



HAL
open science

Analyse de la performance d'une recherche à partir d'un dépôt institutionnel.

Abdou Beukeu Sow

► **To cite this version:**

Abdou Beukeu Sow. Analyse de la performance d'une recherche à partir d'un dépôt institutionnel.. Colloque, Mar 2022, Saint Louis, Sénégal. hal-03616288

HAL Id: hal-03616288

<https://hal.science/hal-03616288>

Submitted on 2 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse de la performance d'une recherche à partir d'un dépôt institutionnel.

Abdou Beukeu SOW

Université de Lille

Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication
(GERiiCO)

Résumé

Cette étude s'est proposé d'analyser deux questions très peu associées l'une à l'autre dans un contexte africain : les archives ouvertes et l'évaluation de la recherche. Elle montre de fait comment évaluer une recherche à partir d'un corpus issu d'un dépôt institutionnel. La méthodologie utilisée à cet effet repose sur les techniques de visualisation des données pour analyser les performances de la recherche dans une institution, d'évaluer ses chercheurs, de développer des partenariats stratégiques, d'identifier et d'analyser les nouvelles tendances de recherche émergentes et de créer des rapports sur mesure. La démarche repose sur trois étapes : le paramétrage de l'outil Dspace, l'extraction des données, et en fin l'analyse par les techniques de visualisation de l'information. Les méthodes utilisées ici sont aussi rigoureuses que celles utilisées par Elsevier avec son outil SciValⁱ pour visualiser la recherche dans le monde.

Les résultats de l'étude ont montré que les dépôts institutionnels peuvent être un outil pour le suivi de la qualité de la recherche et de l'enseignement pour les pays en voie de développement. La plateforme utilisée révèle que la fiabilité et la qualité de ce suivi ne peuvent en effet être espérées que si chaque institution de recherche agrège de manière « exhaustive » toute sa production scientifique dans un dépôt.

Mots-clés : Valorisation de la recherche, archive ouverte, dépôt institutionnel, évaluation de la recherche

Introduction

Les universités africaines sont de plus en plus convaincues de la place des dépôts institutionnels dans la recherche scientifique et l'enseignement supérieur. Les moyens pour la mise en œuvre effective des dépôts semblent cependant inexistantes dans la plupart des pays comme ceux de l'espace Cames (Conseil africain et malgache pour l'enseignement supérieur). Le déploiement et la gestion d'un dépôt institutionnel exigent en effet des ressources humaines, techniques, financières. Cela n'est pas pour autant la chose la plus fondamentale. Les réflexions et les actions menées autour des dépôts institutionnels en Afrique sont axées sur la mobilisation de ces ressources, sur la mutualisation de l'information scientifique et son accès gratuit. Gora Lo (Lo, 2016), met l'accent sur l'accès gratuit des publications des universitaires à travers des plateformes de dépôts numériques. D'autres chercheurs comme Dieyi Diouf (Diouf, 2016) se focalisent sur la nécessité d'identifier et de sauvegarder le potentiel documentaire des institutions de recherche.

La recension sur cette question au Sénégal prouve que l'identification et la mise à la disponibilité de l'Information scientifique et technique (IST) sont la préoccupation majeure des documentalistes et des partisans des archives ouvertes. À un niveau institutionnel plus important, le Cames, pour sa part, s'est chargé de faciliter le travail à ses pays membres en mettant un dépôt commun : le DICAMES¹. Il compte ainsi à travers cette plateforme disposer d'une archive ouverte qui aura comme fonction autre d'évaluer la recherche dans les pays de son espace (pays de l'espace Cames : PEC). Pour cela le Cames compte sur les ressources humaines existantes dans les différentes universités.

L'évaluation de la recherche est en effet un pan très peu pris en compte dans les réflexions. Les communications les plus récentes s'y intéressent peu. Le mouvement du libre accès en Afrique, ses opportunités et défis sont les thématiques les plus abordées par la communauté scientifique.

Évaluer la recherche à partir d'un dépôt institutionnel est capital pour les pays en développement absents des plus grandes bases de données d'indexation. Cette absence est la conséquence logique de leur mauvais rang dans les classements internationaux. L'usage des bases de renom pour analyser la recherche n'est pas en soi l'aspect le plus déterminant, mais le choix du corpus à partir duquel cette recherche est effectuée.

¹ Archive numérique institutionnelle du Conseil africain et malgache pour l'enseignement supérieur (CAMES).

Au-delà de données statistiques fournies par les outils de dépôt, nous nous proposons de montrer comment cartographier une recherche à partir d'un dépôt réalisé avec l'outil Dspace². Les méthodes utilisées ici sont les mêmes que celles utilisées par SciVal³ pour visualiser les performances de la recherche dans une institution, d'évaluer ses chercheurs, de développer des partenariats stratégiques, d'identifier et d'analyser les nouvelles tendances de recherche émergentes et de créer des rapports sur mesure.

Méthodologie

Un projet de cartographie de la recherche passe par une phase amont où sont fixées les grandes orientations de l'analyse. Celles-ci permettent en effet de déterminer la méthodologie. Ici, nous analyserons deux aspects : les interactions entre les directeurs de thèse ou rapporteurs et doctorants dans un premier temps et dans un second les sujets ou domaines de recherche. Ces deux aspects seront considérés comme notre projet d'analyse bibliométrique. Notre démarche repose sur trois étapes : le paramétrage de l'outil Dspace, l'extraction des données, et en fin l'analyse par les techniques de visualisation de l'information.

Le paramétrage de Dspace

Le paramétrage de Dspace s'effectue principalement au niveau des métadonnées. Dspace intègre la sémantique du *Dublin Core* employée par de nombreuses organisations pour l'application du protocole d'interopérabilité OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*). Les métadonnées par défaut ne permettent pas de recueillir toutes les informations nécessaires à notre projet d'analyse bibliométrique. Ici, nous avons en effet besoin d'identifier les auteurs (doctorants), les directeurs de thèse et les rapporteurs. Nous avons ainsi intégré les métadonnées « encadrant » pour les mémoires, « Directeur de thèse », « rapporteur » pour les thèses et « mémoire master » pour définir un nouveau type de document. Ces deux actions sont exécutées sur le fichier « *input-forms.xml* » du répertoire *dspace/config/*.

² Dspace est un logiciel libre servant à la construction d'archives ouvertes et de bibliothèques électroniques. Il est fréquemment utilisé par les Universités ou les Organismes de recherche pour stocker des collections de publications scientifiques.

³ SciVal est un outil développé par Elsevier. Il donne accès aux performances de recherche de plus de 20 700 instituts de recherche et de leurs chercheurs associés, de 234 pays à travers le monde. Il permet de trouver des collaborateurs pour stimuler des solutions innovantes à des problèmes complexes.

```

<dc-element>contributor</dc-element>
<dc-qualifier>encadrant</dc-qualifier>
<repeatable>true</repeatable>
<label>Encadrant</label>
<input-type>name</input-type>
<hint>Enseignant encadrant memoire</hint>
<required></required>
</field>
<field>
<dc-schema>dc</dc-schema>
<dc-element>contributor</dc-element>
<dc-qualifier>directeur</dc-qualifier>
<repeatable>true</repeatable>
<label>Directeur de these</label>
<input-type>name</input-type>
<hint>Directeur de these</hint>
<required></required>
</field>
<field>
<dc-schema>dc</dc-schema>
<dc-element>contributor</dc-element>
<dc-qualifier>rapporteur</dc-qualifier>
<repeatable>true</repeatable>
<label>Rapporteur</label>
<input-type>name</input-type>
<hint>Rapporteur</hint>
<required></required>
</field>

<field>
<dc-schema>dc</dc-schema>

```

Figure 1 : Formatage fichier input-forms.xml

```

<pair>
<displayed-value>Memoire de master</displayed-value>
<stored-value>Memoire de master</stored-value>
</pair>

```

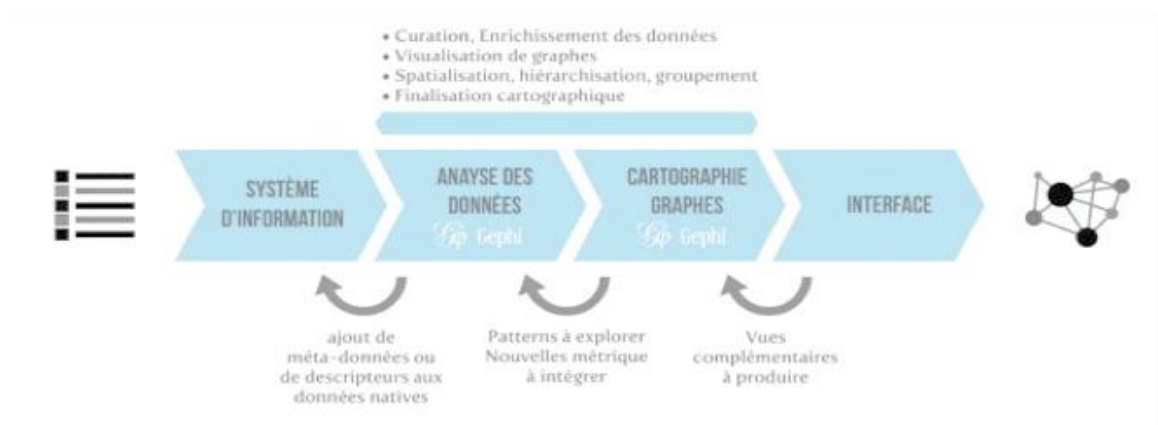
Figure 2 : Ajout des documents de type mémoire de master

Une autre action doit par ailleurs être réalisée pour l'analyse de sujet ou domaine de recherche. Cette action peut se faire de deux manières. La première consiste à définir et fixer au préalable les différents domaines de recherche. Cela se fait à partir du fichier « *input-forms.xml* » du répertoire *dspace-release/config/*. Cette méthode écarte toute possibilité d'erreur.

Les actions qu'il convient de mener dans le but de fournir des indicateurs dépendent des objectifs d'analyses.

L'extraction des données

La plateforme Dspace à partir de laquelle le corpus est extrait est structurée en plusieurs collections, dont celle des thèses. Cette dernière est composée de 39 thèses que nous avons extraites sous format CSV. Le fichier exporté se présente en plusieurs colonnes correspondant au *Dublin Core element set*⁴. Nous avons procédé au nettoyage en vérifiant l'encodage et les séparateurs. Le fichier est ensuite importé dans Gephi. Ainsi, après l'exportation des données brutes, la plupart des interventions seront faites avec le logiciel Gephi. Celui-ci présente des fonctionnalités qui permettent de manipuler les données et d'observer les graphes sur différentes vues.



5

Figure 3 : Chaîne de production des graphes (source)

La cartographie

La troisième et dernière phase est le traitement du fichier importé. D'importantes manipulations sont faites en s'appuyant sur certaines fonctionnalités du logiciel Gephi. Utilisé par beaucoup d'institutions scientifiques en France comme aux États-Unis, Gephi permet de réaliser des représentations de corpus d'informations bien structurées. Il présente des fonctionnalités qui offrent la possibilité d'observer les données sur différents plans, de réaliser des filtres en fonction des éléments statistiques ou bien encore d'identifier des clusters et de faire des focus (vue à grande échelle) sur eux. Grâce à ses multiples fonctionnalités, Gephi offre la possibilité d'analyser des données en regroupant les éléments observés en « clusters » (thématique, période, groupe de recherche). Parmi les fonctionnalités les plus utilisées pour la production de graphes : La fonction spatialisation et la fonction partition.

⁴ Composé des quinze propriétés de base relatives au contenu (titre, sujet, description, source, langue, relation, couverture), à la propriété intellectuelle (créateur, contributeur, éditeur, gestion des droits) et à l'instanciation (date, type, format, identifiant de la ressource).

⁵⁵ Étude cartographique: le "véhicule autonome". Clichés cartographiques autour de 481 notices bibliographiques consacrées à la recherche française dans le domaine du véhicule autonome.

La fonction spatialisation

La fonction spatialisation permet d'associer des coordonnées spatiales aux nœuds d'un graphe en vue de produire une visualisation. Elle s'effectue via un « algorithme de spatialisation », généralement un « force vector » appliquant des principes physiques d'attraction et/ou de répulsion des nœuds en fonction de leurs liens. Ces algorithmes de spatialisation sont basés sur les forces d'attraction et de répulsion. Nous avons utilisé l'algorithme « Force Atlas 2 » pour une meilleure qualité des interprétations.

La fonction partition de Gephi

La fonction partition offre la possibilité d'analyser séparément les données par mots-clés, auteurs, institutions, références, etc. Cette fonction nous permettra de faire un filtre sur les années afin de voir la structure chronologique des éléments à analyser. Nous pourrons aussi manipuler des structures, des formes et des couleurs pour révéler des propriétés cachées.

Éléments statistiques

En plus de la visualisation des graphes, Gephi fournit de manière automatique des valeurs statistiques sur les éléments à analyser. Certains éléments comme la « centralité » renseignent sur l'importance des acteurs au sein d'un réseau. La centralité de proximité est liée à plusieurs éléments : l'accès à l'information, le pouvoir, au prestige, l'influence, le statut social, etc.

Résultats et discussion

Résultats

La figure 4 représente le réseau des doctorants et directeurs de thèse en jaune. La taille des nœuds des acteurs est proportionnelle au nombre de thèses encadrées pour les enseignants. L'épaisseur des liens est proportionnelle au nombre de fois qu'un enseignant encadre un doctorant. Le graphe révèle que certains enseignants ont encadré plus de thèses que d'autres. Nous pouvons en effet constater que les nœuds représentant les enseignants « *Sock, Oumar* » ; « *Gaye, Amadou Thierno* » ; « *Azelino, Dorothé* » ; « *Fabrice, Bernard* » ont une taille plus importante que les autres nœuds jaunes. Cela signifie qu'ils ont encadré

plus de thèses que les autres présents sur la carte. Chaque doctorant ayant soutenu une seule thèse, la taille des nœuds gris devrait être la même. L'exception constatée pour les acteurs « Cisse, Mady » et « Thiaw, Lamine » s'explique par le fait qu'ils sont devenus directeurs de thèses et ont été doctorants sous la direction d'un directeur présent sur la carte.

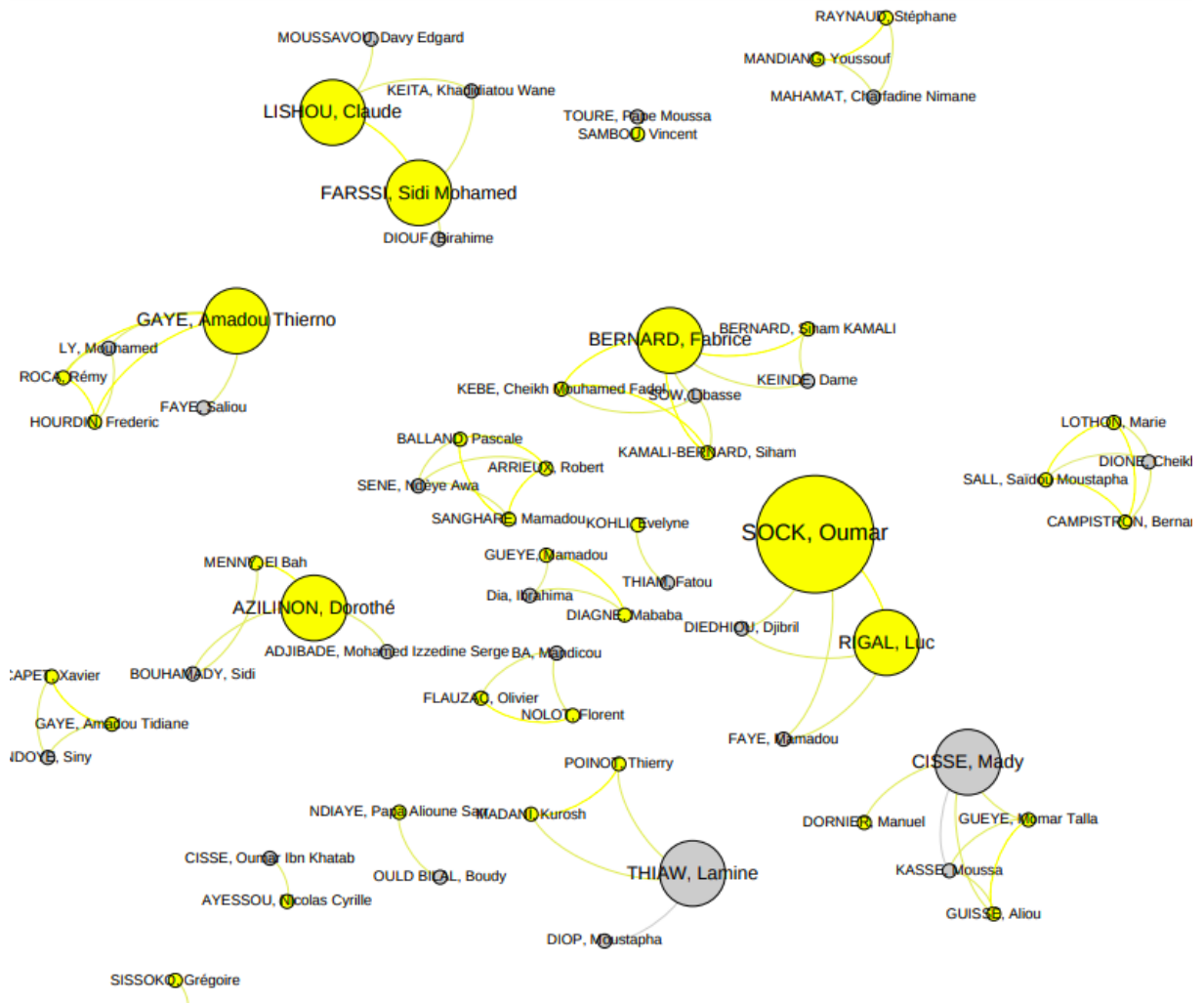


Figure 4 : Réseaux doctorants, Directeur de thèse

La figure 5 fait apparaître trois clusters correspondant aux trois domaines dans lesquels les thèses ont été soutenues (Physique chimie biologie, Génie électrique, génie informatique).

L'absence d'un domaine sur la carte signifie, soit par le fait qu'aucune thèse de ce domaine n'a été enregistrée dans la base, soit par le fait que le domaine n'est pas couvert par le dépôt institutionnel. La différence de taille des clusters et de label (nom des nœuds) prouve une forte présence des thèses dans le domaine « Physique, chimie et biologie » suivi de « Génie informatique ».

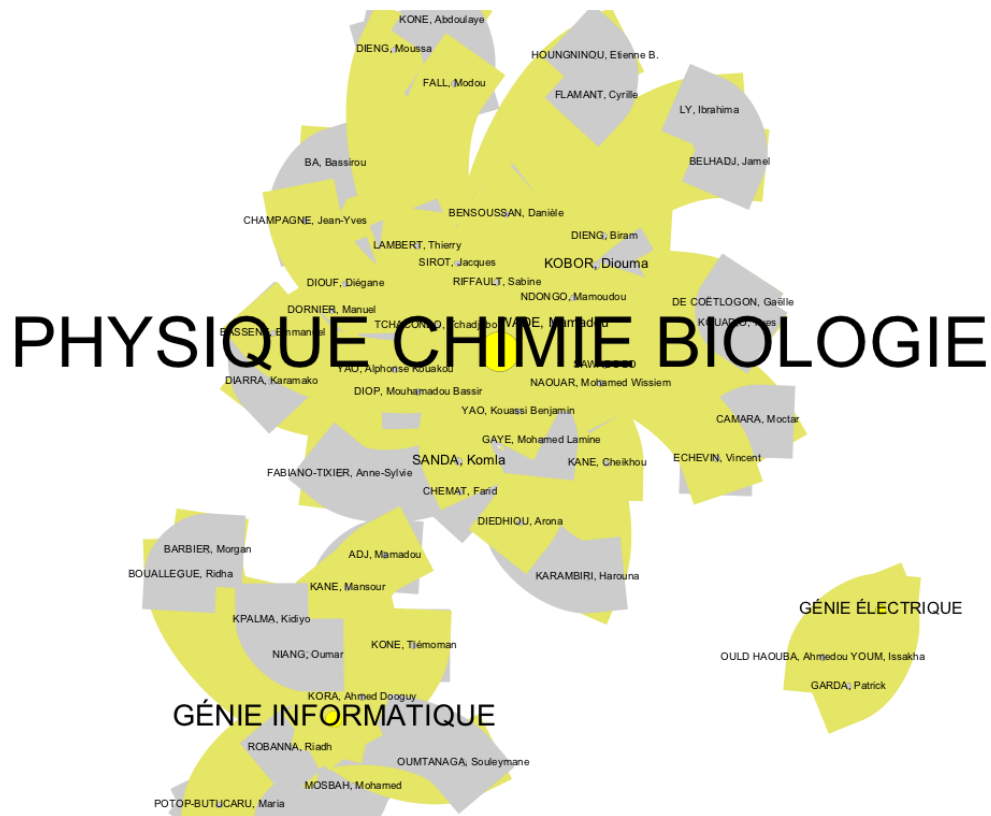


Figure 5 : Formation de trois clusters (Sujet)

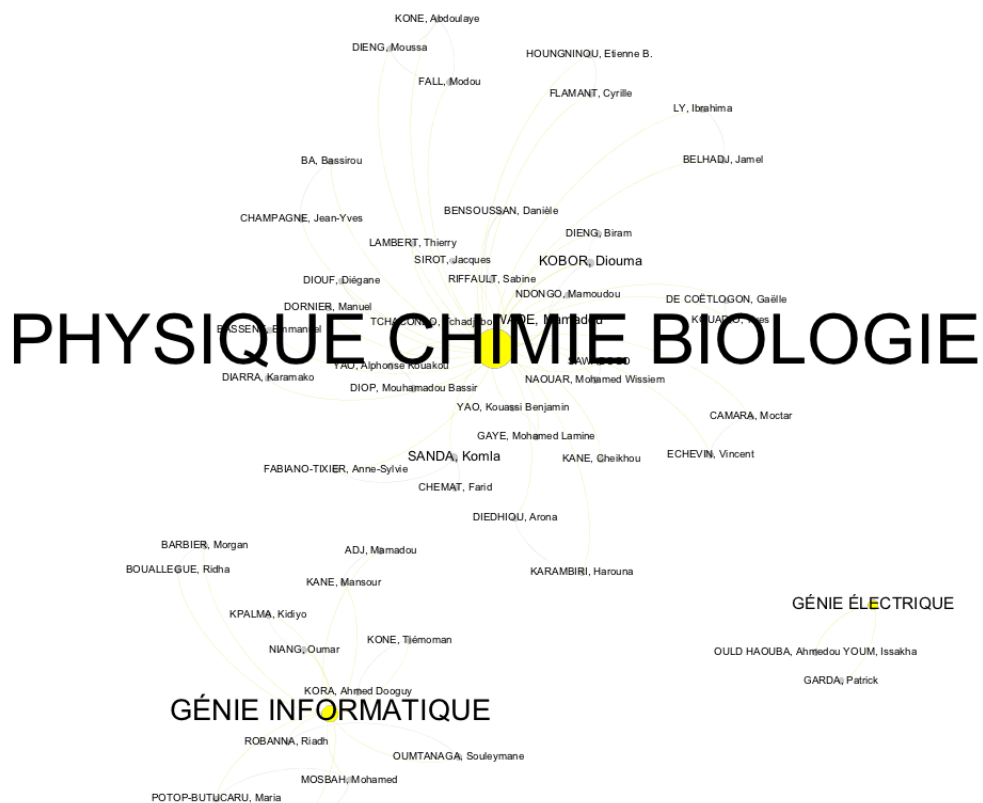


Figure 6 : Réseau auteurs, domaines de recherche

Discussion

La figure 4 montre les interactions qui peuvent avoir lieu entre doctorants et directeurs de thèses ou entre rapporteurs. Elle peut ainsi avec un corpus plus représentatif fournir des indicateurs sur la performance des chercheurs et surtout sur la qualité de l’enseignement dans une institution universitaire africaine. L’analyse des sujets (figures 5 et 6) pour sa part donne une meilleure compréhension de la recherche dans un domaine. Elle permet d’identifier quels sont les domaines présents et ceux qui émergent.

Les bases de données bibliographiques de renom à partir desquelles sont faites les analyses bibliométriques intègrent au-delà des métadonnées Dublin Core d’autres types d’indicateurs notamment les citations, les affiliations des auteurs. Comme pour les métadonnées intégrées dans notre projet d’analyse, il est possible d’en intégrer d’autres pour arriver à avoir des indicateurs et cartographier les citations et les collaborations qui peuvent se nouer entre pays et institution de recherche (universités, écoles, laboratoires, etc.).

Dans aucune version de Dspace, les références bibliographiques et les affiliations auxquelles appartiennent les auteurs ne sont pas prises en compte. Leur intégration ou la création d'une collection « auteurs cités » pourrait permettre de mesurer la performance des chercheurs et des institutions de recherche.

Les universités et institutions de recherche de l'espace Cames ne disposent pas d'outils qui leur permettent de faire des analyses bibliométriques exhaustives. La plupart des revues dans lesquelles elles publient sont des revues hors de leur espace. Le Cames pour mieux évaluer la recherche et la qualité de l'enseignement dans ces pays se propose de concevoir des outils à cet effet. Parmi ces outils le DICAMES. Ce dépôt est présenté sur son site comme l'un outil destiné à la conservation et à la diffusion de toute la production scientifique des universités de l'espace CAMES.

Les résultats de ce travail révèlent qu'un tel dépôt peut servir d'outils d'évaluation pour les universités et institution au même rang que le CAMES.

Le dépôt actuel du Cames ne fournit pas assez d'information pour en faire un outil d'évaluation. Dans ce dépôt comme dans d'autres, les aspects documentaires et bibliométriques sont très peu considérés. Il apparait clairement que la mise en place d'un dépôt institutionnel destiné à évaluer une recherche nécessite une définition des objectifs documentaires et bibliométriques.

Pour les pays en voie de développement notamment ceux absents des bases de données de renom, Dspace peut ainsi constituer au-delà de la sauvegarde de leur patrimoine documentaire un outil de mesure et de prise de décision.

<input type="checkbox"/> Pan African Medical Journal	(563) >	<input type="checkbox"/> Journal Francais D Ophtalmologie	(54) >	<input type="checkbox"/> Journal De Mycologie Medicale	(36) >	<input type="checkbox"/> Travel Medicine And Infectious Disease	(27) >
<input type="checkbox"/> Plos One	(549) >	<input type="checkbox"/> Infectious Diseases Of Poverty	(53) >	<input type="checkbox"/> Journal Of Medical Case Reports	(36) >	<input type="checkbox"/> Agriculture And Food Security	(26) >
<input type="checkbox"/> Malaria Journal	(304) >	<input type="checkbox"/> New Microbes And New Infections	(53) >	<input type="checkbox"/> Nonlinear Dynamics And Chemotherapy	(36) >	<input type="checkbox"/> BMC Medicine	(26) >
<input type="checkbox"/> Plos Neglected Tropical Diseases	(251) >	<input type="checkbox"/> Cahiers Agricultures	(51) >	<input type="checkbox"/> Ecology And Evolution	(35) >	<input type="checkbox"/> Geo Eco Trop	(26) >
<input type="checkbox"/> Scientific Reports	(231) >	<input type="checkbox"/> Science Of The Total Environment	(51) >	<input type="checkbox"/> American Journal Of Primatology	(34) >	<input type="checkbox"/> International Journal Of Science	(26) >
<input type="checkbox"/> Parasites And Vectors	(177) >	<input type="checkbox"/> BMC Pregnancy And Childbirth	(50) >	<input type="checkbox"/> Biotechnology Agronomy And Society And Environment	(34) >	<input type="checkbox"/> International Journal Of Tuberculosis And Lung Disease	(26) >
<input type="checkbox"/> Medecine Et Sante Tropicales	(176) >	<input type="checkbox"/> ACM International Conference Proceeding Series	(49) >	<input type="checkbox"/> Evidence Based Complementary And Alternative Medicine	(34) >	<input type="checkbox"/> Jci Insight	(26) >
<input type="checkbox"/> American Journal Of Tropical Medicine And Hygiene	(157) >	<input type="checkbox"/> European Physical Journal Plus	(49) >	<input type="checkbox"/> Infection Genetics And Evolution	(34) >	<input type="checkbox"/> Monthly Notices Of The Royal Astronomical Society	(26) >
<input type="checkbox"/> BMC Research Notes	(140) >	<input type="checkbox"/> Food Science And Nutrition	(49) >	<input type="checkbox"/> International Journal Of	(34) >	<input type="checkbox"/> New England Journal Of Medicine	(26) >
<input type="checkbox"/> BMJ Open	(129) >	<input type="checkbox"/> World Journal Of	(48) >	<input type="checkbox"/> International Journal Of	(34) >	<input type="checkbox"/> Academic Pro	(26) >
<input type="checkbox"/> BMC Public Health	(123) >						
<input type="checkbox"/> Lecture Notes Of The Institute For Computer	(121) >						

Figure 7 : Revues dans lesquelles les PEC publient le plus dans Scopus⁶

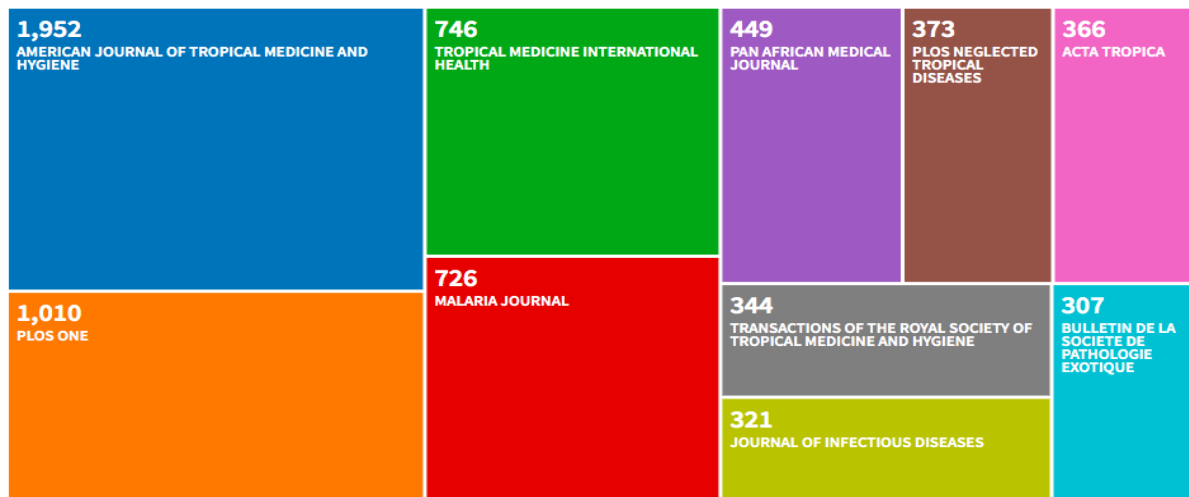


Figure 8 : Revues dans lesquelles les PEC publient le plus dans Web Of science

⁶ Nous avons effectué une recherche par « affiliation pays » dans la base de données Scopus pour chacun des pays de l'espace Cames. Nous avons combiné ensuite les résultats par l'opérateur booléens AND.

Conclusion

Cette étude s'est intéressée à deux questions très peu associées l'une à l'autre dans un contexte africain : les archives ouvertes et l'évaluation de la recherche. Les nombreux travaux sur les dépôts institutionnels se sont longtemps préoccupés des aspects devenus de moins en moins problématiques. La place des dépôts dans le contexte global d'un marché de la publication scientifique et de la diffusion des résultats de la recherche est devenue en effet presque certaine pour les documentalistes africains.

Les résultats de cette étude prouvent que les dépôts institutionnels peuvent être un outil pour le suivi de la qualité de la recherche et de l'enseignement pour les pays en voie de développement.

La fiabilité et la qualité de ce suivi ne peuvent en effet être espérées que si chaque institution de recherche agrège de manière « exhaustive » toute sa production scientifique dans un dépôt. Le calcul du H-index en est une illustration parfaite. Le H-index comme tous les indicateurs bibliométriques est calculé sur la base d'un corpus. Ainsi pour les bases de données bibliographiques dont les revues indexées sont majoritairement Américaines ou Britanniques, les chercheurs africains et leur institution de rattachement sont mal classés. Une analyse sur la base d'un corpus hétérogène ne reflète pas la réalité. Ainsi les pays en développement, pour disposer d'indicateurs fiables devront procéder à l'identification et à la sauvegarde de leur patrimoine documentaire scientifique. L'inaccessibilité des bases de données bibliographiques

et la visibilité des revues des pays africains ne devraient plus être une contrainte pour évaluer la recherche africaine.

Cette étude comporte quelques limites, la base de données utilisée n'est pas assez alimentée pour observer les différentes interactions. La fiabilité des résultats est certes démontrée, mais l'exhaustivité, l'enregistrement correct des publications dans reste à apprécier.

Références

Diouf, D. (2009). *Quelle méthodologie pour l'archivage et la diffusion électronique de la documentation scientifique et technique*. <https://www.theses.fr/2009TROY0037>

Lo, G. (2016). Les dépôts numériques dans les universités d'Afrique noire Francophone . *La Revue Maghrébine de Documentation et d'information*, 25. <https://hal-auf.archives-ouvertes.fr/hal-01418854>

ⁱ SciVal est un outil développé par Elsevier. Il donne accès aux performances de recherche de plus de 20 700 instituts de recherche et de leurs chercheurs associés, de 234 pays à travers le monde. IL permet de trouver des collaborateurs pour stimuler des solutions innovantes à des problèmes complexes.