

L'USAGE DES BANQUES DE DONNÉES À L'INSTITUT NATIONAL DE RECHERCHE PÉDAGOGIQUE

Problématique et réalisations

Philippe Champy

Ressources fondamentales dans l'organisation des nouveaux systèmes d'information, les banques de données ont réalisé, grâce à la télématique et au minitel, une indéniable percée dans le discours de la communication. Présentées comme les gigantesques réservoirs d'un savoir accumulé, leur évocation donne souvent du crédit au mythe d'une possible transmission mécanique des connaissances.

L'opinion selon laquelle elles peuvent révolutionner non seulement l'accès à l'information mais aussi la diffusion des savoirs mérite d'être nuancée. En exposant la problématique et les réalisations de l'INRP en ce domaine, notre but est de préciser la place qu'occupent ces nouveaux outils eu égard aux missions d'information de ce centre de ressources spécialisé et à ses missions de recherche.

1. L'univers des possibles

Avant d'entrer dans le vif du sujet, il nous paraît nécessaire de rappeler succinctement en quoi consistent les banques de données et quel

Communication documentaire

Perspectives documentaires en éducation, n° 23, 1991

univers de possibilités elles ouvrent à ceux qui les produisent comme à ceux qui les utilisent.

1.1. Des banques de données polymorphes...

Une banque de données est un ensemble de fichiers informatisés dont les connexions et la gestion assurent trois types d'opération élémentaire : a) le stockage des données, b) leur recherche rétrospective et c) leur rediffusion ou édition. Ces trois facettes fonctionnelles peuvent être adoptées comme des points d'observation pour rendre compte des réalisations obtenues grâce à l'informatisation.

À la première fonctionnalité correspond l'image des banques de données comme **gisements d'informations** traitées et organisées méthodiquement pour une mise à jour et une exploitation ultérieures (point de vue du producteur). À la deuxième fonctionnalité correspond l'image des banques de données comme **outils de recherche**, consultables localement ou à distance, à partir de critères multiples liés à la nature des données elles-mêmes (point de vue des utilisateurs). Enfin, à la troisième fonctionnalité correspond l'image des banques de données comme **sources éditoriales** permettant la diffusion multimédia d'informations contrôlées (point de vue de l'éditeur).

Chaque point de vue met en lumière un aspect de la réalité des banques de données. Mais ce n'est que la combinatoire de ces différents points de vue qui permet d'avoir une représentation d'ensemble donnant la mesure véritable des possibilités offertes aux producteurs comme aux utilisateurs.

Il arrive que les producteurs de banques de données minorent, lorsqu'ils ne l'ignorent pas, la troisième facette. Ils mènent souvent une politique éditoriale implicite en la présentant, conformément à l'usage documentaire traditionnel, sous les traits d'une politique de « sous-produits » ou de « services annexes », alors que les productions ou services liés à la rediffusion constituent une part importante du chiffre d'affaire global. Selon une opinion qui avait cours au milieu des années 70, l'accès en ligne à distance est encore conçu comme la finalité première, quasi constitutive de la notion de banque de données, les autres modalités de diffusion n'étant considérées que comme dérivées ou connexes.

Ce point de vue mérite d'être modifié pour deux raisons. D'abord, en réduisant les banques de données à ce qui est leur aspect le plus spectaculaire (l'accès en ligne à distance), il ne rend pas compte avec exactitude du processus technique réel de leur production et de leur

utilisation. Valable dans une optique de marketing, une telle présentation est fallacieuse sur le plan descriptif. Plus fondamentalement, il paraît évident que les producteurs ont tout à gagner à prendre conscience de la fonction éditoriale inhérente aux banques de données et à ajouter à leurs savoir-faire traditionnels (ressortissant à la documentation et à l'informatique) les savoir-faire de l'édition, tant sur le plan de la politique éditoriale (définition des produits et des contenus, étude des publics, etc.) que des techniques éditoriales (maquette, fabrication, promotion, etc.). Le développement d'interface de consultation fondée sur l'« hypertexte », c'est-à-dire sur la possibilité pour l'utilisateur de naviguer d'une information à une autre quel que soit son type (texte, image ou son), suppose un travail éditorial sur l'information qui nécessite la maîtrise de ces savoir-faire. De même, les caractéristiques du vidéotex sur minitel obligent à un travail éditorial qui n'avait pas lieu d'être lorsque les banques de données n'étaient interrogeables qu'en accès professionnel à partir de terminaux informatiques.

Une telle modification de point de vue peut favoriser une meilleure perception des possibilités en situant clairement les banques de données au confluent d'un triptyque fonctionnel : gisement de données, outil de recherche, source éditoriale.

Or, compte tenu d'une certaine marge de pondération entre ces trois grandes fonctionnalités, pondération qui dépend des objectifs poursuivis, il existe de sensibles différences morphologiques. Par exemple, il est parfaitement concevable qu'une banque de données ne mobilise que les deux premières fonctionnalités (stockage et recherche des données) par souci d'assurer la confidentialité des informations qu'elle contient. Dans ce cas, la fonction de rediffusion ou d'édition sera nulle ou limitée à des formes élémentaires et à un « public » très restreint. À l'inverse, dans le cas où une banque de données est utilisée comme un support éditorial, la fonction de rediffusion est hypertrophiée.

Au rebours d'une représentation courante distinguant les banques uniquement à partir des domaines qu'elles couvrent, il faut donc souligner le polymorphisme qui règne dans ce nouvel univers de l'information. Comme il sera montré plus loin, les réalisations de l'INRP sont une illustration tout à fait représentative de ce polymorphisme.

D'autres critères permettent également d'effectuer des distinctions d'ordre typologique. La nature des données est le plus évident. On peut distinguer les banques de données bibliographiques (qui accumulent des références sur un domaine défini ou en provenance d'un fonds documentaire spécialisé) et les banques de données factuelles

(qui contiennent des informations brutes, directement assimilables). Parmi ces dernières, on peut distinguer les banques de données numériques (sur lesquelles des calculs peuvent être effectués) et les banques de données en texte intégral (qui donnent accès aux documents complets sur lesquels des recherches lexicographiques et morpho-syntaxiques peuvent être réalisées). D'autres distinctions plus techniques existent encore qu'il n'est pas utile de préciser ici.

1.2. Des banques de données autoformatrices...

Ces grandes distinctions morphologiques et typologiques permettent de situer les réalisations les unes par rapport aux autres. Elles constituent une sorte de savoir minimum pour le consommateur de banques de données soucieux de dominer l'offre d'information.

Si les banques de données favorisent l'accès à l'information, pour autant celle-ci ne se livre pas immédiatement. Son accès suppose des capacités de conceptualisation (pour formuler une requête, sélectionner des critères d'interrogation, etc.) et la mobilisation de savoir-faire (travail sur clavier-écran, lecture cursive, etc.). Pour ces raisons, les banques de données supposent un utilisateur éclairé qui sache « frapper à la bonne porte » pour obtenir l'information qu'il recherche et utiliser à bon escient les techniques mises à sa disposition. On peut établir un parallèle entre l'utilisateur de banques de données et le lecteur de bibliothèque : ce dernier doit aussi savoir se rendre au bon endroit, savoir s'orienter dans le lieu et utiliser les fichiers et les autres outils de repérage à sa discrétion s'il veut rapidement trouver les documents susceptibles de répondre à son attente.

La démocratisation de l'utilisation des banques de données, grâce à la télématique et au vidéotex, comporte donc un effet induit d'autoformation. Certes, l'accès aux banques de données est de moins en moins réservé à des professionnels, il est de plus en plus convivial, mais l'effort fourni par le public pour les utiliser conduit paradoxalement à une certaine professionnalisation de sa part. En consommant des banques de données, ce dernier apprend à mieux dominer l'offre, à mieux connaître l'outil et acquiert plus ou moins intuitivement les rudiments des techniques documentaires mises en œuvre pour les produire. Même accessible par minitel, l'information ne se donne qu'à celui qui fait l'effort de la rechercher. En outre, l'assimilation de l'information par son consommateur suppose un effort d'intellection d'autant plus grand que le domaine est nouveau pour lui. C'est un lieu commun de remarquer que, même diffusée à domicile comme l'eau

courante grâce à la « fée Télématique », l'information ne s'absorbe pas comme on boit un verre d'eau !

1.3. Des banques de données ubiquistes...

Produits de l'informatisation, les banques de données permettent une démultiplication de l'offre d'information grâce aux passerelles jetées entre trois sous-systèmes techniques à vocation spécialisée : a) les gestionnaires de bases de données, b) les serveurs télématiques et c) les systèmes de publication assistée par ordinateur. Bénéficiant d'une synergie croissante entre ces sous-systèmes et des bienfaits des réseaux de télécommunication, les informations contenues dans une banque de données originale, produite en un lieu précis, jouissent a priori du don d'ubiquité via différents canaux et/ou supports de diffusion.

Ce don d'ubiquité permet de répondre à la grande variété de la demande, des publics et de leurs usages en multipliant les formes de l'offre à partir d'un gisement unique de données. D'où la possibilité d'une multitude de produits ciblés : accès en ligne (professionnel ou simplifié, sur contrat ou en libre accès, cher ou bon marché), publications imprimées (intégrales ou sélectives, périodiques ou occasionnelles, papier ou microforme, de haut niveau ou artisanales), accès sur de nouveaux supports comme le CD-ROM, le CDI, le DON, etc. (autant de sigles appelés avec d'autres à entrer petit à petit dans le langage courant...).

La contrepartie de cette grande souplesse éditoriale est la nécessité d'adapter l'information au canal et/ou au support choisi. Cette adaptation passe par différentes opérations : d'abord, le tri des informations pertinentes parmi la totalité des données stockées en usant de critères bien définis ; ensuite, le choix du degré de visibilité de ces données (total, partiel, en clair, codé, symbolique, etc.) ; enfin, éventuellement, un remodelage des informations pour les rendre plus assimilables par tel ou tel public visé. En sciences humaines et sociales, il est possible, par exemple, de réindexer automatiquement des références bibliographiques pour passer d'un vocabulaire scientifique à un vocabulaire moins spécialisé et moins ardu, accessible à un public plus large.

Cette ubiquité échappe la plupart du temps au consommateur qui ignore que ce qu'il consulte sur son minitel et ce qu'il lit dans une bibliographie ou dans un répertoire imprimé peut provenir de la même source.

2. L'usage des banques de données à l'INRP

Lorsque la production de banques de données a été à l'ordre du jour à l'Institut national de recherche pédagogique, il y a une dizaine d'années, cet univers des possibles n'avait pas encore pris toute la dimension que les années 80 lui ont donnée. Mais les grandes orientations en matière de banque de données tracées à l'époque n'ont pas été révisées car elles correspondaient à des tendances profondes que les innovations techniques apparues depuis lors n'ont pas mises en cause.

2.1. *Problématique générale*

Quatre considérations expliquent l'apparition des banques de données à l'INRP :

1°) le souci de moderniser le travail documentaire en vue d'accroître l'offre d'information en provenance du Centre de documentation recherche (CDR) et de la Bibliothèque (les fonds de ces deux services, à maints égards uniques, pouvant être considérés comme des fonds de référence au plan national) ;

2°) la nécessité de mettre davantage en valeur les autres ressources documentaires existant à l'INRP et d'élaborer à partir d'elles de nouveaux outils d'information ;

3°) la nécessité de répondre aux besoins des chercheurs engagés dans des recherches documentaires spécialisées de longue haleine ;

4°) le souhait d'utiliser les nouvelles technologies pour mettre en œuvre de nouveaux modes de transfert d'information vers de nouveaux publics au sein des communautés scientifique et éducative.

Deux types d'acteurs ont exprimé initialement ces besoins : d'une part, le Centre de documentation recherche qui, avec de faibles moyens, réalisait déjà des productions documentaires en utilisant les méthodes traditionnelles, et, d'autre part, des équipes de recherche, en premier lieu le Service d'histoire de l'éducation, puis ultérieurement des équipes des Départements « Didactiques des disciplines » et « Politiques, pratiques et acteurs de l'éducation » qui étaient intéressées à utiliser l'informatique dans le cadre de projets d'inventaire documentaire spécialisé.

Dans le premier cas, il s'agissait d'une problématique classique d'informatisation visant à remplacer par des processus informatisés d'anciens processus de gestion manuelle. Au début des années 80, cette question était déjà largement abordée dans le secteur professionnel de

la documentation et les premiers progiciels du marché avaient déjà fait leurs preuves. Dans le second cas, l'informatisation était la condition de la faisabilité de projets nouveaux, originaux et plus ou moins atypiques, émanant directement d'équipes de recherche. L'expérience acquise dans le secteur de la documentation pouvait servir de source d'inspiration à leur mise sur pied mais aucun « clonage » pur et simple n'était envisageable.

Dans la pratique, la création d'un dispositif de production de banques de données à l'INRP, à l'occasion de l'informatisation du CDR, a permis de répondre à la majeure partie des premiers besoins exprimés. Ce dispositif a ensuite été étoffé avec de nouvelles fonctionnalités pour répondre à une gamme plus étendue de besoins, les plus sophistiqués provenant des projets à finalité de recherche.

2.2. Problématique des banques de données au CDR

Après une première expérimentation début 1981 (1), c'est véritablement à partir de 1982 que la problématique des banques de données a été élaborée. Entre-temps, le Service d'histoire de l'éducation de l'INRP avait mis sur pied une version initiale de la banque de données EMMANUELLE (voir plus loin) en utilisant un logiciel spécialement écrit à cet effet par le Centre informatique recherche (CIR) de l'INRP. Cette banque de données a été redéployée ultérieurement sur le nouveau dispositif créé avec l'informatisation du CDR.

L'étude préalable à cette informatisation a duré de 1982 à 1984. Durant ces deux années, il a fallu d'abord procéder à une étude interne de faisabilité qui tienne compte des moyens disponibles à l'INRP, en particulier des moyens informatiques, et définir les grands objectifs qui seraient poursuivis.

2.2.1. Gestion de bibliothèque ou banque(s) de données ?

Centre de documentation spécialisé en sciences de l'éducation gérant un fonds documentaire d'environ 20 000 documents non périodiques et de 170 périodiques vivants (20 % de ce fonds qui s'étend sur 25 ans étant en anglais), le CDR, sous l'impulsion de Jean Hassenforder, a d'emblée conçu son informatisation documentaire non pas comme une informatisation de la gestion bibliothéconomique (2), mais comme la production (et l'aide à la production) de banques de données.

Ce choix a d'abord été dicté par des raisons politiques. La position particulière du CDR dans le contexte français de la recherche en éducation, en particulier son immersion dans la recherche vivante à l'INRP où travaillent environ 120 chercheurs entourés de plus d'un millier d'enseignants associés, lui a conféré de vastes missions d'information dépassant le public de ses lecteurs/utilisateurs. Centre bibliographique de référence en matière de sciences de l'éducation, le CDR était aussi co-agence nationale du réseau européen EUDISED, en collaboration avec l'Institut de l'information scientifique et technique (INIST) du CNRS, et recueillait à ce titre une information sans cesse mise à jour par voie d'enquête sur les recherches, les équipes de recherche et les chercheurs en éducation opérant en France (3). Le renforcement et l'optimisation de ces deux missions d'information passaient naturellement par l'informatisation des gisements d'informations correspondants afin d'en accroître la diffusion, et donc par la production de banques de données.

Mais des raisons techniques ont également justifiées ce choix. L'accès au fonds documentaire du CDR, pour des questions matérielles (documents acquis, en très grande majorité, en un seul exemplaire, surface de lecture réduite), a volontairement été limitée aux chercheurs en éducation et aux étudiants avancés. Cette autolimitation a permis jusqu'à présent que les tâches de gestion bibliothéconomique, bien qu'effectuées manuellement, soient maintenues à un niveau raisonnable, compte tenu des ressources humaines du CDR, sans que leur charge ne porte préjudice aux missions générales d'information du centre, c'est-à-dire entre autres à la production de banques de données.

Même si, dans la pratique, l'opposition entre les deux types d'informatisation (gestion bibliothéconomique *versus* banque de données) est aujourd'hui moins marquée qu'il y a cinq ou dix ans, grâce aux passerelles programmables entre les progiciels spécialisés dans la gestion de chaque type de tâches, il n'est pas inutile, sur le plan méthodologique, de les séparer. En matière d'informatisation, les solutions aussi « globales » qu'indistinctes favorisent souvent la confusion, la perte de temps et la contre-productivité.

2.22. Le projet EMILE du CDR

Une fois l'orientation en faveur de la production de banques de données documentaires retenue, il s'est agi de situer le projet dans son environnement national et international.

Dans le cadre de la banque multidisciplinaire FRANCIS, le CNRS développait depuis 1972 un important secteur Sciences de l'éducation

correspondant au *Bulletin signalétique* 520 ; plusieurs organismes plus ou moins proches de l'INRP et de taille variable formulaient aussi des projets de banques de données ; en outre, dans le sillage de la mission confiée à M. Roland Carraz, circulait à l'époque (1983) l'idée d'une grande banque de données sur l'éducation qui pourrait fédérer les divers projets institutionnels autour du CNRS ou d'un éventuel groupement à définir, à l'instar du ERIC nord-américain.

L'étude préliminaire mena à quelques conclusions « stratégiques » qui ont orienté la définition et la mise en œuvre du projet :

1°) Dans l'environnement institutionnel français, il fallait distinguer entre deux types de partenaires : ceux qui agissaient dans un secteur similaire à celui du CDR et avec lesquels une coopération pouvait s'instaurer ou se renforcer (en premier lieu le CNRS), et les autres dont des projets techniquement similaires se déployaient dans des milieux et avec des objectifs différents. En tant qu'outils d'information, les banques de données n'échappent pas au pluralisme institutionnel qui correspond à une certaine spécialisation. Il est donc logique qu'en matière d'éducation, la recherche soit traitée en elle-même et qu'elle possède ses propres outils d'information (4).

2°) Il fallait ensuite distinguer deux types de banques de données pour y voir plus clair dans l'offre et la demande d'information en sciences de l'éducation : a) les banques « généralistes » qui couvrent l'ensemble de la recherche en éducation sous tous ses aspects, et b) les banques « spécialisées » aux domaines beaucoup plus restreints. Les premières sont conçues et produites par les structures documentaires classiques (centres de documentation ou bibliothèques spécialisées) dans un but de large information, les secondes le sont par des équipes de recherche ou des réseaux d'équipes dans un but d'information très spécialisée et, éventuellement, avec une visée d'évaluation ou de bilan sectoriel de la recherche.

3°) Enfin, tout en reconnaissant la prééminence des banques de données bibliographiques, il ne fallait pas négliger les banques de données factuelles, ces dernières délivrant une information directement utilisable qui ne renvoie pas aux documents primaires. Cette indépendance à l'égard des fonds documentaires qui caractérise les banques factuelles s'accorde bien à la philosophie de l'accès télématique en ne pénalisant pas l'utilisateur trop éloigné des documents primaires eux-mêmes.

Le projet qui fut baptisé alors EMILE (5), s'orienta donc a) vers la production de banques de données « généralistes » (les banques de données EMILE1 et EMILE2) et b) vers l'aide aux équipes de recherche

désireuses d'utiliser le savoir-faire documentaire à leur profit (les banques DAF – ex EMILE3 – et EMILE4).

Ces orientations trouvèrent un écho favorable auprès du CNRS puisque deux conventions furent signées en 1984 entre l'INIST et l'INRP autour d'EMILE1 et d'EMILE2, conventions qui ont été renouvelées en 1990 et 1991. Par ailleurs, la DBMIST du Ministère de l'Éducation nationale apporta son soutien au projet en permettant de résoudre la question du logiciel documentaire nécessaire à la production des banques sur l'ordinateur du Centre informatique recherche (CIR) (6). Enfin, le CIR s'engagea très efficacement dans la réalisation informatique du projet en devenant le gestionnaire du système de production de banques de données de l'INRP et en développant les nouveaux programmes indispensables à son fonctionnement (7).

2.3. Problématique des banques de données de recherche

Plusieurs équipes de recherche, successivement ou simultanément, ont été amenées à formuler à des moments précis de leur travail à l'INRP, des projets qui ont débouché sur des banques de données. Etant donné l'originalité de chaque projet, il est difficile de dégager un paradigme unique qui serait celui de « la banque de données de recherche » par excellence. Toutefois, il convient de noter les traits communs aux banques spécialisées.

1° Une **définition épistémologique rigoureuse du champ et/ou de l'objet**. Lorsqu'une équipe se penche sur son domaine de recherche pour en capitaliser les acquis, elle doit d'abord prendre du recul par rapport à ses propres recherches. Elle doit borner son champ en le conceptualisant et en définissant ses objets de connaissance. D'où une approche fondatrice résolument épistémologique, indépendante de toutes considérations pratiques, pour tracer les limites du champ et le profil des objets inventoriés.

Par contraste, les banques de données « généralistes » produites par les centres de documentation spécialisés sont fortement conditionnées par les problèmes pratiques inhérents à la constitution et à la gestion des fonds documentaires qui leur servent de source. Indépendantes de l'entretien d'un fonds, les banques de données de recherche n'ont pas de telles contraintes. Elles puisent leur matière dans les fonds documentaires existants. Elles recherchent partout où elle est susceptible de se trouver l'information correspondant à la définition du champ et de l'objet avec une exigence maximale d'exhaustivité.

2°) **Un ensemble de langages documentaires ajustés au champ et à l'objet.** La définition initiale des banques de données de recherche s'articule autour de langages documentaires spécialisés (plan de classement, lexique ou thésaurus, typologie, nomenclature, etc.), élaborés par les chercheurs eux-mêmes pour servir d'outils à leur travail d'indexation. Ceux-ci sont les plus aptes à définir les outils qui permettront les traitements ultérieurs sur les données recensées. Inspirée par le savoir-faire de la documentation, la constitution de ces langages documentaires intervient comme la concrétisation de la démarche d'investigation épistémologique initiale puisqu'il s'agit, à travers eux, de formaliser non seulement les concepts-clés du champ et les types d'objet à recenser, mais aussi toute une série d'informations de contexte liées à la problématique que l'équipe de recherche entend poursuivre sur le corpus constitué.

3°) **Des visées scientométriques et d'évaluation.** Le recueil méthodique et quasi exhaustif des données et leur indexation avec des langages documentaires à visée scientifique confèrent aux banques de données de recherche la légitimité d'un corpus. Traitements statistiques et analyse de données peuvent dès lors s'y appliquer avec rigueur. Dans ces banques, la recherche « documentaire » peut donc aller bien au-delà de la simple recherche multicritère habituelle. Gisement de données à haute valeur ajoutée, elles sont une source fiable pour les études scientométriques et bibliométriques (8) qui servent de base aux états de question, aux études de tendance et à toute recherche concrète d'ordre épistémologique et comparatif sur la constitution d'un champ de recherche particulier.

Pour autant, il serait erroné de penser que ses traits communs soient suffisamment unificateurs pour définir un paradigme unique de la « banque de données de recherche ». Si l'on entend par ce vocable, une banque conçue et réalisée par des chercheurs dans le cadre de leur recherche et ayant des finalités de recherche, deux paradigmes principaux émergent. Le premier paradigme est la « bibliographie spécialisée » ; le second, le « recueil de ressources spécialisé ».

2.31. Les bibliographies spécialisées

L'objectif premier des bibliographies spécialisées est d'inventorier la littérature de recherche dans un domaine circonscrit. Elles sont souvent de nature rétrospective afin de permettre un ample regard sur le domaine traité. Dans les banques de données DAF et EMILE4 (voir plus loin), c'est l'année 1970 qui a été prise comme borne inférieure. La ban-

que de données EMMANUELLE 5 remonte au siècle dernier, alors qu'HISTED commence en 1976.

Le travail de repérage et d'analyse systématique de la littérature de recherche est une sorte d'antidote à l'insatisfaction ressentie quant à la circulation de l'information dans le milieu de la recherche, rapide mais insuffisamment structurée. Les « collègues invisibles » qui se constituent entre les chercheurs dans le cours de leurs activités et de leurs contacts professionnels sont des réseaux spontanés par où transitent avec célérité les informations au sein de micro-communautés scientifiques. Mais ces réseaux informels fonctionnent sur la base de l'hyperspécialisation et reproduisent la fragmentation des savoirs à laquelle ils participent. Non institutionnalisés, ils sont peu aptes à servir de support à la capitalisation et à la synthèse des savoirs.

Produire une bibliographie spécialisée au nom d'une communauté scientifique donnée, c'est d'abord se donner un moyen de dépasser les limites trop étroites et le morcellement du travail hyperspécialisé de la recherche. C'est aussi un moyen de rompre l'isolement relatif des collègues invisibles les uns par rapport aux autres, en particulier d'un pays à l'autre. Enfin, c'est affirmer la cohérence d'un domaine relativement jeune ou en émergence en capitalisant ses résultats, ses problématiques et ses méthodologies. En tant qu'outil émanant d'une communauté de chercheurs en voie d'institutionnalisation, une bibliographie spécialisée constitue une promesse de synergie. La capitalisation qu'elle opère de l'intérieur peut faire prendre conscience aux chercheurs concernés de leur commune appartenance à un champ nouveau et de la nécessité, par delà les frontières nationales et linguistiques, de nouer des coopérations qui transcendent la parcellisation. À ce titre, elle devient un « lieu » de visibilité et de légitimation et plaide en faveur d'une meilleure reconnaissance institutionnelle sur la base d'acquis dûment inventoriés.

2.32. Les recueils de ressources spécialisés

En dehors des bibliographies spécialisées, les ressources traitées par la recherche en éducation et mises en ordre à des fins d'exploitation ultérieure sont très diverses : corpus de documents (officiels, pédagogiques, etc.), fonds d'archives, collections d'objets scolaires, séries d'épreuves, de tests, etc. La constitution de recueils méthodiques de ces ressources est une activité de recherche utile pour la communauté scientifique. Elle rend possible l'étude scientifique en donnant accès à une matière première jusqu'alors enfouie ou connue de façon extrêmement lacunaire. C'est le cas en particulier pour les recherches d'ordre

historique (voir plus loin la banque de données EMMANUELLE). Ces recueils sont la condition pour que de nouvelles approches multidisciplinaires puissent se déployer et renouveler les questions sur une base documentaire fiable, tendant à l'exhaustivité.

Ces recueils de ressources peuvent également intéresser chercheurs et enseignants. Dans le domaine de la didactique des disciplines par exemple, il est clair qu'un recueil d'épreuves testées dans le cadre de recherches expérimentales et se trouvant donc pédagogiquement validées, peut être une précieuse source pour le didacticien chevronné. Il lui permet de bénéficier de la capitalisation des observations issues de la recherche. Mais ce même recueil peut aussi bénéficier à l'enseignant en lui permettant d'acquérir une formation didacticienne qui conjugue épreuves pratiques et notions théoriques dans le commentaire des épreuves (voir plus loin la banque de données DIDAC).

Des considérations précédentes, il ressort que les banques de données de recherche peuvent jouer un rôle de structuration et de légitimation au service des champs ou des objets de recherche nouveaux, au même titre que les revues, les colloques, les associations, etc. Elles constituent un moyen supplémentaire à la disposition d'une communauté scientifique spécialisée pour affirmer son existence, faire connaître ses travaux, rendre visible ce qui circule seulement dans le « collègue invisible » des chercheurs et nouer des coopérations internationales fructueuses.

3. Les réalisations de l'INRP

Jouant sur la gamme des possibilités offertes par la technique des banques de données, les réalisations de l'INRP en la matière sont extrêmement variées. Comme il a déjà été souligné, on trouve à la fois des banques de données bibliographiques et factuelles, des banques « généralistes » et « spécialisées », des banques de données classiques et des banques de données vidéotex, des banques à fort volume et des « micro-banques » de dimension plus réduite. Même si chaque projet a obéi à une rationalité explicitée en terme d'objectifs et d'utilité par les différentes équipes de production en concertation avec les équipes informatiques, en particulier celle du Centre informatique recherche de l'INRP. Il n'y a pas eu de plan d'ensemble. Une des missions du service Banques de données récemment créé est précisément d'assurer la coordination technique des projets et d'organiser leur valorisation, chaque équipe de production demeurant le maître d'œuvre de son projet.

3.1. Les banques de données « généralistes »

❑ EMILE1 : Qui publie quoi ?

Cette banque contient les références bibliographiques du fonds du Centre de documentation recherche, depuis 1985, soit actuellement plus de 5 000 références à des documents non périodiques. Au terme d'une convention avec l'Institut de l'information scientifique et technique (INIST) du CNRS, le CDR transmet, à titre onéreux, les données d'EMILE1 qui ressortissent strictement aux sciences de l'éducation (soit de l'ordre de 80 % de son fonds non périodique) à la banque multidisciplinaire FRANCIS du CNRS. Cette coopération va dans le sens d'une amélioration de la couverture bibliographique au plan national dans le secteur des sciences de l'éducation puisque l'INIST traite pour sa part les articles de périodiques.

Outre les champs bibliographiques traditionnels, chaque référence comporte un résumé, des mots-clés, un indice de classement thématique et un indice de contenu (qui situe le document sur l'échelle de la recherche). Le lexique de mots-clés est celui de FRANCIS-Sciences de l'éducation. Par contre, le plan de classement en sciences de l'éducation et la typologie de contenu sont propres au CDR de l'INRP.

FRANCIS est disponible sur le serveur français QUESTEL moyennant contrat. Un accès par minitel est possible sans contrat sur le 36.29.36.01. Ce numéro donne également accès à la banque de données PASCAL consacrée aux sciences dures et à la psychologie.

Une version sélective d'EMILE1, limitée aux documents français non souterrains, sera également accessible par minitel sur le 36.16 (code INRP) à partir de 1992.

Enfin, une sélection des références majeures issues d'EMILE1 paraît trois fois par an dans la revue *Perspectives documentaires en éducation* à la rubrique Bibliographie courante sous le titre « Ouvrages & rapports ».

❑ EMILE2-ERTEL : Qui recherche quoi ?

Cette banque, co-produite par l'INIST du CNRS et l'INRP, contient le signalement des recherches françaises, des organismes ou équipes de recherche et des chercheurs en éducation depuis 1974. Les données, recueillies grâce à une enquête annuelle menée directement auprès des équipes de recherche, sont traitées à l'INRP dans la banque de données EMILE2 par une équipe conjointe INIST-INRP.

Dans un second temps, elles sont transmises au réseau européen EUDISED qui dispose d'une banque de données contenant des signalements analogues en provenance de 23 pays européens. La banque EUDISED est disponible moyennant contrat sur le serveur ESA-IRS de l'Agence spatiale européenne (9). Une sélection des descriptifs de recherche (environ un millier par an) paraît dans le *Bulletin EUDISED R & D* (10).

La banque française EMILE2 compte actuellement environ 2 000 signalements de recherche, 1 800 signalements de chercheurs et 500 signalements d'équipes de recherche en activité. Chaque descriptif de recherche comporte, entre autres, l'intitulé de la recherche, la date de début et de fin de la recherche, les membres qui composent l'équipe, le nom et les coordonnées de l'organisme où l'équipe travaille, des mots-clés, plusieurs indices de classement thématique, un résumé (hypothèse, méthodologie, résultat) et un signalement succinct des publications afférentes. Le thésaurus utilisé est le *Thésaurus européen de l'éducation* (auparavant *Thésaurus EUDISED*) dont la troisième édition doit paraître en juin 1991.

Une forme simplifiée d'EMILE2 est en cours d'élaboration afin de rendre les données accessibles par minitel sur le 36.16 (code INRP) à partir de l'automne 1991. Dans une première étape, ne seront disponibles que les recherches en cours. Ce service portera le nom de ERTTEL (Éducation Recherche Télématique).

3.2. Les banques de données « spécialisées »

Il a paru intéressant de présenter les réalisations correspondantes par domaine de recherche plutôt que par type (voir plus haut le développement 2.3.).

Histoire de l'éducation

☐ EMMANUELLE : Les manuels scolaires depuis 1789

Élément central dans un programme plus large consacré à l'étude des manuels scolaires, la banque de données EMMANUELLE est un recueil de ressources qui s'attache à recenser toutes les éditions de manuels scolaires publiés en France depuis 1789, dans toutes les disciplines et à tous les niveaux d'enseignement (11). Cette banque concerne environ 80 000 titres et plusieurs centaines de milliers d'éditions.

Etant donné son objet, on pourrait penser qu'EMMANUELLE est une bibliographie spécialisée plutôt qu'un recueil de ressources (voir plus haut le développement 2.3). En fait, les objectifs de la banque comme ses principes d'établissement la définissent clairement comme un recueil de ressources. Son objectif est d'établir un catalogue collectif des manuels scolaires conservés à la Bibliothèque nationale, à la bibliothèque de l'INRP, au Musée national de l'éducation (12) et, ultérieurement dans d'autres fonds, et, sur cette base, de fournir des études statistiques sur leur production et leur conservation. D'où un travail de collecte des données en rapport direct avec cet objectif : a) les notices sont établies pour l'essentiel à partir des grandes recensions et catalogues nationaux et non à partir des documents primaires eux-mêmes, b) elles se rapportent le plus souvent à un ensemble de documents primaires considéré comme les éditions successives d'un même manuel scolaire, c) elles comportent une majorité de champs descriptifs visant à préciser le contexte de production du manuel plutôt que son contenu, l'indexation étant limitée à la discipline et au niveau d'enseignement (au moyen d'un lexique contrôlé).

La banque de données EMMANUELLE est en passe de constituer le standard de description bibliographique international des manuels scolaires. Des conventions ont été signées ou sont en cours de négociation avec plusieurs partenaires étrangers soucieux de recenser et de préserver leurs collections de manuels scolaires nationaux (Brésil, Canada, Colombie, Espagne, Grèce, Mexique, Suisse, Vénézuéla). Elles visent à permettre l'échange de données, à susciter les études comparatives et ainsi à développer la coopération scientifique internationale dans ce domaine.

La collection EMMANUELLE éditée par l'INRP et les Publications de la Sorbonne depuis 1987 a pour fonction de publier les données de la banque par corpus disciplinaire. Près de 8 000 références correspondant aux corpus de latin, grec, italien, allemand sont ainsi disponibles (13).

Il est prévu en outre un accès télématique par minitel sur le 36.16 (code INRP) à partir de l'automne 1991.

□ EMMANUELLE 5 : Bibliographie internationale sur les manuels scolaires

Partie intégrante du programme EMMANUELLE, EMMANUELLE 5 est une bibliographie spécialisée, rétrospective et courante, recensant les études et recherches de toute nature effectuées sur les manuels scolaires en France et, dans le cadre du Conseil de l'Europe, dans le do-

maine européen. Chaque document est l'objet d'un résumé et l'indexation est assurée par un thésaurus spécifique élaboré par les chercheurs. Ce projet sera opérationnel en 1992.

☐ HISTED : Bibliographie d'histoire de l'éducation

La bibliographie d'histoire de l'éducation a vu le jour en 1979 avec la revue *Histoire de l'éducation*. Elle paraît chaque année sous la forme d'un numéro double de la revue (14). Les notices bibliographiques sont indexées avec un plan de classement thématique spécialement conçu pour l'histoire de l'éducation. Au total, cette bibliographie compte près de 15 000 références, dont les 3 000 plus récentes sont déjà saisies dans la banque de données. Une saisie rétrospective des années antérieures à 1988 est prévue ainsi qu'un accès télématique par minitel sur le 36.16 (code INRP) à partir de 1992.

Didactiques

☐ DAF : Bibliographie des recherches en didactique et acquisition du français langue maternelle

Cette banque, appelée antérieurement EMILE3, répertorie méthodiquement les documents produits par la recherche dans le domaine de la didactique et de l'acquisition du français langue maternelle. Elle est réalisée par une équipe de recherche de l'INRP en collaboration avec une équipe québécoise de l'Université de Montréal (15). Des équipes belge et suisse participent également à DAF : le Centre d'études et de documentation pour l'enseignement du français (CEDOCEF) de Namur et l'Institut romand de recherches et de documentation pédagogiques de Neuchâtel (IRDP).

Actuellement, DAF compte 3 500 références couvrant la période de 1970 à nos jours. Elle est en constante mise à jour. Les descripteurs qui indexent les documents sont tous définis, ce qui confère à cette banque un caractère terminologique (16).

Une version simplifiée de DAF a été élaborée pour le public français sous le nom de DAFTEL (Didactique et Acquisition du Français Télématique). La principale modification porte sur le thésaurus qui a été réduit à 150 termes pour permettre l'interrogation par minitel. Jusqu'à la fin de l'année 1991, l'accès se fait par le 36.15 (code SUNK) ou le 36.16 (code SUNIST), avant d'être rendu possible sur le 36.16 (code INRP) à partir de l'automne 1991. À l'intention du public étranger, en particulier nord-américain, DAF est disponible également dans sa forme originale sur le serveur canadien SDM (17).

❑ **DIDAC : Ressources en éducation mathématique**

La banque de données DIDAC est un recueil de ressources en éducation mathématique dont la conception remonte à 1985. Une première maquette a été élaborée en vidéotex sur le serveur SUPER puis testée auprès d'un échantillon d'une centaine d'utilisateurs en 1987 (18). La banque a pris son essor à partir de 1990.

DIDAC combine trois types de ressources : a) des références bibliographiques aux travaux de recherche menés dans le domaine de l'éducation mathématique en France et à l'étranger depuis une vingtaine d'années, b) des épreuves, directement issues de ces recherches, qui ont fait l'objet d'une validation pédagogique, c) des définitions qui permettent d'éclairer le commentaire didactique et statistique de chaque épreuve.

Un thésaurus spécialisé THELEM a été élaboré pour cette banque de données à partir de 1988. Il sert à l'indexation des trois ressources signalées et permet la recherche documentaire multicritère.

Par ailleurs, une messagerie et un journal d'information télématiques sont associés à DIDAC qui est accessible sur le 36.15 (code DIDAC) (19).

Formation des enseignants

❑ **EMILE4 : Bibliographie des recherches sur la formation des enseignants**

Cette banque spécialisée, créée en 1988, répertorie méthodiquement, en utilisant un thésaurus spécifique, les documents produits par la recherche francophone, et dans une certaine mesure non-francophone, dans le domaine de la formation des enseignants de 1970 à nos jours (20). Elle est réalisée par une équipe de recherche de l'INRP en collaboration avec des équipes québécoise et belge. Elle est en cours de constitution et sera également accessible par minitel le moment venu. Une première publication d'environ 1200 références bibliographiques doit paraître en 1991.

4. Perspectives

Deux axes de développement se présentent à partir de ces premiers acquis. De nouvelles banques de données, surtout spécialisées, peuvent voir le jour dans les années futures et bénéficier des techniques mises au point par le Centre informatique recherche de l'INRP. À cet

égard, l'ouverture d'un serveur télématique par l'INRP à partir de l'automne 1991 constitue un événement important pour la valorisation des banques de données existantes et pour l'éclosion de nouveaux projets. Elle trace également de nouvelles perspectives pour la diffusion des banques de données et l'utilisation de la communication télématique (21).

En permettant, sur une plus large échelle qu'auparavant, de développer l'édition télématique à l'INRP, ce serveur va également relancer l'intérêt du public – notamment des enseignants – pour les ressources de l'Institut. La mise en place des IUFR qui doit contribuer à structurer la demande d'information en provenance des enseignants en formation, suppose en retour une offre d'information mieux ciblée, plus abondante, d'un abord plus aisé. En accédant plus facilement qu'auparavant aux ressources de l'INRP, les formateurs d'enseignants, mais aussi les cadres du système éducatif et les documentalistes du vaste réseau documentaire de l'Éducation nationale, seront mieux à même de s'informer sur les résultats de la recherche comme sur le cours de la recherche vivante. Ils pourront ainsi mieux remplir un rôle tout à fait indispensable de relais entre la recherche et la pratique, entre les innovations et les terrains. Une bonne circulation de l'information scientifique parmi les médiateurs que sont ces publics ciblés est une condition indispensable pour que des productions à orientation strictement vulgarisatrice voient le jour grâce au dialogue instauré entre chercheurs et praticiens. C'est dans cet esprit que l'INRP conçoit la mise en place de son serveur et qu'il suit avec attention l'apparition des nouveaux supports optiques (CD-ROM, etc.) avec l'intention, le moment venu, seul ou en collaboration avec d'autres organismes, à recourir à ces nouveaux médias pour contribuer à mieux diffuser les ressources uniques accumulées dans ses banques de données au service de la recherche et des enseignants.

Philippe Champy

*Chef du service Banques de données
Département Ressources et communication
Institut national de recherche pédagogique*

Notes

- (1) Voir Champy, Philippe.- *Rapport préliminaire relatif à l'informatisation du Centre de Documentation Recherche de l'INRP*. Paris, INRP, mai 1981, 22+IV p.
- (2) La gestion bibliothéconomique recouvre les tâches d'acquisition, d'enregistrement et de catalogage des documents, de bulletinage des périodiques

et de gestion des prêts. Le volume annuel de ces tâches au CDR correspond à environ 800 à 900 nouvelles acquisitions, 400 fascicules bulletins et 3000 opérations de prêts en direction de 400 lecteurs encartés.

- (3) Depuis janvier 1991, cette mission est dévolue au service « Banques de données » du département « Ressources et communication » auquel appartient également le CDR.
- (4) Voir Champy, Philippe.- « L'offre d'information documentaire en sciences de l'éducation : essai de clarification », *Perspectives documentaires en sciences de l'éducation*, n° 3, 1984, pp. 29-37.
- (5) Voir Champy, Philippe.- *EMILE. Projet de banque de données sur la recherche en éducation : étude de contexte et description du projet*. Paris, INRP, janvier 1984, 34 p.
- (6) L'INRP utilise le progiciel JLB-DOC de la société JLB-Informatique, fonctionnant sur mini-ordinateur (en particulier UNIX) et sur micro-ordinateur (MS-DOS). Ce logiciel permet de développer seize banques de données en parallèle. De 1984 à 1990, les banques ont été produites sur un Mini6 BULL (sous GCOS6) et, à partir de 1991, sur un HP 9000 (sous UNIX). Certaines sont produites également sur micro-ordinateurs avec le logiciel CDS-ISIS de l'UNESCO dont la licence d'utilisation est gratuite pour les organismes publics.
- (7) L'équipe du CIR à l'instigation de laquelle ces développements ont été rendus possibles est composée de Jean-Pierre Houillon, chef du CIR, Dominique Lenne et Jean Pelé. Les principaux développements concernant les banques de données ont porté sur l'amélioration de l'ergonomie du progiciel JLB-DOC, sur l'apport de fonctionnalités nouvelles et la conception de modules logiciels spécialisés pour l'importation et l'exportation de données (aux formats QUESTEL, TEXTO, ISIS, MARC, etc.), la composition et le balisage automatique pour la PAO, les traitements statistiques bibliométriques.

Le CIR déploie ses activités dans de nombreux autres domaines tels que le traitement des résultats de recherche (statistiques, analyse de données), la télématique (serveur), les bases de données relationnelles (gestion), la PAO, etc. De manière générale, il joue un rôle moteur dans l'introduction de l'informatique et des nouvelles technologies à l'INRP, et assure formation et assistance aux utilisateurs, notamment aux chercheurs, dans la définition et la conduite des projets nécessitant une informatisation.

- (8) Sur les techniques de la scientométrie, lire : Courtial, Pierre.- *Introduction à la scientométrie : de la bibliométrie à la veille technologique*.- Paris : Anthropos, 1990.- 135 p.
- (9) Pour plus d'information, contacter ESA-IRS, 8-10 rue Mario Nikis, 75738 Paris Cedex 15. Tél.: (1) 42.73.72.03.
- (10) Le *Bulletin EUDISED R & D* est diffusé par Saur Verlag, Postfach 71 10 19, D-8000 München 71.
- (11) Sur le programme EMMANUELLE que le Service d'histoire de l'éducation consacre aux manuels scolaires, lire : Choppin, Alain.- « L'histoire des manuels scolaires : une approche globale ».- *Histoire de l'éducation*, n° 9,

- 1980, pp. 1-25 ; Choppin, Alain.- « Le programme informatique EMMA-NUELLE », *Internationale Schulbuchforschung*, n° 2, 1991, pp. 87 sq. ; *Etapes de la recherche*, n° 27, 1991. Pour de plus amples informations, contacter Alain Choppin, Service d'histoire de l'éducation, Département Mémoire de l'éducation, INRP, 29 rue d'Ulm, 75230 Paris Cedex 05. Tél.: (1) 46.34.91.08.
- (12) En outre, le musée alimente une banque de données, en cours de constitution depuis plusieurs années, recensant ses ressources : des fonds de gravures et estampes, jeux et jouets, livres pour la jeunesse et livres de prix, et du matériel pédagogique tel que cartes, globes terrestres, instruments scientifiques, vues fixes sur verre et photographies, mobiliers scolaires, travaux et cahiers d'élèves. Dans l'avenir, cette banque de données qui compte plusieurs dizaines de milliers d'items sera également accessible sur le serveur de l'INRP.
- (13) Choppin, Alain, dir.- *Les manuels scolaires en France de 1789 à nos jours*.- Paris : INRP ; Publications de la Sorbonne.- (Emmanuelle.)
 Les manuels de grec.- 1987.- 180 p.
 Les manuels d'italien.- 1987.- 92 p.
 Les manuels d'italien.- 1988.- 528 p.
 Les manuels d'allemand.- À paraître en 1991.
 Chaque volume comporte en introduction une étude de la production des manuels scolaires dans la discipline concernée.
- (14) *Histoire de l'éducation*, n° 4 (1979), n° 7-8 (1980), n° 11-12 (1981), n° 15-16 (1982), n° 19-20 (1983), n° 23-24 (1984), n° 27-28 (1985), n° 31-32 (1986), n° 35-36 (1987), n° 39-40 (1988), n° 43-44 (1989), n° 47-48 (1990).
- (15) Sprenger-Charolles, Liliane ; Gagné, Gilles ; Lazure, Roger ; Ropé, Françoise.- « Inventaire analytique et états de question des recherches en didactique du français (langue maternelle) : présentation de la recherche », *Perspectives documentaires en sciences de l'éducation*, n° 4, 1984, pp. 31-44 ; « Propositions pour une typologie des recherches », *Perspectives documentaires en sciences de l'éducation*, n° 11, 1987, pp. 49-71.
- (16) Gagné, Gilles ; Sprenger-Charolles, Liliane ; Lazure, Roger ; Ropé, Françoise.- *Recherches en didactique et acquisition du français langue maternelle*.- Bruxelles : De Boeck-Wesmael ; Paris : Editions universitaires, 1989.
 Tome 1 : cadre conceptuel, thésaurus, lexique des mots-clés.- 200 p.
 Tome 2 : répertoire bibliographique.- 497 p.
 Répertoire bibliographique : mise à jour 1.- Montréal : SDM, PPMF ; Paris : INRP, 1990.- 251 p.
- (17) Pour plus d'information, contacter SDM, 1685 rue Fleury Est, Montréal, Québec H2C 1T1. Tél.: (514) 382-895.
- (18) Pour un exposé très complet sur DIDAC, lire : Gilis, D. ; Athenour, C.- « Que peuvent apporter l'outil télématique et les banques de données à l'enseignement et à la recherche en didactique des disciplines ? Un exemple de synergie entre producteur d'informations et serveur de données. », *Recherches en didactique des mathématiques*, vol. 10, n° 2-3, 1990, pp. 327-368.
- (19) Autre accès : 36.15 code SCOL.R. Pour de plus amples informations, contacter Daniel Gilis, Unité de didactique des mathématiques, Département Di-

dactiques des disciplines, INRP, 29 rue d'Ulm, 75230 Paris Cedex 05. Tél.: (1) 46.34.91.28.

- (20) Bourdoncle, Raymond ; Adamczewski, Georges ; Lessard, Claude ; Ropé, Françoise ; Tournier, Michèle.- « Vers un inventaire analytique des travaux sur la formation des enseignants et des formateurs », *Perspectives documentaires en sciences de l'éducation*, n° 15, 1988, pp. 101-115.
- (21) Nous n'avons pas signalé la banque de données d'information consacrée à l'INRP (organigramme, publications, recherches, manifestations, etc.) qui sera accessible par minitel sur le 36.16 (code INRP), ainsi que les autres services télématiques (messageries, forums, etc.) qui seront également disponibles à partir de l'automne 1991.