



**HAL**  
open science

# De l'usage du Web de données pour une recherche efficace sur des ressources disséminées et hétérogènes : la mise en place d'un portail de recherche fédérée pour le Musée National du Sport

Brice Ravet

## ► To cite this version:

Brice Ravet. De l'usage du Web de données pour une recherche efficace sur des ressources disséminées et hétérogènes : la mise en place d'un portail de recherche fédérée pour le Musée National du Sport. domain\_shs.info.docu. 2011. mem\_00679884

**HAL Id: mem\_00679884**

**[https://memic.ccsd.cnrs.fr/mem\\_00679884v1](https://memic.ccsd.cnrs.fr/mem_00679884v1)**

Submitted on 16 Mar 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

Ecole Management et Société-Département CITS

INTD

MEMOIRE pour obtenir le  
Titre professionnel "Chef de projet en ingénierie documentaire" INTD  
niveau I

Présenté et soutenu par

*Brice Ravet*

le 14 novembre 2011

De l'usage du Web de données pour une recherche  
efficace sur des ressources disséminées et  
hétérogènes

La mise en place d'un portail de recherche fédérée pour  
le Musée National du Sport

Jury

Cécile Kattnig  
Laurène Rolland-Bertrand

**Promotion XLI**

# Remerciements

Je tiens à remercier Laurène Rolland-Bertrand et Circé Krouch-Guilhem pour m'avoir accordé leur confiance et mis dans les meilleures dispositions pour mener à bien ma mission mais aussi tous les autres membres du Musée National du Sport avec lesquels les rapports professionnels et humains furent aussi bien enrichissants que constructifs.

Merci à Cécile Kattnig qui a su m'orienter dans mon travail et m'apporter des pistes de réflexion intéressantes dans le cadre de ce mémoire.

Merci enfin à mes proches et mes camarades de promotion qui ont fait preuve d'une patience infinie et n'ont pas hésité à m'apporter leur soutien au cours de cette année.

# Notice

## **Description bibliographique :**

RAVET Brice. De l'usage des outils du Web de données pour une recherche efficace sur des ressources disséminées et hétérogènes : la mise en place d'un portail de recherche fédérée pour le Musée National du Sport. Mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire », Paris : INTD, 2011, 187 p.

## **Résumé :**

Les portails documentaires d'institutions culturelles sont en plein essor car ils représentent une solution pour renouer le contact avec le public, valoriser leurs fonds ainsi que leurs activités et mutualiser leurs ressources avec d'autres institutions. Ces portails proposent, pour la plupart, une interface de recherche fédérée permettant à l'utilisateur, via une interface unique, d'interroger simultanément plusieurs bases de données. L'auteur de ce mémoire, dans le cadre d'une réflexion relative à la mise en place d'un portail de recherche fédérée pour le Musée National du Sport, définit les portails comme des outils permettant de simplifier la recherche d'information grâce à la recherche fédérée dont il rappelle les contraintes techniques ainsi que les limites sémantiques. Ces limites sémantiques doivent être surpassées par l'usage d'outils traditionnels de l'information-documentation : les vocabulaires contrôlés et plus particulièrement les thésaurus pour garantir la pertinence des résultats de recherche mais qui doivent, d'une part, s'adapter aux besoins de chercheurs d'information qui ne sont plus majoritairement des spécialistes et, d'autre part, devenir interopérables avec d'autres systèmes d'organisation des connaissances dans un contexte de partage des ressources électroniques et d'essor du Web de données. Web de données dont l'auteur définira les enjeux, le support sémantique que sont les ontologies, ainsi que les standards sur lesquels il repose avant d'en présenter les apports dans le cadre de la mise en place d'une recherche fédérée sur des bases de données hétérogènes.

## **Mots-clés :**

Recherche d'information, recherche en ligne, portail documentaire, base de données, recherche fédérée, langage contrôlé, thésaurus, norme, système de gestion des connaissances, Web sémantique, Web 3.0, ontologie, RDF, OWL, XML, interopérabilité, référentiel, métadonnée, patrimoine culturel, musée.

# Table des matières

|   |    |
|---|----|
| Remerciements .....   | 2  |
| Notice .....  | 3  |
| Table des matières .....  | 4  |
| <br>  |    |
| Introduction.....   | 10 |
| <br>  |    |
| Première partie Les portails de recherche fédérée : des vitrines mutualisées des activités et des ressources des institutions culturelles ..... | 14 |
| <br>  |    |
| 1 Définir un portail : sa nature, ses composantes, ses objectifs .....  | 15 |
| 1.1 Le portail : un accès simplifié vers des ressources disséminées et disparates ....  | 15 |
| 1.1.1 Le portail : un site web, un point d'entrée aux composantes diverses .....  | 15 |
| 1.1.2 Une brève histoire des portails .....   | 19 |
| 1.1.3 Les apports d'un portail .....  | 20 |
| 1.1.4 Portail et institution culturelle : de la vitrine au dialogue.....  | 22 |
| 1.2 L'essor des portails culturels : un panorama de l'offre muséale et patrimoniale en ligne  | 24 |
| 1.2.1 Identification des portails et de leurs diverses fonctionnalités (liste non exhaustive) .....   | 24 |
| 1.2.1.1 Les portails identifiés pour cet audit.....   | 24 |
| 1.2.1.2 Les modalités de recherche proposées.....   | 25 |
| 1.2.1.3 Visualisation des résultats.....  | 30 |
| 1.2.1.4 Exploitation des résultats/accès aux documents.....   | 31 |
| 1.2.1.5 Gestion du profil des utilisateurs.....   | 33 |
| 1.2.2 Observations .....  | 35 |
| 1.2.2.1 Pour la recherche de documents .....  | 35 |
| 1.2.2.2 Pour l'exploitation des résultats de recherche .....  | 42 |
| 1.2.2.3 Pour l'accessibilité du site .....  | 46 |
| 1.2.3 Les portails d'institutions culturelles : un retard français .....  | 47 |
| 1.3 Mettre en place un portail : une réflexion à mener sur les contraintes potentielles   | 49 |

|       |   |    |
|-------|---|----|
| 1.3.1 | Le CMS comme l'alpha et l'oméga d'une solution portail : une vision erronée et partielle                                    | 49 |
| 1.3.2 | L'ergonomie : un élément à ne pas négliger dans la mise en place d'un portail   | 50 |
| 1.3.3 | Des compétences à mobiliser   | 52 |
| 1.3.4 | Le lancement d'un portail : un projet nécessitant une réflexion en amont  | 53 |
| 2     | Mettre en place une recherche fédérée : réfléchir l'interrogation des bases de données hétérogènes via une interface unique | 54 |
| 2.1   | Définir la recherche fédérée  | 54 |
| 2.1.1 | Les principes d'une recherche fédérée   | 54 |
| 2.1.2 | L'utilité principale de la recherche fédérée : simplifier la démarche de l'utilisateur                                      | 55 |
| 2.1.3 | Sortir d'une interface de recherche « à la Google »   | 57 |
| 2.2   | Un procédé reposant sur des protocoles d'interrogation : la nécessité d'adopter une solution technique adaptée              | 62 |
| 2.2.1 | Présentation des protocoles d'interrogation synchrone et asynchrone   | 62 |
| 2.2.2 | Comparatif des forces et faiblesses des protocoles existants  | 64 |
| 2.2.3 | La solution ODBC  | 65 |
| 2.2.4 | De la fédération des contenus à la fédération des contenants : les « dataware-houses »                                      | 67 |
| 2.3   | La pertinence des résultats : le talon d'Achille des interfaces de recherche fédérée  | 67 |

Deuxième partie Les langages contrôlés à l'heure du Web de données : quand les outils traditionnels garantissent la viabilité des interfaces de recherche des portails en ligne ..... 70

|       |  |    |
|-------|--|----|
| 1     | Le rôle des thésaurus dans la structuration des informations en ligne : le retour des langages contrôlés pour assurer la pertinence de la recherche en ligne | 71 |
| 1.1   | Le thésaurus : définition d'un outil de structuration des informations   | 71 |
| 1.1.1 | Le thésaurus : un langage d'indexation contrôlé  | 71 |
| 1.1.2 | L'essor des thésaurus dans un contexte d'avancée des outils et supports de communication   | 73 |
| 1.1.3 | Le thésaurus : un outil demeurant l'apanage des documentalistes  | 75 |
| 1.2   | Le thésaurus face à Internet : d'un abandon progressif à une hybridation avec les interfaces de recherche statistique  | 76 |
| 1.2.1 | La croissance de l'Internet grand public et le déclin relatif des thésaurus  | 76 |
| 1.2.2 | Des thésaurus plus que jamais nécessaires pour les nouveaux profils de chercheurs d'information confrontés au déluge   | 77 |

|           |  |     |
|-----------|--|-----|
| 1.2.3     | Du thésaurus à l'indexation au thésaurus à la recherche : l'usage d'un vocabulaire contrôlé pour développer des interfaces de recherche simples.....                                       | 78  |
| 1.3       | Le thésaurus face à la mutualisation et à l'agrégation d'informations en ligne : vers des thésaurus interopérables .....   | 82  |
| 1.3.1     | La nécessité d'une interopérabilité sémantique dans le cadre d'une recherche fédérée   | 82  |
| 1.3.2     | Exemples de mise en place de langages contrôlés communs au sein de projets pluri-institutionnels .....   | 83  |
| 1.3.2.1   | La Banque de Données Santé Publique (BDSP) .....   | 83  |
| 1.3.2.2   | Le projet OTAREN du SCEREN-CNDP .....  | 84  |
| 1.3.3     | Les initiatives anglo-saxonnes pour faire évoluer les normes de gestion des thésaurus vers plus d'interopérabilité.....  | 85  |
| 1.3.4     | La constitution de thésaurus interopérables exploitables sur le Web : la norme ISO 25964 .....   | 88  |
| 2         | Des outils sémantiques au service d'une recherche fédérée efficace : les apports du Web de données pour simplifier l'accès à des ressources de plus en plus nombreuses et hétérogènes..... | 91  |
| 2.1       | D'un Web de pages à un Web de données : rendre les ressources en ligne compréhensibles par des machines .....  | 91  |
| 2.1.1     | Une terminologie à définir .....   | 91  |
| 2.1.2     | Du Web sémantique au Web de données : définition des enjeux.....   | 92  |
| 2.1.3     | Un langage pour les machines : la structuration des métadonnées au sein des couches du Web sémantique .....  | 95  |
| 2.1.3.1   | L'URI .....  | 96  |
| 2.1.3.2   | Le XML.....  | 97  |
| 2.1.3.3   | RDF .....  | 97  |
| 2.1.3.4   | Les autres couches.....  | 100 |
| 2.2       | Les ontologies comme fondement sémantique du Web de données.....   | 100 |
| 2.2.1     | Une terminologie floue dans la littérature scientifique .....  | 100 |
| 2.2.2     | Ontologies et thésaurus : deux systèmes d'organisation des connaissances distincts   | 103 |
| 2.2.3     | Les ontologies : un pilier du Web sémantique permettant la compréhension des contenus du Web par les machines .....  | 104 |
| 2.2.3.1   | Les ontologies pour organiser et donner du sens aux métadonnées....  | 104 |
| 2.2.3.2   | Des standards pour générer des ontologies : OWL et SKOS .....  | 105 |
| 2.2.3.2.1 | OWL .....  | 105 |
| 2.2.3.2.2 | SKOS .....   | 107 |

|  |   |     |
|--|---|-----|
| 2.2.4  | Les ontologies et le Web : un retour à la médiation des professionnels de l'information-documentation .....                                 | 109 |
| Troisième partie Mettre en place un portail de recherche fédérée au sein du Musée National du Sport : mettre en valeur des ressources hétérogènes via la mise en place d'une interface d'interrogation unique..... |   |     |
|  |   | 112 |
| 1  | Les contraintes techniques de mise en place du portail : réfléchir à l'interrogeabilité pertinente de nos différentes bases de données..... | 116 |
| 1.1  | Rendre les répertoires de ressources interrogeables en ligne .....  | 116 |
| 1.2  | Les concordances à établir entre nos bases de données .....   | 117 |
| 1.2.1  | La concordance des champs .....   | 117 |
| 1.2.2  | La concordance des valeurs : la mise en place de thésaurus transversaux... ..   | 118 |
| 1.2.2.1  | Normaliser les procédures d'indexation .....  | 118 |
| 1.2.2.2  | Mettre en place un moteur de recherche sémantique .....   | 119 |
| 1.3  | Le dédoublement des notices .....   | 123 |
| 1.4  | Le paramétrage des champs affichés à la présentation des notices .....  | 123 |
| 2  | Présentation du portail : ce qu'il doit proposer au futur visiteur.....   | 126 |
| 2.1  | Le contenu éditorial .....  | 126 |
| 2.1.1  | La publication d'articles .....   | 126 |
| 2.1.2  | L'arborescence éditoriale du portail .....  | 130 |
| 2.1.3  | Principes généraux de navigation .....  | 130 |
| 2.2  | Les modalités de recherche sur le portail du Musée National du Sport .....  | 132 |
| 2.2.1  | La recherche simple .....   | 132 |
| 2.2.2  | L'exploration de l'arborescence thématique .....  | 132 |
| 2.2.2.1  | La thématique « Les pratiques sportives » .....   | 133 |
| 2.2.2.2  | La thématique « Les sportifs » .....  | 135 |
| 2.2.2.3  | La thématique « Les événements sportifs ».....  | 136 |
| 2.2.3  | La recherche avancée .....  | 138 |
| 2.2.4  | Le filtrage des résultats par typologie .....   | 140 |
| 2.3  | Les fonctionnalités attendues.....  | 141 |
| 2.3.1  | Les fonctionnalités essentielles.....   | 141 |
| 2.3.2  | Les fonctionnalités optionnelles.....   | 147 |
| 3  | Les préconisations pour mener le projet de portail du Musée National du Sport.....  | 152 |
| 3.1  | Les différentes étapes constituant le plan d'action .....   | 152 |
| 3.2  | Définition des solutions envisageables .....  | 153 |
| 3.2.1  | La solution OAI-PMH .....   | 153 |
| 3.2.2  | La solution Z39.50 (SRU/SRW).....   | 154 |



|       |  |     |
|-------|--|-----|
| 3.2.3 | La solution ODBC.....  | 155 |
| 3.3   | Une piste à suivre : la solution de l'éditeur Armadillo, compromis entre outil de gestion de contenu et interface de recherche fédérée reposant sur des outils sémantiques | 156 |
| 3.3.1 | Audit des éditeurs de solutions portail et critères d'évaluation.....  | 156 |
| 3.3.2 | La sélection d'une solution « tout en un ».....  | 157 |
|       | Conclusion.....  | 160 |
|       | Bibliographie.....   | 164 |
|       | Annexes.....   | 185 |
|       | Annexe 1 Questionnaire pour les professionnels chargés de conduire le projet de portail de leur institution.....   | 186 |



# Introduction

Dans l'ensemble de la littérature scientifique consacrée aux portails Internet, ce dernier outil est défini comme un point d'accès unifié à des ressources disséminées sur le Web. Le portail, dont le premier exemple concret remonte à Yahoo!, doit permettre de simplifier la recherche d'information par l'utilisateur en lui offrant sur une interface unique des points d'accès thématiques sur des ressources identifiées et validées. En effet tous les auteurs s'accordent sur le fait que l'utilisateur, face au déluge informationnel auquel il est confronté sur le Web, a besoin d'être aiguillé dans sa recherche d'information car si celui-ci a l'impression de pouvoir « tout trouver » sur Internet, il trouvera en fait « de tout » et par conséquent même des informations peu voire pas pertinentes, rendant ainsi difficile sa récolte d'informations.

Dans un contexte plus restreint comme celui des institutions culturelles le but d'un portail est, plus que de communiquer avec leurs visiteurs potentiels, de mettre en valeur et mutualiser leurs ressources ainsi que leurs activités selon Patrick Brélions et Bruno Texier (27, BREBION, TEXIER). Pour ce faire les portails d'institutions culturelles doivent fournir une expérience de navigation aussi bien originale qu'intuitive tout en permettant une interrogation simplifiée de leurs ressources comme le préconisent notamment Laure Bourgeaux et Amélie Fresneau pour le portail de la Bibliothèque Publique d'Information (26, BOURGEOUX, FRESNEAU).

Par conséquent une des briques essentielles des portails, outre le système de gestion de contenu (*Content Management System*), est le module de recherche fédérée qui est présenté comme une fonctionnalité permettant à l'utilisateur de mener une requête sur plusieurs répertoires de ressources. Cette recherche fédérée est donc essentielle dans la mesure où, comme pour le portail, elle simplifie la tâche de l'utilisateur et constitue un point d'accès unique aux ressources qui évite à l'utilisateur d'avoir à sélectionner et à interroger successivement les répertoires de ressources pour trouver l'information voulue. Néanmoins des auteurs, tels David Stern (24, STERN), soulignent que toute interface de recherche fédérée fait face au problème de la pertinence des résultats trouvés dans la mesure où les bases de données interrogées peuvent posséder des langages d'indexation différents.

D'où l'intérêt à porter aux vocabulaires contrôlés, notamment les thésaurus. En effet si leur usage apparaît nécessaire à l'interrogation pertinente de bases de données, ces outils demeurent selon Sylvie Dalbin (43, DALBIN) l'apanage d'une recherche experte par des professionnels de l'information-documentation et des chercheurs/scientifiques et restent par conséquent opaques aux nouveaux profils d'utilisateurs d'informations influencés par le mode de recherche proposé par les moteurs de recherche grand public tels que Google. Par ailleurs dans le cadre de projets de portails et de mutualisation des ressources, il semble

nécessaire de réfléchir à l'interopérabilité des thésaurus entre eux et avec d'autres vocabulaires contrôlés selon Hélène Zysman et Hélène Rabault (53, ZYSMAN, RABAULT), cette nécessité apparaissant d'autant plus évidente avec l'essor du Web de données.

Le Web de données, s'il est une notion de plus en plus abordée dans les métiers de la documentation, s'illustre par une dénomination assez floue. Tantôt *Web de données*, tantôt *Web sémantique* et parfois *Web 3.0*, il possède une définition néanmoins partagée par l'ensemble des auteurs le définissant comme un Web où la machine parvient à « comprendre » le sens des contenus des pages. Cependant Asuncion Valderrama (64, VALDERRAMA) et Catherine Legg (59, LEGG) estiment que ce « miracle » ne peut être rendu possible que par l'usage en amont d'un savoir-faire info-documentaire, c'est-à-dire par la construction de systèmes d'organisation des connaissances (SOC). Ces SOC devenant, dans le contexte du Web, des ontologies.

Les ontologies, si elles sont entourées selon Bruno Menon (61, MENON) d'un flou terminologique et fonctionnel dans la littérature scientifique qui traite de plus en plus des ontologies dans leur aspect de « fondement sémantique » du Web de données, permettent de structurer le Web non plus sur des pages mais autour de concepts au sein desquels les informations sont sémantiquement liées ou potentiellement reliables à d'autres informations grâce à la souplesse des standards du Web de données tels que RDF, OWL ou encore SKOS. Or dans le cadre de la mise en place d'un portail de recherche fédérée pour le Musée National du Sport, le recours aux outils du Web de données est une idée qui peut être jugée pertinente dans la mesure où ceux-ci sont susceptibles de répondre aux besoins et aux attentes relatifs à ce futur portail. En effet, à l'égard des futurs utilisateurs, il apparaît nécessaire de simplifier la recherche, et même de la rendre intuitive et agréable, en réglant le problème d'interopérabilité sémantique due à l'indexation différenciée sur les bases de données du Musée. Ainsi la valorisation des fonds du Musée National du Sport doit nécessairement passer par la structuration de ses ressources autour de concepts clairement identifiables par l'utilisateur afin de lui permettre une recherche simple et pertinente. Par ailleurs la souplesse et l'ouverture des SOC permises par les outils du Web de données s'avèreraient judicieuses dans le cadre d'une possible mutualisation des ressources du Musée National du Sport avec d'autres institutions culturelles et sportives telles que l'INSEP ou le Musée Olympique du CIO...

Ces réflexions nous amènent à nous demander dans quelle mesure les outils et modes de représentation du Web de données représentent un intérêt dans le cadre de la mise en place d'un portail de recherche fédérée visant à la mise en valeur des fonds d'une institution culturelle.

Ainsi nous nous intéresserons dans un premier temps à définir les portails Internet, leurs fonctionnalités, leurs composantes et leurs objectifs dans un contexte culturel tout en développant une réflexion sur l'une de ses composantes essentielles, la recherche fédérée, dont nous étudierons les problématiques techniques et sémantiques.

Cela nous amènera dans un deuxième temps à appréhender le rôle des vocabulaires contrôlés, et plus particulièrement les thésaurus, pour les solutions de type portail documentaire dans un contexte d'émergence du Web de données.

Enfin nous définirons dans la dernière partie de ce mémoire les préconisations que nous avons apportées dans le cadre de la mise en place du futur portail de recherche fédérée du Musée National du Sport.

Première partie  
Les portails de recherche fédérée :  
des vitrines mutualisées des activités  
et des ressources des institutions  
culturelles

# 1 Définir un portail : sa nature, ses composantes, ses objectifs

---

## 1.1 Le portail : un accès simplifié vers des ressources disséminées et disparates

### 1.1.1 Le portail : un site web, un point d'entrée aux composantes diverses

La notion de portail désigne un site web qui offre un point d'accès unique à de multiples services et ressources documentaires. Un portail peut être aussi bien généraliste et s'adresser de ce fait à un public divers et varié, que thématique et s'adresser à des spécialistes, ou communautaire et s'adresser à un groupe d'individus réunis par un intérêt commun. Enfin un portail peut aussi être commercial ou institutionnel en étant l'outil d'une entreprise privée ou publique. Un portail documentaire institutionnel répond à des besoins spécifiques des institutions et de leurs usagers. En ce qui concerne l'information, les portails peuvent aussi bien intégrer des informations internes qu'externes, structurées (bases de données, documents) ou non structurées (mails, messages de forums, notes etc.).

Sandrine Trillaud (38, TRILLAUD) distingue quant à elle deux types de portails :

- les portails horizontaux, qui s'adressent à l'ensemble des utilisateurs (Yahoo!, par exemple, ou bien la partie d'un portail d'entreprise accessible à l'ensemble du personnel)
- Les portails verticaux, également appelés « portails de niche » ou « sous-portails », qui s'adressent à un public spécialisé, traitent d'un thème précis ou d'un métier particulier

Il convient néanmoins de distinguer les portails « classiques » des portails qualifiés de « documentaires ». Le portail documentaire est un site web ayant une fonction centralisatrice, comme tout autre portail, mais ce qui fait sa distinction si l'on s'inspire d'Aurélia Giusti, c'est qu'il met l'accent sur la recherche fédérée qui permet, via une interface de recherche unique, d'opérer une recherche sur un ensemble de bases de données documentaires et de voir les résultats de cette requête s'afficher sur une seule et même interface, celle du portail (20, GIUSTI). De ce fait un portail documentaire se distingue d'un



portail « classique » dans la mesure où il bénéficie d'une valeur ajoutée documentaire : il n'est pas seulement une fenêtre vers des ressources, un ensemble de liens, mais il permet à l'utilisateur de construire une recherche pertinente par la réalisation, en amont, d'un paramétrage des modalités de recherche, d'une réalisation des bases de données liées aux portails par exemple.

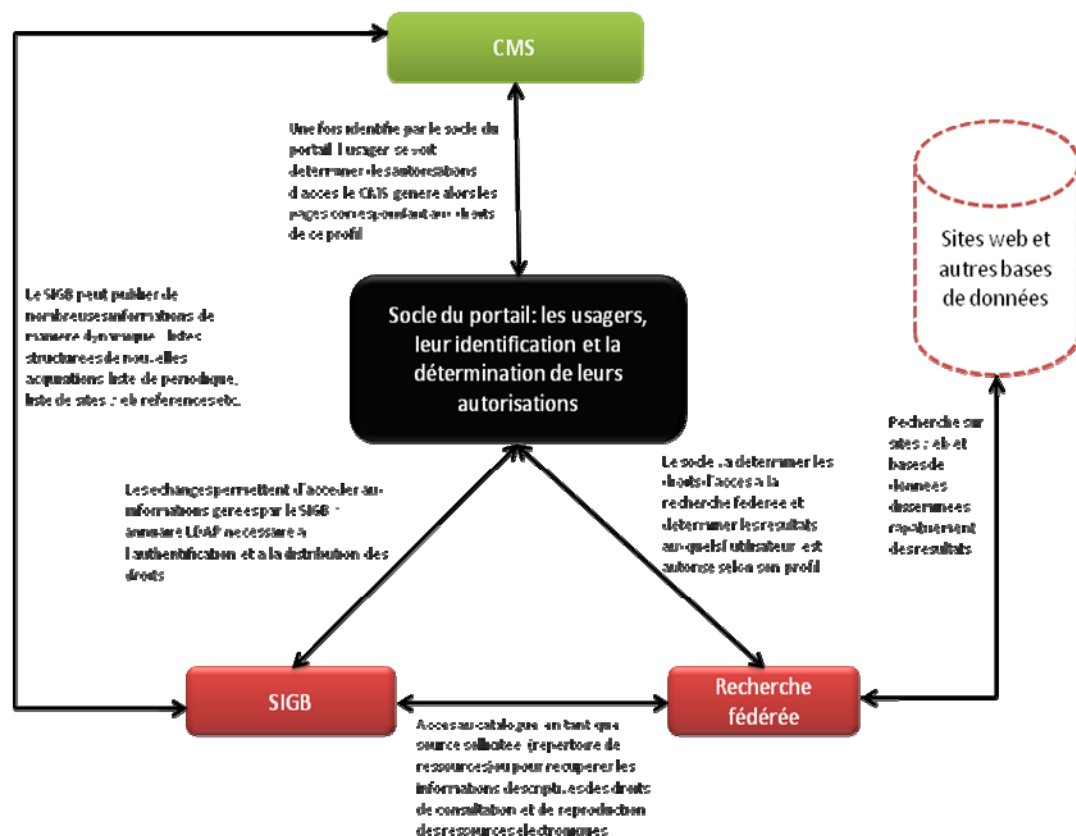
Le portail en tant qu'outil ne doit pas être vu comme un bloc, mais comme un assemblage de composants (37, MAISONNEUVE). Parmi ces composants nous retrouvons :

- La gestion des profils d'utilisateur : un logiciel permet l'identification des usagers (qui doivent renseigner un identifiant au portail), l'authentification de leur identité (les usagers doivent saisir un mot de passe que le logiciel vérifie dans la base des usagers répertoriés, le plus souvent un annuaire LDAP) et enfin la détermination des autorisations d'accès (détermine quels sont les droits d'un usager selon son profil).
- Le système de gestion de contenu est un logiciel permettant de gérer de manière aisée et efficace les pages statiques et dynamiques d'un site web. Il dispose de plusieurs fonctions : gérer l'intervention de plusieurs rédacteurs sur un document ; séparer les opérations de rédaction et de mise en forme pour la publication sur le web ; offrir des services de communication entre les usagers et les bibliothécaires (forums de discussion, intervention dans un blog, chat etc.) ; organiser l'accès des utilisateurs du site web à des contenus structurés (FAQ, fils RSS) ; structurer le portail en un ensemble de pages HTML agréable à consulter et d'un usage facile. Ces fonctions doivent permettre une administration facilitée du site, aussi bien pour la création du site que pour sa mise à jour régulière. Elles doivent simplifier d'une part l'ajout de textes et d'illustrations et d'autre part la mise en forme graphique de l'ensemble. Un système de gestion de contenus dispose de multiples fonctions :
  - Gestion du « workflow » des documents, de leur rédaction à leur publication par un ou plusieurs contributeurs
  - Séparation des opérations de rédaction et de mise en forme pour publier en ligne. Les contributeurs ne s'occupent que de la partie rédactionnelle, l'aspect programmation (code HTML) est automatisée grâce à des feuilles de style CSS (*Cascading Style Sheets*) et au XSL (*eXtensible Stylesheet Language*) qui permettent de définir le format d'affichage des données et de choisir leur ordre d'apparition tout en

facilitant la mise en forme et en garantissant le respect de la charte graphique (une présentation homogénéisée)

- Services de communication avec les usagers (forums de discussion, fenêtre de *chat* etc.)
  - Organisation de l'accès des utilisateurs du portail à des contenus structurés (fils RSS, FAQ etc.)
  - Structuration du portail en un ensemble de pages HTML, agréable à consulter et d'un usage facile
- 
- La recherche fédérée est le deuxième composant essentiel d'un portail d'un centre de documentation ou d'une bibliothèque. Il s'agit d'un service proposé à l'utilisateur lui permettant en une seule requête de consulter plusieurs sources de données, hétérogènes tant du point de vue des formats que des contenus. Un logiciel de recherche fédérée dispose de trois fonctions essentielles : présentation structurée des sources d'information interrogeables afin de faciliter la sélection par l'utilisateur des bases à solliciter (ce qui, dans l'optique d'une interface accessible à des utilisateurs profanes en matière de recherche, ne constitue pas une priorité, l'hétérogénéité des répertoires de ressources devant être confinée au back-office) ; saisie de la requête de l'utilisateur et transmission de celle-ci aux bases qu'il a sélectionnées dans une syntaxe adaptée aux caractéristiques de chacune ; affichage des résultats de la recherche, avec ou sans dédoublonnage, avec ou sans regroupement des réponses apparentées.
  - Le système de gestion de bibliothèque ou le système documentaire. Selon Marc Maisonneuve, il s'agit d'une « brique » essentielle mais pourtant tout à fait extérieure au portail. Ce SIGB peut d'une part :
    - Etre le gestionnaire de la base des usagers que le portail utilisera pour les contrôles d'identification et pour la détermination des droits
    - Etre la source d'information proposée à l'utilisateur de manière directe (consultation de son compte, de son panier de notice, de son historique de recherche etc.) ou indirect (l'interrogation du catalogue via le moteur de recherche fédérée)

Ces composants sont en constante interaction pour rendre le portail dynamique et interactif. Ces interactions peuvent être représentées selon le schéma suivant, inspiré de l'ouvrage de Marc Maisonneuve (37, MAISONNEUVE) :



Audrey Saint-Georges (35, SAINT-GEORGES) résume un portail autour de fonctions de base. Dans un premier temps la gestion des contenus web facilitant le processus de publication d'information et de mise en ligne de documents grâce aux fonctions de CMS. Puis le moteur de recherche fédérée sur un ensemble de ressources acquises, produites et indexées dans de multiples bases de données. Enfin les services personnalisés liés à la recherche documentaire (impression des notices, panier des résultats, historique des recherches, exports des résultats etc.).

### 1.1.2 Une brève histoire des portails

Le concept de portail est né au début des années 1990 avec le site de Yahoo ! qui en est le modèle fondateur (38, TRILLAUD). Il présente à l'utilisateur un ensemble d'informations provenant de l'Internet et classées par typologie. Le portail permet de répondre au problème de la surabondance des informations sur le web :

- Accès plus rapide à l'information pertinente par rapport à la navigation pure et dure
- Pré-traitement de l'information (par intervention humaine pour Yahoo !)

Le succès de Yahoo ! incite donc des entreprises et des institutions à développer leur propre portail à partir de la fin des années 1990. En effet dès 1999 les institutions et les entreprises disposent de leur propre offre de portail qui s'adresse à un public identifié pour lui donner un point d'accès unique à l'information l'intéressant tout en évitant le bruit caractéristique des recherches d'information sur les moteurs de recherche afin d'éviter de se perdre dans la masse des résultats.

Les bibliothécaires et professionnels de l'information-documentation ont pu voir dans les interfaces de recherche sur Internet une concurrence à leur rôle de médiateur (35, SAINT-GEORGES). En effet aux yeux des utilisateurs, le recours à un professionnel ne semble plus indispensable puisque le moteur de recherche leur permet de trouver eux-mêmes l'information. Or la recherche d'un « profane » sur le web conduit à une profusion d'informations dont une très grande partie n'est pas pertinente et donc non utilisable, ce qui amène à constater que l'on ne trouve donc pas tout, mais de tout sur Internet. Un portail réalisé par un professionnel de l'information-documentation lui permet donc toujours de mener sa recherche, mais cette fois-ci dans un espace au préalable défini et balisé par ce professionnel de manière à lui permettre d'obtenir des résultats pertinents.

Si la première génération des sites de bibliothèques demeure un outil de communication administrative avec les usagers (horaires, présentation des services, accès, contacts etc.), on assiste à une évolution progressive vers des sites proposant à partir de 1992 des catalogues interrogeables en ligne (les OPAC, *Online Public Access Catalog*), des ressources électroniques externes, des documents numérisés téléchargeables, des services de bibliothèque (renseignements, réservations, bulletins bibliographiques etc.). Les portails, qui peuvent être considérés comme la dernière génération de sites, ont hérité des ressources et services des précédentes générations mais leur spécificité réside dans la possibilité de mener une recherche fédérée sur l'ensemble des ressources de la bibliothèque et de personnaliser

le contenu et la forme de l'interface. Selon Claire Scopsi, les portails pour bibliothèques et centres de documentation sont apparus en Amérique du Nord vers 2004 (36, SCOPSI). Elle définit ceux-ci comme très orientés vers la publication de contenus (CMS), intégrant des outils de recherche fédérée permettant de traduire dynamiquement les termes d'une requête dans le langage d'interrogation de sources diverses.

### **1.1.3 Les apports d'un portail**

Les portails permettent de contourner les difficultés liées à la recherche d'information sur Internet telles que la richesse, la variété, la crédibilité et l'instabilité de l'information (35, SAINT-GEORGES). Le portail permet une gestion de contenus hétérogènes qui seront diffusés auprès d'un public large mais catégorisé. Un portail doit donc permettre :

- De constituer un point d'accès unique et centralisé à des ressources multiples internes et externes, avec accès ciblés, ainsi qu'à des applications éventuellement
- D'avoir des ressources pas forcément structurées et de formats pouvant être hétérogènes, la diversité devant être gommée par le portail à l'affichage

Ce qui caractérise le mieux un portail est le fait qu'il soit une entrée agréable et pratique à utiliser vers les informations, c'est-à-dire qu'il est un lieu de transition vers les ressources (29, COURTOIS). Le principal besoin de l'utilisateur auquel répond le portail est de lui éviter de perdre du temps à repérer l'information. Pour cela, le portail doit offrir une sélection organisée d'accès à l'information pour que l'utilisateur n'ait pas à effectuer un travail de repérage des sources potentiellement intéressantes. Cependant, contrairement à ce que préconise Raphaële Courtois dans le cadre d'un portail à l'usage d'un cabinet d'avocats, le portail doit donner envie à l'utilisateur d'y passer du temps si l'on se trouve dans une optique de valorisation des fonds et des travaux d'institutions culturelles. En effet, dans un contexte muséal, le portail constitue une vitrine virtuelle, d'où la nécessité d'une part d'inciter l'utilisateur à « flâner » dans les collections numérisées, comme il le ferait lors d'une visite physique du Musée, et d'autre part d'attiser sa curiosité de façon à l'encourager à se rendre au Musée. Il demeure toutefois que le temps passé à la recherche doit être court, afin de ne pas décourager la découverte.

La navigation au sein des portails est naturellement héritière de celle du Web, elle est donc hypertextuelle. Ce type de navigation réside dans la rapidité des accès à l'information, un

simple clic suffisant pour passer d'une page, article, notice à l'autre. L'hypertextualité est d'autant plus efficace dans une information contextualisée, c'est-à-dire organisée par thématiques, domaines de connaissance et dont la navigation en son sein est explicite (liens facilement repérables et identifiables sur une page, dans un article ou une notice).

Par conséquent le portail apporte une solution à la masse et à la variété des informations, ainsi qu'à leurs localisations disséminées. Or le portail a aussi pour but de replacer l'information dans son contexte, de la relier à d'autres informations pertinentes et de répondre à des questions variées autour d'un même sujet. Il doit de ce fait permettre de :

- Présenter une navigation intuitive à travers la hiérarchisation des sujets
- Apporter du contexte à la recherche et d'éliminer les ambiguïtés
- Suggérer des liens à des informations connexes

Nous pouvons définir un projet de portail comme une volonté de fédérer dans un système d'information unique des ressources internes ou externes préalablement sélectionnées ; d'unifier et simplifier l'accès à des ressources diverses ou non, et proposer de façon intégrée des outils de traitement des contenus ; de personnaliser l'offre d'informations et les modes d'accès d'un groupe d'utilisateurs, voire d'un individu (36, SCOPSI). Ainsi le portail facilite la création de multiples formes de services documentaires tels que les catalogues, les bibliothèques, les newsletters etc. Les applications de type portail disposent également de fonctions d'administration des contenus et des utilisateurs.

Néanmoins il convient de préciser que les portails peuvent prendre des formes diverses car ils peuvent viser des objectifs très différents en terme de communauté, de type de ressources, d'accès aux ressources, de personnalisation directe par l'utilisateur, etc.

Dans le domaine scientifique, le lancement du projet de portail TGE Adonis par le CNRS illustre les motivations qui ont mené la communauté des chercheurs en sciences humaines et sociales (SHS) à ce type de solution. En effet le CNRS lance en 2007 le projet de portail Adonis afin de permettre un accès unifié aux documents numériques et données des sciences humaines et sociales. Les objectifs majeurs d'Adonis sont d'assurer une transdisciplinarité à l'intérieur des sciences humaines et sociales, d'assurer une dimension internationale à la recherche en SHS (participation aux projets DARIAH, CESSDA et CLARIN) et de faire reconnaître l'excellence des recherches à travers des publications. Pour ce faire les équipes de recherche en SHS doivent se saisir de cette opportunité pour mutualiser

intelligemment les moyens attribués. Le président du comité de pilotage d'Adonis, Michel Spiro, définit le TGE comme un outil mis au service d'une communauté scientifique, c'est-à-dire un outil structurant et performant pour cette communauté. C'est un outil coûteux qui aurait dépassé les investissements possibles pour un seul laboratoire, mais dans le cadre d'une mutualisation des ressources il représente en fait un outil d'un excellent rapport qualité/prix permettant le travail en réseau (30, DASSA, PIERRAT). De ce fait nous pouvons observer qu'un portail comme le TGE Adonis répond aux problématiques de visibilité et de viabilité de la production scientifique, favorise la coopération scientifique et permet de surmonter les restrictions budgétaires par une mise en commun des ressources ainsi que des moyens humains et financiers. Ainsi, dans un contexte scientifique, un méta-portail tel que TGE Adonis peut-être défini comme un espace virtuel d'accès à de grands ensembles fédérés de ressources, utiles à une communauté et organisés par une mutualisation des moyens. La communauté scientifique peut, grâce à des outils d'exploitation adaptés, explorer, interroger, analyser, annoter, et personnaliser l'information.

En outre le portail présente des avantages non négligeables dans d'autres domaines, tels que dans le secteur banque-finance-assurances où le Crédit Agricole a également opté pour une solution de type portail (28, BREBION). En effet après avoir regroupé ses documentalistes et ses études économiques dans un service unique, la banque a mis en place un portail documentaire pour faciliter les recherches en octobre 2009. Celui-ci permet d'accéder à la quasi-totalité de ses ressources documentaires en fédérant l'accès à différents intranets documentaires déjà existant (Intradoc, Infosecteur, Kiosqueéco) et en proposant une fonctionnalité de recherche transverse pour tous ces sites. Le portail permet à l'utilisateur de mener une recherche sur un intranet en particulier ou sur l'ensemble des intranets, ce qui permet d'avoir différents types et supports d'informations relatives à la requête de l'utilisateur.

#### **1.1.4 Portail et institution culturelle : de la vitrine au dialogue**

Les portails culturels apparaissent comme des vitrines prestigieuses et incontournables pour les collectivités de l'Etat. Ceux-ci se seraient multipliés en quelques années à l'initiative de villes, communautés urbaines, régions, entreprises (quand elles disposent d'un patrimoine qu'elles veulent mettre en lumière) et de l'Etat (27, BREBION, TEXIER). Nous pouvons voir dans le vaste mouvement de numérisation des fonds le moteur de la création des portails. Par exemple l'Institut National de l'Histoire de l'Art (INHA) met en ligne une partie croissante de son fonds au profit d'un vaste projet de numérisation des collections (en 2009 200.000

pages correspondant à autant d'objets ont été numérisées et mises en ligne selon Patrick Beauchet-Filleau, cofondateur d'Arkhênum, prestataire de numérisation).

Par ailleurs Marc Maisonneuve observe que de nombreuses bibliothèques et centres de documentation bâtissent aussi des portails riches et variés, telles que la Bibliothèque Publique d'Information, la Bibliothèque Municipale de Lyon ou encore la Bibliothèque de la Cité des Sciences et de l'Industrie (31, MAISONNEUVE).

Toutefois, si les problématiques diffèrent selon la nature des institutions culturelles, le Web permet de dépasser les différences, notamment entre bibliothèques et musées. En effet pour ces deux types de structures, le Web ne représente plus seulement une vitrine, mais aussi un espace de communication privilégié avec leurs utilisateurs pour permettre de « répondre à deux axes de développement fort des besoins utilisateurs : la personnalisation et le dialogue » (32, NUTTIN). Le portail permet donc aux institutions culturelles de communiquer et de valoriser le patrimoine de manière interactive. Plus que des outils de gestion, les institutions culturelles demandent des outils de diffusion en ligne.

Le principe des portails culturels est d'apporter une réponse transparente en terme de fonds sollicité par les diverses requêtes de l'internaute. Il permet aussi d'ajouter une valeur ajoutée documentaire par les équipes de documentalistes par le biais de publication de produits tels que des dossiers documentaires, des bulletins bibliographiques, des revues ou panoramas de presse etc. Par ailleurs face à la concurrence des portails commerciaux et grand public, et face aux moteurs de recherche fédérée comme Google, Yahoo !, Exalead etc., les portails documentaires des institutions patrimoniales ont dû s'adapter et proposer un accès aux contenus simplifié, intuitif et ergonomique. L'objectif d'un portail est donc de permettre à l'utilisateur d'accéder à des contenus hétérogènes rapidement et simplement.

De plus ces portails répondent à une volonté de mutualisation des ressources et des moyens dans le secteur culturel, ce qui représente une démarche pragmatique à une période de restrictions budgétaires. En outre dans le cadre d'un projet interinstitutionnel, un projet de portail permet de mettre en valeur les fonds d'une structure peu visible sur le Net, et ce notamment grâce à la possibilité d'une recherche multibases (sur l'ensemble des bases de données des institutions participant au projet).

Les portails culturels les plus récents sont plus aboutis que leurs prédécesseurs, notamment en ce qui concerne l'ergonomie et la gamme des services proposés (feuilletage des catalogues, comptes lecteurs, flux RSS etc.). Néanmoins force est de constater que les réalisations les plus abouties sont portées par des établissements plutôt bien dotés en



moyens techniques et humains (31, MAISONNEUVE). Le portail est un outil exigeant dont la mise en œuvre s'accompagne de quelques écueils qu'il convient d'éviter.

## 1.2 L'essor des portails culturels : un panorama de l'offre muséale et patrimoniale en ligne

### 1.2.1 Identification des portails et de leurs diverses fonctionnalités (liste non exhaustive)

#### 1.2.1.1 Les portails identifiés pour cet audit

| Nom du portail   | Nature du portail   | Public visé              |
|--|---|--------------------------|
| <a href="#">Portail de la Musique Contemporaine</a>                            | Moteur de recherche dans les ressources d'organismes français principalement consacrées à la création musicale savante d'après 1945                 | Tous publics             |
| <a href="#">Portail du Museum Nationale d'Histoire naturelle</a>               | Le portail MUSSI permet d'accéder aux différentes ressources documentaires du Muséum National d'Histoire Naturelle                                  | Chercheurs/ Spécialistes |
| <a href="#">Portail des documents techniques de l'eau</a>                      | Ce portail permet d'accéder à une documentation technique sur l'eau et les milieux aquatiques produite en français                                  | Spécialistes             |
| <a href="#">Portail documentaire de la Cité de l'Histoire de l'Immigration</a> | Accès aux ressources de la médiathèque Abdelmalek Sayad et prochainement à celles de bases partenaires  | Tous publics             |
| <a href="#">Portail du Ministère de l'Écologie et du développement durable</a> | Accès aux ressources du Ministère de l'Écologie, du Développement durable, des Transports et du Logement  | Spécialistes             |
| <a href="#">Portail des arts de la marionnette</a>                             | Permet d'avoir accès aux ressources de différents acteurs des Arts de la Marionnette  | Tous publics             |
| <a href="#">Catalogue de la Médiathèque de la Cité de la Musique</a>           | Accès aux différentes ressources de la Médiathèque de la Cité de la Musique (œuvres musicales, documents, bibliographies)                           | Tous publics             |
| <a href="#">Musée Guimet</a>   | Musée des arts asiatiques   | Tous publics             |
| <a href="#">Musée du Louvre</a>  | Site du Musée du Louvre   | Tous publics             |
| <a href="#">Musée Olympique</a>  | Musée mis en place par le CIO   | Tous publics             |
| <a href="#">ECPAD</a>  | Etablissement de Communication et de Production Audiovisuelle de la Défense   | Tous publics             |
| <a href="#">Musée Mc Cord</a>  | Musée de recherche à vocation éducative qui se consacre à la préservation, à l'étude, à la diffusion et à la mise en valeur de l'histoire du Canada | Tous publics             |
| <a href="#">Musée Nicéphore Niépce</a>   | Musée de la photographie  | Tous publics             |

|   |   |              |
|---|---|--------------|
| <a href="#">Brooklyn Museum</a>                 | Musée du quartier de Brooklyn (New-York, Etats-Unis)  | Tous publics |
| <a href="#">The Metropolitan Museum of Arts</a> | Le MET dispose d'une collection de plus de deux millions d'œuvres d'arts couvrant plus de 5000 ans d'histoire | Tous publics |

### 1.2.1.2 Les modalités de recherche proposées

| Nom du portail  | Recherche simple   | Recherche thématique   | Recherche avancée   | Aide à la recherche   |
|---|--|--|---|---|
| <a href="#">Portail de la Musique Contemporaine</a>             | Oui :<br>-possibilité de faire une recherche par typologie de document               | Non  | Oui :<br>-possibilité de faire une recherche dans une bibliothèque et/ou un fonds particulier ou dans l'ensemble des bibliothèques et des fonds<br><br>-possibilité de faire une recherche par typologie de document<br><br>-dans la recherche par type de document, il est possible de combiner plusieurs critères de recherche et d'opérer des requêtes booléennes (avec l'usage des opérateurs ET, OU, SAUF) | Oui :<br>-Aide à la recherche affichée à l'écran<br><br>-transcodage des index<br><br>-Possibilité d'affiner ses résultats de recherche selon des critères de forme, genre, type, localisation, nom |
| <a href="#">Portail du Museum National d'Histoire naturelle</a> | Oui :<br>-possibilité de faire une recherche dans le catalogue d'une bibliothèque en | Oui :<br>-au sein de la recherche simple et multicritères il est possible de faire une | Oui :<br>-possibilité de combiner plusieurs critères de recherche et d'opérer des   | -Transcodage des index<br><br>-Troncature avec le caractère « * »<br><br>-Fenêtre d'aide à  |

|   |   |   |  |   |
|---|---|---|--|---|
|   | <p>particulier ou dans l'ensemble des catalogues des bibliothèques</p> <p>-possibilité de faire une recherche par typologie de document</p> <p>-possibilité de faire une recherche par langue</p> | <p>recherche par thématique, la plupart des bibliothèques étant définies par leurs spécialités respectives (ex : entomologie, cryptogamie etc.)</p> | <p>requêtes booléennes (avec l'usage des opérateurs ET, OU, SAUF, OU EXCLUSIF)</p>   | <p>la recherche</p> <p>-Possibilité d'affiner ses résultats de recherche</p>  |
| <p><a href="#">Portail des documents techniques de l'eau</a></p>                      | <p>Oui :</p> <p>-La recherche s'effectue sur l'ensemble des sources</p>   | <p>Non</p>  | <p>Oui :</p> <p>-Possibilité de combiner plusieurs critères de recherche (mot, auteur, date, titre, identifiant documentaire etc.)</p>   | <p>Oui :</p> <p>-Possibilité de modifier la liste des bases sur laquelle la requête doit être menée</p> <p>-Troncature autorisée</p> <p>-Possibilité de filtrer les résultats (par auteur, date de publication, nuage de tags, organismes diffuseurs, principaux contributeurs)</p> |
| <p><a href="#">Portail documentaire de la Cité de l'Histoire de l'Immigration</a></p> | <p>Oui</p>  | <p>Oui :</p> <p>-Possibilité de faire une recherche dans le classement thématique des ressources</p>  | <p>Oui :</p> <p>-Possibilité d'associer plusieurs critères de recherche avec des opérateurs ET/OU</p> <p>-Possibilité d'associer plusieurs termes au sein d'un critère avec des opérateurs ET/OU</p> | <p>-Index disponible pour certains critères</p> <p>-Prise en compte de la synonymie des termes</p> <p>-Transcodage des index</p> <p>-Paramétrage du degré de pertinence</p> <p>-Paramétrage des fonds à interroger</p> <p>-Consulter et les recherches effectuées</p>               |

|   |   |  |   |   |
|---|---|--|---|---|
|   |   |  |   | -Ajouter des champs   |
| <a href="#"><u>Portail du Ministère de l'Écologie et du développement durable</u></a> | Oui   | Oui :<br>-Possibilité de sélectionner une des thématiques apparaissant au dessus de la zone de recherche | Oui :<br>-Possibilité d'associer plusieurs critères de recherche (sujet, auteur, mots du titre, lieu, date, type de document) avec les opérateurs ET, OU, SAUF            | -Possibilité d'ajouter un champ de recherche<br><br>-Possibilité de combiner ses recherches<br><br>-Accès à l'index de chaque critère<br><br>-Troncature autorisée avec « * » pour remplacer une chaîne de caractères et « ? » pour remplacer un caractère<br><br>-Fenêtre d'aide à la recherche accessible |
| <a href="#"><u>Portail des arts de la marionnette</u></a>                             | Oui   | Oui :<br>-Possibilité de naviguer dans une arborescence thématique (« Exploration thématique »)          | Oui :<br>-Possibilité de combiner plusieurs critères de recherche avec les opérateurs ET, OU, SAUF<br><br>-Possibilité de faire une recherche sur une source ou plusieurs | -Index accessible pour certains critères de recherche<br><br>-Affinage des résultats possible selon différents critères   |
| <a href="#"><u>Catalogue de la Médiathèque de la Cité de la Musique</u></a>           | Oui :<br>-possibilité de faire une recherche simple de documents ou une recherche simple d'œuvres musicales | Non  | Oui :<br>-Possibilité de combiner plusieurs critères de recherche avec les opérateurs ET, OU, SAUF  | Oui :<br>-troncature automatique proposée pour la recherche avancée de documents<br><br>-accès à un index pour les critères « auteur, interprète ou sujet » et « mots-clés »  |
| <a href="#"><u>Musée Guimet</u></a>   | Oui   | Non  | Non   | Non   |
| <a href="#"><u>Musée du</u></a>   | Oui sur   | Non  | Oui, en affinant  | Possibilité d'affiner   |

|  |  |   |  |   |
|--|--|---|--|---|
| <a href="#">Louvre</a>                 | l'ensemble des bases du Musée et sur les pages du site.                            |   | sa recherche il est possible de sélectionner les critères dans une liste   | sa recherche  |
| <a href="#">Musée Olympique</a>        | Oui : recherche sur les pages du site  | Oui : dans la recherche sur le lecteur multimédia, plusieurs thématiques sont proposées (« au-delà du sport », « légende », « rivalités », « meilleurs d'entre nous » etc.) | Oui :<br>-Possibilité de rechercher un médaillé olympique, un record, un résultat en combinant plusieurs critères (« sport », « olympiade », « médaille », « nom », « record » etc.)<br>ou alors des documents multimédia (photos, vidéos) par l'association de plusieurs critères | -Pour la recherche simple, surlignage des caractères saisis lors de la requête dans les résultats<br><br>-Pour la recherche avancée, liste de choix pour les critères ; possibilité d'affiner ses critères de recherche       |
| <a href="#">ECPAD</a>                  | Oui : recherche sur les bases de données du site : documents photo, vidéos, textes | Non   | Non  | Oui :<br>-Autocomplétion lors de la saisie d'un terme<br><br>-Nuage de mots-clés pour la recherche  |
| <a href="#">Musée Mc Cord</a>          | Oui : recherche sur la collection d'images du Musée                                | Non   | Oui : sélection au préalable des collections ciblées, puis recherche multicritères (mots-clés, objet, sujet, artiste etc.)   | Oui :<br>-Liste de choix pour la plupart des critères de recherche<br><br>-Recherche dans les collections grâce à une « mind map » ( <a href="#">le réseau conceptuel</a> )<br><br>-Rubrique d'aide à la recherche accessible |
| <a href="#">Musée Nicéphore Niépce</a> | Oui : recherche sur les pages du site  | Non   | Non  | Oui :<br>- possibilité d'affiner sa recherche en spécifiant la rubrique visée   |

|  |  |   |  |  |
|--|--|---|--|--|
|  |  |   |  | (« articles », « liens web », « contacts », « catégories », « sections », « fils d'actualité », « commentaires »)  |
| <a href="#"><u>Brooklyn Museum</u></a>                 | Oui :<br>-recherche sur les pages du site grâce à un moteur de recherche Google<br><br>-recherche dans les collections | Oui :<br>-possibilité d'effectuer une recherche d'œuvre sur les thématiques servant à classer les œuvres d'art (par ex : « Arts of Africa », « American Arts », « European Arts » etc.) → il suffit de cliquer sur l'une des images de la mosaïque pour afficher la notice liée<br><br>-en passant par la recherche avancée qui propose de faire une recherche par type de collection | Oui :<br>-Recherche sur plusieurs critères possible (type de collection, titre, datation de l'œuvre, tags etc.)  | Oui :<br>-Liste de choix pour certains critères (ex : type de collection)<br><br>-Liste de mots-clés liés aux résultats affichés   |
| <a href="#"><u>The Metropolitan Museum of Arts</u></a> | Oui :<br>-recherche sur les pages du site  | Oui :<br>-Possibilité de faire une recherche sur un type spécifique de collection (« Asian Art », « Egyptian Art » etc.)  | Oui :<br>-Recherche multicritères (Artiste, Titre, Pays/Culture d'origine, Matériaux utilisés, Date de création etc.) avec combinaison par opérateurs booléens (ET, OU)<br><br>-Sélection de la localisation des résultats | Oui :<br>-Affichage du nombre d'articles trouvés par rubrique suite à une recherche simple<br><br>- Affichage du nombre d'articles trouvés par thématique d'œuvres suite à une recherche avancée |

|  |  |           |  |
|--|--|-----------|--|
|  |  | escomptés |  |
|--|--|-----------|--|

### 1.2.1.3 Visualisation des résultats

| Nom du portail   | Liste paginée | Tri des résultats   | Historique des recherches |
|--|---------------|---|---------------------------|
| <a href="#">Portail de la Musique Contemporaine</a>                            | Oui           | Non   | Non                       |
| <a href="#">Portail du Museum National d'Histoire naturelle</a>                | Oui           | Oui   | Oui                       |
| <a href="#">Portail des documents techniques de l'eau</a>                      | Oui           | Oui : tri automatique par pertinence  | Non                       |
| <a href="#">Portail documentaire de la Cité de l'Histoire de l'Immigration</a> | Oui           | Oui mais le tri est à paramétrer au moment de la requête :<br>-pertinence<br>-type<br>-année<br>-auteur<br>-titre<br>-date d'édition croissante<br>-date d'édition décroissante | Oui                       |
| <a href="#">Portail du Ministère de l'Ecologie et du développement durable</a> | Oui           | Oui :<br>-par base<br>-par titre<br>-par auteur<br>-par date  | Oui                       |
| <a href="#">Portail des arts de la marionnette</a>                             | Oui           | Oui   | Oui                       |
| <a href="#">Catalogue de la Médiathèque de la Cité de la Musique</a>           | Oui           | Oui   | Non                       |
| <a href="#">Musée Guimet</a>   | Non           | Non   | Non                       |
| <a href="#">Musée du Louvre</a>  | Oui           | Oui   | Non                       |
| <a href="#">Musée Olympique</a>  | Oui           | Oui par critère   | Non                       |
| <a href="#">ECPAD</a>  | Non           | Oui par rubriques concernées :<br>-Brèves<br>-Photos<br>-Vidéos<br>-Rendez-vous<br>-etc.  | Non                       |
| <a href="#">Musée Mc Cord</a>  | Oui           | Oui :<br>-par pertinence<br>-par présence de description<br>-par artiste<br>- par objet<br>-par date  | Non                       |
| <a href="#">Musée Nicéphore Niépce</a>   | Non           | Oui :<br>-possibilité de trier les résultats selon divers critères (chronologique, pertinence etc.)   | Non                       |
| <a href="#">Brooklyn Museum</a>  | Non           | Non   | Non                       |

|   |     |     |     |
|---|-----|-----|-----|
| <a href="#">The Metropolitan Museum of Arts</a> | Oui | Non | Non |
|---|-----|-----|-----|

#### 1.2.1.4 Exploitation des résultats/accès aux documents

| Nom du portail   | Impression de la notice | Envoi par mail   | Export            | Accès au document primaire  | Syndication/Alertes pour les nouvelles notices | Panier des notices |
|--|-------------------------|--|-------------------|---|--|--------------------|
| <a href="#">Portail de la Musique Contemporaine</a>                            | Non                     | Non mais partage possible sur les réseaux sociaux, sites de « social bookmarking » | Non               | Oui possibilité de consulter un extrait ou l'intégralité du document original (texte, audio, vidéo) | Non  | Non                |
| <a href="#">Portail du Museum National d'Histoire naturelle</a>                | Non                     | Oui  | Non               | Non   | Non  | Oui                |
| <a href="#">Portail des documents techniques de l'eau</a>                      | Non                     | Non  | Oui au format PDF | Oui   | Non  | Non                |
| <a href="#">Portail documentaire de la Cité de l'Histoire de l'Immigration</a> | Oui                     | Non  | Non               | Non   | Non  | Oui                |
| <a href="#">Portail du Ministère de l'Écologie et du développement durable</a> | Oui                     | Non  | Non               | Oui pour certains documents   | Oui  | Oui                |
| <a href="#">Portail des arts de la marionnette</a>                             | Non                     | Non  | Non               | Non, mais définition précise des droits sur le document original                                    | Non  | Non                |
| <a href="#">Catalogue</a>  | Oui                     | Oui,   | Non               | Oui :   | Non  | Non                |



|  |     |  |   |  |   |  |
|--|-----|--|---|--|---|--|
| <a href="#">de la Médiathèque de la Cité de la Musique</a> |     | partage sur réseaux sociaux possible                               |   | -Accès à l'ensemble des documents liés et ce quel que soit leur support (audio, vidéo, texte)  |   |  |
| <a href="#">Musée Guimet</a>                               | Non | Non  | Non   | Non  | Non   | Non  |
| <a href="#">Musée du Louvre</a>                            | Oui | Oui  | Non   | Non  | Non   | Non  |
| <a href="#">Musée Olympique</a>                            | Non | Oui, partage sur les réseaux sociaux possible                      | Non   | Non  | Non   | Non  |
| <a href="#">ECPAD</a>                                      | Non | Oui partage possible sur les réseaux sociaux                       | Non   | Non  | Oui, possibilité de s'abonner aux derniers résultats pour une requête | Non  |
| <a href="#">Musée McCord</a>                               | Oui | Oui, de plus possibilité d'envoyer l'image au format carte postale | Oui possibilité de télécharger l'image                                    | Non mais :<br>- possibilité d'acheter une reproduction de l'image originale<br>-Accès à des documents liés (jeux, dossiers thématiques etc.) | Non   | Oui, possibilité en outre de constituer une sélection d'images |
| <a href="#">Musée Nicéphore Niépce</a>                     | Non | Non  | Non   | Non  | Non   | Non  |
| <a href="#">Brooklyn Museum</a>                            | Oui | Oui, partage sur réseaux sociaux possible                          | Oui, possibilité de télécharger l'image sous plusieurs tailles ; lien URL | Non  | Non   | Oui avec le « Set »  |

|   |     |   |                       |   |   |                          |
|---|-----|---|-----------------------|---|---|--------------------------|
|   |     |   | permanente de l'image |   |   |                          |
| <a href="#">The Metropolitan Museum of Arts</a> | Oui | Non mais partage sur réseaux sociaux possible | Non                   | Oui, le numéro d'accès d'une œuvre est donné dans la notice de celle-ci | Non, syndication aux articles du site et à l'actualité du musée | Oui : « My Met Gallery » |

### 1.2.1.5 Gestion du profil des utilisateurs

| Nom du portail   | Codes d'accès   | Profils différenciés   | Customisation  |
|--|---|--|--|
| <a href="#">Portail de la Musique Contemporaine</a>                            | Non : accès indifférencié   | Non  | Non  |
| <a href="#">Portail du Museum National d'Histoire naturelle</a>                | Oui mais l'authentification sur le site ne confère que des droits concernant le prêt de documents | Oui mais peu de fonctionnalités différent entre les membres inscrits et les visiteurs du portail   | Non  |
| <a href="#">Portail documentaire de l'eau</a>                                  | Non   | Non  | Non  |
| <a href="#">Portail documentaire de la Cité de l'Histoire de l'Immigration</a> | Oui, possibilité de s'identifier avec un identifiant et un mot de passe                           | Oui : -L'accès identifié permet de créer un profil de recherche  | Non  |
| <a href="#">Portail du Ministère de l'Ecologie et du développement durable</a> | Non   | Non  | Non  |
| <a href="#">Portail des arts de la marionnette</a>                             | Oui : nécessité d'entrer un login et un mot de passe pour ouvrir son compte                       | Oui, mais à ce stade de développement il n'est pas possible de distinguer les fonctionnalités propres aux utilisateurs identifiés et aux visiteurs | Oui, la présentation des résultats est paramétrable par l'utilisateur                |
| <a href="#">Catalogue de la Médiathèque de la Cité de la Musique</a>           | Non   | Non  | Non  |
| <a href="#">Musée Guimet</a>   | Non   | Non  | Site consultable en plusieurs langues : français, anglais, japonais, chinois, coréen |
| <a href="#">Musée du Louvre</a>  | Oui : l'accès à son espace personnel est possible en  | Oui, mais les fonctionnalités propres aux  | -Plusieurs langues disponibles (français, anglais, chinois,                          |

|  |  |   |   |
|--|--|---|---|
|  | renseignant son adresse mail ainsi que son mot de passe  | utilisateurs identifiés n'ont pas pu être testés  | japonais)   |
| <a href="#"><u>Musée Olympique</u></a>                 | Non  | Non   | -Affichage en langue anglaise possible<br><br>-Possibilité de paramétrer l'affichage des résultats de recherche dans la rubrique « lecteur multimédia »           |
| <a href="#"><u>ECPAD</u></a>                           | Non  | Non   | Non   |
| <a href="#"><u>Musée Mc Cord</u></a>                   | Oui : nécessité de renseigner son adresse mail et son mot de passe pour accéder à son espace personnel (« mon McCord »)        | Oui : « Mon McCord » permet de regrouper ses images préférées, de commenter, « tagger », annoter les images   | Possibilité de zoomer l'image, de l'annoter (pour les utilisateurs enregistrés)   |
| <a href="#"><u>Musée Nicéphore Niépce</u></a>          | Non  | Non   | -Affichage en langue anglaise possible  |
| <a href="#"><u>Brooklyn Museum</u></a>                 | Oui : il est nécessaire de renseigner son e-mail ainsi qu'un mot de passe afin de s'identifier en tant que membre du « Posse » | Oui : faire partie du « Posse » permet à l'utilisateur de se constituer un panier de notices (« set ») et de soumettre des mots-clés (« tags ») aux notices   | Non   |
| <a href="#"><u>The Metropolitan Museum of Arts</u></a> | Oui : il est nécessaire de renseigner un identifiant et un mot de passe afin d'avoir accès aux fonctionnalités « My Met »      | Oui : les utilisateurs bénéficiant d'un compte « My Met » ont accès à des fonctionnalités telles qu'un panier des notices (« My Met Gallery »), un calendrier des événements customisé afin de n'être informés que de ce qui les intéresse (« My Met Calendar ») etc. | Affichage en plusieurs langues possible dans la rubrique « Plan your visit » (français, chinois, japonais, coréen, italien, espagnol, portugais, russe, allemand) |

## 1.2.2 Observations

Cet audit nous a permis de faire un point sur les fonctionnalités propres aux portails permettant à l'utilisateur d'effectuer une recherche fédérée sur des ensembles de ressources différents. Nous nous intéressons ici aux fonctionnalités que nous jugeons intéressantes et originales.

### 1.2.2.1 Pour la recherche de documents

- Possibilité de faire une recherche par type de documents, par type de fonds (Portail de la Musique Contemporaine).

The screenshot shows the search interface of the Portail de la Musique Contemporaine. At the top, there is a navigation menu with 'accueil', 'rechercher', 'parcourir', 'à propos', and 'agenda'. Below this, a search bar is present with a 'RECHERCHER' button. The search bar is surrounded by various filters, including 'TYPE' and 'PROVENANCE', which are highlighted with red and blue circles. A list of search tips is visible on the right side of the search area. The footer includes logos for Sacem, Cdmc, and other partners.

- Possibilité d'opérer une recherche avancée : association de plusieurs critères de recherche par des opérateurs booléens ET, OU, SAUF (voire OU EXCLUSIF).

- Possibilité de faire une recherche thématique (Portail des Arts de la Marionnette, Portail des collections des Musées de France, Muséum National d'Histoire Naturelle, Portail documentaire de la Cité de l'Immigration et Portail du Ministère de l'Ecologie et du Développement durable).

The screenshot displays the PAM website interface. At the top, the logo 'PAM Portail des Arts de la Marionnette' is visible, along with navigation links for 'français', 'Visiteur', 'M'identifier', and 'Créer votre compte'. Below the header, there are filters for 'Tout (14727)', 'Images (8081)', 'Audiovisuel (474)', 'Objets (3534)', 'Textes (1503)', 'Identités (467)', 'Evénements (612)', and 'Oai (56)'. The main content area shows search results for 'La poétique de l'espace', with a grid of 15 items. The first row contains five items, each with a thumbnail image and the text 'La poétique de l'espace, Notice sans visuel'. The second row contains five items, with the first two having thumbnails and the last three having the text 'La poétique de l'espace, Notice sans visuel'. The third row contains five items, all with the text 'PAM Notice sans visuel'. A left sidebar menu is open, showing 'Exploration thématique' with sub-categories like 'Histoire et théorie', 'Types de marionnettes', and 'Marionnettes dans le monde'. The footer includes links for 'Aide et accessibilité', 'Lettre d'information', 'Mentions légales', 'Crédits', and 'Nous écrire'.

<http://marionnettes.arnadillo.fr/app/photopro.sk/marionnettes/doclist?>

- Index disponible pour certains critères (Portail documentaire de la Cité de l'Immigration et Portail du Ministère de l'Ecologie et du Développement durable).

**Cit** nationale de l'histoire de l'immigration

**PORTAIL DOCUMENTAIRE**

Accès public  
Le 24 Juin 2011

Accueil | Aide | S'identifier

Recherche rapide  **OK** Recherche avancée

**Index**

**Auteur / Interprète**

Recherche commençant par

AL-HUSRI ABOU KHALDOUN  
CONFEDERATION GENERALE DU TRAVAIL (CGT)  
DUPONT  
L'ENTRE-PONT  
LEVEQUE JEAN-JACQUES  
RESEAU D'INFORMATION SUR LES MIGRATIONS INTERNATIONALES (PARIS)  
SAVIDAN PATRICK  
STRASBOURG. CONSEIL CONSULTATIF DES ETRANGERS. COMMISSION DE SCOLARISATION  
TARRAGO LLIBERT  
"CASSANDRE"-HORSCHAMP (PARIS). POLE DE RESSOURCES ART ET SOCIETE  
108  
391 ASSOCIATION ORO  
6E CONTINENT  
9E COLLECTIF DE SANS-PAPIERS

**OPTIONS**

Linguistique [ ? ]

Expressions

Pertinence [ ? ]

Opérateurs

Tri des réponses par

**Valeurs par défaut**

Formes flechées  Synonymes  Orthographe

Mots dans l'ordre 10 caractères maxi. entre les mots

2. Nombre de termes correspondant à la question

ET ENTRE champs OU DANS les champs

Pertinence 10 Résultats par page

Texte libre

Titre

Auteur / Interprète

Année entre  et

Année (pour collections du musée) entre  et

Classement thématique

Classement thématique archives Sayad

Sujets

Cole

Catégorie

Support

Éditeur

Localisation

Matériaux / techniques (pour collections du musée)

**Champs supplémentaires**

[Limiter la recherche +](#)

Terminer

- Possibilité de faire une recherche par langue du document (Portail du Muséum National d'Histoire Naturelle).

MUSÉUM NATIONAL D'HISTOIRE NATURELLE  
Catalogue des bibliothèques

Recherche simple | Recherche multicritères | Recherche alphabétique | Recherche numérique | Historique | Infos

Vous êtes en train de consulter : Catalogue commun des bibliothèques du Muséum

recherche simple

Choisir un index de recherche et taper un ou plusieurs mots

Recherche : Mots de l'auteur et/ou du titre

Exemple : bonnier flore  
Résultat : retrouve entre autre «La grande flore de Gaston Bonnier» à partir d'un ou plusieurs mots de l'auteur et/ou du titre

Limitations

|                  |          |  |
|------------------|----------|--|
| Type de document | Langue   | Bibliothèques                            |
| tous documents   | toutes   | Toutes les bibliothèques                 |
| livres           | français | Bibliothèque centrale du Muséum          |
| périodiques      | anglais  | Bibliothèque du Nouveau Musée de l'homme |
| articles         | allemand | Anatomie comparée                        |
| thèses           | espagnol | Anthropologie maritime                   |

Limitation par date

Date publication = <input type="text"/> ET <input type="text"/>

Tri

Aucun

Horizon Portail d'Information 3.048 2001-2003 Dynix Tous droits réservés.

- Affinage des résultats par sélection de critères (Portail des documents techniques de l'eau, Portail des Arts de la Marionnette).

4. MAITRISE DES POLLUTIONS DIFFUSES

Auteurs : GITTON C / AELB / AGENCE DE L'EAU LOIRE BRETAGNE  
Date de publication : 1998-10-01  
Accès aux documents : [http://esidoc\[...\]sult/DOC00014676](http://esidoc[...]sult/DOC00014676)  
Résumé : Etude dont le but est d'apprendre à : Evaluer la dimension d'un plan [...]Jelà de la maîtrise des pollutions ponctuelles en agissant principalement sur l'azote et le phosphore. [suite]  
[Aperçu HTML] - [Documents similaires] - [Afficher la notice complète]

5. POLLUTION DE L'ETANG DE CHABAUD LATOUR

Auteurs : DANCETTE B / JOURNET JM / AGENCE DE L'EAU ARTOIS PICARDE  
Date de publication : 1977-09-01  
Diffuseur des documents : Agence de l'Eau Artois Picardie  
Résumé : LES RESULTATS DES MESURES REALISEES DEPUIS 1976 MONTEMENT [...]QUE LA VIE P[...]RATON CLASSIQUE, ALORS QUE LA POLLUTION AZOTEE [...]DEMANDE UNE INTERVENTION AUPRES DES POLLUEURS BEGES. [suite]  
[Aperçu HTML] - [Documents similaires] - [Afficher la notice complète]

6. Pollution du piézomètre de St-Maclou (Eure).

Auteurs : VILLEY M. / MINISTERE DE L'INDUSTRIE DE LA POSTE ET DE TELECOMMUNICATIONS  
Date de publication : 1997-01-01  
Diffuseur des documents : BRGM  
Résumé : Le suivi de l'évolution des nappes d'eau souterraine est réalisé [...]pe définir ainsi s'il s'agit d'une pollution ponctuelle par le haut, ou d'une pollution de la nappe. [suite]  
[Aperçu HTML] - [Documents similaires] - [Afficher la notice complète]

7. LA POLLUTION PAR LES PRODUITS PHYTOSANITAIRES. INTERVENTION A L'IFFMA

Auteurs : BABUT Marc/ERM  
Date de publication : 1998  
Diffuseur des documents : Agence de l'Eau Rhin Meuse  
[Documents similaires] - [Afficher la notice complète]

8. Pollution des sols et des sédiments: quels impacts ?

Auteurs : LAPERCHE DOROTHÉE  
Date de publication : 2010-01-01  
Diffuseur des documents : INIST-CNRS  
[Documents similaires] - [Afficher la notice complète]

9. LA POLLUTION DES EAUX PLUVIALES

Auteurs : PORTALES J / AGENCE DE L'EAU SEINE NORMANDIE / ESIGEC CHAMBERY / AGENCE DE L'EAU SEINE NORMANDIE  
Date de publication : 1996-05-01  
Diffuseur des documents : Agence de l'Eau Seine Normandie

Principaux organismes diffuseurs de document

- BRGM (4897)
- Agence de l'Eau Rhône Méditerranée et Corse (1976)
- Agence de l'Eau Rhin Meuse (340)
- Agence de l'Eau Artois Picardie (378)
- Agence de l'Eau Seine Normandie (287)
- Agence de l'Eau Adour Garonne (143)
- Offices de l'Eau Réunion (70)
- ONEMA (47)
- INIST-CNRS (32)

Principaux contributeurs

- Bureau de Recherche Géologiques et Minières (4897)
- Agence de l'Eau Rhône Méditerranée et Corse (1976)
- Institut Français de Recherche et d'Exploitation de la MER (842)
- Agence de l'Eau Loire Bretagne (815)
- CEMAGREF (710)
- Agence de l'Eau Rhin-Meuse (540)
- Agence de l'Eau Artois Picardie (378)
- Offices de l'Eau Seine Normandie (302)

- Historique des requêtes et combinaison de requêtes (Portail du Ministère de l'Écologie et du Développement durable).

Portail documentaire du ministère de l'Écologie, du Développement Durable des Transports et du Logement

Accueil Recherche avancée Ressources Aide Voir aussi Préférences F.A.Q.

Portail Internet Répertoire Internet Annuaire Inter administration

**Recherche** Résultats de la recherche

Nouveautés Recherche simple Recherche avancée Historique Panier Aide

**Recherches** Aucune recherche enregistrée.

Page 1 / 285 Document 1 à 20 / 5699

Tout sélectionner Tout désélectionner Ajouter la sélection au panier

**Combiner des recherches**

**Filter les résultats :**

**Par thèmes :**

- Administration
- Construction
- Droit
- Economie
- Environnement
- Equipements - Infrastructures
- Habitat - Logement
- Sciences et techniques
- Sciences Humaines
- Tourisme
- Transports
- Urbanisme - Aménagement
- Tous les thèmes

**Par types de documents :**

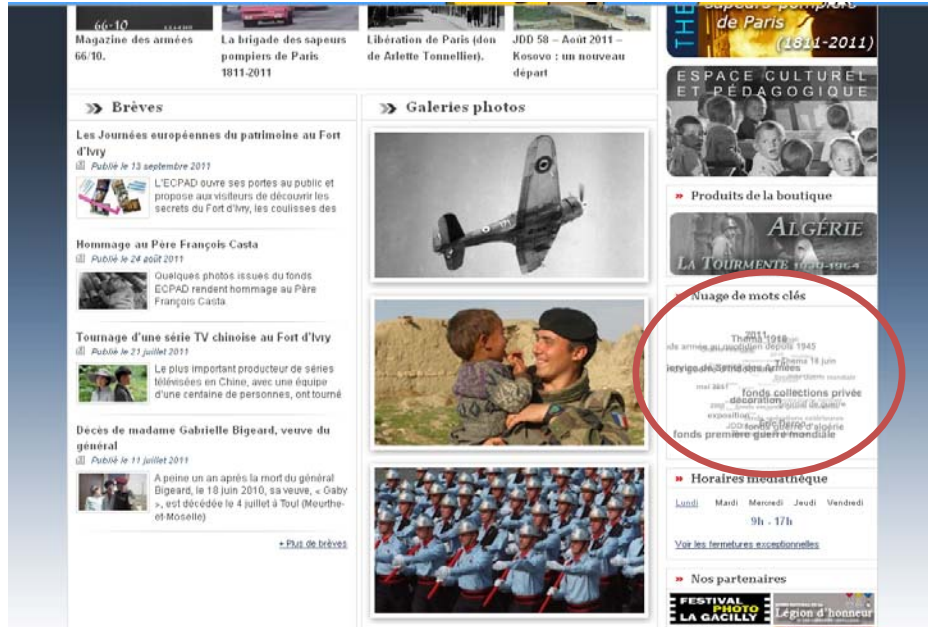
- Actes de congrès
- Cartes et atlas
- Document multimédia
- Documents de référence
- Etudes et rapports
- Livres
- Normes et documents techniques
- Périodiques
- Site Web
- Textes officiels
- Tous les types

|    | Base           | Titre  | Auteur  | Date |
|----|----------------|--|---|------|
| 1  | Base du CRDALN | Etat (L') de santé des eaux côtières et marines méditerranéennes. Conférence technique du 8 juin 2011 à Marseille.-              | BOISSERY, Pierre<br>GUESPEREAU, Martin<br>CHARDIN, Nicolas      | 2011 |
| 2  | Base du CRDALN | Seine (La), hier, aujourd'hui et demain.-  |   | 2011 |
| 3  | Base du CRDALN | Avec le GIP Seine-Aval, la Seine a son "ange gardien".-  |   | 2011 |
| 4  | Base du CRDALN | Odeurs, bruits, pollutions : les collectivités face aux nuisances. Dossier.-   | BAUER, Delphine   | 2011 |
| 5  | Urbamet        | Odeurs, bruits, pollutions : les collectivités face aux nuisances. Dossier.-   |   |      |
| 6  | Base du CRDALN | Autosurveillance des réseaux : les collectivités doivent s'équiper av  |   |      |
| 7  | Base du CRDALN | Marées (Les) vertes en Bretagne : pour un diagnostic partagé, garant d'une action efficace.-                                     | GUYOMARC'H, Jean-Paul<br>LE FOLL, François<br>GUYARD, Christian | 2011 |
| 8  | Base du CRDALN | Analyse : la mesure de conductivité. Conductimétrie : un paramètre de suivi de qualité et d'alerte.-                             | GUYARD, Christian   | 2011 |
| 9  | Base du CRDALN | Nouvelle (Une) voie d'élimination du sélénium dans les eaux potables : essais réalisés pour le Siaep de la vallée de l'Aubetin.- | COUSIN, Alain<br>RENSON, M. A.<br>ROUSSELIN, A.                 | 2011 |
| 10 | Base du CRDALN | Intégrer la dépollution dans le processus d'aménagement : réflexions à partir de cas étrangers.-                                 | LECOMPTÉ, Yannick<br>VERHAGE, Roelof                            | 2011 |
| 11 | Urbamet        | Intégrer la dépollution dans le processus d'aménagement : réflexions à partir de cas étrangers.-                                 | LECOMPTÉ, Yannick<br>VERHAGE, Roelof                            | 2011 |
| 12 | Base du CRDALN | Environnement littoral et marin. Edition 2011.-  | COLAS, Sébastien<br>BOITARD, Corinne (coord.)                   | 2011 |
| 13 | Urbamet        | Environnement littoral et marin. Edition 2011.-  | COLAS, Sébastien<br>BOITARD, Corinne (coord.)                   | 2011 |
| 14 | Base du CRDALN | Substances dangereuses et eaux pluviales.-   | CAULLIEZ, Dorotheé  | 2011 |

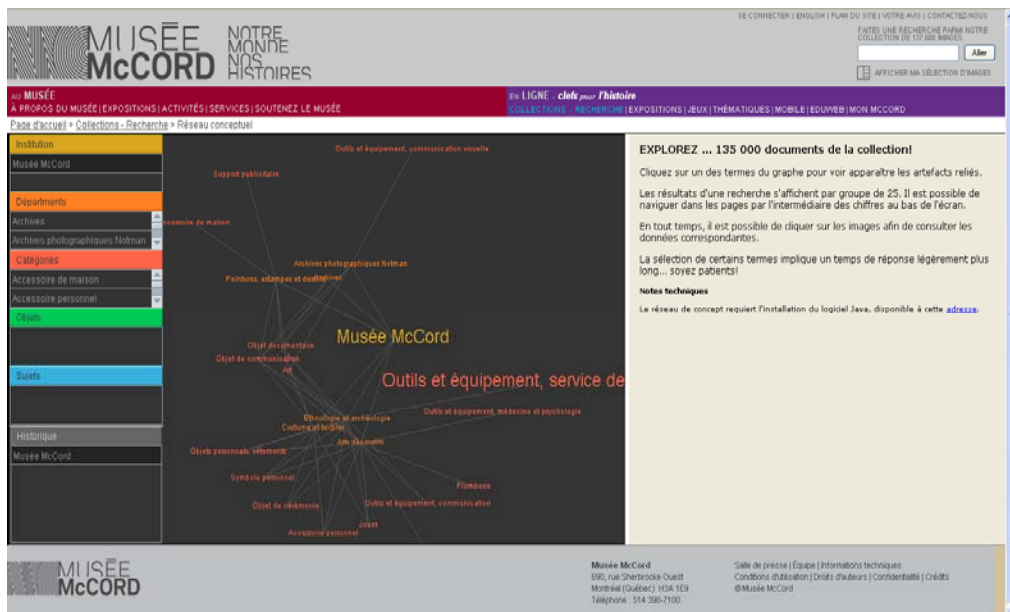
http://cataloguecd�.documentation.developpement-durable.gov.fr/cataloguecd�/document.jsp?id=Cdu-0202591



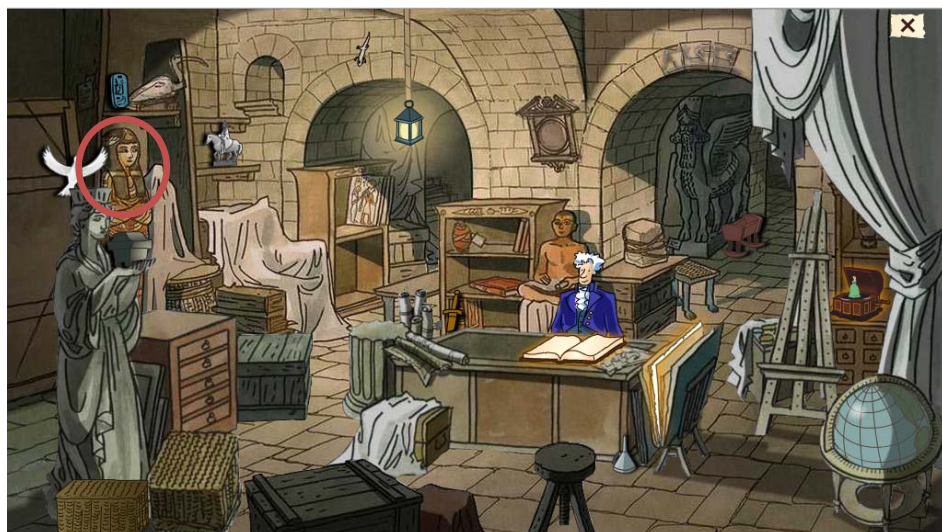
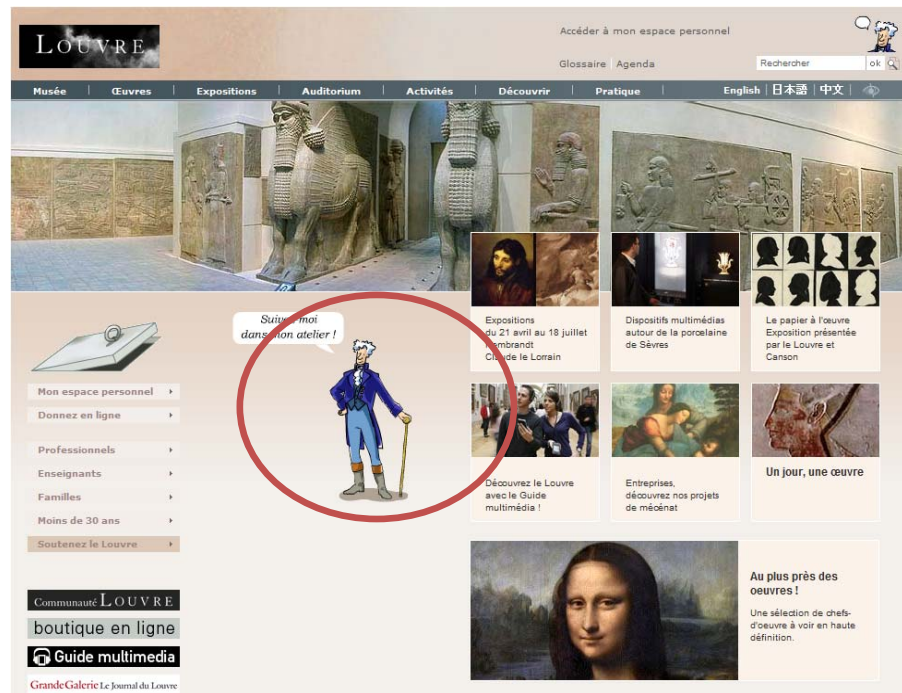
- Possibilité de faire une recherche grâce à un nuage de mots-clés (*tag cloud*) sur le site de l'ECPAD.



- Possibilité de naviguer au sein d'une « mind map » pour effectuer une recherche dans la collection McCord : le « réseau conceptuel ».



- Le site du Musée du Louvre propose une interface de découverte pour les enfants : « La visite de l'atelier », en cliquant sur le petit personnage animé en page d'accueil pour y accéder, permet d'accéder à un atelier dessiné à partir duquel plusieurs fiches d'œuvres présélectionnées (lien vers la rubrique « Œuvres choisies ») sont accessibles en cliquant sur des objets de l'atelier.



### 1.2.2.2 Pour l'exploitation des résultats de recherche

- Possibilité de consulter un extrait ou l'intégralité du document original que ce soit du texte, de l'audio ou de la vidéo (Portail de la Musique Contemporaine).

The screenshot displays the CDMC (Centre de documentation de la musique contemporaine) website interface. The main content area shows the search results for 'Bad tape nostalgia' by Favory, Jean-Baptiste. The record details include: Type: audio; Genre(s): oeuvre(s) musicale(s); Forme(s): disque compact, dvd, cd-rom. A red circle highlights the 'Ecoutez un extrait' button under the 'ACCÈS EN LIGNE' section. Below the main record, a table lists other records:

| numéro | titre   | année | accès    |
|--------|---|-------|----------|
| 6      | Lénine, Staline et la musique 2 - Les Dissonances | 2011  | [icônes] |
| 7      | FAVORY, Jean-Baptiste - Bad tape nostalgia        | 2011  | [icônes] |

The footer contains logos for various partners including the French Republic, SACEM, CDMC, the Centre de documentation de la musique contemporaine, the Centre de la musique contemporaine, the Centre de la musique contemporaine, ensemble intercontemporain, ircam Centre Pompidou, and MUSCÉE MONTAIGNE. The URL at the bottom is [www.musiquecontemporaine.fr/record/oi:cdmc.asso.fr:aloes0109119?language=fr](http://www.musiquecontemporaine.fr/record/oi:cdmc.asso.fr:aloes0109119?language=fr).

- Accès à l'ensemble des documents liés au document consulté qu'ils soient sur un support texte, audio ou vidéo : les relations entre les différents documents ont été préétablies selon un modèle FRBR (Catalogue de la Médiathèque de la Cité de la Musique et le Portail des Arts de la Marionnette).

**MÉDIATHÈQUE** CONTACT À PROPOS PLAN VOUS CHERCHEZ OK

**OEUVRES MUSICALES** RECHERCHE SIMPLE RECHERCHE AVANCÉE

Index sur autorités: bach et Type de document: Oeuvres musicales  
349 résultat(s) Page 1/30

**ACCUEIL**  
**CATALOGUE**  
DOCUMENTS  
OEUVRES MUSICALES  
RECHERCHE SIMPLE  
RECHERCHE AVANCÉE  
BIBLIOGRAPHIES  
**CONCERTS**  
**CONFÉRENCES**  
**MUSÉE DE LA MUSIQUE**  
INSTRUMENTS DE MUSIQUE  
DOSSIERS  
PÉDAGOGIQUES  
GUIDE PRATIQUE DE LA MUSIQUE  
RÉPERTOIRE DE SITES

1) Bach, Johann Sebastian / Weinen, Klagen, Sorgen, Zagen, Fa mineur, BWV 12  
2) Bach, Johann Sebastian / Variations Goldberg, BWV 988  
3) Bach, Johann Sebastian / Wie schön leuchtet der Morgenstern, fa majeur, BWV 1  
4) Bach, Johann Sebastian / Quatorze canons, BWV 1087  
5) Bach, Johann Sebastian / L'offrande musicale, BWV 1079  
6) Bach, Wilhelm Friedemann / Sonate, La mineur, F n° 8  
7) Bach, Wilhelm Friedemann / Douze polonaises, No 1 en do majeur. No 2 en do mineur. No 3 en ré majeur. No 4 en ré mineur. No 5 en mi bémol majeur. No 6 en mi bémol mineur. No 7 en mi majeur. No 8 en mi mineur. No 9 en fa majeur. No 10 en fa mineur. No 11 en sol majeur. No 12 en sol mineur, F 12  
8) Bach, Carl Philipp Emanuel / Six sonates, no 1 en la mineur, no 2 en la bémol majeur, no 3 en si bémol majeur, no 4 en mi mineur, no 5 en mi bémol majeur, no 6 en si mineur, H 30-H 26  
9) Bach, Johann Christian / Six sonates, Op. 5  
10) Bach, Johann Sebastian / L'art de la fugue, BWV 1080  
11) Bach, Johann Sebastian / Prélude et fugue en mi bémol majeur, Mi bémol majeur, BWV 552  
12) Schoenberg, Arnold / Transcription du "Prélude et fugue pour orgue" BWV 552 de Jean-Sebastian Bach, Mi bémol majeur

Page 1/30

**Document(s) numérique(s)**  
**Concert(s) audio**  
Cantate "Weinen, Klagen, Sorgen, Zagen" BWV 12 / Johann Sebastian Bach : Cantus Colln : Konrad Junghanel, direction : Sabine Goetz, soprano : Elisabeth Popien, alto : Hans Jörg Mammel, ténor : Wolf Matthias Friedrich, basse : Bas Ramselaar, basse, 21 min 28 sec  
**CD Audio(s)**  
Cantatas BWV 12. 74. 177 / Johann Sebastian Bach, comp. - Holland Boys Choir - Netherlands Bach Collegium : Pieter-Jan Leusink, dir. - Ruth Holton, soprano - Svyta Buvvalda, alto - Nico van der Meel (BWV 74) - Knut Schooch (BWV 12) - Marcel Beekman (BWV 177), ténor - Bas Ramselaar, basse, 1 cd (1h 11 min)  
Weinen, Klagen, Sorgen, Zagen, BWV 12 - J.S. Bach - Helen Watts, alto - Adalbert Kraus, ténor - Wolfgang Schöne, basse - Bach-Ensemble, Helmuth Rilling, dir., 26 min 24 sec  
**Partition(s)**  
Weinen, Klagen, Sorgen, Zagen BWV 12 - Kantate zum Sonntag, Jubilate / Johann Sebastian Bach  
Cantate, BWV 12 - Weinen, Klagen, Sorgen, Zagen

- Partage des notices sur des réseaux sociaux et de « Social Bookmarking » (Portail de la Musique Contemporaine).

**MUSIQUE CONTEMPORAINE** accueil rechercher parcourir à propos agenda

notice NOTICE 7 sur 16097 Ajouter aux favoris

Cette ressource est disponible chez l'organisme suivant : Cdmc - Centre de documentation de la musique contemporaine

Titre Bad tape nostalgia  
Nom(s) Favory, Jean-Baptiste (compositeur)  
Publication 2011  
Langue(s) français

ENVOYER LA NOTICE PARTAGER

IDENTIFIANT OAI oai:cdmc.asso.fr:aloes:0109119  
DATE DE LA NOTICE 2011-05-05 16:11:30  
IDENTIFIANT PORTAL 56293  
CONTACT Nous écrire

24 juin / L'Autre rive - Zadi Moutaka - Ensemble Musicatreizze - La Vieille Charité (Marseille)  
24 juin / Pascal Dusssin - Orchestre Philharmonique de Radio France - Salle Pleyel (Paris)  
24 juin / What's new saison 4 - TM+ - Maison de la Musique (Nanterre)

NOUVELLE RECHERCHE

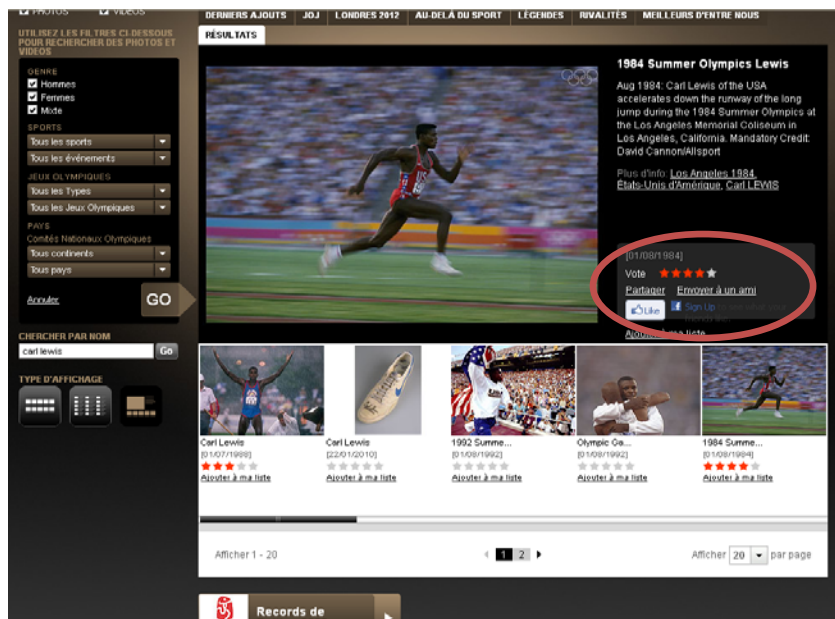
DATE ACCÈS

dien, pour o 2011  
he et techno 2011  
nce IV b, pot 2011  
s, Exercices 2011

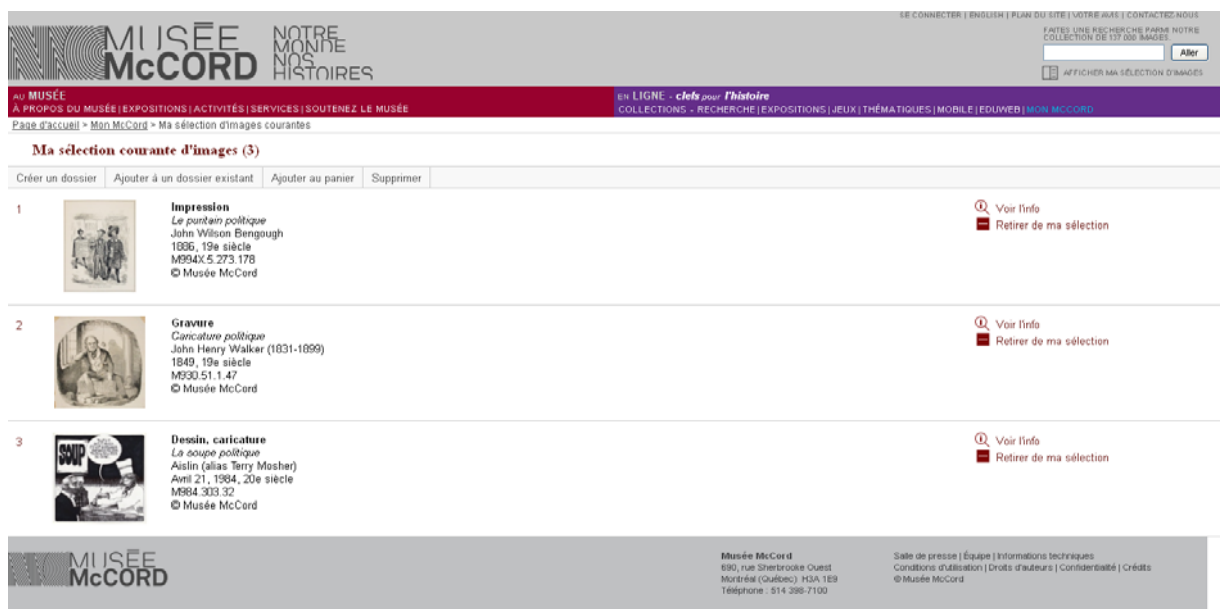
- Syndication sur les mots-clés permettant aux utilisateurs d'être mis au courant des nouvelles ressources quand une notice concernée a été créée ou modifiée (Portail du Ministère de l'Ecologie et du Développement durable).

- Possibilité de « tagger » (adjoindre un mot-clé) une notice pour les membres enregistrés (appartenant au « Posse ») du site du Brooklyn Museum.

- Le Musée Olympique donne la possibilité aux utilisateurs de noter les photos et vidéos.

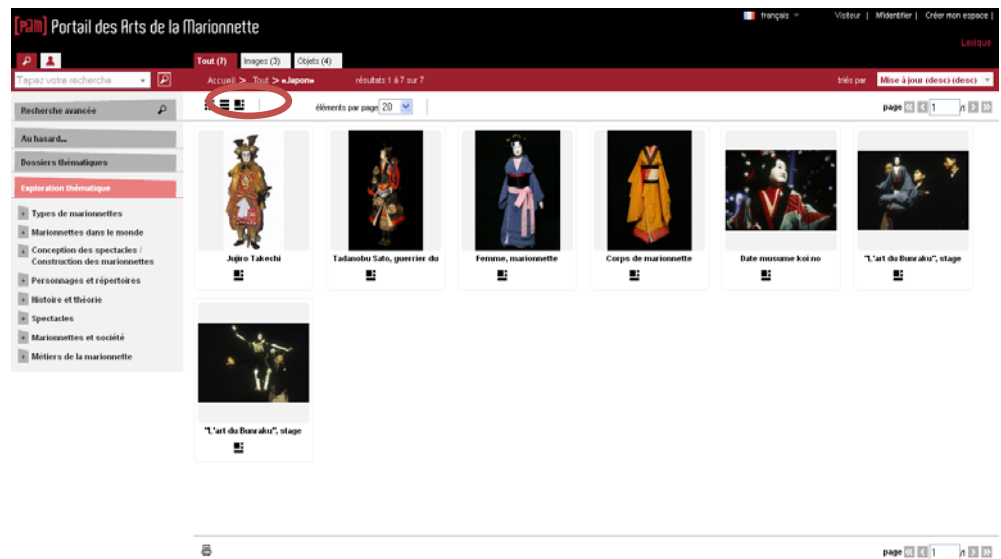


- Le site du Musée Mc Cord permet d'enregistrer ses résultats de recherche et de les gérer (création de dossiers, gestion de ces dossiers, suppression) au sein de « Ma sélection d'images ».

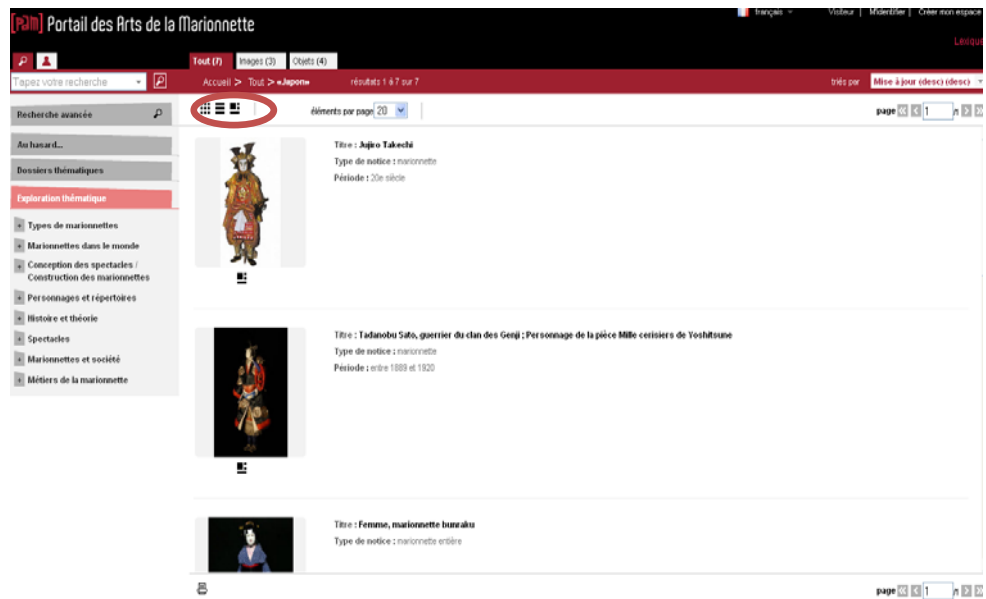


- Le Portail des Arts de la Marionnette permet à l'utilisateur de modifier le mode d'affichage des résultats, ce qui représente une fonctionnalité intéressante pour un portail dont la majeure partie du fonds demeure iconographique.

Présentation avec le premier mode d'affichage :

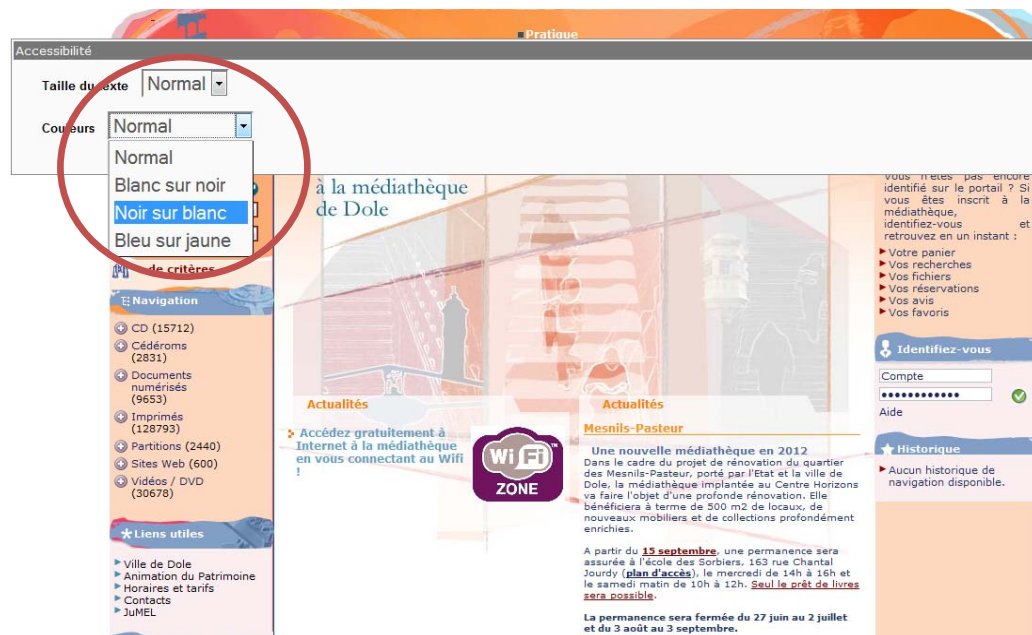


Présentation avec le deuxième mode d'affichage :



### 1.2.2.3 Pour l'accessibilité du site

- Possibilité de changer l’affichage des pages pour les personnes en situation de handicap visuel (Médiathèque de Dole).



### 1.2.3 Les portails d’institutions culturelles : un retard français

Il ressort de cet audit que l’on retrouve des fonctionnalités récurrentes qui définissent l’identité et la raison d’être d’un portail culturel par la valorisation et la communication des ressources rendues possibles par :

- Des modes de recherche classiques : recherche simple (en plein texte sur les pages du site et sur les répertoires de ressources) et recherche avancée (multicritères sur les répertoires de ressources).
- Un affichage et une exploitation des résultats de recherche permettant à l’utilisateur de naviguer rapidement et d’opérer une sélection parmi ces résultats : pagination des résultats, tri, sauvegarde dans un panier, export des notices, historique des recherches, partage sur les réseaux sociaux ou envoi par mail.

L’interface des portails existants est globalement accessible à tous les publics mais celle de leurs modes de recherche conserve en général un aspect très « technicien », surtout en ce



qui concerne le module de recherche avancée : l'utilisateur ne sort pas réellement d'une interface « bibliothéconomie ». Très peu de ces portails proposent une recherche thématique, ce qui est pourtant utile pour encourager les visiteurs néophytes (qui ne savent pas forcément quel type d'information est disponible) à rester sur le site. Très peu de ces portails proposent un principe de « folksonomie », leurs visiteurs ne peuvent donc pas participer à l'indexation des documents, ce qui est dommageable à deux niveaux. D'une part le portail se prive de l'expertise de certains visiteurs (or celle-ci pourrait s'avérer utile pour un portail culturel) et d'autre part cela ne contribue pas à la fidélisation des visiteurs qui à l'heure du Web 2.0 souhaiteraient davantage d'interactivité. Les seules fonctionnalités du Web 2.0 étant assez répandues sur la sitosphère culturelle demeurent la syndication aux actualités des institutions et le partage des notices sur les réseaux sociaux.

Des efforts ont néanmoins été faits par les musées nord américains tels que le Metropolitan Museum of Art (MET), le Musée Mc Cord et le Brooklyn Museum qui ont su proposer :

- une interface graphiquement agréable pour l'utilisateur
- une ergonomie intuitive, l'utilisateur peut accéder très rapidement et facilement à l'information voulue
- des modes de recherche originaux : exploration d'arborescences thématiques, *mind mapping*, rebonds vers des résultats liés
- une participation des utilisateurs à la vie du site et de l'institution : folksonomie, notation, commentaires sur les photos et les vidéos

Ces institutions ont compris que la valorisation de leurs fonds n'était pas suffisante pour mettre en lumière leurs activités et que leur dynamisme reposait aussi sur la communication avec l'utilisateur qui, en plus de se voir proposer une navigation intuitive et chatoyante, se sent devenir acteur de la vie de l'institution en communiquant, notant et indexant (dans une proportion certes restreinte).

Cependant quelques sites français parviennent à sortir des canons de la bibliothéconomie en proposant une expérience originale aux utilisateurs :

- la recherche par « exploration thématique » ou « au hasard » sur le Portail des Arts de la Marionnette
- une interface de découverte des collections pour les enfants proposée sur le site du Musée du Louvre
- la recherche par nuage de mots-clés sur le site de l'ECPAD
- le paramétrage de l'affichage des résultats sur le Portail des Arts de la Marionnette

- le filtrage des résultats au moyen d'un nuage de mots-clés sur le Portail des documents techniques de l'eau
- la syndication à des mots-clés proposée par le Ministère de l'Ecologie

## **1.3 Mettre en place un portail : une réflexion à mener sur les contraintes potentielles**

La mise en place d'un portail nécessite donc la mobilisation de deux types de compétences, celles du bibliothécaire ou du documentaliste, d'une part, pour définir ses priorités de service, et celles de l'informaticien, d'autre part, pour définir le domaine du possible et proposer les solutions techniques adaptées. Il convient donc de nuancer le discours des éditeurs de solutions portail selon qui leur produit peut régler l'ensemble des problèmes et contraintes au sein de l'institution.

### **1.3.1 Le CMS comme l'alpha et l'oméga d'une solution portail : une vision erronée et partielle**

La brique CMS, il est vrai, permet de faciliter la mise en place de la première version du portail et de mettre à jour son contenu assez facilement par une gestion des contenus (publication, suppression, modification, workflow de validation etc.) et de la charte graphique (organisation des « blocs à l'écran, des couleurs, des dimensions sans passer par du code HTML mais grâce aux feuilles de style CSS et au XSL intégrant déjà du code) simplifiée. Le CMS peut être défini comme un outil web permettant à son utilisateur de créer et de mettre en ligne du contenu sans connaissance technique approfondie de code HTML ni des langages de programmation. Le CMS repose sur le paradigme d'une dissociation du contenu et de sa présentation. Cet outil s'articule sur une structure modulaire qui fonctionne sur la base de *plugs-in* venant se greffer à l'existant pour répondre à des besoins précis. Il permet en outre la gestion des différents profils des utilisateurs, est un éditeur wysiwyg- « *what you see is what you get* », comparable à l'interface d'un traitement de texte- et dispose de fonctionnalités d'export des données. Or un CMS ne saurait remplacer un véritable logiciel de SIGB en ce qui concerne la gestion d'un répertoire de ressources attaché à un portail, les solutions CMS ne pouvant assurer qu'une gestion documentaire simple. Par conséquent, si le CMS est indubitablement une solution simple pour mettre en place un portail, il ne doit demeurer qu'une brique devant être interopérable avec les autres briques (SIGB, moteur de recherche, bases de données externes etc.).

Le portail est, dans les faits, un logiciel qui doit faire preuve d'une souplesse extrême (37, TOSCA, MAISONNEUVE, TOUITOU). Il doit en effet prendre en compte les outils existants au sein de l'institution, notamment son SIGB, et doit s'appuyer sur une architecture fonctionnelle adaptée aux caractéristiques de l'ensemble des logiciels installés. Toutefois si le portail doit prendre en compte la complexité du SIGB, il doit permettre la liberté éditoriale des publications. Cette souplesse pose la question de la nature du logiciel à utiliser : un logiciel pouvant être paramétré de manière simple afin d'être adapté à l'architecture fonctionnelle des logiciels déjà installés dans la structure, un logiciel (voire des briques *open source*) qui seront plus adaptés aux besoins mais qui demanderont un paramétrage plus complexe. Quoiqu'il en soit, si le portail est un outil simple à mettre en place et à mettre à jour, il nécessitera sur le long terme les compétences de techniciens informatiques (mise à jour des connecteurs, du SIGB, de la charte graphique etc.). Le portail est donc « un outil à qui l'on demande beaucoup mais qui est également très exigeant. Le service qu'il permet de proposer aux usagers demeurera probablement en rapport avec les moyens humains (mobilisés) » (37, TOSCA, MAISONNEUVE, TOUITOU).

### **1.3.2 L'ergonomie : un élément à ne pas négliger dans la mise en place d'un portail**

L'aspect visuel ne doit en aucun cas être négligé dans la mesure où il donne une première impression à l'utilisateur. Le graphisme est en effet un facteur du dynamisme du portail.

Outre l'aspect visuel (le rendu à l'écran), l'ergonomie repose tout autant sur la navigation entre les différentes pages web. De ce fait il est aussi nécessaire de réfléchir à l'arborescence des pages du portail afin permettre à l'utilisateur de naviguer aisément sur celui-ci. Cela suppose de réfléchir à l'organisation de l'information contenue dans le portail en thématiques et sous-thématiques pertinentes afin que l'utilisateur puisse trouver ce qu'il recherche en navigant sur les différentes pages en identifiant visuellement le chemin à prendre.

De ce fait l'ergonomie demeure un élément indispensable pour optimiser la recherche d'information par l'utilisateur. Nous pouvons définir le terme d' « ergonomie des interfaces » comme « la discipline qui vise à étudier et à améliorer les modes d'interaction entre l'homme et les outils informatiques » (26, BOURGEOUX, FRESNAU, p.2). Cette ergonomie des interfaces s'appuie essentiellement sur l'analyse des processus mentaux afin d'adapter les outils au fonctionnement cognitif des utilisateurs ainsi que les caractéristiques physiques de

l'interaction homme-machine. Nous pouvons déterminer que l'utilisabilité d'un système repose sur :

- Son efficience→la possibilité pour l'utilisateur d'atteindre le but qu'il s'est fixé sans perdre trop de temps
- Sa facilité d'apprentissage→l'aisance et la rapidité avec lesquelles l'utilisateur apprend à se servir de l'outil
- Sa facilité de mémorisation→la capacité pour l'utilisateur de se souvenir des différentes étapes et du fonctionnement général de l'outil
- Sa fiabilité→la capacité du système à prévenir et à gérer les erreurs
- La satisfaction de l'utilisateur

C'est donc pourquoi Laure Bourgeaux et Amélie Fresnau ont mené des tests d'utilisabilité auprès des utilisateurs du portail de la BPI en 2008 dans l'optique de l'ouverture du nouveau site web en décembre 2008. Elles retiennent de ceux-ci plusieurs enseignements :

- La multiplicité des chemins d'accès à l'information et des possibilités de recherche ne fait que dérouter l'utilisateur qui ne parvient pas à en déterminer les subtilités.
- La nécessité d'une exploration thématique au vu de l'appréhension de certains utilisateurs à saisir un terme dans le moteur de recherche
- La terminologie employée sur l'interface doit être pensée pour être comprise par les futurs utilisateurs qui ne comprennent pas forcément le langage des professionnels de l'information-documentation

Elles observent que les attitudes des utilisateurs sont forgées par leur usage des outils de recherche en ligne (moteurs généralistes, librairies en ligne, sites de vente aux enchères etc.), ce qui nécessite donc pour les professionnels de l'information-documentation de « réfléchir à de nouvelles manières de concevoir et d'améliorer les interfaces de recherche et de consultation, associant pleinement les utilisateurs selon un processus dynamique, participatif et itératif » (26, BOURGEOUX, FRESNAU, p.5).

Guillaume Nuttin remarque par ailleurs que les catalogues adaptés aux enfants se multiplient au sein des portails de bibliothèques (mais aussi de Musées : Musée des Arts et Métiers) L'auteur affirme que les espaces consacrés aux jeunes publics sur les portails culturels s'élargissent (Arts et métier, MET, Louvre, Cités des Sciences etc.) mais qu'il est nécessaire de réfléchir à une ergonomie simplifiée (textes plus courts, illustrations significatives, pages aérées, taille de typographie importante etc.) car même si les enfants sont des « digital natives », leurs compétences en recherche demeurent limitées (33, NUTTIN).

### 1.3.3 Des compétences à mobiliser

La mise en place d'un portail documentaire exige du documentaliste la maîtrise de trois éléments :

- la création de pages web
- la construction intellectuelle et ergonomique de l'ensemble
- le repérage d'information, qui suppose la maîtrise de la recherche de l'information, la connaissance des outils et de leur évolution.

Sans être une opération délicate, la mise en oeuvre du CMS nécessite à la fois une adaptation de l'organisation et l'acquisition de compétences informatiques spécifiques. En effet, pour gérer au mieux l'intervention de plusieurs rédacteurs sur un document, le CMS a besoin de s'appuyer sur une procédure formalisée précisant qui peut proposer un article, qui peut y apporter des corrections, qui peut en valider le contenu ou la mise en forme, qui décide de sa mise en ligne, puis de sa mise à jour ou de son retrait... C'est la première difficulté organisationnelle : il faut définir les procédures de publication avant de penser au déploiement de l'outil. Par ailleurs, si le CMS affranchit effectivement le webmestre de la manipulation des balises HTML, il lui demande néanmoins de sérieuses connaissances en informatique comme l'utilisation de XSL, le langage de description de feuilles de style, la maîtrise du langage XML ou de la technique de la syndication de contenu (Atom, RSS, etc.). La réussite du déploiement du CMS va donc mobiliser de multiples compétences : rédactionnelles, organisationnelles et informatiques. C'est un travail d'équipe pour lequel les savoir-faire des bibliothécaires et des documentalistes sont les bienvenus mais pas nécessairement au coeur des besoins.

En outre la fonctionnalité de recherche fédérée est une des premières choses que les utilisateurs s'attendent à voir sur un portail dans la mesure où un moteur de recherche constitue un moyen rapide d'accéder à l'information sur le site et dans les répertoires de ressources. Or si les compétences documentaires semblent tout à fait essentielles pour la qualité du service de recherche fédérée, notamment pour la présentation structurée des sources d'information, le paramétrage des logiques de dédoublement ou de grappage des références, etc., ce sont les compétences informatiques, connaissances des technologies du Web et des spécificités des bases des éditeurs, qui pourront garantir dans la durée le bon fonctionnement du service. Par exemple si nous nous intéressons à la traduction de la requête dans une syntaxe spécifique à chaque base, nous constatons qu'elle s'appuie sur un

connecteur développé par le fournisseur du logiciel ou plus rarement par un tiers. La charge et la complexité de l'exploitation de la recherche fédérée sont directement en rapport avec le nombre de bases auxquelles l'accès est proposé, ce qui demeure le domaine d'intervention de techniciens informatiques et non des documentalistes.

### **1.3.4 Le lancement d'un portail : un projet nécessitant une réflexion en amont**

Le portail dispose de fonctionnalités qui en font un élément essentiel dans la valorisation d'un fonds et dans la communication avec les internautes, mais il ne s'agit pas, comme nous l'avons déjà fait remarquer, d'une solution livrée clé en main. En effet la mise en place d'un portail nécessite une réflexion qui est primordiale pour en garantir le succès :

- Savoir cibler les publics, leurs besoins, leurs compétences.
- Savoir organiser l'information sur le portail, la hiérarchiser autour de thématiques pertinentes et compréhensibles par l'utilisateur.
- Réfléchir à l'accessibilité des ressources sur les bases de données, ce qui revient à se demander comment les rendre interrogeables en ligne (prévoir les contraintes techniques, la concordance des langages d'indexation et des formats des notices etc.).
- S'assurer de la performance du logiciel qui sera utilisé pour l'intégration du portail.
- Estimer les coûts, de l'investissement initial aux frais de maintenance et de formation.
- Elaborer un planning réaliste prenant en compte les différentes phases du projet, les équipes concernées au cours de celles-ci, les modalités de validation des étapes
- Impliquer le personnel tout au long du projet.
- Rédiger un cahier des charges pour expliciter les fonctionnalités attendues et les contraintes à prendre en considération.

Même si les technologies sont en constante évolution, l'approche méthodologique demeure inchangée, « trois recommandations restent éternellement valable : s'enquérir des besoins, repérer les contraintes et orientation de son organisme et analyser les pratiques des utilisateurs » (36, SCOPSI).

## 2 Mettre en place une recherche fédérée : réfléchir l'interrogation des bases de données hétérogènes via une interface unique

---

### 2.1 Définir la recherche fédérée

#### 2.1.1 Les principes d'une recherche fédérée

La recherche fédérée est l'un des piliers du portail documentaire, avec l'outil de gestion de contenus (CMS). La recherche fédérée est présente systématiquement dans tous les portails des bibliothèques et des centres de documentation. La recherche fédérée est le service qui permet, en une seule requête, de consulter plusieurs bases de données hétérogènes que ce soit en terme de format ou de contenus. Un logiciel de recherche fédérée dispose selon Marc Maisonneuve (37, TOSCA, MAISONNEUVE, TOUITOU) de trois fonctions essentielles :

- La présentation structurée des sources d'information interrogeables afin de faciliter la sélection par l'utilisateur des bases à solliciter
- La saisie de la requête de l'utilisateur et la transmission de celle-ci aux bases qu'il a (ou non) sélectionnées dans une syntaxe adaptée aux caractéristiques de chacune
- Affichage des résultats de la recherche, avec ou sans dédoublonnage, avec ou sans regroupement des réponses apparentées

Il apparaît utile de distinguer la recherche fédérée et les métamoteurs de recherche. En effet, un métamoteur permet de lancer une requête, via une interface unique (par exemple les moteurs de recherche Exalead ou encore Allplus), sur plusieurs moteurs de recherche dont les résultats seront « rapatriés » sur l'interface de recherche unique sans distinction. Un moteur de recherche fédérée est un moteur unique qui interroge simultanément plusieurs répertoires de ressources de façon synchrone ou asynchrone. La recherche fédérée peut être aussi définie comme l'interrogation simultanée, à l'aide d'un formulaire, de ressources numériques hétérogènes (20, GIUSTI). Cette hétérogénéité signifie que les ressources interrogées peuvent être :

- De nature différente :
  - Des ressources primaires non structurées → travaux universitaires en texte intégral, articles de périodiques
  - Des ressources secondaires structurées → métadonnées de catalogues de bibliothèque, notices de bases de données

- De format différent : en MARC, EAD etc.

La recherche fédérée constitue la fonctionnalité qui distingue le portail dit « documentaire » des autres portails en ligne. En effet elle permet de chercher simultanément dans plusieurs bases de données documentaires. La requête de l'utilisateur est envoyée à chacune des bases sélectionnées (quand le choix lui est proposé) et les résultats provenant de toutes ces bases de données sont présentés dans une seule et même interface, celle du portail. En plus des bases de données bibliographiques et plein texte, il est possible d'envoyer des requêtes à des catalogues de bibliothèques, de musées, des dépôts institutionnels ou toute autre ressource similaire (20, GIUSTI). Il apparaît nécessaire d'opérer dès à présent une distinction entre recherche fédérée et recherche unifiée, car même si elles reposent toutes deux sur une même logique, interroger plusieurs répertoires de ressources et communiquer les résultats via une interface unique, leurs champs d'action diffèrent :

- La recherche fédérée permet d'interroger des ressources internes à l'institution ET des ressources externes (d'autres institutions)
- La recherche unifiée ne permet d'interroger QUE les répertoires de ressources au sein d'une même institution

### **2.1.2 L'utilité principale de la recherche fédérée : simplifier la démarche de l'utilisateur**

A l'occasion des tests menés en 2008 sur différents utilisateurs dans le cadre de la refonte du portail de la BPI, il a été souligné que ceux-ci sont facilement déroutés par la présence de deux interfaces de recherche nommées « Interroger le catalogue » et « Elargir la recherche ». En effet une majorité d'entre eux ont mal interprété la nature de ces deux fonctionnalités, certains considérant par exemple qu'il s'agissait de la même chose, d'autres pensant que renseigner les deux interfaces permettrait de combiner des critères de recherche. D'où l'intérêt pour le portail de la BPI de ne proposer qu'une seule interface de recherche réunissant l'ensemble des fonctionnalités et contenus disponibles sur le portail, ce qui explique le développement d'un moteur de recherche fédérée. Les usagers ont d'ailleurs exprimé une perte de repères lorsqu'ils devaient passer d'une base de données à une autre (26, BOURGEOUX, FRESNAU).

Bruno Menon, dans le cadre de la journée d'étude ADBS « Optimiser l'accès à l'information, une opportunité pour les langages documentaires », se fonde sur des études réalisées sur



les comportements des internautes utilisant un moteur de recherche. Il en retient une utilisation croissante des moteurs de recherche qui, en ce qui concerne les usages d'Internet, vient juste après celle du courrier électronique. Par ailleurs il note que les internautes demeurent généralement fidèles à un ou deux moteurs de recherche et font confiance aux résultats qu'ils leur livrent et ce quelle que soit la nature de l'information (commerciale, scientifique etc.). Or il est intéressant de remarquer que les internautes conservent des démarches de recherche très simples :

- Pas d'utilisation d'opérateurs booléens (ET, OU, SAUF)
- Peu de mots utilisés dans une requête
- Prise en compte des premiers résultats affichés uniquement
- Peu de fonctionnalités avancées sont utilisées : hormis la catégorisation des résultats, les utilisateurs boudent par exemple la recherche multilingue et la représentation des résultats sous forme cartographique

Au cours de cette journée Michel Bernardini présentait le système LEOnard développé aux Etudes économiques de BNP-Paribas afin de proposer une interface de recherche « grand public » pour satisfaire les membres de l'entreprise, étant conscient que « 90% des utilisateurs cherchent sur Google en première indication » (48, MENON). Au centre de ce système permettant une recherche fédérée sur les différentes bases de BNP Paribas (Intradoc, Kiosqueéco, Infosecteur) se trouve le logiciel Polyspot dont les technologies statistiques permettent de traiter et de rechercher des documents de formats et de langues hétérogènes.

La recherche fédérée représente donc un outil judicieux à mettre en place pour une raison : simplifier la recherche d'information de l'utilisateur. Ce dernier, lorsqu'il a plusieurs bases de données à disposition, préfère utiliser une interface unique plutôt que d'avoir à manipuler différentes interfaces, ce qui nécessite de sa part un apprentissage pour chaque nouvelle interface. Il existe donc un besoin pour les internautes d'une interface simple car unique. En outre l'internaute ne peut en aucun cas être susceptible de savoir sur quelle base de données l'information qui l'intéresse peut se trouver, il peut ainsi être découragé par l'idée d'avoir à en consulter plusieurs avant de trouver ce qu'il recherche, par l'échec et par le temps passé. Une interface d'interrogation unique permet dans une certaine mesure, de contourner le problème du choix des ressources à interroger, celles-ci ayant été présélectionnées par le responsable des collections. En outre, les utilisateurs sont davantage intéressés par les modes d'interrogation simple que par les fonctions de recherche avancée demeurant l'apanage des spécialistes et professionnels de l'information-documentation, probablement en raison de la « googlisation » des méthodes de recherche des internautes. D'où l'intérêt de confiner au « back office » le choix des bases à interroger et la traduction

des requêtes dans des termes compris par les différents répertoires de ressources, l'utilisateur lambda n'étant intéressé que par le résultat final.

Or simplifier la recherche pour l'utilisateur n'est pas le seul avantage d'une recherche fédérée. Cette fonctionnalité permet également de trouver des sources d'informations « cachées » et ainsi d'accroître la visibilité des bases de données et des ressources disponibles dans une institution, l'internaute interrogeant alors des bases dont il ignorait l'existence (20, GIUSTI). La plupart des chercheurs se contentent d'un nombre restreint de bases de données qu'ils jugent fiables et pertinentes pour leur domaine de recherche. La recherche fédérée peut toutefois leur permettre d'élargir leurs perspectives en leur proposant, sans qu'ils aient à le demander, de nouvelles sources d'informations et par conséquent de nouvelles ressources.

Le moteur de recherche fédérée peut aussi offrir des fonctionnalités d'exploitation. Il permet de mémoriser les requêtes, de consulter l'historique des recherches et de définir des alertes basées sur une requête définie par l'utilisateur, mémorisée par le portail et déclenchée périodiquement pour prévenir l'utilisateur des nouveautés.

La recherche fédérée demeure donc un outil clé d'une solution de type portail dans la mesure où elle répond à l'objectif d'aider les internautes à trouver l'information pertinente parmi le déluge d'informations numériques hétérogènes et dispersées sur le Web en opérant en back office une présélection des ensembles de ressources à interroger. L'ambition est de faire accéder l'utilisateur à des contenus hétérogènes sans avoir à multiplier les clics, sans avoir à faire à des interfaces de recherche complexes destinées avant tout à des spécialistes.

### **2.1.3 Sortir d'une interface de recherche « à la Google »**

Si les utilisateurs ont besoin d'une interface de recherche simple, il demeure néanmoins possible de proposer à ceux-ci une expérience de recherche plus intuitive qui parvienne à sortir du carcan de la « recherche à la Google » tout en demeurant simple :

- Proposer, comme le site de l'ECPAD, une recherche par nuage de mots-clés dynamique et interactif. Dynamique car la taille des mots-clés grossit dans le nuage parallèlement au nombre de notices dans lesquelles ils apparaissent, ce qui permet à l'utilisateur de comprendre en un regard quelles sont les thématiques les plus couvertes par les ressources du site et donc si celui-ci est susceptible de lui proposer l'information qu'il recherche. Interactif car un simple clic sur un mot-clé du nuage

permet de lancer une requête sur l'ensemble des bases de données du site (vidéo, photos, articles etc.), ce qui épargne à l'utilisateur la réflexion sur le terme juste à saisir et garantit la pertinence des résultats. Le nuage de mots-clés est accessible en bas à droite sur toutes les pages du site, il suffit de cliquer sur l'un des termes pour lancer la requête (ici « Fonds Seconde Guerre Mondiale ») :

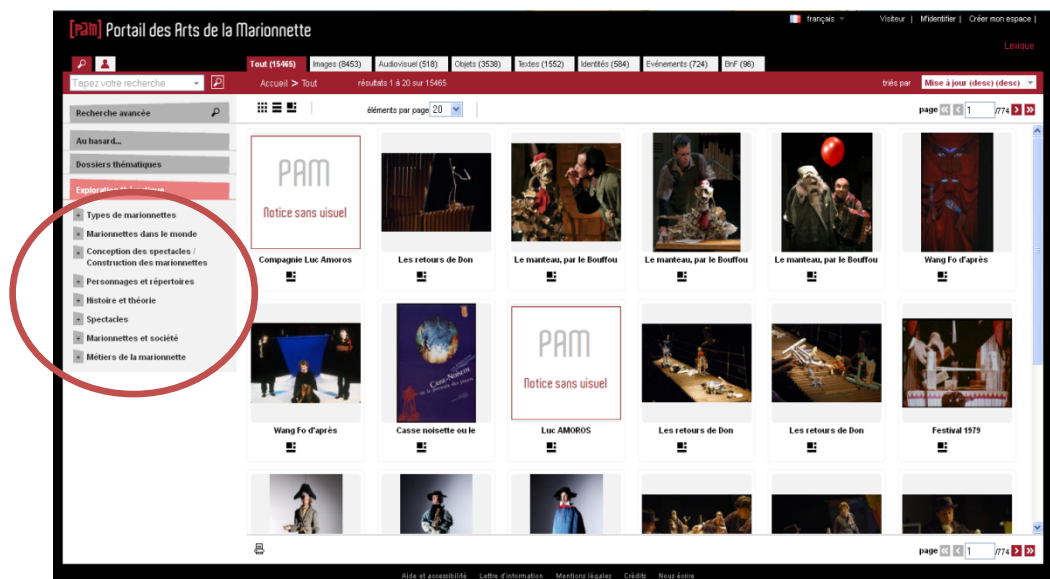


Les résultats en rapport avec le mot-clé sélectionné dans le nuage s'affichent à l'écran :

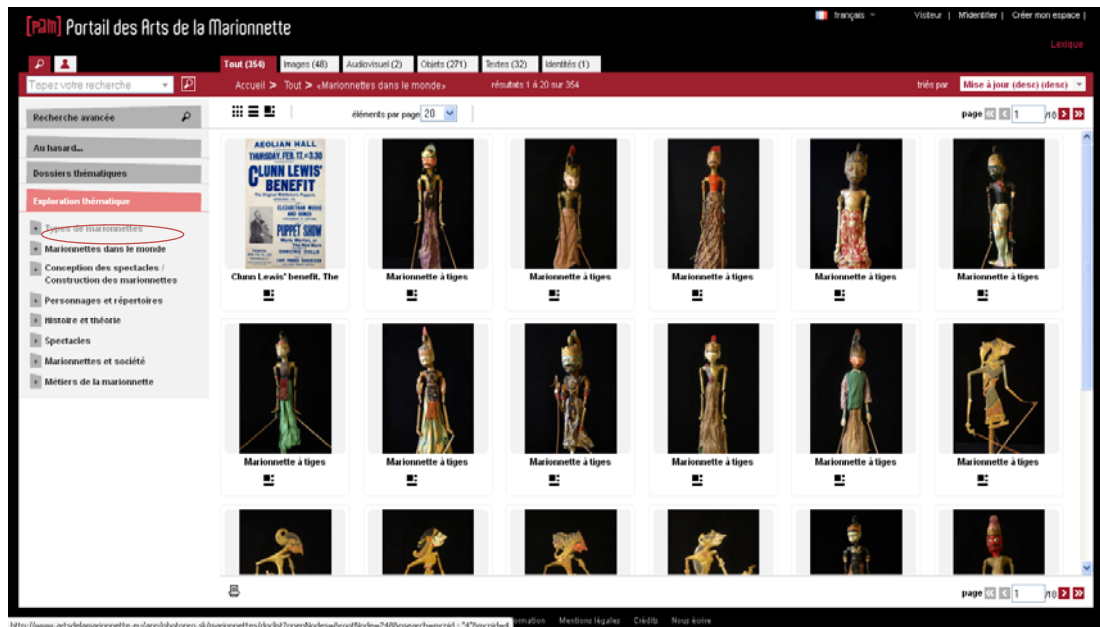


- Proposer, comme le Portail des Arts de la Marionnette, une recherche via une arborescence thématique. « L'exploration thématique » permet à l'utilisateur de naviguer dans l'arborescence des mots-clés en partant des termes génériques (les plus larges) pour arriver aux termes les plus spécifiques. D'une part cela permet d'épargner, comme avec le nuage de mots-clés de l'ECPAD, la réflexion quant au terme à saisir. D'autre part cela permet de satisfaire aussi bien l'utilisateur qui n'a qu'une vague idée de ce qu'il recherche que l'utilisateur qui lui en a une idée précise, le premier sélectionnera volontiers le terme le plus générique et pourra ensuite spécifier sa recherche à la vue des termes spécifiques liés, le second sélectionnera le terme le plus spécifique et élargira son champ de recherche avec un terme plus générique. Ainsi, en plus d'être simple, ce mode de recherche thématique est en mesure de satisfaire les différents profils d'utilisateurs.

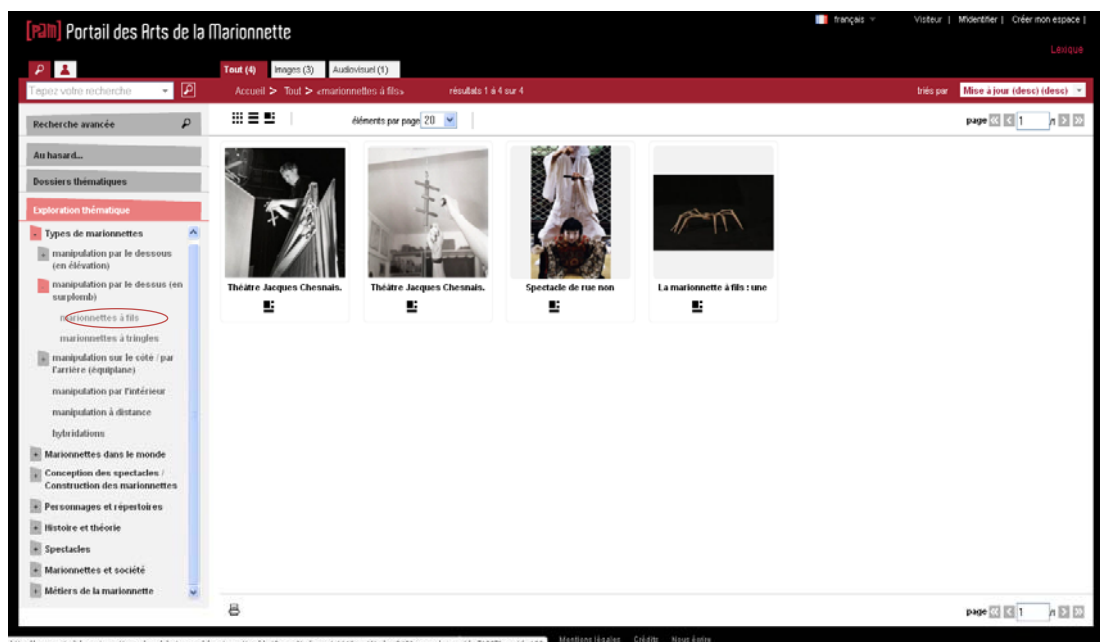
Sur le bloc gauche de l'écran, les termes génériques à sélectionner :



Si l'on clique sur le mot clé « Marionnettes dans le Monde », l'ensemble des notices indexées avec ce mot-clé (ou avec un terme spécifique lié) s'affichent à l'écran (18 pages de résultats au total) :



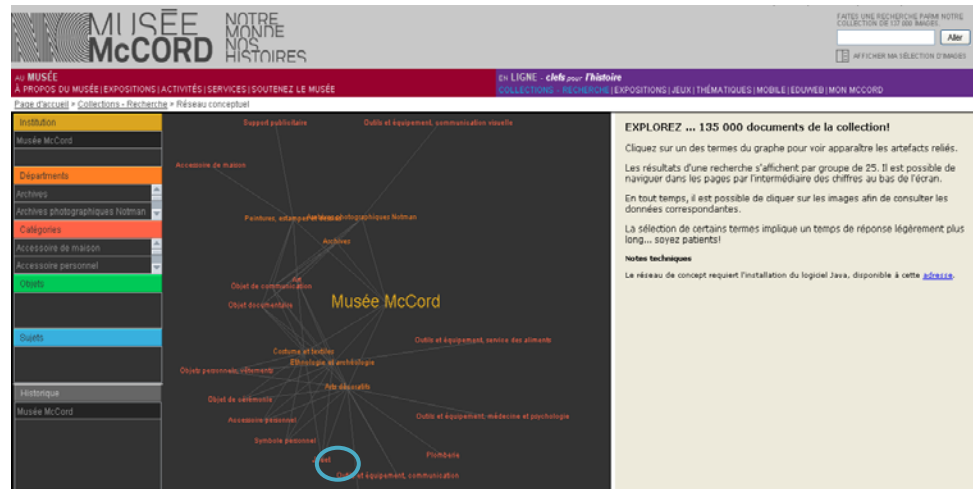
Si l'on souhaite avoir un résultat plus fin, et donc moins d'occurrences, il suffit de « développer » l'arborescence en cliquant sur l'icône « + » à gauche du mot-clé voulu. Une fois que l'on a atteint le niveau le plus spécifique, on obtient des résultats plus précis (une page contenant quatre résultats pour le mot-clé « Marionnettes à fils ») :



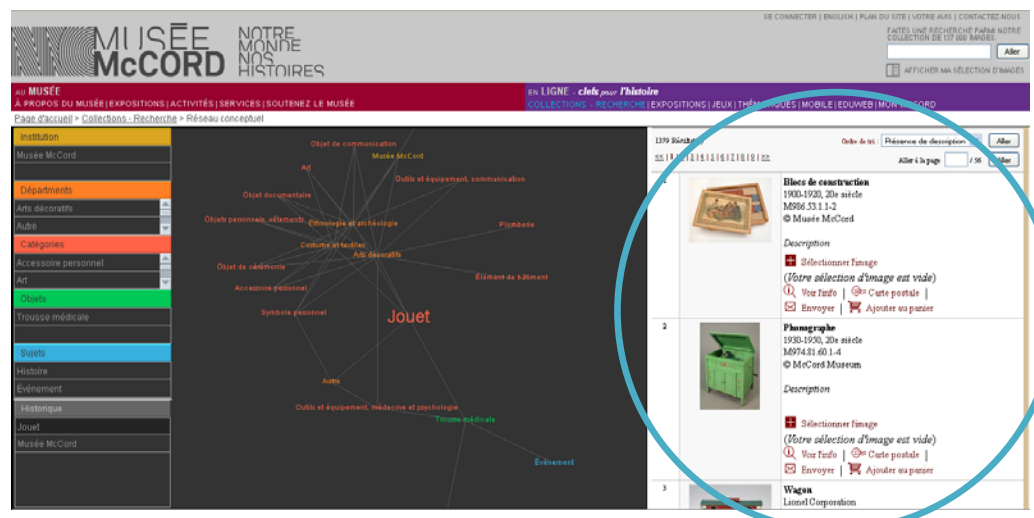
- Proposer, comme le Musée Mc Cord, une recherche via une « Mind Map », c'est-à-dire dans une représentation graphique des thématiques du site et de leurs

relations, ce qui demeure un mode de représentation efficace car permettant à l'utilisateur d'appréhender la globalité du microcosme thématique couvert par le site en quelques clics et en quelques coups d'œil.

Le « réseau conceptuel » permettant la recherche sur les pages et les bases de données du site :



Lorsque l'on clique sur l'un des mots-clés, une nouvelle « sous-arborescence » lui étant liée apparaît à l'écran. Voici le résultat si l'on clique sur le mot-clé « Jouet » (entouré en bleu dans le précédent écran). Les résultats liés apparaissent à droite de l'écran :



## 2.2 Un procédé reposant sur des protocoles d'interrogation : la nécessité d'adopter une solution technique adaptée

### 2.2.1 Présentation des protocoles d'interrogation synchrone et asynchrone

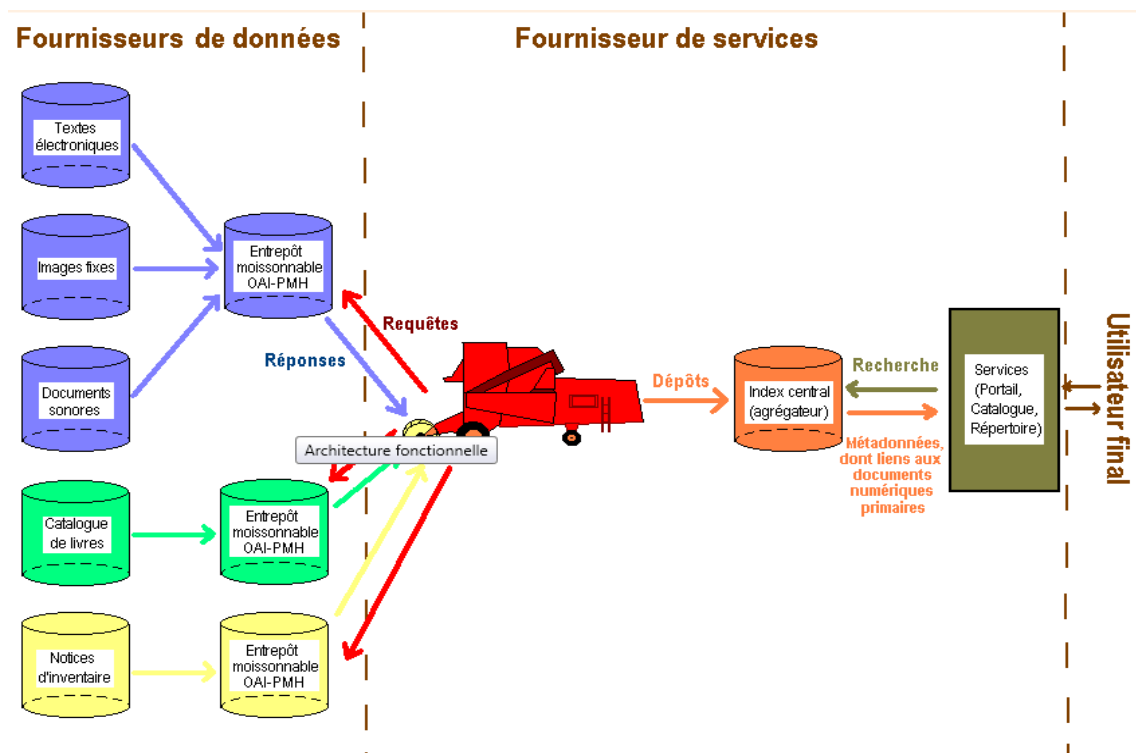
On distingue deux grandes tendances dans les protocoles permettant la recherche fédérée sur plusieurs bases de données, toutes deux reposant sur une syntaxe d'échange XML (*eXtensible Markup Language*) :

- Les protocoles d'interrogation asynchrone → le protocole OAI-PMH (*Open Archive Initiative Protocol for Metadata Harvesting*) qui consiste en une fédération des contenus au sein d'un entrepôt OAI commun à l'ensemble des répertoires de ressources concernés : les différentes bases de données sont transformées en entrepôts OAI qui sont moissonnés par un programme appelé « moissonneur », ce « moissonneur » ne conserve que les dernières métadonnées ajoutées depuis sa dernière moisson et les « stocke » dans un entrepôt OAI commun qui permettra d'alimenter l'index consultable en ligne. De ce fait les bases de données ne sont jamais interrogées directement. Par ailleurs les bases de données n'ont pas besoin d'être moissonnées en ligne, il est en effet également possible de faire un export XML sur l'entrepôt commun qui lui doit être interrogeable en ligne.

Le schéma suivant, réalisé par François Nawrocki<sup>1</sup>, permet de mieux comprendre le fonctionnement du protocole OAI :

---

<sup>1</sup> NAWROCKI, François. Le protocole OAI et ses usages en bibliothèques [en ligne]. Mise en ligne le 28 janvier 2005, dernière mise à jour le 15 février 2005 [consulté le 9 août 2011].  
<http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>

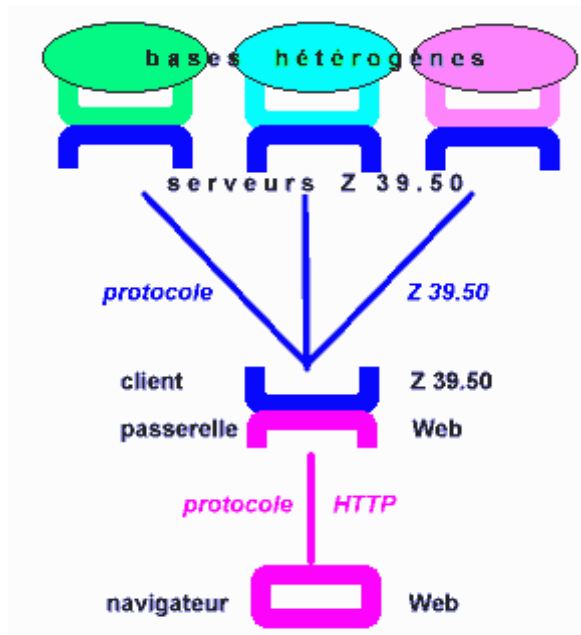


- Les protocoles d'interrogation synchrone → le protocole Z39.50 évoluant progressivement aujourd'hui grâce aux protocoles SRU/SRW (« Search/Retrieve via URL/Web »). L'interrogation se fait directement en ligne sur les diverses bases de données concernées : la requête est traduite en Z39.50 à la base de données par un serveur Z39.50 qui fournit une réponse traduite en Z39.50 par ce même serveur. SRU/SRW ont été développés dans l'optique de retranscrire les procédures pour les rendre conformes à celles du Web : il s'agit, pour résumer, de transcrire les requêtes Z39.50 dans une seule URL (selon une technique REST pour le SRU et une encapsulation SOAP pour le SRW). Le protocole Z39.50 nécessite donc que les bases de données soient interrogeables en ligne.

Le schéma suivant, réalisé par Dominique Lahary<sup>2</sup>, permet de mieux comprendre le fonctionnement du protocole Z39.50 :

<sup>2</sup> LAHARY, Dominique. La norme Z3950 [en ligne]. Mise en ligne octobre 1998, dernière mise à jour le 18 mars 2003 [consulté le 9 août 2011].  
<http://www.lahary.fr/pro/z3950/index.htm>





## 2.2.2 Comparatif des forces et faiblesses des protocoles existants

Le tableau suivant dresse un comparatif des avantages et des inconvénients des protocoles OAI-PMH et Z39.50 dans le cadre de la mise en place d'un portail documentaire. Selon les contextes et les ambitions d'un projet de portail documentaire, l'une de ces solutions apparaît davantage adaptée.

|                   | Protocole OAI  | Protocole Z39.50   |
|-------------------|--|--|
| Le fonctionnement | Les notices de chaque base sont basculées dans un entrepôt OAI, cet entrepôt est moissonné selon une fréquence prédéfinie. La moissonneuse « dépose » ses récoltes dans un index central qui sera interrogé par l'utilisateur  | Chaque base de données est interrogée par un serveur propre qui traduit les requêtes et les réponses en Z39.50   |
| Les avantages     | <ul style="list-style-type: none"> <li>-Rapidité de la réponse car une requête est adressée à une base « fédératrice », <b>la fédération a donc été effectuée en amont</b></li> <li>-Aucun risque d'altération pour les bases d'origine</li> <li>-Aucune surcharge sur les serveurs des bases d'origine</li> </ul> | <ul style="list-style-type: none"> <li>-Interrogation en temps réel, ce qui permet d'avoir accès aux dernières notices créées ou modifiées → <b>protocole synchrone</b></li> <li>-Prise en compte de divers formats de notices structurés (tels que le MARC, le GRS-1etc.), non structurés (SUTRS, etc.),</li> </ul> |

|                                  |   |  |
|----------------------------------|---|--|
|                                  | -Aucun besoin de faire basculer une base de données locale en OPAC Web, seul l'entrepôt de stockage doit être accessible en ligne   | formats d'image et de son  |
| Les inconvénients                | <p>-Le résultat d'une requête reflète l'état « figé » des répertoires de ressources au dernier moment où ils ont été moissonnés, les informations les plus récentes ne sont donc pas accessibles→il s'agit d'un <b>protocole asynchrone</b></p> <p>-Le format Dublin Core (qui est le format préconisé par le protocole OAI-PMH, même si d'autres formats sont pris en compte) est un format relativement pauvre (15 champs de métadonnées pour un Dublin Core non qualifié) par rapport à la richesse des métadonnées liées à des collections muséales</p> | <p>-Fait peser une charge supplémentaire sur les serveurs des bases de données</p> <p>-Temps de réponse plus long, une requête est dupliquée pour chaque base</p>              |
| Les contraintes de mise en place | <p>-Si les bases de données ne sont pas en OPAC web (interrogeables en ligne), cela nécessite que les notices de tous les répertoires de ressources soient exportées dans l'entrepôt dans une syntaxe XML et de préférence dans un format Dublin Core non qualifié.</p> <p>-Configurer le moissonneur</p>   | -Nécessite que chaque base soit accessible en ligne, d'où le besoin de basculer les répertoires de ressources en OPAC Web pour les rendre interrogeables par un serveur Z39.50 |

### 2.2.3 La solution ODBC

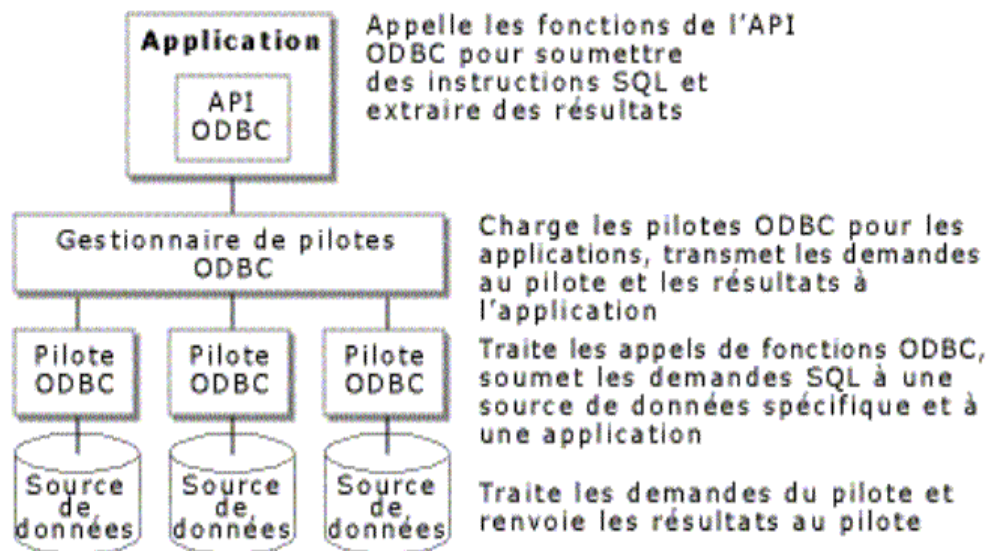
Outre les protocoles OAI et Z39.50, il existe un logiciel permettant d'interroger simultanément plusieurs bases de données en synchrone et prenant en compte tous les formats de notices sans restriction : l'ODBC (*Open Database Connectivity*).

L'architecture ODBC est formée de quatre composants :

- Une interface de programmation d'application (API) qui appelle les fonctions ODBC pour effectuer une connexion à une source de données, envoyer et recevoir des données, puis effectuer la déconnexion.

- Un gestionnaire de pilotes qui fournit des informations à une application (par exemple, une liste des sources de données disponibles), charge les pilotes dynamiquement en fonction des besoins et vérifie les arguments et la transition d'état.
- Un pilote qui traite les appels de fonctions ODBC et gère tous les échanges entre une application et une base de données relationnelle spécifique. Si nécessaire, le pilote peut convertir la syntaxe SQL standard dans le langage SQL natif de la source de données de destination.
- Une source de données qui comprend les données et le moteur de base de données associé.

L'application utilise l'API ODBC pour se connecter à une source de données, soumettre des instructions SQL, extraire des données et se déconnecter. Un gestionnaire de pilotes, situé entre l'application et les pilotes ODBC, décide du pilote à charger et gère les communications lors des appels aux fonctions des pilotes. Enfin, les pilotes mettent en œuvre les fonctions de l'API ODBC pour la base de données concernée.



Par conséquent le logiciel ODBC permet d'accéder de façon uniformisée à des données stockées dans différents formats de bases de données. L'ODBC repose sur des connecteurs spécifiques et configurables qui permettent d'interroger simultanément différentes bases de données et d'extraire les résultats selon les formats propres à ces bases de données (contrairement au protocole OAI qui restitue l'information dans un format pauvre), ce qui pourrait s'avérer judicieux dans un contexte muséal dans la mesure où les répertoires de

ressources dans les musées contiennent des notices riches en métadonnées qui seraient, dans le cadre d'un moissonnage OAI, forcément appauvries par un affichage en Dublin Core.

#### **2.2.4 De la fédération des contenus à la fédération des contenants : les « dataware-houses »**

Au-delà du choix entre protocole d'interrogation synchrone ou asynchrone, la recherche fédérée peut techniquement s'appuyer sur une fédération non de contenus, mais de contenants. C'est-à-dire que ce ne sont pas les données qui sont mises en commun mais leurs bases de données d'origine au sein d'entrepôts de données (*dataware houses*).

Il s'agit d'une technologie déjà largement utilisée dans l'informatique décisionnelle en entreprise. Cela consiste à centraliser dans une base de données commune des systèmes métier aux structures très hétérogènes. Cette option permet aux acteurs d'un projet de mutualisation de garder les applicatifs métier de leurs applications, mais les contenus pertinents pour le portail sont mutualisés dans une base de données commune. La difficulté principale réside alors dans la synchronisation des données entre cette base matricielle et chaque application. L'un des avantages de cette solution est d'être souple, c'est-à-dire qu'elle permet d'accueillir tout nouvel acteur du portail grâce à une structure de données facilement paramétrable et à géométrie variable (16, BREBION, BEL). Par exemple le logiciel doit autoriser la création de nouveaux champs non prévus initialement si un nouvel acteur, un musée, possède comme champ dans son application propre de gestion de fonds un champ « fragilité ».

### **2.3 La pertinence des résultats : le talon d'Achille des interfaces de recherche fédérée**

La recherche fédérée ne constitue pas une solution miracle face à la surabondance de l'information. Il ne faut pas considérer qu'une interface de recherche unique permettant d'interroger simultanément plusieurs bases de données hétérogènes soit infaillible en terme de pertinence des résultats. En effet, une recherche simultanée ne résolve en rien la complexité due à l'hétérogénéité des bases de données. Par exemple des logiciels de « clusterisation » tels que MetaLib ou WebFeat entraînent une perte significative d'efficacité par rapport à une recherche base par base (24, STERN). Par ailleurs il apparaît que les utilisateurs des logiciels de ce type demeurent frustrés par le temps de réponse pour

l'affichage des résultats, ce qui les oblige à limiter le nombre de bases de données à interroger et à se passer d'outils d'exploitation de résultats plus puissants.

La contrainte de l'intégration de sources hétérogènes de données en une interface commune ne consiste pas uniquement à résoudre le problème du protocole d'échanges et d'homogénéisation des formats documentaires, elle consiste aussi à exprimer les relations entre les données des sources intégrées. La recherche fédérée pose le problème du bon mot-clé à utiliser quand il n'existe pas de thésaurus transversal permettant l'interrogation de l'ensemble des catalogues (20, GIUSTI). Dans un système d'information bien structuré, l'utilisateur doit trouver tous les documents en rapport avec le mot-clé qu'il a employé. Or s'il n'existe pas de transversalité des thésaurus entre les différentes bases de données, le moteur de recherche ne « comprendra » pas que « Cassius Clay » (le terme saisi par l'utilisateur) correspondra dans les bases interrogées aux termes « Cassius Clay », « Mohammed Ali », « Muhammad Ali ». Les résultats ne prendront en compte que les notices d'une base de données qui présenteront comme mot-clé « Cassius Clay ». Par conséquent la recherche d'information ne sera que partiellement pertinente.

De ce fait cela pose le souci de l'interopérabilité. Aurélia Giusti définit l'interopérabilité comme « la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre. En d'autres termes l'interopérabilité est la capacité à dialoguer, à interagir, et à échanger des données (avec le minimum de perte d'information et de fonctionnalités) de deux systèmes disposant de différentes caractéristiques en terme de matériels, logiciels, structures de données et interfaces » (20, GIUSTI, p.25). Par conséquent, si l'on souhaite donner l'accès à plusieurs répertoires de ressources via une interface d'interrogation unique, cela suppose l'interopérabilité des bases de données et des outils d'indexation communs ; si le langage d'indexation n'est pas commun, il est alors indispensable de réfléchir à l'interopérabilité sémantique afin de surpasser l'hétérogénéité sémantique. Par conséquent, dans le cadre de la mise en place d'un portail de recherche fédérée, il est conseillé d'identifier les vocabulaires utilisés dans l'ensemble des bases de données concernées afin de déterminer s'il est possible d'utiliser un vocabulaire contrôlé commun à toutes. Si tel n'est pas le cas, l'élaboration d'une table d'équivalences est à envisager en plus d'un travail de normalisation des métadonnées dans les notices.

*Nous avons remarqué dans cette partie que les apports des portails documentaires sont indéniables : simplicité, rapidité, fonctionnalités permettant de renouer le contact avec les publics. La recherche fédérée, si elle représente une brique essentielle des portails documentaires, ne constitue cependant pas une solution « clé en main » rendant possible ex nihilo l'interrogation de répertoires de ressources disséminés et hétérogènes. En effet, outre l'interopérabilité technique à mettre en place, il est nécessaire, pour ne pas dire vital, de réfléchir à l'interopérabilité sémantique. Car s'il est nécessaire de rendre matériellement possible la recherche sur plusieurs répertoires de ressources, il ne faut pas omettre de la rendre intellectuellement pertinente. De ce fait nous allons constater dans la partie suivante qu'une réflexion autour des vocabulaires contrôlés, en particulier les thésaurus, organisant les connaissances dans chaque répertoire de ressources doit être menée afin de définir leur utilité dans le cadre d'une recherche fédérée alors que les technologies du Web sémantique progressent...*

Deuxième partie  
Les langages contrôlés à l'heure du  
Web de données : quand les outils  
traditionnels garantissent la viabilité  
des interfaces de recherche des  
portails en ligne

# 1 Le rôle des thésaurus dans la structuration des informations en ligne : le retour des langages contrôlés pour assurer la pertinence de la recherche en ligne

---

## 1.1 Le thésaurus : définition d'un outil de structuration des informations

### 1.1.1 Le thésaurus : un langage d'indexation contrôlé

L'indexation a pour but de faciliter l'accès au contenu d'un document ou d'un ensemble de documents à partir d'un sujet ou d'une combinaison de sujets. On distingue l'indexation en langage libre, où les notions sont représentées par un ou plusieurs mots de la langue naturelle, de l'indexation en langage contrôlé dans laquelle le choix des mots est fixé par un langage documentaire. Il s'agit d'une médiation effectuée par l'indexeur entre le document et l'utilisateur du système, qui nécessite une analyse du document. Les langages documentaires se répartissent en deux grandes catégories : les langages à structure hiérarchique, les classifications et les langages à structure combinatoire, les thésaurus.

Les langages documentaires présentent plusieurs fonctions (57, DESFRICHES DORIA) :

- Une fonction normalisatrice permettant d'harmoniser l'expression de la clé de recherche dans la recherche par sujet.
- Une fonction d'ajustement ou de navigation liée à la complexité de notion de sujet d'un document et à la difficulté d'en déterminer les limites, et qui permet à l'utilisateur de s'approcher par appariements successifs, de son sujet de recherche. La navigation est rendue possible par les liens sémantiques et linguistiques entre les termes d'un langage documentaire.
- Une fonction d'aide à la sélection des documents les plus pertinents par l'observation de l'indexation effectuée dans les notices.

La nécessité d'un contrôle terminologique de l'expression du sujet des documents, à travers un langage documentaire, est fermement établie en théorie puisqu'elle découle des phénomènes de polysémie et de polynymie qui sont nombreux dans la langue.



Les thésaurus établissent un vocabulaire contrôlé, une terminologie standardisée dans laquelle chaque concept est représenté par un terme, un « descripteur », qui est utilisé pour indexer et pour rechercher de façon pertinente. Le thésaurus doit permettre au professionnel d'identifier tous les descripteurs qui pourraient être attachés au document selon ceux que les utilisateurs seraient susceptibles de choisir pour leur requête. Un bon thésaurus fournit, à travers sa hiérarchie enrichie par les relations associatives entre concepts, une carte sémantique pour les professionnels de l'information-documentation et pour leurs utilisateurs.

Les thésaurus consistent en une sélection de concepts enrichis d'informations relatives à leurs relations sémantiques (relations hiérarchiques et relations associatives). Les descripteurs représentent des concepts : la différence existant entre un « mot » et un « concept » est que différents mots peuvent avoir le même sens et qu'un même mot peut avoir différents sens, le concept lui n'a qu'un seul sens. Par exemple le mot « avocat » possède deux sens : le légume d'une part, la profession d'autre part. L'utilité d'un thésaurus est de prendre en compte les synonymies en utilisant les relations de type « Employé »/ « Employé pour » et les homonymies en spécifiant le sens entre parenthèses : « Avocat » (légume), « Avocat » (profession). En outre le thésaurus parvient à établir la structure d'un champ de connaissance grâce à l'emploi des « Termes associés », « Termes génériques », « Termes spécifiques ».

Un thésaurus est un langage combinatoire structuré à partir de relations :

- Relations d'équivalence permettant de renvoyer les divers synonymes d'un concept vers le descripteur choisi, les relations entre descripteurs sont indiquées par « EP » (« Employé pour ») alors que les relations entre un descripteur et un non descripteur sont indiquées par « EM » (« Employer »)
- Relations hiérarchiques situant les descripteurs dans leur environnement sémantique : Terme générique (TG)/Terme spécifique (TS)
- Relations associatives indiquant une proximité sémantique entre des termes situés dans des hiérarchies différentes, ces relations sont indiquées par « TA » (« Terme Associé »)
- Les notes d'application précisant le sens dans lequel le descripteur doit être employé

La conception classique des thésaurus oppose le langage naturel au langage contrôlé. Le langage naturel est constitué des mots figurant dans les titres, les résumés ou les documents primaires et sert aussi aux utilisateurs des moteurs de recherche pour composer leurs requêtes. Les langages contrôlés permettent de contraindre l'indexation pour éviter les

problèmes d'homonymies, polysémies, synonymies, la variabilité des formes et structures lexicales. Ces derniers visent à améliorer la pertinence des recherches en évitant le silence documentaire (mesuré par le taux de rappel correspondant au rapport entre le nombre de documents pertinents trouvés et le nombre de documents pertinents existant dans le système) et le bruit documentaire (mesuré par le taux de précision correspondant au rapport entre le nombre de documents pertinents trouvés et le nombre total de documents trouvés par le système).

« Un thésaurus est une liste d'autorité organisée de descripteurs et de non descripteurs obéissant à des règles terminologiques propres et reliés entre eux par des relations sémantiques (hiérarchiques, associatives ou d'équivalence). Cette liste sert à traduire en langage artificiel dépourvu d'ambiguïté des notions exprimées en langage naturel »<sup>3</sup>. Un descripteur est un terme retenu dans le thésaurus pour représenter sans ambiguïté une notion contenue dans un document ou dans une recherche documentaire.

### **1.1.2 L'essor des thésaurus dans un contexte d'avancée des outils et supports de communication**

Tout au long de l'évolution des systèmes d'information ont été développés des outils de description du contenu des documents à des fins de recherche documentaire. Ces outils ont pris de nombreuses formes : taxinomies, classifications, thésaurus.

L'apparition des langages documentaires est liée à plusieurs facteurs. En premier lieu la croissance du volume des documents (depuis la fin du 19<sup>ème</sup> siècle marquée par un foisonnement éditorial), puis l'essor de nouveaux supports d'inscription et enfin le développement de nouvelles modalités de communication de l'information. Dans les années 1970, les langages documentaires et l'informatique se développent conjointement, l'informatique permettant de contrôler les langages et d'aider à l'interrogation car les langages informatiques se définissent par leur capacité à effectuer des calculs et opérations complexes telles que le croisement de critères de recherche. Ainsi les thésaurus se développent, ces langages combinatoires ont été encouragés par le développement informatique et le besoin de croiser des critères, ce qui permet de naviguer dans les bases de données de façon non linéaire. Les langages documentaires sont donc nés à l'occasion

---

<sup>3</sup> HUDON M. Le thésaurus : conception, élaboration, gestion. Québec, éditions ASTED, 1994. Collection Clé en main.

des évolutions techno-culturelles conduisant à repenser les questions de l'organisation des connaissances et de l'accès à l'information.

Or si leurs origines précèdent l'apparition des ordinateurs, le développement des thésaurus a bel et bien suivi la diffusion des techniques informatiques. Les thésaurus documentaires à usage informatique trouvent leurs origines dans les années 1950. En effet, pour résoudre la « crise de la documentation scientifique », de grandes banques de références documentaires sont créées et mettent à disposition des fonds d'articles de revues, de rapports techniques, de communications de congrès etc. Or très vite les index produits pour répondre à des usages papier (KWOC, KWIC) ne suffisent plus à répondre aux besoins. C'est alors que le traitement documentaire à partir de descripteurs, puis de thésaurus documentaires apportent une solution concrète et efficace pour accéder à distance à cette documentation (43, DALBIN). Ainsi les thésaurus sont présents dès l'origine de l'informatique et ce grâce aux possibilités conjointes de cette technologie et de la logique combinatoire. Parmi les thésaurus pionniers nous pouvons citer le *MeSH* (base Medlards) et le *Chemical Engineering Thesaurus*. Le but de cette association est de capitaliser les ressources scientifiques et techniques tout en les rendant accessibles via les réseaux de télécommunication.

Le principe du « *Descriptor* » est établi dès les années 1950 par Calvin Northrup Mooers du MIT dans le cadre de solutions mécanisées pour la recherche d'information. Cela a conduit à l'élaboration de thésaurus comme dictionnaires de descripteurs.

Les thésaurus doivent permettre :

- De concevoir un système compact ne demandant pas de répertoire *a priori* de nombreux concepts, mais autorisant l'expression d'un grand nombre de sujets. De nouvelles formes d'indexation sont alors mises au point telles que les « descripteurs » de Calvin Mooers ou les « unitermes » de Mortimer Taube, et ce avec les éléments mécanographiques permettant les combinaisons lors des recherches.
- Prendre compte de la synonymie en harmonisant le vocabulaire employé par les auteurs, celui employé par les indexeurs et enfin celui employé par les utilisateurs.
- Guider les indexeurs et les utilisateurs dans le choix des termes appropriés, ce qui suppose la structuration sémantique des termes par les liens hiérarchiques et associatifs.

Par conséquent le thésaurus peut être défini comme un langage structuré, contrôlé et combinatoire. La logique de recherche booléenne reposant sur la logique combinatoire des thésaurus est parfaitement adaptée aux opérations de base effectuées par l'ordinateur, d'où un succès immédiat et durable des langages documentaires en contexte informatisé (61, MENON).

### **1.1.3 Le thésaurus : un outil demeurant l'apanage des documentalistes**

Contrairement aux classifications et vedettes matières qui sont maîtrisées par les bibliothécaires, les thésaurus demeurent l'apanage des documentalistes. Les thésaurus demeurent les langages documentaires les plus universels malgré leur caractère le plus souvent spécialisé (61, MENON).

Le thésaurus demeure avant tout un outil réservé aux professionnels de l'information-documentation, aux chercheurs et aux scientifiques face à des interfaces de recherche qui ont été initialement conçues pour traiter des requêtes claires et précises. Or il apparaît que le nombre de chercheurs d'information s'est accru depuis les années 1950, et que ce nouveau public n'est plus spécialiste de la recherche d'information. Face au mécontentement de ces nouveaux utilisateurs, les professionnels de l'information-documentation ont donc dû développer des outils de recherche « plein texte » sur l'ensemble des notices sans réelle évaluation de l'efficacité de ce mode de recherche et sans outils indicateurs de pertinence (filtre, pondération, expansion) pourtant nécessaires (44, DALBIN).

Par ailleurs les utilisateurs sont davantage intéressés par le contenu des documents (l'information utile) que par les documents en eux-mêmes. Deux approches peuvent être envisagées selon Sylvie Dalbin pour développer des thésaurus utiles aux utilisateurs:

- Travailler au plus près des producteurs de documents en poussant à une structuration et à un formalisme plus pointus dès leur rédaction. Ainsi un nombre important de métadonnées est produit à la source et peut être exploité au sein des systèmes documentaires, ce qui fait émerger des besoins importants et autonomes de développement et de maintenance de vocabulaires contrôlés (nomenclatures, thésaurus, taxonomies).
- Travailler au plus près des utilisateurs d'informations, ce qui implique de leur proposer des accès à une très grande variété de ressources documentaires par le

biais d'un langage de recherche proposant des accès cohérents, transparents et autonomes par rapport aux langages sources.

## **1.2 Le thésaurus face à Internet : d'un abandon progressif à une hybridation avec les interfaces de recherche statistique**

### **1.2.1 La croissance de l'Internet grand public et le déclin relatif des thésaurus**

Avec l'arrivée du Web au début des années 1990, les serveurs de bases de données spécialisées ont cherché à étendre leur clientèle aux internautes en développant des interfaces de recherche similaires à celles des moteurs de recherche, notamment en les simplifiant à l'extrême par une recherche en texte intégral qui s'oppose à la recherche par descripteurs demeurant le propre des spécialistes qui se référaient jusque là à des thésaurus sur papier ou intégrés à l'interface. En effet Sylvie Dalbin note un mécontentement des utilisateurs d'information face à la complexité et la multiplicité des interfaces des systèmes. Ils souhaitent être informés simplement et rapidement. La plupart de ces utilisateurs de banques de données ne sont effectivement pas des professionnels de la recherche documentaire ou des chercheurs/scientifiques pour qui le modèle documentaire, fondé sur des requêtes claires et concises, était conçu à la base dans les années 1950. De ce fait les thésaurus et les interfaces d'interrogation pour spécialistes ne leur conviennent pas. Pour remédier à ce problème, les professionnels de l'information-documentation ont mis au point des interfaces de recherche en plein texte sur l'ensemble des notices sans avoir au préalable évalué l'efficacité d'un tel mode de recherche et sans lui avoir adjoint des outils de filtrage, de pondération et d'expansion des résultats.

L'évolution du paysage documentaire a donc un impact sur le rôle des thésaurus et celui des vocabulaires contrôlés en général : baisse des activités d'indexation au profit d'autres activités face à l'augmentation des pratiques d'acquisition de ressources et de métadonnées associées, mutualisation des ressources interrogeables à distance structurées par des langages différents, développement de filières d'indexation et de classification automatiques aux dépens d'une indexation humaine (jugée plus longue et coûteuse), des documents multi-formats et de plus en plus plurilingues etc.

Ainsi l'usage des thésaurus pour la recherche demeure sous exploité dans les bases de données spécialisées. En effet il est fréquent, selon Bruno Menon, que les ressources soient

indexées avec un thésaurus alors que seule une recherche par requête libre est proposée sur le site Internet (61, MENON). Ce cas de figure est typique des ressources produites selon une approche documentaire traditionnelle, puis exportées vers un système qui n'en prend pas en compte la structure et les spécificités afin de proposer à un large public, qu'une recherche documentaire classique pourrait rebuter, une recherche simplifiée et un accès à des ressources hétérogènes n'étant pas indexées pour certaines. Par conséquent les potentialités des thésaurus sont sous-exploitées dans la mesure où les descripteurs assignés aux documents sont ni plus ni moins traités comme des éléments du texte intégral.

« Le World Wide Web aurait dû être pour les langages documentaires une formidable occasion d'ouverture, de diffusion et de popularisation, mais cette occasion n'a pas été saisie faute d'une attention suffisante prêtée à la conception ergonomique d'interfaces adaptées à leur utilisation par des non professionnels. » (61, MENON, p.26). En effet les moteurs de recherche sur Internet ne font pas tout, les pratiques de traitement intellectuel de l'information ont toujours leur place.

### **1.2.2 Des thésaurus plus que jamais nécessaires pour les nouveaux profils de chercheurs d'information confrontés au déluge**

Après une période de quasi-monopole dans les bases de données professionnelles, puis une seconde période au sein des bases documentaires des centres documentaires, les thésaurus et les descripteurs ont été relativement oubliés dans les années 1990 en France pour revenir progressivement ces dernières années.

Cela peut s'expliquer par le fait que la recherche libre n'est pas, contrairement à ce qu'un public profane pourrait penser, le mode de recherche le plus simple. En effet la recherche libre fait peser sur l'utilisateur du système l'essentiel de la charge de travail intellectuel nécessaire à l'élimination des problèmes posés par la polysémie et la polynymie (plusieurs termes partageant un même sens). L'utilisateur doit donc, face à des résultats peu ou pas pertinents, réfléchir à la reformulation de sa requête afin d'éliminer toute ambiguïté des termes employés. Se passer d'un langage contrôlé s'avère donc contre-productif.

La massification de l'information sur le web introduit la nécessité pour les interfaces orientées utilisateurs, plus simples et plus souples, d'articuler un ensemble de métadonnées construit à partir d'un vocabulaire contrôlé. La simplification de la recherche d'information passe non seulement par une interface à l'ergonomie simplifiée en *front office* mais aussi par

un recours aux vocabulaires contrôlés en *back office* dans la mesure où ceux-ci sont les plus à même d'assurer la pertinence des résultats.

Toutefois on observe désormais une multiplicité des profils d'utilisateurs et de leurs pratiques. Outre les chercheurs scientifiques demeurant le modèle de référence dans la conception de systèmes documentaires, le panel des profils s'est élargi. Dans l'environnement professionnel, de nouveaux chercheurs d'information apparaissent et disposent de besoins plus ponctuels et plus diffus. Or ces nouveaux profils ne sont pas coutumiers d'une recherche via un thésaurus et préfèrent se voir proposer des interfaces de recherche plus abordables d'un point de vue ergonomique. Celles-ci doivent être développées pour répondre :

- A des utilisateurs dont l'activité professionnelle n'est plus centrée sur la recherche d'information mais sur une production autre : alimenter un système d'information, rédiger un rapport, répondre à une sollicitation ponctuelle etc.
- A des utilisateurs ayant besoin d'informations de nature très variable : norme, statistiques, image, manuel d'utilisation, article, schéma etc.
- A des utilisateurs attendant d'un système documentaire des fonctionnalités diverses : syndication, outils statistiques, « text mining », extraction de données etc.
- A des utilisateurs devant faire face à une masse de sources à exploiter : modalités de consultation plus simples et ergonomiques

Le besoin de ces nouveaux utilisateurs pourrait se résumer comme suit : « être informé en faisant le moins d'efforts et si possible sans même avoir à effectuer une recherche ».

### **1.2.3 Du thésaurus à l'indexation au thésaurus à la recherche : l'usage d'un vocabulaire contrôlé pour développer des interfaces de recherche simples**

Le thésaurus a été largement utilisé pour la recherche d'information en miroir du thésaurus utilisé pour l'indexation, c'est-à-dire que l'indexation des documents reposait sur le même thésaurus qui servait à la recherche de ceux-ci. Or dans un contexte de développement de la recherche en ligne et de diversification des profils d'utilisateurs, les années 1980 voient les spécialistes anglo-saxons réfléchir à des thésaurus exclusivement pour utilisateurs (*user*

*thesaurus*), des thésaurus à la recherche. Les spécialistes ont défini les caractéristiques de ces outils : un « superthésaurus » ou encore une liste de synonymes, des réserves de termes, des thésaurus traditionnels élargis, des thésaurus fusionnés avec des traits particuliers, des thésaurus indépendants du thésaurus à l'indexation (42, DALBIN).

La plupart des utilisateurs désirent passer outre la phase complexe de formulation d'une requête à partir de la sélection de descripteurs dans un thésaurus car ils préfèrent porter leur attention à l'examen des résultats et non à la requête elle-même. Pour simplifier l'utilisation d'un thésaurus, il pourrait être envisagé, selon Sylvie Dalbin (44, DALBIN) de ne pas utiliser celui-ci dans sa forme hiérarchique mais pour ses équivalences, c'est-à-dire que le professionnel de l'information-documentation ne doit pas se focaliser sur les relations entre termes génériques et termes spécifiques mais développer les relations d'équivalence entre descripteurs (langage contrôlé) et non descripteurs (en langage naturel). Ainsi dans le cadre d'une recherche dite « plein texte » ne seraient exploités que les descripteurs de la notice auxquels sont associés les non descripteurs synonymes qui ont été saisis par l'utilisateur. Ce type d'interface permettrait de lancer une requête sur les champs de type « descripteurs » (champ renseignant les descripteurs attribués par l'indexeur à la notice) contenant du langage contrôlé ainsi que sur les champs de type « titre » et « résumé » contenant du langage naturel. D'où une interface hybride entre simplicité d'une recherche plein texte en langage naturel et pertinence d'une recherche à partir d'un langage contrôlé car l'utilisateur n'a qu'à saisir un ou plusieurs mots dans une barre de recherche tandis qu'en *back office* les professionnels de l'information-documentation ont déjà développé les équivalences potentielles entre ses mots en langage naturel et les descripteurs présents dans le thésaurus.

Néanmoins Sylvie Dalbin met en garde contre ce type de solution mettant de côté la prise en compte des relations hiérarchiques dans la mesure où ce type de relation permet une fonctionnalité d'expansion des résultats et donc d'éviter le silence. En effet si un utilisateur du futur portail du Musée National du Sport désire obtenir tous les documents traitant du football en France, il ne verra s'afficher que les notices présentant les descripteurs « Football » et « France » avec le type de solution précédemment exposé. Or si aucun document du fonds ne traite spécifiquement du football en France mais que quelques documents traitent plus généralement du football en Europe de l'Ouest (et donc susceptibles de traiter en partie du football dans l'Hexagone), l'écran de résultat des recherches n'afficherait tout de même aucune occurrence car des informations intéressantes pour l'utilisateur demeureraient invisibles. Or avec une prise en compte des relations hiérarchiques au sein d'un thésaurus, la recherche aurait été élargie à un niveau plus général : « France » → « Europe de l'Ouest » et le silence ainsi évité.



En outre si les thésaurus ne sont pas proposés en *front office* à l'utilisateur au moment de sa requête, ils peuvent être soumis à ce dernier en aval, c'est-à-dire à l'affichage des résultats afin de pondérer ceux-ci à partir des occurrences ou co-occurrences des descripteurs ou en organisant les lots de résultats par rapport aux domaines d'appartenance des descripteurs. Les thésaurus peuvent toutefois être proposés à la recherche s'ils sont présentés sous forme graphique et intuitive, à la manière d'une arborescence thématique qui se dévoile à mesure que l'utilisateur clique sur un descripteur, ou encore à la manière d'un nuage de mots-clés interactif. Une ergonomie simple et soignée ne donne pas l'impression à l'utilisateur profane de manipuler un thésaurus pourtant complexe.

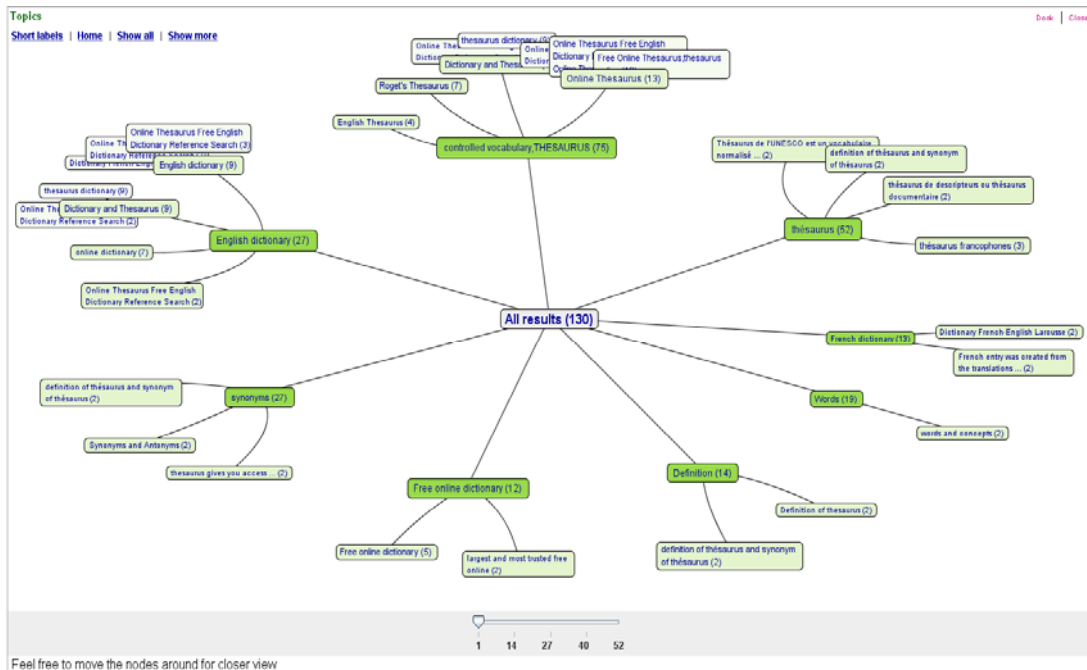
Il est donc nécessaire de s'intéresser à l'usage des thésaurus dans les interfaces de recherche. Sylvie Dalbin a présenté quelques approches au cours de la Journée d'Etude ADBS du 20 septembre 2007 (42, DALBIN) afin d'allier simplicité de l'interface et pertinence des résultats :

- L'exploitation du thésaurus comme dictionnaire pour permettre à l'utilisateur de sélectionner les termes de sa requête, ce qui s'avère utile s'il ne connaît pas la terminologie du domaine concerné.
- L'exploitation du thésaurus sous forme de listes de synonymes, l'utilisateur pouvant ainsi retrouver le descripteur associé au terme qu'il a envisagé en langage naturel.
- L'exploitation en *back office* du thésaurus sous forme d'un dictionnaire de synonymes, le moteur de recherche traduisant le mot saisi par le descripteur synonyme qui apparaît dans les notices.
- L'exploitation du thésaurus pour relancer, reformuler ou étendre une recherche : le moteur de recherche propose, avec la liste des résultats, une liste des descripteurs synonymes ou « proches » de la requête afin que l'utilisateur puisse effectuer un rebond. Nous pouvons citer en exemple le moteur de recherche « AllPlus » qui propose une liste de mots-clés ainsi qu'une représentation graphique des mots-clés associés, le «*Topic Graph* ».

Suite à la requête « Thésaurus » sur le moteur de recherche « Allplus », est affichée la liste des résultats sur la partie centrale de l'écran, les termes associés (« topics ») sur la partie gauche, les médias associés sur la partie droite (vidéos, images, blogs, « news » etc.) :



La représentation cartographique des termes associés est accessible en cliquant sur l'onglet « topic graph » situé sur le bloc gauche de l'écran (encadré en rouge sur l'écran précédent) :



Par conséquent les thésaurus doivent donc évoluer vers des thésaurus à la recherche, n'étant non plus des outils d'indexation et de recherche pour des spécialistes, mais aussi des outils d'aide à la recherche pour un public profane. L'idée directrice est de rendre le vocabulaire efficace à la recherche et de ce fait facilement lisible et accessible pour que l'utilisateur soit naturellement guidé dans sa recherche d'information.

Cependant malgré des progrès indéniables depuis ces dernières années, force est de constater que l'ergonomie des interfaces de recherche dans des ressources indexées à l'aide d'un thésaurus demeure encore assez traditionnelle avec une prise en compte très limitée des pratiques et des difficultés réelles des utilisateurs (44, DALBIN). Les documentalistes et les utilisateurs finaux doivent pouvoir explorer le thésaurus avec leurs propres mots sans s'y perdre, le vocabulaire contrôlé devant se positionner par rapport au moteur de recherche et non par opposition à celui-ci.

### **1.3 Le thésaurus face à la mutualisation et à l'agrégation d'informations en ligne : vers des thésaurus interopérables**

#### **1.3.1 La nécessité d'une interopérabilité sémantique dans le cadre d'une recherche fédérée**

Le problème de l'intégration de sources hétérogènes de données en une interface commune ne consiste pas seulement à résoudre le problème du protocole d'échanges et d'homogénéisation des formats documentaires, il consiste aussi à exprimer des relations entre les données des sources intégrées. De ce fait la recherche fédérée pose le problème du bon mot-clé quand il n'existe pas de thésaurus transversal permettant l'interrogation pertinente de l'ensemble des répertoires de ressources. Il est vain de donner accès à plusieurs sources de données en une interface unique s'il n'existe pas d'interopérabilité sémantique permettant de surpasser l'hétérogénéité sémantique des informations traitées par les diverses applications. Par conséquent le problème posé par la recherche multi-sources est la pertinence des résultats trouvés dans la mesure où différents répertoires de ressources peuvent utiliser différents langages d'indexation (différents thésaurus). Un même concept est susceptible d'être représenté par des descripteurs hétérogènes selon le thésaurus utilisé, et donc des documents pertinents peuvent demeurer invisibles à l'utilisateur selon le descripteur sélectionné.

La notion d'interopérabilité est une notion relativement émergente en France, sa définition la plus précise est donnée par le Grand Dictionnaire terminologique<sup>4</sup> : « *Capacité que possèdent des systèmes informatiques hétérogènes à fonctionner conjointement, grâce à l'utilisation de langages et de protocoles communs, et à donner l'accès à leurs ressources de façon réciproque* ». Appliqué à la terminologie ce terme peut donner naissance à un nouveau terme : l'interopérabilité terminologique que nous pourrions définir comme suit : « *Capacité que possèdent plusieurs langages contrôlés hétérogènes à fonctionner conjointement de façon à donner l'accès, via une interface de recherche unique, à leurs ressources de façon réciproque* ».

### **1.3.2 Exemples de mise en place de langages contrôlés communs au sein de projets pluri-institutionnels**

#### **1.3.2.1 La Banque de Données Santé Publique (BDSP)**

La Banque de Données Santé Publique (BDSP) est un service d'information en ligne fondé sur une couverture multidisciplinaire de la santé publique au début des années 1990. Elle est totalement gratuite depuis 2006. La BDSP est le fruit d'un travail collaboratif dans le cadre d'un réseau de coopération documentaire constitué d'une quarantaine d'organismes. Elle propose des modalités d'échange et de partage du travail, ainsi que l'exploitation de services d'information conçus, développés et alimentés en commun.

Parmi les différents outils et services en ligne figure un élément clé, le « Thésaurus Santé Publique » (TSP) créé en 1992, sa troisième version (TSP 3) datant de 2001. Le TSP peut être gratuitement et librement utilisé, dans le cadre de services informationnels, par des centres de documentation ou des bibliothèques travaillant dans le secteur sanitaire et social. En 2007 est sortie sa quatrième version qui voit l'adjonction de 870 descripteurs et de 856 nouvelles relations de synonymies (51, VIGNON).

La particularité du thésaurus de la BDSP, outil de travail fédérateur, est que tout nouvel acteur adhérent peut garder son propre vocabulaire pour des raisons historiques, pour des raisons de spécificités de sa couverture thématique ou pour ne pas perturber des investissements avec implications terminologiques dans d'autres projets documentaires. Cependant, les 350.000 références bibliographiques (en 2007) constituant la base documentaire de la BDSP sont toutes indexées et interrogeables sur le champ « mot-clé »

---

<sup>4</sup> Office québécois de la langue française < [www.oqlf.gouv.qc.ca/ressources/gdt.html](http://www.oqlf.gouv.qc.ca/ressources/gdt.html) >

avec les descripteurs du TSP ainsi qu'en langage Pascal (INIST-CNRS) alors qu'en amont elles sont livrées par plusieurs membres utilisant divers vocabulaires. La pertinence des recherches est alors rendue possible par l'existence d'une table de correspondances entre chaque terme du vocabulaire spécifique et les descripteurs BDSP. Grâce à des programmes informatiques conçus et développés en interne, une comparaison des vocabulaires et des propositions de correspondances entre chaque descripteur sont réalisées. Toutefois cette automatisation affiche des limites et nécessite donc une relecture humaine, des corrections, des ajouts et une validation par un professionnel. Ce type de passerelle instaure une interopérabilité des langages qui permet d'encourager la mutualisation des ressources entre institutions sanitaires et sociales dans la mesure où la livraison et l'importation de ressources documentaires sont facilitées sans qu'il y ait à changer de langage d'indexation.

### **1.3.2.2 Le projet OTAREN du SCEREN-CNDP**

Pour rechercher une information sur l'éducation, il est nécessaire de consulter des centres de ressources proposant comme langage contrôlé le thésaurus « Motbis » (utilisé par les CDI des collèges et lycées, par le réseau du SCEREN-CNDP) ou le répertoire RAMEAU (BNF, bibliothèques universitaires) voire les deux à la fois (IUFM). Il est donc devenu nécessaire d'élaborer une interface de recherche permettant d'accéder harmonieusement aux deux langages afin que l'utilisateur puisse trouver l'information documentaire pertinente quel que soit son langage d'indexation. Par ailleurs l'utilisateur peut élargir sa recherche sur le thésaurus multilingue de l'ONU (UNBIS), celui de l'UNESCO, le « Thésaurus Européen de l'Education » (TEE) etc. Or il n'est pas viable de se limiter à des procédures de réindexation physique entre les thésaurus et donc à une compatibilité descendante/appauvrissante. C'est pourquoi il semble nécessaire d'envisager un système d'équivalences conceptuelles entre les représentations de concepts présentes dans chacun de ces langages contrôlés, sans passer par une phase de modification de l'indexation initiale. Ces interfaces doivent demeurer des outils de recherche et non d'indexation tout en étant élaborées en cohérence avec les langages d'indexation pré-existants qui demeurent les véritables outils d'indexation propres à chaque répertoire de ressources.

Les équipes du SCEREN-CNDP ont donc développé (entre 2004 et 2006) et lancé (en 2006) le prototype OTAREN (Outil Thématique d'Aide à la Recherche pour l'Education Nationale) permettant l'exploration d'une famille de langages de recherche susceptibles d'être proposés par le CNDP à diverses catégories d'utilisateurs. Fin avril 2006 OTAREN comportait 7989 notions déclarées équivalentes à 59825 autres termes parmi lesquels 29804 « constructions » RAMEAU (46, FEYLER).

### 1.3.3 Les initiatives anglo-saxonnes pour faire évoluer les normes de gestion des thésaurus vers plus d'interopérabilité

En 2005, la publication successive de nouvelles versions de normes anglo-saxonnes ANSI/NISO Z39.19 et BS 8723 ainsi que l'apparition de nouvelles formes de langages documentaires (ontologies, taxonomies) ont amené l'ADBS à poser la question de l'évolution de la norme française AFNOR Z 47-100 et de la norme internationale ISO 2788. En effet l'essor des nouvelles technologies du Web a conduit à une évolution de la présentation des informations à l'écran dans la mesure où le public est plus large. Les thésaurus, qui pour la plupart datent des années 1980, font débat car les nouveaux profils de chercheurs d'information ne sont pas spécialistes de la recherche et ne maîtrisent donc pas la recherche par thésaurus. D'autres outils tels que les moteurs de recherche en texte intégral, les ontologies et les taxonomies constituent la préoccupation majeure des professionnels de l'information-documentation. Dès 2002, dans les colloques internationaux, certains professionnels anglo-saxons échangent à propos de l'« archaïsme » des thésaurus existants. En 2003 la nécessité de restructuration de la norme américaine de gestion de thésaurus apparaît comme une évidence. En juillet 2005, l'organisme de normalisation américain, la *National Information Standards Organization* (NISO) adopte la norme ANSI/NISO Z39.19, tandis qu'en 2005/2006 la *British Standard Institution* (BSI) adopte la norme BS 8723. Cette nouvelle version permet d'élargir le périmètre d'application de la norme de gestion de thésaurus à d'autres langages contrôlés tels que les taxonomies. La norme américaine Z39.19 et la norme britannique BS 8723 prennent donc en compte les thésaurus mais aussi les *synonym rings* (anneaux de synonymes), les taxonomies et les lexiques car étant des langages qui partagent avec les thésaurus un même système de choix des termes et le même type de réflexion préalable à la mise en place de relations dans le but d'assurer aux utilisateurs une recherche plus facile (40, CHICHEREAU).

La norme ANSI/NISO Z39.19 et la norme BS 8723 se caractérisent par :

- La possibilité de donner plus de détails en ce qui concerne le **choix des termes**, par exemple pour l'emploi de qualificatifs (huit cas) ou pour les notes d'application (notions d'historique ou de note réciproque).
- La possibilité de donner plus de détails sur les **types de concepts** (sept au total) : les choses et leurs parties physiques, les matériaux, les activités ou processus, les événements, propriétés, disciplines, unités de mesure et noms propres.

- La gestion des **formes singulier/pluriel** : distinction entre ce que l'on peut dénombrer (*how many ?*) qui amène le pluriel, et ce qui ne peut pas l'être (*how much ?*) qui amène le singulier ; en France, d'après l'AFNOR le choix est simple car tous les descripteurs doivent être au singulier sauf les mots qui sont naturellement au pluriel dans la langue.
- Un traitement plus détaillé des **noms propres** : la norme les considère comme un type de concept et souligne donc l'importance de leur contrôle.
- La **décomposition des termes composés** : un terme composé est constitué d'un cœur (ou d'une tête) et d'un modificateur (ou distinctif) :
  - si le cœur est une partie ou une propriété du distinctif, le terme est à décomposer (ex : « moteur d'avion » doit être décomposé en « moteur » et « avion » car le cœur « moteur » est une partie du distinctif « avion »)
  - si les termes comprennent une action transitive, il est nécessaire de décomposer (ex : « moulage de l'acier » en « moulage » et « acier ») de même si l'action est intransitive (ex : « migration d'oiseau » en « migration » et « oiseau »)
  - le terme ne doit pas être décomposé : s'il s'agit du vocabulaire usuel (ex : « traitement de données »), si la décomposition entraîne une perte de sens ou une ambiguïté, si le terme composé contient un ou plusieurs noms propres, si le distinctif a perdu son sens d'origine ou s'il suggère une ressemblance (ex : « structure en arbre »), si le distinctif ne définit pas une sous-classe ou si le terme principal est un mot vide (ex : « théorie »)
- **Elargissement, enrichissement et formalisation des types de relation** pris en compte, ce qui marque d'une part la volonté d'ajouter des précisions terminologiques et des explications plus rigoureuses et d'autre part de prendre en compte les spécificités de chaque domaine d'application d'un langage contrôlé ainsi que les contraintes terminologiques incontournables :
  - Relations d'équivalence : synonymie, variantes lexicales, synonymies et renvoi au terme générique. Le « *synonym ring* » décrit quant à lui une relation d'équivalence entre plusieurs termes sans désigner de terme préférentiel, ce qui peut s'avérer utile en recherche

- Relations hiérarchiques plus détaillées afin de mieux faire comprendre la nature de ce type de relation
  - Relations associatives plus nombreuses et dont les caractéristiques sont mieux précisées : action/produit, action/propriété etc.
  - « Relations sur mesure » : nouveau type de relation qui donne la possibilité de créer des types de relation autres que les précédents afin de mieux répondre à un besoin particulier
- Une **présentation à l'écran plus précise** selon les besoins de l'utilisateur :
    - Présentation des relations
    - Présentation hiérarchique
    - Présentation graphique
    - Utilisation d'un thésaurus pour la recherche
- L'**interopérabilité** permettant à deux ou plusieurs systèmes d'échanger entre eux des données. Il s'agit de la partie la plus novatrice, celle qui pose le problème occulté dans les normes antérieures. L'idée est de permettre aux usagers de mener une recherche dans des bases de données différentes en utilisant le même langage. Les normes éditées par l'ISO et par l'AFNOR dans les années 1980 ne traitent pas de l'interopérabilité en tant que telle dans la mesure où à l'époque l'environnement technologique imposait plus de limitations que d'ouverture. Ces nouvelles normes préconisent d'établir des ponts entre les concepts, les termes et les relations qui appartiennent à deux vocabulaires différents. Toutefois le problème des formats d'échange et des protocoles n'est pas abordé.

Le potentiel des langages contrôlés n'est pas négligeable à une époque où la production, les besoins et les flux d'information croissent de plus en plus vite. Il est donc nécessaire d'organiser et présenter une structure du langage qui soit immédiatement compréhensible. L'évolution des normes relatives aux thésaurus donne aux professionnels de l'information-documentation des clés pour faire évoluer leurs pratiques de mise à disposition de l'information.



### 1.3.4 La constitution de thésaurus interopérables exploitables sur le Web : la norme ISO 25964

Cinq pays ont voté pour travailler sur la norme ISO 25964 en 2007 : les Etats-Unis, la France, le Royaume-Uni, l'Espagne et l'Allemagne. La norme ISO 25964 est une norme qui positionne clairement les thésaurus par rapport à d'autres types de vocabulaires contrôlés, qui est adaptée aux évolutions des systèmes d'information et qui de ce fait est en phase avec le Web sémantique et ses standards car permettant :

- Une typologie plus fine des relations
- Une gestion d'attributs, qui peuvent être personnalisés
- Un plus grand formalisme dans la description des objets métier
- Des concepts devenant des ressources utilisables dans une optique de base de connaissance et de Web sémantique

L'interopérabilité est désormais possible entre deux ou plusieurs thésaurus ou entre thésaurus et d'autres vocabulaires : classifications, référentiels de classements, taxonomies, vedettes matière, ontologies (correspondance avec OWL), terminologies, listes d'autorité, anneaux de synonymes.

Cette nouvelle norme est centrée sur le thésaurus et s'intitule « Thesauri and interoperability with other vocabularies ». Son but est la gestion électronique de thésaurus pour retrouver tout type de ressource. Cette norme est en cours d'élaboration, elle se constitue de deux parties :

- La partie 1 a été publiée en juillet 2011 et concerne le thésaurus pour la recherche d'information
- La partie 2 sera publiée en 2012 et concernera l'interopérabilité avec des vocabulaires contrôlés

Cette norme va remplacer les normes internationales ISO 2788 : 1986 (thésaurus monolingues) et ISO 5964 : 1985 (thésaurus multilingues) et les normes françaises NF Z47-100 : 1981 (thésaurus monolingues) et NF Z47-101 : 1990 (thésaurus multilingues).

Cette nouvelle norme entend aboutir à un effort de formalisation et de sémantisation des thésaurus en structurant ceux-ci selon un modèle de données métier sous forme d'un schéma de classes UML (*Unified Modeling Language*, un langage graphique de modélisation des données) constitué :

- D'un ensemble de classes d'objets (« concepts », « termes », « notes » etc.)

- D'un ensemble de propriétés pour chaque classe (« identifiant », « libellé », « langue », « dates », ajout possible d' « attributs », etc.)
- D'un ensemble de classes de relations (équivalence, hiérarchique, associative, top level) disposant d'un attribut « rôle » pour préciser la nature de la relation (« cause »/ « effet », « relation hiérarchique partitive » etc.)

L'ISO 25964 prend en compte le multilinguisme :

- Chaque terme est associé à une langue
- Les termes exprimant un même concept sont en relation d'équivalence cross-lingue pouvant être explicite (ex : **fr** *aide économique* **en** *economic aid* / **es** *ayuda económica*). Ces équivalences peuvent être plus graduées : équivalence exacte, équivalence inexacte ou proche, équivalence partielle, non-équivalence.

L'interopérabilité des langages documentaires repose sur un « mapping » des concepts avec divers degrés d'équivalence, comme l'illustre ce tableau d'Hélène Zysman et Hélène Rabault (50, RABAULT , ZYSMAN, p.12) :

**Interopérabilité : mapping des concepts**

|    |  |   |
|----|--|---|
| EQ | Equivalence (sans préciser la nature)          | Eau EQ Water<br>Commande EQ Control   |
| =  | Equivalence exacte                             | Eau =EQ Water<br>Téléphone mobile =EQ téléphone cellulaire<br>Eau =EQ Eau                                       |
| ~  | Equivalence inexacte (sans préciser la nature) | Commande ~ EQ Control<br>Horticulture ~ EQ Jardinage<br>Education secondaire ~EQ Education + Ecoles secondaires |
| BM | Equivalence vers un concept plus vaste         | Roi BM monarque   |
| NM | Equivalence vers un concept moins vaste        | monarque NM rois  |
| RM | Autre équivalence                              | Education RM Puniton  |

Journée BNF / AFNOR - norme thesaurus ISO 25964 27/05/2011 12

La norme ISO 25964 est en grande partie compatible avec le format SKOS (*Simple Knowledge Organization System*). Le SKOS, dont la dernière version date de 2009, est un langage qui a été développé par le W3C et permettant l'exploitation, sur le web, de

systemes d'organisation de concepts (SOC) tels que les thesaurus, les classifications etc. Le schéma SKOS definit des classes et des proprietés en indiquant que l'objet principal géré est le concept représenté par des termes avec au moins un terme préférentiel (un par langue), des termes alternatifs et des termes équivalents. Le langage SKOS prévoit également un ensemble de notes. Les concepts sont hiérarchisés et liés entre eux par des relations associatives et peuvent être regroupés dans des schémas de concepts. SKOS prévoit en outre des éléments de « mapping », d'interopérabilité, pour mettre en correspondance un terme du thesaurus avec un élément extérieur appartenant à un autre vocabulaire contrôlé. Par ailleurs SKOS prévoit des relations sémantiques précises. Plus qu'un concurrent à la nouvelle norme ISO 25964, il permet son implémentation.

## 2 Des outils sémantiques au service d'une recherche fédérée efficace : les apports du Web de données pour simplifier l'accès à des ressources de plus en plus nombreuses et hétérogènes

---

### 2.1 D'un Web de pages à un Web de données : rendre les ressources en ligne compréhensibles par des machines

#### 2.1.1 Une terminologie à définir

Tim Berners-Lee donne une définition du Web sémantique : "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."<sup>5</sup>

Michel Gagnon (8, GAGNON) distingue :

- le **Web actuel** constitué de documents, basé essentiellement sur du HTML, permettant une recherche par mot-clé et utilisable par l'humain
- le **Web sémantique** constitué de connaissances, basé sur du XML et du RDF(S), permettant une recherche par concepts et utilisable par la machine

La notion de « Web sémantique » se décline en de nombreux termes qui, s'ils semblent proches, représentent des approches techniques différentes d'un même sujet. Ainsi, « Web sémantique », « Web 3.0 » et « Web de données » ne représentent pas la même chose. La notion de « Web 3.0 » est contestée dans la mesure où elle suppose que le Web sémantique ne serait qu'une mise à jour du Web (à la manière d'un logiciel), ce que Tim Berners-Lee a réfuté (7, DECLERCQ). Ce dernier préfère la notion de « Web de données » (« Web of data ») supposant qu'il est nécessaire de quitter le Web actuel, le Web de documents qui traite des pages, pour se tourner vers un Web qui gère des données. Ainsi les moteurs de recherche plutôt que de transmettre comme résultats des pages où se trouvent les termes demandés indépendamment de leur localisation sur la page, vont soumettre des documents dans lesquels les termes recherchés sont liés. D'ailleurs Emmanuelle Bermes (3, BERMES)

---

<sup>5</sup> Berners-Lee, Tim ; Hendler, James ; Lassila, Ora. The semantic Web, *Scientific American*, Mai 2001.

distingue la notion de « Web sémantique », concernant les applications des technologies sémantiques propres à une institution, de celle de « Web de données » se définissant comme un concept de décloisonnement et de partage des technologies sémantiques à l'échelle du Web.

Le Web sémantique n'a pas vocation à remplacer le Web actuel mais doit plutôt être considéré comme une brique supplémentaire permettant de structurer l'information contenue sur le web (11, PIERRE). L'objectif de cette nouvelle brique demeure l'absence de structuration des milliards de documents sur le Web qui empêche l'utilisateur d'avoir accès à la plupart de ceux-ci. C'est pourquoi le Web sémantique est souvent défini comme une infrastructure juxtaposant au Web actuel des documents structurés par des langages (pour exprimer la connaissance, pour décrire les relations entre les connaissances, pour décrire les conditions d'utilisation, les garanties) et des dispositifs permettant de trouver des ressources (14, TRONCY). Le Web sémantique repose sur des langages et des outils permettant de représenter formellement de la connaissance afin que la machine ait accès au sens des documents.

### **2.1.2 Du Web sémantique au Web de données : définition des enjeux**

Le but du Web sémantique est aussi vieux que le Web : créer un medium universel pour l'échange de données. Rendre les données sur le Web compréhensibles par les machines est devenu une priorité pour plusieurs institutions, entreprises et communautés d'individus. Le Web ne peut atteindre son véritable potentiel que s'il devient un endroit où les données peuvent être partagées et comprises par des automates et des humains.

Le problème majeur de l'Internet est la dispersion des données. Le Web sémantique permet aux ordinateurs de « comprendre » le sens des pages web en suivant les liens hypertextes des documents web vers des ontologies spécifiques. Poursuivant le développement de la toile, le Web sémantique doit être considéré dans une logique d'utilisation du Web par les machines et non exclusivement par l'homme. Ainsi toute donnée doit être accompagnée d'informations permettant à des machines de l'exploiter, d'où la nécessité d'ajouter des métadonnées qui sont des données contenant de la sémantique car expliquant la donnée à laquelle elles se réfèrent et le contexte de celle-ci. Les métadonnées peuvent être :

- externes à l'objet, comme par exemple les « *Topic Maps* » étant des représentations cartographiques de métadonnées sous la forme d'un réseau sémantique

- internes à l'objet, ce que propose le langage RDF

Quant aux bases de données en ligne, le problème est qu'elles ne sont pas visibles par les moteurs de recherche, elles ne se sont interrogeables que par des individus et leurs ressources demeurent donc invisibles sur un Web caché. Des protocoles d'interrogation multi-sources (Z39.50, SRU/SRW, OAI-PMH) demeurent des protocoles spécifiques à la recherche sur le Web caché et non sur le Web dans sa globalité. Ainsi le Web sémantique doit permettre l'interrogation simultanée de données structurées (dans les bases de données) et non structurées (contenus des pages web) en ajoutant une « couche » structurée de métadonnées à l'ensemble de ces données, « couche » devant être compréhensible aussi bien par les machines que par les humains. Ainsi le Web sémantique tend à « décloisonner les silos d'information » afin de permettre à l'utilisateur ainsi qu'aux machines de mener une requête sur l'ensemble des silos (3, BERMES). Le Web de données entend créer un espace documentaire dans lequel les ressources du Web sont décrites et reliées entre elles : leurs liens sont « typés », c'est-à-dire qu'est renseignée la nature de la relation entre les ressources (ex : « Victor Hugo » *est l'auteur de* « Les Misérables »).

Le Web sémantique devrait permettre, selon Véronique Mesguich (10, MESGUICH), non seulement aux humains d'interagir entre eux, mais offrirait également une meilleure organisation de l'accès aux énormes gisements d'informations numérisées disponibles en modélisant le cheminement du cerveau humain et en donnant aux bases de données la possibilité d'échanges à travers des applications interopérables. L'enjeu du Web sémantique est de faciliter et d'améliorer la recherche d'information grâce à une représentation du sens du contenu des ressources manipulables par des programmes. Car si l'efficacité des moteurs de recherche web a pu être améliorée, le type d'indexation pratiqué par ces outils demeure en revanche fondé sur une approche typographique et non linguistique de la langue. En effet les moteurs web demeurent incapables de désambiguïser un terme ou de prendre en compte les polysémies ou les homonymies, les métadonnées incorporées dans l'en-tête des pages web ne sont donc pas toujours fiables. Or avec le Web sémantique, les documents contenus sur le Web seraient davantage structurés et formalisés et leur contenu serait décrit par un balisage adéquat. Le Web sémantique permet ainsi d'ajouter une couche de métadonnées aux informations existantes.

Les technologies du Web sémantique sont de plus en plus appliquées à un large spectre d'applications au sein desquelles une connaissance de domaine est modélisée et formalisée afin de servir de support à des traitements très diversifiés (des « raisonnements ») effectués par la machine. Ces représentations peuvent être rendues compréhensibles par l'homme pour assurer un couplage optimal entre raisonnement humain (cognitif) et mécanique

(sémantique formelle) afin de confier à l'homme et à la machine des tâches complémentaires (6, CHARLET, LEGER). Les machines deviennent donc capables d'interpréter et d'exploiter le contenu des données web car les documents web ont un sens, ils sont compréhensibles par un ordinateur et par conséquent un ordinateur est capable de comprendre le sens d'un document. Ainsi Sébastien Declercq définit le Web sémantique comme « un web offrant la possibilité d'exploiter l'entièreté des relations sémantiques entre des termes usuels » (7, DECLERQ, p. 2).

Ainsi le Web sémantique peut être considéré comme un dispositif permettant de structurer les informations du Web de façon à les rendre manipulables et « compréhensibles » par des agents logiciels. L'objectif de ces agents est de faciliter l'utilisation des informations et services du Web en libérant l'internaute d'une partie de sa charge cognitive, et de donner l'impression d'un système homogène et cohérent en mobilisant automatiquement et de manière transparente les multiples ressources, sites et services nécessaires à l'accomplissement d'une tâche. Le Web sémantique peut, dans l'idéal, proposer diverses fonctionnalités intéressantes selon Brunon Menon (63, MENON) :

- la recherche généraliste avec un moteur de recherche sémantique doté de capacités de « raisonnement » en s'appuyant sur la description formalisée et la mise en relation des différentes sources d'information pour traiter « intelligemment » les requêtes et présenter en une seule étape des résultats complets.
- l'exploitation et la combinaison de ressources pour accomplir une tâche spécialisée : des outils dédiés associent dialogue avec des sources hétérogènes, description des préférences des utilisateurs et raisonnement basé sur des connaissances métier pour synthétiser l'information requise.
- l'offre de services web plus complets, avec des outils qui identifient, activent et combinent différents services pour mener à bien des opérations plus ou moins complexes de la vie quotidienne ou professionnelle, comme l'organisation d'un voyage, la souscription d'un contrat d'assurance, etc.
- la navigation sémantique, qui profite de la sémantisation des hyperliens pour orienter l'internaute dans son parcours du réseau.

De ce fait après une période de « tissage » du Web apparaît une période de « démêlage » d'un réseau dense et complexe. Ainsi les principes du Web sémantique sont que :

- Tout objet du Web possède une étiquette
- Toute étiquette est lisible par les agents logiciels et par les êtres humains
- Toute étiquette représente fidèlement l'objet
- Toute étiquette est disposée dans un espace commun de lisibilité, espace explorable virtuellement par l'ensemble des humains et des agents logiciels

Les maîtres mots du Web sémantique sont donc sémantique, distribution, ouverture et partage. La sémantique permet une plus grande précision dans les requêtes et les réponses tandis que l'ouverture permet l'échange et le partage de ressources toujours plus complètes et toujours plus précises. Toutefois intégrer ses ressources au Web de données nécessite de lier celles-ci à des ressources extérieures d'autres acteurs/institutions spécialisés dans le même domaine.

### **2.1.3 Un langage pour les machines : la structuration des métadonnées au sein des couches du Web sémantique**

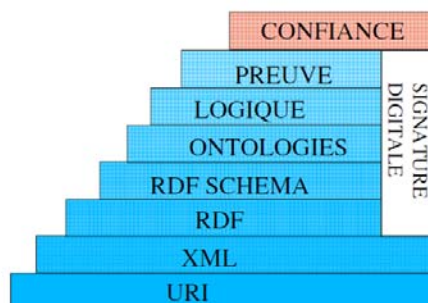
Les métadonnées sont des données sur les données dont la fonction est de qualifier un document ou une information. Cependant l'usage de ces métadonnées dépasse les traditionnelles données bibliographiques ou d'indexation. En effet elles permettent également de gérer le cycle de vie des documents, de produire et diffuser l'information par différents canaux. Or le Web sémantique s'appuie sur ces métadonnées pour augmenter d'une part la qualité des recherches et d'autre part pour offrir de nouveaux services (2, BAUDRY DE VAUX, DALBIN). Néanmoins si les métadonnées sont nécessaires au Web Sémantique, elles doivent néanmoins être encodées et intégrées dans des schémas sémantiques afin de permettre aux ordinateurs de les exploiter de façon efficace.

En effet le Web sémantique est un langage pour les machines (4, BERMES). Comme toute langue elle est composée :

- D'une grammaire : le RDF
- D'un vocabulaire : le RDFS/OWL, Ontologies
- De règles d'écriture : RDF/XML, RDFa
- De moyens de communication : SPARQL



Par ailleurs d'un point de vue technique, il est notable que le Web sémantique est structuré en plusieurs couches, que Michel Gagnon (8, GAGNON) représente comme suit :



### 2.1.3.1 L'URI

L'**URI** (*Universal Resource Identifier*) identifie une entité à laquelle on se réfère sur le Web, mais qui n'est pas nécessairement une ressource disponible sur le Web. Les URI sont des identifiants web dont les exemples le plus connus sont les URL qui identifient la ressource par le moyen d'y accéder (les URL peuvent être pérennes). Cependant les URI, contrairement aux URL, n'ont pas obligation à donner accès au document car demeurent avant tout des identificateurs, à la manière de l'ISBN pour un ouvrage. La non-obligation pour les URI de donner accès à la ressource contribue à la robustesse du Web sémantique car les documents RDF peuvent être consultés hors ligne et parce que le risque de liens cassés disparaît (14, TRONCY). L'URI présente l'intérêt que l'objet référencé n'a pas besoin de se trouver ou d'être décrit dans le document lui-même. L'usage des URI permet le partage des descriptions entre plusieurs documents. Par exemple un ouvrage (défini par son URI) peut être décrit dans un fonds tandis la biographie de son auteur (auteur défini par une autre URI) peut se trouver dans un autre fonds. Le langage RDF constitue donc un graphe de relations entre les informations disséminées, entre leurs URI respectives. Une URI doit être validée par l'IANA, organisme de contrôle des URI car il est nécessaire qu'une ressource ne soit identifiée que par une URI unique.

Une URI peut se construire selon la syntaxe suivante :  
<<http://mondomaine.org/ressource/345>>

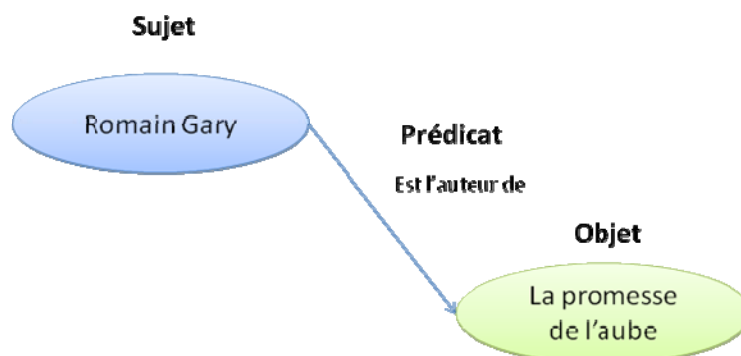
### 2.1.3.2 Le XML

Le **XML** (*eXtended Markup Language*) qui est un langage de balisage extensible, un métalangage qui permet de définir des balises pour des documents. Il s'agit de la norme pour la représentation d'information structurée.

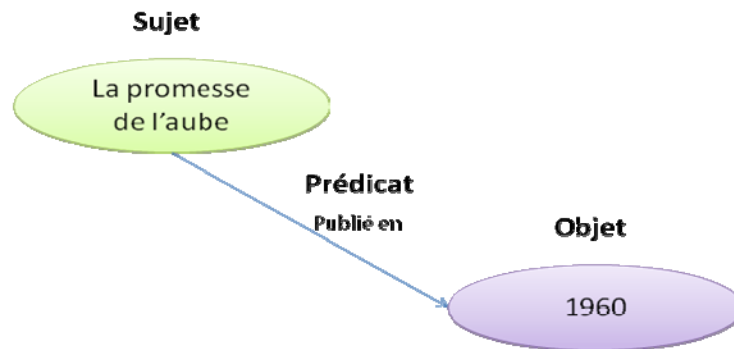
### 2.1.3.3 RDF

**RDF** (*Resource Description Framework*) est un langage recommandé par le W3C (*World Wide Web Consortium*) fondé sur les notions de ressources et de relations entre les ressources. **RDF** est un modèle de graphes censé représenter les diverses ressources du Web et leurs métadonnées afin d'en permettre le traitement automatique. Le modèle RDF favorise l'interopérabilité en annotant les documents ou applications non structurés sur le Web. Ce schéma RDF est représenté par un ensemble d'énoncés (*statements*). Chaque énoncé est un triplet <S, P, O> : S est le sujet, P est le prédicat (le type de relation entre le sujet et l'objet), O est l'objet (une ressource ou une autre donnée liée au sujet en question, la nature de ce lien étant définie par le prédicat). Le **sujet** est toujours une URI : toute « chose » (sujet) sur laquelle on souhaite faire des assertions doit avoir une URI. Les « choses » ont toujours un « type », c'est-à-dire une **classe**. Le **prédicat** est toujours une URI : il permet d'exprimer les **propriétés** des « choses » ou les **relations** des « choses » entre elles. L'**objet** peut être un texte (**littéral**) ou une URI. Les **classes** et les **propriétés** sont déclarées dans des **vocabulaires** pour être réutilisées.

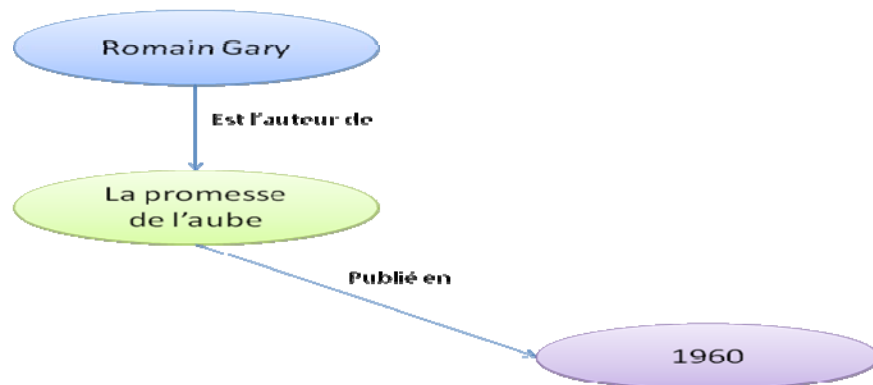
Pour illustrer la logique de construction d'un schéma RDF, nous pouvons prendre l'exemple de l'écrivain Romain Gary qui est l'auteur de *La promesse de l'aube* :



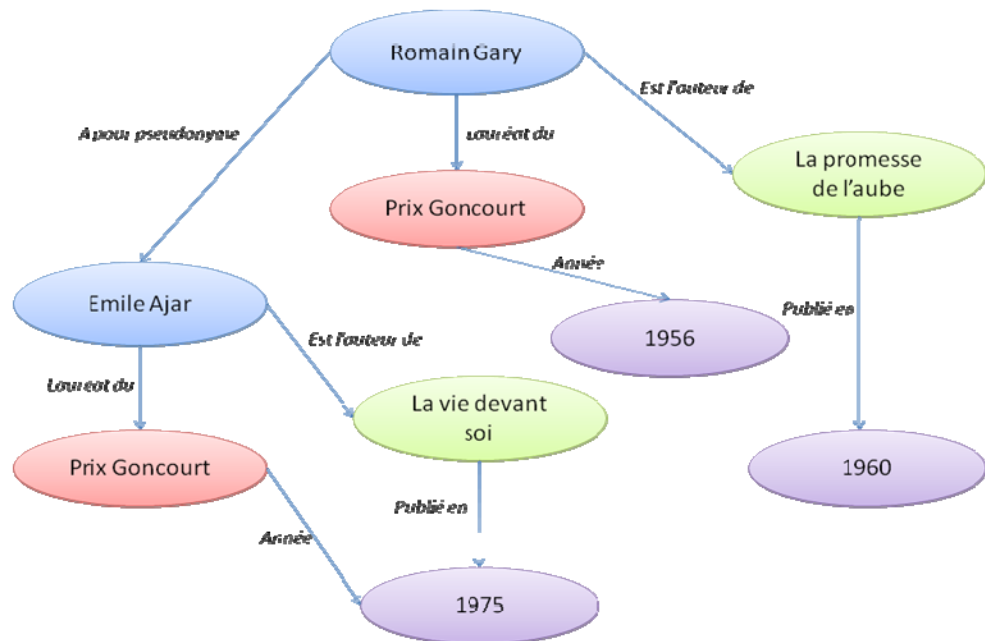
*La promesse de l'aube* est également un sujet en plus d'être un objet lié à Romain Gary :



Ainsi l'assertion « Romain Gary est l'auteur du roman *La promesse de l'aube* publié en 1960 » peut être représentée dans le schéma RDF suivant :



Or Romain Gary, lauréat du Prix Goncourt en 1956, a également écrit sous des noms d'emprunt tels qu'Emile Ajar, récompensé du Prix Goncourt en 1975 pour *La vie devant soi* :



*NB : Les termes employés pour nommer les ressources ne correspondent pas sur les schémas précédents à des URI, tel que cela devrait être le cas pour l'ensemble des « sujets », des « prédicats » et aussi pour certains « objets » (ceux susceptibles de devenir des « sujets » doivent obligatoirement être identifiés par une URI).*

Le RDF est un format très ouvert et malléable car il permet d'ajouter des annotations à propos de n'importe quelle ressource. L'intérêt du RDF est de permettre à une communauté d'utilisateurs de partager les mêmes métadonnées pour des ressources mutualisées. L'un des atouts du RDF étant la souplesse dans la façon de décrire une donnée, qu'il s'agisse d'une définition standardisée au niveau mondial ou bien d'une définition interne à une entreprise.

RDF est donc le modèle qui sous-tend le Web de données. Chaque ressource, identifiée par une URI (nécessaire à l'identification par une machine), est autonome tout en étant liée à d'autres. Chaque sujet est lié à un objet, objet pouvant lui aussi être un sujet lié à un autre objet etc.

#### 2.1.3.4 Les autres couches

Les autres couches structurant le Web sémantique sont : les **ontologies** ; la **logique** qui est le langage permettant d'exprimer des règles de raisonnement, ces règles servant à déduire de nouveaux faits à partir de faits existants ; les **preuves** qui sont une suite d'applications de règles qui permettent de déduire un nouveau fait ; le **niveau de confiance** que l'on attribue aux agents automatiques, ces derniers devant expliquer clairement comment ils arrivent à leurs conclusions (preuves) et garantir la fiabilité et l'origine des informations utilisées (signature digitale).

Grâce à cette structure le Web de données présente, selon Emmanuelle Bermes (3, BERMES), plusieurs avantages :

- **Pas de conversion/ « mapping »** : chacun utilise son propre format de métadonnées, ce qui n'empêche pas de rassembler des triplets <s,p,o> au format RDF
- **Pas de données redondantes** : chacun crée les données dont il a besoin et récupère les informations qui existent déjà
- **Pas de moissonnage/collecte de métadonnées** : les données sont disponibles directement sur le Web
- **Pas de problème de traçabilité** : grâce aux URI il est possible de revenir aux données sources, quelle que soit leur origine
- **Pas de développements spécifiques** : tout repose sur des standards ouverts comme RDF, SPARQL etc...aucun besoin d'apprendre de nouveaux langages et protocoles

## 2.2 Les ontologies comme fondement sémantique du Web de données

### 2.2.1 Une terminologie floue dans la littérature scientifique

Le terme « ontologie », dans sa définition philosophique, se définit comme l'étude de l'être en tant que tel, ses propriétés intrinsèques. Les « ontologies » sont quant à elles considérées comme des objets techniques et opérant dans divers univers tels que l'intelligence artificielle, la linguistique ou l'ingénierie des connaissances (11, PIERRE). Les ontologies sont définies comme une spécification formelle, explicite, d'une conceptualisation partagée. Celles-ci ont été rapprochées des vocabulaires contrôlés dans la mesure où elles

peuvent servir à normaliser un corpus de termes et à expliciter les liens sémantiques qui les lient. Les ontologies constituent un cadre conceptuel ou un mode de représentation dans lequel les thésaurus peuvent être englobés mais ne représentent pas des langages documentaires en elles-mêmes (61, MENON). Or même si elles ne sont pas des langages documentaires, leur processus d'élaboration est proche de celui des thésaurus car nécessitant de normaliser les termes d'un fonds lexical, d'analyser ce fonds sous un angle linguistique, de l'uniformiser sémantiquement et d'établir les relations entre les termes (11, PIERRE).

Le terme « ontologie » est apparu dans son acception informationnelle à la fin des années 1990 dans le domaine de l'ingénierie des connaissances et de l'intelligence artificielle, ce terme désignant selon Jacques Chaumier les artefacts élaborés dans le cadre d'une modélisation conceptuelle apte à jouer un rôle de référentiel conceptuel (55, CHAUMIER). Les travaux sur les ontologies se sont particulièrement développés depuis et ont pris leur essor avec le Web sémantique. Le terme d'ontologie peut se référer à une compréhension partagée d'un domaine d'intérêt qui peut être utilisé comme cadre unificateur pour surpasser les problèmes d'interopérabilité technique entre différents systèmes et d'interopérabilité sémantique entre différents langages.

Selon Jacques Chaumier une ontologie est constituée de :

- une ou plusieurs taxonomies ordonnées en classes et sous-classes composées d'instances représentant des individus ou des objets
- des types d'attributs ou de propriétés pouvant être attachés à ces objets
- des types de relations entre les concepts d'une taxonomie
- des axiomes ou des règles d'inférence permettant de définir les propriétés de ces relations

Selon Catherine Legg c'est plutôt à partir de la fin des années 1950 que l'ontologie commence à investir les sciences informatiques dans la mesure où les bases de données devenaient de plus en plus sophistiquées et ambitieuses et que commencèrent à se poser des problèmes pour les gestionnaires de bases de données (59, LEGG). Par exemple est-ce qu'une entreprise est attachée à une localisation ? (une multinationale pouvant avoir plusieurs localisations et donc plusieurs identités dans un répertoire de ressources, par exemple dans une base de données le constructeur Toyota peut exister sous la forme « Toyota », la maison mère, et sous la forme de ses diverses filiales « Toyota Amérique », « Toyota Europe » etc.) ou une entreprise est-elle considérée comme une entité s'abstrayant d'un cadre géographique ? (la multinationale se définissant alors par une identité

unique quelles que soient ses implantations dans le monde : « Toyota »). Les ontologies susciterent alors l'intérêt de la recherche dans le champ de l'intelligence artificielle.

Une ontologie peut être définie, selon Elin K. Jacob, comme une conceptualisation partielle et simplifiée du monde tel qu'il est supposé exister pour une communauté d'individus, une conceptualisation créée dans un but explicite. Dans l'environnement du Web une ontologie n'est pas seulement une structure conceptuelle mais une structure concrète qui modélise la sémantique d'un domaine dans un langage compréhensible par les machines (58, JACOB).

Bruno Menon reconnaît lui-même le flou terminologique entourant les ontologies. Selon lui le terme « ontologies » (dont la forme au pluriel s'oppose au singulier « ontologie » désignant la discipline philosophique) a été utilisé dans des contextes tels que l'intelligence artificielle, l'ingénierie linguistique ou la gestion des connaissances. Ce terme, s'il est de plus en plus fréquemment utilisé dans la littérature scientifique en raison de l'essor des travaux sur le Web sémantique dont il est présenté comme l'un des piliers, demeure flou de par la diversité des définitions lui étant attribuées et apparaît donc polysémique (61, MENON). Ce flou terminologique s'accompagne d'un « flou fonctionnel », c'est-à-dire que les avis divergent sur les fonctions et la nature même d'une ontologie. Par exemple Bruno Menon constate que parfois les ontologies sont considérées comme jouant, dans le cadre du Web sémantique, un rôle similaire à celui des langages documentaires dans les systèmes d'information bibliographiques puisqu'elles servent de cadre de contrôle et de référence pour l'expression des métadonnées et leur interprétation par des applications « intelligentes ». Or plus rarement elles peuvent aussi être considérées comme incluant des langages documentaires et des taxonomies, voire leur servir de modèle pour leur conception et perdre ainsi tout rôle de sémantisation au profit d'un rôle d'organisation.

Cependant Véronique Mesguich apporte une définition synthétique et assez juste des ontologies dans le cadre du Web. En effet elle présente les ontologies comme des systèmes d'organisation des connaissances (SOC) qui correspondent à la fois à un vocabulaire organisé, à la manière d'un thésaurus, mais également à la formalisation explicite des relations créées entre les différents termes du vocabulaire (10, MESGUICH) permettant aux ordinateurs de communiquer aux ordinateurs, contrairement aux thésaurus qui sont des outils permettant une communication d'humain à humain.

## 2.2.2 Ontologies et thésaurus : deux systèmes d'organisation des connaissances distincts

Une analogie peut être faite entre ontologies et thésaurus dans la mesure où la structuration de concepts en réseaux et la normalisation de leurs expressions sont des points communs irréfutables. Toutefois ces deux instruments possèdent leurs spécificités. Si les thésaurus servent de base à la construction d'ontologies, ils demeurent avant tout des outils de médiation documentaire et sont destinés à un usage humain. Quant aux ontologies, elles doivent servir à la représentation de multiples aspects des ressources numériques et sont donc orientées vers un usage par les machines. Bruno Menon considère que : « les normes pour les thésaurus fixent la liste des relations utilisables et la forme des termes, mais laissent une certaine souplesse dans les formats et les présentations utilisés ; pour les ontologies, on a en revanche une normalisation très stricte des formats, mais une grande ouverture dans la définition des relations nécessaires aux applications visées et dans le type de termes qui y figurent » (63, MENON, p.389). Ainsi si l'utilisateur d'un index est humain il est préférable d'utiliser un thésaurus, si l'utilisateur doit être une machine, comme dans le cadre du Web sémantique, les ontologies s'imposent (56, CYROT, PREUSS). La différence entre thésaurus et ontologie est que le premier est limité en ce qui concerne la spécification du type de relation entre les termes (limité à la hiérarchie et à l'association) tandis que la seconde apporte plus de sémantisation sur les liens.

En outre nous pouvons constater qu'il existe une différence entre thésaurus et ontologies concernant le degré de formalisation (ou de conceptualisation). En effet le thésaurus ne doit pas être composé de concepts trop spécifiques afin de permettre l'indexation de documents différents mais traitant de sujets proches avec les mêmes descripteurs afin d'éviter le silence documentaire, mais ces mêmes concepts ne doivent pas non plus être trop génériques pour éviter le bruit documentaire lors de la recherche. Or pour les ontologies l'ambition n'est pas l'indexation de documents mais la récupération d'informations, leur niveau de formalisation dépendant ainsi de l'usage recherché par ses concepteurs et du public visé (57, DESFRICHES-DORIA). Par ailleurs les processus de recherche proposés diffèrent entre les thésaurus et les ontologies. Il n'existe **pas d'analyse sémantique** dans un **thésaurus** car celui-ci repose sur une logique combinatoire de termes, c'est-à-dire de **combinaison syntaxique** de descripteurs pour la recherche, et propose des résultats dans lesquels le(s) terme(s) saisi(s) a (ont) été indexé(s). En revanche la recherche sur les **ontologies** repose sur des calculs logiques à partir d'un terme saisi en **langage naturel** et présente les **résultats liés sémantiquement** qui ne contiennent pas obligatoirement le(s) terme(s) de la requête, il n'y a donc pas de contrainte sur l'utilisation d'un terme plutôt qu'un autre (57, DESFRICHES-DORIA).



Enfin l'une des caractéristiques intéressantes des ontologies demeure leur ouverture, c'est-à-dire que des ontologies peuvent faire référence à des concepts introduits dans d'autres ontologies en définissant la nature des relations entre les ressources de ces différentes ontologies (équivalence, hiérarchie, incompatibilité) (7, DECLERCO). Cette ouverture permet donc une incrémentation constante des ontologies et des liens sémantiques entre les ressources du Web.

## **2.2.3 Les ontologies : un pilier du Web sémantique permettant la compréhension des contenus du Web par les machines**

### **2.2.3.1 Les ontologies pour organiser et donner du sens aux métadonnées**

Selon Bruno Menon le rôle central qui va être confié aux métadonnées dans le Web sémantique sera amplifié par rapport à une approche « catalogage et indexation » dans la mesure où la définition qui en a été proposée est révélatrice de la finalité descriptive des métadonnées : « information associée à une ressource du web, permettant d'en favoriser l'utilisation par un agent humain, du fait de son exploitation par un agent logiciel » (63, MENON, p.389). Les fonctions des métadonnées dans le Web sémantique dépassent les dimensions signalétiques et thématiques et peuvent aussi servir de support à la gestion des droits, au recueil d'annotations diverses telles que des commentaires et recommandations, à la qualification des hyperliens, à la définition de parcours de lecture ou d'assemblage de documents à la carte etc.

Or si les métadonnées structurent le Web sémantique, elles ne sont en elles-mêmes pas suffisantes. En effet les métadonnées sont transparentes pour les humains mais demeurent dénuées de sens pour des machines. Si le XML, langage structuré, remplace progressivement le HTML en tant que langage du Web, ces langages ne se préoccupent que de la structure physique des documents web et de ce fait leurs mots-clés manquent de « relief sémantique » qui pourrait permettre l'interprétation du sens du document par l'ordinateur. De ce fait pour être exploitées automatiquement, les métadonnées doivent être entièrement explicites, c'est-à-dire suivre un modèle et être exprimées dans un vocabulaire clairement et formellement défini grâce aux ontologies qui modélisent les connaissances nécessaires à la description et au traitement d'un ensemble de ressources. Dans ces ontologies sont représentées les valeurs qu'il est possible de donner aux métadonnées et l'interprétation que les systèmes peuvent en faire, c'est-à-dire les concepts d'un domaine, les relations qu'ils entretiennent et la sémantique de ces relations (63, MENON).

Par conséquent des standards sont requis pour encoder et représenter sémantiquement un savoir. Le Web sémantique propose grâce aux ontologies de donner un sens aux contenus ainsi que de créer des liens entre ceux-ci et de laisser la possibilité d'ajouter/liier de nouveaux savoirs (58, JACOB). Les ontologies sont des systèmes de métadonnées et constituent ainsi l'artère centrale de métadonnées du Web sémantique, ontologies dans lesquelles des schémas classificatoires complexes sont créés et enregistrés pour permettre aux agents électroniques d'interpréter intelligemment des données, faire des liens et réaliser des tâches (9, GREENBERG).

Les ontologies jouent donc un rôle fondamental dans le Web sémantique. En effet la structuration des données ne suffit pas à les rendre intelligibles par un ordinateur, les regroupements de données doivent être complétés par du sens pour devenir réellement exploitables. Ce qui explique pourquoi, selon Jacques Chaumier (55, CHAUMIER), les ontologies ont été développées parallèlement au développement des métadonnées.

### **2.2.3.2 Des standards pour générer des ontologies : OWL et SKOS**

#### *2.2.3.2.1 OWL*

Si les ordinateurs « comprennent » les documents, c'est-à-dire en connaissent la nature (vidéo, texte, audio, image etc.) et « savent » comment les traiter, ils n'en comprennent pas le contenu. Or le Web sémantique doit permettre aux ordinateurs d'appréhender le sens d'un document en exploitant les données fournies au format sémantique, l'ordinateur assimilant à une chaîne de caractères une entité numérique compréhensible par lui. Ce format sémantique consiste en la création de liens sémantiques puissants entre les termes, liens dont le caractère pluridirectionnel tisse une toile sémantique nommée « ontologie ». Ainsi une ontologie peut être définie comme un « super thésaurus » offrant la possibilité de multiplier les relations sémantiques à souhait. Car si le RDF propose une représentation relationnelle des connaissances d'un domaine, il ne possède aucune valeur sémantique intrinsèque, ne propose aucune interprétation du contenu. Le modèle RDF doit s'appuyer sur l'utilisation d'éléments appartenant à des ontologies et des schémas de métadonnées qu'il va identifier par une URI afin de conférer une valeur sémantique à une ressource.

Le langage OWL, compatible avec RDF, permet de représenter des ontologies dans le Web sémantique. Il a pour but d'étendre RDF en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour ce faire il fournit un mécanisme permettant de spécifier les classes dont les instances sont des ressources, comme les propriétés. OWL

s'écrit toujours à l'aide de triplets RDF, en définissant la sémantique de nouveaux mots-clés. En outre OWL dispose également d'un grand nombre de constructeurs permettant de contraindre avec plus de finesse la définition :

- des classes en les définissant comme union, intersection, complémentaire d'autres descriptions ou comme l'ensemble d'un certain nombre d'individus.
- des domaines de relations en spécifiant le type de toutes les valeurs ou d'un certain nombre de celles-ci.
- des relations en les déclarant transitives, symétriques ou en spécifiant leur inverse.

Le OWL (ou *Web Ontology Language*), basé sur le RDF, est un langage web proche de la machine, ce qui le rend difficilement compréhensible pour un humain. Le OWL est fondé sur une syntaxe XML et a été promu au rang de standard du Web en 2004 par le W3C. Ce standard permet la rédaction de pages web complètes, ce qui permet une utilisation plus en profondeur des termes repris dans le document. Ce type de métadonnées permet des liens plus en profondeur et plus précis entre les termes d'un document, offrant ainsi la possibilité de faire transparaître sur le Web les relations sémantiques entre les termes. Les métadonnées principales sont les « Class » (proche des « termes génériques » des thésaurus) qui combinées à des « SubClassOf » (proche des « termes spécifiques » des thésaurus) permettent de créer des relations hiérarchiques entre les termes. Par ailleurs il existe des métadonnées qui offrent des possibilités de relations associatives, ce sont les métadonnées « Property » qui permettent de démultiplier les relations sémantiques car étant multipliables à l'infini à l'instar des métadonnées du XML. Par exemple il est possible de créer, si besoin est, des propriétés telles que « *est\_née\_en* », « *a\_eu\_lieu\_en* », « *a\_pour\_pseudonyme* ». Par conséquent la puissance du OWL réside dans la possibilité de créer le nombre de métadonnées souhaitées en passant des plus simples (les « Class ») aux plus complexes telles que « *Property:est\_l\_auteur* ». Ainsi, grâce à ces métadonnées, un terme est renseigné de manière multiple, en augmentant de ce fait sa sémantisation (7, DECLERQ).

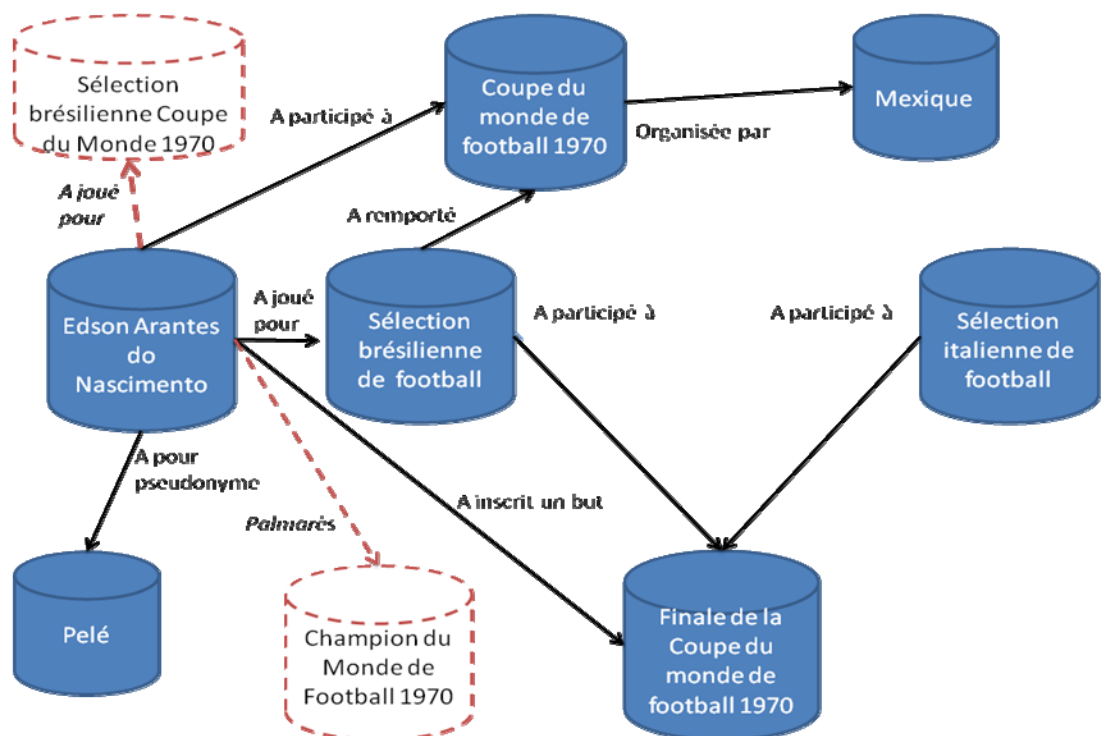
Prenons l'exemple suivant : nous signalons à l'ordinateur que :

- 1) *Pelé a participé à la Coupe du Monde de Football 1970*
- 2) *La sélection brésilienne a remporté la Coupe du Monde de Football 1970*
- 3) *Pelé faisait partie de la sélection brésilienne de football*
- 4) *Le Brésil et l'Italie ont participé à la finale de la Coupe du Monde de football 1970*
- 5) *Pelé a inscrit un but lors de la finale de la Coupe du Monde de Football 1970*
- 6) *La Coupe du Monde de football 1970 s'est déroulée au Mexique*

### 7) Edson Arantes do Nascimento a pour pseudonyme Pelé

L'ordinateur, grâce au standard OWL, « comprend » que : *Le joueur de la sélection brésilienne Edson Arantes do Nascimento, également connu sous le nom de Pelé, a inscrit un but contre l'Italie lors de la finale de la Coupe du Monde de football 1970 au Mexique que le Brésil a remportée, permettant à ce joueur de devenir champion du Monde.*

Graphiquement nous pouvons représenter les assertions précédentes (en bleu) ainsi que les déductions « effectuées » par l'ordinateur (en rouge pointillé) de la manière suivante :



Ainsi OWL est un système puissant d'exploitation des ontologies qui permet justement la création déductive de liens entre les termes. Par ailleurs le OWL rend possible l'automatisation des associations entre différentes ontologies. De ce fait la modélisation d'une nouvelle base de connaissances n'est plus un acte isolé mais partagé.

#### 2.2.3.2.2 SKOS

SKOS (*Simple Knowledge Organization System*) est défini comme un modèle simple par opposition au langage OWL qui est plus à même de représenter des structures sémantiques plus riches mais par conséquent plus complexe à utiliser. SKOS est un modèle proposé par le W3C afin d'exprimer des systèmes d'organisation des connaissances d'une manière qui

s'intègre dans le Web sémantique. SKOS a pour objectif de fournir un modèle et le vocabulaire permettant de représenter la structure et le contenu des schémas conceptuels tels que les thésaurus, les schémas de classification, les glossaires et tous types de vocabulaires contrôlés. Ainsi SKOS offre la possibilité de définir des concepts et de les relier entre eux, un schéma conceptuel décrit en SKOS utilisera des propriétés identifiantes, descriptives, structurantes ou de gestion. Pour cela SKOS définit un certain nombre de classes et de propriétés RDF et peut donc être vu comme une application RDF (14, TRONCY). A l'instar de OWL, SKOS a pour ambition de représenter des connaissances sur le Web de façon à ce que les ordinateurs soient en mesure de « comprendre » leur sens. La différence entre OWL et SKOS est que le premier fournit un modèle logique et orienté objet puissant, rendant ainsi très complexes les ontologies à produire, alors que le second est orienté vers des applications plus modestes et peut satisfaire de nombreuses applications (14, TRONCY).

Selon Michèle Lénart (60, LENART) le formalisme de représentation utilisé par SKOS repose sur les graphes RDF. Le concept constitue le centre du graphe auquel peuvent être attachés en tant que propriétés RDF :

- Des indications portant sur le concept lui-même :
  - Termes préférentiels ou alternatifs, les équivalents dans d'autres langues
  - Termes cachés, utiles pour gérer les variantes correspondant à des fautes d'orthographe courantes
  - La représentation par une image
  
- Les différents types de notes : notes de définition et d'application (*scope note*), exemples, notes historiques, etc.
  
- Les relations sémantiques : la hiérarchie et les associations entre les termes.

Les extensions de SKOS (*SKOS Extensions*), qui complètent le noyau de SKOS (*SKOS Core*) comprenant les classes et les propriétés de base, proposent une représentation plus fine des relations (tout/partie ; classe/instance) et précisent certains attributs de concepts (ex : une note historique).

## 2.2.4 Les ontologies et le Web : un retour à la médiation des professionnels de l'information-documentation

L'essor des technologies Internet et les mutations informationnelles qui en découlent n'ont pas, comme le souligne Brunon Menon (61, MENON), suscité de nouveaux langages documentaires. L'évolution du World Wide Web et de ses avatars (intranets, extranets, places du marché électroniques) ont engendré des outils dont les structures et les fonctions s'apparentent à celles des langages documentaires classiques. Par exemple les sites du Web marchand et les intranets d'entreprise s'accompagnent de la création de cadres de présentation et de découverte des ressources, les taxonomies, tandis que parmi les fonctionnalités du Web 2.0 se retrouve la possibilité pour les utilisateurs d'indexer un document ou une page avec leurs propres mots-clés (*folksonomie* : la taxonomie des utilisateurs/compagnons).

Par conséquent les professionnels de l'information-documentation possèdent les savoirs que les ingénieurs informatiques, les entrepreneurs et d'autres recherchent pour appréhender ce qu'est le Web sémantique. Des points communs existent entre les professionnels de l'information-documentation qui créent des taxonomies et les ingénieurs informatiques qui créent des ontologies, ontologies et taxonomies étant ainsi souvent considérées comme des synonymes : les ingénieurs informatiques qualifient d'« ontologies » les hiérarchies de vocabulaires structurés déployant les relations existant entre les termes alors que les professionnels de l'information-documentation les qualifient de « taxonomies ». Cependant une grande partie des concepteurs de sites web n'est pas spécialiste en matière de linguistique et de sémantique, ce qui ne permet pas d'envisager une révolution dans la conception de pages web. D'où le rôle prédominant des professionnels de l'information-documentation dans la mesure où, ayant déjà acquis les connaissances en gestion de métadonnées et en automatisation d'outils descriptifs, ils sont les plus susceptibles de jouer le rôle d'intermédiaire entre linguistes et informaticiens. Les documentalistes, selon Sébastien Declercq, sont aptes à aider la mise en application d'un système sémantique (7, DECLERCQ).

La démultiplication des outils sémantiques (par exemple le moteur de recherche d'ontologies « *Swoogle* ») rend la recherche d'information plus performante mais aussi plus complexe, d'où l'importance du rôle des documentalistes qui doivent mettre leur savoir-faire au service des utilisateurs afin de les faire accéder à l'information pertinente grâce au développement d'ontologies facilité par le standard OWL.

Derrière la « simplicité » du Web sémantique du point de vue de l'utilisateur, il y a en *back office* un travail conséquent de création d'ontologies, d'annotations, de mise en relation et d'organisation qui demeure en grande partie réalisé par des humains, les professionnels de l'information-documentation. Chronologiquement Asuncion Valderrama (64, VALDERRAMA) distingue trois modèles de système d'information au cours desquels les professionnels de l'information-documentation ont joué des rôles différents :

- Le modèle classique suppose une forte intermédiation entre l'information et l'utilisateur car il existe un fort besoin d'expertise professionnelle pour diverses tâches (indexation, informatique, services de références etc.).
- Le modèle moderne ne repose sur aucune forme d'intermédiation car l'utilisateur accède à l'information via une interface unique (exemple des moteurs de recherche grand public sur le Web tels que Google).
- Le modèle post-moderne introduit par le Web sémantique et dans lequel il existe une médiation cachée, en arrière-plan, car le Web sémantique réintroduit le besoin d'une expertise de type documentaire à travers les pratiques d'annotation et de gestion de métadonnées.

Asuncion Valderrama considère que ces trois modèles coexistent et nécessitent des compétences professionnelles différentes, même si le modèle post-moderne ressemble au modèle classique de par le rôle de médium joué par le professionnel de l'information-documentation (64, VALDERRAMA). C'est pourquoi elle reprend la formule de Bruno Menon en qualifiant le Web sémantique de « vin vieux dans des outres neuves » dans la mesure où le rôle du documentaliste « post-moderne » n'a pas véritablement changé par rapport à celui du documentaliste « classique », la différence reposant sur la visibilité de son rôle, explicite pour le premier, implicite pour le second mais tout autant fondamental pour permettre à l'utilisateur de trouver l'information pertinente. En effet il est nécessaire pour constituer des ontologies de structurer et de gérer des ressources, de créer et d'exploiter des métadonnées. Si le modèle moderne, issu de la généralisation d'Internet et de l'émergence du Web social, avait effacé l'influence du professionnel de l'information documentation, le Web sémantique lui permet de mettre une nouvelle fois en valeur son savoir-faire tout en donnant l'illusion à l'utilisateur qu'il demeure l'unique acteur de sa recherche.

*Nous avons observé que les outils et les technologies du Web sémantique constituaient non pas une menace mais une authentique opportunité pour les langages documentaires, et en particulier les thésaurus, de jouer un rôle incontournable dans l'accès à l'information dans la mesure où, couplés avec des outils sémantiques en constant progrès, ils permettent de structurer des systèmes d'organisation des connaissances souples, ouverts et partagés qui donnent du sens aux contenus du Web et par extension à toute interface permettant à l'utilisateur de mener une recherche sur Internet. Les portails de recherche fédérée ne font pas exception et doivent pleinement profiter d'une part de l'efficacité incomparable des thésaurus en matière de pertinence des résultats de recherche, si tant est qu'un effort ait été fait sur la démocratisation de l'usage des thésaurus orientés utilisateur, et d'autre part ne pas négliger les apports des technologies sémantiques dans la mesure où celles-ci garantissent l'interopérabilité sémantique entre différents langages d'indexation. A la lumière de ces observations il convient désormais de définir les préconisations que nous apportons dans le cadre du futur portail de recherche fédérée du Musée National du Sport.*



## Troisième partie

Mettre en place un portail de recherche fédérée au sein du Musée National du Sport : mettre en valeur des ressources hétérogènes via la mise en place d'une interface d'interrogation unique

Projet lancé en 1963 et physiquement inauguré (au Parc des Princes) en 1988, le Musée National du Sport est consacré aux pratiques sportives et à leur histoire. Il conserve, diffuse et valorise l'une des plus importantes collections au monde composée de plus de 100 000 pièces dont plus de 2000 accessoires de sports parmi les plus anciens, 17300 affiches et 3200 dessins.

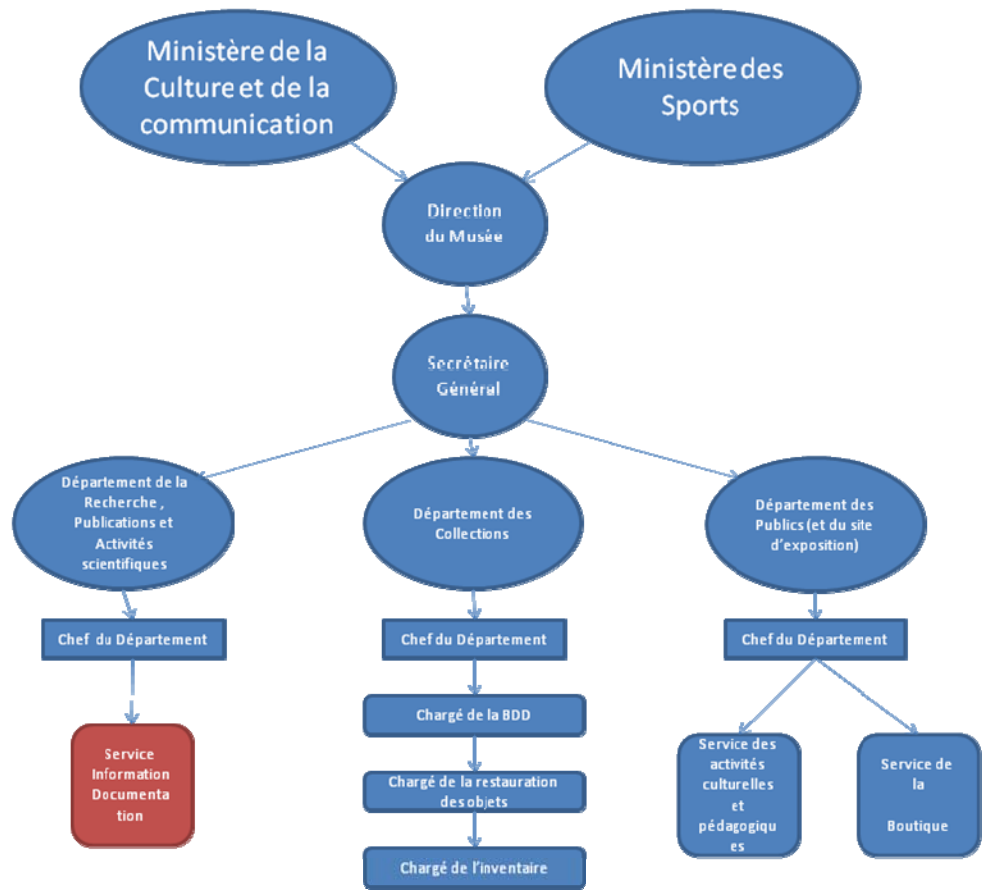
Dans le cadre de l'ouverture du nouveau Musée National du Sport à Nice prévue pour l'année 2013 (les collections étant exposées à ce jour au 93, avenue de France-75013 Paris), le service information-documentation de cette institution souhaite mettre en place un portail en ligne permettant une recherche fédérée sur l'ensemble des bases de données du Musée :

- Micromusée qui gère les notices du fonds « Collections »
- Armadillo qui gère le fonds iconographique (images animées et images fixes)
- Cadic Intégrale qui gère le fonds documentaire (ouvrages et périodiques) et le fonds « Archives »

Ce portail viendrait se « greffer » au site web du Musée National du Sport et permettrait de :

- mettre en valeur ses divers fonds qui seraient interrogeables et consultables via une interface unique et intuitive, ce qui est essentiel dans la mesure où le public ciblé est le grand public, peu au fait des méthodes de recherche « avancées » sur le Web et donc demandeur d'une expérience simplifiée.
- Mutualiser les ressources du Musée avec celles d'autres institutions ayant le sport pour thématique (INSEP, Musée Olympique etc.).
- Proposer une charte graphique qui sorte des canons de la bibliothéconomie : les ressources du Musée National du Sport étant composées d'une part non négligeable d'images, il serait pertinent de proposer un rendu visuel mettant en valeur celles-ci.

Par ailleurs il apparaît nécessaire d'apporter des précisions sur l'organisation du Musée National du Sport et la position du Service Information Documentation au sein de celui-ci afin de mieux saisir les contraintes organisationnelles à prendre en compte dans la conduite d'un projet pour cette structure. Nous avons jugé pertinent de présenter ce contexte sous la forme de l'organigramme suivant :





# 1 Les contraintes techniques de mise en place du portail : réfléchir à l'interrogeabilité pertinente de nos différentes bases de données

---

## 1.1 Rendre les répertoires de ressources interrogeables en ligne

En effet dans un premier temps, si l'on envisage une recherche fédérée, cela implique que nos trois bases de données soient interrogeables en ligne, à la manière d'un OPAC web, ou exportables dans un entrepôt OAI qui alimentera un index interrogeable en ligne. Après nos entretiens téléphoniques avec les prestataires de nos bases de données, il en ressort les observations suivantes :

- Le logiciel Armadillo version 9.90 (société Armadillo) est interrogeable en ligne et peut donc faire l'objet d'un moissonnage OAI ou d'une interrogation selon un protocole Z39.50.
- Le logiciel Cadic Intégrale version 5.4 (société Cadic) est également interrogeable en ligne selon les modes d'interrogation cités précédemment. Toutefois, suite à notre entretien avec l'un des membres du Service de documentation de la Cité de l'Immigration, le logiciel Cadic Intégrale a posé de nombreux problèmes quant au moissonnage OAI mis en place pour le portail lancé en 2009, ce qui nous incitera à demander un test d'export XML vers un entrepôt OAI si nous venions à adopter ce protocole pour notre portail.
- Le logiciel MicroMusée (société Mobydoc), en l'état actuel des choses, n'est interrogeable qu'en local. Néanmoins il suffirait, selon le service d'assistance technique de Mobydoc, de mettre à jour notre répertoire de ressources et d'installer la dernière version du module OPAC Web afin de rendre notre base de données Micromusée interrogeable en ligne. Or il s'avère que la mise en place de l'OPAC Web est au point mort depuis 2009 suite à des désaccords quant à la ventilation des données à l'affichage, ce qui rend une interrogation selon le protocole Z39.50 et le protocole OAI-PMH impossible. Depuis, les demandes de démonstrations d'exports XML de la part du SID (Service Information Documentation) du Musée National du Sport sont restées sans réponse, empêchant ainsi la perspective de constituer un

entrepôt OAI. Il conviendra donc de rétablir la communication avec la société Mobydoc afin de lancer l'OPAC Web ou de réaliser un export XML des notices Micromusée et donc de pouvoir lancer le projet de portail.

## 1.2 Les concordances à établir entre nos bases de données

Rendre nos bases de données consultables en ligne ne suffit pas à mettre en place une réelle recherche fédérée dans la mesure où les bases de données possèdent des champs d'indexation différents (couvrant des besoins différents) et des procédures d'indexation différentes. Il est en effet nécessaire de réfléchir à la concordance qui peut exister entre les champs, c'est-à-dire définir les champs des différentes bases de données contenant des valeurs communes, et entre les valeurs, c'est-à-dire les différentes valeurs employées pour un même concept.

### 1.2.1 La concordance des champs

La réflexion sur les critères de recherche à proposer pour notre éventuel module de recherche avancée nous a déjà amené à définir la concordance des champs entre nos diverses bases de données. C'est pourquoi nous employons ici le tableau qui sera présent dans le chapitre consacré à la recherche avancée.

| Type de champ                  | Champ concerné sur Cadic | Champ concerné sur Armadillo | Champ concerné sur Micromusée  |
|--------------------------------|--------------------------|------------------------------|--|
| Titre                          | Titre                    | Titre                        | Désignation  |
| Auteur/responsable/créateur    | Auteur                   | Auteur                       | Création→Auteur  |
| Date                           | Date de parution         | Date d'exécution             | Date de création ainsi que l'ensemble des champs de type « date »  |
| Discipline sportive concernée  | Genre                    | Discipline                   | Indexation→Sujet/Thème   |
| Personnalité (sportive ou non) | Personnalités            | Personne représentée         | Indexation→Personne/Collectivité   |
| Événement sportif concerné     | Événements               | Événement                    | Indexation→Événement   |
| Type de document               | Type de document         | Type de document             | Désignation (ce champ donne aussi bien des informations sur le créateur de l'œuvre que sur le support de celle-ci) |

## **1.2.2 La concordance des valeurs : la mise en place de thésaurus transversaux**

Ces thésaurus doivent permettre de surmonter une difficulté majeure : l'absence de normalisation dans l'indexation des champs, et ce que ce soit entre les champs des différentes bases ou au sein même de ces bases. En effet malgré les chartes de saisie qui ont été rédigées pour Cadic et Armadillo nous déplorons l'absence de normalisation qui fait encourir le risque de rendre des notices « invisibles » à certains utilisateurs. Si l'on prend comme exemple l'événement « Jeux Olympiques d'été de Tokyo de 1964 », l'on s'aperçoit en consultant les index des différentes bases de données que celui-ci a été saisi de manières différentes :

- Jeux Olympiques d'été de Tokyo, 1964
- Tokyo, Jeux Olympiques, 1964
- Jeux Olympiques de Tokyo, 1964
- Jeux Olympiques 1964

Or si le mot-clé que l'utilisateur a sélectionné dans l'arborescence thématique ne correspond pas avec certains termes synonymes présents dans les notices susceptibles de l'intéresser au sein des bases de données du Musée National du Sport (par ex : « Jeux Olympiques d'été de Tokyo, 1964 » ; « Tokyo, Jeux Olympiques, 1964 » ; « Jeux Olympiques de Tokyo, 1964 ») la requête aboutira à du silence. Il en sera de même pour la recherche simple et la recherche avancée car si le terme saisi par l'utilisateur ne correspond pas aux divers termes synonymes employés dans les bases de données, les résultats ne prendront pas en compte certaines notices qui auraient pu s'avérer pertinentes.

Pour remédier à ce problème, il existe deux solutions.

### **1.2.2.1 Normaliser les procédures d'indexation**

Normaliser les procédures d'indexation permettrait d'uniformiser la saisie. Un terme aura donc la même chaîne de caractères sur toutes les notices quelle que soit la base d'origine. Cependant un problème subsisterait : les futures notices seraient normalisées, mais il resterait un lourd travail de normalisation des notices existantes qui entraînerait une charge de travail supplémentaire pour les personnels du Musée National du Sport (risques de

réticences, risque de lenteur d'un tel projet dans la mesure où les personnels doivent poursuivre leurs propres activités et donc retard sur le lancement du portail).

### 1.2.2.2 Mettre en place un moteur de recherche sémantique

Mettre en place un moteur de recherche sémantique s'avère judicieux dans le cadre d'un projet de portail donnant accès à des ressources hétérogènes. En effet le moteur de recherche ne se limite plus à la reconnaissance des chaînes de caractères mais à la reconnaissance de concepts et donc de tous les termes qui « entourent » ce concept. Pour ce faire un moteur de recherche sémantique pourrait s'appuyer, dans notre contexte, sur un thésaurus transversal. Ainsi un mot-clé sélectionné par l'utilisateur doit correspondre à un descripteur synonyme dans les différentes bases de données. Par exemple si l'utilisateur sélectionne « rugby » alors que le descripteur « ovalie » est présent sur certaines notices, la transversalité doit permettre de traduire « rugby » par « ovalie » dans la requête.

Cela suppose l'élaboration d'une terminologie avec des outils automatiques de traitement linguistique développés notamment par la société Mondeca (qui élabore des thésaurus à partir d'éléments textuels qui seront analysés et croisés avec des thésaurus existants) spécialisée dans le Web sémantique, la gestion des thésaurus, taxonomies, terminologies et bases de connaissance multilingues. Voir aussi le moteur de recherche sémantique *Intuition* développé par la société Sinequa ou encore les moteurs *Lingway*, *Spirit...*

Cette transversalité possède des avantages non négligeables :

- Ne rend plus indispensable la normalisation des procédures d'indexation, ce qui se traduit par la suppression d'une charge de travail qui s'avèrerait chronophage. En effet une fois les thésaurus construits pour chacun des ensembles de champs communs, la sélection d'un terme dans la liste de choix permettra à l'utilisateur de voir s'afficher l'ensemble des notices présentant des descripteurs synonymes dans les différentes bases de données.
- Evite tout risque de silence : si l'utilisateur sélectionne « Jeux Olympiques d'été de Tokyo, 1964 », les notices présentant dans le champ *Événement* les termes « Jeux Olympiques d'été de Tokyo, 1964 » ; « Tokyo, Jeux Olympiques, 1964 » ; « Jeux Olympiques de Tokyo, 1964 » ; « JO de Tokyo » ; « JO 1964 » seront présentes parmi les résultats.

Nous présentons dans les tableaux ci-après une ébauche de thésaurus transversaux pour les typologies de champs définis dans le cadre de la concordance entre les champs des



différentes bases de données. Il ne s'agit là que de suppositions pouvant servir de base à la réflexion nécessaire à la mise en place de thésaurus transversaux. Ceux-ci permettraient la reconnaissance de termes mal orthographiés ou incomplets (comme le permettent la reconnaissance des chaînes de caractères et les troncatures) et des termes synonymes (sauf pour le thésaurus « type de document » qui consiste en un rassemblement de plusieurs termes en un nombre restreints de typologies).

Thésaurus « Les pratiques sportives »

| Valeur normalisée | Valeurs associées dans les bases de données   | Valeurs potentiellement associables     |
|-------------------|---|---|
| Boxe thaïlandaise | - <i>Boxe thaïlandaise</i> (Cadic, Micromusée)<br><br>- <i>Muay Thai</i> (Armadillo, Micromusée)  | - <i>Boxe thai</i><br><br>- <i>Thai</i> |
| Tennis de table   | - <i>Tennis de table</i> (Armadillo, Cadic, Micromusée)<br><br>- <i>Ping-pong</i> (Armadillo, Cadic)<br><br>- <i>Ping pong</i> (Micromusée)                           | - <i>Pongisme</i>                       |
| Boxe Française    | - <i>Savate</i> (Armadillo, Cadic, Micromusée)<br><br>- <i>Boxe française</i> (Armadillo, Cadic, Micromusée)<br><br>- <i>Savate</i> <i>Boxe française</i> (Armadillo) | - <i>Boxe pieds poings</i>              |

Thésaurus « Les sportifs »

| Valeur normalisée | Valeurs associées dans les bases de données   | Valeurs potentiellement associables |
|-------------------|---|-------------------------------------|
| Pelé              | - <i>Pelé</i> (Micromusée, Armadillo, Cadic)<br><br>- <i>Arantès do Nascimento, Edson</i> (Micromusée)<br><br>- <i>Edson Arantes do Nascimento dit Pelé</i> (Armadillo) | - <i>Roi Pelé</i>                   |
| Ali, Mohammed     | - <i>Ali, Muhammad</i> (Micromusée, Armadillo)<br><br>- <i>Clay, Cassius Marcellus</i>  | - <i>The greatest</i>               |

|         |   |               |
|---------|---|---------------|
|         | <i>Junior</i> (Micromusée)<br>- <i>Ali, Mohammed</i> (Armadillo, Cadic)<br>- <i>Clay, Cassius</i> (Armadillo)<br>- <i>Ali, Mohammed, Clay Cassius</i> (Cadic) |               |
| Ronaldo | - <i>Ronaldo</i> (Micromusée)<br>- <i>Ronaldo Luis Nazario de Lima</i> (Armadillo)  | Non déterminé |

Thésaurus « Les événements sportifs »

| Valeur normalisée                          | Valeurs associées dans les bases de données  | Valeurs potentiellement associables  |
|--|--|--|
| Jeux Olympiques d'été de Berlin, 1936      | - <i>Berlin 1936</i> (Armadillo, Micromusée)<br>- <i>Jeux Olympiques d'été 1936</i> (Armadillo)<br>- <i>Jeux Olympiques Berlin 1936</i> (Armadillo)<br>- <i>Jeux Olympique (sic) de Berlin, 1936</i> (Armadillo)<br>- <i>Jeux Olympique berlin (sic) 1936</i> (Armadillo)<br>- <i>Jeux Olympiques 1936</i> (Cadic) | - <i>JO 1936</i><br>- <i>JO Berlin</i><br>- <i>JO Berlin 1936</i><br>- <i>JO Allemagne Nazie</i><br>- <i>JO IIIe Reich</i>   |
| Coupe du Monde de Football de la FIFA 1986 | - <i>Coupe du Monde de Football, 1986</i> (Armadillo)<br>- <i>Coupe du Monde de Football de la FIFA, 1986</i> (Armadillo)<br>- <i>Coupe du Monde 1986</i> (Cadic)  | - <i>Mondial 1986</i><br>- <i>Mundial 1986</i><br>- <i>Mexique 1986</i><br>- <i>Mexico 1986</i><br>- <i>Mondial Mexique 1986</i><br>- <i>Coupe du Monde Mexique 1986</i> |
| Roland-Garros                              | - <i>Roland Garros</i> (Micromusée, Cadic)   | - <i>Grand Chelem France</i><br>- <i>Tournoi du Grand Chelem</i>   |

|  |   |   |
|--|---|---|
|  | - <i>Internationaux de France</i> (Micromusée, Armadillo, Cadic)<br><br>- <i>Roland-Garros</i> (Cadic)<br><br>- <i>Tournoi de Roland Garros</i> (Armadillo) | <i>France</i><br><br>- <i>Grand Chelem terre battue</i> |
|--|---|---|

Thésaurus « Type de document »

| <b>Valeur normalisée</b> | <b>Valeurs associées</b>  |
|--------------------------|---|
| Images fixes             | <i>Photographie</i> (Micromusée)<br><br><i>Image</i> (Micromusée)<br><br><i>Estampe</i> (Micromusée)<br><br><i>Photo d'identité</i> (Armadillo)<br><br><i>Poster</i> (Armadillo)<br><br><i>Carte postale</i> (Armadillo)                          |
| Audiovisuel              | <i>Documentaire</i> (Micromusée)<br><br><i>Films d'actualités</i> (Micromusée)<br><br><i>Vidéogramme</i> (Micromusée)<br><br><i>Bande sonore</i> (Armadillo)<br><br><i>Disque 78 tours</i> (Armadillo)<br><br><i>Film pédagogique</i> (Armadillo) |
| Objets                   | <i>Ballon</i> (Micromusée)<br><br><i>Maillot de bain</i> (Micromusée)<br><br><i>Raquette</i> (Micromusée)<br><br><i>Médaille</i> (Armadillo)<br><br><i>Paire de patins à glace</i> (Armadillo)<br><br><i>Sac</i> (Armadillo)                      |
| Documents                | <i>Livre</i> (Cadic)<br><br><i>Titre de Périodique</i> (Cadic)<br><br><i>Périodique</i> (Armadillo)<br><br><i>Livre</i> (Armadillo)   |

|        |   |
|--------|---|
|        | <i>Livret</i> (Micromusée)<br><i>Périodique</i> (Micromusée)  |
| Autres | A déterminer : dans cette « catégorie » seraient classés les types de documents n'ayant aucune correspondance avec les autres « catégories » (Images fixes, Audiovisuel, Objets, Documents) |

### 1.3 Le dédoublonnage des notices

La mise en place des logiciels Cadic Intégrale et Armadillo a été beaucoup plus récente que la base de données Micromusée. De ce fait des notices Micromusée ont depuis été reprises sur les deux nouvelles bases et par conséquent certains items du Musée National du Sport possèdent à l'heure actuelle une notice sur la base d'origine (Micromusée) et une notice sur l'une des deux nouvelles bases. Or si nous mettons en place une recherche fédérée sur l'ensemble de nos bases, il existe un risque d'affichage de doublons sur l'écran des résultats. De ce fait il conviendra de mettre en place une fonction de dédoublonnage des notices qui privilégierait les notices les plus récentes (celles rédigées sur Cadic Intégrale et sur Armadillo) aux notices les plus anciennes (rédigées sur Micromusée).

Privilégier les notices les plus récentes repose sur l'observation suivante : si des notices présentes sur Micromusée ont été reprises sur Armadillo ou Cadic, c'est parce que les champs et modes de saisie sur ces deux derniers outils sont plus adaptés à la nature de l'item traité (documentation et archives pour Cadic Intégrale, image fixe ou animée pour Armadillo), d'où des notices plus pertinentes.

### 1.4 Le paramétrage des champs affichés à la présentation des notices

Il sera nécessaire de sélectionner les champs de métadonnées qui seront les plus intéressants pour l'utilisateur du portail. En effet, ce dernier n'aura aucun intérêt à avoir des informations de gestion du type « localisation », « identité du récoleur/catalogueur », « date du catalogage », « constat d'état », le « numéro d'inventaire » etc. Il sera davantage intéressé par la mention de responsabilité, la définition du contenu, le contexte de création, le support et le format de l'objet ou encore les items liés dans nos fonds.

Nous dressons donc le tableau suivant qui fait part de nos suggestions à propos des champs à conserver à l'affichage des notices pour chacune des bases concernées.

|  |   |   |
|--|---|---|
| Micromusée   | Armadillo   | Cadic Intégrale   |
| <b>Désignation :</b> Titre de l'item ; donne des informations sur le support des items présents dans Micromusée                | <b>Titre :</b> Titre de l'item  | <b>Titre :</b> Titre de l'item  |
| <b>Auteur :</b> Mention de responsabilité  | <b>Auteur :</b> Mention de responsabilité   | <b>Type de document :</b> donne des informations sur le support de l'item   |
| <b>Date de création :</b> Donne une estimation de la date de création quand celle-ci n'est pas connue                          | <b>Description :</b> description neutre du contenu de l'image fixe ou animée  | <b>Auteur :</b> Mention de responsabilité   |
| <b>Mesures :</b> donne des informations sur le format des items  | <b>Mots-clés :</b> Informations extrasportives telles que la position, l'âge, le sexe du ou des individus représenté(s), la situation etc. → les équipements sportifs représentés font partie des mots-clés | <b>Date de parution :</b> date de parution  |
| <b>Inscription/marques :</b> donne des informations sur les éléments annexes à l'item  | <b>Personne représentée :</b> la personne représentée (s'il s'agit d'un sportif ou d'un non sportif connu)  | <b>Langue :</b> renseigne sur la langue du texte  |
| <b>Description analytique :</b> donne une description physique des items   | <b>Evénement :</b> définit l'événement sportif concerné par l'item  | <b>Résumé :</b> résume en une ligne le contenu de l'item  |
| <b>Sujet/thème :</b> la discipline couverte (mais aussi des informations extrasportives : position, âge, sexe, situation etc.) | <b>Discipline :</b> définit la discipline couverte  | <b>Editeur :</b> renseigne sur l'identité de l'éditeur (utile pour avoir un éventuel contact)   |
| <b>Personne/collectivité :</b> la personne représentée ou la collectivité représentée  | <b>Date d'exécution :</b> Donne une estimation de la date de création quand celle-ci n'est pas connue   | <b>Genre :</b> la discipline couverte   |
| <b>Evénement :</b> définit l'événement sportif concerné par l'item   | <b>Couleur :</b> définit la coloration de l'image fixe ou animée → noir et blanc, en couleur  | <b>Descripteur :</b> catégorise le contenu intellectuel de l'item, quelques mots-clés concernent des événements sportifs (Tour de France, Coupe Davis, Jeux |

|  |   |   |
|--|---|---|
|  |   | Olympiques d'été, Jeux Olympiques d'hiver, Jeux Paralympiques) mais ceux-ci demeurent rares et peu précis (l'année d'organisation ne figure jamais) |
| <b>Objet associé :</b> donne des informations sur des items liés | <b>Négatif positif :</b> informe si l'image fixe ou animée est un négatif ou un positif               | <b>Événements :</b> précise l'événement sportif concerné  |
|  | <b>Orientation :</b> définit l'orientation de l'image (paysage, portrait)                             | <b>Personnalités :</b> indique les personnalités sportives (ou non) concernées  |
|  | <b>Mesure chantier :</b> définit le format du support de l'image fixe ou animée                       |   |
|  | <b>Lieu de création :</b> donne des informations sur le contexte géographique de la création          |   |
|  | <b>Exécutant :</b> donne des informations sur les droits attachés à l'item                            |   |
|  | <b>Communicabilité :</b> donne des informations sur les droits attachés à l'item                      |   |
|  | <b>Utilisation :</b> donne des informations sur les droits attachés à l'item                          |   |
|  | <b>Autres documents photographiques du fonds :</b> donne les cotes des documents photographiques liés |   |

Pour chaque notice présentée, il serait intéressant d'ajouter un champ qui présenterait les notices et les documents numérisés liés (par des mots-clés communs) à la notice consultée.

## 2 Présentation du portail : ce qu'il doit proposer au futur visiteur

---

### 2.1 Le contenu éditorial

#### 2.1.1 La publication d'articles

Même si la fonction essentielle du futur portail sera la recherche, il sera nécessaire de pouvoir publier et gérer des contenus sous forme d'articles. En effet nous prévoyons de publier trois types d'informations :

- Des informations sur les sportifs
- Des informations sur les événements sportifs
- Des informations sur les pratiques sportives

Ces contenus ont avant tout pour but de mettre en valeur les images et les vidéos appartenant au fonds du Musée National du Sport. L'article n'est qu'un « prétexte » pour que l'utilisateur puisse découvrir ce fonds. Cependant la publication de contenus à intervalles réguliers permettra également de donner un aspect vivant au portail ; la principale contrainte dans la publication d'articles sera la sensibilisation des membres du Musée à cette tâche supplémentaire.

#### Les sportifs

Présentation de figures du sport moderne sous forme d'articles illustrés par un diaporama de photos issues de notre fonds ainsi que d'une ou plusieurs vidéos (également issues de notre fonds). Seul le contenu du dernier article en date sera publié, la liste des articles précédents (ayant été archivés) sera accessible via un lien en dessous de l'article consulté. Reste à savoir à quelle fréquence les articles seront publiés (mensuelle ? hebdomadaire ?). A côté de cet article sera proposée une recherche par arborescence thématique concernant les sportifs (mode de recherche explicité dans la partie 2.2).

De ce fait le contenu de cette page pourrait ressembler au « zoning » présenté page suivante :

## Les sportifs

Présentation du contenu de la rubrique *Les sportifs*

### Portrait d'une légende:

#### Marcel Cerdan : Le rêve inachevé

Article présentant la carrière du boxeur Marcel Cerdan

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Photo du fonds  
image sur Marcel  
Cerdan

(cliquer pour  
accéder au  
diaporama → ouverture  
d'une nouvelle  
fenêtre)

Vidéo du fonds sur  
Marcel Cerdan

(cliquer pour la lancer  
dans le  
lecteur → ouverture  
d'une nouvelle fenêtre)

Ecran de recherche dans  
l'arborescence thématique des  
sportifs

Consulter nos anciens articles

← Cliquer sur cette fenêtre pour  
accéder à la liste des anciens  
articles)



## Les événements sportifs

Présentation d'événements sportifs illustrés par un diaporama de photos issues de notre fonds ainsi que par une ou plusieurs vidéos (également issues de notre fonds). Seul le contenu du dernier article en date sera publié, la liste des articles précédents (ayant été archivés) sera accessible via un lien en dessous de l'article consulté. A côté de cet article sera proposée une recherche par arborescence thématique concernant les événements sportifs. De ce fait le contenu de cette page pourrait ressembler au « zoning » suivant :

## Les événements sportifs

Présentation du contenu de la rubrique *Evénements sportifs*

**Retoursur :**

**Berlin 1936**

Article présentant les Jeux Olympiques d'été de Berlin organisés par le IIIe Reich

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Photo du fonds image sur les JO d'été de 1936  
(cliquer pour accéder au diaporama → ouverture d'une nouvelle fenêtre)

Vidéo du fonds sur les JO d'été de 1936  
(cliquer pour la lancer dans le lecteur → ouverture d'une nouvelle fenêtre)

### Ecran de recherche dans l'arborescence thématique des événements sportifs

Consulter nos anciens articles

← Cliquer sur cette fenêtre pour accéder à la liste des anciens articles)

## Les pratiques sportives

Présentation de pratiques sportives (règles, conditions de pratique, matériel, histoire) illustrées par un diaporama de photos issues de notre fonds ainsi que par une ou plusieurs vidéos (également issues de notre fonds). Seul le contenu du dernier article en date sera publié, la liste des articles précédents (ayant été archivés) sera accessible via un lien en dessous de l'article consulté. A côté de cet article sera proposée une recherche par arborescence thématique concernant les pratiques sportives.

De ce fait le contenu de cette page pourrait ressembler au « zoning » suivant :

## Les pratiques sportives

Présentation du contenu de la rubrique *Les pratiques sportives*

**Sport à la Une :**

**Le rugby à XIII**

Article présentant le rugby à XIII

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Photo du fonds image sur le Rugby à XIII  
(cliquer pour accéder au diaporama → ouverture d'une nouvelle fenêtre)

Vidéo du fonds sur le Rugby à XIII  
(cliquer pour la lancer dans le lecteur → ouverture d'une nouvelle fenêtre)

**Ecran de recherche dans l'arborescence thématique des pratiques sportives**

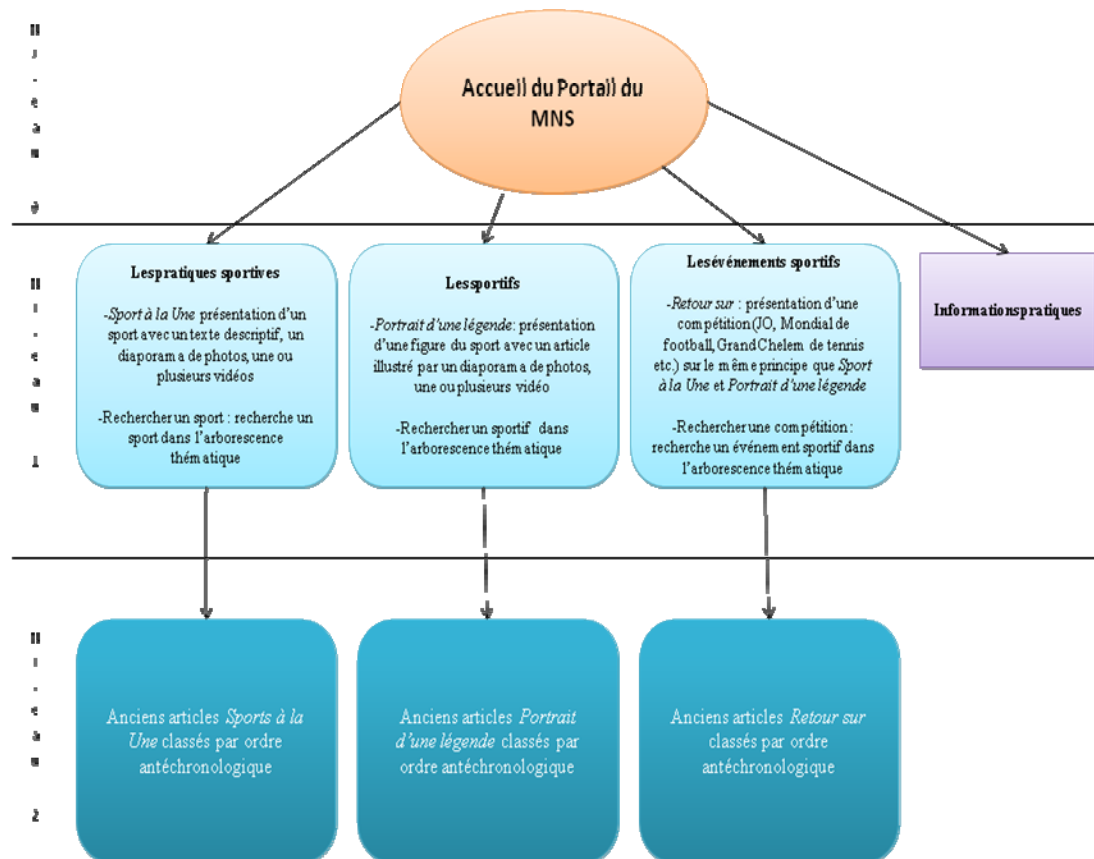
Consulter nos anciens articles

←

Cliquez sur cette fenêtre pour accéder à la liste des anciens articles)

## 2.1.2 L'arborescence éditoriale du portail

Le portail, dans la mesure où il n'a pas vocation à se substituer au site déjà existant, sera donc limité en pages de contenus. Nous proposons ici une ébauche de l'arborescence des rubriques :



Il s'agit donc d'une structure de base du futur portail envisagé, il est à gager que celle-ci évoluera selon les fonctionnalités qu'il sera ou non possible de mettre en place sur celui-ci (expositions virtuelles, nuage de tags, jeux). L'écran de recherche avancée devra quant à lui être accessible sur toutes les pages du site en cliquant sur le lien présent en dessous de la barre de recherche simple (qui sera elle-même présente sur l'ensemble des pages du site).

## 2.1.3 Principes généraux de navigation

Il n'est pas tout à fait possible, à ce stade de notre réflexion, de se faire une idée sur le type de navigation qui sera adopté pour le portail. D'une part il ne s'agit pas d'une priorité, la recherche fédérée sur nos trois bases de données constituant le cœur de notre réflexion.

D'autre part nous n'avons pas encore décidé du contenu définitif de ce portail et donc de l'organisation de celui-ci.

Néanmoins, que les propositions en terme de contenu qui ont été apportées dans ce document venaient ou non à être adoptées, nous pourrions énoncer comme principe essentiel le fait de permettre à l'utilisateur de pouvoir se situer sur le portail. Cela reviendra à mettre en place des fonctionnalités qui aideront l'utilisateur à trouver l'information voulue sans difficulté :

- Un plan de site interactif accessible sur toutes les pages.
- Un fil d'Ariane dynamique et interactif qui permettra à l'utilisateur de savoir où il se trouve dans l'arborescence du site et de pouvoir « revenir sur ses pas » (utile s'il y a plus de deux niveaux d'arborescence).
- Des rubriques toujours accessibles, c'est-à-dire permettant à l'utilisateur de naviguer de rubrique en rubrique sans avoir à revenir en arrière.
- Une surbrillance de la rubrique parcourue qui permet d'identifier la page sur laquelle l'utilisateur se trouve (très utile si l'utilisateur arrive sur le portail par une des pages du site et non par la page d'accueil).
- Un bandeau titre qui demeure le même sur l'ensemble du site pour que l'utilisateur puisse identifier le portail sur lequel il est arrivé (dans le cas où il n'est pas arrivé par la page d'accueil)
- Une distinction chromatique des informations : dans chaque rubrique il serait judicieux d'utiliser une couleur (ou deux) qui diffère du reste du site (permet à l'utilisateur de comprendre d'un coup d'œil à quel type d'information il a affaire).
- Une barre de recherche simple (avec un lien vers l'écran de recherche avancée) accessible sur toutes les pages du portail.

## 2.2 Les modalités de recherche sur le portail du Musée National du Sport

La recherche fédérée sur les trois bases de données du Musée National du Sport pourra être effectuée par l'utilisateur selon trois modalités : la recherche simple, la recherche avancée et l'exploration de l'arborescence thématique.

### 2.2.1 La recherche simple

La recherche simple se fera en plein texte. L'utilisateur saisira une chaîne de caractères dans une barre de recherche clairement identifiable dans l'interface du site. La recherche se fera sur l'ensemble des champs des trois bases de données ainsi que sur l'ensemble des pages du portail.

### 2.2.2 L'exploration de l'arborescence thématique

L'utilisateur pourra naviguer dans une arborescence thématique, c'est-à-dire dans un thésaurus constitué :

- De termes génériques exclusifs constituant le niveau le plus général (large) dans l'arborescence → ex : **Athlétisme** -> Epreuves du Lancer -> Lancer du poids
- De termes génériques ET spécifiques constituant le ou les niveau(x) intermédiaire(s) dans l'arborescence (ils sont un terme spécifique par rapport à un terme générique mais sont aussi un terme générique par rapport à des termes spécifiques) → ex : Athlétisme -> **Epreuves du Lancer** -> Lancer du poids
- De termes spécifiques exclusifs constituant le niveau le plus précis dans l'arborescence → ex : Athlétisme -> Epreuves du Lancer -> **Lancer du poids**

Ce type d'exploration devra être mis en place pour trois types de thématiques :

- Une thématique concernant les sportifs
- Une thématique concernant les événements sportifs
- Une thématique concernant les pratiques sportives

Nous avons sélectionné ces trois thématiques pour plusieurs raisons :

- D'un point de vue intellectuel, nous considérons que ces thématiques seront celles qui seront les plus susceptibles d'intéresser l'utilisateur dans sa démarche de recherche sur nos fonds.
- D'un point de vue technique, ces trois thématiques sont couvertes dans chacune de nos bases de données, d'où la pertinence de la mise en place d'une recherche fédérée.

La mise en place de ce type de recherche nécessite donc la constitution, pour chacune de ces thématiques, d'un thésaurus transversal qui prenne en compte la correspondance entre les différents termes utilisés pour une même expression sur les différentes bases.

A l'écran seront affichés par défaut les seuls termes génériques exclusifs (le premier niveau de l'arborescence) afin de ne pas surcharger l'écran puis, une fois que l'utilisateur aura cliqué sur un « bouton » figurant à gauche du terme voulu, la liste des termes de deuxième niveau pourra s'afficher (avec des boutons également cliquables pour afficher le troisième niveau) selon un procédé de menu déroulant. Pour lancer la recherche avec le mot-clé voulu, il suffira à l'utilisateur de cliquer sur celui-ci. Nous représentons l'affichage à l'écran de la recherche par exploration thématique ainsi qu'une illustration graphique des thésaurus liés à ces thématiques.

#### **2.2.2.1 La thématique « Les pratiques sportives »**

L'exploration de l'arborescence thématique des pratiques sportives pourra s'effectuer sur la page où sera publié le dernier article sur une pratique sportive, comme illustré dans le « zoning » suivant :

## Les pratiques sportives

Présentation du contenu de la rubrique Les pratiques sportives

**Sport à la Une :**

**Le rugby à XIII**

Article présentant le rugby à XIII

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Photo du fonds image sur le Rugby à XIII  
(cliquer pour accéder au diaporama → ouverture d'une nouvelle fenêtre)

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Vidéo du fonds sur le Rugby à XIII  
(cliquer pour la lancer dans le lecteur → ouverture d'une nouvelle fenêtre)

Consulter nos anciens articles

**En savoir plus sur...**

**Athlétisme (TG)**

Lancer (TS/TG)  
 Disque (TS)  
 Javelot (TS) ← Il suffira de cliquer sur un terme pour lancer la recherche sur la discipline concernée  
 Marteau (TS)  
 Poids (TS)

Saut (TS/TG)  
 Saut à la perche (TS)  
 Saut en hauteur (TS)  
 Saut en longueur (TS)  
 Triple saut (TS)

Sprint (TS/TG)  
 100m (TS)  
 200m (TS)  
 400m (TS)

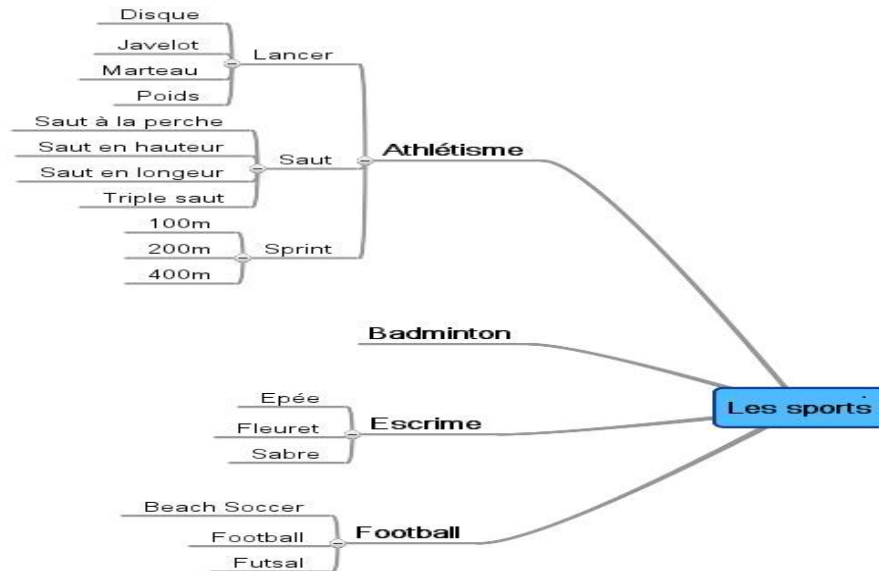
**Badminton (TG)**

**Escrime (TG)**  
 Epée (TS)  
 Fleuret (TS)  
 Sabre (TS)

**Football (TG)**  
 Beach Soccer (TS)  
 Football (TS)  
 Futsal (TS)

Chacun des termes génériques exclusifs (TG) et des termes génériques/spécifiques (TG/TS) sera précédé d'un bouton cliquable faisant apparaître/masquer la liste de ses termes spécifiques (comme un menu déroulant)

Cette arborescence pourrait s'appuyer sur une ébauche de thésaurus pouvant être représentée comme suit :



## 2.2.2.2 La thématique « Les sportifs »

L'exploration de l'arborescence thématique des sportifs pourra s'effectuer sur la page où sera publié le dernier article sur un sportif. L'arborescence reprendra exactement les mots-clés des pratiques sportives, mais comprendra un niveau plus spécifique avec les noms d'athlètes liés aux disciplines. En effet nous considérons que l'utilisateur va rechercher un sportif avec sa discipline de prédilection comme « clé d'entrée » (« si je recherche Marcel Cerdan, je vais forcément trouver son nom en cliquant sur le mot-clé *Boxe* puis *Boxe anglaise* ») comme illustré dans le « zoning » suivant :

### Les sportifs

#### Présentation du contenu de la rubrique Les sportifs

Portrait d'une légende:

Marcel Cerdan : Le rêve inachevé

Article présentant la carrière du boxeur Marcel Cerdan

>Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Photo du fonds image sur Marcel Cerdan  
(cliquer pour accéder au diaporama → ouverture d'une nouvelle fenêtre)

Vidéo du fonds sur Marcel Cerdan  
(cliquer pour la lancer dans le lecteur → ouverture d'une nouvelle fenêtre)

Consulter nos anciens articles

En savoir plus sur...

**Boxe (TG)**  
Boxe anglaise (TG/TS)  
*Carpentier, Georges (TS)*  
*Cerdan, Marcel (TS)* ← Il suffira de cliquer sur un terme pour lancer la recherche sur la discipline concernée  
*Dempsey, Jack (TS)*  
*Louis, Joe (TS)*

Boxe Thaïlandaise (TG/TS)

Boxe française (TG/TS)

**Football (TG)**  
Beach soccer (TG/TS)

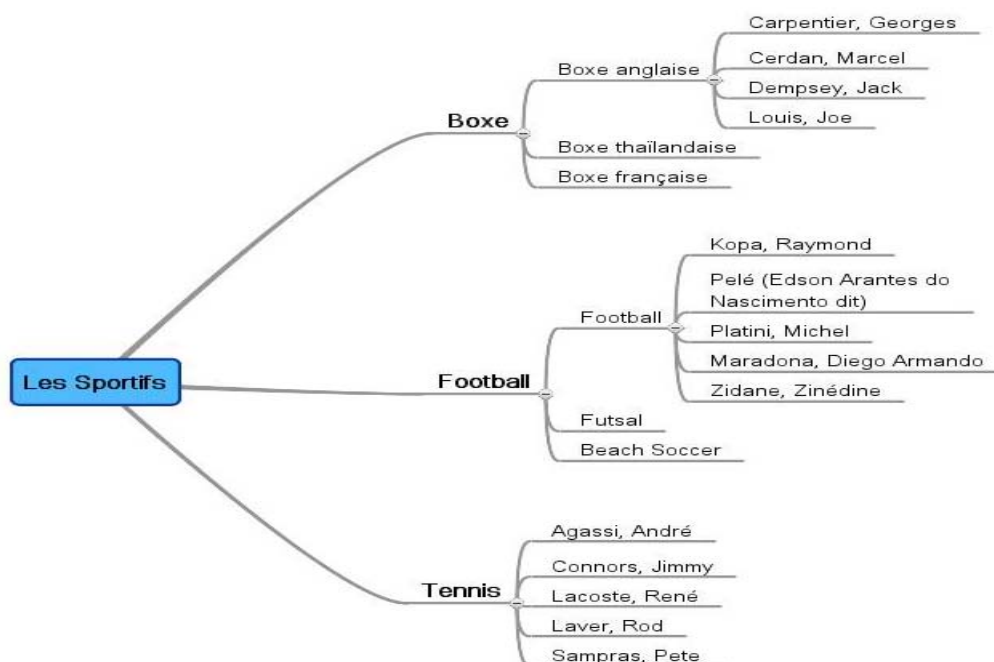
Football (TG/TS)  
*Kopa, Raymond (TS)*  
*Pele (Edson Arantes do Nascimento dit) (TS)*  
*Platini, Michel (TS)*  
*Maradona, Diego Armando (TS)*  
*Zidane, Zinedine (TS)*

Futsal (TG/TS)

Chacun des termes génériques exclusifs (TG) et des termes génériques/spécifiques (TG/TS) sera précédé d'un bouton cliquable faisant apparaître/masquer la liste de ses termes spécifiques (comme un menu déroulant)



Cette arborescence pourrait s'appuyer sur une ébauche de thésaurus pouvant être illustrée comme suit :



### 2.2.2.3 La thématique « Les événements sportifs »

La recherche se fera par discipline au niveau le plus général. Par ailleurs on prendra le soin de ne sélectionner qu'un nombre limité de sports dans un premier temps, dans la mesure où chaque discipline sportive peut être marquée par divers types d'événements. Par exemple le football avec le championnat de France (annuel), la Coupe de France (annuelle), la Ligue des Champions (annuelle), la Coupe de la ligue (annuelle), l'Europa League (annuelle), La Coupe du Monde de la FIFA (tous les quatre ans), le Championnat d'Europe des Nations de l'UEFA (tous les quatre ans), etc.

A un niveau plus spécifique, la recherche se fera par événement une fois que la discipline concernée aura été « cliquée » par l'utilisateur. Un événement sportif qui bénéficie d'une fréquence annuelle (ex : Coupe de France de football, Tour de France cycliste, Roland-Garros etc.) verra ses diverses éditions identifiées par un mot-clé commun (ex : la Coupe de France de Football 1975, la Coupe de France de Football 1984, la Coupe de France de

Football 1995 et la Coupe de France de Football 2003 auront pour mot-clé commun « Coupe de France de football ») afin d'éviter une pléthore de mots-clés.

Les événements biannuels (Championnats du monde d'athlétisme) ou ayant lieu tous les quatre ans (Jeux Olympiques d'été, Jeux Olympiques d'hiver, Coupe du Monde de Football, Coupe du Monde de Rugby) verront chacune de leur édition bénéficier d'un mot-clé propre (ex : les JO d'été de 1936 auront pour mot-clé « Jeux Olympiques d'été de Berlin 1936 »).

L'exploration de l'arborescence thématique des événements sportifs pourra s'effectuer sur la page où sera publié le dernier article sur un événement sportif comme l'illustre le « zoning » suivant :

## Les événements sportifs

### Présentation du contenu de la rubrique Événements sportifs

### Retoursur :

#### Berlin 1936

Article présentant les Jeux Olympiques d'été de Berlin organisés par le IIIe Reich

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit animid est laborum.

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Photo du fonds image sur les JO d'été de 1936  
(cliquer pour accéder au diaporama → ouverture d'une nouvelle fenêtre)

Vidéo du fonds suisses JO d'été de 1936  
(cliquer pour la lancer d'ans le lecteur → ouverture d'une nouvelle fenêtre)

Consulter nos anciens articles

### En savoir plus sur...

**Omnisports (TG)**  
Jeux Olympiques d'été (TG/TS)  
Moscou 1980 (TS)  
Los Angeles 1984 (TS)

Jeux Olympiques d'hiver (TG/TS)  
Grenoble 1968 (TS)  
Sapporo 1972 (TS)

**Football (TG)**  
Coupe du Monde de la FIFA (TG/TS)  
Mexique 1970 (TS)  
RFA 1974 (TS)

Championnat d'Europe des Nations (TG/TS)  
France 1984 (TS)  
RFA 1988 (TS)

Coupe d'Europe des Clubs (TG/TS)  
Championnat de France (TG/TS)  
Coupe de France (TG/TS)  
Coupe de la ligue (France) (TG/TS)

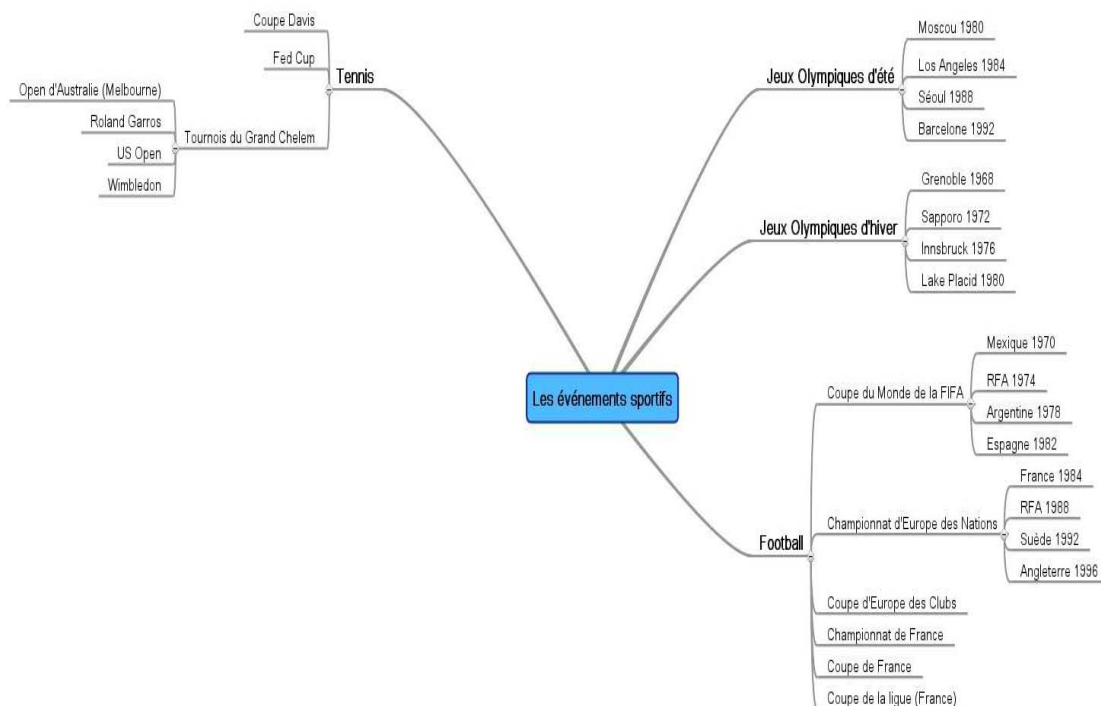
**Tennis (TG)**  
Coupe Davis (TG/TS)  
Fed Cup (TG/TS)

Tournois du Grand Chelem (TG/TS)  
Open d'Australie (Melbourne) (TS)  
Roland Garros (TS)  
US Open (TS)  
Wimbledon (TS)

Il suffira de cliquer sur un terme pour lancer la recherche sur la discipline concernée

Chacun des termes génériques (TG) et des termes génériques/spécifiques (TG/TS) sera précédé d'un bouton cliquable faisant apparaître/masquer la liste de ses termes spécifiques (comme un menu déroulant). Les compétitions annuelles, étant nombreuses, seront identifiées par un descripteur commun

Cette arborescence pourrait s'appuyer sur une ébauche de thésaurus pouvant être illustrée comme suit :



### 2.2.3 La recherche avancée

La recherche avancée devra permettre à l'utilisateur de spécifier un critère de recherche ou d'associer plusieurs critères de recherche afin d'obtenir les résultats les plus précis possibles. Dans le cadre d'une recherche fédérée il sera nécessaire de définir des types de critères communs à nos trois bases de données, en excluant tous les critères concernant les champs de gestion de nos bases de données. Nous pouvons, à ce stade de réflexion, proposer les critères suivants :

| Nom du critère sur le portail  | Champ concerné par la requête sur Cadic | Champ concerné par la requête sur Armadillo | Champ concerné par la requête sur Micromusée   |
|--------------------------------|---|---|--|
| Titre                          | Titre                                   | Titre                                       | Désignation  |
| Auteur/responsable/créateur    | Auteur                                  | Auteur                                      | Création→Auteur  |
| Date                           | Date de parution                        | Date d'exécution                            | Date de création ainsi que l'ensemble des champs de type « Date »  |
| Discipline sportive concernée  | Genre                                   | Discipline                                  | Indexation→Sujet/Thème   |
| Personnalité (sportive ou non) | Personnalités                           | Personne représentée                        | Indexation→Personne/Collectivité   |
| Événement sportif concerné     | Événements                              | Événement                                   | Indexation→Événement   |
| Type de document               | Type de document                        | Type de document                            | Désignation (ce champ donne aussi bien des informations sur le créateur de l'œuvre que sur le support de celle-ci) |

Les critères « titre », « auteur », « date » ne nécessitent pas l'élaboration d'un thésaurus transversal car il n'existe pas de besoin d'établir une concordance entre les diverses valeurs existantes pour ce type de champs. Les critères « discipline sportive concernée », « événement sportif concerné », « personnalité (sportive ou non) », et « type de document » font déjà l'objet d'une réflexion quant à la concordance des valeurs présentes dans les champs concernés sur les trois bases de données.

Les critères de sélection seront proposés aux utilisateurs par le biais d'une liste de choix déroulante. À côté de chacun des critères devra figurer une barre de saisie. Il conviendra de proposer une combinaison des critères de recherche avec les opérateurs booléens « ET », « OU », « SAUF » qui seront sélectionnables au choix grâce à des « boutons radio » disposés à droite de chaque barre de saisie (chaque opérateur devra combiner un critère avec celui qui le suit directement). Par défaut seuls trois critères de recherche seront associables, mais l'on pourrait prévoir une fonctionnalité permettant d'ajouter à l'écran un ou plusieurs critères supplémentaires.

L'écran de recherche avancée pourrait se présenter comme suit :

**Recherche avancée**

Cliquer sur l'icône pour afficher la liste des critères

|   | ET                    | OU                    | SAUF                  |
|---|-----------------------|-----------------------|-----------------------|
| Sélection du critère ▼ Saisir votre recherche | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sélection du critère ▼ Saisir votre recherche | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sélection du critère ▼ Saisir votre recherche | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Ajouter un autre critère

Lancer la recherche

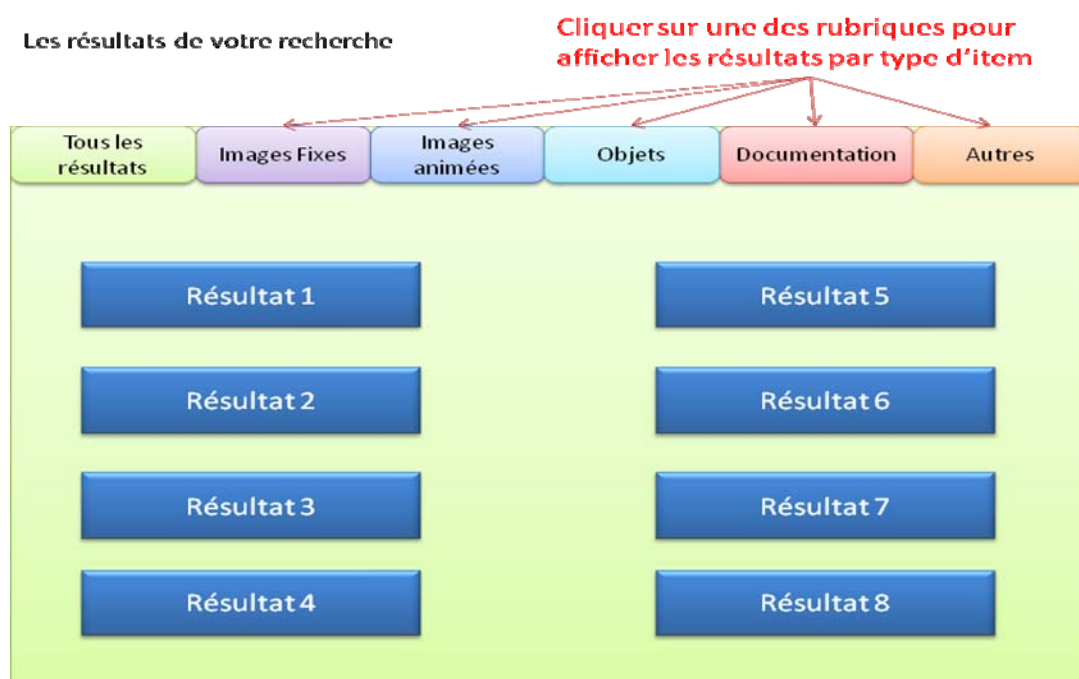
A noter qu'il serait nécessaire de proposer une liste de choix à la place de la barre de recherche si l'utilisateur sélectionne le critère « type de document ». Cette liste de choix devra être constituée des termes qui auront été normalisés dans un thésaurus transversal des types de documents.

## 2.2.4 Le filtrage des résultats par typologie

A l'exception de la recherche avancée qui propose le filtrage par typologie des items en amont, c'est-à-dire au moment de la recherche, une proposition de filtrage des résultats à l'écran de consultation des résultats est à envisager pour la recherche simple et pour la recherche par exploration thématique. Le « rebond » par typologie pourra s'effectuer en cliquant sur des onglets affichés sur l'écran des résultats, ce qui s'avèrerait bien plus original que l'affichage d'une liste de cases ou de « boutons radio » à cocher.

Cela suppose de créer un thésaurus transversal des typologies dans la mesure où le nombre de celles-ci s'avère assez important sur nos diverses bases de données, d'où la nécessité de pouvoir regrouper ces typologies sous un nombre restreint de mots-clés.

Ce filtrage pourrait être affiché à l'écran comme suit :



## 2.3 Les fonctionnalités attendues

### 2.3.1 Les fonctionnalités essentielles

Les fonctionnalités dites essentielles sont celles qui sont jugées comme indispensables sur le futur portail du Musée National du Sport si nous voulons respecter notre objectif qui est de rendre celui-ci accessible au grand public.

| Fonctionnalité   | Objectif  | Contrainte  | Critère d'appréciation minimal requis <sup>6</sup>                       | Un exemple à suivre |
|------------------|---|---|--|---------------------|
| Recherche simple | Permettre à l'utilisateur d'effectuer une recherche sur l'ensemble des pages du | 1) Une barre de recherche rapidement repérable sur les pages du portail<br>2) Prévoir plusieurs | Très bon : la recherche simple sera probablement le mode de recherche le |                     |

<sup>6</sup> Très bon>bon>moyen>médiocre

|   |  |   |   |  |
|---|--|---|---|--|
|   | portail et du répertoire de ressources en entrant une chaîne de caractères dans une barre de recherche   | éléments en <i>back office</i> qui aideront l'utilisateur :<br>-transcodage des index (majuscules/minuscules indifférenciées, accents, caractères spéciaux etc.)<br>-troncature automatique (recherche des termes tronqués)<br>-autocomplétion (suggestion de termes lors de la saisie dans la barre de recherche)  | plus utilisé, elle doit donc être efficace  |  |
| Recherche par navigation dans l'arborescence thématique | La navigation dans l'arborescence thématique permet à l'utilisateur de connaître la logique d'organisation des thématiques (avec les sous-thématiques et les niveaux inférieurs) et d'accéder à une information thématique de plus en plus précise | Bien définir les thématiques, ce qui passe par :<br>-une répartition cohérente du fonds entre les thématiques<br>-la construction d'une arborescence thématique avec différents niveaux (du plus général au particulier)<br>-élaborer un thésaurus transversal, c'est-à-dire qui permette de lancer une requête comprise par les trois bases de données disposant de modes d'indexation différents  | Très bon : ce mode de recherche n'est pas très développé car il nécessite une mûre réflexion (et un travail de réindexation) en amont. Un mauvais classement thématique du fonds rendra celui-ci inaccessible   | <a href="#">Le Portail des arts de la marionnette</a> permet une exploration thématique, c'est-à-dire naviguer dans l'arborescence thématique du fonds. Le <a href="#">Portail de la Cité de l'Immigration</a> propose également cette fonctionnalité  |
| Recherche avancée                                       | Permettre à l'utilisateur de mener une recherche plus pointue sur les répertoires de ressources en combinant des critères de recherche   | 1) Des critères de recherche pertinents mais restreints (quatre ou cinq) afin d'éviter à l'utilisateur de se perdre dans la sélection de ceux-ci<br>2) Une liste de choix ou un accès à l'index pour les critères afin d'aider l'utilisateur dans sa recherche. Néanmoins il doit avoir la possibilité de saisir ses propres critères<br>3) Pouvoir combiner les différents critères de recherche avec des opérateurs booléens ET, OU, SAUF qui seront soit suggérés à l'utilisateur (il ne doit pas les saisir lui-même) soit présents par | Bon : la recherche avancée n'est pas le mode de recherche privilégié par les utilisateurs (sauf pour le public de chercheurs et de spécialistes qui ne constitue pas notre priorité), il doit néanmoins être pris en main facilement et donner des résultats pertinents | Portail documentaire du Ministère de l'Écologie, cliquer <a href="#">ici</a> pour un aperçu de l'écran de recherche avancée<br><br>Musée Olympique avec sa « recherche de médailles » dont les critères sont proches de ceux à mettre en place pour le |

|                             |  |  |   |   |
|-----------------------------|--|--|---|---|
|                             |  | défaut (l'opérateur « ET » semble le plus pertinent dans ce cas)   |   | Portail du Musée National du Sport  |
| Aide à la recherche         | Aide l'utilisateur qui ne sait pas comment construire sa recherche   | 1) Une rubrique d'aide à la recherche facilement identifiable et qui s'ouvre dans une fenêtre pop-up afin de ne pas faire quitter l'écran à l'utilisateur<br><br>2) L'aide à la recherche ne se résume pas à une rubrique d'aide. Cela concerne aussi l'accompagnement pendant la recherche : l'accès à des listes de choix, aux index, l'autocomplétion des champs, le transcodage des index etc. | Très bon : l'utilisateur n'étant pas habitué à faire une recherche sur un site utilisera volontiers toute aide qui lui sera proposée, surtout si celle-ci est facilement accessible | Portail de la Médiathèque François-Mitterrand (Poitiers), cliquer <a href="#">ici</a> pour un aperçu des aides à la recherche (index proposés, rubrique d'aide visible)           |
| Liste des résultats paginée | Si le nombre de résultats d'une recherche est important, il est nécessaire de les répartir sur plusieurs pages. Cela évitera à l'utilisateur de trop « scroller » sur une même page. Une pagination permettra à celui-ci de naviguer entre les pages de résultats. | 1) Une pagination apparente (en haut et en bas de l'écran des résultats) avec des liens hypertextes fonctionnels.<br><br>2) Une répartition pertinente des résultats entre les pages : ne pas avoir trop ou pas assez de résultats sur une page<br><br>3) Définir les champs des notices qui seront affichées dans la liste des résultats  | Bon : une liste paginée participe au confort de l'utilisateur dans sa consultation, cette fonctionnalité ne doit donc pas être négligée   | Site du <a href="#">Musée du Quai Branly</a> sur lequel le nombre de résultats est affiché, les pages de résultats à consulter apparaissent en haut et en bas de la page          |
| Tri des résultats           | Le tri permet à l'utilisateur de modifier le classement des résultats à l'affichage selon divers critères prédéfinis   | Proposer des critères de tri pertinents :<br>-Chronologique<br>-Alphabétique<br>-Pertinence<br>-Période concernée<br>-Localisation<br>-Etc.  | Bon : cette fonctionnalité permet, en cas d'un nombre trop important de résultats, d'affiner sa recherche ou de voir s'afficher les résultats les plus intéressants parmi les       | <a href="#">Portail de la Cité de l'Immigration</a> qui permet de faire un tri sur plusieurs critères :<br>-pertinence<br>-type<br>-année<br>-auteur<br>-titre<br>-date d'édition |



|  |   |   |   |  |
|--|---|---|---|--|
|  |   |   | premières occurrences   | croissante -date d'édition décroissante<br>Néanmoins le tri est proposé au moment de la recherche et non sur l'écran des résultats                       |
| Panier des notices                         | Permet à l'utilisateur de sauvegarder les résultats qui l'intéressent. Il pourra les utiliser ou les consulter plus tard  | Si une fonctionnalité « panier des notices » est créée, il faut penser :<br>- à la durée de conservation des notices (le temps de la session pour un utilisateur non enregistré, permanente pour un utilisateur identifié)<br>- à la gestion des notices au sein du panier (ajout/suppression/classement)<br>- au nombre de notices maximum par panier<br>- à l'éventualité d'un partage des notices avec d'autres utilisateurs | Bon : un panier de notices n'est avant tout qu'un espace de stockage, néanmoins il participe à la fidélisation de l'internaute au site (il y reviendra s'il peut retrouver rapidement le fruit de ses recherches antérieures) |  |
| Impression des résultats                   | Permet à l'utilisateur de disposer d'une copie physique d'une notice ou d'une liste de résultats  | 1) Si une fonction impression des résultats est présente, elle doit être clairement identifiable<br>2) Proposer en outre un affichage de la notice au format PDF afin de permettre à l'utilisateur de sauvegarder celle-ci sur sa machine et de pouvoir l'imprimer plus tard  | Bon : participe au confort de l'utilisateur, certains internautes préfèrent encore lire sur un format papier  |  |
| Customisation de l'affichage des résultats | Permet à l'utilisateur de modifier l'affichage des résultats de sa requête. Cette customisation est surtout à proposer pour des résultats présentant des images | Réfléchir à des modes d'affichage variés, agréables visuellement et pertinents dans la présentation des informations  | Bon : participe au confort de l'utilisateur et à la facilitation de la consultation de l'information  | Le Portail des Arts de la Marionnette et le <a href="#">Musée Olympique</a> proposent à l'utilisateur de modifier l'affichage des résultats de recherche |
| Rebonds sur les notices,                   | L'utilisateur peut aller plus   | Cette suggestion suppose que les notices « liées »  | Bon : Un nombre   | de <a href="#">la</a>  |

|   |  |  |   |   |
|---|--|--|---|---|
| documents numérisés, URL liés à la notice consultée | loin dans sa recherche, se voir proposer des informations auxquelles il n'aurait pas pensé mais qui seraient susceptibles de l'intéresser  | dans les trois bases de données (ainsi que sur d'éventuelles bases extérieures) possèdent un ou plusieurs mots-clés en commun. Il faudra en amont réfléchir au nombre de mots-clés communs à partir duquel des notices pourront être « liées » et donc apparaître en suggestion des autres. En effet un nombre trop faible de mots-clés communs aboutirait à une suggestion trop large, un nombre trop élevé à une suggestion trop restreinte voire nulle.<br>Un thésaurus transversal est à construire dans la mesure où pour un même événement, une même discipline ou un même sportif des valeurs différentes ont été saisies sur les trois bases (parfois même au sein des mêmes bases)→besoin d'un (ou plusieurs) thésaurus pour harmoniser les mots-clés pour l'ensemble des bases | équilibré et une pertinence des notices liées raviront les utilisateurs qui prolongeront leur visite sur le site et seront encouragés pour d'éventuelles futures recherches | Médiathèque de la Cité de la Musique propose un accès à l'ensemble des documents liés à la notice consultée et ce quel que soit leur support (audio, vidéo, texte)<br><br>Le <a href="#">Brooklyn Museum</a> propose, à chaque résultat, des « tags » (mots-clés) associés à celui-ci |
| Accessibilité pour un public handicapé              | Il s'agit de mettre en place pour les internautes atteints d'un handicap visuel une fonctionnalité d'agrandissement des caractères, une modification des couleurs, ainsi qu'une rubrique « accessibilité » (où seront explicitées les commandes clavier pour agrandir les caractères sur les différents navigateurs) | Pas de réelle contrainte à envisager, il s'agit de fonctionnalités de base sur la plupart des sites professionnels.  | Bon : il s'agit de ne pas se priver d'une partie du Portail du Musée et de répondre aux recommandations du W3C en ce qui concerne l'accessibilité des sites Internet        | Sur le portail de la <a href="#">Médiathèque de Dole</a> il est possible de changer les couleurs du site et la taille de la police pour les personnes atteintes d'un handicap visuel  |

|   |   |   |  |  |
|---|---|---|--|--|
| Renseignements précis des droits attachés à l'objet | Permet à l'utilisateur d'appréhender l'ensemble des droits sur l'œuvre. Cela lui sera utile s'il souhaite réexploiter celle-ci  | Cela entraînera une indexation plus précise. Il faudra prévoir un peu plus de temps à la rédaction des nouvelles notices. Par ailleurs le renseignement des champs « droits » pourra être un critère de publication d'une notice sur le portail → si les champs « droits » ne sont pas assez renseignés, une notice ne pourra être publiée  | Très bon : La question de la propriété intellectuelle d'une œuvre est sensible   | Le Portail des Arts de la Marionnette propose des renseignements très précis sur les droits de l'item original consulté  |
| Export des résultats                                | Permet de pouvoir consulter les résultats en dehors du site   | Réfléchir aux formats d'export susceptibles d'intéresser les utilisateurs (fichiers texte ; PDF ; HTML ; XML ; etc.)  | Moyen : la fonction d'export est appréciée par les utilisateurs mais ne revêt pas pour autant une importance capitale, surtout si le portail leur propose un envoi des notices par mail et une impression de celles-ci | Le portail de la <a href="#">Cité de l'Architecture et du Patrimoine</a> propose un export des notices en plusieurs formats :<br>-CSV<br>-ISO 2709<br>-Texte<br>-Procite<br>-EndNote<br>-PDF |
| Historique des recherches                           | Permet de conserver ses requêtes. Cette fonction est utile pour tout utilisateur souhaitant faire les mêmes requêtes à des instants différents (il ne risque plus de les oublier). Cela pourrait en outre servir à combiner ses différentes recherches avec un opérateur booléen ET ou OU | Pour un historique des recherches il faut penser :<br>- à la durée de conservation des requêtes (le temps de la session pour un utilisateur non enregistré, permanente pour un utilisateur identifié)<br>- à une éventuelle gestion de celles-ci au sein d'un panier (ajout/suppression/classement)<br>-au nombre de requêtes maximum par panier<br>-à l'éventualité d'un partage des requêtes avec d'autres utilisateurs | Bon : il s'agit d'une fonctionnalité appréciable mais pas essentielle. La combinaison des recherches est quant à elle exclusive à des usagers expérimentés et ne constitue donc pas une priorité                       | Le <a href="#">portail de la Cité de l'Immigration</a> propose de consulter et de combiner les recherches effectuées dans son interface de recherche avancée                                 |
| Syndication   | L'utilisateur   | La rédaction d'une  | Très bon : une   | Le site du   |

|                                   |   |   |  |  |
|-----------------------------------|---|---|--|--|
| aux actualités du site/Newsletter | recevra par mail ou sur son agrégateur de flux de façon périodique un article sur l'actualité du Musée National du Sport (expositions, acquisitions etc.). L'utilisateur sera ainsi toujours concerné par le Musée sans avoir à y retourner | Newsletter nécessite un travail à fréquence régulière. Par conséquent nécessité de mobiliser des rédacteurs motivés, sinon réfléchir à la possibilité de mettre en place un système semi-automatisé (avec une structure prédéfinie et un contenu identifié) | Newsletter doit d'une part donner une information concise et complète à l'utilisateur et d'autre part susciter l'envie d'aller directement sur le site. Elle doit donc être de qualité | Musée du Louvre propose lors de <a href="#">l'inscription</a> de l'utilisateur un abonnement à la lettre d'information du Louvre |
|-----------------------------------|---|---|--|--|

### 2.3.2 Les fonctionnalités optionnelles

Les fonctionnalités qualifiées d'optionnelles sont celles que nous jugeons intéressantes pour notre futur portail, mais qui peuvent être modifiées, appauvries voire supprimées si les contraintes techniques et économiques liées à leur réalisation risquent de bloquer le projet.

| Fonctionnalité                  | Objectif  | Contrainte   | Critère d'appréciation minimal requis   | Un exemple à suivre  |
|---------------------------------|---|--|---|--|
| Nuage de tags pour la recherche | Permettre à l'utilisateur d'avoir une représentation visuelle des principales thématiques (« tags ») abordées par le portail (la taille de la police est proportionnelle à la récurrence d'un thème dans les notices). Il peut, en cliquant sur un des termes, accéder aux résultats liés | <p>1) Travail de mise en place d'un thésaurus transversal qui permette une correspondance entre les mots-clés du nuage et les mots-clés présents dans les notices</p> <p>2) Rendre le nuage de tags interactif, c'est-à-dire que les mots clés doivent être cliquables et mener aux notices leur étant attachées</p> <p>3) Le fonds étant en constante</p> | Bon : il s'agit d'un mode de recherche très peu répandu sur les sites culturels, la présence d'un nuage de tags basique serait pour le moment suffisant | Le blog « outils froids » propose en bas de page un <a href="#">nuage de mots-clés</a> qui permet d'accéder aux articles liés à ceux-ci<br><br><a href="#">L'ECPAD</a> propose sur sa page d'accueil un nuage de tags pour |

|   |  |  |  |  |
|---|--|--|--|--|
|   |  | incréméntation, il faut donc prévoir un nuage dynamique qui évolue graphiquement parallèlement à l'évolution du fonds (ajout/modification/suppression de notices)  |  | que l'utilisateur puisse mener une recherche par mot-clé   |
| Nuage de tags pour les résultats                          | Cette fonctionnalité est utile pour connaître l'importance des thèmes couverts par la requête de l'utilisateur, ce dernier peut donc apprécier la pertinence de sa requête d'un simple coup d'œil. Permet aussi de trier les résultats par mots-clés | <p>1) Travail de mise en place d'un thésaurus transversal qui permette une correspondance entre les mots-clés du nuage et les mots-clés présents dans les notices</p> <p>2) Rendre le nuage de tags interactif, c'est-à-dire que les mots clés doivent être cliquables et mener aux notices leur étant attachées</p> <p>3) Le fonds étant en constante incréméntation, il faut donc prévoir un nuage dynamique qui évolue graphiquement parallèlement à l'évolution du fonds (ajout/modification/suppression de notices)</p> | Très bon : une mauvaise attribution des mots-clés ou des liens inopérants rendraient l'information invisible à l'utilisateur. Il est donc capital que le nuage de tags soit efficace | Le <a href="#">portail des documents techniques de l'eau</a> propose, dans son écran des résultats, un nuage de tags dynamique et interactif |
| Syndication/alertes par rapport aux requêtes enregistrées | L'utilisateur peut s'abonner aux notices créées/modifiées en relation avec ses requêtes. Il recevra, lorsque des notices en rapport avec ses requêtes seront créées/modifiées, une alerte sur son agrégateur de flux (Netvibes, Google Reader etc.)  | Difficulté pour mettre en place un système automatisé d'alerte par rapport à des requêtes : il existe autant de pages de résultats qu'il existe de requêtes. Par conséquent il serait nécessaire de contourner cette difficulté en créant une syndication par rapport aux mots-clés de la notice. Si l'utilisateur se voit proposer un flux RSS en rapport avec sa requête, la syndication se fera en réalité sur les mots-clés en rapport avec sa requête et présents   | Très bon : Un utilisateur appréciera de pouvoir être informé des nouveautés du fonds du Musée National du Sport en rapport avec ses centres d'intérêt                                | -Le Portail du Ministère de l'Ecologie, propose une <a href="#">syndication par base de données ou par thème</a>                             |

|   |  |  |  |   |
|---|--|--|--|---|
|   |  | dans les notices   |  |   |
| Possibilité de créer un profil de recherche | Le profil de recherche pourrait permettre à l'utilisateur qui s'identifie sur le site de disposer d'un espace personnel dans lequel il pourrait retrouver son historique de recherche, son panier de notices, ses thématiques de prédilection etc.               | <p>1) Réfléchir à la nature des renseignements que les utilisateurs vont devoir donner pour s'enregistrer</p> <p>2) Réfléchir au mode d'identification (login/mot de passe sélectionnés par l'utilisateur ou attribués automatiquement par le Portail ?)</p> <p>3) Savoir combien de profils de recherche, avec l'ensemble des informations y étant présentes, pourront être supportés par le portail</p> <p>3) Réfléchir aux fonctionnalités réservées aux utilisateurs identifiés (panier de notices et historique des recherches d'une durée plus longue ; possibilité de « tagger » les notices ; abonnement à la Newsletter, aux fils RSS etc.)</p> | Bon : les fonctionnalités du profil de recherche satisferont les chercheurs/spécialistes, toutefois ceux-ci ne sont pas le principal public visé par le futur Portail du Musée National du Sport | <p>Le site du Musée du Louvre propose à l'internaute de créer <a href="#">un profil de recherche</a>.</p> <p>Le Brooklyn Museum (avec le « <a href="#">Posse</a> ») et le MET (avec « <a href="#">My Met</a> ») proposent un espace utilisateur faisant bénéficier de fonctionnalités supplémentaires</p> |
| « Tagging » des notices                     | Permet aux visiteurs de « tagger » les notices qu'ils consultent, c'est-à-dire d'ajouter des mots clés s'ils considèrent l'information lacunaire ou erronée. Cela s'avèrerait très utile au Musée National du Sport dans la mesure où certaines personnalités et | <p>1) Réserver le « tagging » aux utilisateurs identifiés (ceux qui disposent d'un profil de recherche) si l'on souhaite éviter d'être débordé par les suggestions des internautes</p> <p>2) Pour les notices prévoir un espace « suggestions » réservé aux utilisateurs : les notices n'ont pas à être modifiées par les utilisateurs mais par</p>  | Bon : le principe du « tagging » des notices n'est pas très répandu sur la sitosphère culturelle. Ainsi un principe de « tagging » basique sera apprécié par les utilisateurs                    | <p>Le <a href="#">Musée Mc Cord</a> propose à ses utilisateurs identifiés de tagger, dans un espace dédié, les images qu'ils consultent</p> <p>Le <a href="#">Brooklyn Museum</a> propose également aux</p>   |

|   |  |  |  |  |
|---|--|--|--|--|
|   | <p>événements restent difficiles à identifier pour des non spécialistes. Ainsi les suggestions des spécialistes/chercheurs constituent une aide non négligeable. En outre le principe du « tagging » permet de créer un sentiment de participation aux activités du Musée chez les utilisateurs identifiés</p> | <p>les membres du Musée National du Sport qui seront tenus au courant des suggestions et qui pourront en tenir compte pour modifier les notices <i>a posteriori</i></p> <p>3) Prévoir un système d'alerte pour que les administrateurs puissent répercuter les suggestions aux catalogueurs et décider d'intégrer ou non l'information aux notices</p> |  | <p>membres du « Posse » de tagger les notices</p>  |
| <p>Calendrier des activités de la structure et des grands événements sportifs du moment</p> | <p>1) Promouvoir les activités du Musée National du Sport</p> <p>2) Donner un aspect vivant au Portail</p>   | <p>Etre régulièrement mis à jour</p>   | <p>Bon</p>   | <p>Le <a href="#">Musée Guimet</a> propose sur la page d'accueil de son site un calendrier de ses activités</p>  |
| <p>Espace/rubrique « découverte » pour faire découvrir des éléments du fonds</p>            | <p>Faire découvrir des éléments du fonds du Musée aux internautes. Cela permet d'identifier la nature des collections possédées</p>  | <p>Etre régulièrement mis à jour, une rubrique « découverte » non renouvelée pourrait donner une impression d'abandon du site.</p>   | <p>Bon</p>   | <p>Le <a href="#">Musée Guimet</a> propose dans sa rubrique « Découverte » de visionner ses dernières acquisitions</p>   |
| <p>Exposition virtuelle</p>   | <p>Proposer aux internautes une exposition en ligne avec une thématique et des objets de collection définis. Cela permet d'abolir les contraintes physiques et de promouvoir les activités du Musée National du Sport. Deux types</p>  | <p>1) Nécessité d'intéresser les différents Départements du Musée à participer régulièrement à de tels projets</p> <p>2) Contrainte technique : nécessité de prévoir divers supports de documents (vidéo, audio, texte) tout en garantissant une navigation fluide</p>   | <p>Bon : les expositions virtuelles constitueraient une vitrine en ligne pour les expositions physiques du Musée</p> | <p>Le Musée du Quai Branly propose des expositions exclusivement pour son site, ce sont les <a href="#">programmes interactifs</a></p> <p>Le Musée Mc Cord propose</p> |

|      |   |   |  |  |
|------|---|---|--|--|
|      | <p>d'exposition pourront être mis en ligne :</p> <p>-Une exposition physique achevée pourra être poursuivie en ligne</p> <p>-Une exposition propre au portail</p>                             |   |  | <p>des <a href="#">expositions virtuelles</a> accessibles en flash ou en HTML</p>  |
| Jeux | <p>Proposer aux internautes des jeux (quizz par exemple) autour des images du fonds du Musée National du Sport. Cela aura pour effet de faire découvrir de façon ludique les collections.</p> | <p>1) Être régulièrement mis à jour (ne pas proposer le même jeu pendant plusieurs mois)</p> <p>2) Varier la nature des jeux, faire preuve d'imagination</p> <p>3) Nécessite une participation régulière des membres du Musée (Département des publics)</p> <p>4) Intéresser la participation avec des gains : entrée gratuite, rencontres etc.</p> | <p>Bon : peut constituer un plus dans l'intérêt du visiteur pour le Musée et ses collections</p> | <p>Le Musée Mc Cord propose des <a href="#">jeux</a> avec les images de son fonds (jeux de rôle, jeux d'association, jeu d'observation, jeux-questionnaires)</p> |

L'ambition première de notre futur portail étant de permettre une recherche fédérée sur l'ensemble de nos bases, les fonctionnalités essentielles concernent avant tout ce qui est lié à la recherche et à l'exploitation des résultats. Les fonctionnalités optionnelles deviendraient plus importantes si notre portail devait être appelé à se substituer au site déjà existant. Néanmoins nous considérons que certaines de ces dernières pourraient être intéressantes sur notre portail afin de refléter leur utilité pour le site existant.



## 3 Les préconisations pour mener le projet de portail du Musée National du Sport

---

### 3.1 Les différentes étapes constituant le plan d'action

Pour chacune des solutions soumises, le plan d'action comportera plusieurs phases communes successives :

- Communiquer sur le projet de portail : il faudra sensibiliser les membres du Musée National du Sport quant à l'intérêt d'un portail de qualité pour mettre en valeur les activités de chacun. Il serait donc judicieux dans un premier temps de s'entretenir avec la nouvelle direction du Musée dont l'arrivée est prévue pour le mois de septembre 2011. En effet le soutien affiché de la Direction sera capital pour rendre le projet concret aux yeux de tous. Dans un deuxième temps mettre en place des réunions de pré-lancement pour communiquer les enjeux du portail aux Départements du Musée ainsi que la participation qui leur sera demandée. Enfin tout au long du projet il sera utile de tenir des réunions (hebdomadaires si possible, mensuelles au moins) afin de communiquer sur l'avancée du projet, de valider les étapes, d'opérer des ajustements etc.
- Mettre en place une équipe pour le projet de portail et en confier le pilotage au SID du Musée National du Sport dans la mesure où ce projet va nécessiter des savoir-faire info-documentaires, une expérience dans la conduite de projet, des contacts privilégiés avec les prestataires des bases de données, une connaissance du chantier de numérisation en cours. Le comité de pilotage sera chargé d'élaborer un planning définitif et de constituer les équipes participant au projet.
- Réfléchir, avec les institutions disposant de fonds sur le sport (INSEP, Musée du CIO par ex.), d'une mise en commun de nos ressources interrogeables par les internautes.
- Elaborer des thésaurus transversaux pour chacun des groupes de champs communs définis précédemment (sportifs, disciplines, événements, types de document), ce qui sera un préalable indispensable car sans transversalité la recherche fédérée n'est pas viable : de nombreuses notices demeureraient invisibles à l'utilisateur. Faire

appel à un éditeur de moteur de recherche sémantique (ex : Mondeca) pour développer en commun ces thesaurus.

- Le choix d'un protocole d'interrogation en fonction des besoins (et des moyens) du Musée National du Sport (OAI, Z39.50, autre ?). Selon le choix validé, des actions nécessaires pour rendre nos bases de données interrogeables en ligne devront être entreprises parallèlement à l'élaboration des thesaurus. Nous expliciterons plus en détail ces actions dans la partie suivante.
- Le choix des notices qui seront accessibles : régler la question du dédoublement et être attentif à la progression du chantier de numérisation du fonds images fixes et vidéo, les fonds non numérisés ne devront évidemment pas être rendus accessibles.
- Définir les contenus du portail et choisir un prestataire pour réaliser le portail.
- Choisir l'hébergeur du futur portail

## **3.2 Définition des solutions envisageables**

Le recherche fédérée devient techniquement possible en se reposant sur un protocole OAI, un protocole Z3950 ou une application ODBC. Pour chacune de ces solutions nous devons rappeler des contraintes spécifiques.

### **3.2.1 La solution OAI-PMH**

Nos bases de données Cadic Intégrale et Armadillo sont, selon nos entretiens avec les éditeurs de ces produits, moissonnables. Le choix d'un protocole OAI ne devrait, idéalement, pas poser de problème pour ces deux bases. Il conviendra néanmoins d'avoir la possibilité de tester le moissonnage afin de s'assurer de l'opérabilité de cette solution.

Cependant notre base de données Micromusée demeure à ce jour en local malgré les demandes du SID du Musée National du Sport auprès de l'éditeur (Mobydoc) pour basculer la base de données en OPAC Web. Face à cette contrainte nous pouvons envisager deux solutions :

- Tenter de rétablir le dialogue avec Mobydoc : selon leur service d'assistance technique le basculement en OPAC Web ne nécessiterait qu'une mise à jour de notre

base de données, le téléchargement et l'installation de la dernière version du module OPAC Web.

- La mise en place d'un moissonnage OAI possède l'avantage de ne pas nécessiter le basculement notre base Micromusée en OPAC Web. Mais il serait alors nécessaire d'exiger de Mobydoc un export XML des notices Micromusée vers notre entrepôt OAI qui alimentera l'index central interrogeable en ligne. Néanmoins cela nécessitera fréquemment des coûts et une charge de travail supplémentaire à chaque fois que nous désirerons exporter des notices.

Dans le cadre d'un moissonnage OAI, il convient de se procurer et de configurer le programme permettant de moissonner les bases : le moissonneur. Cela sera nécessaire si nous désirons un moissonnage automatique pour les bases Cadic et Armadillo (voire MicroMusée si nous pouvons la basculer en OPAC Web).

Par ailleurs, si nous souhaitons mettre en commun les ressources de diverses institutions, le protocole OAI semble le plus adapté aux projets pluri-institutionnels dans la mesure où celui-ci permet d'éviter de créer une charge supplémentaire sur les serveurs des différentes bases de données. Par conséquent, quelle que soit la solution adoptée pour notre projet de portail, nous recommandons le protocole OAI-PMH en ce qui concerne une éventuelle mise en commun de ressources pluri-institutionnelles. En contrepartie cela exigerait de contacter, avant même la validation du projet par la nouvelle direction du Musée National du Sport, les acteurs visés par ce projet de partage afin de réfléchir aux contraintes de mise en place d'un moissonnage OAI.

### **3.2.2 La solution Z39.50 (SRU/SRW)**

Nos bases de données Armadillo et Cadic Intégrale sont, selon leurs éditeurs respectifs, également interrogeables par un protocole Z39.50 (par exemple le logiciel Cadic Intégrale propose un serveur Z39.50). Il conviendrait toutefois de leur demander quels sont leurs clients ayant opté pour une interrogation Z39.50 afin de :

- tester le mode de recherche sur leurs sites
- prendre contact avec ces clients afin d'obtenir des informations sur la mise en place concrète d'une interrogation via un protocole Z39.50

Concernant notre base de données Micromusée, il sera nécessaire de basculer celle-ci en OPAC Web dans la mesure où le protocole Z39.50 repose sur une interrogation synchrone de nos différents répertoires de ressources. Selon le service d'assistance technique de l'éditeur Mobydoc, la base de données Micromusée deviendrait alors interrogeable par ce protocole. Or si la solution Z39.50 (SRU/SRW) est adoptée, la mise en place d'un moissonnage OAI devra tout de même être envisagée dans l'optique d'une mise en commun de ressources pluri-institutionnelles. Cela entraînerait donc des coûts et une charge de travail supplémentaires pour le Musée National du Sport, ce qui nécessite de se demander si les avantages d'une interrogation selon un protocole Z39.50 (fraîcheur des informations trouvées, reconnaissance de plusieurs formats de notices) valent des dépenses et une charge de travail supplémentaires.

### **3.2.3 La solution ODBC**

La solution ODBC permettrait, comme le protocole Z39.50, d'interroger en synchrone plusieurs répertoires de ressources mais en prenant en compte l'intégralité des formats de notices pouvant exister, livrant ainsi des notices fidèles à leur format d'origine en guise de résultat. Cette solution s'avèrerait judicieuse dans la mesure où les notices « muséales » sont très riches en métadonnées.

Néanmoins quatre contraintes sont à prendre en compte pour la mise en place d'un ODBC :

- Basculer Micromusée en OPAC Web car l'ODBC repose sur un mode d'interrogation synchrone
- Mise en place coûteuse
- Une configuration des connecteurs qui s'avère complexe et qui donc nécessiterait de faire appel à un prestataire → coût supplémentaire
- La nécessité de mettre en place un moissonnage OAI en parallèle si le projet de mise en commun des ressources des différentes institutions est lancé

### 3.3 Une piste à suivre : la solution de l'éditeur Armadillo, compromis entre outil de gestion de contenu et interface de recherche fédérée reposant sur des outils sémantiques

#### 3.3.1 Audit des éditeurs de solutions portail et critères d'évaluation

Nous avons, au cours d'un audit sur les éditeurs de portails, pu apprécier des illustrations concrètes de réalisations. Nous avons dressé le tableau récapitulatif suivant :

| Identité                           | Objet  | Exemples de réalisation   |
|------------------------------------|--|---|
| <a href="#"><u>Archimed</u></a>    | Editeur français dont la division « Culture et savoir » propose des solutions portail dans le domaine de la culture  | <a href="#"><u>Portail des médiathèques de Poitiers</u></a><br><br><a href="#"><u>Portail des médiathèques de Massy</u></a><br><br><a href="#"><u>Médiathèque de Dole</u></a><br><br><a href="#"><u>Portail du Museum Nationale d'Histoire naturelle</u></a><br><br><a href="#"><u>Portail documentaire de la Cité de l'architecture</u></a><br><br><a href="#"><u>Catalogue de la Médiathèque de la Cité de la Musique</u></a> |
| <a href="#"><u>Armadillo</u></a>   | Editeur français de logiciels documentaires et proposant une solution portail avec son produit « Armadillo Portail »   | <a href="#"><u>Portail des arts de la marionnette</u></a><br><br>Portail de l'Océan Indien  |
| <a href="#"><u>Mosquito</u></a>    | Concepteur de portails pour les lieux de culture   | <a href="#"><u>Musée Guimet</u></a>   |
| <a href="#"><u>Heliogabale</u></a> | Web Design   | <a href="#"><u>Musée des Arts et Métiers</u></a>  |
| <a href="#"><u>Advisa</u></a>      | Agence de conseil en communication interactive   | <a href="#"><u>Centre Pompidou Metz</u></a>   |
| <a href="#"><u>Cap Gemini</u></a>  | Propose un large éventail de services et de solutions allant du « Conseil en stratégie » à l'« Intégration de systèmes et applications informatiques ». Dispose d'une équipe « Culture » pour fournir des solutions et un accompagnement de projet pour des structures culturelles | <a href="#"><u>Musée du Louvre</u></a>  |

Le critère de sélection le plus important parmi ces éditeurs a été l'interface graphique, c'est donc pourquoi nous ne conseillerons pas la société Archimed qui, même si elle a une expérience non négligeable dans le domaine de la culture, développe des produits à l'aspect très « bibliothéconomie » dont nous souhaitons justement nous détacher.

La société Cap Gemini, qui a quant à elle réalisé un travail que nous considérons comme très abouti d'un point de vue graphique, navigationnel et fonctionnel pour le Musée du Louvre, semble être un éditeur qui n'agit que sur de très gros marchés en menant des projets très ambitieux. Or notre futur budget, qui n'a pas encore été fixé mais qui restera certainement limité par rapport au Musée du Louvre, ainsi que nos ambitions ne seront très probablement pas assez intéressants pour attirer cet éditeur.

Nous ne préconiserons pas la société Heliogabale qui, malgré un travail assez intéressant pour le site du Musée des Arts et Métiers, ne fournit que très peu d'informations sur son propre site.

Les sociétés Advisa et Mosquito semblent proposer des produits intéressants en terme d'interface, de navigation et de fonctionnalités et ont en outre déjà travaillé avec des musées. Par ailleurs ces éditeurs, de par les sites qu'ils développent, semblent beaucoup plus adaptés à un projet tel que la création d'un portail pour une structure disposant de moyens plus limités.

### **3.3.2 La sélection d'une solution « tout en un »**

Armadillo semble être l'éditeur de portail le plus intéressant pour notre projet car celui-ci propose de solutionner plusieurs de nos problématiques :

- Rendre nos trois bases de données interrogeables, dont la base Micromusée. En effet Armadillo s'engagerait à s'entretenir avec les éditeurs des deux autres logiciels pour discuter de l'interopérabilité de leurs produits respectifs. Par ailleurs le portail de l'Océan Indien qu'Armadillo a développé comprenait aussi une base de données Micromusée, ce qui garantit son savoir faire par rapport aux outils développés par Mobydoc.
- Garantir le respect des formats d'origine des notices de nos bases de données : en mettant en place un ODBC qui reconnaisse tous les formats de notices, Armadillo

permettrait de restituer des notices riches en métadonnées qui caractérisent les fonds muséaux.

- Prendre en charge le développement des connecteurs ODBC.
- Développer un moteur de recherche sémantique pour notre site qui permettrait de surmonter le problème du manque de normalisation des procédures d'indexation qui a abouti à avoir plusieurs termes différents pour une même valeur. Cela nécessiterait l'élaboration de thésaurus transversaux en collaboration avec les membres du Musée National du Sport et assurerait une recherche efficace sur nos trois bases de données.
- Développer un portail au graphisme mettant en valeur l'image et à la navigation intuitive.
- Une solution « tout en un » qui pourrait nous faire gagner du temps et de l'argent. Choisir Armadillo nous éviterait de séparer la mise en place de l'interrogeabilité de nos bases de données, le développement du moteur de recherche sémantique (et l'élaboration de thésaurus transversaux qui va de pair) et le développement du portail avec ses fonctionnalités. De ce fait nous estimons que si un même prestataire intervient à toutes les étapes du projet, cela permet d'assurer une continuité et une compréhension de l'ensemble des problématiques qui ne peuvent qu'être bénéfiques quant à la qualité, la cohérence et la rapidité d'élaboration de notre projet. Par ailleurs nous considérons qu'il serait possible de négocier auprès de cet éditeur un « prix de gros » dans la mesure où les sommes investies auprès de lui seraient plus conséquentes. Par ailleurs Armadillo propose un service « à la carte », c'est-à-dire que l'éditeur propose des ajustements aux besoins et fonctionnalités définis dans le cahier des charges en fonction du budget alloué au projet, ce qui garantit une certaine transparence des coûts.
- Le développement d'un moissonnage OAI dans le cadre d'une mise en commun de nos ressources avec celles d'autres institutions. En effet Armadillo se propose, parallèlement au développement d'une application ODBC, de mettre en place un entrepôt OAI pour permettre le moissonnage des ressources du Musée National du Sport si celui-ci envisage de rejoindre un projet interinstitutionnel.

- La connaissance de nos ressources et de nos problématiques puisqu'Armadillo est l'éditeur de notre base de données iconographique et a donc déjà collaboré avec le Musée National du Sport.



# Conclusion

Réfléchir à la mise en place d'un portail de recherche fédérée pour le Musée National du Sport nous a permis de comprendre qu'il était nécessaire, avant même de planifier des tâches, d'adopter une démarche d'investigation auprès de structures ayant déjà réalisé ce type de projet. La veille sur les portails culturels et les entretiens avec les responsables de ces projets nous ont appris que toutes les institutions culturelles devaient faire face aux mêmes contraintes techniques (langages d'indexation et formats de notice différents selon les bases de données, mise en interrogeabilité de celles-ci, mise en place des protocoles d'interrogation) et organisationnelles (susciter l'adhésion des membres de l'institution au projet, budget à respecter, faibles marges de manœuvre concernant les délais, critères de sélection à retenir pour le choix du ou des prestataires, gel des crédits). Ces institutions ont d'ailleurs un but commun, celui de renouer le contact avec le public en lui proposant de découvrir leurs fonds et leurs activités de manière simple et interactive. Simplifier la navigation et la recherche nécessite de réaliser un travail de longue haleine en « back office ». Nous pourrions même déduire que plus le travail de réflexion sur la concordance des champs, l'interopérabilité des langages d'indexation, la charte graphique, les modes de navigation, les modes de recherche, la présentation et l'exploitation des résultats, les fonctionnalités etc. est abouti (et donc long), plus la tâche de l'utilisateur aura été simplifiée.

C'est donc dans cette optique que nous avons mené ce projet de réflexion sur la mise en place d'un portail, réflexion qui a été matérialisée par un cahier des charges. Notre mission a consisté à réfléchir à la possibilité pour le Musée National du Sport de lancer un tel projet d'un point de vue technique. Nous avons pour ce faire effectué un état des lieux de ce que les autres institutions culturelles avaient entrepris avant de nous demander comment le Musée National du Sport pouvait surpasser ses diverses problématiques (indexation non uniformisée malgré les efforts des membres du SID, formats de notices et donc champs de métadonnées différents selon les bases de données, présence de doublons entre ces dernières, une base de données non interrogeable en ligne) afin de proposer le même type de produit aux futurs utilisateurs. Notre futur portail étant destiné à être principalement consacré à la recherche, nous nous sommes surtout consacrés à :

- Rendre toutes les bases de données interrogeables (la base Micromusée devant être basculée en OPAC Web ou à défaut bénéficier pour les notices sélectionnées d'un export XML vers un entrepôt OAI)
- Préserver les formats originaux des notices, même si les champs de gestion peuvent être supprimés car non utiles à l'utilisateur
- Réfléchir à la concordance des champs, c'est-à-dire identifier les champs communs à l'ensemble des bases de données afin de proposer une recherche avancée (selon des critères de recherche) efficace

- Réfléchir à l'interopérabilité sémantique des langages d'indexation sur les diverses bases de données
- Sélectionner le protocole d'interrogation le plus adapté à la situation du Musée National du Sport

La question de l'interopérabilité sémantique a été la plus intéressante selon nous car étant au cœur des problématiques métier sur l'organisation de l'accès à l'information. Nous avons très rapidement déduit au cours de nos recherches sur les contraintes liées à la mise en place d'une recherche fédérée que le talon d'Achille d'une telle fonctionnalité demeurerait le manque de pertinence des résultats. En effet, en l'absence d'un vocabulaire contrôlé de type thésaurus, une recherche sur une base de données perd en pertinence. Et cette contrainte se complexifie si l'interrogation doit être effectuée sur plusieurs bases de données, car chacune pouvant posséder un thésaurus propre. De ce fait la solution qui nous a semblé la plus efficace a été la mise en place d'un « superthésaurus » permettant la concordance entre les divers thésaurus, un « concept » ayant pour équivalents plusieurs « concepts » dans les différentes bases de données. Or une nouvelle contrainte devait être prise en considération : l'attachement des internautes à la « recherche simple » (en plein texte). Car même si nous leur présentions une recherche par thésaurus via une interface simple et intuitive (l'arborescence thématique), beaucoup d'entre eux utiliseraient instinctivement la recherche simple car plus proche de leurs habitudes de recherche (ex : Google). Par conséquent nous devons nous demander comment instaurer une équivalence entre les termes du « superthésaurus » et les termes qui pourraient être potentiellement saisis par les internautes. C'est ainsi que la structuration des informations autour de concepts liés sémantiquement et non plus autour de descripteurs nous a semblé la solution la plus pertinente. De ce fait nous avons logiquement préconisé de se tourner vers les prestataires proposant des solutions sémantiques (Armadillo) afin de résoudre le problème de l'interopérabilité sémantique et ainsi d'optimiser la recherche d'information.

Suite à ces préconisations, il va être nécessaire de :

- Les communiquer à la nouvelle direction du Musée National du Sport, prenant place au cours du mois de septembre, afin de lancer concrètement le projet
- Travailler conjointement avec l'éditeur retenu, Armadillo, afin d'élaborer une ontologie (définition des concepts, des relations sémantiques)
- Obtenir la collaboration des éditeurs des divers logiciels de gestion de bases de données du Musée afin de lever toute contrainte technique
- Sensibiliser les membres du Musée au projet et à ses apports pour tous au-delà des investissements personnels
- Solliciter d'autres institutions sportives et culturelles (INSEP, Musée Olympique etc.) dans l'optique d'une mutualisation des ressources grâce à notre solution sémantique

Ce travail s'inscrit dans la continuité des travaux de Laurène Rolland-Bertrand et de Circé Krouch-Guihem, aujourd'hui membres du Musée National du Sport. En effet si Laurène Rolland-Bertrand a délimité les enjeux et objectifs de ce que devait être le futur Service d'information-documentation du Musée en ce qui concerne la mise en valeur des fonds par des impératifs de gestion (65, BERTRAND), Circé Krouch-Guilhem s'était quant à elle intéressée à la mise en valeur du fonds iconographique du Musée et avait préconisé la mise en place d'un portail de recherche fédérée à cette fin (66, KROUCH-GUILHEM). Nous avons donc cherché dans ce mémoire, ainsi qu'au cours de notre mission de stage, à rester fidèle aux ambitions du SID tout en proposant de nouvelles perspectives pour servir au mieux les intérêts du Musée National du Sport.

Nous pouvons donc conclure que la réflexion relative à ce projet de portail de recherche fédérée a confirmé à nos yeux que le rôle du professionnel de l'information-documentation est devenu déterminant face au besoin de valorisation de ressources sur le Web. Si l'émergence de nouvelles pratiques de recherche sur le Web a semblé estomper la médiation du documentaliste dans l'accès à l'information (moteur de recherche statistique avec le « Web 1.0 », la « folksonomie » avec le Web 2.0), le Web de données a confirmé le besoin d'une réelle structuration sémantique des informations sur Internet. Or cette « structuration sémantique », qui prend la forme d'une ontologie, ne peut apporter du sens au contenu des pages web que grâce au savoir info-documentaire concernant les vocabulaires contrôlés qui, s'ils demeurent des outils traditionnels, semblent plus que jamais nécessaires dans l'élaboration d'un Web compréhensible aussi bien par les machines que par les humains. Si le futur du Web sera sémantique, il incombe au documentaliste de permettre l'évolution d'un Web de pages à un Web de données ouvertes et partagées.

# **Bibliographie**

**La bibliographie a été arrêtée le 15 octobre 2011. Elle comprend 66 références.**

La rédaction des références bibliographiques est conforme aux normes :

- Z44-005. décembre 1987. Documentation. Références bibliographiques : contenu, forme et structure.
- NF ISO 690-2 février 1998 Information et documentation. Références bibliographiques, documents électroniques, documents complets et parties de documents.

Nous avons regroupé les notices selon des rubriques thématiques en rapport avec notre problématique. Celles-ci apparaissent dans l'ordre suivant : Web sémantique, recherche fédérée, portail, thésaurus, ontologie, contexte info-documentaire au Musée National du Sport.

### **Web Sémantique**

[1] ADAMS Katherine. The Semantic Web. Online Mag, juillet/août 2002, p.20-23.

*K. Adams livre une définition du Web sémantique. Elle présente les taxonomies et les ontologies qui reposent sur un vocabulaire structuré qui identifie un mot-clé afin de représenter un concept qui peut être décrit en utilisant plusieurs mots. Les professionnels de l'information possèdent selon elle les connaissances que les informaticiens, les chefs d'entreprise et autres recherchent quand ils essaient d'appréhender le Web sémantique.*

[2] BAUDRY DE VAUX Marie, DALBIN Sylvie. « Journée d'étude ADBS-INTD » : Métadonnées et valorisation de l'information. Documentaliste-Sciences de l'information, 2006, vol.43, n°2, p.144-147.

*L'importance des métadonnées pour la gestion du cycle de vie des documents, pour la production et la diffusion des informations ainsi que pour la valorisation de l'information lorsque les métadonnées sont manipulées/analysées par des outils informatiques. La métadonnée, de par sa nature, est une donnée qui contient de la sémantique dans la mesure où elle donne du sens à la ressource à laquelle elle est attachée. Ainsi une métadonnée est importante dans le cadre du Web sémantique dans la mesure où elle doit pouvoir être analysée/ « comprise » par un agent informatique.*

[3] BERMES Emmanuelle. Le Web de données : perspectives pour les métiers de l'information-documentation (en ligne), 18 février 2010 (consulté le 19 août 2011).

<<http://www.adbs.fr/le-web-de-donnees-perspectives-pour-les-metiers-de-l-information-documentation-79361.htm?RH=1266334869518>>

*Intervention vidéo d'Emmanuelle Bermes sur les perspectives offertes par le Web de données (ou "Web sémantique") aux professionnels de l'information-documentation. Elle définit ce qu'est le Web de données, sa mise en œuvre avec RDF, l'adoption de l'architecture du Web etc.*

[4] BERMES Emmanuelle. Publier les référentiels sur le Web sémantique. *In* Journée AFNOR CG 46/BNF (en ligne), 27 mai 2011 (consulté le 16 août 2011). 25 p.

<[http://www.bnf.fr/documents/afnor2011\\_bermes.pdf](http://www.bnf.fr/documents/afnor2011_bermes.pdf)>

*Emmanuelle Bermes apporte à l'occasion de cette journée une définition synthétique du Web sémantique, de ses outils et de ses langages assez récente.*

[5] BRUN-PICARD Céline. Les instruments d'organisation des connaissances à l'heure du Web sémantique. *BBF*, 2010, n°6, p.72-73.

*L'auteure rappelle que l'objectif des travaux du Web sémantique est de rendre le contenu du Web à la fois lisible par les humains et exploitable par les machines. Cela permet par exemple de donner accès au Web caché, de faciliter la recherche d'information ou encore l'agrégation de données dans un portail d'entreprise. L'ensemble de la démarche repose sur des données structurées ou semi-structurées et représentées dans un formalisme permettant leur manipulation par les machines : les métadonnées.*

[6] CHARLET Jean, LEGER Alain. Applications du Web sémantique. *Information Interaction Intelligence*, 2004, hors-série, p.1-19.

*Les apports des technologies du Web sémantique sont aussi bien qualitatifs, économiques et sociaux comme en témoignent les applications de traduction automatique permettant la traduction du sens d'un texte dans une autre langue et non plus une traduction mot à mot, le traitement automatique des langues permettant à la machine de « comprendre » le sens*

*d'une phrase, rendant possible de ce fait l'interopérabilité de sources hétérogènes afin de permettre un partage des ressources.*

[7] DECLERCQ Sébastien. Ontologie et Web sémantique : l'ère du documentaliste a-t-elle sonné ? Cahiers de la documentation, 2010/2, p.18-21.

*L'article présente les grandes thématiques associées au Web sémantique (définition, mise en évidence des ontologies) ainsi que le langage OWL (Web Ontology Language) permettant la création d'ontologies. Enfin l'auteur affirme que les professionnels de l'information-documentation ont un rôle à jouer dans la gestion de ces nouveaux éléments informationnels.*

[8] GAGNON Michel. Intervention à l'École Polytechnique de Montréal : Introduction au Web sémantique (présentation powerpoint) (en ligne). 84 p. (consulté le 15 août 2011).

<<http://www.cours.polymtl.ca/inf4215/documentation/tutorielWS.pdf>>

*Cette présentation powerpoint est une initiation au Web sémantique : sa création, son utilité, les technologies sur lesquelles il repose, ses outils etc.*

[9] GREENBERG Jane, SUTTON Stuart, GRANT CAMPBELL D. Metadata : fundamental component of the Semantic Web. Bulletin of the American Society for Information Science and Technology, avril/mai 2003, vol.29, n°4, p.16-18.

*Les auteurs présentent les métadonnées comme les fondements mêmes du Web sémantique. Ils donnent en outre des définitions au sein d'un petit dictionnaire du Web sémantique (RDF, Ontologie, XML).*

[10] MESGUICH Véronique. Le Web sémantique : utopie ou réalité ? Netsources (en ligne), novembre-décembre 2007, (consulté le 7 août 2011). n°71.

<[http://www.bases-publications.com/print/e-docs/00/00/04/04/document\\_article.phtml](http://www.bases-publications.com/print/e-docs/00/00/04/04/document_article.phtml)>

*L'auteure explique que l'enjeu du Web sémantique est de faciliter et d'améliorer la recherche grâce à une représentation du sens du contenu. Elle définit en quoi consiste la structure de base du Web sémantique : langages de représentation ou de programmation, ontologies ou thésaurus et schémas RDF.*



[11] PIERRE Béatrice. L'avenir des langages documentaires dans le cadre du Web sémantique : conception d'un thésaurus iconographique pour le Petit Palais. Mémoire pour l'obtention du titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2010, 115 p.

*Ce mémoire est une étude sur les langages documentaires et leur utilité pour structurer les informations dans l'idée d'un Web sémantique. Cette réflexion se place dans un contexte muséologique de mise en valeur d'un fonds iconographique.*

[12] POUPEAU Gautier. Perspective pour les référentiels à l'heure du Web de données. *In* Journée d'étude AFNOR CG 46/BNF (en ligne), 27 mai 2011 (consulté le 17 août 2011). 32 p.

< <http://www.slideshare.net/AntidotNet/perspectives-pour-les-rferentiels-lheure-du-web-de-donnes> >

*Cette présentation définit les apports des référentiels constituant l'outil sémantique Antidot (normalisation, multilinguisme, désambiguïsation, hiérarchisation, liaison des données/informations).*

[13] TILLET Barbara B., HARPER Corey. Les vocabulaires contrôlés de la Bibliothèque du Congrès, le fichier d'autorité international virtuel et leur application au Web sémantique. *In* World Library and Information Congress : 73rd IFLA General Conference and Council (en ligne), 19-23 août 2007 (consulté le 17 août 2011), Durban, Afrique du Sud.

<[http://archive.ifla.org/IV/ifla73/papers/147-Tillet\\_Harper-trans-fr.pdf](http://archive.ifla.org/IV/ifla73/papers/147-Tillet_Harper-trans-fr.pdf)>

*Cet article fait le point sur les vocabulaires contrôlés de la Bibliothèque du Congrès et présente le projet VIAF (Fichier d'autorité international virtuel). Il y a tout lieu de penser que les vocabulaires contrôlés pourront rapidement être les éléments constituant du Web sémantique. Au niveau international il sera possible de lier les données fiables des notices d'autorité de chaque pays pour le bénéfice des utilisateurs au niveau mondial. Cet article définit également les technologies et connaissances qui structurent le « Web 3.0 » (SKOS, RDA, FOAF, FRBR, OCLC etc.).*

[14] TRONCY Raphaël. Nouveaux outils et documents audiovisuels : les innovations du Web sémantique. Documentaliste-Sciences de l'information, 2005, vol.42, n°6, p.392-404.

*Présentation des outils et des technologies du Web sémantique (langages RDF et OWL, ontologies, URI, FOAF). Selon l'auteur l'utilisation conjointe de ces technologies avec les outils documentaires traditionnellement employés en indexation multimédia est efficace, il cite plusieurs exemples d'applications concrètes. Il estime que la sémantique permet une plus grande précision des requêtes et des réponses, ainsi qu'une possibilité de partage des ressources.*

### **Recherche fédérée**

[15] BIBLIOPEDIA. SRU/SRW (en ligne). (consulté le 12 juillet 2011).

<<http://biblio.wikia.com/wiki/SRU/SRW>>

*Définition des protocoles SRU/SRW permettant l'interrogation synchrone de plusieurs répertoires de ressources en ligne. Ils succèdent progressivement au Z39.50.*

[16] BREBION Patrick, BEL Laurent. Les portails entrepôts de données vont se généraliser. Archimag, décembre 2008-janvier 2009, n°220, p.25.

*Laurent Bel définit comme l'essence même des portails la mutualisation de ressources souvent hétérogènes administrées dans des solutions métier spécifiques avec des acteurs, des méthodologies d'indexation et des vocabulaires très différents. Pour rendre interrogeable de façon fédérée ces contenus, il distingue deux approches : le rapatriement des informations pertinentes de façon synchrone (Z3950, ODBC, SRU/SRW) et asynchrone (OAI-PMH) ou l'utilisation de dataware-houses (entrepôts de données) qui centralisent des systèmes métier aux structures hétérogènes (on n'entrepose pas des données, mais une réplique des systèmes les gérant).*

[17] CZAPLINSKI Jean-Marc, MORET Yves. Z39.50. Bulletin d'informations de l'ABF (en ligne). 1997, n°174. (consulté le 13 juillet 2011), p.77-81.

<<http://www.enssib.fr/bibliotheque-numerique/revues/afficher-45392>>

*Définition du protocole Z39.50 permettant l'interrogation synchrone de plusieurs répertoires de ressources en ligne.*

[18] DALBIN Sylvie, GIRAUD Odile. SRU/SRW en quelques mots. Documentaliste-Sciences de l'information, 2008, vol.45, n°3, p.26-27.

*Présentation et définition de ces deux protocoles de recherche.*

[19] GAUTHIER Mélanie. Le protocole OAI-PMH et les fonctionnalités de recherche : étude de portails du domaine patrimonial. Mémoire pour l'obtention du titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2007, 135 p.

*Le protocole OAI-PMH a été rapidement adopté après son apparition par de nombreuses institutions culturelles désireuses de créer des ponts entre leurs collections et de se rendre plus visibles sur le Web par l'intermédiaire de portails permettant l'accès à leurs ressources disséminées. Cependant, cette mise en commun dans des index centralisés de métadonnées de provenance et de type divers pose des questions en terme de recherche. Ce travail a pour objectif d'identifier les problèmes les plus fréquemment rencontrés et de fournir des préconisations pour y remédier.*

[20] GIUSTI Aurélia. La recherche fédérée des portails patrimoniaux : quelles solutions documentaires ? L'exemple du MuCEM. Mémoire pour l'obtention du titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2009, 110 p.

*Ce mémoire a pour objectif de présenter les outils utilisés dans la recherche fédérée puis d'analyser les solutions documentaires choisies par différentes institutions culturelles pour répondre à la question de l'harmonisation des mots-clés. La dernière partie fournit des préconisations, à partir d'un exemple concret, pour le futur portail documentaire du MuCEM. Pour finir, des pistes de réflexion sont amorcées concernant les technologies du Web sémantique et les possibilités de mettre en correspondance différentes terminologies.*

[21] LAHARY Dominique. La norme Z39.50 (en ligne). Mis en ligne en octobre 1998, dernière mise à jour le 18 mars 2003 (consulté le 12 juillet 2011).

<<http://www.lahary.fr/pro/z3950/index.htm>>

*Définition du protocole Z39.50 permettant l'interrogation synchrone de plusieurs répertoires de ressources en ligne.*

[22] MICROSOFT DEVELOPER NETWORK. ODBC (Open DataBase Connectivity) (en ligne). Dernière mise à jour le 30 novembre 1999 (consulté le 17 juillet 2011).

<<http://msdn.microsoft.com/fr-fr/library/bb469866.aspx>>

*Définition de l'application ODBC qui permet l'interrogation synchrone de plusieurs répertoires de ressources en ligne.*

[23] NAWROCKI François. Le protocole OAI et ses usages en bibliothèque (en ligne). Mis en ligne le 28 janvier 2005, dernière mise à jour le 15 février 2005 (consulté le 10 juillet 2011).

<<http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>>

*Définition du protocole OAI permettant l'interrogation asynchrone de plusieurs répertoires de ressources en ligne.*

[24] STERN David. Harvesting : power and opportunities beyond federated search. Online, 2009, vol.33, n°4, p.35-37.

*L'auteur nuance les bienfaits de la recherche fédérée car selon lui la pertinence des résultats n'est pas toujours au rendez-vous sans un travail de réindexation des ressources hétérogènes en amont. Il suppose donc que la recherche fédérée doit impérativement s'appuyer sur un index prenant en compte les relations (hiérarchiques, équivalences) entre les mots-clés pour être vraiment efficace.*

## Portail

[25] BERNARDINI Michel, GICQUEL Florence. Outils de recherche et analyse sémantique : le portail de BNP Paribas. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.40-41.

*Lancement du projet LEOnard en janvier 2003 ayant pour but d'accéder facilement, à travers un portail, à des sources hétérogènes, externes comme internes (Internet, intranet, bases métier, bases Notes et bases GED). L'outil doit permettre de fédérer des informations disparates et de présenter des résultats triés et exploitables.*

[26] BOURGEAUX Laure, FRESNEAU Amélie. Moteur ou labyrinthe ? Le portail documentaire de la Bibliothèque publique d'information évalué par ses utilisateurs. BBF, 2009, n°6, p.73-77.

*Les auteurs s'interrogent sur l'expérience de la recherche chez les utilisateurs du portail de la BPI. Ce portail a été développé dans le but de permettre la recherche sur les différents catalogues et bases de données de la bibliothèque, mais aussi pour proposer une interface simple et intuitive. Les tests ont été réalisés auprès de trois types de public : les novices, les habitués et les professionnels qui composent l'essentiel des visiteurs du site. Les auteurs en déduisent la nécessité de créer un processus participatif et itératif et non une navigation « en aveugle » ou au contraire totalement dirigiste.*

[27] BREBION Patrick, TEXIER Bruno. La culture monte au portail. Archimag, décembre 2008-janvier 2009, n°220, p.18-22.

*Présentation des portails comme des vitrines pour les institutions culturelles. Selon les auteurs on a récemment assisté à une multiplication des portails culturels. L'essor de ces portails est corrélé à la numérisation des collections qui disposent ainsi d'un espace virtuel d'exposition. La fédération des ressources à interroger permet de mettre en valeur des fonds méconnus. Cette fédération pose néanmoins des problèmes d'ordre technique.*

[28] BREBION Patrick. Le Crédit Agricole unifie l'accès à sa documentation. Archimag, février 2010, n°231, p.28-30.

*Le portail du Crédit Agricole lancé en 2009 unifie l'accès à la quasi-totalité de sa documentation et propose une fonction de recherche transverse. Cet article présente les modalités d'administration d'un portail. Malgré une fonction de recherche fédérée, pas*

*d'élaboration de thésaurus mais paramétrage, par le logiciel Polyspot, d'un dictionnaire de noms propres notamment pour les acronymes.*

[29] COURTOIS Raphaèle. Dynamisation d'un portail documentaire et valorisation de produits documentaires pour des avocats : le cas du cabinet Lefèvre Pelletier et associés, Avocats. Mémoire pour l'obtention du titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2008, 118 p.

*Ce mémoire apporte une définition d'un portail, de son utilité, des besoins auxquels il répond, des éléments dynamiques où à dynamiser (le visuel, le contenu, la navigation, le moteur de recherche etc.).*

[30] DASSA Michèle, PIERRAT Marie-Joséphine. Le TGE Adonis, un outil pour les sciences humaines et sociales. Documentaliste-Sciences de l'information, 2008, vol.45, n°2, p.4-6.

*Le portail TGE Adonis peut être défini comme une volonté de mutualisation des ressources en sciences humaines et sociales afin de faciliter le travail des chercheurs et de favoriser la publication simplifiée de contenus par la communauté scientifique.*

[31] MAISONNEUVE Marc. Bâtir un portail de bibliothèque ou de centre de documentation, un morceau de bravoure ? Documentaliste-Sciences de l'information, 2007, vol.44, n°3, p.243-247.

*La mise en place d'un portail repose sur le souhait de renouer le contact avec le public en lui offrant une interface nouvelle de recherche. Néanmoins l'élaboration d'un portail suppose de réfléchir à la gestion des contenus (l'outil CMS) et la recherche fédérée ainsi qu'à l'interaction entre ces deux composants.*

[32] NUTTIN Guillaume. Bibliothèques et musées : le meilleur des deux mondes en ligne. Archimag, décembre 2009-janvier 2010, n°230, p.34-36.

*Le Web fait sortir les musées et les bibliothèques de la logique de vitrine pour adopter une logique relationnelle, avec une attitude plus centrée sur l'utilisateur que sur le fonds. Les nouveaux enjeux pour les institutions culturelles sont désormais de communiquer et de valoriser le patrimoine de manière interactive.*

[33] NUTTIN Guillaume. L'enfant, point de convergence d'une réflexion sur l'accessibilité. Archimag, décembre 2008-janvier 2009, n°220, p.24.

*L'auteur affirme que les espaces consacrés aux jeunes publics sur les portails culturels s'élargissent (Musée des Arts et Métiers, MET, Louvre etc.) mais qu'il est nécessaire de réfléchir à une ergonomie simplifiée car même si les enfants sont des « digital natives » leurs compétences en recherche sont limitées.*

[34] NUTTIN Guillaume. Mettez un CMS dans votre gestion documentaire. Archimag, juillet-août 2010, n°236, p. 26-27.

*La solution CMS constitue une solution peu onéreuse, voire gratuite, et surtout simple pour publier du contenu en ligne. Certains logiciels peuvent même permettre une gestion documentaire quand les besoins ne sont pas trop complexes et exigeants. De ce fait, en alliant publication de contenus et gestion documentaire simplifiée, certains CMS (Drupal, Joomla, EZ Publish, Jahia etc.) proposent une solution de type portail accessible d'un point de vue technique et financier.*

[35] SAINT-GEORGES Audrey. Un portail pour améliorer l'information : l'exemple de la Bibliothèque HEC. Mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2008, 87 p.

*L'auteure présente le portail comme un moyen pour les bibliothèques de permettre aux utilisateurs de retrouver l'information utile face à la masse croissante des données sur le Web. En effet c'est la connaissance des besoins des utilisateurs ainsi que les savoir-faire info-documentaires qui confèrent aux personnels de bibliothèque de l'importance dans la mise en place d'un portail. Elle définit également les fonctionnalités d'un portail ainsi que les réflexions à mener au sujet de sa mise en place.*

[36] SCOPSI Claire. Construire un portail ! Oui, mais comment ? Documentaliste-Sciences de l'information, 2008, vol.45, n°4, p.16-17.

*Claire Scopsi présente les objectifs d'un portail (fédérer les accès aux ressources, traiter du contenu) ainsi que les questions à se poser avant de le développer. Elle insiste sur la question de la recherche fédérée sur des bases hétérogènes internes et externes.*

[37] TOSCA Consultants, MAISONNEUVE Marc, TOUITOU Cécile. Logiciels portails pour bibliothèques et centres de documentation. L'offre d'outils de recherche fédérée et de gestion de contenu. Paris, ADBS, 2007. 215p.

*Les auteurs présentent dans cet ouvrage les fonctionnalités des portails, leurs composantes et leurs interactions avant de dresser un panorama des solutions proposées par une dizaine d'éditeurs sur le marché.*

[38] TRILLAUD Sandrine. Mutualiser les informations documentaires de deux intranets en un portail unique et fédérateur : cas de la conception du portail documentaire d'OSEO. Mémoire pour l'obtention du titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2006, 116 p.

*Définition d'un portail documentaire comme un outil fédérateur. Sont présentées ses forces et ses faiblesses ainsi que ses apports.*

## **Thésaurus**

[39] BOURDON Françoise, AMAR Muriel. Référentiels, données d'autorité, thésaurus, ontologies, taxonomies...Pour en savoir plus ! Documentaliste-Sciences de l'information, 2008, vol.45, n°3, p.14-27.

*Le Web sémantique s'appuie sur des référentiels dans la mesure où ceux-ci structurent l'information. L'essor du Web sémantique conduit l'AFNOR (dont Françoise Bourdon) à réfléchir à l'interopérabilité des référentiels.*

[40] CHICHEREAU Dominique *et al.* Les normes de conception, gestion et maintenance de thésaurus : évolutions récentes et perspectives. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.66-74.

*Cet article (publié en 2007) fait suite à la révision de deux normes (américaine et britannique) d'établissement et de développement de thésaurus. Le groupe de travail avance la nécessité de réviser la norme française et internationale pour répondre aux problématiques actuelles des langages documentaires.*



[41] DALBIN Sylvie. Instrument de recherche sur le Web. *In* INRIA. La recherche d'information sur les réseaux. Paris, ADBS Editions, 2002. Chapitre 1, p.11-70.

*Sylvie Dalbin souligne la complexité de la recherche d'information sur le Web face à l'accroissement constant d'informations. Participer à la production d'information, dans ce contexte du Web, avec une visée d'optimisation de la recherche sous-entend d'assimiler les concepts d'hypertexte, de granularité de l'information et de sémantique afin d'optimiser la recherche de cette même information. D'où le développement de travaux de normalisation parallèles à un besoin de mutualisation des méthodes. Il suffirait selon l'auteure d'appliquer les pratiques traditionnelles de l'information-documentation en réfléchissant à la question de l'interopérabilité entre les systèmes et à la prise en compte des normes.*

[42] DALBIN Sylvie. Thésaurus à la recherche. In MENON Bruno. Journée d'étude ADBS : Optimiser l'accès à l'information, une opportunité pour les langages documentaires ? (en ligne). 20 septembre 2007, Paris. Document créé le 7 septembre 2007, mis à jour le 27 septembre 2007 (consulté le 12 août 2011).

<[http://www.atd-doc.com/xmedia/publications/2007\\_DALBIN\\_JE-2007-09-20-text-v3b.pdf](http://www.atd-doc.com/xmedia/publications/2007_DALBIN_JE-2007-09-20-text-v3b.pdf)>

*L'auteure affirme que l'évolution du paysage documentaire a un impact sur le rôle des thésaurus et sur celui des vocabulaires contrôlés en général : baisse des activités d'indexation au profit d'autres activités. Pourtant il est primordial selon elle d'associer les technologies avec les vocabulaires contrôlés afin de permettre une recherche pertinente sur Internet.*

[43] DALBIN Sylvie. Thésaurus et informatique documentaires : des noces d'or. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.76-80.

*Le concept de thésaurus est apparu pour faire face à l'augmentation de la production documentaire à partir des années 1950 et au besoin d'organiser l'accès à cette information. L'avènement et la démocratisation du Web et de son mode de recherche de l'information a occulté l'utilité des langages contrôlés. Toutefois la tendance est récemment à l'hybridation entre interface de recherche web (plus intuitive) et structuration de l'information à partir de langages contrôlés en back office.*

[44] DALBIN Sylvie. Thésaurus et informatique documentaires : partenaires de toujours ? Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.42-55.

*Sylvie Dalbin rappelle ici la place et le rôle souvent occultés du thésaurus dans divers logiciels documentaires pour l'indexation, la recherche et la maintenance terminologique. Elle remarque qu'il est nécessaire, face à des chercheurs d'information qui sont de moins en moins des professionnels de l'information, de simplifier les interfaces de recherche avec par exemple le développement de recherche plein texte plutôt qu'une interface de recherche reposant sur un thésaurus car jugé complexe.*

[45] DESFRICHES-DORIA Orélie, ZACKLAD Manuel. Améliorer la recherche d'information à l'aide d'un thésaurus *ad hoc* : expérimentations et réflexions méthodologiques. Document numérique, 2010, vol. 13, n°2, p.13-40.

*La mise en place d'un thésaurus ad hoc voit ses fondements méthodologiques exposés par les auteurs. Celui-ci implique la pertinence d'une coopération entre les différents SOC (Systèmes d'Organisation des Connaissances). Les auteurs développent ensuite une méthodologie de conception d'un thésaurus ad hoc.*

[46] FEYLER François. De la compatibilité à l'interopérabilité en matière de repérage d'information pertinente : la problématique et l'exemple d'OTAREN. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.84-92.

*Cet article aborde la question de la compatibilité et de l'interopérabilité des langages documentaires, conditions d'un juste repérage de l'information pertinente dans des sources de plus en plus nombreuses et volumineuses. François Feyler explique ici que, lorsqu'on est confronté à une multiplicité de systèmes d'indexation, il faut envisager un système d'équivalences terminologiques entre les représentations de concepts présentes dans chacun des langages contrôlés, et ce sans passer obligatoirement par une phase de modification de l'indexation initiale. Ce qu'illustre le projet OTAREN développé par le CNDP pour l'Éducation Nationale.*

[47] HJORLAND Birger. Semantics and Knowledge Organisation. Annual Review of Information Science and Technology, 2007, vol.41, p.367-405.

*L'auteur présente la sémantique comme l'essence des réflexions des professionnels de l'information quant à l'organisation des connaissances. L'auteur définit l'utilité des thésaurus ainsi que leurs fonctions.*

[48] MENON Bruno. Journée d'étude ADBS : Optimiser l'accès à l'information, une opportunité pour les langages documentaires ? Documentaliste-Sciences de l'information, 2007, vol.44, n°6, p.385-388.

*Bruno Menon livre un compte-rendu de cette journée d'étude au cours de laquelle il ressort un état des lieux des pratiques de recherche sur Internet par les utilisateurs. Il est pour lui indéniable que les opportunités pour les langages documentaires existent à l'heure où il devient nécessaire d'optimiser la recherche d'information sur Internet : cela nécessite qu'ils soient adaptés et intégrés à des dispositifs de recherche forcément flexibles et multimodaux.*

[49] NUTTIN Guillaume. Thésaurus : l'outil des documentalistes gardiens du trésor. Archimag, novembre 2008, n°219, p.30-32.

*Cet article dresse un panorama de l'offre de logiciels de gestion de thésaurus qui sont dorénavant influencés par l'essor des technologies sémantiques. Ces logiciels se distinguent en deux catégories : les modules de gestion de thésaurus d'une solution de type SIGB d'une part et les logiciels autonomes d'autre part. Ces logiciels ont tous des fonctionnalités en commun : outils de création, d'importation, d'export, d'impression, outils de contrôle de descripteurs et de leurs relations qui est une fonctionnalité essentielle.*

[50] RABAULT Hélène, ZYSMAN Hélène. Les schémas de concept et le Web sémantique : la norme ISO 25964 et le Web sémantique. In Journée d'étude AFNOR CG 46/BNF (en ligne), 27 mai 2011 (consulté le 17 août 2011). 17 p.

<[http://www.bnf.fr/documents/afnor2011\\_norme\\_thesaurus.pdf](http://www.bnf.fr/documents/afnor2011_norme_thesaurus.pdf)>

*Présentation de la norme ISO 25964 qui a pour objet de rendre un thésaurus interopérable avec d'autres vocabulaires contrôlés, ce qui est primordial pour préserver l'intérêt des thésaurus face à l'essor du Web sémantique et de ses standards ainsi que dans l'optique d'une mutualisation des ressources sur le Web.*

[51] VIGNON Marie-Edith. Des outils d'indexation pour alimenter en réseau une base de données : l'exemple de la BDSP. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.93-95.

*La banque de données de la santé publique est le fruit d'un travail collaboratif dans le cadre d'un réseau de coopération documentaire constitué d'une quarantaine d'organismes. La BDSP s'appuie sur un thésaurus fédérateur qui n'oblige pas les participants à s'y conformer pour leur propre gestion documentaire. En effet chaque membre peut conserver son propre vocabulaire pour des raisons historiques, pour la spécificité de sa discipline couverte etc. et ce grâce à l'existence d'une table de correspondances entre chaque terme du vocabulaire spécifique de la BDSP et les différents descripteurs concernés sur les autres thésaurus. Cela débouche sur l'utilisation d'un vocabulaire contrôlé permettant une optimisation de la recherche sur des ressources hétérogènes, mais nécessite toutefois du temps et un travail humain conséquent.*

[52] ZACKLAD Manuel, GIBOIN Alain. Systèmes d'organisation des connaissances hétérogènes pour les applications documentaires. Document numérique, 2010, vol.13, n°2, p.7-12.

*Définition des Systèmes d'organisation des connaissances (SOC ou KOS en anglais). Ceux-ci visent à définir des principes de description d'un domaine pour faciliter des opérations de classement et de recherche d'items plus ou moins abstraits : documents, personnes, lieux ou activités. Le souci des SOC est qu'ils sont répartis entre différents champs disciplinaires et secteurs professionnels. A chaque champ ou chaque secteur son propre SOC avec ses théories, ses pratiques de construction. La problématique est donc d'être en mesure de proposer une complémentarité de SOC hétérogènes dans le cadre de projets interinstitutionnels de portails de ressources.*

[53] ZYSMAN Hélène, RABAULT Hélène. Une nouvelle norme pour le thésaurus : pourquoi une nouvelle norme ? (en ligne), 15 février 2011 (consulté le 20 août 2011).

<<http://www.adbs.fr/une-nouvelle-norme-pour-le-thesaurus-1-pourquoi-une-nouvelle-norme--100316.htm?RH=1266334869518>>

*Intervention vidéo d'Hélène Zysman et Hélène Rabault qui présentent les principales nouveautés apportées par la norme ISO 25964, nouvelle norme internationale sur le thésaurus en cours d'élaboration et permettant l'interopérabilité entre les différents langages contrôlés à l'heure du Web sémantique et de la mutualisation des ressources.*

## Ontologie

[54] BOULIER Serge. Quoi de neuf sur les moteurs de recherche ? Documentaliste-Sciences de l'information, 2008, vol.45, n°3, p.20-21.

*L'auteur fait un état des lieux sur les moteurs de recherche existants (2008) en présentant le Web sémantique, ou plutôt Web de données, comme le futur du Web et donc de la recherche car celui-ci pourrait être à l'avenir articulé comme une gigantesque ontologie permettant de résoudre les problèmes grammaticaux, sémantiques et qualitatifs des résultats de recherche. L'auteur pose néanmoins comme prérequis l'évolution des normes actuelles vers un modèle unifié et réellement sémantique.*

[55] CHAUMIER Jacques. Les ontologies : antécédents, aspects techniques et limites. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.81-83.

*Selon l'auteur, qui définit le concept d'ontologie et de langages contrôlés, le développement du Web a accru le besoin d'outils de description des contenus des documents et des ressources : les ontologies sont une nouvelle famille d'outils linguistiques apparus à la fin des années 1990. Les travaux sur les ontologies ont connu un essor avec le Web sémantique. Les ontologies et les thésaurus se ressemblent dans la mesure où ils sont des vocabulaires structurés, utilisés et validés par des acteurs d'un domaine. Dans les deux cas le vocabulaire est doté de relations sémantiques entre les termes qui le composent. La différence fondamentale entre ontologie et thésaurus est que les ontologies reposent sur des sémantiques formelles qui ne sont pas destinées à être interprétées par des humains mais par des machines.*

[56] CYROT Catherine, PREUSS Christian. Réingénierie de thésaurus : une étude de cas. Documentaliste-Sciences de l'information, 2009, vol.46, n°3, p.4-13.

*Les auteurs se posent la question de savoir comment utiliser un thésaurus comme passerelle pour échanger entre organismes partenaires du même domaine d'activité et comment faire évoluer son utilisation vers des outils du Web sémantique. Les auteurs proposent de retravailler les relations associatives entre mots-clés afin de constituer des ontologies précieuses pour une application de type sémantique.*

[57] DESFRICHES-DORIA Orélie. La recherche d'information du grand public : l'évolution des langages documentaires ou l'avenir des ontologies : création d'un thésaurus dynamique pour le site de l'assistance en ligne de France Télécom Orange. Mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2008. 115 p.

*L'auteure propose une typologie et un historique des langages documentaires, présente l'étude préalable à la création d'un thésaurus avant de s'intéresser à un nouveau type de langages documentaires : les ontologies au sein de leur contexte d'utilisation privilégié qu'est le Web sémantique. L'auteure pointe les problèmes spécifiques liés à l'élaboration de ces nouveaux outils et rapproche thésaurus et ontologies.*

[58] JACOB Elin K. Ontologies on the Semantic Web. Bulletin of the American Society for Information Science and Technology, avril/mai 2003, vol.29, n°4, p.19-22.

*Elin K. Jacob définit le concept d'ontologie concernant le Web comme la structure qui modèle la sémantique d'un domaine de connaissances dans un langage compréhensible par une machine. Dans le domaine du Web sémantique, une ontologie est une conceptualisation partielle d'un domaine de connaissance partagée par des utilisateurs. Cet article définit également l'architecture RDF de la (re)présentation des connaissances sur le Web.*

[59] LEGG Catherine. Ontologies on the Semantic Web. Annual Review of Information Science and Technology, 2007, vol.41, p.407-451.

*Catherine Legg présente le concept d'ontologie comme l'essence du Web sémantique dans la mesure où il permet de structurer des domaines de connaissance de façon à rendre des requêtes intelligibles par des machines.*

[60] LENART Michèle. SKOS, un langage de représentation de schémas de concepts. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.75.

*Définition du langage SKOS qui est un langage de représentation de schémas de concepts tels que les thésaurus, les classifications, les taxonomies etc. Il propose un système permettant d'exprimer et de gérer des modules interprétables par les machines dans la perspective du Web sémantique. Il est défini comme simple par opposition au standard OWL qui est plus à même de représenter des structures sémantiques plus riches telles que les ontologies.*

[61] MENON Bruno. Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan. Documentaliste-Sciences de l'information, 2007, vol.44, n°1, p.18-28.

*Bruno Menon étudie la formation et l'évolution des langages documentaires sur un siècle en appréciant leurs qualités et leurs limites. Puis il examine les taxonomies et les ontologies, nouveaux outils de recherche d'information avec l'explosion des technologies internet et les mutations des pratiques informationnelles qui en découlent.*

[62] MENON Bruno. Les systèmes d'organisation des connaissances de Babel aux Linked Data. In Journée d'étude AFNOR CG 46/BNF (en ligne), 27 mai 2011 (consulté le 14 août 2011). 25 p.

<[http://www.bnf.fr/documents/afnor2011\\_intro\\_menon.pdf](http://www.bnf.fr/documents/afnor2011_intro_menon.pdf)>

*L'auteur dresse une histoire des SOC depuis le début de la civilisation à nos jours en affirmant que les motivations dans l'élaboration des SOC sont demeurées identiques à l'heure du Web sémantique, c'est-à-dire se représenter des connaissances hétérogènes de façon structurée et permettre ainsi l'accès facilité à celles-ci. Il définit toutes les formes que peuvent prendre les SOC : thésaurus, classifications bibliographiques, listes des vedettes matière en documentation/bibliothèques, dictionnaires, terminologies en linguistique, taxonomies d'entreprise, ontologies, modèles de données, bases de connaissances en technologie de l'information etc.*

[63] MENON Bruno. Journée d'étude ADBS-IUT Paris V : Le Web sémantique : de nouveaux enjeux documentaires ? Documentaliste-Sciences de l'information, 2003, vol.40, n°6, p.387-391.

*L'auteur présente les fondements du Web sémantique (métadonnées, ontologies etc.), ses apports ainsi que les enjeux qu'il représente déjà en 2003 pour les professionnels de l'information-documentation (standardisation des formats d'encodage des informations, des métadonnées et des ontologies).*

[64] VALDERRAMA Asuncion. Web 3.0 et recherche sémantique : du vin vieux dans des outres neuves ? Documentaliste-Sciences de l'information, 2009, vol.46, n°3, p.20-21.

*Selon l'auteure, le « Web 3.0 » permet au professionnel de l'information de faire face au déluge d'informations généré par le Web 2.0. Elle précise que cette « solution miracle »*

*nécessite en back office un travail énorme de création d'ontologies, d'annotations, de mises en relation et d'organisation réalisé en grande partie par des humains. Selon elle le Web sémantique correspond à un système d'information post-moderne, c'est-à-dire que la médiation du professionnel de l'information-documentation entre l'information et l'utilisateur est cachée (elle est explicite dans le système classique et absente dans le système moderne). Cette médiation requiert des savoir-faire traditionnels (d'où le titre de l'article).*

### **Le contexte info-documentaire au sein du Musée National du Sport**

[65] BERTRAND Laurène. Réflexion sur la définition et l'organisation de la documentation au sein d'un musée : quel service d'information et de documentation au Musée National du Sport ? Mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2008, 141 p.

*Ce mémoire définit ce qu'est la documentation dans un musée et élargit cette définition à toute l'information circulant au sein d'un musée. Le but étant de proposer des solutions à l'organisation de cette information multiple et complexe. Ensuite une analyse marketing de l'information et de la documentation du Musée National du Sport permet de mettre en exergue les opportunités pour le futur musée. Enfin l'auteure apporte des conclusions et propose des scénarios pour la mise en place d'un service d'information et de documentation au sein du Musée National du Sport.*

[66] KROUCH-GUILHEM Circé. Appréhender la valorisation d'un fonds iconographique de musée à partir d'une démarche marketing : quels produits et quels services pour la phototèque du Musée National du Sport ? Mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire ». Paris : INTD, 2009, 108 p.

*L'auteure de ce mémoire vise à élaborer une méthodologie de travail ainsi qu'une méthodologie d'identification et de construction de ses outils, inspirées du marketing documentaire afin d'oeuvrer à la conduite d'un projet de valorisation de fonds iconographique. Cette méthodologie est illustrée d'applications concernant le cas spécifique de la photothèque du Musée National du Sport, alors en cours de constitution et de légitimation.*





# Annexes

# Annexe 1 Questionnaire pour les professionnels chargés de conduire le projet de portail de leur institution

Nous avons mené, outre un audit sur les portails d'institutions culturelles, une enquête auprès des personnes ayant mené, ou du moins participé au pilotage, le projet de portail au sein de leur institution. La quasi-totalité de ces entretiens a été menée par téléphone, deux entretiens physiques (avec les documentalistes de la Cité de l'Immigration et du Musée du Quai Branly) ont pu être réalisés.

## Origine du projet

*-Quand a été mené le projet de site de votre structure ?*

*-Pour quelles raisons avez-vous décidé de lancer ce projet ? (se substituer à un existant ? premier projet de ce genre ?)*

- Valoriser votre fonds et vos activités sur Internet
- Enrichir les activités de votre structure (nouveaux services)
- Se substituer à un site existant pour l'améliorer
- Autres raisons :

*-A quel public souhaitez-vous vous adresser avec ce site ?*

- Tous publics
- Chercheurs/experts
- Jeunes
- Adultes
- Autres

*-Quelles ont été pour vous les fonctionnalités à prioriser pour votre site ?*

## **La dimension humaine du projet**

*-Quelle a été la répartition des rôles dans la conduite du projet ? Combien de personnes ont-elles été mobilisées en interne ? Combien de jours ouvrés par semaine et par personne ?*

*-Avez-vous communiqué sur ce projet en interne ?*

- Oui
- Non

*-Si oui, comment ?*

- Réunions
- Intranet
- Mails
- Réseaux sociaux
- De façon informelle (pauses café, repas)
- Autre

*-Quels ont été les rapports entre les membres du projet en interne ?*

- Très bons rapports : tous les membres de l'équipe ont participé volontairement au projet, les échanges ont été globalement constructifs
- Assez bons rapports : certains membres de l'équipe ne se sont pas montrés convaincus au début du projet, mais la majorité a finalement adhéré et les échanges, bien que tendus, ont été constructifs
- Assez mauvais rapports : certains membres ont affiché des réticences tout au long du projet, les échanges ont été globalement tendus
- Mauvais rapports : la majorité des membres ne s'est pas sentie concernée par le projet, les échanges ont été houleux

## **La gestion du temps**

*-Combien de temps a été nécessaire à la conduite du projet ? Le planning initial a-t-il été respecté ou dépassé ?*

*-Pourriez-vous nous communiquer le retroplanning de votre projet (le MNS s'engage évidemment à ne pas le communiquer en externe) ? Cela nous permettrait de cerner les différentes étapes du projet, les acteurs mobilisés etc.*

*-Quelles étapes vous ont posé problème ?*

### **Le Budget**

*-De quel budget disposiez-vous pour mener ce projet ? Y-a-t-il eu des ajustements pendant le déroulement du projet ?*

*-Y-a-t-il eu des ambitions qui ont dû être revues à la baisse voire supprimées pour des raisons budgétaires ?*

### **Les relations avec le(s) prestataire(s)**

*-Avez-vous développé ce projet en interne ou avez-vous eu recours à un ou plusieurs prestataires ?*

*-Si oui qui ?*

*-Quels critères de sélection ?*

*-Comment s'est passée la collaboration sur le projet ?*

- Très bien : le prestataire s'est montré réceptif, a su répondre aux besoins définis dans votre cahier des charges, s'est montré force de proposition
- Bien : le prestataire, même s'il n'a pu répondre à toutes les exigences, s'est montré honnête en annonçant dès le début ce qu'il serait ou non capable de faire, a proposé des ajustements. La communication a été possible
- Assez mal : le prestataire n'a pas su répondre à plusieurs de vos exigences, la communication a été plutôt compliquée
- Très mal : le prestataire n'a pas su répondre à la plupart de vos exigences, a dissimulé ses lacunes lorsqu'il a accepté le projet. En outre la communication a été très compliquée

*-Assure-t-il une maintenance et une formation du personnel ?*

- Maintenance et formation
- Maintenance
- Formation
- Rien

### **Question technique**

*-Si vous disposez de plusieurs bases de données pour vos ressources, cela a-t-il posé problème dans le cadre de l'instauration d'une recherche fédérée (à partir d'une interface unique) ? Comment y avez-vous remédié ?*

- Moissonnage OAI
- Z 39.50
- Autres (ODBC)

### **Documents utiles**

*Pourriez-vous nous communiquer les documents suivants ? :*

- Cahier des charges (fonctionnel et/ou technique)
- Budget
- Rapport d'activité
- Note de procédure

