

Comprendre les protocoles d'échange à travers deux exemples : Z39-50 et OAI-PMH

1. L'échange de données sur le web via les protocoles d'échange

L'échange, la circulation, et la transmission d'informations au sein d'un réseau, qui caractérisent l'action de communication sur le web, impliquent que les divers acteurs suivent des règles communes pour se comprendre mutuellement. Dans ce contexte de partage d'informations, la notion d'**interopérabilité** occupe une place centrale : il s'agit de la « capacité de deux systèmes à dialoguer et à interagir" (M. Foulonneau), via un ensemble de règles qui régissent les échanges et permettent un dialogue entre deux machines pour atteindre un objectif donné.

Dans ce cadre, un protocole d'échange peut se définir ainsi : « tout échange d'information repose nécessairement sur un ensemble de conventions partagées entre l'émetteur et le destinataire d'un message : il faut que l'un et l'autre sachent notamment à quel moment commence la communication, selon quelles procédures elle s'effectue, et à quel moment elle se termine » (BnF). Il existe des protocoles de communication et d'échange couramment utilisés sur le web, comme http (Hypertext Transfer Protocol). Un protocole consiste donc en un ensemble de procédures et de formats rendant possible l'échange d'informations. Il permet à deux acteurs, un client et un serveur, d'interagir. Le client demande l'exécution d'un service via une requête normalisée, tandis que le serveur réagit et répond également de façon normalisée à cette demande en exécutant le service.

Les bibliothèques partagent et échangent des données entre elles, structurées selon différents formats (Unimarc, Marc 21, Dublin Core...), en utilisant des protocoles de type Z39-50 ou OAI-PMH, aux modalités et aux finalités différentes, mais qui permettent un accès aux données du web invisible.

1. Le protocole Z39-50

Définition / repères historiques : Z39-50 est à l'origine une norme américaine publiée pour la 1^{ère} fois en 1988, qui a connu deux autres versions depuis (1992 et 1995, avec révision en 2003). Elle est gérée par la [Z39-50 maintenance agency](#), à la Bibliothèque du Congrès. La norme a été reconnue et adoptée en 1997 sous le nom d'ISO 23950 par l'International Organization for Standardization. La version 3 fonctionne avec internet et propose notamment l'interrogation sur les données locales.

La norme Z39-50 est un protocole de recherche et d'échange d'informations bibliographiques en ligne, considéré comme pionnier des protocoles pour l'interrogation à distance. Il consiste en une série de requêtes et de réponses selon un modèle client Z39-50-serveur Z39-50 (*origin* et *target* dans la norme) via une requête simultanée en mode synchrone de plusieurs bases de données bibliographiques à distance (dont les catalogues de bibliothèques), même si celles-ci utilisent des logiciels différents : la requête est réalisée en direct sur un serveur distant et la réponse est immédiate, par la récupération en local des informations bibliographiques sous forme structurée, permettant ainsi leur réutilisation. La réponse se caractérise par l'envoi des résultats trouvés sous forme textuelle ou dans un format structuré (Marc, Unimarc...). L'interrogation se fait donc via un client Z39-50, elle peut aussi être réalisée via le protocole tcp/ip (Transmission Control Protocol et Internet Protocol), ou encore via le protocole web http.

Dérivés du protocole Z39-50 et également maintenus par la Bibliothèque du Congrès depuis 2001, les protocoles SRW (Search/Retrieve for the Web) et SRU (Search and Retrieve via URL) constituent une évolution de Z39-50 adapté aux technologies du web (protocole HTP et format XML).

Avantages, utilité : possibilité de faire des recherches et de créer des catalogues communs virtuels en rassemblant artificiellement des catalogues qui ne fonctionnent pas avec les mêmes logiciels, sans que l'utilisateur en ait conscience et en temps réel.

Limites : fonctionnement de Z39.50 complexe par rapport à celui d'internet (nombre de procédure limité, tendance à généraliser les formats XML) ; coût d'implémentation ; problème d'interopérabilité (requêtes Z39-50 parfois incompatibles entre divers SIGB) ; temps d'interrogation.

Applications : Les bibliothèques utilisent communément cette fonctionnalité via leurs SIGB, ou dans le cadre de catalogues collectifs virtuels (BnF, Sudoc, CCFr, réseau Lectura, réseau Nebis...).



2. Le protocole OAI-PMH

Définition / repères historiques : OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), a été élaboré dans le cadre du mouvement des archives ouvertes (open access) en fédérant les contenus sur le web, d'abord par l'Open Archive Initiative lors de la convention de Santa Fé en 1999, puis en 2002 par la BOAI (Budapest Open Access Initiative), lors d'un appel visant à définir le protocole. Depuis 2002, la version 2 est considérée comme définitive.

OAI-PMH est un protocole de moissonnage facilitant la collecte et l'échange de données, qui vise à "créer, alimenter et tenir à jour [...] des réservoirs d'enregistrement qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine" (F. Nawrocki). L'utilisation du protocole OAI-PMH est libre et repose sur un modèle client-serveur : un fournisseur de données, qui dépose ses métadonnées exposées dans un entrepôt (ou serveur web), communique avec un fournisseur de services (ou agrégateur) qui collecte les données accessibles pour les intégrer à ses propres entrepôts de données, par exemple dans une bibliothèque numérique. Un même établissement peut être à la fois fournisseur de données et de services. Le moissonnage des métadonnées s'appuie sur des standards existants : il s'opère via une requête du client au serveur en HTTP ; le serveur envoie les réponses au format XML, généralement en Dublin Core, parfois en MarcXML.

Avantages, utilité : facilite les échanges de données sur le web entre plusieurs institutions ; accroît la visibilité des collections numériques en signalant des ressources du web invisible ; fournit un accès commun à des ressources réparties sur différents sites, aux origines diverses, en reconstituant virtuellement des collections.

Limites : état des données reflète le dernier passage du moissonneur (pas de mise à jour en temps réel) ; notion de *record* propre au Dublin Core décrit un ensemble fini de données, avec un modèle similaire aux notices papier et calqué sur le web. Une évolution est envisagée, l'OAI-ORE (Open Archives Initiative Object Reuse and Exchange), qui se rapproche du web sémantique.

Quelques applications : les entrepôts de la BnF OAI-NUM (données de Gallica) et OAI-CAT (données du catalogue général) ; le moteur de recherche OAIster, désormais intégré au catalogue OCLC ; l'archive ouverte pluridisciplinaire française HAL (CNRS) ; la plateforme Isidore, spécialisée en SHS...

Z39.50 interroge simultanément des bases hétérogènes, fédère les diverses ressources et les affiche en temps réel ; OAI-PMH moissonne et rend accessible des données collectées simultanément.

Les deux protocoles se complètent et peuvent être utilisés conjointement, en proposant à la fois les résultats de l'interrogation de catalogue distants et du moissonnage de données.

Bibliographie

Textes de référence

- [Agence de gestion de la Z39-50 : Z39.50 International Standard Maintenance Agency](#)
- [Texte de la norme Z39-50 : Z39.50 Text](#)
- [Site de l'OAI : Open Archives Initiative](#)
- [Spécifications du protocole OAI-PMH : The Open Archives Initiative Protocol for Metadata Harvesting](#)

Webographie

- [Protocoles d'échange de données, BnF, 2011](#)
- [Protocole OAI-PMH, BnF, 2012](#)
- [Protocole Z39-50, BnF, 2013](#)
- [Z 39.50, Jean-Marc Czaplinski, Yves Moret, Bulletin d'information de l'abf, 1997, n°174](#)
- [Le protocole OAI-PMH : éléments de compréhension, Invisu, 2014](#)
- [Le protocole OAI et ses usages en bibliothèque, François Nawrocki, Ministère de la Culture, 2005](#)
- [Présentation de l'OAI-PMH : historique, objectifs, règles de fonctionnement, Guillaume Godet, formation BnF, 2013](#)

Fiche créée par Charlotte Vignet, le 05 décembre 2014

Liens vérifiés par l'Esssib, le 20 juillet 2016