

Le Catalogue collectif de France

Entrevue avec Jacques Bourgain (15 décembre 1993)

ABF – Comment se positionne le projet du CCF par rapport à celui de la Bibliothèque nationale de France ?

J. B. – Il s'agit d'une composante politique et fonctionnelle essentielle du projet de la BNF. C'est en quelque sorte la traduction, avec la politique des pôles associés, de la volonté des pouvoirs publics de créer, à la périphérie du bâtiment de Tolbiac qui s'inscrit irréductiblement dans le paysage parisien, une sorte de bibliothèque de recherche immatérielle, accessible de tous les points du territoire (et de l'étranger), traduction palpable par tous les usagers de la dimension réellement nationale de ce chantier.

ABF – Le CCF doit-il donner accès à l'ensemble des ressources documentaires françaises ?

J. B. – Non, pas exactement. Il vise plutôt à permettre à ses usagers d'obtenir les informations de localisation des documents intéressant la recherche, le mot recherche étant pris ici au sens générique du terme, au sens de recherche d'informations. Il ne s'agit pas, par exemple, de localiser pour la énième fois le dernier prix Goncourt (c'est le rôle des catalogues locaux) mais plutôt de permettre au public de savoir où se trouve tel document à des fins de recherche et d'y avoir accès.

ABF – Comment se situe l'étape actuelle dans la réalisation du CCF et comment la conduite du projet est-elle organisée ?

J. B. – J'ai été mis à disposition de l'EPBF¹ en février 1993 pour mener les travaux d'étude et de réalisation du CCF. L'ensemble des études qui avaient déjà été menées par l'établissement, en particulier celles portant sur les fonctionnalités du service à mettre en œuvre, avaient exploré le champ du possible, sans contraintes réelles à ce stade. Un travail a alors été mené pour confronter le résultat de ces travaux aux contraintes budgétaires, techniques et calendaires afin de définir une cible « réaliste » et pragmatique, tout en restant à la mesure du projet de la BNF et des attentes qu'il suscite. Ce travail s'est traduit par la rédaction d'un cahier des charges pour la réalisation du schéma directeur informatique opérationnel (SDIO) du CCF, confié après appel d'offre à la société STERIA. Il a débuté en octobre 1993 et ses travaux seront terminés début mai 1994.

Parallèlement, le directeur du Livre confirmait la BNF dans son rôle de maître d'ouvrage et des instances (comité de pilotage pour les décisions « politiques », groupe de projet pour l'inscription permanente du projet et groupes de travail spécifiques) étaient mises en place pour le suivi du dossier, de telle façon que les différents partenaires potentiels du CCF puissent y faire entendre leur voix, c'est-à-dire le ministère de la Culture et de la Francophonie, le ministère de l'Enseignement supérieur et de la Recherche, la BNF et le Conseil supérieur des bibliothèques.

ABF – Vous en êtes donc à l'étape du schéma directeur informatique opérationnel. Comment les travaux de celui-ci ont-ils été structurés ?

J. B. – De la façon la plus classique qui soit en matière de conduite de projet !

L'étude comprend les étapes suivantes : bilan de l'existant et orientations majeures du projet ; définition de la cible fonctionnelle (indépendamment, à ce stade, des contraintes techniques) ; étude des scénarios techniques de réalisation et approfondissement du scénario retenu ; évaluation budgétaire et établissement du plan d'action ; rapport final de synthèse.

ABF – A la suite de cette étude, quels sont les objectifs assignés au CCF et les principes qui guideront sa réalisation ?

J. B. – Les *objectifs* définis dans le cadre du schéma directeur opérationnel sont les suivants :

- donner, grâce à une interface unique, accès à un ensemble significatif de catalogues collectifs ou de catalogues locaux décrivant les fonds français intéressant la recherche, pour en offrir une vision globale et structurée. A l'ouverture, le CCF donnera accès, au minimum, au catalogue de la BNF, aux catalogues collectifs gérés par le ministère de l'Enseignement supérieur et de la Recherche (Pancatalogue, CCNPS², Téléthèses, Catalogue collectif des ouvrages [CCO] du CNRS), ainsi qu'aux fonds anciens et spécialisés d'une cinquantaine de grandes bibliothèques municipales, soit au total environ 14 millions de notices ;
- améliorer l'accès aux documents en organisant la meilleure continuité possible entre la consultation et les services d'accès aux documents ;
- informer l'utilisateur sur la géographie des fonds disponibles en France et sur l'accessibilité des documents grâce à un

1. EPBF : Établissement public de la bibliothèque de France. Par décret n° 94-3 du 3 janvier 1994, cet établissement a été fusionné avec la Bibliothèque nationale pour former la Bibliothèque nationale de France.

2. CCNPS : Catalogue collectif national des publications en série.

Répertoire national des bibliothèques et centres de documentation (RNBCD). Ce répertoire décrira non seulement les bibliothèques partenaires du CCF, mais aussi l'ensemble des bibliothèques et centres de documentation, avec notamment une description aussi précise et complète que possible de leurs fonds. Il se comportera donc comme un fichier autonome permettant des recherches thématiques et/ou géographiques.

Son point de départ devrait être fourni par les données du RBCCN (Répertoire des bibliothèques du CCN) et d'ORIADOC³ qui seront récupérées puis, pour une partie d'entre elles, mises à jour pour former le premier noyau du RNBCD. A noter que cette récupération de données ne préjuge pas aujourd'hui des modes de gestion et de mise à jour de ce répertoire.

Les *principes* qui guident les travaux du schéma directeur informatique opérationnel sont les suivants :

- *fonctionnalités propres* : le CCF est avant tout un outil d'identification et de localisation des documents. Il n'a donc pas pour vocation de permettre le catalogage partagé ni la dérivation de notices. Il n'a pas non plus pour objectif premier d'être un outil de recherche documentaire. Ainsi il n'offrira pas à l'ouverture une fonctionnalité de recherche par sujets, en raison de l'extrême hétérogénéité des notices auxquelles il donnera accès ;
- *interfaçage avec les services de prêt entre bibliothèques et de fourniture de documents à distance* : la continuité entre la recherche de localisations et un système d'accès aux documents, notamment à distance, sera assurée, ainsi qu'un système de paiement par monétique ou par compte individuel. Cet interfaçage entre le CCF et les services de prêt entre bibliothèques se fera en respect des normes internationales existantes ;
- *interfaçage avec les autres systèmes* : au-delà de ses fonctionnalités propres (identification, localisation, répertoire des bibliothèques, accès aux services de prêt entre bibliothèques), le CCF doit pouvoir s'inscrire comme un maillon des réseaux des systèmes documentaires nationaux et internationaux. Il devra donc pouvoir être interrogé par l'intermédiaire des principaux réseaux de

télécommunication, en particulier en France TRANSPAC et RENATER⁴ et au plan international INTERNET⁵. De même, il permettra, en particulier après consultation du répertoire des bibliothèques, de se connecter à ces systèmes externes *via* ces mêmes réseaux ;

- *principe de non-duplication* : le CCF essaiera de limiter la duplication physique des bases auxquelles il donnera accès, pour éviter des surcoûts importants de gestion et faciliter la cohérence des données accédées grâce au CCF.

ABF – Vous disiez que l'une des étapes consistait à étudier divers scénarios, puis à choisir l'un d'entre eux. Quel est le choix qui a été fait en ce domaine ?

J. B. – Trois scénarios ont été étudiés.

Un fichier centralisé. Le CCF serait constitué d'un seul fichier de 12 à 13 millions de notices, comprenant des données de diverses provenances (BNF, Pancatalogue, CCNPS, Téléthèses, CCO, base des rétroconversions de fichiers de BM, pôles associés à la BNF...), dans des formats différents, avec des différences sensibles de pratiques de catalogage et de libellés des points d'accès selon les sources et les époques de rédaction. Ce scénario présente de nombreux inconvénients. Sur le plan bibliographique, le risque est important, en particulier au niveau de la constitution et surtout de l'exploitation des index, d'aboutir à un fichier difficilement exploitable par les utilisateurs, tant les données qui y seraient rassemblées seraient hétérogènes. Il ne respecte pas le principe de non-duplication des bases existantes qui avait été défini en étape 2 du schéma directeur. A quoi bon en effet dupliquer et maintenir sur plusieurs sites des bases de la taille de celle de la BNF ou du Pancatalogue par exemple, alors qu'elles sont déjà accessibles en ligne ? Exploitation (inutilement) coûteuse en termes de dimensionnement des machines, de taille des équipes d'exploitation et de lourdeur des mises à jour en raison de la taille du fichier.

Pour ces motifs ce scénario n'a pas été retenu par le groupe de projet, ni par le comité de pilotage.

Une solution intermédiaire. Dans ce cas la base de données du CCF ne contient pas une copie intégrale des notices

bibliographiques de ses partenaires mais des notices réduites constituées des principales zones (principalement les points d'accès). Ces notices réduites sont régulièrement chargées et réindexées. Un lien (ou pointeur) permet, à partir d'une recherche fructueuse (ou qui semble l'être) des utilisateurs sur la base des notices réduites, de désigner la ou les notices bibliographiques s'y rapportant (base d'origine et identifiants). L'accès à la notice bibliographique d'origine s'opère alors en allant interroger directement en temps réel le serveur qui la détient, ceci de façon « transparente » pour l'utilisateur. En simplifiant, l'idée est de faire ainsi de la base de données du CCF un serveur d'index permettant un premier niveau de recherches *in situ*, susceptible au mieux de répondre directement aux utilisateurs dans un certain nombre de cas (pour les notices simples et sans ambiguïté), ou en tout cas d'orienter avec sûreté les requêtes vers les serveurs adéquats.

Ainsi ce scénario pourrait présenter un intérêt dans la mesure où la consultation de la notice réduite serait souvent suffisante pour identifier un document sans avoir à accéder à la notice bibliographique complète située sur un serveur distant.

En fait, compte tenu de l'hétérogénéité des données, de leur fort taux probable de doublonnage, des habitudes constatées des chercheurs, il y a tout lieu de penser que le recours à la notice complète sera nécessaire ou souhaité dans la plupart des cas. En outre, la difficulté de constitution, de maintenance et d'utilisation des index est la même que dans le scénario précédent, à ceci près, qui est un risque très important, que l'on peut craindre à terme une incohérence entre les données du serveur CCF (notamment les liens) et celles des serveurs affiliés (notices complètes et index) difficilement rattrapable autrement qu'en rechargeant les données régulièrement. Enfin les charges d'exploitation restent importantes.

Ce scénario n'a donc pas non plus été retenu pour cet ensemble de raisons.

Une solution « décentralisée ». Dans cette hypothèse, le CCF n'a pas la responsabilité du stockage et de la mise à jour des données bibliographiques et de localisation. Il se comporte comme un système d'interfaçage uniforme envers l'ensemble des serveurs associés à qui il adresse en ligne les requêtes de ses utilisateurs. La base de données gérées par le CCF (par exemple les don-

3. ORIADOC : répertoire d'orientation et d'accès aux sources d'information et de documentation (géré par une association loi de 1901, sous la responsabilité de la Documentation française).

4. RENATER : réseau national de la technologie, de l'enseignement et de la recherche.

5. INTERNET : c'est en fait l'interconnexion mondiale de l'ensemble des réseaux locaux ou régionaux de télécommunication (environ 12 500 réseaux dans 120 pays) utilisant le protocole IP [Internet protocol].

nées des rétroconversions des BM) est considérée elle-même comme un serveur externe (ce qui ne signifie pas nécessairement distant). Le système propre du CCF est constitué d'un serveur d'accueil dont la fonction est l'accueil de l'ensemble des usagers, l'analyse et la mise en forme de leur requête puis l'attribution, en fonction de cette analyse, des éléments de recherche à un ou plusieurs serveurs distants. Il met en forme en retour les éléments de réponse et les présente à l'utilisateur. Le dialogue peut naturellement être itératif, soit que la requête soit modifiée par l'utilisateur pour tenir compte des premiers niveaux de résultats, soit qu'elle soit réorientée sur d'autres serveurs. Par ailleurs une fonction d'historisation est offerte à l'utilisateur. Le serveur CCF supporte en outre une fonction de service lui permettant, après une recherche fructueuse de localisation, d'accéder à un ou plusieurs services de PIB distants. Il supporte enfin un certain nombre de fonctions de gestion, notamment de monétique pour le paiement des prestations.

Cette solution suppose naturellement le recours à la normalisation internationale pour permettre le dialogue avec des serveurs distants sous systèmes hétérogènes (norme ISO « search and retrieve » n° 10162 et 10163 et/ou norme ANSI Z 39-50 pour les échanges liés à la recherche bibliographique et de localisations, norme ILL pour les échanges avec les systèmes externes de PIB). Ce recours systématique à la normalisation, lorsqu'elle existe, est d'ailleurs l'un des principes qui sous-tend la mise en œuvre du CCF.

Ce scénario présente les avantages suivants :

- il constitue un cadre de développement et d'exploitation bien structuré en distinguant radicalement les données et les services ;
- il autorise l'évolution des différentes bases de données d'une façon indépendante, ainsi que des services indépendamment de ces bases et minimise les difficultés liées à la présence de formats différents dans les bases de données des partenaires. Sur ce point, le format utilisé par le CCF pour la ou les bases de données qu'il sera amené à gérer sera UNIMARC ;
- le système informatique qui équipe le CCF n'a pas besoin d'une grande puissance ;
- ce système ne requiert aucune duplication de données. L'administration s'en trouve minimisée et on est certain de disposer des données les plus à jour ;
- enfin cette architecture permet de mutualiser les ressources d'infrastructure

(charges d'exploitation, facturation, demandes de prêt, accès réseau...), notamment si le système qui supporte le CCF est hébergé physiquement chez l'un de ses principaux partenaires (SUNIST, BNF...).

Il présente néanmoins des contraintes :
- il nécessite l'adaptation des bases de données partenaires pour le support de la normalisation ISO ou ANSI évoquée précédemment, ce qui représente un effort important. Il faut cependant noter que les deux principaux partenaires, la BNF et les applications gérées par la DISTB, s'orientent fermement dans cette direction dans des délais compatibles avec le calendrier de réalisation du CCF.
- cette architecture décentralisée le rend très dépendant des serveurs distants et notamment de leur disponibilité physique, surtout si, comme cela est prévu, le CCF se donne un rôle de « guichet » d'accès à ses utilisateurs, notamment (mais pas seulement) étrangers *via* INTERNET.

Considérant que les avantages de ce scénario l'emportent largement sur ses contraintes, en particulier parce qu'il était le plus porteur d'avenir en exploitant au mieux les avantages de la normalisation, le groupe de projet comme le comité de pilotage ont décidé de le retenir pour la réalisation du CCF.

ABF - Le CCF permettra-t-il l'accès à d'autres systèmes ?

J. B. - Je disais tout à l'heure que le CCF serait accessible *via* INTERNET, c'est-à-dire en France pour le secteur de la recherche par le réseau RENATER. Il donnera donc accès aux serveurs disponibles sur ces réseaux. En fait il y aura en quelque sorte deux niveaux de services : des services intégrés, avec interface unifiée, pour les bases de données partenaires, allant de la recherche jusqu'à la fourniture de documents et paiement des services par monétique ; un « simple » accès par routage sur les autres serveurs de données accessibles en France ou à l'étranger par INTERNET. Dans ce cas de figure, l'utilisateur se trouvera connecté dans l'environnement du serveur visé (par exemple le fichier de la Bibliothèque du Congrès), avec notamment le mode de dialogue et les services propres à ce dernier.

ABF - L'accès au CCF sera-t-il gratuit ?

J. B. - La réponse sur ce point doit être faite à deux niveaux. Sur le principe, il est décidé que les services du CCF (interrogation, accès au PIB) seront payants, ne serait-ce que parce que l'accès par la fonction kiosque banalisée du minitel

rend cette tarification nécessaire. Quant au(x) niveau(x) de tarification, c'est essentiellement une décision politique qui devra, entre autres, prendre en compte la politique tarifaire de ses partenaires. Il serait souhaitable sur ce point que nous puissions définir cette politique tarifaire à l'échelon national.

ABF - Le CCF fournira-t-il d'autres services ou produits que ceux évoqués précédemment ?

J. B. - Outre les services en ligne dont je viens de parler, il cherchera à définir une ligne de produits. On peut penser que l'édition sur CD-ROM de la base de rétroconversion des BM ou bien du RNBCD peut avoir un intérêt pour certaines catégories de public. De même l'édition sur papier de sous-ensembles régionaux du RNBCD (Répertoire national des bibliothèques et centres de documentation) devrait être utile et motiver les bibliothèques signalées dans ce répertoire pour assurer la fraîcheur des données qui les concernent.

Rappelons aussi que le CCF ne fournira pas de notices aux bibliothèques. Le serveur bibliographique national et/ou la Bibliothèque nationale de France, ainsi que le futur système universitaire à horizon 1997, sont là pour cela.

ABF - Quel est le calendrier de réalisation du CCF ?

J. B. - Le schéma directeur informatique opérationnel devrait être terminé début mai. Juste après seront menées ce qu'on appelle l'étude préalable et l'étude détaillée, ce qui nous mènera jusqu'à la fin mars 1995. Puis commencera la phase de réalisation proprement dite qui devrait se dérouler d'avril 1995 à fin mars 1997, avec une mise en place pilote d'avril 1997 à fin juin 1997 et une dernière phase de généralisation jusqu'à la fin de 1997.

Cependant tout sera fait pour qu'un service minimum soit disponible à l'ouverture de la BNF, soit le noyau du RNBCD permettant sa consultation et, suite à celle-ci, la connexion automatique par reroutage des usagers sur les fichiers collectifs et locaux accessibles en ligne (en particulier celui de la BNF et celui des rétroconversions des BM).