

ANALYSER LES USAGES D'E-BOOKS

Mode d'emploi

Vincent de Lavenne
Université Bretagne Sud



Doc' Camp 2025
(Re)parlons PoIDoc
LILLIAD



No 25.

Collybia radicata.
Collybia à racine.



No 26.

Collybia velutipes.
Collybia à pied velouté. (comestible).



No 7.

cata.
nisse.



Résumé

L'objectif du présent document est de transmettre une méthodologie d'analyse d'une offre d'ebooks sur abonnement, dans le but d'avancer notamment une réflexion sur l'utilisation de la base, les manières de la compléter ou, dans un contexte budgétaire contraint, sur les possibilités de substituer ou de pallier à celle-ci par une autre méthode d'acquisition.

Pour ce faire, la méthode propose quatre parties :

→ l'analyse des données d'usage, qui permet de détecter s'il existe des publics spécifiques à celle-ci, et de pouvoir ainsi compléter l'analyse quantitative avec une forme qualitative

→ l'analyse des données éditeurs, qui permet d'avoir une compréhension plus fine des documents utilisés au sein de la base

→ le calcul d'un taux de couverture chez un fournisseur titre-à-titre, afin de comparer celui-ci à la variété de l'offre d'un bouquet

→ le calcul d'un taux de remplacement, total ou partiel, pour vérifier la validité économique de la substitution d'une offre de bouquet par une pratique d'acquisition au titre-à-titre

Enfin, nous proposerons des éléments de réflexion critique et de conclusion

Travailler les données d'usage

Les données d'usage utilisées dans la méthodologie sont extraites d'EZMesure, en sélectionnant, sur un tableau de bord standard au niveau établissement, la base d'ebooks souhaitée (par exemple, Cyberlibris pour ScholarVox), afin d'analyser les formations et structures les plus utilisatrices, ceci afin de repérer si la répartition des usages se fait de manière homogène entre les formations universitaires et composantes de l'institution.

Dépendamment du paramétrage d'EZMesure, il pourra être nécessaire de paramétrer un croisement pour faire apparaître le nom lié aux codes étapes.

Dans cette méthode, l'ensemble des croisements s'est faite sur Excel, via la fonction RECHERCHEV.

Celle-ci vous permet de comparer une cellule à l'ensemble des valeurs de plusieurs colonnes, et de retourner, dans la cellule où l'opération a été inscrite, la valeur correspondante à la cellule recherchée, ou celle d'une colonne voisine. Par exemple, dans notre cas, nous avons une colonne A dans laquelle étaient classées les code-étapes des formations, une colonne B vide, une colonne C dans laquelle étaient classés les volumes d'accès par formation. Nous avons rajouté des colonnes D et E mentionnant respectivement l'ensemble des codes-étapes de l'université (en D) et leur intitulé (en E).

Pour inscrire dans la colonne B l'intitulé correspondant, nous inscrivons donc dans la cellule : =RECHERCHEV(A2;D:E;2;FAUX).

Code étape	Intitulé	Nombre d'accès	Codes étapes université	Intitulé
ppM983	=RECHERCHEV(A2; D:E;2;FAUX)	2300	ouY748	L1 Sciences de la vie et de la Terre
alkdd78		1457	DdtU4	L2 Sciences de la vie et de la Terre
m1458J		298	ppM983	L3 Sciences de la vie et de la Terre
suh751		5	Fd6Er8	M1 Sciences de la vie et de la terre
			4Csku7	M2 Sciences de la vie et de la terre
			Skp54S	L1 Économie
			Df789r	L2 Économie
			alkdd78	L3 Économie
			Pm63Lm	M1 Bancassurance
			87Amn1	M2 Bancassurance
			suh751	L1 Lettres Classiques
			Bz1Ow	L2 Lettres Classiques

Fig 1 : Exemple de fonction RECHERCHEV pour extraire l'intitulé du code étape dans une liste.

La formule va donc comparer l'intitulé de A2 avec les colonnes D et E. Quand elle trouvera une correspondance (en D), elle rentrera dans B2 le contenu de la deuxième colonne (E) de la même ligne que la correspondance.

Cette formule peut être répétée dans toute la colonne, et permet de faire des graphiques de répartition d'utilisation.

Code étape	Intitulé	Nombre d'accès	Codes étapes université	Intitulé
ppM983	L3 Sciences de la vie et de la Terre	2300	ouY748	L1 Sciences de la vie et de la Terre
alkdd78	L3 Économie	1457	DdtU4	L2 Sciences de la vie et de la Terre
m1458J	M2 Humanités Numériques	298	ppM983	L3 Sciences de la vie et de la Terre
suh751	L1 Lettres Classiques	5	Fd6Er8	M1 Sciences de la vie et de la terre
				M2 Sciences de la vie et de la terre
				L1 Économie
				L2 Économie
				L3 Économie
				M1 Bancassurance
				M2 Bancassurance
				L1 Lettres Classiques
				L2 Lettres Classiques
				L3 Lettres Classiques
				M1 Humanités Numériques
				M2 Humanités Numériques

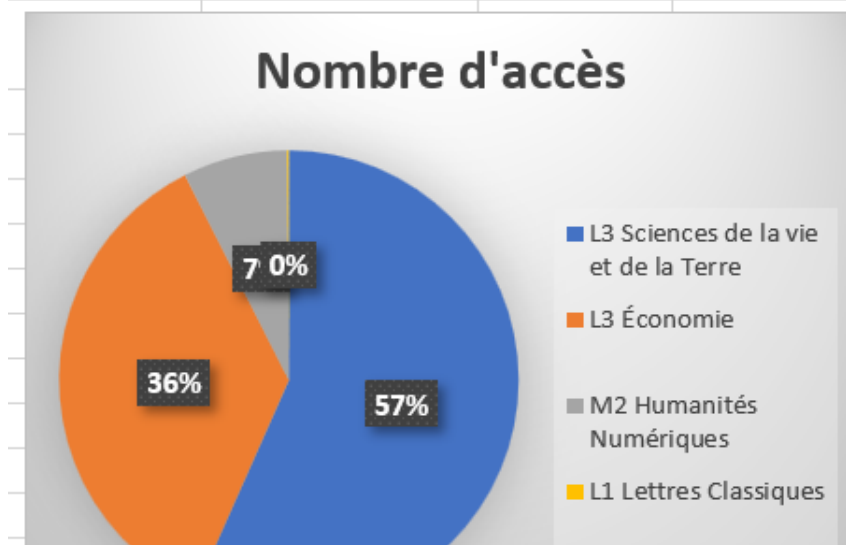


Fig 2 : conception d'un graphique d'utilisation clair avec les nombres d'accès par intitulé

Travailler les données éditeurs

Les données éditeurs pour les bouquets d'ebooks prennent, dans le meilleur des cas, la forme d'un *book report*, portant sur tout ou partie de l'entité documentaire. Dans le cadre de cette méthodologie, les valeurs qui nous ont paru intéressantes sont spécifiquement les champs titre, éditeur (*publisher*) et le nombre d'utilisation d'une période totale (*reporting period total*). L'enquête ne s'est pas spécifiquement basée sur les champs liés à des utilisations dans des périodes spécifiques, ce qui permettrait de mesurer l'évolution de l'utilisation d'un bouquet dans l'année, et de mesurer les pics et les creux (ainsi que l'impact immédiat et l'effet durable de formations documentaires, le cas échéant). Elle ne s'est pas basée quantitativement sur les ISSN qui pourrait

permettre de travailler sur des éventuelles utilisations d'éditions différentes d'ouvrages au sein d'un même bouquet¹.

Les champs retenus ont été intégrés, sous un classeur (de type Excel, OpenOffice), dans un tableau croisé dynamique. Le premier tableau retenu identifiait la somme du *reporting period total* pour chaque *publisher* (champ *publisher* en ligne, somme de *reporting period total* en valeur), et en supplément le nombre d'ISBN pour chaque *publisher*.

Étiquettes de lignes	Somme de Reporting Period Total	Nombre de ISBN
Editeur 1	3179	1266
Editeur 6	1391	642
Editeur 4	1077	517
Editeur 10	872	609
Editeur 8	838	427
Editeur 9	468	355
Editeur 3	350	101
Editeur 112	343	196
Editeur 5	312	161
Editeur 12	304	160
Editeur 2	290	93
Editeur 21	270	146
Editeur 13	205	129
Editeur 14	185	91
Editeur 18	173	76
Editeur 34	158	86
Editeur 7	152	87
Editeur 44	142	90
Editeur 39	97	64
Editeur 15	92	55
Editeur 25	91	60

Fig 3 : sélection de la SOMME du reporting total, du NOMBRE d'ISBN en valeur et de la donnée Publisher en lignes

Ce qui permet d'indiquer à la fois le nombre d'usages pour chaque éditeur, et le nombre de titres sur lequel se répartissent ces usages. Une rapide consultation des données permet de vérifier si une consolidation est nécessaire : les éditeurs peuvent ainsi être identifiés différemment, et nécessiter une forme de regroupement (par exemple : Elsevier et Elsevier Science).

Les données ainsi croisées permettent déjà une compréhension de l'utilisation de la base, qu'il convient de croiser avec une connaissance empirique, afin de répondre à un ensemble de premières questions : y'a-t-il un ou plusieurs éditeurs sur-représentés ? La variété du bouquet se retrouve-t-elle dans les usages ? Elle permet d'identifier la distribution des usages, qui dans notre cas

¹ Ce qui nous aurait permis de juger de l'éventuelle pertinence des options de recherche ou de la mise à jour du contenu)

se situait globalement dans la logique d'une distribution de Pareto², avec un petit nombre d'éditeurs regroupant la très grande majorité des usages, et une « longue traîne ».

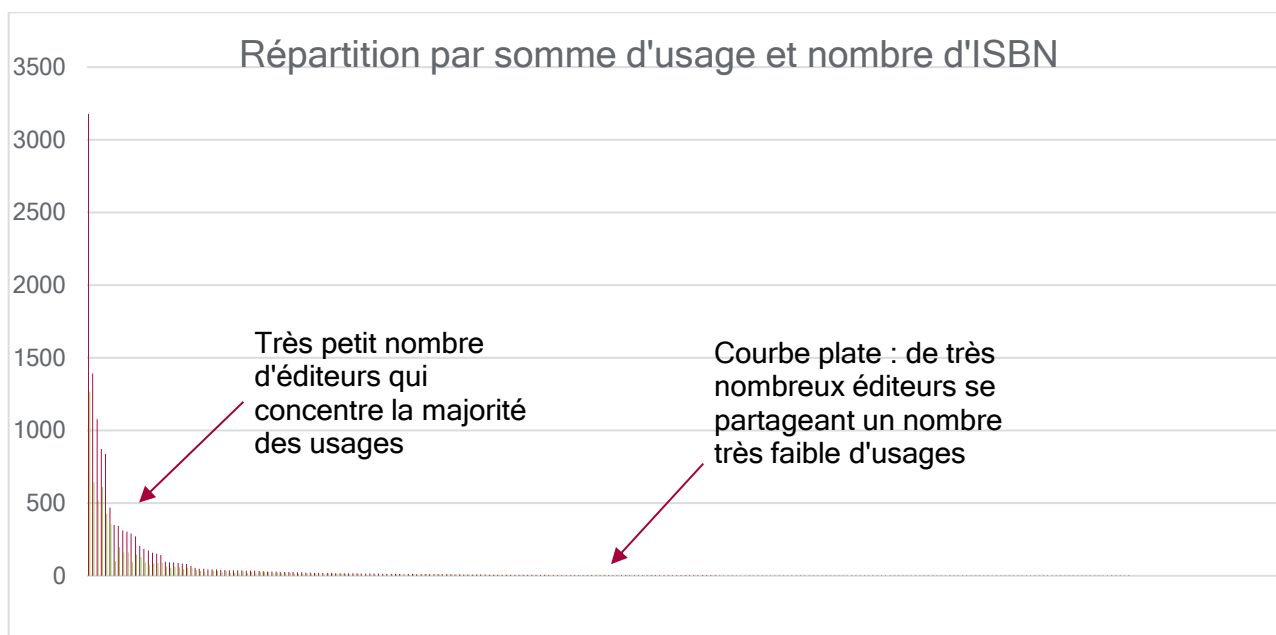


Fig 4 : une visualisation de la répartition des utilisations de la base (somme d'utilisations, nombre d'ISBN)

À la consultation de ces données, nous avons choisi de ne pas calculer un coût de remplacement total, ni de porter une analyse sur des éditeurs qui avaient cumulé **une** utilisation portée sur **un** titre, en partant de la logique que notre mission en tant que bibliothèque universitaire est de desservir, dans la mesure de nos moyens, un public-cible déterminé, et pas de répondre à l'intégralité des besoins individuels des personnes composant ce public-cible.

La détermination de la « ligne de découpe » peut se faire de manière propre à chaque établissement, et, dans notre cas, s'est fait d'une manière empirique, sur le critère des éditeurs dont plus de 60 titres ont été consultés dans la période d'analyse³. En suivant la même méthodologie, il est bien entendu tout à fait possible de porter l'analyse sur l'ensemble du corpus, ou au moins une portion plus large⁴.

À la suite de la détermination de la ligne de découpe, le corpus est donc constitué d'un ensemble d'éditeurs, avec un nombre de titres associés et un nombre total d'usages.

² La distribution de Pareto est généralement ramenée à la formule : 20% des causes entraînent 80% des effets

³ Dans notre cas, cela correspondait à 64% du total des titres, et 63% des usages.

⁴ Dans notre cas, l'une des raisons de l'étude était l'exploration d'éventuelles économies : le choix de la ligne de découpe tient donc aussi du fait de trouver un périmètre dans lequel les montants simulés n'étaient pas hors de portée budgétaire.

Calculer un taux de couverture de l'offre

Afin de proposer une offre alternative, il convient de comparer les données obtenues avec un ou plusieurs fournisseurs pour proposer des simulations de coût. Ces simulations donneront un ordre d'idée du coût de remplacement partiel ou total du contenu.

Dans le cadre de notre étude, nous avons choisi un fournisseur unique de titre-à-titre, et avons commencé par comparer, manuellement (à défaut de bénéficier d'une extraction rapide) l'offre éditoriale à celle du bouquet. Dans le cas où le taux de couverture ne soit pas satisfaisant, il peut être intéressant de rajouter plusieurs fournisseurs, mais cette combinaison devient une source potentielle de complexité de remplacement de l'offre en bouquet.

Calculer un coût de remplacement

Une fois le ou les fournisseurs sélectionnés, il convient de proposer un prix raisonnable sur lequel baser une simulation de coût. Dans le cadre de notre étude, et toujours en l'absence d'extraction exploitable sous forme de tableur, le choix a été fait de déterminer un **coût médian** par éditeur⁵, en vérifiant manuellement que celui-ci n'était pas aberrant par rapport aux documents présents dans le rapport d'usages. Dans la plupart des cas, les prix étaient assez homogènes pour que le prix médian soit une base solide de calcul.

Nous avons choisi de proposer, pour notre part, deux simulations : le **remplacement total** des titres chez les éditeurs retenus, et le **remplacement partiel** chez ces mêmes éditeurs.

Editeur	Nombre de titres utilisés	Prix médian constaté	Nombre d'accès	Coût remplacement total	Modèle économique (si unique ou majoritaire)	Nb d'ouvrages avec + de 5 accès	Coût remplacement partiel (ouvrages avec + de 5 accès)
Editeur 1	1266	42,48 €	3179	=B2*C2	License d'un an, 7 utilisateurs simultanés	171	=H2*C2

Fig. 5 : calcul des coûts de remplacements total et partiel

Pour le remplacement total, nous avons multiplié le **prix médian par le nombre de titres** pour obtenir une estimation. Les enseignements de ce calcul, dans notre cas, sont les suivants : il n'est budgétairement pas possible de faire l'acquisition au titre à titre de tous les documents utilisés sans un surinvestissement extrêmement conséquent sur cette forme de mise à

⁵ Par une forme de vérification empirique : sur 26 pages de résultats, vérifier le prix d'un item au milieu de la 13^e page, et vérifier ensuite par échantillonnage la validité de cette prédiction par rapport à l'offre

disposition⁶. C'est la raison même de l'existence des bouquets, mais c'est aussi ce qui nous amène à proposer une solution de remplacement partiel.

Pour le remplacement partiel, nous avons multiplié le **prix médian par le nombre de titres utilisés au moins cinq fois**. Cette barre correspond à un niveau d'utilisation souhaité pour un document acquis, c'est-à-dire qu'il permet d'éviter la longue traîne des titres utilisés 1 fois seulement. Dans ce cadre, nous nous situons dans une logique palliative, et dans cette logique palliative le remplacement vise donc les principaux usages des documents du bouquet, sans chercher à répliquer l'effet d'abondance que celui-ci propose. Les enseignements de ce calcul sont plus intéressants : le montant de ce calcul, dans notre étude, est intégrable dans notre répartition budgétaire actuelle ; il est même inférieur au coût d'abonnement au bouquet.

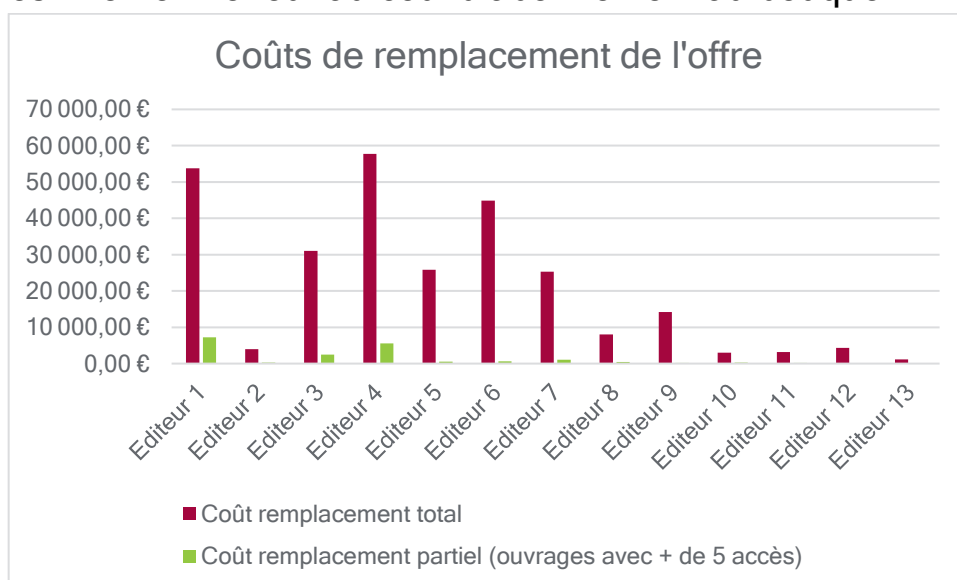


Fig. 6 : visualisation des différences entre les deux modèles de coût de remplacement

En somme, il n'est pas possible de répliquer un nombre certain de propriétés, notamment l'effet d'abondance qui provoque des utilisations de type « catalogue » ou « échantillonnage » documentaire au sein de la base. Il est cependant possible de remplacer les documents qui concentrent la majorité des usages.

Réflexion critique et conclusion

Il est important de rajouter à la réflexion la question des modalités d'achat chez le fournisseur de titre-à-titre. Dans un modèle économique évolutif au niveau du livre électronique, nous avons rencontré dans le cas de l'étude un grand

⁶ Dans l'étude, le coût de remplacement total des éditeurs sélectionnés (donc de 63% des utilisations totales) représente près de 2/3 du budget documentaire total de l'université, et près de trois fois le budget monographies (actuellement essentiellement utilisé pour l'acquisition de livres papier)

nombre de modèles différents de mise à disposition (nombre limité d'utilisateurs simultanés, achat limité dans le temps...) qui changent la nature du calcul selon les documents présents dans le corpus : le renouvellement induit par l'achat de documents de type manuels, fortement évolutifs ou par des livres de recherche pour lesquels le renouvellement est moins crucial, peut changer la donne du calcul ou l'intérêt de remplacer un abonnement par bouquet d'une acquisition au titre-à-titre.

Dans ces deux modalités d'acquisition, ce qui semble précieux reste la connaissance intime de la composition de l'offre par les bibliothécaires ; dans un système où l'acquisition d'ouvrages numériques au titre-à-titre est un miroir de l'acquisition d'ouvrages papiers, cette connaissance sera plus forte, mais nécessitera un effort financier dont le bénéfice est à recalculer fréquemment. Dans le cas de la mise à disposition via l'abonnement à un bouquet, l'absence de gestion des acquisitions et du désherbage éloignent la collection du bibliothécaire, mais celle-ci reste possible par un ensemble d'actions. A minima, si elle ne délimite pas d'éléments nouveaux sur la pertinence économique d'un choix, cette proposition d'étude permet une analyse plus fine de l'utilisation d'une base, un tour d'horizon d'autres modalités d'acquisition permettant d'évaluer l'évolution de l'offre en la matière, et peut amener à des dynamiques de changement sur un terme plus long, contrairement à une obligation d'adapter son offre dans la suite de contraintes budgétaires.