

Les coûts et les délais des processus administratifs. Une interprétation des préceptes de Hammer

Jean-Louis PEAUCELLE

Professeur en sciences de gestion, IAE Paris I

RÉSUMÉ

Le BPR cherche à améliorer les délais et les coûts. Cet article fournit une interprétation des préceptes qualitatifs de Hammer en comparant les différentes manières de traiter des flux aléatoires de travail dans les processus administratifs. Cette comparaison est menée de manière discursive en liant les délais et les coûts dans chaque solution. Les préceptes qualitatifs avancés par Hammer prennent le sens d'une recherche de la solution organisationnelle la meilleure en délais et en coût. Si on ne tient compte ni des effets de productivité, ni des spécialisations des métiers, ni des différences de salaires, le centre d'appels, à un seul niveau, est l'organisation archétypale la meilleure, en termes de coût et délais.

Mots-clés : Centre d'appels, BPR, Processus, Reengineering, Reconfiguration, File d'attente, Hammer, Beckmann, Simulation, Délai, Coût.

ABSTRACT

Business Process Reengineering aims to improve deadlines and costs. This paper provides an interpretation of the Hammer's qualitative principles through a comparison of various methods to treat labor flows in administrative processes. This comparison is carried out in a discursive way connecting the deadlines and the cost for each solution. The Hammer's qualitative precepts appear as the seeking for the best organizational solution, the best under the criterion of deadline and cost. Without productivity effects, nor specialization, nor gap in wages, the call center with a unique level is the best archetypal organization according deadline and cost.

Key-words: Call Center, BPR, Process, Reengineering, Queues, Hammer, Beckmann, Simulation, Deadline, Cost.

1. LE BUSINESS PROCESS REENGINEERING (BPR) COMME NOUVELLE THÉORIE DE L'ORGANISATION

Depuis le début des années 1990, les entreprises occidentales réorganisent leurs activités administratives plus radicalement qu'elles ne l'avaient fait durant les trois décennies précédentes. Dès les années 60, elles ont informatisé largement les tâches administratives. Mais les résultats ont été sans doute insuffisants puisque de nouvelles améliorations semblent possibles avec le BPR (Business Process Reengineering) lancé par Hammer et Champy (1993) et repris par nombre de cabinets d'organisation. Beaucoup d'entreprises lancent des opérations de reconception de leurs processus. Le BPR devient une référence pour l'organisation. Cette réorganisation sera considérée comme caractéristique de notre époque au même titre que le taylorisme pour le début du XX^e siècle. Il faut comprendre ses principes et ses résultats.

Les résultats ne sont pas toujours aussi exceptionnels que les cas exemplaires choisis par les promoteurs du BPR le laisseraient croire. D'après Hall, Rosenthal et Wade (1993) l'amélioration de performance, par le BPR, est faible (moins de 5 %) dans 11 cas sur 20 firmes étudiées. Cependant, dans six cas la performance augmente de 18 %. La réorganisation par le BPR est un changement complexe (Grover, Jeong, Kettinger et Teng, 1995) dont le succès nécessite une gestion spécifique.

La place de l'informatique dans le BPR est particulièrement importante. Ces réorganisations supposent toujours un changement des systèmes d'information. Cette étape est souvent un goulet d'étrangle-

ment du processus de changement. L'informatique impose ses délais. Ses retards sont complètement répercutés sur la réorganisation. Les défaillances des infrastructures et les défauts des logiciels sont directement ressentis dans la nouvelle organisation. Le projet de BPR et l'organisation résultante sont très dépendants de la technologie.

Comme le taylorisme, le BPR peut être étudié sous différents aspects. Tout d'abord, au niveau de la doctrine, on peut examiner les textes des auteurs majeurs du domaine. Ensuite, au niveau des réalités, l'étude des processus réorganisés sous l'égide du BPR renseigne sur les faits. On y constate alors l'efficacité des principes affichés par les auteurs. Un troisième niveau concerne les processus d'action, la manière de réorganiser, les objectifs et les méthodes des acteurs. Ces trois aspects sont importants. C'est le premier qui est choisi dans cet article.

Étudier les principes affichés par des auteurs pourrait n'être que répétition d'idées. Quand on examine les textes sur le BPR, par exemple Hammer (1993 et 1996) ou Davenport et Stoddard (1994), on constate que les idées évoluent, et pas dans le sens d'une plus grande rigueur. Les textes incorporent les enseignements de la pratique où un grand nombre de compromis ont été nécessaires. Curieusement le premier article, celui de Hammer en 1990, paraît le plus clair.

Ces difficultés d'interprétation des textes se retrouvent dans toutes les œuvres des grands organisateurs. Ils sont des hommes d'action. Ils écrivent pour convaincre que leur action sera efficace.

Leurs textes ne sont pas l'explication qui rendrait reproductible leurs manières de faire. Ce sont au contraire des discours pour obtenir la conviction de leurs futurs clients : les dirigeants d'entreprise, relayés éventuellement par les enseignants. Alors, la structure de ces textes obéit à un schéma rhétorique classique en gestion : énoncé de principes généraux et description d'exemples où de considérables gains de performance ont été réalisés. Les cas sont considérés comme une preuve de la pertinence des principes auxquels ils sont accolés.

Mais la relation causale est loin d'être ainsi établie. Le cas est décrit si succinctement qu'on ne sait pas si d'autres facteurs ont pu jouer. La réalité est toujours l'objet de causalités multiples. D'autres facteurs dont on ne parle pas ont peut-être influé heureusement. L'exemple emporte la conviction mais n'est pas la démonstration. Il serait sans doute possible de trouver toujours un contre-exemple où l'action envisagée n'a pas eu d'effet à cause d'autres facteurs qui n'apparaissent pas dans la description de l'exemple.

2. MÉTHODOLOGIE

Dans cet article, nous allons examiner de façon approfondie l'article de Michael Hammer (1990) qui est au point de départ des réflexions sur le BPR. Il critique les informatisations traditionnelles qui ont copié les procédures manuelles antérieures notamment pour ne pas remettre en cause la répartition des rôles entre les divers services. Il avance sept nouveaux principes pour organiser les ser-

vices en faisant "table rase" ("oblitérate") de l'organisation passée.

Cet article de Hammer, dans sa nature prescriptive, semble sortir des sciences de gestion. Mais ces préceptes peuvent être interprétés, d'une part, en tant qu'objectifs fixés aux acteurs, directions d'entreprise et organisateurs. On peut alors s'interroger sur les raisons qu'ils ont d'y adhérer, sur l'écart entre leurs actions et ces préceptes. D'autre part, elles sont intéressantes comme hypothèses intuitives sur les changements efficaces dans les organisations. C'est ce point de vue qui nous intéresse ici. Quelles raisons donner pour justifier les préceptes ?

Ceci serait simple s'il n'y avait au fond que du bon sens dans ces préceptes, s'ils ne reprenaient que des idées couramment admises. Mais Hammer fait une innovation surprenante. Son premier précepte prône une polyvalence des personnes dans les diverses étapes d'un processus. Or depuis l'identification de la division du travail par Adam Smith, personne ne doute de l'intérêt d'une spécialisation des personnes pour augmenter la productivité dans le processus. Comment justifier cette originalité du BPR ?

Il existe plusieurs méthodes pour ce faire. La première est expérimentale. Il faudrait conduire avec rigueur des opérations de BPR en identifiant tous les facteurs et en isolant l'application des principes. Mais justement l'entreprise veut la meilleure amélioration possible et donc qu'on agisse, autant que faire se peut, sur tous les facteurs en même temps. Il est alors difficile d'isoler les préceptes afin d'examiner leur part contributive. Les buts de l'entreprise divergent ici avec ceux du chercheur.

Une autre voie consiste à reproduire le raisonnement par des modèles. Une préconisation de changement prend son sens si l'organisation nouvelle est meilleure que l'ancienne. Donc tout changement repose sur une notion de performance, critères en fonction desquels on juge l'écart entre les deux organisations. Tout changement préconisé dessine une opposition entre une situation ancienne et la nouvelle. La situation ancienne n'est pas toujours explicite, mais il est possible de la reconstituer. La lecture critique d'un texte d'organisation doit identifier les critères de performances, explicites ou non, et l'organisation ancienne qu'on préconise de changer. Le raisonnement par des modèles consiste à vérifier que les écarts de performances sont cohérents avec les préconisations. C'est la démarche suivie dans cet article.

La modélisation qui va être rapprochée du BPR est la théorie des files d'attente. Cette théorie mathématique fournit, dans des cas simples, le délai moyen de service d'un système répondant à une demande aléatoire. Pour les cas plus compliqués, la simulation des processus est nécessaire. Christoph Loch (1998) a déjà vu l'intérêt de rapprocher ces deux domaines de connaissance, BPR et files d'attente. A partir de son expérience en organisation industrielle, il effectue une relecture du livre de Champy et Hammer (1993). Il étudie particulièrement trois structures productives : la division du travail en série, la mise en parallèle de postes polyvalents et l'aiguillage des cas selon leur complexité. Il simule, pour un cas, la courbe de répartition des temps d'accomplissement du travail. La solution qui offre les délais les

plus brefs est celle des postes polyvalents qui sera appelée F (voir figure 1). Le présent article se propose d'approfondir ces résultats.

Tout d'abord quels sont, pour Hammer, les objectifs poursuivis ? "The watchwords of the new decade are innovation and speed, service and quality", écrit-il en 1990. Aurait-il oublié le coût ? Il est certain que non mais il ajoute d'autres critères qui feront la différence entre entreprises. De ces critères retenons le délai. Les organisations prônées par le BPR sont performantes en termes de délais et de coût *simultanément*. Dans tous les exemples, l'amélioration sur ces deux critères est sans cesse réaffirmée. Par exemple, Hammer parle des procédures d'une compagnie d'assurance (Mutual Benefit Life). La productivité est augmentée de 60 % pendant que le délai passe de 25 à 5 jours. Considérons donc que la performance d'une manière de réaliser un processus administratif est constituée de son coût et de son délai pour effectuer le travail.

3. LE PREMIER PRINCIPE DE HAMMER

Le premier principe énoncé par Hammer (1990) paraît étonnant. Il recommande une polyvalence des personnes de bout en bout dans le processus ("to have one person perform all the steps in a process"). La même personne réalise toutes les étapes du processus. Elle suit le dossier de bout en bout. Les tâches traditionnellement séparées sont intégrées dans un même poste.

Cette organisation nouvelle s'oppose à une organisation ancienne. Il n'est pas difficile de comprendre

qu'il s'agit de la division du travail. Le premier principe de Hammer s'oppose aux habitudes d'une division du travail. Dans la production industrielle, on a depuis très longtemps spécialisé les personnes, comme Adam Smith l'avait remarqué. Les raisons de cette spécialisation étaient :

- l'amélioration de l'habileté sur une tâche réduite (à condition qu'elle ne soit pas contrebalancée par l'ennui) ;
- la faible compétence de la main-d'œuvre disponible (avec une division du travail très poussée, on peut former rapidement les personnes) ;
- les aptitudes inégalement réparties chez les travailleurs (des personnes sans compétence, donc moins payées, exercent les postes les plus simples) ;
- l'adaptation aux machines qui sont les auxiliaires du travail industriel pour des étapes bien précises (les machines permettent une augmentation de productivité).

On sait que la division du travail trouve probablement son origine dans la différenciation des salaires de la chaîne de production (Peaucelle, 1999b). Les ouvriers les plus compétents peuvent travailler de manière polyvalente, ils préfèrent se spécialiser sur les postes les mieux payés. Ils occupent dans la chaîne de valeur les points où leur rémunération est la plus forte.

Certaines de ces raisons continuent d'être vraies dans les entreprises de service en cette fin du XX^e siècle. Les caractéristiques de la population des pays industrialisés et les technologies informatiques rendent ces raisons moins pressantes :

- les employés ont une compétence de base assurée par le système éducatif et on sélectionne, au moment de l'embauche, en fonction de cette compétence ;
- la formation permanente permet d'accroître rapidement la compétence et l'adaptabilité à de nouveaux modes de travail ;
- les machines informatiques ont une polyvalence qui permet aujourd'hui d'accéder à toute la variété des programmes et des ressources depuis un poste de travail standard (alors que les dossiers et archives papier n'étaient accessibles que dans un service unique spécialisé sur une tâche) ;
- la compétence est épaulée et contrôlée par les programmes informatiques (par exemple les systèmes d'assistance peuvent de manière experte apporter la connaissance d'une réglementation appliquée très rarement).

La diversité des compétences reste un argument d'actualité qui peut justifier une place à part pour les personnes disposant d'une compétence particulière (en général avec des diplômes) afin de les utiliser au mieux compte tenu de leur salaire plus élevé. Elle reste un principe de différenciation des tâches. Par exemple, restent spécialisées les tâches de signature par le chef hiérarchique, seule habilité à engager juridiquement l'entreprise.

Si les arguments de la division du travail cessent partiellement d'être vrais, il faut considérer les arguments inverses. Herzberg (1959) plaidait dès la fin des années 50 pour élargir les tâches et les enrichir, c'est-à-dire regrouper les tâches divisées au long de la séquence des opérations et inté-

grer dans l'exécution quelques tâches de contrôle ou fonctionnelles (petite maintenance par exemple). Mais son objectif premier était la motivation et ce n'est qu'au travers de celle-ci qu'il visait la productivité.

La division du travail a un inconvénient bien connu. Il faut assurer le déplacement des matières (ou des dossiers dans les services) d'un poste de travail à un autre. Avec la chaîne, Ford mécanise ce déplacement. Dans les services, on a mis en place des systèmes automatiques de transport des dossiers papiers, plus ou moins perfectionnés. En allant plus loin, le workflow (Gestion Electronique de Documents) transforme tous les dossiers en image informatique et les réseaux assurent leur communication. Ces systèmes automatisent le passage d'un poste de travail à un autre, donc réduisent les délais et suppriment les postes des personnes (en général peu formées) chargées de faire circuler les dossiers. Curieusement Hammer ne fait pas allusion au workflow, il préconise une solution plus radicale, la suppression de la division du travail.

Pour comprendre cette préconisation qui est tant à contre-courant, il faut faire un petit détour. On va examiner les organisations avec division du travail et on va les comparer à la solution de la polyvalence.

4. POLYVALENCE CONTRE DIVISION DU TRAVAIL

L'organisation traditionnelle de division du travail est celle de la chaîne (ou de la procédure). Les postes de travail disposés le long du processus traitent successivement les dossiers qui se sont pré-

sentés au premier poste. Le travail arrive de manière aléatoire. La durée de traitement par chaque poste de travail peut être aléatoire ou fixe. Si cette durée de traitement est aléatoire pour chaque poste, indépendamment des autres postes, il se crée une file d'attente devant chaque poste. C'est la solution C (voir figure 1).

Si la durée de traitement par chaque poste de travail est constante, il n'y a pas de file d'attente entre les postes de travail. C'est la solution E (voir figure 1). Cette solution se produit aussi si la durée de traitement est aléatoire pour le premier poste et rigoureusement égale pour les autres postes.

L'organisation polyvalente préconisée par Hammer est décrite dans la figure 1 (solution F). Les postes de travail se partagent une même file d'attente. Les cas s'orientent vers le poste de travail qui devient libre.

Pour montrer que la solution F (polyvalente) est meilleure en termes de délai et de coût que les solutions C et E (division du travail), trois méthodes sont possibles. On peut construire, à l'aide de la théorie des files d'attente, la relation analytique entre le coût et le délai de chaque solution. On peut aussi, par simulation, construire cette relation dans des cas précis, avec une valeur pour les différents paramètres. Il est enfin possible de raisonner qualitativement.

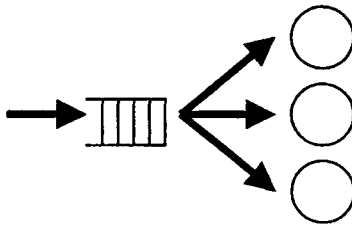
La méthode qualitative est plus agréable à la lecture. La contrainte de rigueur y est moins facile à respecter. Le résultat n'est pas quantifié. Ainsi un écart faible est repéré de la même manière qu'un écart fort.



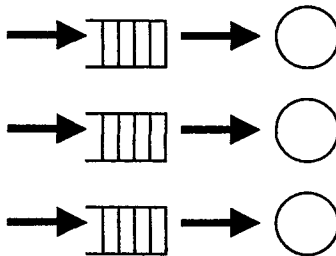
Division du travail avec file d'attente devant chaque poste (solution C)



Division du travail sans file d'attente devant chaque poste (solution E)



Polyvalence avec file d'attente unique devant tous les postes (solution F)



Polyvalence avec une file d'attente devant chaque poste (solution A)

NB : Les lettres (A, C, E, F) adoptées pour désigner les organisations sont celles utilisées dans une communication antérieure (Peaucelle, 1999a).

Figure 1 : Organisation polyvalente et organisation avec division du travail

La simulation a l'intérêt de donner des valeurs chiffrées dans des cas choisis, qui peuvent être très compliqués. Ces résultats, malheureusement, ne sont valables que pour les cas étudiés. Même si ceux-ci sont nombreux, on ne connaît pas la généralité des résultats.

La méthode analytique est possible, ici, avec des hypothèses relativement générales (Peauccelle, 1999a). C'est une application directe de la théorie des files d'attente, en la reformulant pour relier le délai au coût. Il n'y a pas de démonstration mathématique nouvelle. L'intérêt est alors de quantifier l'effet qui sera montré ci-dessous et de préciser les conditions dans lesquelles il est valable. Un modèle mathématique est toujours un résultat plus précis, et donc plus fécond pour les applications. Les personnes qui s'y intéressent trouveront les formules mathématiques sur le site Internet du GREGOR :

(<http://panoramix.univ-paris1.fr/GREGOR/99-06.htm>).

Il n'a pas paru bon de retenir cette approche pour le présent article. La méthode d'exposition sera donc qualitative. En effet, les résultats seront présentés selon un raisonnement en langue naturelle. Pour commencer rappelons les caractéristiques d'une file d'attente avec serveur unique. Nous en déduirons les propriétés des modèles à comparer. Cette organisation est désignée sous la lettre A dans la figure 1. C'est celle d'un poste de travail unique traitant les dossiers de bout en bout, avec sa propre file d'attente. Dès que le poste de travail a fini de travailler sur un cas, il traite le suivant dans la file d'attente. Si la file d'attente est vide, il reste inoccupé. Les cas

nouveaux se placent à la fin de la file d'attente.

Quand le flux de travail dépasse la capacité d'une personne, on installe autant de postes de travail qu'il le faut. Chacun dispose de sa propre file d'attente. Les postes de travail sont souvent spécialisés par territoire géographique (région) ou symbolique (numéros de dossiers, lettres de l'alphabet). On tente alors d'équilibrer les charges des postes de travail.

Cette solution est traitée par la théorie mathématique des files d'attente, classique en recherche opérationnelle. L'aléa d'arrivée des cas et de durée de traitement de chaque cas est exprimé par des lois exponentielles. En résultat, les cas sortent du traitement également selon une loi exponentielle. Les formules mathématiques figurent dans les ouvrages de recherche opérationnelle (par exemple Churchman, Ackoff, Arnoff, 1961). Ils expriment que :

- Pour atteindre un état stationnaire, il faut que la capacité de traitement soit supérieure au flux de cas arrivant.
- Le coût est d'autant plus élevé qu'on a mis en place des capacités excédentaires.
- Avec des capacités excédant de peu les flux à traiter, le coût tend vers un coût minimum correspondant à un traitement continu des dossiers, sans temps mort du serveur, avec une file d'attente infinie.
- Le délai est d'autant plus bref qu'il existe beaucoup de capacités excédentaires. Il tend vers un délai minimum qui est le temps de traitement lui-même.
- Le délai moyen est d'autant plus bref que le coût est élevé.

- Le rapport entre le délai d'attente et le délai de traitement dépend du taux d'occupation des personnes.
- Le délai d'exécution est composé du délai d'attente et du délai de traitement. Il est mesuré par l'inverse de la différence entre la capacité de travail et le flux de cas arrivant.

Cette solution simple (A) sert de référence. On va lui comparer les solutions avec division du travail, en supposant que la productivité reste toujours la même, avec ou sans division du travail. Le temps total pour traiter un cas par trois personnes successivement est le même que le temps mis par une personne polyvalente. Autrement dit, dans la procédure de 3 postes de travail successifs, chaque personne met, en moyenne, un tiers du temps total à traiter un cas. On suppose, de plus, que le salaire des personnes est égal. La personne polyvalente est payée le même prix que les personnes spécialisées. En outre, que le temps de transmission des dossiers entre postes de travail est considéré comme nul.

Dans ces conditions, la division du travail C est totalement équivalente à la solution A, avec trois postes traitant chacun un tiers du flux. Chaque poste de travail fait un tiers des tâches. Il les fait trois fois plus vite que dans la polyvalence A (hypothèse de constance de la productivité). Il fait face à un flux trois fois plus fort que dans la solution A. Le délai d'exécution (traitement plus attente), pour chaque poste, est donc du tiers du délai de la solution A. Ce délai, du tiers, se reproduit trois fois. Donc au total il est égal à celui de la solution A. Globalement, le temps de traite-

ment est donc le même dans les solutions C et A.

La division du travail E est plus performante. En effet, il n'y a plus qu'une seule file d'attente, celle qui est devant le premier poste. Comme le temps de travail du premier poste est réduit (un tiers pour une division du travail entre trois personnes), le temps d'attente dans cette file est réduit d'autant. Donc pour un même coût, le délai est moindre dans la solution E.

Maintenant montrons que la solution F de la file d'attente commune aux postes de travail polyvalents est la meilleure. Il faut la comparer à la solution E, meilleure que C et A.

Dans cette solution E, avec la division du travail, on peut avoir, au même moment, des personnes inoccupées et des attentes dans la file. Les arrivées sont aléatoires. A certains moments, il y a plus de cas que les personnes ne peuvent en traiter. Ces cas prennent place dans la file d'attente. A d'autres moments, peu de cas arrivent. La file d'attente se vide. Si la file d'attente est vide assez longtemps, la première personne reste inoccupée. Son inoccupation se transmet aux postes de travail suivants. Si, à ce moment-là, arrivent deux cas très rapprochés, le premier cas est pris en charge par le premier poste de travail et le deuxième cas attend dans la file. On a alors simultanément des personnes libres, en arrière guichet, et des cas qui attendent dans la file. C'est une source de gâchis qui diminue la performance de la solution E, en termes de coût et de délais.

Au contraire, cette situation n'existe pas dans la solution F. Il n'y a jamais des cas qui attendent

dans la file alors qu'une personne serait inoccupée. C'est impossible puisque la règle est que les personnes se saisissent des cas de la file dès qu'elles sont inoccupées. La polyvalence et la mise en commun des aléas minimisent les gâchis d'une simultanéité d'attentes des cas à traiter et de l'inoccupation des personnes chargées de les traiter.

Dans la théorie des files d'attente, on connaît bien la supériorité de la solution F par rapport à la solution A. La mise en commun des files d'attente permet de réduire l'effet relatif des aléas d'arrivée des cas et de temps de traitement. On a montré un résultat supplémentaire. La division du travail (solutions C ou E ou toute solution intermédiaire) a une performance comprise entre celles de ces deux solutions classiques A et F. La solution F est meilleure que toute solution avec division du travail, sous le critère de délai et de coût simultanément. Si on ne considère que le coût, toutes les solutions sont équivalentes (productivité prise constante par hypothèse).

5. INTERPRÉTATION DES PRINCIPES DU BPR

Après ce détour méthodologique par la théorie des files d'attente, il est possible de revenir aux principes de Hammer. Examinons successivement chacun d'entre eux.

5.1. Polyvalence

("to have one person perform all the steps in a process")

Ce principe s'oppose à une division du travail. Il préconise la solution F par rapport aux solutions C et E de division du travail. En termes de coût et de délais simul-

tanément, ce principe s'interprète facilement comme la dominance de la solution F qui a été montrée au paragraphe précédent.

Les hypothèses qui ont été faites permettent d'éclairer le principe de Hammer. La productivité est constante. S'il est possible de réaliser un gain de productivité considérable par une division du travail, il faut abandonner ce premier principe. Si, par ailleurs, les personnes polyvalentes demandent un salaire beaucoup plus élevé que les personnes spécialisées, ce principe est aussi inapplicable. Comme beaucoup d'organisateur, Hammer élude la question des salaires. Il suppose que les rémunérations restent constantes.

En concevant un poste qui assure toutes les tâches du processus, on change radicalement de modèle d'organisation et les coûts et délais associés sont meilleurs. Bien entendu ce n'est pas toujours possible.

En faveur du regroupement de tâches, on pourrait avancer d'autres arguments. Dans une procédure séquentielle, chaque bureau prend connaissance du dossier. Ce temps de lecture se répète. En supprimant la division du travail, on supprime cette duplication de la lecture. Il est donc possible que le temps total passé par la personne polyvalente pour traiter le dossier de bout en bout, soit inférieur au total des temps éclatés. Les comparaisons faites ci-avant ne prenaient pas en compte cet effet, pas plus que celui inverse d'accroissement de productivité par division du travail.

Dans tous les exemples donnés par les chercheurs ou les consultants, on note la diminution du nombre d'étapes dans les procédures. Par exemple Hammer

(1993) montre que la compagnie de téléphone Bell Atlantic a raccourci le délai de raccordement des nouveaux abonnés de 15 jours à 5 heures en constituant des équipes de 5 à 6 personnes à compétence transfonctionnelle dans un seul lieu, pour toute la chaîne du travail effectuée auparavant par 13 personnes en séquence, installées dans des lieux différents.

Faire traiter les cas de bout en bout par la même personne suppose des outils informatiques puissants pour qu'elle dispose, sur son poste de travail, de tous les éléments du dossier qui est traité à ce moment, quel que soit ce dossier. Ce ne serait pas possible avec des archives papier. L'informatique aide à rendre possible la solution F, comme le dit le deuxième principe.

5.2. Fournir les informations et des aides informatiques partout ("computer-based data and expertise are more readily available")

Le deuxième principe signalé par Hammer (1990) n'est qu'une condition pour que le premier soit réalisable. Il s'agit de fournir toute l'information à la personne qui traite le dossier. L'accès aux bases de données, la construction de systèmes experts spécialisés, la fourniture de documentation en ligne sont des moyens pour rendre compétentes des personnes, sans formation ni mémoire sur le cas qui se présente.

L'organisation à laquelle ce principe s'oppose est celle de bureaux spécialisés (services fonctionnels) qui ont les dossiers, les compétences et l'expertise pour accomplir cette partie de la tâche. C'est

une division du travail qui est liée à la structure fonctionnelle de l'organisation. On a déjà vu qu'elle n'est pas performante.

Les aides informatiques au travail permettent aux opérationnels de faire tout seuls, sans attente, sans plus d'erreurs, ce qu'ils devaient demander aux postes spécialisés (comptabilité, achats...). La division du travail ne peut être abandonnée au profit de tâches globalisées que si ces traitements de l'information sont accessibles aux opérationnels. Si la mise en place de ces outils d'aide n'est que progressive, les transformations organisationnelles suivent ce rythme. Les services informatiques sont ainsi des partenaires obligés de toutes les actions de BPR.

La perspective de l'informatisation associée au BPR n'est donc pas de remplacer les personnes dans leurs tâches mais de déplacer le lieu de traitement. Du service spécialisé, il migre vers les personnes polyvalentes qui traitent les cas de bout en bout. C'est en ce lieu que sont fournies les informations, au moment où les gens en ont besoin. L'informatique communique et suit le rythme de travail des personnes.

Lucas, Berndt et Truman (1996) donnent un exemple de reconfiguration du processus de traitement des valeurs sur support papier déposées par les clients aux agences de Merrill Lynch. Ces valeurs sont envoyées par la poste au centre unique de traitement. Une part du temps était consommée en aller-retour entre le centre et les agences pour des anomalies (documents annexes non fournis, valeurs n'ayant pas cours légal...). Un système expert, accessible dans les agences, permet de consulter toutes les valeurs en circu-

lation et de décrire tous les cas où des documents annexes sont nécessaires (certificat de décès en cas d'héritage par exemple). On réduit alors à la fois le temps passé par les agents (donc le coût) et le délai pour le client (à cause des allers-retours). Des solutions de workflow ont aussi été mises en œuvre. Les valeurs sont scannées (ou microfilmées si elles sont de format trop important). Une part du traitement est faite après reconnaissance optique du document s'appuyant sur une base de données des "valeurs papier" en circulation.

Cette suppression des services spécialisés est radicale. Elle est efficace s'il n'y a pas d'effet de productivité. Avec les aides informatiques, les personnes polyvalentes effectuent le travail, à peu près dans le même temps (avec la même qualité) que les personnes spécialisées. C'est souvent vrai pour les cas simples. Champy et Hammer (1993) amoindriront ce principe en conservant des personnes spécialisées pour traiter les cas complexes.

L'effet de salaire, implicite pour Hammer, joue en faveur du deuxième principe. Les opérationnels sont souvent moins payés que les personnes spécialisées des services fonctionnels dont elles vont accomplir les tâches.

5.3. Intégration du système d'information au monde réel ("subsume information-processing work into the real work that produces the information")

Traditionnellement les activités administratives sont effectuées par des employés spécialisés, distincts des ouvriers productifs. Il s'agit de la distinction bien connue des

cols blancs et des cols bleus. Les personnels des ateliers ne sont pas amenés à tenir des crayons et des papiers. Quand ils le font, on sait que la qualité de leur écriture n'est pas fameuse.

L'informatique renouvelle cette approche. On peut par des terminaux, spécialisés ou non, réaliser la saisie des informations par les personnes en contact le plus direct avec le "monde réel", celui sur lequel portent les informations. Avec cette intégration, quand elle est possible, on supprime les postes administratifs.

La saisie des données par les productifs, par les clients, par les demandeurs exige des systèmes informatiques spécialisés, sécurisés, conviviaux. Ici encore la réalisation informatique est une partie essentielle du BPR.

En intégrant système d'information et monde réel, on supprime des étapes dans une procédure, par automatisation. Au bout de ce principe, on supprime tout traitement manuel de l'information. Après la saisie par le client, toutes les tâches sont prises en charge par la machine.

Cette perspective de totale automatisation est loin d'être neuve. Dans le cas de Bell Atlantic, rapporté par Hammer (1993) et cité plus haut, on veut aller vers un délai zéro d'installation du service téléphonique. Le client saisira lui-même sa demande (par Internet) et ce sont des programmes qui déclencheront les actions sur les centraux téléphoniques pour que le système soit mis à disposition. C'est possible si l'infrastructure, lignes téléphoniques notamment, est déjà en place. La partie administrative, signature de contrat, acceptation du tarif, sera prise en

charge après la mise à disposition du service.

5.4. Centralisation virtuelle ("treat geographically dispersed resources as though they were centralized")

Le quatrième principe concerne les grandes entreprises qui disposent de localisations géographiques sur un territoire : agences, bureaux, boutiques... Si on les regroupe, on bénéficie des économies d'échelle qu'il y a entre le modèle A répété et la mise en commun de la file d'attente (F). On évite ce gâchis d'un serveur inoccupé à un endroit alors qu'il existe des dossiers en attente dans un autre endroit.

L'avantage supplémentaire de la centralisation vient de l'ajustement plus facile aux flux. Les bureaux répartis n'ont jamais une charge complètement égale. Même si au moment de leur création, leurs charges sont similaires. Les évolutions créent des distorsions entre bureaux locaux. La mise en commun de la charge de travail évite l'existence de surcharges locales, avec des charges très faibles en d'autres lieux.

Ces économies d'échelle face aux aléas de l'arrivée des dossiers sont les mêmes que celles entraînées par la centralisation des stocks. Le stock centralisé évite les situations de rupture dans un dépôt local avec des marchandises dans d'autres dépôts.

Bien sûr, il faut que ce regroupement soit possible, c'est-à-dire que le service soit offert localement malgré le regroupement géographique. Les technologies de la communication rendent l'accès à distance possible. Les exemples sont bien connus avec les centres

d'appels qu'on retrouve ici comme un résultat du BPR.

Mais la technologie va plus loin. Il n'est pas nécessaire de déplacer physiquement les personnes en un seul lieu. Puisque leur travail est délocalisé par rapport à la demande, il peut l'être par rapport à l'équipe des personnes substituables. La délocalisation du travail rend possible la constitution d'un service unique sans déplacer les personnes. La centralisation géographique peut être virtuelle. La charge de travail est mutualisée entre tous les serveurs. Dès qu'un dossier attendrait sur son site local, il est proposé aux agences éloignées géographiquement qui sont libres à ce moment. Les centres locaux s'échangent la charge de travail. C'est le regroupement en équipes virtuelles préconisé par Hammer.

Le regroupement permet aussi d'ajuster les effectifs à la saisonnalité de la charge de travail : variation dans la journée, dans la semaine, dans l'année. Quand les effectifs sont limités, voire une seule personne, il n'y a pas de modulation possible. Quand les effectifs sont importants, cet ajustement limite le surcoût découlant d'effectifs pléthoriques à certains moments et les délais d'attente des dossiers lors des périodes de pointe.

Wareham et Neergaard (1998) rapportent le cas de la réorganisation de la compagnie d'assurance mutuelle Alka au Danemark. La réorganisation a eu comme principe directeur le "guichet unique". Chaque client doit avoir un seul point de contact avec Alka, directement par téléphone, pour gérer son contrat (ventes) et un autre pour gérer son sinistre (déclaration, remboursement...). Antérieu-

rement il devait s'adresser successivement par courrier aux divers services spécialisés. A la place de 26 agences locales, ce sont 6 agences régionales puis trois qui assurent les contacts commerciaux. Les auteurs n'indiquent pas les effectifs consacrés à la vente. Ils disent que 56 postes de travail ont été supprimés (sur un effectif total de 375 personnes). Et la compétence (polyvalence) du personnel en place a été largement augmentée. Les procédures, derrière le guichet unique, continuent d'exister. Elles ont été simplifiées avec un maximum de 69 étapes réduit à 30 pour le commercial et, pour les sinistres, on passe de 193 étapes au maximum à 44 seulement. Le délai moyen passe de 32 jours à 6 jours pour régler un sinistre.

5.5. Mise en parallèle des activités
 ("link parallel activities instead of integrating their results")

Ce principe s'oppose à la structure traditionnelle des procédures qui est la séquence, la succession d'étapes. On copie ainsi la succession des postes de travail de la chaîne. Mais, même sur la chaîne des usines, il arrive que deux opérations soient simultanées, par exemple quand elles portent sur des sous-ensembles différents avant leur montage. Pour le traitement d'un dossier, certaines étapes peuvent se dérouler simultanément, d'autant plus aisément que le dossier est dématérialisé. Il est clair que les délais de chaque étape ne s'ajoutent plus. On gagne en délais.

Ce principe de simultanéité des étapes est celui qu'on trouve en production pour le montage des outils et le réglage des machines en "temps caché" (pour réduire les

temps d'arrêt entre deux séries de production). Pour gagner du temps, on commence une étape avant que la précédente soit terminée.

Avec des activités exécutées en parallèle, on retrouve une division du travail qui peut être meilleure que le poste polyvalent (F). C'est le gain principal des applications de Gestion Electronique de Documents (GED). La dématérialisation des documents permet de revoir la séquentialité des procédures. A chaque fois que le parallélisme est instauré, on gagne en délais et il est possible, éventuellement, d'aller plus vite que les postes polyvalents. Il est tout à fait intéressant de remarquer que Hammer n'a pas considéré ce principe comme le premier à exposer. Ce n'est donc pas sans doute pour lui le principal moyen d'améliorer la performance.

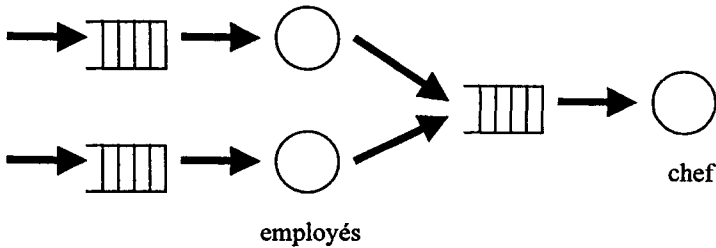
5.6. Décentralisation de la prise de décision et contrôle sur les résultats
 ("put the decision point where the works is performed and build control into process")

Dans son sixième principe, Hammer s'oppose à la division verticale du travail, entre les exécutants et la hiérarchie. Le traitement d'un cas par les opérationnels fait parfois intervenir des niveaux plus élevés, pour des décisions, des autorisations, des signatures officielles. Cet appel à la hiérarchie est tout à fait similaire à la division fonctionnelle du travail que Hammer veut éviter (deuxième principe).

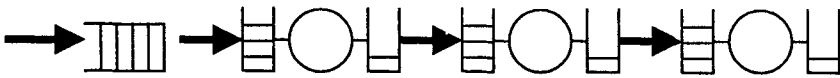
Plusieurs subordonnés traitent des cas et les font vérifier ou contresigner par un chef hiérarchique (voir solution B, figure 2). Il y a une file d'attente devant

chaque agent et devant le chef. La modélisation mathématique des coûts et des délais de cette solution traditionnelle de la hiérarchie a été faite par Beckmann (1988). De manière qualitative, examinons le cas où le chef n'est pas plus payé que ses subordonnés. Supposons en outre qu'il n'y ait pas d'effet de productivité. Le temps de travail de l'employé polyvalent sur un cas (solution F ou A) est partagé entre un

temps de travail par les subordonnés et puis en temps de travail par le chef. Avec les hypothèses classiques d'une arrivée aléatoire des cas et d'un temps de traitement aléatoire, les temps d'attente dans les deux files ont un total égal à l'attente dans la solution A. Avec ces hypothèses, la performance de la solution hiérarchique B est la même que celle de la solution A de polyvalence.



Hiérarchie à deux niveaux (solution B)



Division du travail avec transfert par lots (solution D)

Figure 2 : Organisation avec division du travail entre un chef et ses subordonnés et organisation du travail par lots (batch processing)

Ainsi, la solution B d'une hiérarchie est moins bonne en termes de délais et de coûts que la solution F de postes de travail polyvalents. Si, de plus, on tient compte de l'effet des salaires, cette supériorité s'accroît. Le chef est plus payé que les employés polyvalents. Son remplacement par un employé (pour conserver la charge de travail) permet de réduire les coûts.

Le rôle de la hiérarchie est joué par les outils informatiques qui

encadrent les actions et enregistrent ce qui a été fait. Les personnes sont alors responsables de leurs actes. Toute hiérarchie n'est pas supprimée. Il reste des personnes qui interviennent *a posteriori* en contrôlant les actions enregistrées.

Hammer préconise donc une décentralisation de la décision et du contrôle au point le plus bas, le plus proche du processus du monde réel. La hiérarchie peut ainsi presque complètement dispa-

raître. Ce sont des coûts qui disparaissent, en même temps que les délais sont diminués.

5.7. Saisie unique à la source ("capture information once and at the source")

Le septième principe de Hammer concerne la saisie unique à la source. Il s'agit encore de la suppression d'étapes dans le processus de traitement informationnel. Ce principe a été énoncé depuis longtemps pour la conception des systèmes d'information. Il a pour but d'automatiser une partie du processus informationnel, donc de supprimer du traitement administratif manuel, comme le principe 3 dont au fond il ne diffère pas.

6. LA COHÉRENCE DES PRINCIPES DE HAMMER

Quatre des sept principes de Hammer s'interprètent précisément comme un changement vers la solution la meilleure en termes de coût et de délai, celle de serveurs suivant de bout en bout le traitement des dossiers et partageant une file d'attente commune (F), c'est-à-dire les centres d'appels à un seul niveau avec des agents polyvalents. Le principe 5, préconisant une mise en parallèle des activités, est de même nature. Au lieu de regrouper les tâches sur une seule personne, on gagne encore en délai si on fait exécuter plusieurs morceaux de la tâche globale simultanément. Les principes 3 et 7 (intégration du SI au monde réel et saisie unique à la source) sont au contraire animés d'un esprit très classique d'automatisation. Là où il y a un travail humain, parfois la machine peut l'accomplir. Dans ces conditions

les délais et les coûts sont ceux de l'informatique.

Cette analyse laisse percevoir deux aspects complémentaires de la démarche du BPR. D'une part on réorganise vers une exécution des tâches globalement, de bout en bout (ou on parallélise les tâches découpées). D'autre part, on informatise. Cette informatisation possède son propre but de substituer la machine aux personnes dans le traitement des dossiers (automatisation), mais elle est aussi la condition pour rendre possible la réorganisation du travail humain. L'un et l'autre buts vont de pair. Les exemples décrits par Hammer ne sont pas suffisamment précis pour distinguer ces deux effets.

Le changement déclenché par le BPR est dit "radical". Le sens de cette expression est rarement précisé. Ce caractère radical vient probablement d'un retour sur la division du travail, d'une remise en cause d'un principe couramment admis et qui est contradictoire avec les objectifs de performance en coût et en délai, face à une arrivée aléatoire du travail. On réussit à gagner en délai, donc en productivité organisationnelle, sans modifier les temps passés sur chaque cas. En mettant en commun l'aléa de l'arrivée du travail, la solution de polyvalence F accroît les performances. L'intensification du travail est en revanche perçue par les agents sous la forme d'une réduction des temps de poses occasionnelles, causées par l'absence de travail à faire (file d'attente vide).

Cette analyse omet un mode de travail très fréquent dans les services administratifs, le traitement par *lots*. Ce mode de travail est représenté dans la figure 2 (solu-

tion D). Il y a en général une division du travail. Les cas sont d'abord stockés dans une première file d'attente. Avant le traitement par le premier poste, un lot est constitué, de taille fixe ou correspondant à tous les cas arrivés dans la file (période de traitement fixe, quotidienne, mensuelle, annuelle...). Une fois le lot constitué, le premier poste de travail le traite en entier et le passe au second poste de travail. Le travail se déroule ainsi avec un transfert des cas globalement, pour le lot entier.

Le traitement par lots est une très mauvaise solution en termes de coûts et de délais. La raison est bien connue des organisateurs d'usine. La taille du lot de transfert influe fortement sur les délais. En effet les éléments d'un lot attendent à chaque poste de travail que les autres éléments soient traités. Plus il y a de dossiers dans le lot, plus le délai est long. La taille du lot dépend de la périodicité (à même flux d'arrivée). La périodicité est donc le facteur dominant de l'explication du délai. Cette solution s'améliore avec une accélération de la périodicité. Elle reste toujours moins bonne que le traitement au cas par cas (au fil de l'eau) sur un serveur unique.

Par opposition, il est facile de comprendre que Hammer, implicitement, parle de processus de travail opérant "*au fil de l'eau*". Il est facile d'améliorer le délai d'un processus de traitement par lots en accélérant sa périodicité. Par exemple, on réduit toujours le délai de fourniture des informations comptables en passant d'une périodicité annuelle à un traitement mensuel.

Il est encore plus performant d'effectuer les traitements au fil de l'eau, quand cela est possible. Cependant, il arrive que les procé-

dures par lots ne puissent pas être abandonnées. C'est le cas quand le traitement de chaque dossier interfère avec celui des autres dossiers (classement, répartition de ressource limitée, quotas, concours, budget, postes...) ou lorsqu'on veut traiter la totalité des éléments d'un ensemble (période comptable, territoire géographique).

Le traitement d'un lot est performant si le temps de traitement contient une forte partie fixe, indépendante de la taille du lot. Malgré les délais longs, le traitement par lots reste alors meilleur en termes de coût de revient et de délai parce qu'il économise la répétition des coûts fixes à chaque cas. C'est la problématique des "séries économiques" en gestion de production.

Pour comparer les solutions, les raisonnements de cet article ont été qualitatifs. L'argument principal a consisté à identifier des attentes concomitantes avec une inoccupation de personnes. Avec des hypothèses simplificatrices, un raisonnement quantitatif précis aurait été possible, en utilisant les résultats de la théorie mathématique des files d'attente. Avec les hypothèses classiques de la théorie des files d'attente (voir Peaucelle 1999a), un exemple permet de jauger quantitativement les écarts entre solutions.

Considérons un flux de travail de 45 dossiers à traiter par heure. Supposons que les personnes passent 4 minutes en moyenne pour traiter un dossier, soit une capacité de travail individuelle de 15 dossiers par heure. Supposons que les six solutions, A, B, C, D, E, F, avec 6 personnes, soient possibles. La capacité de traitement globale est de 90 dossiers (6×15) par heure. Le coût

Charge de travail moyenne	45 dossiers par heure	84,1 dossiers par heure donc un coût par dossier moindre de 46,5 %
Solution	Délai moyen de traitement	Délai moyen de traitement
A, B et C	8 minutes	61 minutes
D (lots de 1 heure)	3,5 heures	6,1 heures
E	4,7 minutes	13,5 minutes
F	4,13 minutes	8,25 minutes

Tableau 1 : Exemples de coût et de délais dans les 6 solutions

est le même pour les six solutions. Le délai d'attente moyen dans chacune est donné dans le tableau 1. On a complété les calculs pour une charge de travail plus grande qui permet de diminuer le coût de traitement par dossier.

Si on part d'une solution telle que A, B ou C qui sont équivalentes, on améliore le processus en passant à la solution F. Le gain, à coût constant, s'exprime en diminution du délai de 48 %. En acceptant une augmentation de la charge, le gain peut être au contraire de conserver le délai avec un coût moindre de 46,5 %. Le changement réel peut se situer entre ces deux extrêmes.

Ces calculs sont faits avec des délais de transmission ayant une valeur nulle. Si ces délais ne sont pas nuls, les solutions avec une division du travail sont encore moins bonnes.

Peut-être est-il utile de revenir sur une hypothèse évoquée rapidement au début de cet article. On a considéré les travaux du tertiaire répétitifs, ceux pour lesquels des procédures fixent l'organisation du travail. Dans la revue SIM (Peaucele, 1998), on a déjà présenté le cas du travail pour lequel il n'existe pas de procédures,

mais où un ensemble déterminé de personnes interviennent, par des réunions ou par des échanges désynchronisés sur le réseau. Les outils de travail collaboratif asynchrone sur le réseau (groupware) réduisent le délai d'achèvement du travail collectif, par rapport aux réunions en face-à-face.

Cette analyse confirme le rôle des technologies de l'information dans une démarche de BPR. L'informatique est un passage obligé de la réorganisation. Mais l'essentiel réside dans la réorganisation. Comme disent Davenport et Stoddard (1994), considérer l'informatique comme le moteur du BPR est un mythe. En revanche, elle peut être un inhibiteur qui empêche le changement à cause de retards ou d'impossibilités à construire les nouveaux systèmes. Si le projet de reconfiguration est géré par les informaticiens, il risque sans doute de ne pas saisir la complexité du changement organisationnel. L'échec potentiel dépend alors d'un pilotage insuffisant sur les aspects organisationnels du changement : identification des gains, négociation de leur redistribution partielle aux personnels, qualité du travail découlant d'une relation, non formalisée, des agents avec l'environnement auquel ils sont habitués.

7. SIMULER LES RÉSULTATS DE LA RÉORGANISATION

Les gains de la réorganisation recherchés dans le BPR se mesurent, au moins, en termes de réduction des délais et des coûts. Pour connaître ces gains avant de déclencher le changement, il faut étudier le fonctionnement de la nouvelle organisation. Parfois, il est possible de construire un modèle analytique de la situation avec les modèles des files d'attente. Le plus souvent la situation réelle est suffisamment compliquée pour qu'il faille mettre en œuvre des outils de simulation. On génère une arrivée aléatoire des cas et les conditions de leur traitement dans chaque solution. La simulation permet de calculer les performances de délai et de coût de chaque organisation envisagée. Avant d'entreprendre le changement, on connaît les impacts. Cette approche est déjà suivie par quelques chercheurs et praticiens.

Kim et Kim (1997) modélisent le trajet des malades en consultation à l'hôpital adventiste de Séoul. Les étapes administratives (enregistrement et paiement) sont obligatoires. Puis les malades sont vus par le médecin. Ensuite ils peuvent répéter plusieurs fois des étapes d'examen ou analyses (prépayés) ou délivrance de médicaments (prépayés aussi). Les chercheurs ont observé les lois d'arrivée des patients, les temps de prestations des diverses étapes. La simulation et les observations convergent sur le temps passé par les patients en consultation à l'hôpital : une heure et demie. La simulation permet alors de tester d'autres hypothèses organisationnelles : intégration de l'enregistrement et du paiement, automa-

tisation des tâches administratives et réduction des personnels administratifs, augmentation du nombre de médecins. Le processus ainsi redéfini traite les patients en 20 minutes en moyenne, d'après la simulation. Gain considérable de temps avec des dépenses similaires dont l'article ne parle qu'au travers du taux d'utilisation des personnes. La simulation est un support fort de l'évaluation quantitative d'une réorganisation projetée.

Kumar, Ow et Prietula (1993) adoptent la même méthodologie pour traiter un autre cas d'organisation d'hôpital. Ils s'intéressent à la fixation des horaires des analyses médicales par les équipements spécialisés (radio, IRM, scanner...). C'est un problème d'ordonnement avec des contraintes similaires à celles de la production industrielle. Ils considèrent une gestion centralisée ou non des services (où séjournent les patients) et une gestion centralisée ou non des laboratoires. Ils simulent les quatre situations avec différents flux de patients. Ils abordent ainsi, par la simulation, le problème classique des théories organisationnelles du lien entre structure et technologie (informatique dans ce cas).

La simulation du processus renforce sans doute les chances de succès d'une réorganisation. Elle facilite l'identification des gains potentiels de diverses solutions alternatives. Avec les principes de Hammer et par tâtonnement, on découvre la manière de s'organiser pour avoir le moins possible d'attentes et de personnes inoccupées, malgré les aléas du travail à accomplir. Les logiciels spécialisés de simulation de processus facilitent aujourd'hui cette approche.

8. CONSÉQUENCES PRATIQUES

Ces réflexions donnent une valeur supplémentaire aux principes du BPR. Elles permettent de mieux raisonner sur leur application. Toutes les personnes qui veulent améliorer les processus administratifs peuvent en retirer des conséquences pratiques.

a) Les performances attendues du processus administratif doivent d'abord être identifiées. Il semble qu'aujourd'hui on recherche simultanément des *coûts* et des *délais* réduits, tout en faisant face à des travaux qui surviennent de manière aléatoire. Ces deux dimensions de la performance des processus administratifs sont généralement recherchées. Dans chaque processus, il faut vérifier si ces deux dimensions de la performance sont importantes. Rechercher simultanément un coût faible et des délais réduits conduit à une organisation des processus différente de celle correspondant à un objectif de faible coût.

b) La productivité du travail peut être mesurée en tant que temps passé, en moyenne, par cas traité. C'est le rythme de travail. La *productivité organisationnelle* du travail est aussi le nombre moyen de cas traités par une équipe de personnes. Ces deux définitions diffèrent par l'organisation du travail face aux aléas, par les *temps morts*. On peut augmenter cette productivité organisationnelle sans modifier le rythme de travail, en diminuant les temps morts. L'entreprise cherche à améliorer cette productivité organisationnelle, car c'est elle qui est liée à la performance économique.

c) Pour une organisation donnée, face à une arrivée aléatoire du travail, les délais courts s'obtiennent avec des coûts élevés. *Réduire*

les coûts augmente les délais. Il y a un arbitrage entre les coûts et les délais.

d) Les manières d'organiser les processus administratifs peuvent se classer sur le double critère de délai et de coût. En changeant l'organisation, on peut donc améliorer le délai à coût constant, réduire le coût à délai constant, voire réduire simultanément le coût et le délai. Le classement est indépendant de l'arbitrage entre les coûts et les délais.

e) Le travail peut être organisé en traitant les cas un à un (au fil de l'eau) ou par lots constitués d'un ensemble de cas. Une procédure qui prévoit de constituer des *lots* est toujours *mauvaise* dans la double perspective du coût et du délai. Elle ne se justifie que s'il existe, dans le traitement des cas, un coût fixe indépendant du nombre de cas traités. C'est le concept bien connu en production de "série économique". A chaque fois qu'on réduit ce coût fixe, on réduit la taille de la série et on améliore les délais et les coûts. La réduction de la taille des lots améliore les délais. Dans le travail administratif, ce sont parfois des considérations de logique de traitement qui justifient la notion de lot. Il faut avoir exhaustivement tous les cas d'une période pour établir des comptes, des statistiques, des classements.

f) La *division du travail* dans les procédures où les bureaux fonctionnels spécialisés traitent un aspect des dossiers n'est pas performante en termes de coût et de délai simultanément, si les facteurs de productivité, justifiant classiquement la division du travail, ne jouent pas ou peu. Il faut se souvenir que la division du travail a un intérêt si la producti-

tivité en est augmentée (habileté des agents sur une tâche élémentaire), si les salaires sont très différents (allègement de la charge des personnes à hauts salaires par des assistants moins payés), si la compétence mise en œuvre est si spécifique que peu de personnes l'ont acquise ou en disposent institutionnellement.

g) S'interroger sur les contraintes à la division du travail et tenter de les réduire en introduisant une *polyvalence* permet d'améliorer les processus en termes de délais et de coûts.

h) La mise en commun des cas d'un territoire dans une file d'attente commune à tous les postes de travail améliore les coûts en même temps que les délais. Les *centres d'appels* sont construits sur ce principe. Cette solution, de plus, offre une meilleure flexibilité.

i) Les technologies de l'information sont une aide fondamentale des processus administratifs, une *aide* pour l'exécution des tâches, une aide pour lever les contraintes de *territoire* et *communiquer* l'information, une aide pour *mettre en commun* toute l'information mémorisée, une aide pour *assister l'exécution* de procédures et *contrôler* l'application de règlements, une aide pour assister la *compétence* des personnes. Ces aides suppriment des contraintes qui paraissaient antérieurement inscrites au cœur de la nature des processus eux-mêmes. Une approche globale permet d'améliorer leurs performances, en termes de coût, de délai mais aussi bien souvent en termes de qualité et de flexibilité.

j) Les technologies de l'information sont au cœur de la *reconception des processus*. Le rythme de leur déploiement et la qualité

des solutions techniques conditionnent les succès.

k) Si on veut améliorer un processus administratif en termes de coût et de délai, avec une arrivée aléatoire des cas à traiter, l'intuition ne suffit pas. Les organisateurs doivent étudier soigneusement les lois d'arrivée des cas, leur saisonnalité notamment. Le processus actuel doit être comparé aux résultats de sa *simulation* sur ordinateur. Les solutions envisagées doivent être, de même, étudiées en termes de performance (coût et délai) sur le simulateur. La robustesse aux variations de charge doit aussi être examinée (flexibilité). Si on veut améliorer radicalement les processus administratifs, ces techniques, qui sont celles de l'étude des processus industriels, deviennent pertinentes.

BIBLIOGRAPHIE

Beckmann, M.J. (1988), *Tinbergen Lectures on Organization Theory*, Springer Verlag, Berlin, 252p.

Churchman, C.W., Ackoff, R.L., Arnoff, E.L. (1961), *Éléments de recherche opérationnelle*, Traduit et adapté en français par Jean Lavault, Pierre Rosenstiehl, Patrice Bertier, Bernard Roy, Jacques de Guenin et Francis Piquemal, Dunod.

Davenport, T.E., Stoddard, D.B. (1994), « Reengineering : business change of mythic proportions ? », *MIS Quarterly*, June, p. 121-127.

Davenport, T.H., Beers, M.C. (1995), « Managing information about processes », *Journal of Management Information Systems*, Summer, Vol. 12, n° 1, p. 57-80.

Grover, V., Jeong, S.R., Kettinger, W.J., Teng, J.T.C. (1995), « The implementation of Business Process Reengineering », *Journal of Management Information Systems*, Summer, Vol. 12, n° 1, p. 109-144.

Hall, G., Rosenthal, J., Wade, J. (1993), « How to make reengineering really work », *Harvard Business Review*, Vol. 71, n° 6, p. 119-131.

Hammer, M. (1990), « Reengineering work: don't automate, obliterate », *Harvard Business Review*, July-August, p. 104-114.

Hammer, M., Champy, J. (1993), *Reengineering the Corporation*, Harper Business, NY.

Hammer, M., Stanton, S.A. (1995), *The Reengineering Revolution*, Harper Collins, NY.

Hammer, M. (1996), *Beyond Reengineering*, Harper Collins, NY.

Herzberg, F., Mausner, B., Snyderman, B. (1959), *The motivation to work*, John Wiley.

Kawalek, J.P. (1994), « Interpreting business process re-engineering on organization work-flow », *Journal of Information Technology*, n° 9, p. 276-287.

Kettinger, W.J., Grover, V. (1995), « Toward a Theory of Business Process Change Management », *Journal of Management Information Systems*, Summer, Vol. 12, n° 1, p. 9-30, June, p. 223-240.

Kim, H.W., Kim, Y.G. (1997), « Dynamic process modeling for BPR: A computerized simulation approach », *Information & Management*, n° 32, p. 1-13.

Kumar, A., Ow, P.S., Prietula, M.J. (1993), « Organizational simulation and information systems design: an operations level example », *Management Science*, February, Vol. 39, n° 2, p. 218-240.

Loch, C. (1998), « Operations Management and Reengineering », *European*

Management Journal, Vol. 16, n° 3, p. 306-317.

Loebbecke, C., Jelassi, T. (1998), « Business process redesign at CompuNet, standardizing top-quality service through IT », *Journal of Strategic Information Systems*, Vol. 6, p. 339-359.

Lucas, H.C.Jr, Berndt, D.J., Truman, G. (1996), « A reengineering framework for evaluating a financial imaging system », *Communications of ACM*, Vol. 39, n° 5, p. 86-96.

Peaucelle, J.-L. (1998), « Fixer une réunion ou travailler ensemble sur le réseau : comparaison des délais d'achèvement », *Systèmes d'Information et de Management*, Vol. 3, n° 3, p. 29-47.

Peaucelle, J.-L. (1999a), Hammer avait raison ! Les conditions d'efficacité du BPR, 4^e Colloque de l'AIM, mai, Cergy-Pontoise.

Peaucelle, J.-L. (1999b), « La division du travail : Adam Smith et les encyclopédistes observant la fabrication des épingles en Normandie », *Gérer et Comprendre*, n° 57, p. 36-51.

Smith, A. (1776), *Inquiry into the nature and causes of the Wealth of Nations*, London, 3 volumes.

Sillince, J.A.A., Harindranath, G. (1998), « Integration of requirements determination and business process re-engineering: a case study of an ambulatory care and diagnostic (ACAD) centre », *European Journal of Information Systems*, n° 7, p. 115-122.

Wareham, J., Neergaard, P. (1998), Still guilty of technical determinism ? Reengineering in the insurance sector, 6th European Conference on Information Systems, Aix, France, 4-6 June, p. 1291-1304.