EROS : a European Research Open System

Christian Lahanier et Geneviève Aitken
Centre de recherche et de restauration des musées de France
Département Documentation et Technologie de l'Information
6, rue des Pyramides
75041 Paris Cedex 01
tél : 33 (0)1 40 20 58 71
Christian.Lahanier@culture.fr
genevieve.aitken@culture.fr

Introduction

Toute unité documentaire, bibliothèque, archive ou centre de documentation, a une double mission, dont les termes sont souvent contradictoires : la préservation du document unique (manuscrit, enluminures, dessin, etc.) pour le transmettre et sa communication.
Avec l'apparition du numérique, le support optique ou magnétique s'est substitué au papier et au parchemin. Ainsi, le chercheur comme le simple amateur peut consulter le même document, sans perte d'information et avec une meilleure lisibilité sur écran, s'il est numérisé en haute définition. Grâce à ces nouvelles technologies, le document original est conservé dans des conditions optimales et l'information transmise.

C'est à partir des années 70 que les bases de données se sont développées au niveau international. Autrefois, la quête d'information nécessitait de se déplacer sur le lieu d'archivage, d'interroger des fichiers ou des catalogues manuels dont l'information était classée par auteur, titre ou lieu géographique. Aujourd'hui, le Web permet, quelque soit le site d'archivage de ces bases de données, suite à une interrogation complexe associant simultanément des types d'information tels que l'auteur, le titre, la date, le lieu, de sélectionner le document recherché. De plus, les moteurs de recherche en texte intégral permettent de localiser les occurrences d'un terme dans un texte. Le mode de recherche sémantique et contextuel affine considérablement la qualité des réponses à partir de l'analyse linguistique.

Inscrite dans la société de l'information, la documentation devient un secteur primordial de la recherche. Les documentalistes doivent s'associer aux développements technologiques de l'informatique, du numérique et des réseaux, en apprendre les outils pour les adapter aux besoins.
Ces nouveaux modes d'accès à l'information sont les enjeux d'une évolution sociétale dont les "acteurs" des centres de documentation doivent rester les garants de leur ouverture au plus grand nombre de consultants.

1.- La nature de la documentation scientifique et technique du C2RMF

Pour l'étude et l'analyse des œuvres conservées dans les musées de France, le laboratoire du centre de recherche et de restauration est équipé d'instruments scientifiques : du microscope binoculaire à l'accélérateur de particules. Ces techniques permettent, par exemple, de déterminer les composants chimique de la matière constitutive des objets (bois, céramique, métal, textile, pierre, pigment …) ou de caractériser l'état de conservation de l'œuvre avant intervention par le restaurateur.
Trois types de documents en résultent :

- des images (photographies sous différentes longueurs d'onde, radiographies, coupes stratigraphiques), mais également des graphiques et des spectres,
- du texte : rapports, articles de revues, constats d'état ou notes de commissions de restauration

- et des tableaux de chiffres issus des analyses de composition réalisées avec les techniques spectrométriques.

La documentation est volumineuse et variée. Elle concerne 17.400 peintures et dessins ainsi que 34.000 objets. Après étude en laboratoire soit environ 100 000 photographies (format 18x24 cm), 100 000 radiographies, et 6000 rapports.
A chaque étape importante de la restauration d'une œuvre, des documents témoignent de l'évolution du travail et de la modification -minime soit-elle- de l'œuvre. Ainsi des prises de vue photographiques permettent de retracer les différentes actions du restaurateur, notifiées également dans des rapports où sont mentionnés les matériaux d'intervention utilisés .150 000 diapositives, 20 000 rapports et autant de fiches de santé constituent le fonds documentaire du service de restauration.
La première fonction de ces informations est celle de toute discipline scientifique : étayer le raisonnement qui a conduit à une interprétation ou présidé le choix d'une méthode en permettant de vérifier la justesse du propos ; la seconde est d'illustrer la démonstration par les exemples appropriés.

Le caractère unique de ces documents font la richesse du C2RMF : en effet même la plus légère des interventions fait disparaître quelque chose de l'état originel de l'œuvre; seuls les documents en portent témoignage. Ils peuvent être consultés à tout moment dès qu'une étude ou qu'une intervention s'impose. Constituant un vaste fonds, enrichi progressivement, il devient lui-même –en tant qu'ensemble- un domaine de recherche sur les matériaux constitutifs, les types d'altération ou de traitement.

Des bases de données avaient été développées au laboratoire dans les années 1975, mais conçues plutôt comme outils de gestion des documents. La réflexion commune de quelques responsables de laboratoires européens conscients dans les années 1985 des enjeux de ces développements informatiques et leur volonté de faire circuler l'information scientifique contenue dans des dossiers conservés sur les rayonnages des photothèques institutionnelles ont donné naissance à un ambitieux projet de numérisation et d'informatisation. A la fin des années 1980, la Commission Européenne, convaincue de l'importance de l'apport de ces nouvelles technologies aux applications scientifiques développées dans le cadre des musées a défendu ce projet intitulé NARCISSE (Network Art Research Computer Image SystemS in Europe). Deux axes de recherche le structuraient :
- la mise au point d'un numériseur haute définition, développé par la société Thomson-Broadcast
- la constitution d'un système documentaire multilingue relatif au domaine de la conservation-restauration, en allemand, français et portugais.

Cette volonté de mener en parallèle des recherches sur l'image et sur le texte est une caractéristique persistante du C2RMF puisqu'engagé dans deux nouveaux projets européens, ce double enjeu est au cœur de CRISATEL* et de SCULPTEUR *.
Ainsi pour le premier, l'acquisition de l'image sera effectuée à l'aide d'une caméra à balayage linéaire et filtrage multispectral à la définition de 12000 points par 30000 lignes (cf. article C. Lahanier). En parallèle, l'approche ontologique permettra la lisibilité et l'opérabilité de la base de données EROS
L'analyse d'image et la reconnaissance semi-automatique de contenu par ordinateur devraient permettre d'élaborer de nouvelles formes d'indexation des images pour effectuer des sélections automatiques. Il sera intéressant de confronter ces nouveaux modes d'indexation aux méthodes traditionnelles d'indexation manuelle selon le vocabulaire normalisé, développé précédemment.

2.- Les serveurs HP

3.- L'architecture de la base EROS

Une œuvre peut être associée à plusieurs images ou plusieurs documents, fiche de santé, constat d'état ou mouvement. Un document peut être associé à plusieurs œuvres (cas des rapports d'étude ou des commissions de restauration).

4.- Les logiciels libres et les standards
5.- Les interfaces de recherche et de saisie

Au croisement des savoirs : un vocabulaire multilingue relatif à la conservation-restauration
La mise au point d'un thesaurus multilingue spécifique au domaine de la conservation-restauration reste, dix ans après, un apport majeur du projet NARCISSE. Des séances de travail au niveau international ont réuni trois années durant des spécialistes de la matérialité de l'œuvre d'art, tant pour l'étude en laboratoire qu'en terme de restauration.
Elaboré dès l'origine en allemand, italien, portugais et français, ce choix de concepts sur l'état de l'œuvre, accompagnés des définitions était indispensable pour construire la base de données. La forme thesaurus, défini comme un vocabulaire de termes contrôlés pour l'indexation et structuré de façon à mettre en évidence les relations entre les concepts et normalisé pour sélectionner des occurrences dans une base de données, a été privilégiée car il améliore la qualité de l'information, il élimine les données récurrentes par l'utilisation de la structure hiérarchisée et surtout il fédère des utilisateurs à mettre à jour un système d'indexation.
C'est à partir de l'expérience professionnelle des experts habitués à lire et à interpréter les documents photographiques ou radiographiques, réalisés sur les tableaux, que cette structure a vu le jour. Là réside son originalité : il ne s'agissait pas de décrire l'acte créateur de l'artiste, ou de reprendre les recueils des soi-disant recettes d'atelier (moine Théophile ou le manuscrit de Turquin de Mayerne) encore moins de vouloir recenser la totalité des termes rencontrés dans les dictionnaires (comme d'énumérer tous les types de bois utilisé comme support de l'œuvre). Volontairement restreint (300 mots), ce vocabulaire a été ordonné selon trois critères :

- les éléments constitutifs de l'œuvre
- les altérations
- les interventions

Parmi les techniques d'examen photo et radiographiques une dizaine d'entre elles ont également été définies.

Sachant qu'il est parfois bien difficile de distinguer, à partir de documents visuels, ce qui appartient à l'une ou l'autre catégorie, cette grille d'indexation doit toujours être validée ou remise en question par l'examen direct de l'œuvre. A partir de la description des éléments de l'œuvre originale, des altérations ou des interventions qui ont pu laissé une trace sur le film photo ou radiographique, ce vocabulaire est utilisé pour l'indexation des rapports et des articles.
Premier système documentaire multilingue relatif à la conservation et restauration des peintures et enluminures, il a été publié par les Arquivos Nacionais Torre de Tombo de Lisbonne et ainsi diffusé à l'ensemble de la communauté scientifique.

6.- Le mode de navigation
7.- Les autres moteurs de recherche
8.- Les afficheurs d'images
9.- Les fonctionnalités développées
Conclusion
Bibliographie

Elaboration d'une plate-forme ouverte "Art et Science" pour l'archivage et la recherche documentaire structurant les résultats des études et des restaurations menées depuis plus de 70 ans sur les œuvres d'art.
L'objectif est d'utiliser les nouvelles technologies de l'information pour exploiter la masse de données scientifiques et techniques réalisées au Centre de recherche et de restauration des musées de France (C2RMF) sur le patrimoine des musées, collection d'archives scientifiques uniques qui constitue la mémoire collective de la vie des œuvres.
Les sciences de l'information et de la cognition doivent permettre de gérer l'archivage de masse (7 téra-octets d'images haute définition réalisées sur 17.000 peintures de chevalet et 35.000 objets) de ce patrimoine documentaire unique afin de maîtriser leur contenu subtil et parfois invisible mais présent.
L'exploitation de ces résultats est à même de renouveler les orientations de recherche dans le domaine du patrimoine national.
L'accès multilingue en ouvre la diffusion. L'exploitation informatique de ces données permet d'assurer la préservation du patrimoine culturel et d'en améliorer la compréhension.

Le système EROS du C2RMF

Ce système a été initialisé et développé à partir de 1990 dans le cadre de projets européens pour la conservation et la diffusion des archives photographiques et radiographiques.
Il a été renouvelé en 2001 en collaboration de la mission de la recherche et de la technologie du ministère de la culture. Il utilisant des logiciels libres faisant appel aux technologies du Web et en respectant les nouveaux standards d'interopérabilité et de gestion de contenu. Des techniques avancées de numérisation ont été développées et appliquées pour permettre d'exploiter le contenu des images scientifiques et techniques.
La masse d'information recueillie nécessite aujourd'hui de mettre au point de nouveaux outils d'exploitation en faisant appel aux technologies avancées de reconnaissance de contenu. Cette base utile à la recherche matérielle menée sur les œuvres (technique de fabrication, les matériaux et leur vieillissement etc) sert également à la formation en histoire de l'art et en conservation-restauration ainsi qu'à la préservation des collections.
Ce système est réparti et sera implanté dans le réseau des ateliers de restauration en France afin de partager le savoir et d'assurer la cohérence de la documentation afférente à la connaissance matérielle des collections des musées.
Il sera accessible à partir du site Web du C2RMF http://www.c2rmf.fr en cours de réalisation

Liste des illustrations
L'architecture du système CELARTEM

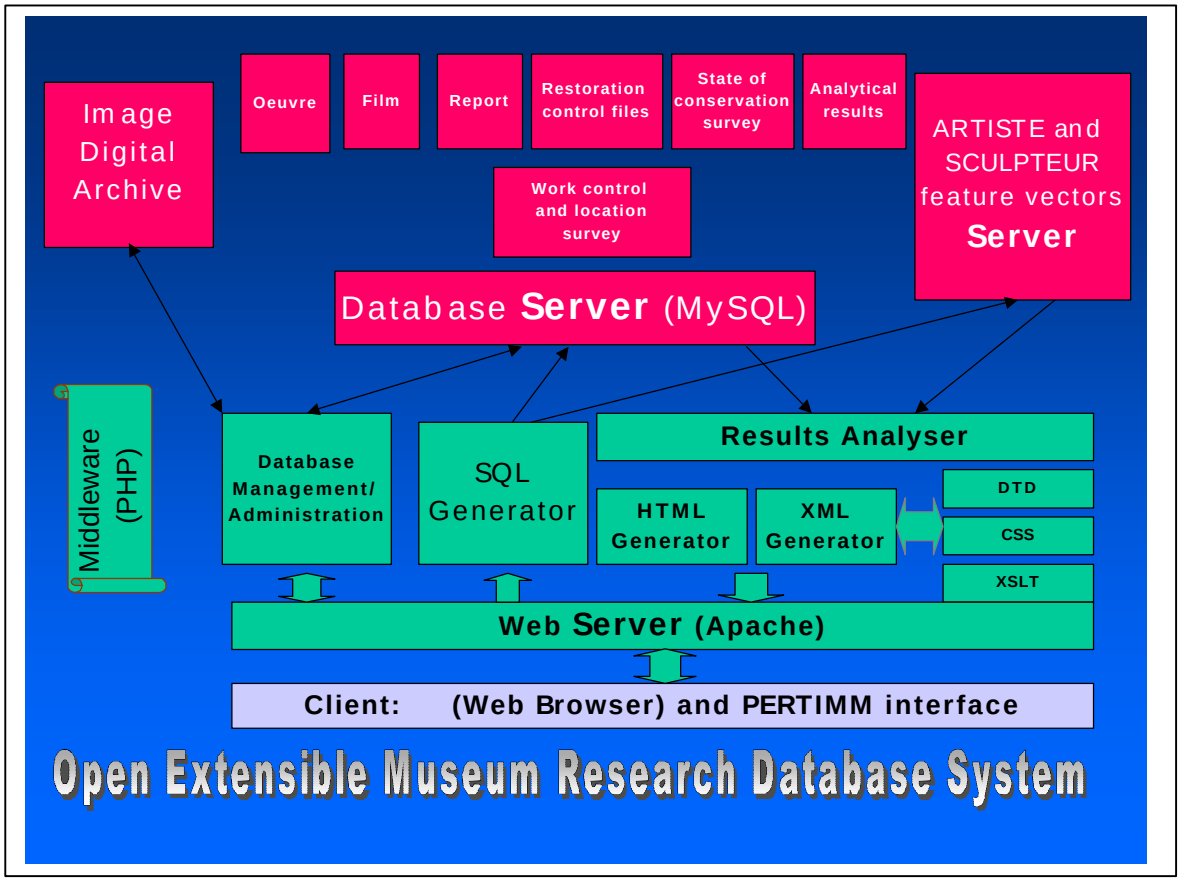

Schéma synoptique de la base EROS

Les interfaces d'interrogation et de saisie avec explication des menus
La gestion des droits
L'implémentation des dictionnaires hiérarchisés (technique pour réaliser une hiérarchie)
La gestion du multilinguisme à partir de codes
PERTIMM
PLS
Icram
Admin
Les serveurs HP

The C2RMF database

- 15.000 paintings and 35.000 objects belonging to 1.200 museums
- 300.000 photographic and X ray scientific files digitised
- 140.000 photographs digitised
- 4.000 scientific study and restoration reports (500 in html format)

Masse de données
1931-2003 : 70 ans d'histoire

Un centre de recherche avec un fonds documentaire exceptionnel sur la matérialité des œuvres :
- 16.000 peintures et 35.000 objets étudiés ou restaurés conservés dans 1200 musées,
- 250.000 photographies grand et moyen format,
- 100.000 radiographies,
- 70.000 analyses physico-chimiques, structurales, isotopiques, minéralogiques et la datation
- 7.000 rapports d'étude et de restauration,
- 8.000 fiches de santé,
- 1.000 rapports de commission de restauration,
- 2.000 articles, thèses, etc…
- 7.000 ouvrages.

La base EROS, développée par Ruven Pillay, comporte aujourd'hui trois catégories d'informations :
- les meta-données historiques, muséologiques et matérielles sur les œuvres,
- les images obtenues par des numérisations de clichés photographiques et radiographiques mais aussi des images électroniques (camera infra-rouge, microscope optique et électronique etc.) et des images numériques par capture directe de l'objet,
- les rapports d'étude et de restauration, fiche de contrôle d'état des œuvre, articles, synthèses électroniques.

- 16.000 peintures et 35.000 objets appartenant à 1.200 musées,
- 220.000 notices de clichés informatisées,
- 140.000 clichés photographiques et radiographiques numérisés
- 4.000 rapports d'étude et de restauration numérisés dont 500 au format html
- 6.600 fiches de santé

A huge amount of digital data
The HP servers configuration

- 2 Hewlett-Packard Ultra-SCSI RAID5 dual 700 MHz serving a total of 6 TeraBytes.
- 1 Hewlett-Packard 6000 quad 550MHz server,
- 3 GB RAM on each machine
- 1 Gbps Ethernet switch linking the servers
- LTO archiving system with 40 tapes each of 200 GB
- 1 DLT recording tape of 80 Gbytes

The server and Web software is installed on the quad server, while the images and databases are distributed across each of the two dual processor RAID servers containing 6 Tbytes.

Logiciels libres et standards utilisés
LINUX version comme operating system
SQL moteur de recherche
MySQL version comme base de données relationnelles
APACHE comme serveur Web
Red Hat version comme xxxxx
JAVA version langage de programmation pour la réalisation des écrans
PERL langage de programmation pour la gestion des script
PHP langage de programmation pour la gestion des tables et des écrans
Mozilla Version comme navigateur
XML langage de programmation pour la modélisation, la visualisation, le stockage et l'acquisition des données. Stockage en XML ou converti automatiquement pour un enregistrement sous SQL server
HTML langage de programmation pour l'édition de pages Web
DTD langage de programmation pour gérer les données dans un texte intégral
JPEG format de compression des imagespour la vignette
TIFF format des images multi-résolution

L'interface est composée de trois parties
1) les informations muséologiques, historiques et matérielles relatives à l'œuvre (39 champs)
2) les informations techniques liées à la gestion des images (20 champs)
3) les informations techniques liées à la gestion des documents textuels, tels que rapports, etc. (20 champs).
4) les informations techniques liées à la régie des œuvres (xxx champs)

Les listes d'autorités multilingues
Les listes d'autorités sont alimentées dynamiquement par le gestionnaire. Chaque terme est associé à un code alphabétique ou alphanumérique (pour les champs multi-critères). Les listes d'autorité dans les autres langues sont élaborées à partir de ces codes qui sont utilisés lors des requêtes SQL.

- gestion de thésaurus
- des documents à partir de listes d'autorité par domaine et par catégorie d'information
- des images à partir de listes d'autorité par domaine et par catégorie d'images (photographie en lumière directe, en lumière rasante, en fluorescence d'ultra-violet, dans l'infra-rouge, aux rayons X, en microscopie optique, en microscopie électronique, etc.)
- des images 2D et 3D à partir de modules d'analyse de forme dans des images (reconnaissance de parquetages, de la forme des châssis etc.)
- des éléments contenus dans des images 2D par détourage et indexation manuelle permettant ensuite de pointer et de localiser l'élément dans l'image
- des images 2D et 3D à partir de vecteurs caractéristiques de leur contenu (couleur, texture, segmentation, proximité etc.)
- classification des images à partir de vecteurs caractéristiques de leur contenu (couleur, texture, segmentation, proximité etc.) et de l'intervention de l'expert qui rejette le classement par niveau de similarité.
- classification de la forme des craquelures de la couche picturale (stable et instable) en prévision de leur évolution et de déterminer leur origine
- recherche de liens sémantiques sur le Web

4 moteurs de recherche
1) SQL opérant sur des tables de metadonnées (textuelles, numériques et ou alphanumériques) et de texte libre.
2) Pertimm opérant sur des index multilingue issus de texte intégral avec interface en français et en anglais:
- en mode simple avec identification de la fréquence et de la localisation du terme recherché dans le texte intégral
- en mode de recherche avancée

3) graphe dynamique 3D d'un modèle de classes ontologiques (SCULPTEUR)
4) ARTISTE et RETIN la liste des vecteurs caractéristiques

4 afficheurs d'images

1) **l'afficheur d'images EROS** permet la visualisation des images haute définition (jusqu'à 12.000x30.000 pixels), multi-spectrales (13 fichiers de l'ultra-violet au proche infra-rouge) multi-résolution (jusqu'à xxxx niveaux de définition) fabriquées avec le logiciel NIP developpé par la National Gallery de Londres (www.xxxxxxxac.uk)
Schéma et description d'une image pyramidale tuilée
Un autre composant essentiel du système de gestion de la base de données consiste à visualiser des images numérisées en haute résolution pouvant atteindre 12.000 pixels x 30.000 lignes et calibrées en couleur représentant des peintures et des vues panoramiques d'objets.
Chaque image est enregistrée avec une vignette au format JPEG et une image pyramidale au format TIFF pour la visualisation de détails. Une applet JAVA permet la consultation d'une image à plusieurs niveaux d'agrandissements en relation avec un serveur "d'images tuilées". Cet afficheur d'images a été développé pour gérer des images 2D dans le cadre du projet européen VISEUM. Il a ensuite été complété pour gérer des images panoramiques d'objets 3D dans le cadre du projet ACOHIR. Il utilise le protocole d'images INTERNET; par conséquent, il est compatible avec les autres afficheurs d'images. Il fonctionne en faisant appel seulement aux tuiles correspondant à la résolution demandée pour voir un détail d'une image. Les tuiles concernées sont ensuite encodées JPEG dynamiquement par le serveur et transférées à l'applet. Ainsi, des images, quelque soit leur taux d'agrandissement, peuvent être consultées rapidement par INTERNET. Les vues panoramiques d'objets 3D sont constituées par 24 images 2D prises à partir de 24 point de vues successifs autour de l'objet de telle sorte que l'on peut naviguer autour de l'objet et agrandir des détails de sa surface.

Ce viewer développé sous Java par Denis Pitzalis permet de visualiser des images pyramidales tuilées soit 2D pour les peintures soit 2D-R pour les objets. Le mode 2D-R comporte un certain nombre de vues panoramiques (24 dans le procédé ACOHIR et 36 avec le procédé SCULPTEUR) prises autour de l'objet et avec différentes inclinaisons de la caméra. On peut donc faire tourner l'objet puis augmenter la résolution pour visualiser un détail. Chacune des images de la vue panoramique est identifiée par un nombre afin de pouvoir la reconstruire et l'exporter pour être réutilisée. Un bouton permet d'afficher directement l'image haute définition. Un autre bouton permet de revenir à l'image plein écran. Un bouton d'aide en ligne apporte des informations sur son fonctionnement.
L'afficheur permet également en couplage avec NIP :
- d'ouvrir plusieurs fenêtres pour comparer une image avec différentes résolution ou plusieurs images d'un même tableau ou des images de plusieurs tableaux simultanément,
- de superposer des images avec recalage (changement de dimensions d'une des images),
- de changer l'opacité d'une des images superposées,
- d'inverser des images,
- de changer l'histogramme des niveaux de gris d'une image

- de transformer une image couleur en niveau de gris

Génération des images. Un script en PHP permet

- de transformer une image tif au format PYR-TIFF avec compression JPEG sans perte et de générer un répertoire portant le numéro de référence de l'œuvre et de stocker l'image dans un répertoire. Elle est ensuite reversée sur les disques Raids dans les baies informatiques.
- de générer une vignette et un répertoire portant le numéro de référence de l'œuvre et de reverser la vignette dans ce répertoire.

2) **l'afficheur d'image SCULPTEUR** permet la consultationd'images 3D.

- vision du maillage de la surface
- vision de la texture
- vision du maillage et de la texture
- modification de la texture
- modification de l'orientation de l'éclairage

3) le viewer Icram permet d'analyser et de visualiser la couleur d'une image numérique dans plusieurs modèles de repères colorimètriques.

- représentation des images multi-spectrales dans un espace couleur et multi-vectoriels avec traitement statistiques (analyse factorielle) pour caractériser la palette d'un artiste, ses mélanges, l'évolution de sa technique par comparaison chronologique de son œuvre. Colantoni

4) **le viewer 3D du CNRC**

Le multilinguisme

- les glossaires multilingues d'indexation des rapports

Le vocabulaire normalisé établi dans le cadre des projets NARCISSE et CRISTAL permet d'indexer les rapports d'étude et de restauration. Il a été traduit en allemand, anglais, catalan, chinois, danois, espagnol, français, italien, japonais, portugais et russe. Ces termes permettent de sélectionner des rapports par leur contenu. Ces termes ont également été définis dans toutes ces langues afin d'éviter toute ambiguïté d'interprétation. Ces définitions sont également consultables en ligne.

- la traduction complète de la base

Les interfaces d'interrogation et de saisie, le fichier de messages, les champs en texte intégral (tel que le titre), ainsi que toutes les listes d'autorités constituées de termes normalisés pour informer et interroger la base, tels que l'auteur, le lieu de conservation, etc. ont été traduites du français en anglais, en japonais, en chinois et partiellement en portugais.

Fonctions principales des modules de gestion

Le système comporte de nombreux modules pour gérer les composants de la base de données.
Ils permettent la :

- Gestion des œuvres (données muséologiques, historiques et matérielles séparées par domaine : peinture, sculpture, céramique, verre, émail, métal, , objet ethnographique etc.) et de créer des relations entre les fiches descriptives d'œuvres ayant des points communs (ex : éléments d'un polyptyque). Les champs d'information qui apparaissent à la suite d'une question ne sont pas tous interrogeables
- Gestion des œuvres composites (comportant plusieurs parties indépendantes et assemblées, tels que le mobilier)
- Gestion des images numériques : film numérisés, numérisation directe (données techniques)

- Gestion des rapports : rapport d'étude, rapport de restauration, rapport des commissions de restauration, fiche de santé, fiche de constat d'état etc. (indexation)
- Gestion des articles et des publications,
- Gestion des synthèses électroniques
- Gestion de l'historique des constats d'état (données muséologiques et techniques)
- Gestion de l'historique des fiches de santé (données muséologiques et techniques)
- Gestion des analyses qualitatives et quantitatives des matériaux constitutifs des œuvres d'art : analyses élémentaires, structurales, isotopiques et moléculaires (données numériques, spectres, courbes, formules chimiques etc.)
- Gestion de l'historique du mouvement des œuvres (données muséologiques et administratives)
- Gestion des droits d'accès différenciés par catégorie d'utilisateur (administrateur, gestionnaire, saisie, consultation) et par catégorie de notices (œuvres, image et documents)
- Gestion dynamique et contrôlée des dictionnaires dont certains sont hiérarchisés (listes d'autorités composées de mots clés ou de termes d'indexation pour les rapports d'étude et de restauration); gestion différenciée par champ d'information,
- Gestion du multilinguisme,
- Gestion des liens dynamiques de mots clés ou de termes d'indexation qui pointent sur les définitions des termes d'indexation ou sur l'adresse de sites Web en relation par exemple avec des auteurs ou des musées
- Gestion des interfaces (menu convivial, navigation intuitive de la page de requête à la page de résultat partiel puis à la page de résultats détaillés, paramétrage du nombre de notices à afficher par page de résultat, boutons interactifs comportant des étiquettes de commentaire)
- Gestion de l'historique des connexions à la base EROS par la fonction Admin. Elle permet de gérer et d'identifier les utilisateurs qui se sont connectés à la base EROS. Des histogrammes représentent ces connections par heure, par jour, par mois, par trimestre, par utilisateur etc.
- Gestion des mises à jour quotidiennes des listes de terme de sélection séparée par domaine situées dans l'écran de requête (voir loupe)
- Gestion quotidienne de la sauvegarde du contenu de la base,
- Gestion de DTD
- Gestion XML

Paramétrage dynamique des fiches descriptives et mises à jour des listes de sélection
Seul l'attribut clé des tables de la base est obligatoire à la saisie. Il correspond, en général, à l'identifiant de l'objet concerné (œuvre, image, document etc). Lors de la saisie d'une notice œuvre, image ou document, il est possible d'afficher le numéro de la dernière notice enregistrée ou de dupliquer automatiquement des fiches à la saisie

Les listes qui figurent à la fin des champs représentées par une loupe et qui sont constituées à partir des occurrences des termes (mots clés) informés dans la base, sont séparées par domaine (peinture, céramique et terre, sculpture, etc.) afin de n'y trouver que les termes relatifs à ce domaine.

Modes d'édition
- Présentation d'une page HTML comportant le dossier d'une œuvre avec liens dynamiques sur tous les termes de la notice. Suite à la notice de l'œuvre comportant une vignette, les images sont ordonnées par leurs dates de réalisation, des plus anciennes vers les plus récentes.

- Présentation du résultat d'une requête sur le choix d'images sous forme de planche (mosaïque) d'imagettes interactives. Les images sont ordonnées chronologiquement selon leurs dates de réalisation, des plus anciennes vers les plus récentes. Les informations liées à la définition de l'image numérique peuvent être consultées avec le bouton droit de la souris (définition, nature, compression etc.).
- exportation du résultat d'une requête au format XML
- tabulation des données au format HTML
- tabulation des données au format texte pour être réutilisé par un tableur.
- Affichage des rapports ou des synthèses selon leur format HTMF ou Word par lien dynamique
- Reconstruction des images sources au format TIFF à partir des images multi-résolution PYR-TIFF

Mode de recherche simple ou avancée à partir des méta-données

Recherche simple

- Mono-champ
- Monochamp et multicritères
- Mono-base (œuvre, ou image ou document)
- Mono-domaine
- Terme libre avec troncature gauche et droite
- Terme exact choisi dans une liste constituée à partir des informations qui se trouvent dans la base. Il est possible de pointer rapidement sur un terme en cliquant sur la première lettre du terme qui se trouve dans l'entête de cette liste.
- Par intervalle de dates (dates négatives avant Jésus-Christ, inférieure ou supérieure à etc.)
- Par intervalle de dimensions
- Tout les termes qui figurent dans l'écran du dossier complet sont interactifs et permettent de reformuler une question sans revenir à l'écran de requête.

Recherche avancée

- Multichamp
- Multichamp et multicritères
- Multi-base (question croisée)
- Multi-domaine
- Recherche complexe sous SQL à partir de formules écrites en PHP

<u>en mode reconnaissance de contenu d'images avec classification et adaptativité</u>

- **l'indexation des images à partir de vecteurs caractéristiques de leur contenu (couleur, texture, segmentation, proximité etc.) avec classification (clustering) permet d'affiner ces groupes par intervention de l'expert qui rejette le classement par niveau de similarité. La soumission d'une nouvelle image permet de comparer son niveau de similarité aux classes existantes. ENSEA**

traitement primaire des images numériques

- recollement de vues multiples d'un objet :
- recollement électronique par identification de deux points similaires dans deux images adjacentes (logiciel VIPS)
- **recollement des images individuelles par reconnaissance des contenus des zones de chevauchement des images HP Palo Alto**

Comment repousser les limites des bases de données

- un vocabulaire d'indexation normalisé, défini et multilingue,
  - des catégories d'information définies et non ambïgües,
  - un système documentaire commun (ateliers de restauration)
  - des formats d'échanges standards (langage XML),
  - des modèle de référence par concepts
    - classement des informations par arbre des connaissances (base de connaissance avec un modèle de référence des concepts (ontologie) Southampton
    - accès graphique par liens entre les classes et les instances (vision globale de la base et des occurrences) Southampton
      (tel que le modèle ontologique du CIDOC CRM valable pour les musées, les archives et les bibliothèques),
      - des moteurs de recherche en texte libre (PERTIMM) par analyse grammaticale, contextuelle, multilingue avec localisation du terme recherché dans la page,
      - la recherche de similarité entre images par vecteur d'indexation,
      - la classification semi-automatique par clustering (RETIN)

Conclusion

La base EROS a été développée en interne pour répondre aux besoins de gestion informatique de documents numériques. Elle est élaborée pour la gestion de contenus issus de l'analyse en laboratoire et de la conservation-restauration des collections de musées. La spécificité de ses tables repose sur ces informations scientifiques et techniques (vocabulaires d'indexation des rapports d'étude et de restauration, données numériques des analyses quantitatives, spectres, courbes, formules chimiques, images techniques sous ultra-violet, dans l'infra-rouge, en lumière rasante, au microscope électronique à balayage). Ces tables comportent également des données administratives telles que la régie du mouvement des œuvres mais aussi les historiques des constats d'état, contrôle périodique des collections.

La saisie et la numérisation, en interne, de la documentation produite au C2RMF depuis plus de 70 ans a nécessité déjà 10 années d'investissement Cette base est unique par son volume d'images (6 téra-octets en ligne) mais aussi par la diversité de son contenu. Elle sera prochainement accessible en mode INTRANET à partir d'une ligne à haut débit (réseau RENATER des Universités et des Centres de Recherche du CNRS à 100 Mbits/s)

Un réseau de coopération s'établit autour d'EROS pour y porter de nouveaux développements tels que :

- la visualisation d'images multi-spectrales par plans monochromes et colorés (projet européen CRISATEL),
- l'accès à la base par un mode graphique 3D dynamique réalisé à partir d'un modèle ontologique (le CRM) développé par le CIDOC du Conseil International des Musées, (projet européen SCULPTEUR)
- la représentation en mode graphique 3D de la couleur des images avec analyse statistique en ligne (Laboratoire LIGIV, Université Jean Monnet) ,
- la réalisation de classes d'images ayant des propriétés sémantiques similaires à partir d'un faisceau de vecteurs caractéristiques du contenu de ces images (Equipe Traitement des Images et du Signal, Ecole Nationale Supérieure d'Electronique et de ses Applications),

- la réalisation d'un moteur de recherche en texte intégral avec analyse contextuelle multilingue (PERTIMM, SYSTAL, Ecole Supérieure d'Ingénieurs en Systèmes industriels Avancés Rhône-Alpes).
- la traduction de la base et des dictionnaires en d'autres langues dans le cadre de l'ICOM-CC (Comité International pour la Conservation)

Ce système pourrait alimenter le réseau des ateliers de restauration en France, point de fusion des données et des recherches. Un groupe s'est constitué pour tenter de mettre en place une plate-forme qui réponde à tous les besoins des utilisateurs dans le domaine de la restauration.

Que ces informations d'un passé proche servent au présent au développement de nouvelles recherches dans le domaine culturel et à la protection du patrimoine mondial. Leur élaboration et leur diffusion sont une nécessité pour la connaissance des œuvres et leur sauvegarde.

Bibliographie

(1) NARCISSE Système documentaire des peintures et enluminures; Lisboa, .Arquivos Nacionais/ Torre do Tombo, nov. 1993, 353p

(2) Christian Lahanier, David Meili, Michel Aubert "Art and Science" a multilingual CD Rom Intelligent Multimedia Information Retrieval SystemS and Management, Rockefeller University, New York USA, October 11-13 1994

(3) Ch. Lahanier, G. Aitken et M. Aubert NARCISSE: une bonne résolution pour l'étude des peintures Techné n°2 1995 pp 178-190

(4) CD Rom NARCISSE Glossaire multilangue. - Lisboa, Arquivos Nacionais- Torre do Tombo, 1993 278p

(5) Séminaire NARCISSE, Actes, Arquivos Nacionais - Torre do Tombo, 1993, 98p

(6) CRISTAL Conservation Restoration Institutions for Scientific Terminology dedicated to Art Learning Network (n° R99/ DGX Bruxelles, programme Raphaël, juin 1999- décembre 2000)

(7) Christian Lahanier[1], Geneviève Aitken [1], Jiro Shindo[2], Ruven Pillay[2], Kirk Martinez [3] and Paul Lewis[3,] EROS: An Open Source Multilingual Research System for Image Content Retrieval dedicated to Conservation-Restoration exchange between Cultural Institutions, Actes du 13ième Congrès Triennal duComité International Pour la Conservation de l'ICOM, Rio de Janeiro, Brésil, du 22 au 28 septembre 2002 (à paraître)

(8) Langres xxxxxxxxxxxxxxxxxxxxxxxx (en cours de publication)

Nancy xxxxxxxxxxxxxxxxxxxxxxxx (en cours de publication

(9) Le Bœuf Patrick : *le modèle CRM pour la documentation muséographique : s'attacher au sens pour ne pas être piégé par la forme*, communication du 20 mai 2003 dans le cadre des journées de l'ADBS Documents numériques vol 3. Ed. Techniques de l'ingénieur mai 2003

(10) CRISATEL : Conservation Restauration Innovation Systems for Image capture and digital Archiving to enhance Training Education and lifelong Learning, (programme IST2-1999 n°20163, septembre 2001-septembre 2004)

(11) SCULPTEUR :