

SYSTÈME D'INFORMATION DOCUMENTAIRE GÉO- RÉFÉRENCÉ SUR LE PATRIMOINE ARCHÉOLOGIQUE ET HISTORIQUE DES ALPES-MARITIMES

**Emmanuel Pelligrino (CEPAM, Valbonne, France), Nader
Boutros, Pierre Dittgen (PASS Technologie, Ivry-sur-Seine, France)**

Abstract

The creation of the « *base de données archéologiques et patrimoniales du département des Alpes-Maritimes* » (*Archeological and patrimonial database of the department of the Alpes-Maritimes*) is a joint project of the general council of the Alpes Maritimes, UMR 6130 of the CEPAM (Centre of studies, Prehistory, Antiquity, Middle age) of CNRS (French National Center for Research) and UNSA (University of Sofia-Antipolis). This project lies within the scope of Partnership for the Alpes-Maritimes heritage and knowledge.

The goal is to build a single application that allow to browse and to locate scientific documentation relating to the archeological sites and the heritage monuments of the department.

Each site or monument is described and located by the integration of geographical references. These geo coordinates make it possible to visualise the sites in their spacial, environmental and chronological context thanks to a documentary and geographical information system associating descriptive data base and thematic maps.

Keywords: Information system, cartography, SVG, SDX, XToGen, PostgreSQL, PostGIS

Résumé

La constitution de la *base de données archéologiques et patrimoniales du département des Alpes-Maritimes* est une entreprise conjointe du Conseil Général des Alpes-Maritimes, de l'UMR 6130 du CEPAM (Centre d'études, Préhistoire, Antiquité, Moyen Âge) du CNRS et de l'UNSA (Université de Nice Sophia Antipolis). Elle s'inscrit dans le cadre de l'action de Partenariat pour la connaissance du Patrimoine des Alpes Maritimes.

L'objectif est de construire un outil de travail unique permettant de connaître et de localiser la documentation scientifique concernant les sites archéologiques et les monuments patrimoniaux du département.

Chaque site ou monument patrimonial est décrit et localisé grâce à l'intégration de références géographiques. Ces dernières permettent de replacer et visualiser les sites dans leur contexte

spatial, environnemental et chronologique grâce à un système à la fois documentaire relationnel et d'information géographique associant en ligne une base de données descriptives et des cartes thématiques.

Mots clés : Système d'information, cartographie, SVG, SDX, XToGen, PostgreSQL, PostGIS

Historique

Nous disposions à l'origine d'une série de bases de données et de fichiers d'origines diverses et de formats informatiques non cohérents :

- la base bibliographique et archivistique constituée par Suzanne Roscian (CEPAM CNRS). Il s'agit de trois fichiers chaînés au format *Filemaker Pro* (CARTABIB, CARTACOM, CARTASIT) contenant plus de 1800 références bibliographiques et sitologiques
 - la base MOBILIER compilant les pré-inventaires des collections des Musées d'Art et d'Histoire de Provence à Grasse, du Musée d'Archéologie et du Musée Picasso à Antibes et des réserves du CEPAM au format *Excel* (Emmanuel Pellegrino, CEPAM CNRS).
 - un abrégé de la base PATRIARCHE des sites archéologiques géo-référencés, fourni par le SRA-PACA au format *Excel*
 - la base EDIFICE, inventaire du patrimoine des Alpes-Maritimes établie par Jean-Claude Poteur (Conseil-Général des A.-M.) à l'origine au format *Access* et *Mapinfo*.
 - INVFULCO établie par Suzane Roscian à partir de l'inventaire des sites préhistoriques de Stéphane Fulconis à l'origine au format *Filemaker Pro*
 - les bases FENCEINT et FSTRUCTU, inventaire des enceintes et structures de hauteur du département des Alpes-Maritimes, élaboré dans le cadre du PCR enceinte dirigé par M. Gazenbeek (CEPAM CNRS) à l'origine au format *Filemaker Pro*
 - l'inventaire des sites de la cité de Vence et de Briançonnet (Laurence Lautier, CEPAM CNRS-UNSA) au format *Excel*.
 - la base des toponymes RIVOLI au format *Texto* (Dominique Trousson, CEPAM CNRS)
 - l'inventaire des inscriptions grecques et latines et base onomastique élaborées par Frédéric Gayet et Pascal Arnaud (CEPAM CNRS-UNSA) au format *Excel*.
 - le fichier des sites pré et protohistorique élaboré par Didier Binder et Bernard Gassin (CEPAM CNRS)
 - le fichier des sites de l'Institut de Préhistoire et d'Archéologie des Alpes-Maritimes.
- Tous ces documents ont été réalisés par des personnes et organismes différents selon des normes de travail et des principes variés.

Mise en œuvre

Durant l'année 2003, l'équipe (Etienne Hoffman) a établi une liste d'autorité géo-référencées de sites en compilant l'ensemble des bases de données sources (CARTA, INVFULCO, FENCEINT, EDIFICE, MOBILIER, inventaire Lautier) dans un tableur au format *Excel*.

Le résultat est un fichier contenant 4865 entrées (sites ou monuments) dont moins de la moitié était géo-référencée ou liée à des références bibliographiques.

les principaux problèmes identifiés ont été :

- l'existence de doublons et de sites fictifs. En effet, les différents auteurs des fichiers sources ont souvent donné des noms différents aux mêmes sites, ou regroupé sous la même appellation des sites différents mais espacés. Il est impossible de contrôler ces informations erronées sans avoir un accès directs aux documents originaux qui ont permis l'élaboration des fichiers sources.
- le terme de « site » est porteur de confusions. On peut entendre ce dernier comme un ensemble de vestiges dans un lieu donné (un village, une habitation, un monument...). On peut aussi faire entrer une notion chronologique dans la définition de ce dernier. Or un site peut avoir été occupé à plusieurs époques successives (une église médiévale implantée sur une nécropole antique par exemple). Faut-il alors considérer qu'on a affaire à un seul ou à plusieurs sites ?
- les fichiers sources ont été réalisés par différents auteurs ayant chacun des centres d'intérêt et un langage propre, il en résulte une confusion dans la définition des vestiges mis au jour sur les sites.
- sous la forme d'un tableur, la liste d'autorité s'avère un système lourd qui ne facilite pas la recherche.

Le choix a donc été :

- de contrôler la liste d'autorité en reprenant la documentation originale.
- de créer plusieurs niveaux d'interprétation pour les sites et monuments :
 - le niveau site, conçu comme un ensemble de structures localisées (une villa, un village, une église isolée, une nécropole...)
 - le niveau entité, subdivision du site d'ordre topologique (l'église à l'intérieur du village, une aire de passage à l'intérieur d'une villa...) ou d'ordre chronologique (le site protohistorique sous le village d'époque romaine, la nécropole antique sous la chapelle médiévale...)
 - le niveau opération
- l'utilisation d'un thésaurus pour décrire les structures mises au jour sur chaque site.
- de confier le développement du système d'exploitation à des personnes spécialisées dans l'architecture des bases de données et respectant les normes et standards d'échanges documentaires. (la société PASS Technologie a été retenue après appel d'offre).

Ces remarques ont déterminé à la fois la méthode de travail et la forme actuelle de la base de données. Un cahier des charges a été défini.

La base actuelle

A partir de la fin 2004, la réalisation de la *Base de données documentaire géo-référencée sur le patrimoine archéologique et patrimonial des Alpes-Maritimes* est entrée dans une nouvelle phase à la suite de la livraison d'une première maquette proposée par la société *PASS Technologie*. Celle-ci a repris les différentes indications notées dans le cahier des charges.

Dans le même temps, Suzanne Roscian et Emmanuel Pellegrino ont été chargés de :

- reprendre la base bibliographique et archivistique (CARTABIB, CARTACOM, CARTASIT) afin de redéfinir les sites d'une part et les entités archéologiques d'autre part. Ces dernières sont conçues comme des subdivisions topographiques ou chronologiques de sites, comme nous l'avons indiqué plus haut.
- reprendre la liste d'autorités (FICHER GLOBAL) en y intégrant les données issues de la bases PATRIARCHE, d'une part et celles de la bases bibliographique d'autre part.
- mettre en forme le fichier mobilier mis au jour sur les sites du département des Alpes-Maritimes.

Les données de la base bibliographique, de la liste des autorités et du fichier mobilier provenant de quatre des communes des Alpes-Maritimes ont été intégrées dans le prototype élaboré par la société *PASS Technologie*. Après les corrections d'usage, le résultat semble satisfaisant.

La base de donnée fonctionne à trois niveaux, un niveau de consultation recherche, où toutes les informations sont accessibles, un niveau de consultation restreint pour le grand public et un niveau de saisie.

Le niveau consultation permet une recherche :

- par commune qui permet de connaître tous les sites et les entités qui y sont implantés.
- par site ou entité archéologique qui permet :
 - d'avoir les principales informations sur chacun d'eux (localisation, vestiges, datation...)
 - d'atteindre toutes les références bibliographiques indexées.
 - de connaître et de localiser le mobilier archéologique éventuellement conservé dans les différents musées et dépôts de fouilles.

Différents renvois permettent de se référer aux sites ou entités archéologiques liées entre eux.

- par type de vestige
- des requêtes en plein texte permettent de faire des sélections plus fines sur les mots contenus dans chacune des rubriques de la base

L'interface cartographique permet de situer tous les sites ou les entités dont on connaît les coordonnées sur des fonds des cartes numérisés et actives. Il s'agit de fonds topographiques ou thématiques (hydraulique, végétation, géologie, voies anciennes etc...). On peut y faire apparaître tous les sites sur une commune, ou les mêmes types de vestiges, les sites de la même période sur l'ensemble du département et procéder à un certain nombre de traitements de l'information spatialisée (utilisant les fonctionnalités du SIG en ligne).

À terme, la saisie de nouvelle données se fera directement sur un serveur placé dans les locaux du CEPAM géré par un administrateur.

Pré-requis technologique

Les données sont enregistrées dans un format ouvert (XML) et standardisées selon les normes de la discipline. L'alimentation du serveur se fait en ligne. La documentation est éditée sur le web d'une part et est exportée en PDF pour édition papier d'autre part. Les documents dont le laboratoire a la responsabilité de production doivent en outre pouvoir être aisément vus (ou « moissonnés » et donc être identifiables au moyen des outils supportant le protocole OAI-PMH [*Open Archive Initiative – Protocol for Metadata Harvesting* : <http://www.openarchives.org>]) dans des portails documentaires permettant la vision simultanée sur requête de type géographique, des bases du ministère de la Culture ou des collectivités territoriales et organismes locaux.

Le projet consiste à mettre en place un serveur de données documentaires sur le patrimoine archéologique et historique du département des Alpes Maritimes. Une partie des données à intégrer au serveur existe déjà et proviennent de bases différentes gérées sur des logiciels différents (FileMaker Pro, MSExcel).

Dans la phase actuelle, le serveur de données est destiné à être en accès sur place (intranet) au laboratoire pour le public dans le cadre du centre de documentation sur les Alpes du Sud, et en ligne (sur accès autorisé) (extranet) à l'usage d'équipes de recherche de différents laboratoires et institutions.

À terme, les données issues du serveur devront pouvoir être visualisables dans différents portails publics par exemple sur le site du Conseil Général des Alpes Maritimes (<http://www.cg06.fr>) ainsi que les bases accessibles en ligne du Ministère de la Culture, ou des portails régionaux ou interrégionaux transfrontaliers en projet.

Serveur d'applications

Le serveur de données documentaires est composé principalement de 3 parties : L'inventaire global des sites, la base bibliographique et la base des collections archéologiques. La base bibliographique et la base des collections archéologiques référencent des sites contenus dans l'inventaire global. L'inventaire global des sites contient des données géoréférencées et est, à ce titre, couplé à un SIG (Système d'Information Géographique).

Les sites contenus dans la base sont composés de 3 parties : 1) une fiche d'identification du site selon les standards reconnus, 2) des informations géographiques permettant de placer le site sur une carte, 3) des descriptions de projets permettant d'introduire à la demande des informations supplémentaires sur les sites qui seront ensuite visualisables sur une carte.

Les données bibliographiques de la base contiennent des références de lieux liés aux sites (une bibliographie peut être associée à plusieurs lieux). Les collections archéologiques référencent également des sites. Les données bibliographiques ainsi que les collections archéologiques sont des documents structurés (XML) utilisant les schémas en vigueur et basés sur des normes de la discipline comme : la fiche d'indexation minimale pour les sites archéologiques [La fiche internationale d'indexation minimale pour les sites archéologiques résulte de la collaboration entre le Comité pour la Documentation (CIDOC) du Conseil International des Musées (ICOM) et du groupe de documentation en archéologie du Conseil de l'Europe. http://www.object-id.com/heritage_fr/int_fr.html].

Les sites de la base sont associées à des informations sur les projets de recherche. Ces informations sont basés sur une structure de données identique pour chaque projet de recherche. Une étude plus approfondie de la gestion des projets de recherche et des champs nécessaire devra s'appuyer sur un ensemble de projets existants pour valider la généricité du système.

Technologies sous-jacentes

PASS Technologie a bâti le serveur de données documentaires à l'aide de la boîte à outils SDX - Système de Documentation en XML (<http://adnx.org/sdx>). SDX est un logiciel libre qui permet de construire des applications Web documentaires où la recherche joue un rôle

important. Basé sur l'infrastructure Cocoon2 (<http://cocoon.apache.org/>) de la fondation Apache (<http://www.apache.org>), il permet de construire des sites Web puissants gérant des documents au format XML. Le projet SDX est né à la fin de l'année 2000 sous l'impulsion de la mission de la recherche et de la technologie du ministère de la culture et de la communication. Depuis, il a atteint une maturité certaine et sa base d'utilisateurs ne cesse de s'étoffer.

En complément, le projet XToGen - **XML To Generated** Web Applications (<http://xtogen.tech.fr>), également promu par le Ministère de la Culture et de la Communication est utilisé afin de disposer rapidement d'une maquette fonctionnelle. Cette surcouche à SDX, développé par PASS Technologie, ajoute des fonctionnalités d'administration avancées aux applications SDX.

PASS Technologie a développé de nombreuses solutions basées sur ces technologies : SDX/XToGen. Celles-ci sont de natures diverses : Base de données d'experts en Chimie (CNRS, Paris), Base de données de clichés d'archives archéologiques (MCC-CNRS, Le Caire), Base de données bibliographiques (IRPMF, Paris), Répertoire des langues d'Europe et de Méditerranée (DGLFLF, Paris).

Les applications SDX/XToGen permettent la diffusion de documents mais aussi leur gestion complète : édition, suppression, création de nouveaux documents en ligne. Les documents sont visibles sous forme résumée sous forme de résultats de recherche (recherche par index, recherche plein-texte ou recherche avancée), au format détaillé sous forme de page Web exportable au format PDF pour une impression de qualité. La génération de brochure de documents PDF est également possible. Aux fonctionnalités documentaires et aux documents XML proprement dit, des documents multimédia (images, son, vidéo) peuvent être associés.

Import de données

Si le serveur de données documentaires est destiné à devenir le serveur de données de référence et permettre la saisie des documents en ligne, il devra, dans un premier temps, être alimenté avec les données des bases existantes.

Ces données proviennent de fichiers MSEXcel, et FileMaker Pro. Des scripts de conversion sont développés pour permettre l'alimentation de la base commune à partir des bases existantes.

La réflexion sur la structuration des données de la base documentaire s'inspire de la structure adoptée pour les bases existantes, une certaine réorganisation de l'information a été nécessaire afin de profiter au mieux des fonctionnalités du système.

Fonctionnalités géographiques

La couche géographique a été projeté pour utiliser les normes et standards en vigueur. PASS Technologie promeut l'utilisation des protocoles édictés par l'OGC - Open Geospatial Consortium - <http://www.opengeospatial.org/> :

- WFS (Web Feature Service) permettant la création, la suppression, la mise à jour et l'interrogation des bases cartographiques et des fichiers cartographiques du système par un langage unique normalisé en XML : GML (Géographic markup language) ;
- WMS (Web Map Service) permettant la génération et l'affichage de vue cartographiques sous forme de couches provenant de sources hétérogènes (données vectorielles et raster).

Au vu des contraintes de sécurité liées aux contenus vectoriels chargés dans le système, l'architecture implémentée ne permettra pas d'utiliser ces services dans l'état du projet actuel.

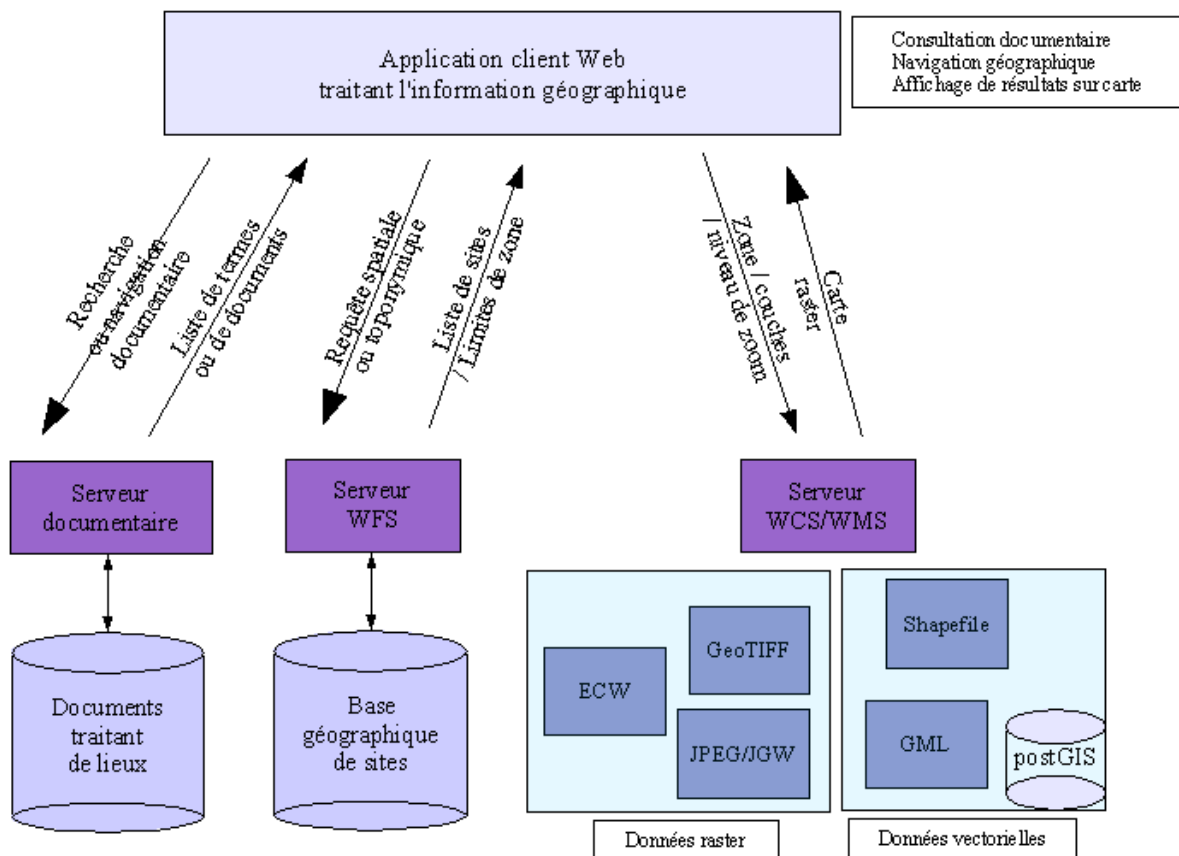


Illustration 1 Organisation des serveurs

Le schéma ci-dessus présente une configuration idéale en client/serveur des différentes composantes d'un système d'information intégré diffusant des informations sur internet et disposant d'un espace restreint d'accès et de saisie d'information par groupe de travail.

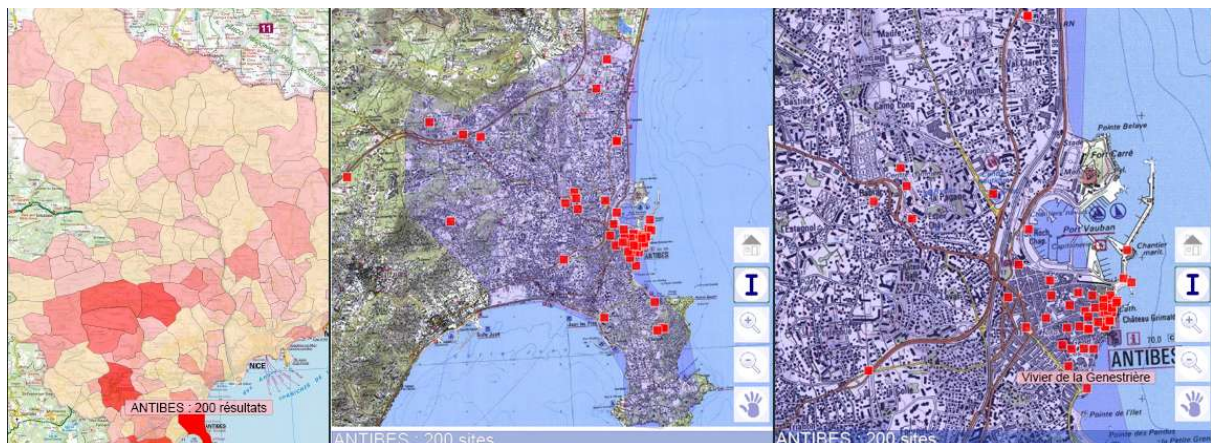
Les différentes composantes du système composant la partie serveur sont détaillées ici pour plus de compréhension mais peuvent être réunies sur un même serveur pour plus de

commodité. Les différents types d'information (documents, lieux, informations purement géographiques) sont servis par des serveurs spécialisés :

- Serveur documentaire : pour la recherche plein texte et structurée sur les documents ;
- Serveur WFS : pour les recherches spatiales ou toponymiques sur les sites, actuellement cette recherche est réalisé par des cartes interactives issues de recherche documentaires ;
- Serveur WCS/WMS : pour l'affichage de fonds de carte personnalisé (niveau de zoom, boites englobantes, couches ...), actuellement cette fonctionnalité est couverte par une interrogation spatiale directe à la couche géographique (PostGIS) de la base de données associée (PostgreSQL).

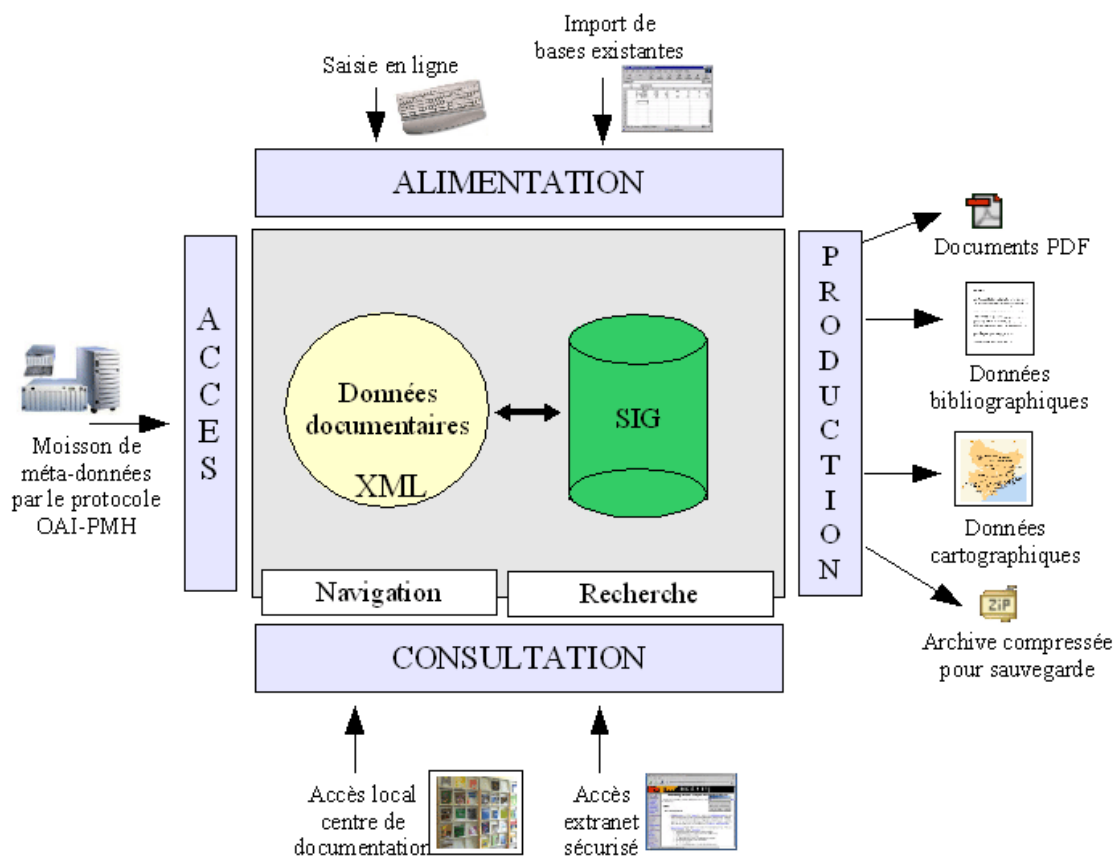
Le client (application Web fonctionnant dans un navigateur Web), propose une interface simple à l'utilisateur et se charge d'effectuer les différentes requêtes aux composantes serveur pour offrir une réponse géolocalisée.

Un site peut être défini par un point, un linéaire ou une surface. Cette information est intégrée au système par la possibilité d'importer des fichiers au formats Shapefile (*.shp) [Format de fichier géographique défini par la société ESRI et qui est devenu un des standards de fait des échanges cartographiques. (<http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>)] dans la base de données géographiques. Des fonds de carte raster, des photos aériennes ou satellite géolocalisées et spécifiques pourront être intégrés pour des projets de recherche par des procédure d'import dédiées.



L'interface actuelle de consultation et d'administration de couches géographiques se base donc sur le module géographique « PostGIS » de la base de données relationnelle « PostgreSQL » et la norme SVG - Scalable Vector Graphics (<http://www.w3.org/TR/SVG/>) - pour l'affichage des cartes zoomables et interactives.

Architecture globale du système



Ce schéma suivant présente une architecture logique globale pour le serveur documentaire. Les données sont au centre du serveur : données documentaires (bibliographie, collections archéologiques, sites) liées à une base SIG pour le géoréférencement.

Le système est accessible aux utilisateurs du centre de documentation et consultable en ligne à des utilisateurs autorisés (extranet sécurisé). Des méta-données (au format Dublin Core - <http://dublincore.org/>) sont disponibles via le protocole OAI-PMH à des moissonneurs externes. Ceci permettra l'intégration future de données sélectionnées provenant de la base documentaire dans des portails externes.

La base est alimentée par l'import de bases existantes et également par saisie et modification en ligne. Ces deux opérations font l'objet d'une politique de sécurité empêchant tout accès non souhaité aux données par une personne étrangère au système.

Le système permet la production de documents et résultats de recherche au format PDF pour l'impression. Des données bibliographiques peuvent également être extraites de la base documentaire pour être importées dans des logiciels spécialisés. Les résultats de recherche peuvent être visualisés sur des cartes géographiques. L'administrateur a la possibilité d'effectuer des sauvegardes des données sous forme d'archives compressées.