

# Z39.50

par Jean-Marc Czaplinski et Yves Moret

*Bibliothèque nationale de France*

*Direction de l'informatique et des nouvelles technologies*

## Introduction

Qu'est-ce que Z39.50 ? C'est un protocole d'interrogation de bases de données bibliographiques, documentaires, ou autres. Ce protocole a déjà fait l'objet de plusieurs versions – 3 à ce jour – mises au point au sein du ZIG (Z39.50 Implementors Group) et présentées aux diverses instances de normalisation nord-américaines et mondiales. Cette norme définit un protocole d'interrogation et de réponse suivant un mode client-serveur, devant permettre à tout utilisateur disposant d'un logiciel client Z39.50 d'interroger toute base d'information accessible par un serveur conforme à Z39.50. Z39.50 s'appuie sur les normes de communication TCP/IP et se trouve de ce fait parfaitement adapté à Internet. Il est donc facile d'imaginer que l'usage d'un logiciel client conforme à Z39.50 et fonctionnant sur un système relié à Internet permette d'interroger toute base bibliographique reliée à Internet et mettant en œuvre un logiciel serveur conforme à Z39.50.

## Historique

### *Z39.50 et Search and retrieve*

Il s'agit des dénominations des standards ou des normes relatives à la recherche et la récupération d'information, par consultation à distance de bases de données.

Bien qu'aujourd'hui plus souvent citée, Z39.50 est une norme nationale américaine, développée sous les auspices du NISO (National Information Standards Organization), qui est un comité de

l'ANSI (American National Standards Institute). Search and retrieve (SR) ou ISO 10162-10163 est la norme – le couple de normes – internationale élaborée dans le cadre de l'ISO (International Standards Organization). En fait, il s'agit de la partie fonctionnelle (description des services du protocole de recherche d'information) ISO 10162 et de la spécification du protocole technique (modalité d'organisation des données échangées) ISO 10163.

Les deux normes ont été définies pour un environnement OSI. Néanmoins, Z39.50, parce qu'implémentée le plus souvent en environnement Internet, utilise directement les couches TCP/IP. C'est la tendance très largement majoritaire aux États-Unis. C'est pour cette raison que Z39.50 jouit d'un plus grand succès que SR, surtout que SR ne représente qu'un sous-ensemble de Z39.50, qui dispose d'une plus grande richesse fonctionnelle.

Les normes SR ont été approuvées en 1991 et publiées en 1993, et sont actuellement en cours de révision.

La maintenance de deux normes n'est pas sans poser de problèmes, que ce soit d'un point de vue intellectuel (éviter le plus possible les divergences éventuelles, dans la mesure où il s'agit d'apporter une réponse satisfaisante à un même besoin) que d'un point de vue industriel et économique : les développements correspondant à leur implémentation sont coûteux et le marché encore relativement étroit. Dans cette perspective, la convergence et la fusion de SR et Z39.50 en une seule norme ont été évoquées, et l'ISO a décidé d'examiner Z39.50-1995 pour en faire éventuellement la deuxième version de SR. La nouvelle norme devrait être enregistrée sous le code ISO 23950.

### *Les trois versions de Z39.50*

La version 1 date de 1988. Elle n'a que peu de rapport avec les versions 2 et 3. Elle correspond cependant à la problématique de recherche automatisée d'informations bibliographiques. Cette préoccupation a longtemps évolué sans modèle commun, au gré des réalisations des producteurs de systèmes. L'intérêt est donc de disposer des moyens permettant de s'affranchir des différences liées aux particularités des systèmes et des formats. Au milieu des années 1980, une première tentative a abouti à la mise au point et à l'approbation de cette première version de Z39.50. Plusieurs sociétés ont réalisé des développements correspondant à son implémentation ou à celle d'une variante, dont la plus notable est WAIS (Wide Area Information Services).

La norme a été révisée de manière très profonde pour la version suivante (version 2 ou 1992). Une des évolutions les plus notables est la recherche d'un alignement avec SR. En outre, cette version impose le recours à une syntaxe abstraite, qui permet le dialogue client-serveur dans la transmission des unités d'information : ASN1, notation de syntaxe abstraite numéro 1. La Bibliothèque nationale du Canada a largement participé à ces travaux, en assurant notamment la réalisation d'un client Z39.50, avec la société Kinetics, dont elle propose les sources gratuitement.

La version 3, stabilisée en juillet 1995, reprend la version 2 en l'enrichissant d'un certain nombre de fonctions et de services. On peut noter que si la version 2 comptait une cinquantaine de pages, il en faut près de 150 pour l'actuelle version !

Ce que l'on nomme aujourd'hui Z39.50 correspond donc à la version 3, ou au

minimum à la version 2. Il serait erroné de considérer que la version de 1988 peut encore être retenue dans le cadre des débats sur les accès distants aux catalogues de bibliothèques.

Z39.50 est en évolution permanente. Les travaux correspondant aux développements de la norme se tiennent dans le cadre du ZIG (Z39.50 Implementors Group), qui réunit depuis 1990 les représentants d'institutions et de sociétés. La Z39.50 Maintenance Agency, à la Bibliothèque du Congrès, se charge de la mise à jour du registre des implémentations ainsi que de la préparation et la diffusion de la documentation.

## Rappel des grands principes de Z39.50

### Protocole d'interrogation client-serveur

Z39.50 définit un protocole d'interrogation entre un client posant une requête d'interrogation et un serveur effectuant la recherche dans ses bases et renvoyant le résultat trouvé. Le protocole fournit ainsi la nature et le format d'un ensemble de

questions et de réponses – appelés services – sans aucune hypothèse sur la nature, la structure, le format bibliographique et le jeu de caractères employés dans les bases bibliographiques ou documentaires consultées.

Ce protocole applique le principe de fonctionnement client-serveur : un logiciel client Z39.50 s'adresse à un logiciel serveur Z39.50.

### Rôle du logiciel client

Le logiciel client doit prendre en charge les fonctions suivantes :

- le dialogue homme-machine avec l'utilisateur : présentation, interface d'interrogation, ergonomie générale, formatage et présentation des résultats ;
- les fonctionnalités complémentaires à l'interrogation : impression, mémorisation, export, etc. ;
- la traduction en Z39.50 des requêtes utilisateurs d'interrogation, l'adressage des requêtes vers les serveurs interrogés, le décodage des résultats retournés en Z39.50.

Le logiciel client reçoit toutes les informations constituant le résultat dans des formats structurés de stockage ou d'échange classiques ; il a ensuite la

charge de traiter ces informations « brutes » en local pour les présenter à l'utilisateur.

Les seules contraintes pour un réalisateur de logiciel client conforme à Z39.50 sont les suivantes :

- les requêtes « quittant » son logiciel doivent être conformes à Z39.50 (nature, format, structure) donc compréhensibles par un serveur Z39.50 ;
- le logiciel doit être capable de décoder et traiter tout résultat retourné en Z39.50.

### Rôle du logiciel serveur

Le logiciel serveur doit prendre en charge les fonctions suivantes :

- le décodage Z39.50 des requêtes reçues ;
- l'interprétation de cette requête, puis son exécution sur les bases interrogées (la requête est généralement adressée en interne sur le serveur au logiciel de gestion de la base de données) ;
- le formatage puis le codage du résultat trouvé en Z39.50, puis l'adressage du tout au client.

Le logiciel serveur doit envoyer toutes les informations constituant le résultat dans des formats structurés de stockage ou d'échange classiques.

## Exemple de dialogue Z39.50 au cours d'une session

Client	Services Z39.50 échangés	Serveur
Demande ouverture de la session	Init	Init Acceptation ouverture et paramétrage session
Demande d'informations sur les bases accessibles	Explain	Explain Fourniture d'informations sur les bases accessibles
Envoi d'une requête de consultation XXX	Search	Search Exécution de la requête XXX Mémorisation du résultat ( <i>result set XXX</i> )
Demande des premiers éléments contenus dans le <i>result set XXX</i> Travail en local du poste client : affichage, traitements...	Present	Present Retour statut et identifiant du <i>result set XXX</i> Envoi du nombre de notices demandé du <i>result set XXX</i>
Envoi d'une requête de consultation YYY	Search	Search Exécution de la requête YYY Mémorisation du résultat ( <i>result set YYY</i> )
Demande des premiers éléments contenus dans le <i>result set YYY</i> Travail en local du poste client : affichage, traitements...	Present	Present Retour statut et identifiant du <i>result set YYY</i> Envoi du nombre de notices demandé du <i>result set YYY</i>
Demande de suppression du <i>result set YYY</i>	Delete	Delete Suppression mémorisation du <i>result set YYY</i>
Demande des derniers éléments contenus dans le <i>result set XXX</i>	Present	Present Envoi du nombre de notices demandé du <i>result set XXX</i>
Clôture session	Close	Close Clôture session

BnFjmc

## Z39.50 et l'échange de données bibliographiques

### Z39.50 véhicule des formats structurés de données bibliographiques

Le cœur de ce protocole est bien sûr constitué par les services de recherche et de restitution de résultats. La richesse du protocole permet de supporter aussi bien des recherches simples (monocritère) que des recherches plus complexes (multicritères, recherches pondérées, etc.) en fonction des capacités des serveurs interrogés. Les critères de recherche supportés sont nombreux (plus d'une centaine !) et portent aussi bien sur des informations habituellement présentes dans des notices bibliographiques, autorité ou de données locales.

La norme permet la restitution de résultats dans une multitude de formats, donc les principaux formats MARC.

C'est là que réside tout l'intérêt de Z39.50 par rapport au sujet qui nous intéresse : l'échange d'informations bibliographiques en ligne. Conformément à Z39.50, un serveur Z39.50 peut transférer à tout client Z39.50 les informations demandées, non seulement sous une forme textuelle mais aussi sous diverses formes structurées déjà rencontrées dans les domaines bibliothécaire ou documentaire comme les formats MARC, UNIMARC, INTERMARC, USMARC, etc., ISO 2709 sont des notions parfaitement véhiculées par Z39.50. Comme le standard comprend aussi des fonctions de contrôle et vérification d'habilitation, il peut même être envisagé de limiter ou non ces transferts en fonction de l'identification du client.

Z39.50 constitue donc bien un moyen d'interroger à distance une base bibliographique et de récupérer en local des informations bibliographiques sous une forme structurée classique permettant dans un deuxième temps la réutilisation de ces informations en local.

Attention cependant aux « profils » d'implémentation retenus par chacun : formats de données supportés, formats MARC utilisés si les notices sont renvoyées en MARC, jeux de caractères, etc.

Les informations rapatriées sur le poste local le sont dans le cadre de l'architecture client-serveur, qui sous-tend Z39.50. Ce n'est cependant pas forcément le mode idéal pour le transfert en masse de données, même bibliographiques.

Les seules contraintes pour un réalisateur de logiciel serveur conforme à Z39.50 sont les suivantes :

- le logiciel doit être capable de décoder et traiter toute requête Z39.50 ;
- les résultats « quittant » son logiciel doivent être conformes à Z39.50 (nature, format, structure) donc compréhensibles par un client Z39.50.

### Contenu du standard

On trouve défini dans ce standard un ensemble de services, en fait des couples de questions (adressées par un client) réponses (renvoyées par un serveur).

Les principaux services sont les suivants :

- ouverture d'une session d'interrogation par un client, avec négociation des paramètres de la session (taille maximum des messages par exemple, etc.) ;
- obtention des informations générales sur la (ou les) base(s) accessible(s) sur le serveur (nom, description, horaires préférentiels, index disponibles, etc.) ;
- recherche à partir d'une équation de recherche (plusieurs syntaxes disponibles) accompagnée d'une précision sur la nature de la réponse attendue (nombre de notices trouvées, mémorisation du résultat trouvé, listes partielles ou complètes en fonction de seuils préférentiels) et du format préférentiel des données retournées (format MARC, format textuels, etc.) ;
- restitution du résultat d'une recherche (intégralité ou sous-ensemble) ;

- parcours d'un index à partir d'une clé de recherche avec retour des entrées d'index correspondantes ;

- tri et fusion d'un ou plusieurs résultats de précédentes recherches ;
- vérification d'habilitation de consultation (mots de passes, etc.) ;
- évaluation et décompte du coût d'une session de consultation ;
- clôture d'une session de consultation.

### Quelques informations supplémentaires

Z39.50 n'est pas figé : le ZIG examine régulièrement des demandes de modification (nouveaux formats de données, nouvelles grammaires de requêtes, nouveaux critères d'accès, etc.) et d'extension des fonctions couvertes.

Z39.50 n'est pas dédié à l'interrogation de catalogues de bibliothèques : il peut servir à interroger toute base d'information documentaire et véhiculer des informations autres que des notices MARC.

Z39.50 est dédié à la consultation pure : il ne comprend pas de services de mises à jour de bases de données.

Pour étendre son champ d'application, certains implémenteurs ajoutent au standard des fonctions complémentaires (mise à jour, demande de documents ILL, etc.) au moyen de services « propriétaires » ou non encore validés, définis et structu-

rés de manière cohérente avec les services standards. Mais cela reste en dehors du standard (pour le moment !).

### Les « profils » d'implémentation de Z39.50

L'éventail des fonctionnalités offert par Z39.50 (une demi-douzaine de grammaires de requêtes, une centaine de critères d'accès supportés, de multiples formats de restitution des données) en fait toute la richesse potentielle, mais celle-ci reste difficile à atteindre.

En effet, implémenter dans un logiciel serveur ou dans un logiciel client toutes les fonctionnalités avec tous leurs paramètres de fonctionnement représente une tâche à la fois énorme et complexe à réaliser. Pour un logiciel client ou un logiciel serveur, cela voudrait dire être capable de supporter toutes les grammaires de requêtes, tous les critères d'accès, tous les formats de résultats potentiels donc les formats MARC (une douzaine), mais aussi tous les jeux de caractères habituels (ISO 5426, ALA, ISO 8859, ANSI).

Dans ces conditions, tenter d'imposer à court terme le support intégral du contenu du standard à tous les acteurs du monde bibliographique (bibliothèques, fournisseurs de logiciels, etc.) aurait été irréaliste. Aussi, la liste des services, paramètres, formats de données, etc., de

Z39.50 précise pour chacun d'entre eux si son support est obligatoire ou optionnel (la grande majorité).

Le risque est ainsi devenu très grand de rencontrer un client et un serveur incapables de dialoguer efficacement tout en étant conforme à Z39.50 car n'ayant pas retenu les mêmes options. Pour faire face à ce problème, la notion de « profil » a été introduite : elle permet de définir précisément quelles options sont réellement retenues et implémentées dans un logiciel client ou dans un logiciel serveur.

On trouve maintenant publiée une liste de « profils » par les principaux implémenteurs de Z39.50, ce qui permet d'ailleurs de constater que personne ne cherche à couvrir l'intégralité de la norme, surtout en ce qui concerne les différents formats de données supportés. Ainsi, il reste probablement à trouver un client Z39.50 universel capable de travailler avec des notices résultats dans tous les formats MARC supportés par la norme.

### Z39.50 et l'interrogation des catalogues à distance

Z39.50 n'est pas la seule manière de consulter des catalogues de bibliothèques à distance. Depuis plusieurs années déjà, un certain nombre de fournisseurs proposent des extensions permettant la

consultation d'un OPAC par Vidéotex notamment.

*Les clients passifs : Minitel, Telnet (VT 100, 220, 3270, etc.)*

Ce sont des accès très classiques, utilisant des terminaux ou des émulations de terminaux passifs. Plusieurs dizaines de catalogues de bibliothèques de lecture publique sont accessibles par Minitel. Par émulation Telnet, on peut consulter BN-OPALE, BN-OPALINE, les catalogues de certaines BU ou BM. Il n'y a pas ici de « langage universel » d'interrogation : c'est en fonction du système sollicité que telle touche du clavier correspond à telle commande.

### Les accès par le Web

Le nombre de catalogues accessibles par le Web ne cesse d'augmenter : les interfaces proposées aux usagers sont souvent bien plus conviviales et ergonomiques que celles des générations antérieures. Elles permettent, grâce à l'utilisation d'ancres ou de liens hypertextes, une « navigation » intuitive lors de la consultation, qui permet de rebondir d'une information à l'autre. Le Web a recours au protocole HTTP, de type client-serveur, particulièrement simple et robuste.

Ce type d'accès permet la réalisation d'OPAC particulièrement attrayants, sans

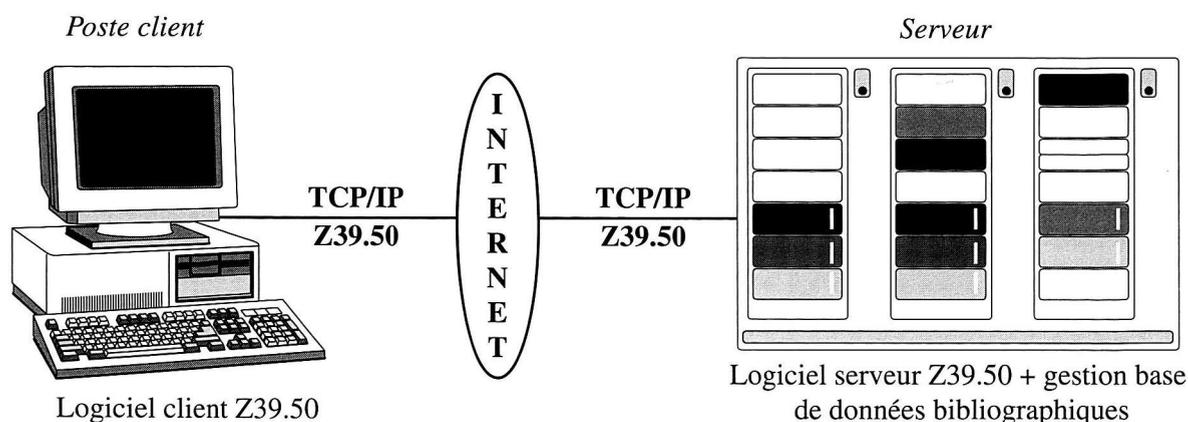
pour autant recourir à l'utilisation de Z39.50. Pourtant, certains serveurs Web permettent d'obtenir la consultation de catalogues *via* Z39.50. Il convient donc pour clarifier ce point un peu obscur de revenir sur l'architecture client-serveur.

### L'utilisation et les implémentations de Z39.50

La norme n'utilise pas les termes de client et de serveur, mais d'*origin* et *target*, origine et cible. C'est de l'origine que partent les demandes et de la cible qu'arrivent les réponses, même si au cours d'une session, les deux points dialoguent sous forme de questions réponses pour échanger des informations relatives notamment aux conditions techniques de la session.

Dans le cas le plus simple, l'origine, et la partie logicielle correspondante, est installée sur le poste de l'utilisateur, la partie cible étant localisée « devant » la base de données interrogée. Ce qui apparaît à l'écran correspond donc à des interfaces homme machine de type Windows. Il n'y a pas de problème particulier lié à ce schéma de fonctionnement, sauf la contrainte posée par le poste de travail (type et configuration suffisante pour assurer une charge d'exploitation relativement gourmande).

### Solution classique client-serveur Z39.50



Une autre solution consiste à télécharger sur un poste client Web, équipé du logiciel client correspondant de type Mosaic, Netscape, Internet Explorer, un *plug in*, un programme complémentaire, qui permet de disposer du logiciel origine, sur le poste de l'utilisateur, à partir du client Web. Ce type de dispositif, annoncé au cours de 1995, a été réalisé sous forme de version bêta, mais ne semble pas avoir connu pour l'instant, un grand succès.

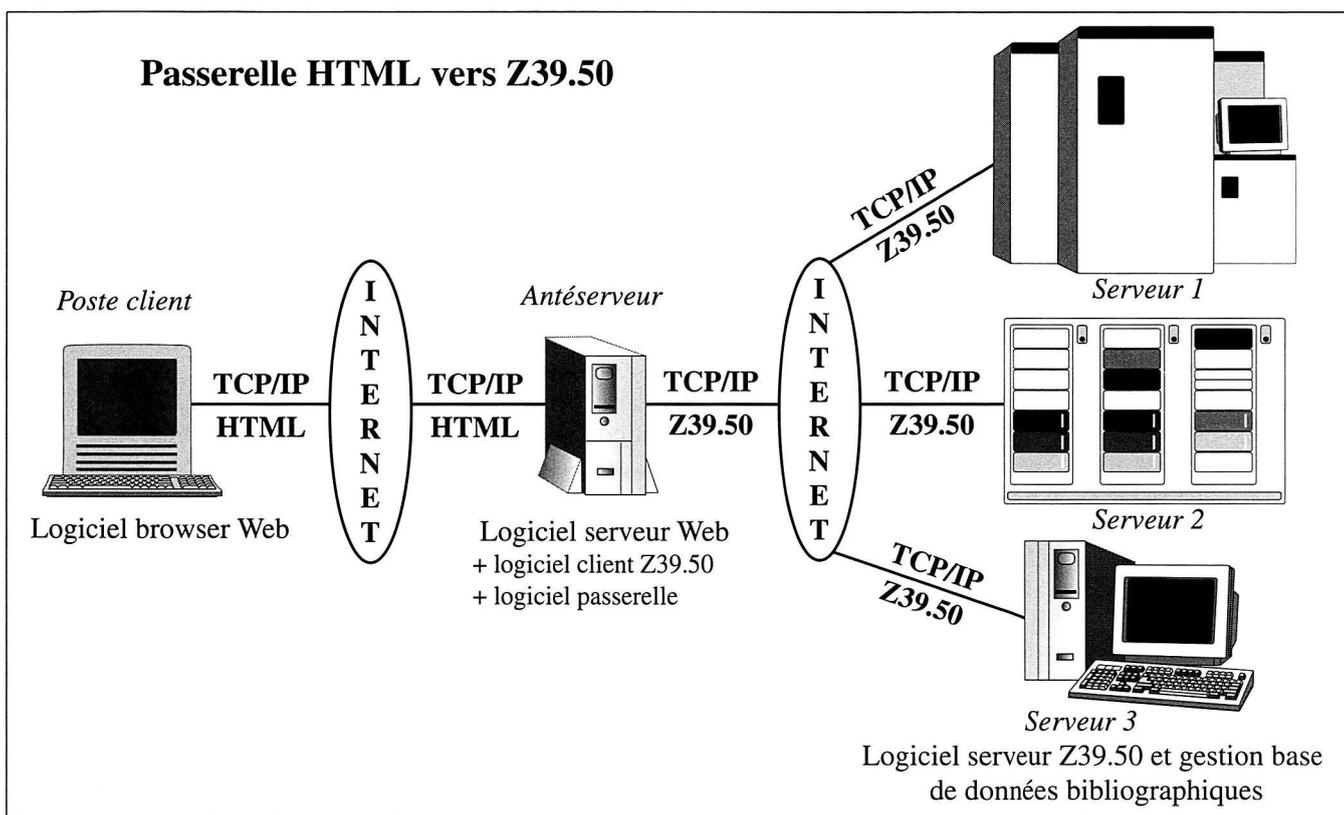
Le type d'implémentation, sans doute le plus répandu actuellement, consiste à « déporter » la partie origine de Z39.50. Dans cette configuration, l'utilisateur dispose sur son poste d'un client Web. Il accède

par une transaction HTTP à un serveur Web, qui offre parmi ses services, l'accès au client ou origine Z39.50. Il y a donc traduction des messages envoyés par l'utilisateur pour les rendre compréhensibles par l'origine, qui dialogue ensuite avec la cible Z39.50. Les réponses reçues de la cible par l'origine sont ensuite renvoyées sur le poste de l'utilisateur, après traduction. Ce système, en quelque sorte « à deux étages », présente plusieurs avantages. Il permet à l'utilisateur de s'affranchir d'un certain nombre de problèmes matériels : les clients Web ont été développés pour tout type de plate-forme (Mac, PC 16 et 32 bits, stations Unix). La maintenance et l'évolution du logiciel origine Z39.50 sont assu-

rées en un site central, ce qui s'avère d'autant plus intéressant que la plupart des développements sont étroitement liés à des projets encore en cours. Enfin, cela permet d'offrir un « bouquet » de catalogues consultables. Il s'agit ici donc d'une passerelle ou d'un guichet (*gateway*).

Vu la richesse de la norme, le nombre de possibilités d'implémentation qu'elle offre tout en respectant la conformité à la version 1995, la plus ou moins grande couverture de fonctions et de services est sans doute liée à la capacité qu'ont les postes de travail de permettre une correcte exploitation de la partie logicielle correspondante.

### Passerelle HTML vers Z39.50



### La BNF, le CCF et Z39.50

Z39.50 sera implémenté dans le système informatique de la BNF sous la forme de serveur Z39.50 : il s'agit d'un moteur de recherche supplémentaire s'appuyant sur l'infrastructure technique auparavant développée pour les autres types d'accès : graphique Windows, texte VT100, Web ou Minitel 40 colonnes.

La version V3 du standard a été retenue. Il sera possible de consulter des notices bibliographiques accompagnées de leurs données locales et des notices d'autorité.

Au vu de la richesse potentielle de la norme Z39.50 et afin d'harmoniser le type de service offert, un « profil » commun a été élaboré par la BNF, le CCF et le ministère chargé des bibliothèques universitaires. Ce profil, « CCFr », devrait permettre le dialogue avec le système informatique de la BNF, le CCF (Catalogue collectif de France) et le SU (Système universitaire, mis en place par l'ABES).

Dans le cadre du CCF, c'est une implémentation de type « passerelle » qui a été retenue, permettant à l'utilisateur de consulter un catalogue virtuel à partir de son poste de travail équipé de son navigateur ou butineur favori. La passerelle comprendra, outre le serveur Web et le client Z39.50, un dispositif permettant de dédoubler et présenter les notices envoyées en réponse par les trois catalogues consultés : celui de la BNF, celui du SU et celui constitué par les produits des campagnes de conversion rétrospective des catalogues des fonds anciens, locaux, régionaux ou particuliers d'une cinquantaine de bibliothèques municipales.