Document generated on 07/11/2025 10:40 p.m.

Documentation et bibliothèques

Données bibliométriques et informations

Roland Ducasse

Volume 24, Number 4, December 1978

URI: https://id.erudit.org/iderudit/1055127ar
DOI: https://doi.org/10.7202/1055127ar

See table of contents

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)
2291-8949 (digital)

Explore this journal

Cite this article

Ducasse, R. (1978). Données bibliométriques et informations. *Documentation et bibliothèques*, *24*(4), 165–173. https://doi.org/10.7202/1055127ar

Article abstract

The optimisation of the efficiency of an information service requires that its performance be measured. However, the evaluation of a library requires the use of bibliometric measurements. After having defined the nature of bibliometric measurements, the author emphasizes their importance for an analysis of the complicated documentary situations existing in libraries. He demonstrates the usefulness of these bibliometric measurements, in the management of documentary systems, with the help of two analytical patterns: the Buckland model which permits the determination of the number of copies to be purchased for any given title, and another model which is used to measure the degree of integration achieved by the library in its “milieu”.

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 1978

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Données bibliométriques et informations

Roland Ducasse
Assistant UPTEC-ILTAM
Université de Bordeaux III

L'optimisation de l'efficacité d'un système d'information nécessite la mesure de ses performances. Or, l'évaluation d'une bibliothèque requiert l'emploi de données bibliométriques. Après avoir défini la nature des données bibliométriques, l'auteur souligne leur importance pour une analyse approfondie des situations documentaires complexes existant dans les bibliothèques. Il démontre l'utilité des données bibliométriques dans la gestion des systèmes documentaires à partir de deux modèles analytiques: le modèle de Buckland qui permet de déterminer le nombre d'exemplaires à acquérir pour un titre donné et un modèle qui sert à mesurer le «taux de pénétration» d'une bibliothèque dans son environnement.

The optimisation of the efficiency of an information service requires that its performance be measured. However, the evaluation of a library requires the use of bibliometric measurements. After having defined the nature of bibliometric measurements, the author emphasizes their importance for an analysis of the complicated documentary situations existing in libraries. He demonstrates the usefulness of these bibliometric measurements, in the management of documentary systems, with the help of two analytical patterns: the Buckland model which permits the determination of the number of copies to be purchased for any given title, and another model which is used to measure the degree of integration achieved by the library in its "milieu".

Para optimizar un sistema de información, es preciso medir su eficacia, y la evaluación de una biblioteca requiere el uso de datos bibliométricos. Después de explicar y definir los datos bibliométricos, el autor subraya su importancia en un análisis profundizado de situaciones complejas relativas a la documentación en las bibliotecas. Demuestra la utilidad de los datos bibliométricos en la gestión de sistemas de documentación con dos modelos: el modelo de Buckland, que permite determinar la cantidad de ejemplares de un volumen particular que la biblioteca debe adquirir, y un modelo que se utiliza para medir el «coeficiente de penetración» de una biblioteca en su medio circundante.

Définir la bibliothèque comme un système d'information, c'est avant tout mettre l'accent sur la complexité d'une organisation composée d'individus, de matériel et de procédures, dont la finalité première reste la mise en relation permanente et rapide de l'information et de l'usager. Organisation complexe, car les responsables de sa configuration doivent *également* tenir compte de l'aptitude du système à traiter l'information documentaire en amont, i.e. l'offre, comme de son aptitude à répondre à la demande en aval, qu'elle soit formulée ou non. Or, les procédures de traitement

(sélection, acquisition, classification, catalogage), pour itératives qu'elles soient, portent sur des unités dont les caractéristiques de production, de distribution, d'audience, de durée sont radicalement différentes. Difficulté spécifique à laquelle s'ajoute l'accroissement quasi-exponentiel[1] de la production internationale qui draine vers chaque bibliothèque un flux virtuel de documents dont le volume est sans commune mesure avec la capacité réelle d'absorption.

Comment dans ces conditions faire en sorte que la fonction relationnelle, prépondérante, ne diminue ni en intensité, ni en valeur et qu'à une demande, même faible, le système réponde positivement?

Comment optimiser cette fonction dans le cadre étroit des ressources pécuniaires que l'on connaît en règle générale, cadre qui sanctionne durement toute erreur, tout investissement improductif?

On connaît la réponse: elle s'inscrit dans une rationalisation rigoureuse des procédures de sélection et d'acquisition; elle implique la définition préalable des objectifs du système; elle ne néglige pas de contrôler a posteriori la validité et l'efficacité des décisions prises, en vue d'en corriger au besoin la nature ou l'ampleur.

On connaît moins les moyens. Ils passent selon nous par la capacité du système à auto-évaluer ses performances, et c'est le point sur lequel nous voudrions insister ici en montrant le contexte théorique dans lequel il s'insère.

L'évaluation de l'aptitude d'un système d'information à atteindre les objectifs qu'on lui a fixés en matière de développement culturel, d'éducation, de services, de recherche scientifique... est généralement fondée sur l'utilisation effective du système par la population concernée. Cette utilisation peut être mesurée.

Elle le sera le plus souvent à partir des données bibliométriques endogènes, i.e. fournies par chaque composante du sous-système de circulation des documents, en liaison étroite avec celles issues des différents services: acquisition, référence, matériel. Ces données, l'analyste les confrontera à celles, exogènes, ayant trait à l'environnement du système d'information (les clientèles, mais aussi la superstructure documentaire), données dont il n'est sans doute plus nécessaire de souligner l'importance.

Nous entendons par donnée bibliométrique:

toute information (ensemble d') de nature quantitative, normalisée, ayant trait aux différents éléments du système (documents, usagers, transactions) qui, après avoir été saisie, donne aux principaux responsables un état actualisé dont l'analyse engendrera la prise de décision.

Si nous examinons les termes de cette définition, nous remarquons:

— *la nature quantitative de l'information*. C'est en effet une grandeur mesurable: elle porte généralement sur des effectifs, des dimensions, des durées, des coûts.

— *la normalisation nécessaire*. L'identification et la description précise des données, au préalable, sont des impératifs absolus. La normalisation s'inscrit le plus souvent dans la codification des données.

— *la portée sur les différents éléments du système d'information*. Les acquisitions, le catalogage, les transactions, les collections, les usagers peuvent être l'objet d'une description et d'une analyse quantitative.

— *après avoir été saisie...* Une donnée n'existe que lorsqu'elle a été saisie ou collectée, c'est-à-dire lorsqu'on a procédé

1. J. Meyriat, «Les non-livres», in *Le livre français: hier, aujourd'hui, demain...*, sous la dir. de J. Cain, R. Escarpit et H. J. Martin, Paris, Imprimerie nationale, 1972, p. 320.

à son enregistrement selon des techniques que nous n'examinerons pas ici, et dans le cadre de procédures strictement définies[2].

— *donne aux principaux responsables...* Dans un système traditionnel, la décision appartient le plus souvent au conservateur, au bibliothécaire, lequel, face à l'absence de données objectives (ou à leur rareté) a tendance à fonder le processus décisionnel uniquement sur son expérience professionnelle; avec un système bibliométrique la décision pourra être décentralisée, distribuée entre plusieurs destinataires selon leurs responsabilités.

— *un état actualisé...* La fréquence de saisie des données variant avec leur type (et selon le coût induit), ce n'est pas tant la vitesse de traitement qui importe, mais plutôt la régularité et la ponctualité de leur échéance fonctionnelle. En fait, l'actualisation des états doit être liée aux opérations de contrôle, des transactions par exemple, de la gestion (audit), ou commandée durant la phase d'instruction du processus de prise de décision.

— *dont l'analyse engendrera la prise de décision.* La production d'états statistiques n'est pas une fin en soi; elle implique conséquemment la démarche analytique. En fait, à un ensemble de données bibliométriques doit correspondre un mode d'analyse dont on se sera au préalable assuré de la pertinence par rapport à la nature de la décision. Par exemple, la distribution des fréquences d'emprunt des ouvrages d'une catégorie donnée sera confrontée à une distribution théorique[3], (Poisson, binomiale négative) permettant de déterminer la probabilité d'utilisation des ouvrages au cours d'une période donnée et donc de décider une modification éventuelle de la durée du prêt, du développement de la collection, etc.

Données naturelles et informations

Cette définition des données bibliométriques sera nécessairement complétée par la remarque suivante qui distingue deux types de données, selon que leur origine est:

— *naturelle,* c'est-à-dire que la valeur de la variable observée est obtenue à partir d'un simple dénombrement; par exemple, on dira que le nombre de transactions mensuelles (sorties, retours, réservations) est une donnée bibliométrique naturelle.

— *élaborée,* c'est-à-dire que la valeur de la variable observée résulte d'un traitement préalable, d'un rapport (ratio), d'une analyse; par exemple, on dira que le nombre de transactions corrigé des variations saisonnières est une donnée bibliométrique élaborée.

En d'autres termes, nous soulignons ainsi le caractère métabolique des données, qui ne sont pas intrinsèquement des informations, *mais seulement des sources d'information*. Celles que nous appelons des données élaborées représentent en fait, au point de vue opérationnel, les véritables informations.

Si nous insistons sur la différence entre le concept de donnée et celui d'information, c'est parce que nous avons pu constater que la mise en place d'une base de données naturelles (ou d'un système comparable) constituait assez souvent le stade ultime de la gestion automatisée.

C'est oublier nous semble-t-il les préceptes de la théorie de Shannon[4] déjà esquissés par R.V.L. Hartley[5]. Pour lui, en effet, une information est le moyen de sélectionner certains signes (données) dans un ensemble de signes équiproba-

2. Pierre d'achoppement de maintes tentatives de rationalisation des choix de gestion, la saisie des données, si elle est certes facilitée par l'automatisation, n'en demeure pas moins praticable à partir de systèmes traditionnels: Type Newark par exemple pour la circulation des documents.
3. Voir à ce sujet l'ouvrage fondamental de Philip Morse, *Library Effectiveness: a Systems Approach*, Cambridge, Mass., M.I.T. Press, 1968, 207 p.
4. C.E. Shannon, «A Mathematical theory of communication», *Bell System Technical Journal*, vol. 27 (July 1948), 379-423, (October 1948), 623-656.
5. R. V. L. Hartley, «Transmission of information», *Bell System Technical Journal*, vol. 7 (July 1928), 535-563.

bles. Ainsi, pour un message composé de N signes pris dans un ensemble A comprenant un nombre d'éléments S, l'ensemble des messages possibles est égal à:

$$A_N^S = S^N$$

et la quantité Q d'information nécessaire pour en sélectionner un est égale à:

$$Q = N \log S$$

la quantité majeure d'informations d'une source, c'est-à-dire son *entropie*, devenant:

$$Q = \sum_{i=1}^{n} Pi. \log \frac{1}{Pi}$$

Nous pourrions dès lors dire qu'élaborer des données, c'est énoncer toute l'information qu'elles contiennent, ou encore, en reprenant l'expression utilisée par Robert Escarpit à propos de la lecture objective d'un texte, que cela revient à «épuiser l'entropie de la source.»[6]

En résumé une donnée n'a d'importance pour un système qu'en fonction de son efficience informationnelle, ce que Ph. Dumas[7] exprime dans la formule:

information = donnée + signification

la signification étant la marque de sa finalité par rapport à un objectif, une décision, etc.

C'est pourquoi il n'y a pas lieu d'insister particulièrement sur une typologie des données bibliométriques naturelles, à synergie faible, quand le bibliothécaire ne les charge pas de signification, mais plutôt sur le processus de construction des données élaborées ou données transformées en informations.

6. Robert Escarpit, *Théorie générale de l'information et de la communication*, Paris, Hachette, 1976, p. 126.
7. Ph. Dumas, *Une approche développementielle de l'étude d'un système d'information*, Aix en Provence, C.E.R.O.G., 1975.

Typologie des données bibliométriques naturelles (sous-système de circulation)

Une typologie élémentaire des données bibliométriques naturelles susceptibles d'être collectées, régulièrement ou occasionnellement par un système de circulation, pourrait revêtir la forme suivante (page 169), qui ne distingue pas le type de bibliothèque (de jeunesse, de lecture publique, d'université):

Comme il est facile d'en juger à la lecture de ce tableau, l'établissement des données naturelles revient finalement à restituer sous la forme d'états (outputs) les éléments descripteurs (items) sélectionnés à l'entrée du système (inputs).

Un traitement positif de l'information ne saurait se limiter à ce qui n'est finalement qu'une réorganisation, un reclassement, une ventilation de l'ensemble de départ. Phase strictement descriptive, elle permet certes une meilleure lecture de «l'information», mais ne l'exploite guère.

Le schéma suivant, relatif à l'économie d'un système de prêt automatisé (page 170), montre que l'analyste peut — *et doit* — intervenir de manière inductivo-déductive au niveau de la base de données en sollicitant celles qui lui paraissent nécessaires à la résolution de tel problème, au contrôle de telle opération ou en vue d'une décision particulière. S'assurant de leur pertinence, puis définissant leur mode de traitement et d'analyse, le bibliothécaire lance la phase d'élaboration.

Comme le souligne ce schéma, l'élaboration des données bibliométriques est doublement liée:

— à l'organisation de la base de données;
— à une démarche analytique inductivo-déductive.

Dès lors, établir la typologie des données bibliométriques (ici dans le cas du sous-système de circulation) relève soit de la

Données naturelles

Ensemble des documents

- Nombre total de documents
- Nombre de documents par catégorie (livres, périodiques, cartes, etc.)
- Nombre de titres
- Nombre de documents par collection
- Nombre de documents par section
- Nombre de documents par cote
- Nombre de documents par format
- Nombre de documents par support
- Nombre de documents par zone de stockage
- Nombre de documents selon leur origine (don, échange, fournisseur)
- Nombre de documents selon le pays d'édition
- Nombre de documents selon la date d'édition
- Nombre de documents selon la date d'acquisition
- Nombre de documents selon la langue
- Nombre de documents selon le mode d'accès (direct, réserve)
- Nombre de documents par descripteur

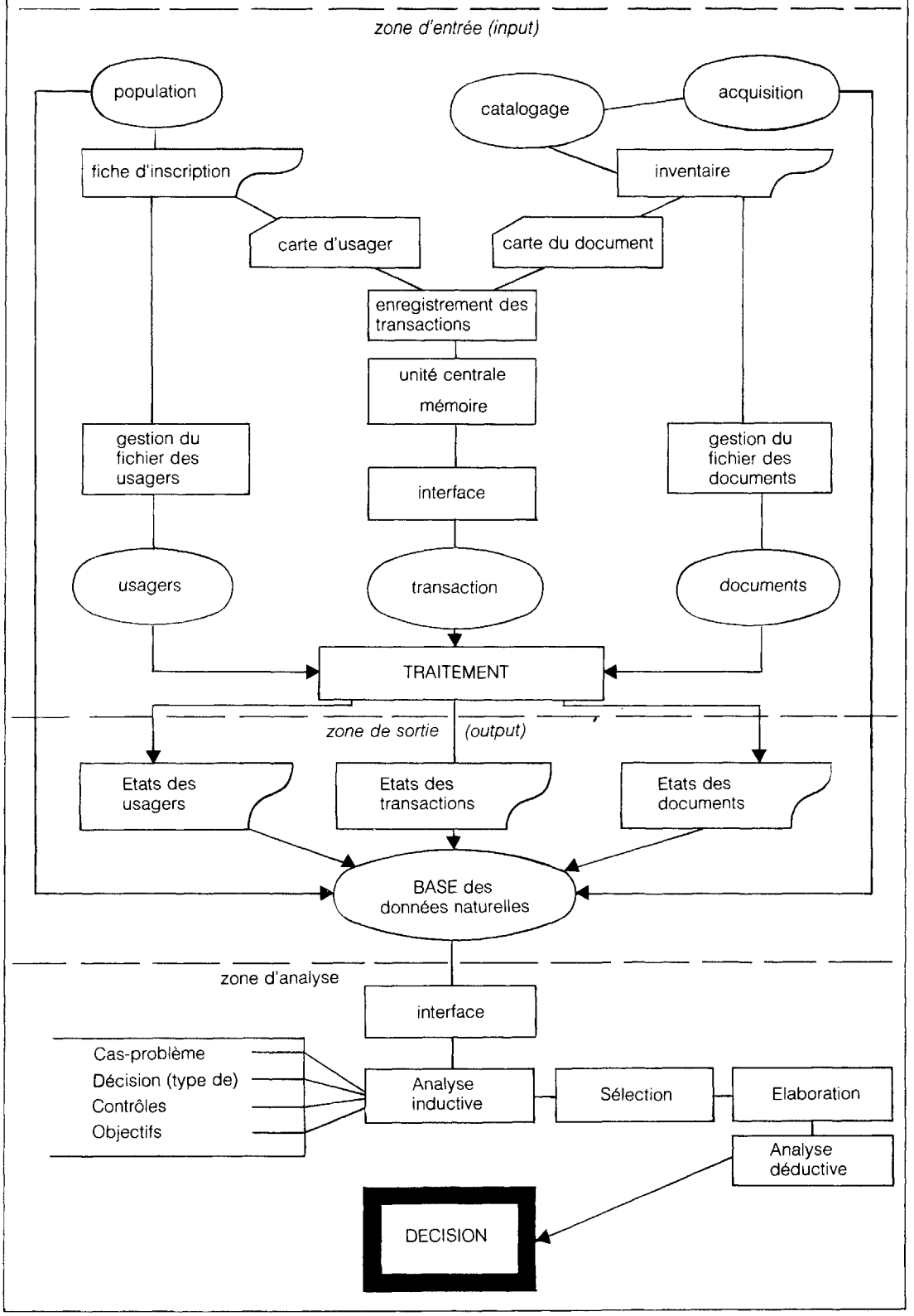
Ensemble des usagers

- Nombre d'usagers selon l'âge
- Nombre d'usagers selon le sexe
- Nombre d'usagers selon la catégorie socio-professionnelle
- Nombre d'usagers selon le secteur d'activités (agricole, industriel, commercial)
- Nombre d'usagers selon le niveau d'études
- Nombre d'usagers selon la langue
- Nombre d'usagers selon le pays d'origine
- Nombre d'usagers selon le statut (enfant, adolescent, adulte, étudiant, chercheur)
- Nombre d'usagers selon le lieu de résidence (ville, zone, quartier, rue)
- Nombre total d'inscrits
- Nombre d'inscrits par section

Ensemble des transactions

- Nombre total de transactions (au cours d'une période de référence)
- Nombre de transactions selon la date
- Nombre de transactions selon le type (sortie, retours, lecture sur place, réservations)
- Nombre de transactions selon le lieu d'enregistrement (section, annexe)
- Nombre de transactions selon les item décrivant
  - l'ensemble des documents
  - l'ensemble des usagers

Economie générale d'un système de prêt automatisé



combinatoire intra et inter-groupes[8] des données naturelles relatives aux trois fichiers de base (ensemble des documents, ensemble des usagers, ensemble des transactions), soit de l'analyse de l'ensemble des décisions. Il va de soi que la seconde possibilité n'est guère praticable in abstracto et que la première serait pour le moins fastidieuse. Aussi nous bornerons-nous à l'exposé sommaire de deux cas relevant de l'une et de l'autre.

Premier cas: Données bibliométriques élaborées dans un processus décisionnel: application au choix du nombre d'exemplaires à acquérir pour un document donné[9].

La décision relative à la quantité d'exemplaires à acquérir pour un document donné représente une difficulté majeure pour la gestion des bibliothèques. En effet, si un nombre insuffisant d'exemplaires est acquis, la bibliothèque ne satisfait pas pleinement le besoin d'information des usagers et assure une qualité de service passablement médiocre: attente, frustration, renoncement. Dans le cas contraire, si un trop grand nombre d'exemplaires est acquis, il s'ensuit un gaspillage financier car le coût d'acquisition, de traitement et de conservation aurait pu être reporté sur d'autres documents ou services. Ce problème est particulièrement important dans les bibliothèques universitaires qui doivent servir équitablement des étudiants suivant les mêmes programmes et manifestant souvent au même moment le même besoin.

Dans un premier temps, l'analyse inductive aura pour rôle la mise en évidence des principaux facteurs, dont la mesure constituera ensuite l'ensemble des données.

Soit, a priori, la quantité q d'exemplaires d'un ouvrage donné A dont l'acquisition a été jugée nécessaire; elle se définit comme le nombre d'exemplaires suffisant à la satisfaction du besoin d'information exprimé par certains usagers. La demande est satisfaite dès lors que l'usager reçoit immédiatement l'ouvrage requis.

Si nous appelons n: le nombre de demandes relatives à l'ouvrage A, on voit que q la quantité d'exemplaires de A doit être telle que:

$$o \leqslant q \leqslant n:$$

Précisant que la demande n_i s'entend par période de prêt affectée à l'ouvrage A, on dira que P(n): est la fréquence relative des n_i demandes de A.

Lorsque $P(n_i)$ est inférieur ou égal à q pour une période donnée, la disponibilité de l'ouvrage est maximale et toutes les demandes sont satisfaites. Inversement, si $P(n_i)$ est supérieur à q, alors une partie de la demande est insatisfaite et conséquemment se crée une file d'attente.

Pour séduisante que soit l'idée de la détermination du nombre d'exemplaires q par la seule fréquence des demandes $P(n_i)$, on conçoit aisément les difficultés (financières, de stockage) qu'une telle démarche entraînerait. Aussi doit-on garder à l'esprit un troisième facteur: la durée du prêt qui, théoriquement, permet de régler la circulation de A et partant la satisfaction de la demande. A une demande faible correspondrait une longue durée de prêt et, inversement, à une demande forte correspondrait une durée de prêt plus courte. Ceci pour assurer la rotation maximale de l'ouvrage A.

A ce stade de l'analyse, on pourrait songer à élaborer un type d'information (ratio, indice, modèle) qui permettrait, connaissant la fréquence des demandes et la durée du prêt, de calculer le nombre d'exemplaires à acquérir, nombre minimisé par une rotation rapide par exemple.

Buckland, longtemps attaché de recherche à la bibliothèque universitaire de

8. Une combinatoire »aveugle» ne crée pas d'information. Que tirer par exemple du rapport Nombre de titres vs. Nombre d'usagers de plus de 50 ans?
9. Pour un développement, voir M. K. Buckland and I. Woodburn, «An analytical approach to duplication and availability», *University of Lancaster Library Occasional Papers no. 2*, Lancaster, 1968, 24 p.

Lancaster (Grande Bretagne)[10] a essayé avec succès un modèle de prévision qui fait intervenir un quatrième facteur qu'il appelle *standard of service*, c'est-à-dire un taux de disponibilité immédiate de l'ouvrage A.

Ce taux, dont le niveau est fixé par le responsable de la bibliothèque, donne la probabilité qu'a le lecteur de trouver immédiatement sur les rayons l'ouvrage recherché. Généralement établi autour de 80%, Buckland l'obtient par l'équation suivante, qui lie les quatre facteurs:

— nombre de demandes;
— nombre d'exemplaires;
— durée du prêt;
— *standard of service*;

telle que:
Disponibilité immédiate

$$(\text{en } \%) = \frac{100\,[ni\ \Sigma P(ni) + \Sigma q\ P(ni)]}{\Sigma ni\ P(ni)}$$

$$\text{avec } P(ni) = \frac{e^{-\bar{x}} \cdot \bar{x}^{n}}{n\,!}$$

et e = 2,71828

Le modèle se construit comme suit: pour un ouvrage donné A, on observe la distribution des fréquences de la demande au cours de périodes de prêt successives, ce qui permet de calculer la moyenne $\bar{x}$ par période; puis, on recherche la loi de distribution théorique qui ajuste au mieux la distribution observée. Ainsi que nous l'avons laissé entendre plus haut, il est fréquent d'ajuster la distribution des sorties d'ouvrages par une loi de Poisson de paramètre $\bar{x}$ ou moyenne de la distribution observée. Le modèle de Buckland faisant varier le nombre d'exemplaires et le taux de disponibilité immédiate, il est alors facile de déterminer le nombre d'exemplaires en fonction du taux de disponibilité immédiate requis, et, réciproquement, connaissant le taux de disponibilité immédiate obtenu avec un certain nombre d'exemplaires de A, d'augmenter celui-ci afin qu'il atteigne le niveau souhaité[11].

Ainsi, face à un ensemble de données naturelles collectées par le système, un bibliothécaire peut opter pour différentes stratégies, dans notre cas, combiner le nombre de demandes, le nombre d'exemplaires, la durée du prêt ou ajouter à ceux-ci une contrainte liée à un objectif particulier comme la disponibilité immédiate par exemple.

A l'université de Lancaster, le système de prêt porte sur vingt-deux périodes par semaine: chaque jour de la semaine en a quatre (matin, après-midi, soirée, soir au lendemain) et le samedi en comporte deux (matin et de midi au lundi matin), le bibliothécaire ne souhaitant pas modifier sa périodicité; l'adoption d'une stratégie portant uniquement sur la disponibilité immédiate des ouvrages a entraîné un accroissement du prêt d'environ 100%.

On retiendra de l'exposé ci-dessus l'intérêt que peuvent présenter les modèles analytiques pour la gestion des systèmes d'information. Construits à partir de données judicieusement collectées, ils aident, pour le moins, à clarifier une conceptualisation des facteurs impliqués dans des processus complexes comme la bibliothéconomie en connaît de nombreux.

Deuxième cas: Données bibliométriques élaborées par combinaison de données naturelles: application au calcul du «taux de pénétration» d'une bibliothèque dans la population à desservir.

Caractérisée par son implantation, sa dimension, la nature de ses collections et son environnement socio-culturel, la bibliothèque est un système d'information créé pour répondre à la demande d'une population donnée.

10. De 1967 à 1969 l'Office for Scientific and Technical Information a financé une équipe de recherche en bibliothéconomie à la bibliothèque de l'université de Lancaster. Dirigés par A. G. Mackenzie, les travaux de cette équipe ont atteint une notoriété mondiale.

11. On trouvera dans Michael K. Buckland, *Book Availability and the Library User*, New York, Pergamon Press, 1975, 196 p. un extrait de la table proposée par l'auteur.

Dans quelle mesure cet objectif essentiel est-il atteint? Pour le savoir[12], la bibliothèque peut opérer de différentes manières:

— en interrogeant l'ensemble de la population par la voie d'enquêtes, de sondages, etc.

— en interrogeant le système lui-même au niveau de l'enregistrement des inscrits (fichier d'inscription) ou des transactions (lecteurs actifs).

La seconde méthode est de loin la plus commode. Elle permet en effet d'obtenir très rapidement l'état statistique des usagers (éventuellement des actifs) et en les rapportant à celui de la population potentielle[13] de mesurer l'impact, l'audience de la bibliothèque dans la population.

Le taux de pénétration se définira alors comme le rapport (ratio) entre le nombre d'usagers[14] U et le nombre d'individus dans la population-mère N, tel que:

Taux de pénétration

$$(\text{en } \%) = \frac{100\ U}{N}$$

Distribué selon les différentes catégories de données naturelles relatives à la population, le taux de pénétration permettra de mettre en évidence les variations d'impact qui peuvent exister à son niveau, voire d'en comparer les résultats avec ceux obtenus par d'autres bibliothèques présentant des caractéristiques similaires.

Ces deux cas nous ont permis de mettre en évidence la nécessité, tant pour la recherche que pour la gestion, de dépasser le stade d'une élaboration primaire et contingente de données bibliométriques, pour construire une véritable typologie d'*indicateurs statistiques* qui constituerait par ailleurs le tableau de bord de l'entreprise qu'est à bien des égards une bibliothèque.

12. Le taux de pénétration traduit la réponse du système au besoin d'information en termes généraux seulement.
13. La population potentielle est généralement appréhendée à partir d'états officiels: recensement, inscription à l'université.
14. Donnée bibliométrique naturelle qui se distingue parfois de celle concernant les actifs (inscrits ayant emprunté au moins un ouvrage durant la période de référence).