

Commission documentation FAMDT samedi 29 mai 2004 Toulouse

Présents :

Stéphanie Coulais (UPCP-Métive, Parthenay)
Jean-Louis Neveu (UPCP-Métive, Parthenay)
Mikaël O'Sullivan (Mémoires vives, Anost)
Jean-Luc Ramel (Dastum, Rennes)
Laura Thomas (CIMP, Céret)
Edith Debru (Conservatoire Occitan-CMDT, Toulouse)
Daniel Vissière (Conservatoire Occitan-CMDT, Toulouse)
Bénédicte Bonnemason (Conservatoire Occitan-CMDT, Toulouse)

Excusés :

Perrine Deljesus (CMTRA, Lyon)
Véronique Ginouvès (MMSH, Aix-en-Provence)
Frédéric Bianchi (Archives départementales du Cantal)
Laetitia Nicolas (Alpes de Lumière)

Ordre du jour :

- 1) Un point sur la formation de la veille : Collecte et analyse de la littérature orale
- 2) L'acquisition du logiciel Alexandrie
- 3) La mise en ligne des catalogues

1) Un point sur la formation des 27 et 28 mai 2004 : Collecte et analyse de la littérature orale

Il serait bon de revenir sur la thématique de l'analyse de la littérature orale mais avec une partie du temps de formation consacrée à des travaux pratiques. Est-ce qu'une suite pourrait s'envisager dans le cadre de la commission documentation où nous sommes plusieurs à avoir besoin de mettre en application la théorie et de savoir utiliser très concrètement les divers catalogues existant. La Famdt pourrait-elle être initiatrice de ces stages de formation ?

2) L'acquisition du logiciel Alexandrie

a) Acquisition

- L'idée de Mikaël qui serait que la FAMDT achète le logiciel et que chaque centre acquiert un poste supplémentaire (moins cher que le logiciel) n'est pas retenue car il faudrait qu'un centre gère la base principale ce qui évidemment est impossible.

- Mikaël lit l'accord cadre proposé par GB Concept pour l'achat groupé du logiciel Alexandrie. Discussion sur les configurations nécessaires pour les postes qui reçoivent Alexandrie. Daniel propose de mettre sur papier les questions qu'il faut poser à GB Concept pour qu'on obtienne des réponses précises et les transmet à Mikaël.

A voir également, la question de la mise à jour. Est-ce que celle-ci fournit les mises à jour gratuitement ?

- Tarifs proposés dans l'accord cadre par GB Concept :
2380 euros TTC : monoposte et maintenance pour 1 an
350 euros HT : maintenance pour les années suivantes

b) Mise en place d'une formation à Aix-en-Provence

Discussion autour d'une date qui conviendrait à tous. La semaine du 27 septembre pourrait convenir. Reste à la soumettre à Véronique qui a proposé d'organiser cette formation dans son organisme et qui avait dans un premier temps proposé la fin du mois d'août.

3) La mise en ligne des catalogues

a) Vers quoi s'oriente l'UPCP-Métive

Texte de Jean-Louis Neveu

La société ARGIA (Parthenay, Deux-Sèvres)

Collaboration avec UPCP-Métive dans le cadre du projet européen « Phonomedia Tradidio » (www.phonomedia.org) et développeur de la base de données multimédias en ligne du CERDO (www.cerdo-metive.org)

Philosophie : non concurrentielle car trop d'énergie, focalisation sur la réalisation (ARGIA se définit comme un concepteur et assembleurs de solutions).

La logique d'argia est donc de proposer une solution alternative aux logiciels propriétaires sur la base d'une utilisation des applications « open source » mais en tenant compte des besoins métier du client. C'est la solution d'assemblage :

- utilisation systématique des applications libres pour ce qui concerne les architectures applicatives (serveurs d'application – Apache-Tomcat - système d'exploitation – Linux)
- utilisation des composants métier développé en « open source » (gestion d'accès, de caddy voire de moteur de recherche)
- développements spécifiques à la demande

Réalisation dans un environnement collaboratif : « Open source » sous licence GPL (qui n'est pas une zone de non droits)

La licence GPL (GNU General Public License) est une "licence de logiciel libre" : "licence de logiciel" car elle définit un cadre d'utilisation fixé par un programmeur sur son logiciel. "libre" car le cadre d'utilisation défini par cette licence est - contrairement aux licences classiques - de s'assurer qu'un certain nombre de "libertés" seront respectées sur le logiciel sous licence GPL (ou sur tout logiciel en dérivant). Ces libertés sont les suivantes :

- ▶ **liberté d'exécuter le logiciel, et ce pour n'importe quel usage ;**
- ▶ **liberté de modifier le logiciel pour l'adapter à ses besoins (dans la pratique, cela nécessite l'accès au code source) ;**
- ▶ **liberté de redistribuer des copies, soit gratuitement, soit contre rémunération ;**
- ▶ **liberté de distribuer des versions modifiées afin que la communauté du logiciel libre puisse en profiter.**

Outils et langages

Langage java (libre de droit) auprès de SUN

Le choix s'est fait sur sa capacité de développement. Selon ARGIA : plus grande capacité et plus pérenne que les autres particulièrement avec Apache (logiciel serveur http gratuit tournant sur un serveur Unix et depuis peu sur Windows NT) avec sa composante Tomcat qui est développé dans un environnement ouvert et participatif (distribué sous licence Apache)

Application de mise en ligne

L'objectif d'Argia était d'élaborer une méthodologie très structurée propre à gérer relativement aisément les différentes composantes utilisables dans un site en ligne tant en pages statiques que dynamiques. Imaginer une application de type conceptuel sans être tributaire d'un développement informatique (programmation) et par conséquent de gagner en autonomie de gestion.

La solution BOC (Back Office Contenu)

BOC : Système permettant la gestion autonome et indépendante des éléments de contenu, de navigation et de présentation

- organisation du contenu indépendant de la présentation. Structuration de l'information, des données qui doivent être parfaitement repérables et maintenables (format des données, éléments multimédias...)
- navigation qui doit se faire sur au moins deux axes (vertical et horizontal) + une multitudes d'autres : thématiques, multicritères, guidée...
- Présentation (banque de données graphiques aisément triturables au fil du temps, des messages et des besoins clients) : réalisation par un tiers
par Template : Le mot réservé "template" est toujours placé au début de la définition et de la déclaration préliminaire d'une fonction template. La suite est constituée d'une liste de paramètres

formels séparés par des virgules, délimitée par une paire de signe d'inégalité ("< >"). Cette liste est appelée liste de paramètres formels de la fonction template. Elle ne doit pas être vide.

Par modèles de présentation (par exemple des feuilles de style XSL qui permettent de mettre en forme un flux XML)

A côté du Boc mais en lien étroit viennent s'ajouter une gamme d'outils modulables en fonction des besoins (outils puiser dans le catalogue open source ; moteur de recherche full text Synequa d'Intuition) actuellement payant).

Cf. également les documents distribués.

b) La MMSH : Lecture du texte de Véronique Ginouvès et de Jean-Christophe Peyssard
Cf. texte de V.G. et de J.-C.P. en pièce jointe

c) Au Conservatoire Occitan

Texte de Daniel Vissière

L'ensemble des catalogues est géré à l'aide du progiciel TEXTO.

Afin de gérer certaines insuffisances de ce programme, un certain nombre de palliatifs ont été utilisés.

Démarrage de l'outil internet.

Le choix du plan d'action a été de créer un site classique, mais avec les ouvertures logicielles destinées à la mise en ligne du catalogue.

Les choix faits ont été :

Utilisation d'un hébergeur mettant une bande passante suffisante pour ce démarrage, le langage PHP et un accès à une base de données MySql (qui étaient compris dans l'offre avec des limitations qui ne nous gênaient pas)

Nous reviendrons plus loin sur les raisons qui ont dicté ces choix.

Au cours de la mise en place du site, ont été menées les mises en place d'un intranet privé pour le Conservatoire Occitan, un intranet semi-public (accès limité aux associations partenaires) qui se limite à l'heure actuelle à un forum destiné à recueillir les souhaits des fonctionnalités en vue d'un portail commun, ainsi que les bases de la mise en ligne des catalogues.

Le cahier des charges de la mise en ligne du catalogue

Côté gestionnaire :

- Récupération des données TEXTO (base non relationnelle) en tenant compte des palliatifs mis en place.
- Prise en compte de l'imminence de l'arrivée d'un nouveau progiciel (Alexandrie)
- Utilisation d'une base de données relationnelle
- Droits d'accès sécurisés et hiérarchisés pour l'accès aux données
- Possibilité d'attribution de droits précaires (heure, journée, semaine...)
- Possibilités de saisies de certains champs
- Synchronisation avec les données saisies in situ automatique, périodique ou à la demande.
- Capacité à cumuler les résultats de requête sous différents serveurs et/ou bases de données.

Côté utilisateur

- Interface conviviale, mais sobre afin de ne pas pénaliser les accès modem RTC
- Limiter, voire éviter l'utilisation de technologies pénalisante en terme de ressources (Flash, Java, ...)
- Assurer la possibilité de consultation en mode texte (rapidité, accessibilité aux malvoyants)
- Offrir des modes de recherches conviviaux, et efficaces.
- Offrir des modes de recherches plus élaborés pour un public initié

Les choix

Utilisation d'un hébergeur, dans un premier temps hébergement mutualisé

Cette solution est de loin la moins coûteuse. En effet, un hébergement in situ impliquerait

Un coût matériel (nécessité d'avoir deux serveurs redondants)

Un coût salarial (embauche d'un salarié pour la maintenance du système)

Un coût élevé pour les lignes spécialisées France Télécom

On pourrait bien sûr penser à une connexion ADSL, mais ce serait oublier que si le débit vers l'utilisateur est élevé, dans le sens inverse, c'est à dire en mode serveur, il est beaucoup plus faible...

L'offre des technologies utilisées côté hébergement est très large.
Côté système d'exploitation, le choix se réduit pratiquement à Microsoft et Linux.
Notre longue expérience, et les retours d'utilisateurs professionnels nous ont conduit à éliminer l'utilisation d'une plate-forme sous Microsoft.
Nous avons donc choisi une plate-forme Linux.
Pour le moteur de base de données, et pour les mêmes raisons que précédemment, nous avons éliminé une solution Microsoft (manque de sécurité, mauvaises compatibilités entre versions, lourdeur des drivers ODBC,...)
Nous avons opté pour MySQL, dont la fiabilité n'est plus à démontrer, et dont les quelques limitations (par exemple absence de procédure stockées) ne posaient pas problème dans l'utilisation envisagée.
Au niveau du langage de programmation, le choix a été fait sur PHP, qui, là encore, possède une excellente fiabilité.
Ces choix permettent de garantir une pérennité au développement.
De plus, nous avons fait le choix de modulariser la structure applicative, et d'utiliser le plus possible la notion d'objet.
Pour la mise à jour des bases (Alexandri -> MySQL, nous avons envisagé dans un premier temps d'utiliser un CRON chez l'hébergeur (des solutions existant lorsque l'hébergeur ne le propose pas).
Finalement nous avons opté pour la mise en place d'un micro serveur interne sous Linux qui a pour rôle :

Un serveur de fichier permettant au réseau interne actuel fait de Mac et de PC de pouvoir collaborer simplement

- Un collecteur de mails qui assure un filtrage des mails parasites
- Un espace miroir des bases de données
- Une gestion de mise à jour régulières des bases mises en ligne.

L'accès des divers postes du Conservatoire Occitan à l'internet est réalisé à l'aide d'un routeur /pare feu au travers d'une connexion ADSL haut débit.

Structuration des tables

Comme nous l'avons évoqué plus haut, le moteur de base étant MySQL, nous utilisons une structure de tables relationnelles.
Lors des choix d'organisation des tables, nous avons essayé d'être proche de ce qui est prévu avec l'utilisation d'Alexandrie.
Mais il faut insister sur le fait qu'il n'y a pas de problèmes ou de difficultés majeures à synchroniser des bases, même si elles sont différentes.
La seule contrainte est que toute information que l'on souhaite placer dans la base de destination doit pouvoir être extraite de la base d'origine, sinon, il faut la saisir manuellement.
Par ailleurs, il n'y a aucune difficulté à gérer des infos contenues dans différents moteur de base de données, il suffit que les propriétaires de ces bases soient d'accord et donnent les moyens d'accéder à tout ou partie de ses données !
En fait cette pratique est monnaie courante sur le réseau internet !

Les types de données

Le principal type de données à manipuler, dans notre cas est du texte. Dans ce cadre, seul le choix du format de stockage des dates doit attirer notre attention. Mais les solutions existent et sont parfaitement gérables dans notre application.
De même, les formats de stockage des nombres (qui s'avère délicat dans des applications de comptabilité ou bancaire) n'a aucune incidence dans notre cas.
Les données de type image, vidéo, son ... sont stockées dans des fichiers dédiés et leur emplacement vectorisé au niveau de la base. Nous avons fait ce choix pour avoir une grande souplesse de gestion, mais d'autres méthodes peuvent être mises en œuvre.

Les droits d'accès aux données

Nous avons porté une grande attention à ce problème. En effet, il s'est avéré difficile de déterminer les modalités et les restrictions de consultation.
En conséquence, nous avons mis en place une structure hiérarchisée d'une très grande souplesse qui permet de gérer, dans un cadre cohérent, des droits d'accès

- Au niveau de l'accès au site
- Au niveau de l'accès aux tables
- Au niveau de l'accès aux champs

De plus, la gestion des droits inclue un vecteur temporel qui accroît la souplesse de la gestion

- Durée illimité
- Durée limité
- Heures limitées

Enfin, il est aussi possible d'accorder plusieurs droits avec des durées différenciées, et/ou des rétrogradations d'accès.

Le visuel

Nous nous sommes fixé un certain nombre de critères

Les temps de chargement doivent être rapides

Les opérations qui risquent de demander du temps (par exemple une recherche complexe) doivent être signalées

La présentation doit être claire et donc éviter les surcharges d'informations ou d'images

Respecter les critères de contraste afin d'éviter les textes sombres sur fond sombre, car il est difficile de savoir comment l'ordinateur de l'utilisateur final est réglé.

Proposer la possibilité de consultation en mode texte (pour les connexions lentes ou les malvoyants).

Cette partie est en cours d'élaboration, car nous pensons qu'il était indispensable d'avoir un développement logiciel des fonctionnalités fiable, la visualisation bien qu'étant la seule partie significative pour les utilisateurs n'en est pas moins que la partie visible de l'iceberg.

d) A Dastum

Texte de Jean-Luc ___

La mise en ligne des archives sonores à Dastum :

Lors de l'acquisition par Dastum du logiciel Alexandria en 1999, nous pensions (naïvement sans doute) avoir là LA solution à plusieurs de nos problèmes, sinon à tous : informatisation du catalogue (création-modification des fiches documentaires, fichiers son et texte en pièces jointes,...) ; saisie de fiches doc. dans les antennes ; consultation du catalogue sur internet.

Après plusieurs années de fonctionnement (ou de flottement, d'interrogation parfois), nous pouvons tirer les premiers enseignements :

L'OUTIL UNIQUE UNIVERSEL n'existe sans doute pas !

Alexandrie en utilisation régulière, en local, par des permanents salariés documentalistes OK. Donc, on garde Alexandria pour nos besoins au centre de doc. de Rennes.

Pour les données produites à l'extérieur par les antennes ou des individus isolés, Alexandria n'est pas adapté (quand bien même cela serait possible économiquement) : trop difficile à prendre en main pour quelqu'un qui fait cela de temps en temps ; un outil comme le tableur Excel peut remplir tout à fait convenablement ce rôle.

Pour ce qui est de la consultation de la consultation des archives sonores via internet, nous ne sommes pas satisfait de notre expérience Module Web Alexandria (là encore en dehors de toute considération économique) : non fiabilité, dépendance de GB Concept pour l'évolution,...

C'est pour cela que nous avons mis au point un nouvel interface d'interrogation/consultation des archives sonores autour des outils/langages PHP et MySQL. Concrètement, nous faisons un export de nos données Alexandria pour les intégrer dans une nouvelle base SQL. Le résultat auquel nous avons abouti nous satisfait entièrement :

Ça marche.

Nous en sommes les auteurs des outils et nous pouvons les faire évoluer.

Ça coûte bien moins cher !

VOIR LE SITE :

<http://ns2621.ovh.net/~archivesx/accueil>

e) Discussion

Certains centres ont bien avancé dans le projet de mise en ligne (UPCP-Métive, Conservatoire Occitan) et Dastum en est au stade de la finalisation. D'autre part, Mémoires Vives et le CIMP sont des organismes qui n'ont – pour l'heure – ni le temps ni les moyens de mettre en place un tel chantier. Mais si tous les centres n'en sont pas au même point, la volonté de penser les choses dans le cadre d'un projet collaboratif est ce vers quoi tout le monde veut tendre. La solution évoquée est de solliciter la FAMDT pour l'embauche d'une personne chargée d'effectuer un état des lieux ainsi que de vérifier si une solution commune est envisageable. Les centres qui ont avancé, présents autour de la table, sont d'accord pour faire profiter de leur expérience. Reste à savoir comment cela peut être possible ? Cela sera une des missions du coordinateur.

Décision est prise de soumettre à la FAMDT un argumentaire afin qu'elle puisse demander le financement d'une mission et recruter la personne idoine ? Le travail sur l'argumentaire est alors entrepris et plusieurs lignes sont rédigées. Par ailleurs, Jean-Louis se propose pour parler de notre projet à Pierre-Olivier Laulanné, directeur de la FAMDT.

Argumentaire (qui reste à développer et à compléter) :

- Besoin de consolider le réseau
- Après avoir effectué le tour des différents centres il est clair qu'il existe une volonté commune du réseau de la FAMDT d'avancer sur des projets libres et ouverts de mise en ligne des catalogues. Les centres envisagent de continuer d'utiliser une solution propriétaire pour la gestion de leur base par l'utilisation ou l'acquisition du logiciel Alexandria¹.
- Il devient impératif, voire urgent que quelque chose se mette en place.
- Pour les sites qui composent actuellement le pôle associé FAMDT à la BnF, obligation de mettre en ligne leur catalogue des archives sonores comme le stipule la convention.
- Recrutement d'une personne avec comme critère minimum la disponibilité et la compétence
- Mission :
 - o Faire un état des lieux
 - o Comprendre les objectifs de chaque de centre
 - o Faire une synthèse faisant apparaître les éléments communs et les éléments divergents et comprenant une proposition de la procédure la mieux adaptée au réseau (facteurs économiques, techniques et autres).

Le profil d'un chargé de mission est également évoqué. Qui serait capable de mener ce genre de mission ? Une société ? Bénédicte pense que ce n'est pas une bonne chose de solliciter ARGIA comme Jean-Louis le proposait. Autre solution possible, avoir recours à un étudiant. Edith connaît des organismes universitaires à qui l'on pourrait s'adresser. Et enfin, les noms de Daniel Vissière et de Jean-Christophe Peyssart sont cités. Ils ont tous les deux pour qualités de bien connaître la question (en terme de compétences techniques) ainsi que les fonds concernés et les exigences documentaires des organismes. Daniel évoque alors la possibilité de mettre en place un forum. Cette solution paraît intéressante mais elle n'est pas retenue comme pouvant remplacer le recrutement d'un chargé de mission.

Compte rendu réalisé par Edith Debru et Bénédicte Bonnemason
Toulouse, 10 juin 2004

¹ Ces deux phrases sont extraites du texte de la MMSH *Mise en ligne des catalogues d'archives dans le cadre du réseau FAMDT : point de vue de la MMSH*.