.).
- **Appui aux politiques scientifiques** : éclairer les décisions de financement, recrutement ou structuration disciplinaire.
- **Cartographie des connaissances** : visualiser les réseaux thématiques, identifier les domaines émergents et repérer les zones de convergence scientifique.
- **Analyse de tendance** : suivre l'évolution temporelle de la production scienti-

fique et des collaborations.

Toutefois, une utilisation excessive ou mal contextualisée des indicateurs scientométriques peut engendrer des biais, en favorisant la quantité au détriment de la qualité ou en invisibilisant certaines productions non indexées [4].

1.2 Revue des sources de données

Pour cette étude, les données ont été extraites et croisées à partir de trois sources principales :

Web of Science (WoS) : plateforme historique de Clarivate, WoS propose des données structurées incluant les facteurs d’impact, les références croisées et les indicateurs issus du Journal Citation Reports.

Scopus : éditée par Elsevier, cette base couvre un large spectre disciplinaire, avec une forte représentation des sciences appliquées, sociales et de l’ingénierie. Elle propose des formats d’export compatibles avec les outils de traitement de données.

HAL (Hyper Articles en Ligne) : archive ouverte nationale française, HAL permet de valoriser les publications des établissements publics. Les données HAL ont été intégrées dans une seconde phase via un script Python exploitant l’API HAL, avec un ciblage spécifique sur les productions associées à DigitAg et ses partenaires institutionnels.

Conformément aux recommandations d’Echchakoui [4], nous avons entrepris une procédure de fusion WoS/Scopus/HAL en plusieurs étapes : export des métadonnées, harmonisation des champs, dédoublonnage, puis agrégation dans une base relationnelle. Ce choix permet de bénéficier de la complémentarité des bases, en combinant la couverture internationale (WoS/Scopus) et la représentativité nationale et institutionnelle (HAL). La méthodologie d’intégration HAL est détaillée dans la section 1.5.

1.3 Sources de données et formats - WOS & SCOPUS

Dans le cadre de cette étude scientométrique, les données ont été collectées en utilisant un ensemble de mots-clés représentatifs des thématiques liées à l’agriculture nu-

mérique, tels que : « Digital agriculture », « Smart farming », « Precision agriculture », « E-agriculture », « Agriculture 4.0 », « Agri-tech », « Farming 4.0 », « Agricultural innovation », « Digital farming », « Intelligent agriculture », « Climate-smart agriculture », « Data-driven agriculture », « Sustainable digital farming », « Robotic farming », « Precision breeding » et « Farm automation ». À partir des principales sources :

Web of Science (WoS) : Les données ont été extraites sous forme de fichiers texte. Un script Python a été développé pour parser ces fichiers et les convertir en fichiers CSV, facilitant ainsi leur intégration dans Power BI.

Scopus : En raison du volume important de publications disponibles depuis 1960, plusieurs fichiers CSV ont été téléchargés directement depuis la plateforme.

1.4 Traitement et nettoyage des données

Le nettoyage et la transformation des données ont été réalisés principalement à l'aide de Power Query dans Power BI et Python. Cette étape a permis de :

- **Sélection des champs utiles :** Titre, auteurs, affiliations, mots-clés, année, financement, DOI.
- **Dédoublonnage :** basé sur le titre nettoyé (minuscules, sans ponctuation) et la date. Cela permet d'identifier les doublons inter-bases comme le recommande Echchakoui [4].
- **Standardisation :** les noms d'auteurs, institutions et mots-clés sont normalisés pour éviter les ambiguïtés liées à la casse, aux abréviations ou à la langue.
- **Préparation aux analyses :** les champs de type chaîne (auteurs, affiliations, mots-clés) ont été convertis en formats tabulaires séparés pour alimenter les modèles relationnels.

Ce prétraitement garantit une cohérence des données en amont des analyses biométriques (co-citation, co-auteur, clustering thématique), comme recommandé par Zupic & Čater [7].

1.5 Intégration des publications HAL

Dans une seconde phase du travail, nous avons élargi le corpus d'analyse en intégrant les publications issues de l'archive ouverte nationale HAL, afin de compléter la sous-représentation des productions françaises constatée dans WoS et Scopus. L'objectif principal était d'identifier les publications liées à l'agriculture numérique et/ou aux partenaires du programme **DigitAg**.

a) Requête ciblée via l'API HAL

Un script Python a été développé pour interroger dynamiquement l'API HAL. Trois types de requêtes complémentaires ont été formulées :

- **Requête thématique** : recherche des publications contenant des mots-clés représentatifs de l'agriculture numérique dans le titre ou le résumé (ex. : *digital agriculture*, *smart farming*, *precision agriculture*, etc.).
- **Requête institutionnelle** : filtrage par noms d'unités ou structures partenaires de DigitAg (UMR, UR comme MOISA, TETIS, ITAP, etc.).
- **Requête financement** : détection des publications mentionnant explicitement *DigitAg* ou l'identifiant ANR-16-CONV-0004.

Ces requêtes ont été encodées en syntaxe SOLR et exécutées par lot avec pagination, en respectant un *timeout* entre les appels pour ne pas surcharger l'API.

b) Métadonnées extraites

Pour chaque publication, les métadonnées suivantes ont été extraites : titre, auteurs, affiliations, nom des structures (**structName**), résumé, journal, année, DOI, URL, éditeur, mots-clés, langue, type de document, etc.

Un champ supplémentaire **Source_HAL** a été ajouté, indiquant l'origine de l'entrée (par exemple : "AgriNumerique", "DigitAg_Funding", "DigitAg_UMR").

À partir de cette colonne, une nouvelle variable booléenne **Funding_DigitAg** a été construite. Elle indique si la publication est explicitement associée à un financement DigitAg (**true**) ou non (**false**). Cela permet, dans Power BI ou toute autre plateforme, de filtrer facilement les publications selon leur lien financier avec le programme.

c) Statistiques de récupération

Le script a permis de récupérer un total de **5 456 publications HAL** pertinentes. Parmi celles-ci, **513 publications** ont été identifiées comme ayant un **financement DigitAg**, grâce à la colonne `Funding_DigitAg`.

d) Nettoyage et enrichissement dans Power BI

Après importation dans Power BI, les données HAL ont été nettoyées à l'aide de Power Query :

- Suppression des doublons et des champs inutiles ;
- Extraction des UMR à partir de la colonne `structName` ;
- Séparation entre **unités de recherche (UMR)** et **institutions** grâce à une table de correspondance personnalisée ;
- Création d'un champ `Institution_nettoyé` prenant en compte les variantes orthographiques et les doublons d'affiliations (ex. : INRAE, Institut Agro, INRIA...).

Un traitement particulier a été appliqué à l'Institut Agro et à l'INRIA, permettant de regrouper les centres spécifiques tout en évitant les doublons (par exemple : suppression de INRIA si un centre INRIA précis est déjà mentionné).

e) Intégration dans MySQL

Les données nettoyées ont été structurées dans une base relationnelle MySQL, à l'image de WoS/Scopus. La table centrale `publication_hal` est reliée aux entités secondaires (auteurs, affiliations, mots-clés) par des jointures via des identifiants.

Cette modélisation rend possible l'exploitation croisée avec les autres bases et l'export vers Cortext.

Chapitre 2

Modélisation des données scientométriques

2.1 Objectifs de la modélisation

L’objectif de cette étape est de structurer les métadonnées bibliographiques issues des bases WoS, Scopus et HAL, afin de permettre des analyses scientométriques avancées. Il s’agit de représenter les entités scientifiques (publications, auteurs, institutions, mots-clés, financements) dans un modèle relationnel, puis de les exploiter sous forme de graphe pour identifier les structures cognitives et collaboratives.

Selon Zupic et Čater [7], les méthodes bibliométriques se répartissent en deux grandes familles :

- **Analyse de performance** : basée sur des indicateurs quantitatifs comme le nombre de publications, les citations, ou l’indice h.
- **Cartographie scientifique** : reposant sur des techniques de co-citation, co-auteurs, co-occurrence de mots, ou réseaux.

2.2 Présentation des sources et stratégie d’intégration

Trois bases ont été mobilisées :

- **WoS et Scopus** : bases internationales riches en métadonnées (DOI, affiliations, citations, mots-clés...).
- **HAL** : archive ouverte française complémentaire, utile pour repérer les publica-

tions nationales, les affiliations précises (UMR) et les financements spécifiques à DigitAg.

La stratégie d'intégration repose sur :

- La création de schémas relationnels distincts pour chaque source.
- La normalisation des entités (titres, DOI, auteurs, institutions).
- La construction d'une table centrale `WOS_SCOPUS_HAL` permettant la fusion des publications via leurs DOI et titres.

2.3 Modélisation relationnelle sous MySQL

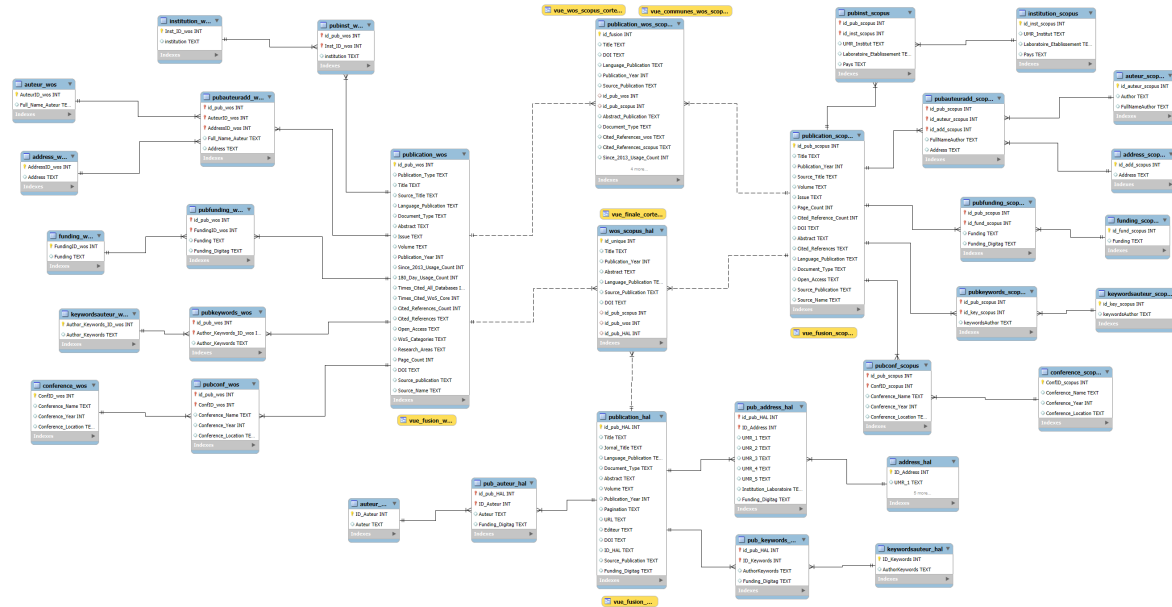


FIGURE 2.1 – Schéma relationnelle de la base de données

2.3.1 Structuration des sources WOS & SCOPUS

Chaque source (WoS, Scopus) dispose :

- d'une table **Publication**,
- de tables secondaires pour **Auteurs**, **Institutions**, **Adresses**, **Mots-clés**, **Conférences**, **Financements**,
- et de tables de jointure gérant les relations n-n entre publications et entités.

2.3.2 Gestion des doublons inter-bases

Une table de correspondance **publication_wos_scopus** lie les publications présentes dans les deux bases internationales. La fusion s'appuie sur le **DOI** ou le **Title** nettoyé. Elle est enrichie de champs comme les citations ou les domaines de recherche.

2.3.3 Intégration des publications HAL

- Création d'une table **Publication_HAL** et de ses entités associées : **Auteur_HAL**, **Address_HAL**, **Keywords_HAL**.
- Ajout d'un champ **Funding_DigitAg** à différents niveaux pour repérer les publications soutenues par ce programme.

2.3.4 Table centrale de fusion

La table **WOS_SCOPUS_HAL** centralise les publications issues des trois sources :

- Elle intègre un identifiant unique **id_fusion**.
- Elle conserve les références croisées (**id_pub_wos**, **id_pub_scopus**, **id_pub_HAL**).
- Elle permet de générer des exports standardisés à destination de Cortext Manager.

2.3.5 Vue finale pour export Cortext

Les vues **vue_fusion_wos**, **vue_fusion_scopus** et **vue_fusion_hal** alimentent **vue_finale_cortext**, à partir de laquelle un fichier CSV est généré pour l'analyse réseau.


```

CREATE OR REPLACE VIEW vue_finale_cortext AS
SELECT
    p.Title,
    COALESCE(w.Authors, s.Authors, h.Authors) AS Authors,
    COALESCE(w.Affiliations, s.Affiliations, h.Affiliations) AS
        Affiliations,
    COALESCE(w.Keywords, s.Keywords, h.Keywords) AS Keywords,
    COALESCE(w.Funding, s.Funding, '') AS Funding,
    COALESCE(w.Funding_Digitag, s.Funding_Digitag, h.Funding_Digitag)
        AS Funding_Digitag,
    p.Abstract AS Abstract,
    p.DOI,
    p.Publication_Year,
    p.Source_Publication AS Source
FROM WOS_SCOPUS_HAL p
LEFT JOIN vue_fusion_wos w ON p.id_pub_wos = w.id_pub_wos
LEFT JOIN vue_fusion_scopus s ON p.id_pub_scopus = s.id_pub_scopus
LEFT JOIN vue_fusion_hal h ON p.id_pub_HAL = h.id_pub_HAL;
UPDATE WOS_SCOPUS_HAL
SET Source_Publication =
    CASE
        WHEN id_pub_wos IS NOT NULL AND id_pub_scopus IS NOT NULL AND
            id_pub_HAL IS NOT NULL THEN 'wos&scopus&hal'
        WHEN id_pub_wos IS NOT NULL AND id_pub_scopus IS NOT NULL THEN '
            wos&scopus'
        WHEN id_pub_wos IS NOT NULL AND id_pub_HAL IS NOT NULL THEN 'wos&
            hal'
        WHEN id_pub_scopus IS NOT NULL AND id_pub_HAL IS NOT NULL THEN '
            scopus&hal'
        WHEN id_pub_wos IS NOT NULL THEN 'wos'
        WHEN id_pub_scopus IS NOT NULL THEN 'scopus'
        WHEN id_pub_HAL IS NOT NULL THEN 'hal'
        ELSE 'unknown'
    END;
SELECT * FROM vue_finale_cortext
INTO OUTFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/
    publication_cortext_wos_scopus_hal.csv'
FIELDS TERMINATED BY ',';
ENCLOSED BY '"'
LINES TERMINATED BY '\n';

```

2.4 Représentation par un graphe orienté pondéré

La base relationnelle est transposée en un graphe orienté pondéré $G = (V, E)$, selon la logique de Callon et al. [2] et des travaux de Loconto et al. [5] en socio-technologie des controverses, où :

- V est l'ensemble des sommets : *Publications, Auteurs, Institutions, Mots-clés, Financements*.
- E est l'ensemble des arêtes orientées et pondérées, représentant les liens entre entités (ex. : un auteur est lié à une publication).

2.4.1 Définitions : Graphes orientés et pondérés

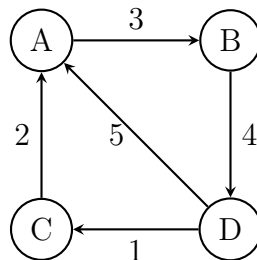
Graphe orienté : [1, 3].

- Un graphe est dit *orienté* si ses arêtes, appelées **arcs**, possèdent un **sens de parcours**.
- Un **chemin** est une succession d'arcs mis bout à bout.
- Un **circuit** est un chemin fermé dont les arcs sont tous distincts.

Graphe pondéré : [1, 3].

- Un graphe est dit *étiqueté* si ses arêtes ou arcs portent des **étiquettes** (mots, lettres, symboles, nombres, etc.).
- Lorsque ces étiquettes sont des **nombres**, le graphe est dit **pondéré**, et les étiquettes sont appelées **poids**.
- Le **poids d'une chaîne** (ou d'un chemin) est la **somme des poids** des arcs qui le composent.

FIGURE 2.2 – Exemple de graphe orienté et pondéré



Lecture : Le graphe ci-dessus est orienté (les arcs ont un sens) et pondéré (chaque arc

est affecté d'un poids numérique). Par exemple, un chemin de A vers D via B a un poids total de $3 + 4 = 7$.

2.5 Construction des matrices d'adjacence

À partir des relations dans la base, on construit des matrices d'adjacence binaires $A \in \{0, 1\}^{n \times m}$ [6] :

- n : nombre de publications.
- m : nombre d'entités associées (auteurs, mots-clés, institutions, etc.).

$$A_{i,j} = \begin{cases} 1 & \text{si } i\mathcal{R}j \\ 0 & \text{sinon} \end{cases}$$

Remarque : $i\mathcal{R}j$: Le noeud i en relation avec le noeud j .

2.5.1 Exemples de tables d'adjacence

Matrice Publication–Auteur–Adresse

Publication	Auteur A	Auteur B	Auteur C	Adresse 1	Adresse 2
Pub1	1	1	0	1	0
Pub2	0	1	1	1	1
Pub3	1	1	1	1	1

Lecture : Pub1 est écrite par les auteurs A et B, affiliés à l'adresse 1 ; Pub2 est écrite par les auteurs B et C, affiliés aux adresses 1 et 2 ; Pub3 est écrite par les auteurs A, B et C, affiliés aux adresses 1 et 2.

Matrice Publication–Mot-clé

Publication	Agri-tech	Digital agriculture	Data-driven agriculture
Pub1	1	1	0
Pub2	1	0	1

Lecture : Pub1 utilise les mots-clés Agri-tech et Digital agriculture. Pub2 utilise Agri-tech et Data-driven agriculture.

Matrice Publication–Financement

Publication	#DigitAg	CNRS	UM
Pub1	1	1	0
Pub2	0	1	1

Lecture : Pub1 a été financée par #DigitAg et le CNRS ; Pub2 par le CNRS et l’Université de Montpellier.

Ces matrices sont les points de départ pour l’analyse réseau avec Gephi ou VOSviewer, et pour le calcul de mesures de centralité, de modularité ou de densité Newman [6].

L’approche en graphe permet de croiser les dimensions sociales (collaborations) et cognitives (thématiques), rejoignant l’analyse de controverses développée par Loconto et al. [5] dans leur étude sur les débats agroécologiques.

Conclusion

Ce chapitre a présenté l’architecture relationnelle de la base de données scientométriques, intégrant WoS, Scopus et HAL. Grâce à une structuration rigoureuse et une fusion cohérente des métadonnées, cette base permet des analyses comparées, tant sur les dynamiques de publication que sur les structures de financement, notamment celles liées à DigitAg.

Chapitre 3

Analyse et interprétation des résultats

3.1 Objectif du chapitre

Ce chapitre présente les principaux résultats issus de l’analyse scientométrique réalisée à l’aide de Power BI, MySQL, des scripts Python et Cortext manager. Il s’appuie sur un corpus de 150668 publications extraites, nettoyées et fusionnées à partir des bases de données Scopus, Web of Science et HAL, selon la méthodologie décrite précédemment. L’analyse s’intéresse à l’évolution temporelle des publications, à la typologie documentaire (Scopus, WoS et HAL), aux sources de financement, aux institutions affiliées (INRAE, CIRAD, etc.), et au rôle structurant de DigitAg.

3.2 Évolution temporelle des publications

L’analyse temporelle de la production scientifique constitue une étape clé pour comprendre l’émergence progressive d’un champ de recherche. À l’aide d’un graphique d’aire cumulée, la progression annuelle du volume de publications liées à l’agriculture numérique a été modélisée depuis les années 1960.

Les résultats montrent trois phases distinctes :

- **Phase 1 — Expansion lente (1960–2000)** : La production reste relativement marginale durant cette période, avec moins de 500 publications par an dans le monde. Le champ est alors embryonnaire, centré sur des aspects techniques

spécifiques (capteurs, automatisation de tracteurs).

- **Phase 2 — Croissance progressive (2000–2015) :** La période est marquée par l'introduction des technologies d'information et de communication (TIC) dans l'agriculture, ainsi que par les premiers projets structurants autour de la « precision agriculture ». On observe un doublement quasi tous les 5 ans du volume de publications.
- **Phase 3 — Accélération et explosion (2015–2024) :** Dès 2015, la courbe devient exponentielle. Cette explosion est corrélée à l'émergence des concepts d'« agriculture numérique », à la généralisation des capteurs IoT, à l'intelligence artificielle appliquée à la gestion agricole, et au lancement de programmes tels que DigitAg en France en 2017. Le pic atteint en 2024 (plus de 37 000 publications) reflète la cristallisation du champ au niveau international.

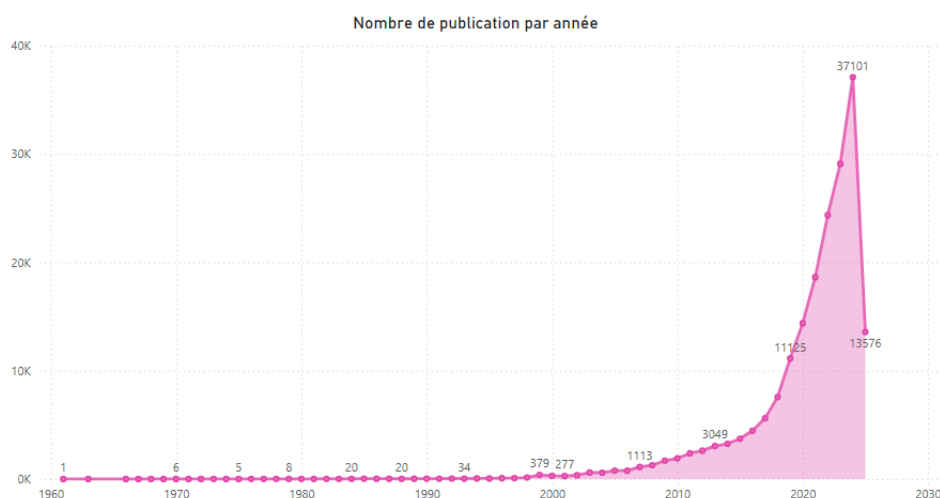


FIGURE 3.1 – Evolution des publications scientifique wos et scopus

Cette dynamique confirme l'hypothèse selon laquelle l'agriculture numérique est devenue un domaine stratégique d'innovation, suscitant l'intérêt croissant des chercheurs, des institutions et des financeurs publics.

3.3 Typologie des documents et des sources

Une typologie documentaire a été réalisée afin de mieux cerner les modalités de diffusion scientifique dans le domaine. À partir d'un graphique combiné (barres empilées par type de document et courbe représentant la source), trois types de publications sont apparus comme dominants :

- **Articles scientifiques (Articles)** : Représentant environ 80 % du corpus, ils constituent le principal canal de diffusion académique. Ce mode de publication est favorisé dans les bases croisées WoS/Scopus et concerne essentiellement des revues à comité de lecture.
- **Communications en conférences (Conference Papers)** : Plus présentes dans Scopus, elles traduisent une orientation technologique, notamment dans les disciplines de l'ingénierie, de la robotique et des technologies de l'information. Elles sont courantes dans les publications issues de projets européens ou asiatiques.
- **Chapitres d'ouvrage (Book Chapters)** : Bien que minoritaires, ils occupent une place significative dans certaines disciplines, notamment les sciences sociales ou les projets collaboratifs transversaux. Ils sont souvent présents dans Scopus, mais rarement indexés dans WoS.

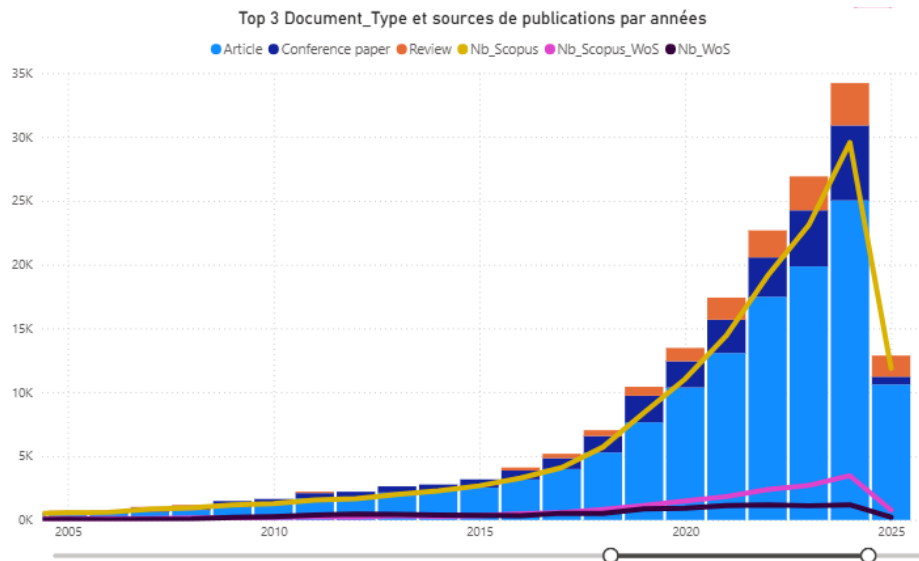


FIGURE 3.2 – Top 3 types de documents et sources de publications par années

Par ailleurs, on note une part croissante de publications figurant simultanément dans Scopus et WoS, signe d'une meilleure visibilité et d'une reconnaissance scientifique accrue. Cette convergence améliore la fiabilité des mesures bibliométriques et traduit une standardisation progressive des pratiques de publication dans le champ.

3.4 Analyse des financements – Focus sur DigitAg

Afin d’analyser les principales sources de financement associées aux publications en agriculture numérique, une visualisation en anneau (donut chart) a été conçue à partir des métadonnées issues des bases bibliographiques. Ce graphique repose sur une mesure DAX personnalisée permettant d’extraire dynamiquement les cinq financeurs les plus fréquemment mentionnés, à laquelle a été ajoutée manuellement la mention explicite du programme DigitAg [ANR-16-CONV-0004], afin de garantir sa visibilité malgré une fréquence d’apparition plus faible.

Cette méthode permet de contourner une limitation observée dans Power BI, où les parts visuelles de chaque catégorie apparaissent équivalentes, indépendamment de leur fréquence réelle. Pour remédier à cela, une mesure intermédiaire a été introduite pour recalculer précisément le nombre de publications par financeur, uniquement pour les entités sélectionnées, assurant ainsi une représentation proportionnelle plus fidèle à la réalité.

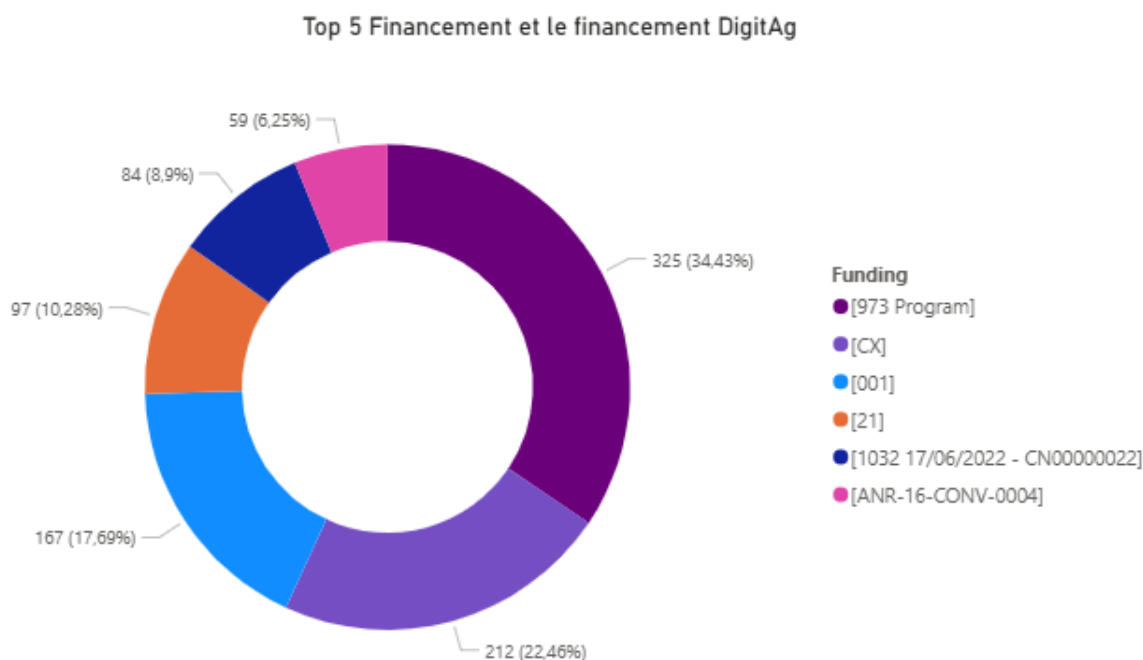


FIGURE 3.3 – Top 5 financements et le financement DigitAg

Il convient toutefois de souligner certaines limites méthodologiques concernant les données analysées. Tout d’abord, cette étude repose exclusivement sur les métadonnées extraites des bases **Scopus** et **Web of Science (WoS)**. Les publications présentes dans la base **HAL** n’ont pas encore été intégrées en raison de contraintes techniques

et temporelles. Cette absence peut partiellement expliquer la sous-représentation des organismes nationaux de recherche (ONR) français tels que l'INRAE, le CIRAD ou l'IRD, ainsi que de l'Agence nationale de la recherche (ANR), principal financeur institutionnel français. Ces acteurs sont souvent mieux référencés dans les archives ouvertes nationales.

Par ailleurs, la qualité de l'extraction automatique des mentions de financement présente des limites. Ces informations sont généralement indiquées entre crochets [] ou parenthèses () dans les champs textuels, ce qui permet leur détection via des expressions régulières. Cependant, certains cas particuliers, notamment dans des publications chinoises, comportent des parenthèses imbriquées (ex. : **(Supported by the National Program (973) Grant No. ...)**), ce qui complique fortement l'extraction automatique. Malgré l'utilisation de scripts de nettoyage adaptés, ces structures peuvent entraîner des erreurs de segmentation, voire l'omission de certaines mentions. Ces incertitudes doivent être prises en compte dans l'interprétation des résultats présentés.

Les résultats observés à ce stade de l'étude montrent :

- Le programme chinois [973 Program] apparaît comme le financeur principal, avec **325 publications**, soit **34,43 %** du total analysé. Ce chiffre reste toutefois susceptible d'être sous-estimé ou surestimé en raison des difficultés d'extraction évoquées.
- Le programme [CX] suit avec **212 publications (22,46 %)**, puis [001] avec **167 publications (17,69 %)**.
- Les financements [21] (**97 publications**, soit **10,28 %**) et [1032 17/06/2022 - CN0000022] (**84 publications**, **8,9 %**) complètent le Top 5.
- Enfin, le programme DigitAg ([ANR-16-CONV-0004]) est associé à **59 publications**, soit **6,25 %** du total, ce qui justifie son intégration ciblée dans le visuel.

Bien que numériquement inférieur aux grands financeurs internationaux, DigitAg joue un rôle stratégique dans le paysage français. Son inclusion volontaire dans l'analyse vise à dépasser une lecture strictement quantitative, en soulignant sa fonction structurante dans le développement de l'agriculture numérique.

De plus, des analyses préliminaires montrent que DigitAg est fréquemment associé à d'autres financeurs comme l'INRAE, le CNRS ou des agences nationales. Cette co-occurrence reflète un positionnement transversal du programme, orienté vers le soutien à des projets collaboratifs multi-acteurs. Une exploration croisée avec les affiliations

des auteurs permettrait de mieux caractériser ces dynamiques, ce qui fera l’objet d’un approfondissement dans la section suivante.

Par ailleurs, il convient de souligner certaines limites méthodologiques concernant les données analysées jusqu’à présent. L’essentiel de cette analyse repose exclusivement sur les métadonnées extraites des bases **Scopus** et **Web of Science (WoS)**. Or, de nombreuses publications françaises — notamment issues de l’INRAE, du CIRAD, de l’IRD ou de l’Institut Agro — sont davantage représentées dans la base nationale **HAL**.

C’est pourquoi l’intégration récente des données issues de HAL constitue une avancée majeure : elle permet de compléter l’analyse des financements par une approche institutionnelle et territoriale, en exploitant des champs détaillés tels que les affiliations, les UMR (Unités Mixtes de Recherche) et les structures partenaires. Cette nouvelle source permet de corriger partiellement la sous-représentation des acteurs français dans les bases commerciales, et d’enrichir la compréhension des dynamiques de financement dans un cadre national.

La section suivante revient en détail sur cette dimension, à travers des visualisations spécifiques portant sur les institutions, les UMR et les publications financées par DigitAg référencées dans HAL.

3.5 Analyse géographique de la répartition des publications

Pour compléter les analyses thématiques, relationnelles et institutionnelles, une visualisation géographique a été réalisée à l’aide de Power BI. Cette carte permet de représenter les 30 pays ayant le plus contribué à la production scientifique en agriculture numérique, en se basant sur le nombre de publications issues du corpus fusionné WoS/Scopus.

Comme le montre la figure 3.4, la production scientifique se concentre très fortement dans un petit nombre de pays, illustrant les déséquilibres mondiaux en matière de recherche sur l’agriculture numérique.



FIGURE 3.4 – Répartition géographique des publications scientifiques (Power BI)

- **La Chine** domine largement le classement avec plus de 32 000 publications, ce qui témoigne d'un effort massif d'investissement dans la recherche agricole et technologique.
- **L'Inde** suit avec plus de 22 000 publications, traduisant une forte dynamique de recherche sur des technologies adaptées aux systèmes agricoles tropicaux.
- **Les États-Unis** occupent la troisième place avec plus de 18 000 publications, se distinguant par une approche multidisciplinaire (durabilité, IA, politiques agricoles).
- Des pays européens comme **l'Allemagne**, **l'Italie** et **l'Australie** dépassent chacun les 7 000 publications.
- **Le Brésil**, **l'Espagne** et le **Royaume-Uni** affichent chacun un volume supérieur à 5 000 publications.
- **La France**, bien qu'en retrait, contribue avec environ 4 000 publications, notamment à travers des structures comme INRAE, CIRAD ou l'Institut Agro Montpellier (anciennement Montpellier SupAgro).

Cette visualisation renforce les analyses lexicales précédentes en révélant la dimension géographique de la production scientifique. Elle permet de mieux comprendre la distribution des efforts de recherche à l'échelle mondiale et de situer la France dans ce paysage globalisé.

Elle constitue également un outil pertinent pour identifier les pôles régionaux de spécialisation, les zones de collaboration scientifique potentielle, ainsi que les marges d'amélioration en matière de visibilité scientifique pour certains pays.

3.6 Analyse avec Cortext Manager

Dans cette dernière phase, nous avons utilisé la plateforme **Cortext Manager** pour analyser les dynamiques lexicales et thématiques du corpus bibliographique fusionné. Cette analyse s'inscrit dans une démarche de cartographie scientifique, complémentaire aux visualisations temporelles et quantitatives précédemment présentées.

3.6.1 Exportation des données vers Cortext

Une fois les métadonnées nettoyées, harmonisées et fusionnées dans MySQL Workbench (cf. section 2.2), une vue consolidée appelée `vue_finale_cortext` a été générée pour agréger, pour chaque publication, les champs clés suivants :

— `Title, Authors, Affiliations, Keywords, Funding, Abstract, DOI, Publication_Year, Source.`

Cette vue a été exportée au format CSV via la commande SQL (cf. section 2.2)

Le fichier produit respecte les exigences de Cortext Manager, notamment pour l'import du champ multi-valeurs ("Keywords", "Authors", "Affiliations", etc.) séparé par "****".

3.6.2 Analyse lexicale et temporelle avec Epic Epoch

L'analyse dans Cortext a porté sur la dynamique des mots-clés, des sources et des domaines thématiques sur une période longue (1960–2025), en combinant des algorithmes d'extraction lexicale et de mise en réseau. En particulier, le module **Epic Epoch** a été utilisé pour visualiser l'évolution temporelle des concepts dominants.

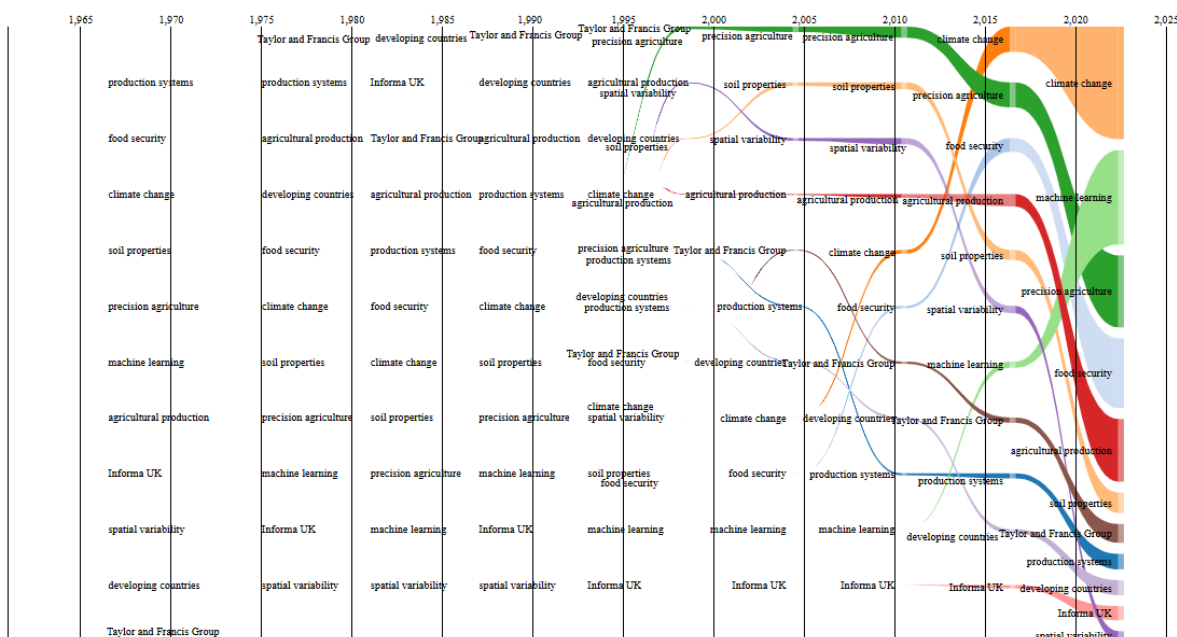


FIGURE 3.5 – Visualisation des thématiques dominantes avec Epic Epoch dans Cortex Manager

La figure 3.5 représente les principales thématiques scientifiques (mots-clés) identifiées dans les publications. Chaque colonne correspond à une période (époque), et chaque nœud à un concept fréquent dans les textes. Les flux entre les colonnes matérialisent la continuité ou l'évolution des thématiques dans le temps.

Les résultats les plus significatifs sont :

- **Precision agriculture**, **Soil properties** et **Spatial variability** sont présents depuis les années 1990 et restent des axes forts.
- **Climate change** et **Food security** s'imposent massivement après 2010, reflétant des préoccupations globales.
- **Machine learning** émerge après 2015, signe d'une intégration croissante de l'intelligence artificielle dans le domaine agricole.

Ces trajectoires thématiques révèlent une hybridation croissante du champ, où se croisent enjeux techniques, environnementaux et sociaux. L'outil Epic Epoch permet ainsi de visualiser non seulement les thématiques dominantes à chaque époque, mais aussi les continuités conceptuelles et les bifurcations stratégiques.

3.6.3 Cartographie des réseaux termes–auteurs

Dans le prolongement des analyses lexicales et temporelles, une cartographie de réseau croisant les termes scientifiques et les auteurs les plus représentés a été réalisée avec **Cortext Manager**. Cette visualisation repose sur un algorithme de co-occurrence basé sur le *Chi2* et une mesure de proximité entre les entités textuelles (auteurs et termes).

La figure 3.6 met en évidence plusieurs structures significatives dans la production scientifique autour de l’agriculture numérique. On y observe :

- Des **clusters thématiques** centrés sur des notions telles que *precision agriculture*, *machine learning*, *soil moisture*, *climate change*, ou encore *IoT*.
- Des **groupes d’auteurs fortement connectés** à certains mots-clés, révélant une spécialisation : par exemple, les auteurs ZHAO CHUNJIANG, ZHANG YU, WANG WEI ou encore LI DAOLIANG sont fortement liés aux thématiques liées à l’intelligence artificielle, aux capteurs et aux systèmes d’aide à la décision.
- Une **distribution géographique implicite** : de nombreux auteurs du réseau sont affiliés à des institutions chinoises, ce qui reflète la forte contribution de la Chine à ces thématiques.

La construction de ce graphe permet de visualiser les interdépendances entre acteurs et thématiques, et d’identifier les figures structurantes du champ. Les co-occurrences entre auteurs et mots-clés révèlent non seulement des proximités scientifiques, mais aussi des stratégies de positionnement, des écoles thématiques et des partenariats internationaux implicites.

Ce type de visualisation offre un point d’entrée pertinent pour de futures analyses de réseaux de co-auteurs, ou pour suivre l’évolution de groupes de recherche autour de concepts émergents.

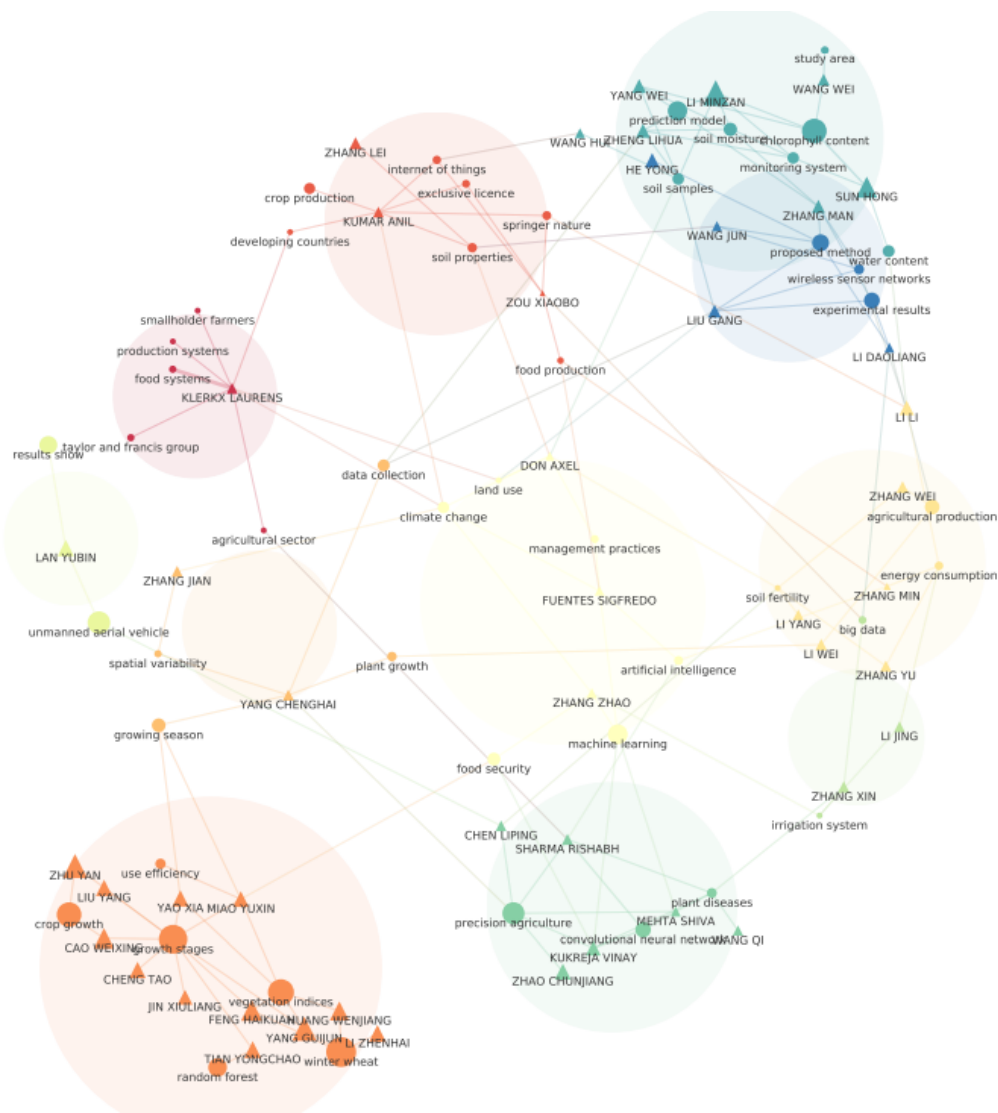


FIGURE 3.6 – Réseau de co-occurrence entre termes scientifiques et auteurs (Cortext Manager)

3.6.4 Analyse croisée des mots-clés par pays

En complément de l’analyse diachronique menée à l’aide du module Epic Epoch, une analyse croisée entre les pays contributeurs et les mots-clés dominants a été réalisée. Pour cela, une matrice de type *heatmap* a été générée, permettant de visualiser les variations d’occurrence des concepts en fonction des principaux pays producteurs de publications en agriculture numérique. La figure 3.7 présente :

- En ordonnée : une sélection des 15 pays les plus représentés dans le corpus.
- En abscisse : les mots-clés les plus fréquents, standardisés pour éviter les doublons sémantiques.

- En couleur : une échelle de chaleur (**red-blue**) indiquant la fréquence relative par rapport à la moyenne globale. Le rouge indique une surreprésentation, le bleu une sous-représentation.

Cette matrice met en évidence des spécialisations géographiques marquées :

- **La Chine** apparaît fortement liée à des concepts d'intelligence artificielle appliquée à l'agriculture, avec une nette surreprésentation des termes **deep learning** et **smart agriculture**.
- **Les États-Unis** se démarquent par une orientation marquée vers les enjeux de durabilité et d'adaptation climatique, via des mots-clés comme **climate change**, **food security** et **sustainability**.
- **L'Inde** survalorise les concepts d'agriculture connectée, notamment **IoT** et **smart farming**, mais est en retrait sur les thématiques IA ou durabilité.
- **Le Brésil, l'Espagne et l'Italie** présentent des occurrences spécifiques sur **precision agriculture**, **remote sensing** ou **soil moisture**, illustrant une orientation plus territoriale ou expérimentale.
- **La France** et les pays d'Europe occidentale présentent des profils équilibrés, souvent en lien avec les thématiques transversales (**agriculture**, **remote sensing**, **food security**).

Cette visualisation conforte les résultats de l'analyse temporelle et relationnelle, en y ajoutant une dimension géopolitique. Elle permet de mettre en lumière la manière dont chaque pays s'approprie des concepts spécifiques, reflétant à la fois des priorités scientifiques nationales, des stratégies de financement, et des contextes agroécologiques différenciés.

Par exemple, la matrice met en évidence que le changement climatique et la durabilité émergent clairement comme des thématiques majeures dans la recherche française, en cohérence avec les priorités affichées par des établissements tels que l'INRAE et le CIRAD, notamment en matière de biodiversité et de transition agroécologique.

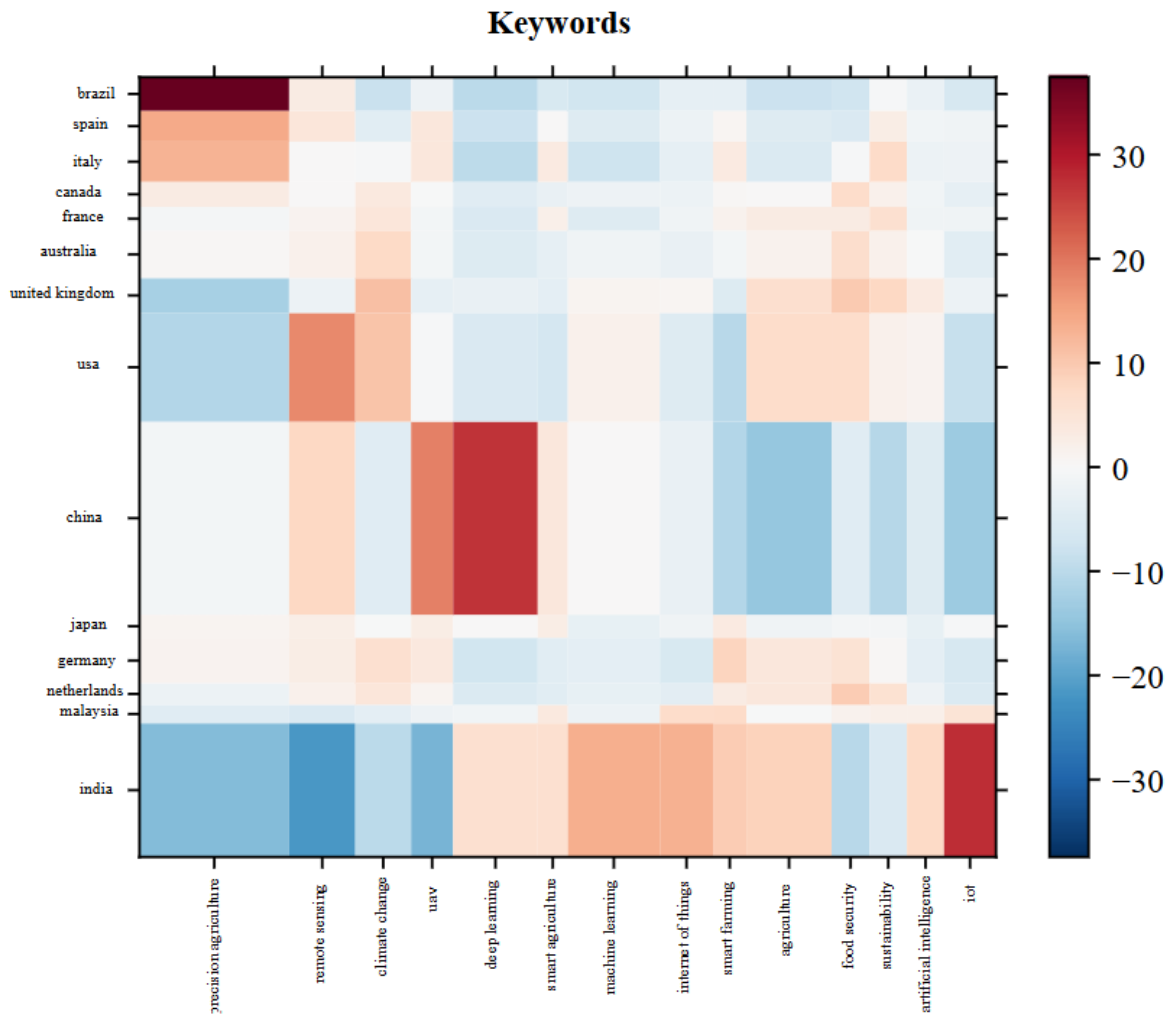


FIGURE 3.7 – Corrélation entre mots-clés dominants et pays contributeurs

3.7 Analyse des publications issues de HAL

3.7.1 Contexte et objectifs de l'intégration de HAL

L'intégration des données issues de la plateforme HAL (Hyper Articles en Ligne) constitue une étape essentielle pour enrichir notre corpus scientométrique, initialement fondé sur les bases Web of Science (WoS) et Scopus.

Cette démarche répond à plusieurs objectifs :

- **Compléter les bases commerciales** : certaines publications, notamment francophones ou issues d'institutions publiques françaises, ne sont pas référencées dans

WoS/Scopus.

- **Renforcer la représentativité nationale** : HAL valorise la production des organismes français tels que l'INRAE, le CIRAD, l'IRD ou les universités, souvent sous-représentés dans les bases internationales.
- **Identifier des publications liées à DigitAg** : l'exploitation des métadonnées HAL permet de repérer des publications mentionnant explicitement un financement du programme DigitAg.

Au total, **5456 publications** ont été extraites depuis HAL et intégrées à la base relationnelle modélisée dans MySQL. Parmi celles-ci, **513 publications** sont étiquetées comme étant financées par DigitAg, soit **environ 9,40 %** du corpus HAL analysé.

L'interface Power BI a ensuite permis de visualiser cette sous-partie spécifique, à travers une série de graphiques exploratoires centrés sur les institutions contributrices, les types de documents et la temporalité des publications.

3.7.2 Typologie des publications HAL financées par DigitAg

L'analyse des types de documents présents dans les publications HAL financées par DigitAg permet de mieux comprendre la stratégie de diffusion scientifique soutenue par le programme. À partir du sous-corpus de 513 publications identifiées comme financées, la répartition par Document_Type met en évidence les profils suivants :

- **Articles scientifiques** : 200 publications, soit **41,9 %** du total — confirmant la priorité accordée à la production académique de qualité, valorisée via HAL ;
- **Communications scientifiques** : 159 publications (**33,3 %**) — illustrant un fort engagement dans la dissémination lors de colloques ou congrès ;
- **Thèses** : 75 publications (**15,7 %**) — soulignant un investissement important dans la **formation par la recherche**, caractéristique des programmes structurants comme DigitAg ;
- **Posters** : 25 publications (**5,2 %**) — traduisant des activités de communication intermédiaire ou pédagogique, souvent issues de projets collaboratifs ;
- **Autres types** : 25 publications également (**5,2 %**) — comprenant rapports techniques, mémoires ou documents non conventionnels.

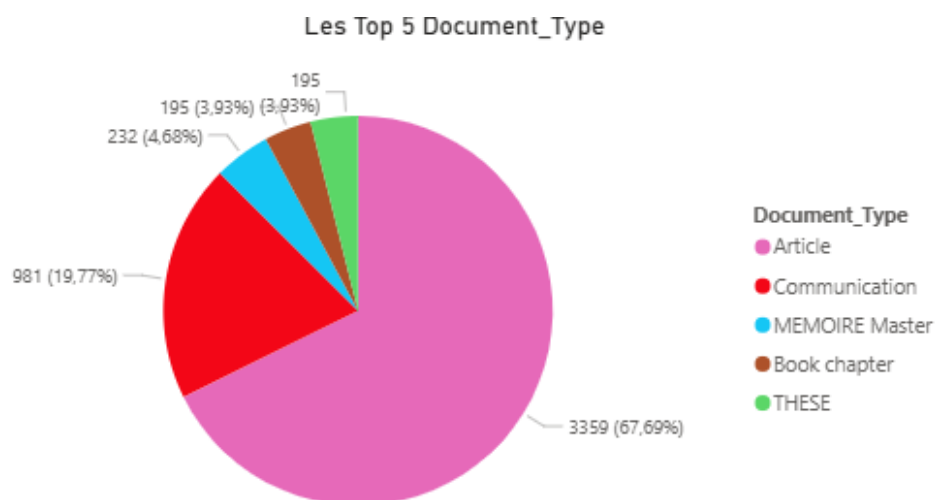


FIGURE 3.8 – Répartition des publications HAL financées par DigitAg selon le type de document

Cette diversité documentaire montre que DigitAg ne soutient pas uniquement la production d’articles, mais favorise également la communication scientifique et la formation doctorale. Par comparaison avec WoS/Scopus, où la typologie est plus restreinte, HAL permet de mieux capter les formes variées de production issues de la recherche française, en particulier dans les établissements partenaires du programme.

3.7.3 Évolution temporelle des publications HAL financées par DigitAg

L’analyse diachronique des publications financées par DigitAg dans la base HAL révèle une dynamique croissante jusqu’en 2022, suivie d’un léger repli, puis d’un regain en 2025. Deux indicateurs complémentaires ont été mobilisés pour cette analyse :

- Le **nombre absolu** de publications par année (en rouge),
- Le **taux de financement relatif** par rapport à l’ensemble des publications HAL de chaque année (en bleu).

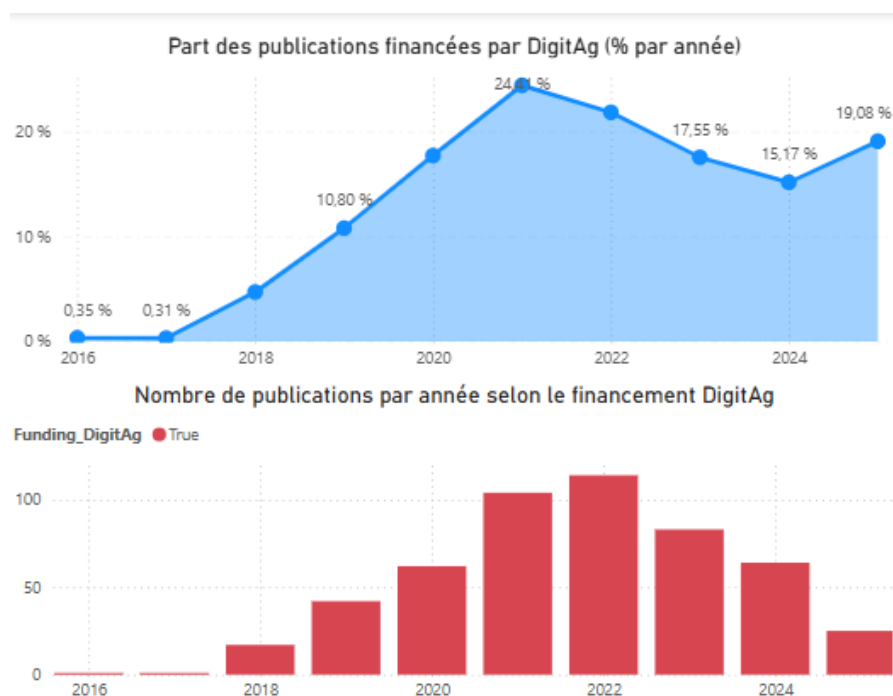


FIGURE 3.9 – Évolution annuelle des publications HAL financées par DigitAg (en valeur absolue et en pourcentage)

Les résultats montrent une montée en puissance du financement DigitAg jusqu'en 2022 :

- En **2016–2018**, les publications financées restent marginales ($<1\%$) ;
- Entre **2019 et 2022**, le taux de financement progresse fortement, atteignant un pic en 2021–2022 (jusqu'à **24 %** des publications HAL de l'année) ;
- En **2023**, une baisse relative est observée (15,17%), probablement liée à un décalage de dépôt ou à une variation du financement ;
- En **2025**, on constate un rebond à **19,08 %**, ce qui pourrait refléter un recentrage du programme ou un meilleur référencement dans HAL.

Cette double visualisation permet de croiser les dynamiques quantitatives et relatives, mettant en lumière la trajectoire d'influence croissante du programme **DigitAg** dans la production scientifique déposée dans HAL. L'année 2022 constitue un point culminant, en volume comme en part relative, ce qui suggère un effet de structuration maximal à cette période.

3.7.4 Répartition institutionnelle des publications HAL financées par DigitAg

L'analyse institutionnelle des publications HAL bénéficiant d'un financement de DigitAg permet d'identifier les principaux laboratoires et établissements impliqués dans le programme. Deux niveaux d'agrégation ont été mobilisés : les unités mixtes de recherche (UMR) et les institutions de rattachement.

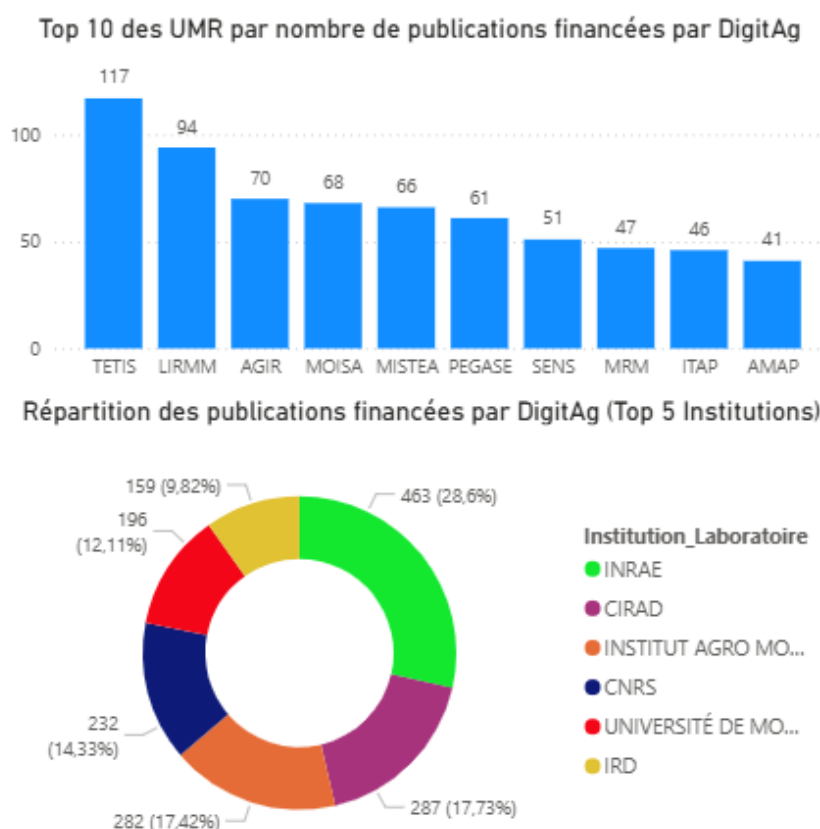


FIGURE 3.10 – Top 10 des UMR et Top 5 institutions pour les publications HAL financées par DigitAg

UMR les plus contributrices : Le classement des 10 principales UMR montre une forte concentration de la production financée :

- **TETIS** arrive en tête avec **117 publications**, confirmant son rôle central dans les thématiques de géomatique, télédétection et systèmes d'information appliqués à l'agriculture.
- Elle est suivie par le **LIRMM (94 publications)**, spécialisé en robotique et

intelligence artificielle.

- Les UMR **AGIR**, **MOISA**, **MISTEA**, **PEGASE**, **SENS**, **MRM**, **ITAP** et **AMAP** apparaissent également comme des contributeurs réguliers.
- Cette diversité illustre l’ancrage interdisciplinaire du programme DigitAg, à l’interface entre sciences agronomiques, sociales et numériques.

Principales institutions partenaires : L’analyse par établissements met en évidence une forte représentation d’acteurs académiques français :

- **INRAE** domine avec **463 publications**, soit près de **29 %** du total.
- **CIRAD** et l’**Institut Agro Montpellier** suivent, avec respectivement **287** et **282 publications**.
- Le **CNRS**, l’**Université de Montpellier** et l’**IRD** complètent ce top 5, tous partenaires historiques du LabEx.

Ces résultats confirment que le financement DigitAg est fortement structurant pour les institutions partenaires du consortium, en cohérence avec son objectif de fédérer la recherche française en agriculture numérique. L’intégration des données HAL permet ici de mieux capturer cette dimension institutionnelle, souvent absente des bases bibliographiques commerciales.

3.8 Synthèse des résultats

L’étude scientométrique présentée dans ce mémoire repose sur un corpus consolidé de **150668 publications** issues de trois sources complémentaires : Web of Science, Scopus et HAL. Cette intégration a permis d’élargir le périmètre d’analyse, en rendant visible une partie de la production scientifique francophone et institutionnelle initialement absente des bases commerciales.

Grâce à une chaîne de traitement combinant modélisation relationnelle (MySQL), visualisation interactive (Power BI), et analyse lexicale/réseau (Cortext Manager), plusieurs résultats majeurs ont pu être dégagés :

- Une **croissance exponentielle** de la production scientifique en agriculture numérique depuis 2015, avec un pic de publications observé en 2024.

- Une **typologie documentaire dominée par les articles scientifiques**, mais avec une diversité plus marquée dans HAL (thèses, posters, communications non indexées).
- Une **géographie de la recherche mondiale** fortement polarisée : Chine, Inde et États-Unis en tête, tandis que la France reste visible mais relativement sous-représentée dans WoS/Scopus.
- Des **thématiques structurantes** identifiées grâce à l’analyse textuelle : *precision agriculture, IoT, climate change, remote sensing, machine learning*.
- Une meilleure caractérisation du programme **DigitAg**, avec :
 - 513 publications HAL financées identifiées, contre 59 dans WoS/Scopus,
 - Un **taux de financement HAL estimé à 10,63 %**, contre moins de 1 % dans les bases commerciales,
 - Une représentation claire des institutions et UMR bénéficiaires (TETIS, LIRMM, INRAE, CIRAD, Institut Agro Montpellier).

L’intégration des publications issues de HAL a permis de combler un biais important et de fournir une vue plus complète, notamment sur la production scientifique nationale, les unités de recherche, et la formation par la recherche (via les thèses). La méthodologie adoptée offre ainsi une lecture multi-niveaux — temporelle, thématique, institutionnelle, géographique — du champ de l’agriculture numérique.

Conclusion générale

Ce mémoire a présenté une démarche complète d’analyse scientométrique du champ de l’agriculture numérique, avec un focus sur le programme **DigitAg**. L’approche repose sur l’exploitation conjointe des bases Scopus, WoS et HAL, enrichies, fusionnées et modélisées dans une architecture relationnelle (MySQL), puis analysées par Power BI et Cortext Manager.

Trois volets méthodologiques ont structuré cette étude :

- **Un traitement rigoureux des données** : extraction, nettoyage, désambiguïsation, normalisation, fusion, modélisation relationnelle, export vers outils d’analyse.
- **Une analyse multidimensionnelle** : dynamique temporelle, typologie des publications, financement, affiliation institutionnelle, géographie scientifique, co-termes et co-auteurs.

- **Une mise en lumière des logiques scientifiques et institutionnelles** : structuration des thématiques dominantes, réseaux d’auteurs, poids différencié des institutions et des pays.

L’étude a permis de formuler plusieurs constats clés :

- Le champ **agriculture numérique** connaît une dynamique de croissance rapide, portée par l’essor des technologies numériques et des enjeux de transition agroécologique.
- La **France** est présente mais sous-visibilisée dans les bases bibliographiques commerciales, d’où l’importance d’inclure des sources comme HAL pour une meilleure représentativité.
- Le programme **DigitAg**, s’il reste peu visible dans WoS/Scopus (59 publications), est bien mieux représenté dans HAL (513 publications), notamment via des thèses, posters, et communications.
- **Les UMR comme TETIS, LIRMM, MISTEA, AGIR** et les institutions comme **INRAE, CIRAD, Institut Agro Montpellier** sont les principaux contributeurs au sein de la recherche financée par DigitAg.

Cette étude constitue une base solide pour :

1. Concevoir des tableaux de bord de suivi de la production scientifique à l’échelle d’un programme ou d’un institut,
2. Aider à la valorisation des productions non indexées dans les bases commerciales,
3. Approfondir l’analyse des effets institutionnels et thématiques du financement, en croisant affiliations, mots-clés, et collaborations.

Les prochaines étapes pourront porter sur :

- L’intégration plus fine des co-auteurs, UMR et financements dans les analyses réseau,
- Le suivi diachronique de la structuration d’un « écosystème DigitAg »,
- L’élargissement à d’autres corpus comme PubMed, AgEcon ou OpenAlex pour capter des niches disciplinaires spécifiques.

En définitive, cette recherche s’inscrit dans une logique d’outillage stratégique pour les institutions scientifiques, en mobilisant des méthodes de veille, d’intelligence scientifique et de visualisation exploratoire, au service d’une meilleure compréhension des dynamiques de production et de financement dans un champ en pleine structuration.

Bibliographie

- [1] Claude Berge. *Théorie des graphes et ses applications*. Dunod, 1981.
- [2] Michel Callon, Jean-Pierre Courtial, and Hervé Penan. La scientométrie : une introduction à la mesure de la science. *La Découverte*, 1986.
- [3] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction à l'algorithmique*. Pearson, 3ème édition edition, 2009.
- [4] Sami Echchakoui. Why and how to merge scopus and web of science during bibliometric analysis : The case of salesforce literature. *Journal of Marketing Analytics*, 8 :165–184, 2020.
- [5] Allison Loconto, Marion Desquilbet, Théo Moreau, Denis Couvet, and Bruno Dorin. The land sparing–land sharing controversy : Tracing the politics of knowledge. *Land Use Policy*, 96 :103610, 2020.
- [6] M. E. J. Newman. *Networks : An Introduction*. Oxford University Press, 2010.
- [7] Ivan Zupic and Tomaž Čater. Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3) :429–472, 2015.