

**UN METACATALOGUE
D'ARCHIVES SONORES
INEDITES POUR LA BnF ET LE
PÔLE ASSOCIÉ FAMDT**

**Damien Belvèze, Carine Chamoin, Sébastien
Caudron, Etienne Cavalié, Laurent Séguin,
Nicolas Tocquer**

Sous la direction de Jean-Michel Salaün
Enseignant à l'ENSSIB

Remerciements

Nous tenons à remercier en premier lieu M. Pierre-Yves DUCHEMIN, directeur du Département des ressources documentaires de l'Enssib, dont nos innombrables sollicitations n'ont jamais altéré la patience. Nous adressons aussi nos remerciements à M. Jean-Michel SALAÜN, notre tuteur, qui a nous a dispensé d'emblée des conseils avisés.

Nous ne saurions passer sous silence l'aide que nous ont apportée MM. Jean-Paul ROUX-FOUILLET, chargé de cours à l'Enssib, et Michel BOTTIN, président de l'Association pour la documentation numérique en XML, en mettant à notre service leurs compétences informatiques. Nous leur exprimons ici notre gratitude, de même qu'à Mmes Martine BARUCHEL, Elizabeth CHERHAL, Nicole CYVOCT, et MM. Géraud BULLY, Antoine MENSOR, Martin SÉVIGNY dont l'expertise a alimenté notre réflexion.

Nous sommes également redevables à Mmes Florence CLAVAUD, responsable du service des nouvelles technologies au Centre historique des Archives nationales, Emmanuelle JANÈS-OBÈRE, chargée de la gestion de projets et des développements informatiques à la Médiathèque de l'Institut Pasteur, Fabienne QUEYROUX, responsable de la collection des manuscrits contemporains à l'Institut de France ainsi qu'à MM. Patrice VERRIER et Rodolphe BAILLY, du Centre de documentation de la cité de la Musique, Michel FINGERHUT, directeur de la Médiathèque de l'IRCAM, et Uwe DIEROLF, responsable du KVK à la Bibliothèque universitaire de Karlsruhe.

À la BnF, nous tenons à exprimer particulièrement notre gratitude à M. Pascal CORDEREIX, responsable du projet, pour sa disponibilité et la cordialité de son accueil. Nous exprimons aussi notre reconnaissance à Marie-Pierre BODEZ et Danièle BRANGER, ainsi qu'à M. Pierre ECHEGARAY, du Département de l'Audiovisuel, qui ont bien voulu nous faire profiter de leur expérience, ainsi que Mmes Sara AUBRY et Annie IMBAUD du Département de la Bibliothèque numérique, Mme Véronique FALCONNET et M. Yves MORET du Catalogue collectif de France, et Mme Sophie MAZENS, du Département du Dépôt légal.

Nous remercions enfin Mmes Bénédicte BONNEMASON Stéphanie COULAIS, Véronique GINOUVÈS et MM. Jean-Louis NEVEU, Pierre-Olivier LAULANNÉ et Jean-Luc RAMEL, de la FAMDT, pour leur précieuse collaboration.

Sommaire

INTRODUCTION	7
PARTIE 1 – ANALYSE DE L’EXISTANT ET DÉFINITION DE LA PRESTATION	9
1. OBJECTIFS VISÉS.....	10
1.1. Contexte	10
1.2. Le projet	10
1.3. Objectifs définis dans la lettre de mission	11
2. LES ACTEURS DU PROJET.....	12
3. ANALYSE DES FONDS ET PERTINENCE DU PROJET	13
3.1. Les associations et la BnF	13
3.2. Coup de sonde dans les fonds de la FAMDT	13
3.3. Topographie des fonds de la BnF	15
3.4. Comparaison des catalogues	16
3.4.1. Les champs descriptifs des notices	16
3.4.2. Thésaurus et classifications	17
3.4.3. Le format informatique des notices	18
3.5. Dissonances et harmonie.....	18
3.6. Le terrain d’entente du métacatalogue.....	21
4. REDÉFINITION DE LA PRESTATION	23
4.1. Analyse de l’existant / connaissance du terrain	23
4.2. Besoins et attentes	25
4.3. Choix des solutions techniques les mieux adaptées aux contraintes.....	27
5. CONTRAINTES	28
5.1. Hétérogénéité relative des catalogues.....	28
5.2. Hétérogénéité des acteurs.....	29

5.3. Contraintes de mise en place	30
5.4. Contraintes de fonctionnement.....	31
5.5. Contraintes d'évolutivité.....	31
COMPTE-RENDU DÉCISIONNEL DU 1^{ER} COMITÉ DE PILOTAGE	33
PARTIE 2 – PRÉSENTATION DES SCÉNARIOS	34
1. LE MÉTACATALOGUE : UNE PREMIÈRE ESQUISSE	36
1.1. Elaboration d'une liste des champs d'interrogation	36
1.2. Les autres fonctionnalités de l'interface	37
2. LES SCÉNARIOS	38
2.1. Z39.50 : le poids de l'expérience	38
2.1.1. Définition technique.....	38
2.1.2. Avantages	38
2.1.3. Inconvénients.....	39
2.1.4. Investissements	39
2.2. OAI-PMH : un moissonnage balisé	40
2.2.1. Définition technique.....	40
2.2.2. Avantages	42
2.2.3. Inconvénients.....	43
2.2.4. Investissements	43
2.3. Les connecteurs : autonomie, simplicité et fonctionnalité.....	44
2.3.1. Définition technique.....	44
2.3.2. Avantages	45
2.3.3. Inconvénients.....	46
2.3.4. Investissements	46
2.4. Le métamoteur XML : une collaboration active	47
2.4.1. Définition technique.....	47
2.4.2. Avantages	48
2.4.3. Inconvénients.....	48
2.4.4. Investissements	48
3. TABLEAU DE COMPARAISON DES DIFFÉRENTES SOLUTIONS	50
COMPTE-RENDU DÉCISIONNEL DU 2^{ÈME} COMITÉ DE PILOTAGE	52

PARTIE 3 – CAHIER DES CHARGES..... 53

1.1. Une phase préparatoire pour la FAMDT.....	56
1.1.1. La mise en ligne des catalogues	56
1.1.2. L'acquisition d'un progiciel adéquat	56
1.2. Le choix d'une DTD	57
1.2.1. Critères de choix	57
1.2.2. Les formats envisageables.....	57
1.3. La conversion des notices au format XML	58
1.3.1. Du côté de la FAMDT	58
1.3.1.1. Une phase de concertation.....	58
1.3.1.2. Les modalités et les étapes de la conversion	59
1.3.2. Du côté de la BnF	59
1.3.2.1. Enrichir les notices pour permettre leur sélection	59
1.3.2.2. Renseignements des droits d'accès aux métadonnées	60
1.3.2.3. Montée en charge progressive	60
1.3.2.4. Programmes à prévoir pour la conversion au format XML..	61
1.4. La question de l'hébergement des données.....	62
1.4.1. Stockage des données ou conversion à la volée ?	62
1.4.1.1. Le stockage des données au format XML	62
1.4.1.2. La conversion des données au format XML « à la volée » ..	62
1.4.2. Qui héberge les données ?	63
2. MISE EN PLACE DE SDX	64
2.1. Paramétrage de l'application SDX.....	64
2.1.1. L'exploitation public/professionnel	64
2.1.2. Les modes d'interrogation	64
2.1.3. La gestion des bases de données en XML : de SDX à Cocoon	64
2.1.4. La mise en place du portail.....	65
3. MAINTENANCE ET PERSPECTIVES	66
3.1. L'alimentation des bases existantes.....	66
3.2. La numérisation des documents et leur accessibilité.....	66
3.2.1. Communication à distance des métadonnées	66
3.2.2. Communication à distance des données numérisées.....	67
3.3. L'intégration de nouvelles bases	69
3.3.1. Tous les types de documents	69
3.3.2. Indépendance	70
3.3.3. D'un XML à l'autre.....	70
4. TABLE DES COÛTS ET CALENDRIER	71
4.1. Préalables : mise en ligne des catalogues qui ne le sont pas encore	72

4.2. La question de la répartition des catalogues	75
4.2.1. Base XML commune.....	75
4.2.2. Bases XML éclatées.....	76
4.2.3. XMLisation à la volée (sans stockage des données XML)	78
4.3. Mise en place du portail d'interrogation.....	79
4.4. Calendrier synthétique	81
COMPTE-RENDU DÉCISIONNEL DU 3^{ÈME} COMITÉ DE PILOTAGE	82
CALENDRIER DE LA PRESTATION	83
BUDGET DE LA PRESTATION.....	86
1. BUDGET PRÉVISIONNEL DE LA PRESTATION.....	86
2. BUDGET RÉALISÉ DE LA PRESTATION.....	87

Introduction

Le projet de catalogue collectif d'archives sonores, suscité par le Département de l'Audiovisuel de la Bibliothèque nationale de France, est destiné à mettre en valeur les archives sonores conservées en France portant sur le folklore national. La réflexion qui l'a lancée s'est tenue dans le cadre de la convention de pôle associé entre la BnF et quatre partenaires membres de la Fédération des Associations de Musiques et de Danses Traditionnelles (FAMDT), à savoir : Dastum à Rennes, UPCP-Metive à Parthenay, le Conservatoire Occitan à Toulouse et la Maison Méditerranéenne des Sciences de l'Homme (MMSH) à Aix-en-Provence.

Deux exigences essentielles pour l'avenir de ce catalogue ont dès le départ orienté les recherches : d'une part la possibilité à terme d'accéder par Internet aux documents sonores numérisés à partir des notices contenues dans les catalogues ; d'autre part l'extension ultérieure à de nouveaux partenaires et acteurs, étrangers à la convention de pôle associé.

Le projet avait l'attrait d'être extrêmement novateur dans le cadre des documents sonores, puisque très peu d'entreprises comparables ont pu être repérées aux Etats-Unis, et aucune en France. Il a eu le double intérêt de nous confronter au problème de la coordination de partenaires différents lors de la mise en place d'un réseau, et de nous introduire dans la complexité des formats de description documentaire qui sont en pleine mutation avec les ressources fournies d'une part par le langage XML et ses applications, et d'autre part par les réflexions portant sur la difficulté à décrire des documents autres que textuels, qui suscitent souvent moins d'intérêt lors de la normalisation des descriptions, et lors de l'élaboration des outils pour traiter les notices descriptives.

Deux difficultés essentielles ont été soulevées durant la première partie du calendrier. Leur résolution était la condition à l'élaboration de scénarios propres à répondre à la demande du commanditaire.

D'une part, il importait de connaître la nature des collections et des catalogues des cinq partenaires concernés par le projet. Cette investigation a été rendue particulièrement difficile par l'éloignement géographique des partenaires les uns par rapport aux autres et par rapport à Paris ; et par leurs différences institutionnelles, qui rendaient compte à la fois d'une grande disparité dans les fonds conservés et dans leur description, et d'une volonté croissante de converger vers de mêmes pratiques de catalogage.

D'autre part, il fallait se renseigner sur les moyens techniques, informatiques, aptes à remplir les conditions émises par le commanditaire dans sa lettre de mission et ultérieurement. Ces recherches ont été d'autant plus difficiles qu'aucun de nous n'avait reçu au préalable de formation informatique propre à nous fournir une connaissance générale de la description documentaire, étudiée d'un point de vue technique.

Cette double tâche s'est révélée extrêmement formatrice sur bien des plans.

Le document suivant reprend donc en trois parties les différentes problématiques initiales du projet, ainsi que ses étapes chronologiques : une analyse de l'existant d'abord, aboutissant à une reformulation de la prestation demandée en fonction de la nature des partenaires ; la description de scénarios issus de cette analyse et des possibilités techniques ; l'approfondissement du scénario retenu, afin d'en faciliter au mieux la mise en place, dans un calendrier qui dépasse le cadre de notre réflexion.

Partie 1 – Analyse de l'existant et définition de la prestation

Objet du présent document

Ce document décrit la nature de la prestation que le groupe de gestion de projet de l'ENSSIB, composé de six conservateurs-stagiaires, s'engage à fournir au commanditaire, selon le calendrier établi par l'ENSSIB. Il s'appuie d'une part sur la lettre de mission, d'autre part sur l'analyse de l'existant, commencée lors du premier stage passé au sein du Département de l'Audiovisuel, du 9 au 13 février 2004, et qui s'est poursuivie jusqu'au second stage, effectué au même endroit du 29 mars au 2 avril 2004. Il sera soumis à l'approbation du commanditaire.

La première partie porte sur la connaissance progressivement acquise des différents partenaires, de leurs fonds, de leurs modes de fonctionnement, de leurs statuts respectifs, des difficultés qui leur sont propres dans le cadre de ce projet de catalogue collectif. Cette description aboutit à une redéfinition indispensable de la prestation.

1. Objectifs visés

1.1. Contexte

La convention que la Bibliothèque nationale de France a signée en 1998 avec les membres de la FAMDT incluait comme perspective l'établissement d'un catalogue collectif des archives sonores qui serait inclus dans le CCFr. Ce projet n'a pas encore abouti parce que le protocole de ce catalogue collectif ne prévoit pas l'intégration de documents sonores. Par ailleurs, le format UNIMARC s'est avéré assez lourd d'utilisation et peu maniable pour le catalogage des archives sonores dans les centres de documentation. C'est au vu des dernières évolutions techniques de la gestion et de la recherche documentaires (généralisation du XML, des DTDs – EAD notamment – et de l'OAI-PMH) que le Département de l'Audiovisuel a jugé opportun de le relancer en faisant appel à un groupe de travail de l'ENSSIB.

1.2. Le projet

Le projet vise principalement à concevoir un catalogue collectif consultable en ligne des archives sonores que détiennent le Département de l'Audiovisuel de la BnF et les quatre phonothèques citées plus haut, afin de faciliter pour un public dont la composition est encore incertaine la consultation de leurs fonds. La solution retenue se devra d'être évolutive et de permettre techniquement une mise en ligne de quelques fonds numérisés au cas où cela deviendrait possible ; elle ne doit pas être non plus un obstacle à l'intégration d'autres ressources, tant nationales comme les fonds du Musée national des Arts et Traditions Populaires, qu'européennes.

1.3. Objectifs définis dans la lettre de mission

Compte tenu de leurs histoires respectives et de leur enracinement territorial, les cinq partenaires ont des pratiques et des savoir-faire qui leur sont propres, tout en suivant des règles communes de catalogage exposées dans le *Guide d'analyse documentaire du son inédit* publié en 1991 par la FAMDT et par l'AFAS¹. Aussi la lettre de mission reçue en janvier 2004 nous invitait-elle à explorer les possibilités de mutualiser les catalogues en présence afin d'en élargir l'accès. Voici quels en étaient les termes :

« Depuis 1998, dans le cadre d'une convention BnF de pôle associé, quatre organismes de la Fédération des Associations de Musiques et Danses Traditionnelles, FAMDT (le Conservatoire occitan, Dastum, Metive, la Maison Méditerranéenne des Sciences de l'Homme), produisent les notices bibliographiques d'une sélection d'enregistrements sonores de leurs fonds. En 2003, ce « catalogue » virtuel compte les descriptions d'environ 2000 heures enregistrées. Chaque ensemble est toutefois isolé dans l'outil documentaire de chaque centre. De plus, ce traitement documentaire, tout en utilisant les mêmes règles de catalogage, n'est pas strictement homogène d'un centre à l'autre. De Bookline d'Archimed à un protocole OAI, etc., le projet consiste donc à étudier et à proposer les moyens de mutualiser ce traitement documentaire et d'en permettre l'interrogation. »

¹ Association Françaises d'Archives Sonores.

2. Les acteurs du projet

Bibliothèque nationale de France :

Département de l'Audiovisuel :

- Pascal Cordereix, chef du Service des documents sonores, responsable du projet
- Elizabeth Giuliani, adjointe au Directeur du Département
- Marie-Pierre Bodez, responsable de la Section de coordination du catalogage

Département de la coopération :

- Véronique Falconnet, responsable du Catalogue collectif de France
- Corinne Bouquin, chargée de mission pour le suivi des pôles associés

Fédération des Associations de Musiques et de Danses Traditionnelles (FAMDT) :

- Pierre-Olivier Laulanné, président de la FAMDT.
- Bénédicte Bonnemason, responsable du Service documentation du Conservatoire occitan-Centre des musiques et danses traditionnelles Toulouse-Midi-Pyrénées.
- Jacques Michenaud, responsable de Dastum.
- Jean-Louis Neveu, responsable du Cerdo-Metive.
- Véronique Ginouvès, responsable de la phonothèque du Centre de documentation de la MMSH.

Enssib :

- Jean-Michel Salaün, tuteur.
- Groupe projet :
 - Carine Chamoin, chef de projet
 - Damien Belvèze, chroniqueur

- Étienne Cavalié, planification et budget
- Sébastien Caudron, planification et budget
- Laurent Séguin, communication externe
- Nicolas Tocquer, communication interne

3. Analyse des fonds et pertinence du projet

3.1. Les associations et la BnF

La BnF et les associations (type Loi de 1901 pour trois d'entre elles) réunies dans le cadre de la FAMDT sont dépositaires de fonds sonores inédits. Nés dans les années 1970 du désir de sauvegarder, de valoriser et de promouvoir l'accès à la culture régionale dans chacun des territoires couverts (Poitou-Charentes-Vendée, Bretagne, monde occitan, l'espace méditerranéen français et, plus largement, de l'Europe méridionale), les fonds constitués par les associations réunies dans le pôle associé témoignent chacun d'une identité régionale. Les collections touchent à différents thèmes s'organisant autour d'axes forts : littérature orale et ethnomusicologie ; mémoire du vécu et mémoire historique ; techniques et savoir-faire ; identité linguistique et culturelle. S'alimentant à l'origine des collectes d'un réseau d'associations, d'artistes et de militants couvrant les différents territoires, certains des membres du pôle associé ont depuis longtemps développé des relations étroites avec les Universités (Rennes, Aix-Marseille) et la recherche.

3.2. Coup de sonde dans les fonds de la FAMDT

Les quatre centres de la FAMDT qui forment le pôle associé de la BnF conservent donc tous, depuis une trentaine d'années, des documents relatifs aux arts et aux traditions populaires². Un patrimoine tout à fait remarquable s'est ainsi accumulé, dont les collections sonores constituent sans conteste le point d'orgue puisqu'elles s'attachent, sans discrimination, à capter aussi bien les chansons, les

² L'association Dastum a été créée en 1972, le Conservatoire occitan en 1971, la phonothèque de la MMSH date de 1979, et UPCP-Metive réunit des fonds collectés depuis 1969.

contes, les proverbes ou les devinettes, que les récits de vie, ou encore les techniques et les savoir-faire traditionnels, l'objectif étant de se faire l'écho le plus fidèle possible de la diversité et de la richesse expressive des terroirs concernés. Ces banques sonores, qui comptent bandes magnétiques, cassettes, DAT, 33 tours, 45 tours, 78 tours et CD, sont complétées par un florilège de supports écrits – monographies et revues, mais encore feuilles volantes, manuscrits, tapuscrits – et iconographiques – tirages, diapositives, cartes postales, affiches, plans – issus bien souvent de collectes qui sont le mode d'acquisition privilégié de ces magnéto-photo-phonothèques.

Il va sans dire que ces documents, très souvent uniques, recueillis dans l'urgence pour éviter qu'ils ne sombrent dans l'oubli, constituent de véritables raretés. Revers de la médaille : cette dimension volontiers muséographique, ce statut ambigu à la croisée des bibliothèques et des musées, cette spécificité des fonds – en particulier sonores – posent un certain nombre de problèmes de description catalographique qui rendent essentielle la coopération des uns et des autres dans une harmonisation des pratiques.

Composition des fonds de Dastum, de UPCP-Metive, de la MMSH et du Conservatoire occitan

		Dastum	UPCP-Metive	MMSH	Conservatoire occitan
Fonds écrits	Revue	30	370	1600	180
	Monographies	1500	5500	84 000	3000
	Documents inédits	7000 feuilles volantes, manuscrits...	450 dossiers inédits (mémoires, thèses...)	?	1500 partitions
Fonds iconographiques	Photographies, plans, Cartes postales, affiches	27 000 photographies	50 000 photographies et cartes postales	1300 cartes et plans	8000

Fonds Sonores	Edités	1500 heures	1100 cassettes et CD	3500 cassettes, vinyles, CD	1200 heures
	Inédits	4000 heures	8000 heures + 800 heures d'archives video-cinéma	4000 heures	1400 heures
	Numérisés	2500-3000 heures	En cours	?	En cours (une station de numérisation)
Logiciel de catalogage utilisé		Alexandrie	?	Alexandrie	Texto
Indexation		Thésaurus	?	?	?

	Dastum	UPCP-Metive	MMSH	Conservatoire occitan	BnF
Nombre d'heures	4000 h	8000 h	4000 h	1400 h	?
Nombre d'heures numérisées	2500 h-3000 h	790 h	1500 h	450 h-500 h	?
Nombre de notices	40 000	?	?	442 h	de 500 à 1000
Enrichissement annuel	?	150 h	150 h	100 h	?

3.3. Topographie des fonds de la BnF

Le projet de métacatalogue ne porte pas, loin s'en faut, sur l'ensemble des collections conservées à la Bibliothèque nationale de France mais sur une petite partie d'entre elles : les documents sonores de collectes effectuées sur le territoire français. A cet égard, le Département de l'Audiovisuel se veut l'héritier direct des Archives de la Parole, de la Phonothèque nationale et, à travers elles, de l'œuvre d'hommes tels que Ferdinand Brunot, Jean Poirrot, Hubert Pernot ou Roger Dévigne, dans la volonté expresse qui fut toujours la leur de constituer un atlas phonographique et d'élaborer une géographie sonore des parlers, des langues, des dialectes et des patois des régions françaises. Aussi le Département de l'Audiovisuel de la BnF peut-il se prévaloir de posséder des témoignages uniques

de la vie dans les Ardennes franco-belges, le Berry, le Limousin, l'est de la France, les Alpes et la Provence, le Languedoc-Roussillon, les Pyrénées orientales et ariégeoises, la Vendée, la Normandie, la Bretagne ou encore la Corse. A quoi il faut ajouter les enregistrements des conversations de F. Brunot avec des artisans parisiens en 1912-1913. Dans l'état actuel des choses, l'on dispose d'environ 500 à 1000 notices d'items sachant que certains fonds majeurs, comme le fonds Massignon légué en 1985 à la BnF, ou le fonds Quilici sont en cours de traitement.

Ce tour de France des archives sonores inédites a pris un nouvel élan en 1998³ lors de la signature de la convention de pôle associé entre la BnF et les quatre membres de la FAMDT, puisqu'il s'est agi de mettre en pratique la notion de réseau documentaire, appliquée en l'occurrence à l'univers du patrimoine oral. Si bien qu'aujourd'hui, aux collections de la BnF sus-mentionnées viennent s'ajouter les 8000 heures d'enregistrements inédits de Metive, les 4000 heures de la MMSH, les 4000 heures de Dastum, et les 1400 heures du Conservatoire occitan, le tout formant un ensemble remarquable en matière d'ethnomusicologie et d'ethnographie en général⁴.

3.4. Comparaison des catalogues

La mise en place d'un catalogue collectif suppose une certaine convergence dans les pratiques de catalogage. A cet égard, trois écueils s'élèvent traditionnellement, qui concernent les champs descriptifs qui figurent sur les notices, le vocabulaire employé (thésaurus, listes, classifications) et enfin le format informatique des notices.

3.4.1. Les champs descriptifs des notices

S'agissant des documents sonores, il existe une norme internationale de description, l'ISBD(NBM), qui a pour équivalent national la norme Z44.066 publiée par l'AFNOR. Mais cette norme ne porte que sur les documents sonores édités et reste par conséquent trop peu développée pour les archives sonores inédites qui requièrent, faut-il le rappeler, un traitement spécifique. Le *Guide d'analyse documentaire du son inédit*, rédigé par la commission Documentation de

³ Le premier accord de pôle associé est signé du 30 octobre 1998.

⁴ Voir à ce propos le document du premier comité de pilotage.

la FAMDT, fut le premier ouvrage en langue française entièrement consacré aux questions de normalisation des pratiques de catalogage du son inédit⁵. Aussi est-ce à l'aune des règles qui y sont exposées que nous avons comparé les notices émanant des différents partenaires en présence. Le *Guide* préconise deux niveaux de catalogage : celui de la notice générale, qui décrit un support et son contenu ; celui d'autre part de l'item, qui est l'unité documentaire fondamentale (le conte, la chanson, le récit de vie...). Il est ainsi possible de concevoir une structure hiérarchisée où une notice générale – notice-mère – permettrait de donner accès à des notices-filles correspondant aux différents items qui figurent sur le support. En la matière, les pratiques diffèrent d'un centre à l'autre. La MMSH, Metive et le Conservatoire occitan considèrent en effet que l'unité documentaire fondamentale est *l'enquête orale* et effectuent un traitement par corpus, puis par enquête ou support, puis éventuellement par item. L'association Dastum a, quant à elle, adopté une description par item uniquement et ne rédige pas de notices générales. Malgré tout, force est de constater l'homogénéité des champs descriptifs utilisés par les uns et les autres.

3.4.2. Thésaurus et classifications

En ce qui concerne le vocabulaire utilisé, les quatre centres témoignent également d'une étroite collaboration puisque tous utilisent les thésaurus élaborés par la FAMDT et publiés dans le *Guide*, à savoir le thésaurus pour instruments de musique traditionnels dans l'Hexagone et la Corse, ainsi que le thésaurus pour les langues et dialectes dans l'Hexagone et en Corse. A quoi s'ajoute, pour les descripteurs thématiques, l'utilisation systématique du thésaurus Ethnophoto⁶. Des classifications sont également utilisées, notamment en ce qui concerne l'analyse typologique des contes et des chants⁷. Chacun des centres a pu par ailleurs

⁵ B. BONNEMASON, V. GINOUVÈS, V. PÉRENNOU, *Guide d'analyse documentaire du son inédit pour la mise en place de banques de données*, Saint-Jouin-de-Milly-Paris, Modal-AFAS, 2001. Cet ouvrage a pris le relais de : BIBLIOTHÈQUE NATIONALE, DÉPARTEMENT DE LA PHONOTHÈQUE NATIONALE ET DE L'AUDIOVISUEL, *L'Oral en fiche : manuel de traitement des phonogrammes*, Paris, AFAS, 1985, qui contient un important chapitre consacré au document inédit (p. 49-68).

⁶ Philippe RICHARD, Brigitte LOZZA, *Ethnophoto : thésaurus pour l'analyse de la photographie ethnographique du domaine français*, Paris, Editions de la Maison des sciences de l'homme, 1997.

⁷ Antti AARNE, Stith THOMPSON, *The types of the folktale : a classification and bibliography*, 2e éd., Helsinki, Suomalainen Tiedekatemia-Academia scientiarum fennica, 1981 ; Paul DELARUE, Marie-Louise TENÈZE, *Le conte populaire français : catalogue raisonné des*

procéder à des ajouts de termes qui ressortissent aux spécificités linguistiques et culturelles des régions concernées. Il est clair par conséquent que ces pratiques communes dans l'utilisation d'un vocabulaire contrôlé permettent aux quatre centres, là encore, de sonner à l'unisson.

3.4.3. Le format informatique des notices

Reste l'épineux problème du format informatique des notices, que nous n'avons, pour le moment, pas consulté⁸. Quoi qu'il en soit, le(s) format(s) en question sont étroitement liés aux logiciels de catalogage utilisés par les uns et les autres. Deux des quatre centres – le Conservatoire occitan et Metive – utilisent pour l'heure le logiciel Texto qui n'est plus commercialisé et n'offre, par là même, que peu de perspectives d'évolution. Malgré tout, des exports sont possibles au format ASCII ou en ISO 27.09. Compte tenu des limites de Texto, il est question de le remplacer par Alexandrie, qu'utilise déjà Dastum, et qui permet des exports facilités en UNIMARC, voire en XML, à condition toutefois de s'équiper des modules adéquats. Le logiciel utilisé par la MMSH est Loris, de la société Ever, mais, là encore, les regards semblent se tourner vers Alexandrie. Il y a donc de fortes chances pour qu'à court terme, les quatre centres du pôle associé disposent du même logiciel de catalogage, ce qui, en l'occurrence, faciliterait grandement le partage et l'échange des données, bref, l'interopérabilité des catalogues.

3.5. Dissonances et harmonie

La FAMDT, et en particulier les quatre centres qui nous concernent, a depuis plusieurs années déjà engagé une réflexion sur une harmonisation des pratiques de description et de catalogage des archives sonores inédites compte tenu de leur spécificité. Si un document sonore édité peut être facilement décrit – auteur, titre, éditeur, date de publication, etc – il n'en va pas de même pour les enregistrements inédits, dont la variété des formes et des supports n'a d'égale que l'approximation catalographique car les éléments de description sont bien souvent laissés à la libre

versions de France, Paris, G.P. Maisonneuve et Larose, 1985, 4 vol. ; *Répertoire des chansons françaises de tradition orale*, tome 1 : *La poésie et l'amour*, tome 2 : *La vie sociale et militaire*, Patrice Coirault, Geogres Delarue, Yvette Fédoroff et Simone Wallon éd., Paris, Bibliothèque nationale de France, 1996-2000.

⁸ Il s'agit bien du format professionnel des notices, et non des notices en affichage public que nous avons, en revanche, pu consulter.

appréciation de l'enquêteur. Face à cette situation inconfortable, pour ne pas dire anarchique, la FAMDT a édité un *Guide d'analyse documentaire du son inédit* qui édicte un certain nombre de règles descriptives élémentaires à l'intention des centres d'archives sonores. Il doit être fait mention notamment du genre de document, du contexte d'enregistrement, de la forme et de l'expression musicale, du nom de l'enquêteur, des interprètes, de la date d'enregistrement, de la langue utilisée, etc., dans une sorte de cascade hiérarchique qui va du général – le support d'enregistrement – au particulier – l'item, c'est-à-dire, l'unité documentaire (le chant ou le conte par exemple). Les archives sonores inédites requièrent par conséquent une analyse tout à fait spécifique, une finesse descriptive que le catalogage « classique » est incapable de mettre en œuvre.

En examinant les catalogues de Dastum et de Metive, il est apparu que les instructions de la FAMDT étaient scrupuleusement respectées. S'agissant de l'association bretonne, le catalogue propose onze entrées possibles, c'est-à-dire onze champs d'interrogation, à savoir :

Interprète
Titre
Expression
Genre
Danse
Terroir
Mot clef
Collecteur
Commune d'enregistrement
Événement
Langue

L'utilisateur peut mener sa recherche en remplissant librement un ou plusieurs de ces champs, ou bien en accédant à des index. En cliquant sur « **genre** » par exemple, une liste d'entrées s'affiche, à savoir : chanson, conte-légende-récit, expression vocale, etc. Libre à l'utilisateur de sélectionner le genre qui lui

convient. Pour certains champs, la liste est arborescente et hiérarchisée⁹, ce qui affine la recherche. Une fois la requête formulée et lancée, une liste d'items s'affiche, dont il est possible de visualiser les notices détaillées. Le document numérisé n'est cependant consultable qu'à la médiathèque de Rennes et dans les antennes Dastum, pour éviter les piratages intempestifs et les diffusions illicites des documents.

La configuration du catalogue de Metive diffère quelque peu du modèle breton. Neuf champs d'interrogation mènent au catalogue :

Fonds
Thème
Pays
Terroir
Lieu d'enregistrement
Mot du titre
Auteur, interprète, enquêteur
Année
Mot clef

L'utilisateur peut, ici aussi, accéder à des listes en cliquant sur chaque entrée, mais celles-ci ne sont pas arborescentes. Ce qui n'invalide pas pour autant ce catalogue, loin s'en faut, car la requête donne accès à la notice descriptive du support – par exemple une cassette audio – décrit de la façon suivante : résumé, lieu, date, enquêteur, informateur, circonstances, analyste, nombre de pistes de l'original, durée totale. Pour les documents numérisés, il est possible d'accéder à la description détaillée des items, et ce, conformément aux instructions du *Guide* de la FAMDT.

Les catalogues de la Maison Méditerranéenne des Sciences de l'Homme (MMSH) et du Conservatoire occitan, qui ne sont pas accessibles par Internet, n'ont pas pu être consultés. Mais il est fort probable qu'ils se situent eux aussi dans cette lignée catalographique qui tend vers une description très fine des

⁹ Cf. *Guide d'analyse documentaire du son inédit*, p. 44 et suiv.

supports jusqu'aux items, puisqu'ils utilisent Alexandrie et Texto, deux logiciels préconisés par la FAMDT. L'heure n'est donc pas au repli solipsiste ni aux pratiques atomisées mais bien à l'harmonisation des pratiques, d'autant qu'une réflexion a d'ores et déjà été engagée à la fois sur la mise en place de thésaurus¹⁰ et sur la mutualisation des ressources, comme le confirment ces quelques mots de Véronique Ginouvès et de Véronique Pérennou dans leur préface au *Guide* : « La mise en place d'outils partagés, de catalogues accessibles en tout point du globe laisse entrevoir des perspectives nouvelles et exaltantes pour des archives qui jusqu'ici étaient quasiment inaccessibles »¹¹. Une mutualisation des catalogues des centres associés à la Bibliothèque nationale de France semble donc tout à fait possible, d'autant qu'aucun obstacle technique ne s'est, pour l'instant, élevé contre cette entreprise, bien au contraire : les formats utilisés sont compatibles¹², la description catalographique est conforme aux normes de la FAMDT et la numérisation est en cours dans au moins trois centres, ce qui ouvre la voie vers une consultation directe, en ligne, des documents. Parmi ces « outils partagés » évoqués par Mmes Ginouvès et Pérennou, il en est un qui suscite un enthousiasme lui aussi partagé : le métacatalogue, particulièrement défendu par M. Cordereix, du Département de l'Audiovisuel de la BnF.

3.6. Le terrain d'entente du métacatalogue

Que le projet de métacatalogue ait été proposé par Pascal Cordereix aux élèves conservateurs de l'ENSSIB n'est sans doute pas anodin, car la BnF a tout à gagner d'un tel outil. Force est en effet de constater le caractère lacunaire du catalogue Bn-Opale-plus qui ne répond que de façon incomplète aux exigences des documents sonores malgré les adaptations dues au format INTERMARC. La mise en place d'un métacatalogue serait un moyen pour la BnF de valoriser l'accès à ses ressources puisqu'il s'agirait pour elle de tendre vers la précision descriptive des quatre centres de la FAMDT. Ce faisant, les documents sonores obtiendraient leurs

¹⁰ Le Guide sus-mentionné fait état de deux listes, la première concernant « les instruments de musique traditionnels dans l'hexagone et en Corse », la seconde portant sur « les langues et dialectes dans l'hexagone et en Corse », à quoi il faut ajouter le projet encore à l'étude d'un thésaurus d'ethnographie française sous les auspices de Dastum.

¹¹ *Op. cit.*, pp. 5-6.

¹² Propos rapporté par Nicolas Tocquer à l'occasion d'un entretien informel avec Jean-Luc Ramel de Dastum.

lettres de noblesse dans cette institution où ils font pour l'heure figure de parent pauvre. Inversement, la FAMDT aurait accès aux ressources de la BnF qui viendraient en complément de ses propres fonds. Un métacatalogue semble par conséquent un terrain d'entente idéal où chacun – BnF et FAMDT – pourrait tirer parti de sa relation avec l'autre dans ce qui n'est ni plus ni moins qu'un échange de bons procédés.

Au-delà de ces considérations politiques, la mise en place d'un métacatalogue permettrait aux centres de la FAMDT d'étendre leur public qui demeure pour l'essentiel dans les limites de leurs aires respectives de collectage. Or la valorisation du patrimoine passe par un partage des ressources, par la mise place de catalogues collectifs et de réseaux¹³. Il n'est pas interdit en effet d'imaginer des musiciens bretons – qui forment l'essentiel du public de Dastum – s'inspirant du répertoire provençal ou poitevin et vice-versa. Le monde de la recherche pourrait également tirer un considérable bénéfice de cette mutualisation, et les archives sonores en sortiraient grandies en obtenant le statut de source scientifique à part entière¹⁴. Encore faudrait-il, pour ce faire, que les fonds interrogeables ne se limitent pas aux documents sonores, mais incluent toute la « littérature grise » qui les complète.

Comment un tel partage des ressources peut-il être mis en œuvre ? Un métacatalogue est, littéralement, un catalogue qui se situe au-delà ou en-deçà d'autres catalogues, et qui permet de les interroger simultanément quand bien même ils utiliseraient des formats différents. Une première semaine de stage passée à la BnF du 9 au 13 février 2004 nous a permis de prendre la mesure des moyens techniques permettant cette mutualisation des ressources, qui vont de la solution toute prête par le biais du logiciel Bookline que propose la société Archimed, jusqu'à la conversion des notices au format XML. Les données XML présentent en effet l'avantage de répondre aux réquisits descriptifs des documents sonores en vertu de leur caractère extensible, hiérarchique et interopérable, ce qui n'est pas le cas du format UNIMARC. Il s'agirait donc de dégager une structure

¹³ Comme le soulignent à juste titre V. Ginouvès et V. Pérennou : « Sans surestimer la fonction des autoroutes de l'information dans l'avenir, on peut imaginer que les réseaux deviendront un outil de promotion du patrimoine oral », *op. cit.*, p. 6.

¹⁴ Cf. le chapitre intitulé « L'oral comme source scientifique » du *Guide* de la FAMDT, p. 11.

logique de description du document sonore, de déterminer en quelque sorte, des métadonnées auxquelles serait associée une DTD (Document Type Description), la question étant de savoir si une telle DTD existe qui correspondrait aux archives sonores et à leur spécificité.

Les lignes qui précèdent ont tenté d'asseoir la légitimité du projet qui nous a été proposé. Il nous appartient désormais d'étudier les modalités de ce chantier ambitieux.

4. Redéfinition de la prestation

L'esprit du projet tel qu'il nous a été présenté au cours de notre première visite à la BnF était essentiellement technique puisqu'il s'agissait de trouver un procédé idéal permettant l'interrogation des cinq catalogues à l'aide d'une seule interface, si elle existe. Ce choix mérite d'autant plus un cas d'étude que les solutions possibles sont multiples et que les procédés techniques sont en pleine évolution.

Après réflexion il nous est apparu que la recherche puis le choix d'une solution technique adéquate est une démarche qui ne peut se faire que si deux travaux préalables ont été menés à bien : d'une part une définition précise de ce que l'on veut faire (les besoins, les résultats et les avantages attendus de ce métacatalogue), et d'autre part une connaissance tout aussi précise de l'existant à partir duquel ce métacatalogue sera créé.

Dans un deuxième temps, il est également apparu que la dimension humaine du projet était plus importante que ce qui était conçu dans l'ordre de mission. Ceci nous a conduits à intégrer une partie d'analyse des besoins des publics et des centres.

Le travail qui nous est donné à faire présente donc deux parties : l'élaboration d'un cahier des charges à partir d'une analyse de l'existant et des besoins, ainsi que la recherche de solutions techniques donnant lieu à des scénarios.

4.1. Analyse de l'existant / connaissance du terrain

Cette partie vise à recueillir les données permettant l'élaboration d'un cahier des charges.

- Règles de catalogages en vigueur dans chaque centre.
- Niveau de traitement documentaire (voir notamment si le niveau d'analyse jusqu'à l'item prôné par le guide de traitement documentaire de la FAMDT est généralisé).
- Paramètres techniques à prendre en compte pour le métacatalogue (format des données, logiciel utilisé et ses compatibilités...).
- Publics fréquentant ces centres, leurs habitudes, leurs thèmes de recherche ou leurs intérêts, leur localisation géographique par rapport au centre qu'ils fréquentent.
- Les ressources en personnel.
- Les ressources financières de chaque centre et celles que la BnF est prête à mettre dans le projet (ce qui inclut la disponibilité ou non du service informatique pour adapter un logiciel libre ainsi que la possibilité d'héberger le catalogue collectif).
- Les projets de numérisation en cours, leur philosophie (logique de conservation ou projet de diffusion).
- Le rythme annuel d'enrichissement des collections.
- La politique suivie par les centres en réponse aux problèmes juridiques posés par la communication des archives sonores.

Le but de cette partie du travail est de déterminer la compatibilité des notices des différents centres et de celles de la BnF à la fois dans le catalogage et dans le niveau de traitement documentaire, pour voir tout d'abord si le catalogue commun est rendu possible par les efforts d'homogénéisation qui ont été faits, et dans ce cas, de décider si le

catalogue s'arrête au signalement de la notice générale, ou s'il va plus loin dans la description (item), les champs de recherche qu'il prendra en compte, et enfin de savoir si le travail qui a été fait permet ou non d'envisager dans un proche avenir un accès direct aux documents.

4.2. Besoins et attentes

A côté de l'étude de faisabilité technique (catalogage, traitement documentaire, divers paramètres techniques...), il est important que le prestataire et chacun des acteurs décident de la configuration du catalogue qu'ils souhaitent. En effet, le champ des possibilités est très vaste car un très grand nombre de paramètres peuvent être pris en compte, il y a donc des choix à opérer en amont de la recherche technique.

Le premier choix à opérer concerne le périmètre du métacatalogue. Pour la BnF, doit-on inclure tous les enregistrements sonores, y compris édités, s'ils intéressent l'ethnologie régionale, ou se limite-t-on aux archives sonores inédites ? Pour les quatre centres de la FAMDT se pose la question de toute la documentation (photographies, témoignages écrits, archives de chercheurs) également conservée par ces centres. Le pôle associé ne porte que sur les enregistrements sonores, mais la mise en contexte du document sonore est importante pour sa compréhension et son analyse. Se pose aussi la question des documents audiovisuels qui peuvent couvrir la même thématique que les enregistrements sonores, voire les compléter. Il va sans dire que le choix de la solution technique aboutissant à la réalisation de ce catalogue commun ne sera peut-être pas la même s'il couvre un seul ou plusieurs types de documents. Ce choix est donc relativement important pour l'aspect final du projet.

Un autre élément important et influent sur l'ampleur à donner à ce catalogue collectif, est la question du public destinataire. Quel public le prestataire et les acteurs du projet souhaitent-ils toucher ? Le public actuel est-il intéressé ? Le public potentiel est-il un public spécialisé demandeur d'un outil de recherche très performant, ou cherche-t-on plutôt à atteindre par le biais du catalogue une plus grande visibilité envers le grand public ? La question du public pourrait amener à se demander dans quelle mesure ce catalogue collectif trouverait réellement un

écho. On peut raisonnablement partir du présupposé que la mutualisation des catalogues est légitime, puisque la publication des catalogues des pôles associés est prévue dans la convention qu'ils ont signée avec la BnF en 1998 et que celle-ci constitue la conséquence logique des subventions qu'ils perçoivent. Par contre, une étude qualitative des publics actuels et des publics potentiels que ce catalogue commun serait susceptible de toucher est intéressante. La connaissance de leurs habitudes, de leur localisation par rapport aux différents centres, de leurs pratiques, de leur thèmes de recherche permettrait notamment de discuter sur une base concrète de la pertinence du projet (est-ce que la comparaison entre les différents us et coutumes des régions est fréquente ? Le public est-il suffisamment dispersé dans l'espace pour justifier ce catalogue ?). Cette étude permettrait surtout de bien cibler les champs et les aspects à prendre en compte en ce qui concerne l'aspect à donner à ce catalogue collectif (public de chercheur demandeur d'un instrument de recherche très pointu, grand public amateur désirant localiser une notice). Pour définir au mieux les besoins, le groupe a prévu de se livrer à une étude du public.

Quelle est l'évolution souhaitée de ce catalogue : accès direct aux documents sonores pour tous, ou seulement en extranet ? Accès à un panel d'enregistrements sur internet ? Elargissement souhaité et probable à d'autres centres ?

En clair il s'agit de recueillir les données concernant ce qui est voulu, sur l'aspect et le niveau du catalogue qui est souhaité par tous. A partir de là, le groupe établira le cahier des charges :

- Contraintes techniques du cahier des charges (format des données, travail à faire sur chaque catalogue, nécessité d'un procédé technique permettant plus ou moins de hiérarchie dans les données...).
- Contraintes humaines : les centres et la BnF ont-ils les moyens de retravailler leurs notices, de les saisir par la suite directement en XML, etc.
- Contraintes juridiques : l'accès aux différents niveaux des documents (notice générale, notice de l'item, document lui-

même, nom de l'informateur etc.) est-il juridiquement envisageable ?

- Cette partie du travail vise également à dire si ce qui est voulu est faisable compte tenu de l'existant. Ce travail devrait aboutir à un cahier des charges permettant la recherche une fois le champ des possibles délimité.

4.3. Choix des solutions techniques les mieux adaptées aux contraintes

En amont de cette partie du projet, un travail préalable aura été réalisé, qui permettra de déterminer les points forts et les points faibles des solutions possibles (certains procédés techniques nécessitent plus de remaniement des notices que d'autres, certains permettent une plus grande souplesse, etc.).

Ce travail de recherche aboutit à la présentation de plusieurs scénarios prenant en compte différents paramètres (coûts, complexité, possibilités d'évolutions....), après la mise à l'écart des techniques semblant inadaptées (différence d'échelle, mauvaise compatibilité avec les documents sonores...). Le groupe se réserve la possibilité de conclure qu'en l'état actuel des procédés techniques, aucune solution ne permettrait de remplir les conditions du cahier des charges.

Différents paramètres sont à prendre en compte dans le choix d'une solution technique : est-ce qu'il vaut mieux publier sur Internet les catalogues des deux associations qui ne l'ont pas encore fait pour que le métacatalogue aille y puiser, ou bien vaut-il mieux que les notices soient fournies à l'hébergeur ? Quelles seraient les implications diplomatiques des différentes solutions envisagées ? La spécificité du traitement documentaire des enregistrements sonores inédits implique-t-elle que le développement en interne d'un procédé est rendu obligatoire, ou est-il plus simple de faire appel à un prestataire ? Dans tous les cas, quelle serait la disponibilité du service informatique pour travailler sur ce projet éclaté entre plusieurs centres ?

Il est à signaler que compte tenu du temps imparti à la réalisation de ce projet, le groupe ne pourra en aucun cas se livrer à une réelle étude de marché des différents prestataires et des différentes solutions. Si nous tâchons de jauger du

coût financier de chaque solution afin d'aider au choix, nous ne pourrions pas présenter un devis complet et précis de chaque étape des scénarios.

Ces scénarios doivent être présentés lors du 2^{ème} comité de pilotage, à l'issue duquel le commanditaire choisira le scénario lui convenant le mieux.

5. Contraintes

Le présent paragraphe n'a pas pour rôle d'apporter nécessairement de nouveaux éléments mais, dans un souci de clarté, de reprendre ceux qui ont été considérés comme incontournables pour l'élaboration du présent projet.

5.1. Hétérogénéité relative des catalogues

Coexistent indéniablement des pratiques variées entre les partenaires, tant sur plan du catalogage que sur les ambitions concernant le niveau de collaboration. Sur ces deux aspects néanmoins, il est important de souligner que ne cesse de croître une volonté de coordination et d'homogénéisation.

Au regard des catalogues existants et des divergences dues aux histoires respectives, l'accent a dû être mis, pour la définition des champs d'interrogation du métacatalogue, sur le dénominateur commun. Il a donc fallu définir une liste de renseignements que l'on pouvait retrouver dans tous les catalogues concernés, éventuellement sous diverses appellations.

La situation des fonds de la BnF est très particulière : s'est révélée en effet la nécessité de créer une base de données spécifique ne contenant que les notices de documents concernés par le présent projet. Il est en effet impossible, étant donné le contenu, et surtout la masse des notices existant aujourd'hui sur Bn-Opale plus, d'y sélectionner à la volée, lors d'une recherche, les seuls documents d'archives sonores concernant le folklore français. Cette extraction préalable ne peut être faite que manuellement par un conservateur, qui doit par la suite veiller à l'enrichissement progressif de cette base à mesure de la création de nouvelles notices par les catalogueurs.

Le projet de métacatalogue d'archives sonores représentant en soi une entreprise ambitieuse, il s'est révélé impossible de profiter du travail réalisé pour

repenser l'homogénéité des catalogues et aboutir à une ressemblance parfaite dans la structure même des notices et les modes de catalogage. Cette impossibilité a plusieurs motifs :

1. Cela n'a pas été demandé aux élèves de l'ENSSIB, ni au moment de l'exposition première du projet, ni lors du premier comité de pilotage.

2. Les recherches sur les outils existants qui servent à constituer un catalogue collectif, permettent d'acquérir des connaissances et des compétences qui ne sont d'aucun intérêt lorsqu'il s'agit de définir des manières de cataloguer des documents d'archives sonores issus de collecte. Il n'était pas possible de réutiliser la documentation accumulée pour cette nouvelle orientation, les problématiques ne se recoupant pas.

3. Les difficultés à s'entendre sur les modalités d'un catalogue collectif et les choix à faire à ce niveau-là sont suffisants pour ne pas y ajouter en supplément une réflexion sur les manières de cataloguer.

4. Lancer une réflexion sur une utilisation plus riche des outils informatiques pour le catalogage des documents sonores, retarderait considérablement la constitution d'un métacatalogue, dont un des atouts est, à ce jour, son degré d'innovation au regard de ce qui existe (et surtout de ce qui n'existe pas encore) actuellement dans le monde. Il va de soi que cet atout risque de disparaître très rapidement.

5.2. Hétérogénéité des acteurs

Les acteurs sont très diversifiés, avec de profondes différences de statut entre, d'une part les associations membres de la FAMDT, et d'autre part le Département de l'Audiovisuel de la Bibliothèque nationale de France. Il faut également souligner les particularismes au sein du pôle associé FAMDT puisque, par exemple, la MMSH dépend du CNRS. Ce fossé engendre des contraintes de type institutionnel dont il faut impérativement tenir compte. Il est en effet évident qu'un tel projet ne saurait se concrétiser sans la participation totale et volontaire de tous ses membres. Cela implique une capacité d'action et un droit de décision réels, et ce pour chacun des partenaires.

En outre, les besoins ne sont pas nécessairement convergents. Les missions sont différentes, les modes de fonctionnement également. Par-dessus tout, aucun des partenaires n'est autonome : tous sont intégrés dans une institution plus vaste, avec laquelle ils doivent mener une politique cohérente.

Sur la question de la dispersion des acteurs, une des contraintes majeures à laquelle se sont trouvés confrontés les élèves de l'ENSSIB, a été, de façon regrettable, l'éclatement géographique des différents partenaires. Dans la plupart des cas, il fut en effet impossible de se rendre sur place, faute de moyens financiers. De multiples problèmes, notamment pour la connaissance réelle des fonds de chaque centre, sont nés de cette situation.

5.3. Contraintes de mise en place

Les élèves de l'ENSSIB ont dû prendre en compte plusieurs éléments contraignants pour la mise en place du futur métacatalogue (ou catalogue virtuel).

Il s'agit d'abord d'un problème de conditions de travail non pour la mise en place du projet lui-même, mais pour la réflexion préalable. En effet, et c'est un problème dont nous n'avons pas réellement pu prendre la mesure, à cause de l'éloignement géographique, il fallait tenir compte des compétences techniques et informatiques des responsables des fonds d'archives sonores, dont les expériences et les acquis sont différents. Dans le même domaine, il était indispensable de considérer l'importance du parc informatique de chacun, ainsi que la question d'une présence éventuelle d'un responsable informatique disponible à proximité des centres : de nouveaux problèmes de maintenance vont en effet se poser.

Dans ce même cadre des contraintes pour l'élaboration du projet, le chiffrage budgétaire de chacun des scénarios s'est révélé extrêmement difficile et souvent hasardeux. En effet, les sociétés prestataires de services contactées ont demandé, afin de pouvoir faire un devis, de disposer de tous les éléments rentrant dans le projet. Or il fallait conjointement rassembler ces éléments et en déduire les coûts – ce qui est à peu près impossible. En outre, de nombreuses alternatives comprises au sein de chaque scénario peuvent faire varier considérablement les investissements nécessaires, en particulier concernant les questions de

programmation informatique, selon que celle-ci peut être faite en interne, ou doit être demandée à une société privée.

Par ailleurs, pour l'installation ultérieure (et le fonctionnement) du métacatalogue, il faut demeurer conscient du fait que le mode de fonctionnement budgétaire des associations diffère beaucoup de celui que l'on trouve habituellement dans le monde des bibliothèques : elles sont généralement alimentées par des subventions issues pour l'essentiel des collectivités locales.

5.4. Contraintes de fonctionnement

Pour la mise en place du futur portail d'interrogation, un des principaux enjeux réside dans la politique du service qui, à la BnF, gère le Catalogue Collectif de France. En effet, du fait de sa structure actuelle, celui-ci ne peut abriter des sous-catalogues. Mais il désire vivement s'orienter dans cette direction, et héberger dans l'avenir un certain nombre de catalogues collectifs constitués soit autour de thèmes précis, soit par le statut des institutions qui y participent. Or cet avenir ne se situe pas avant 2007.

Mais à partir de cette date, il est fort probable que le CCFr cherche à intégrer le catalogue collectif des documents d'archives sonores. Il faut donc prévoir l'emplacement du portail d'interrogation jusqu'en 2007, mais aussi une capacité de ce portail à être déplacé. Cela implique un peu les aspects techniques, mais plus encore des questions politiques.

5.5. Contraintes d'évolutivité

Enfin, deux exigences de principe ont été posées dans la capacité du catalogue collectif à évoluer.

D'une part, il faut que les outils informatiques utilisés et la répartition des rôles, permettent une numérisation des documents sonores dont les notices sont accessibles par le métacatalogue. Les modalités quant à l'accès de ces documents numérisés¹⁵ restent évidemment à préciser, en fonction des partenaires, mais aussi en fonction de chacun des documents eux-mêmes.

¹⁵ Accès possible par n'importe qui en ligne, ou à partir de certains postes seulement, ou à partir d'une certaine date seulement, etc.

D'autre part, il est important que le projet ne soit pas limité aux seuls acteurs d'aujourd'hui, mais dispose d'une capacité d'extension à de nouveaux partenaires, qu'il s'agisse de musées, de membres de la FAMDT, de centres d'archives ou d'autres bibliothèques. Pour chaque scénario envisagé, il faut tenir compte de la *possibilité* d'un outil à intégrer aisément un nouveau partenaire, ainsi que du *coût* de cette intégration, qui n'est pas le même, tant s'en faut, d'un scénario à l'autre.

Compte-rendu décisionnel du 1^{er} comité de pilotage

Partie 2 – Présentation des scénarios

Objet du présent document

Cette seconde partie traite des scénarios proposés lors du second comité de pilotage. Ils ont été élaborés à l'issue de la réflexion portant sur l'existant, et des connaissances accumulées sur l'évolution des moyens informatiques à disposition. L'exposé de ces scénarios est cependant précédé d'une description dans les grandes lignes des propriétés du catalogue collectif, dans ses fonctionnalités tant sur le plan documentaire pur que dans ses autres aspects.

Méthode suivie

Les investigations visant à mieux connaître les partenaires n'ont pas cessé jusqu'à l'élaboration proprement dite des scénarios, de même que les recherches effectuées pour se familiariser avec les langages informatiques. Il est extrêmement important toutefois de préciser que le projet, considéré dans un premier dans ses seuls aspects techniques, s'est révélé parallèlement problématique et formateur sur le plan institutionnel. Les difficultés à mettre en place un réseau informatique buttent en effet souvent sur la nécessité de laisser à chacun le sentiment de pouvoir s'exprimer. Nous avons donc intensifié nos relations avec les membres de la FAMDT qui, étant dans l'impossibilité de se rendre sur Paris au moment des comités de pilotage, risquaient de se sentir exclus de l'évolution de notre réflexion, qui les concernait au premier chef.

1. Le métacatalogue : une première esquisse

1.1. Elaboration d'une liste des champs d'interrogation

Les champs d'interrogation sont destinés à figurer dans le formulaire de recherche tel qu'il apparaîtra sur le portail d'interrogation. Il va sans dire que seule une concertation entre les différents centres de la FAMDT et la BnF pourra dire quels sont les champs à retenir ainsi que leur correspondance. Toutefois la réponse des centres au questionnaire qui leur a été envoyé, ainsi qu'une première réflexion publiée sur le sujet¹⁶ nous a permis de suggérer une liste des champs interrogeables ainsi que leurs correspondances dans les différents catalogues.

Les notices des quatre centres d'archives sonores et celles de la BnF ne comportent pas les mêmes champs descriptifs. Le métacatalogue se heurte donc ici à une première difficulté consistant en l'élaboration d'une base commune qui ne peut se faire qu'au prix d'une perte de finesse descriptive.

Les différents champs suggérés sont les suivants : Titre, Responsabilité intellectuelle (auteur, collecteur, interprète...), Mot clef, Lieu (lieux cités ou lieu d'enregistrement), Date (de l'enregistrement ou période citée), Fonds/collecte et Langue.¹⁷

Un champ de recherche de type « moteur de recherche », opérant une recherche en texte libre sur tous les champs des notices, y compris ceux qui sont en texte (résumé, notes, incipit...) et qui sont donc plus difficilement intégrables à un formulaire de recherche standard, permettrait de compenser les limites posées par ce formulaire commun.

Bien que limités, ces champs posent un problème relatif aux notices de la BnF qui ne sont pas indexés par sujet. On essaiera dans un premier temps de pallier ce manque en faisant correspondre pour les notices de la BnF le champ « Mot clef » à une recherche par mot dans tous les champs de la notice.

¹⁶ Cf. l'article en ligne de Véronique Ginouvès : <http://afas.mmsb.univ-aix.fr/cc/INTRO.HTM>.

1.2. Les autres fonctionnalités de l'interface

Comme pour la question précédente, seule une concertation entre les différents partenaires pourrait légitimement décider des fonctionnalités à mettre en place sur le portail de recherche. Ces choix sont en effet en partie techniques car la possibilité de limiter la recherche à tel ou tel critère nécessite une grande uniformité dans l'attribution de ce critère, mais aussi en partie stratégiques (simple interface commune d'interrogation des catalogues, ou véritable portail offrant en plus un certain nombre de service ?). Rappelons que la possibilité que de nouveaux partenaires viennent se joindre à ce catalogue virtuel implique que l'interface d'interrogation ne soit pas « verrouillée » sur une seule forme de description, même si l'uniformité relative des notices des centres de la FAMDT rend cela tentant.

Les fonctionnalités souhaitées sont globalement classiques : possibilité de limiter la recherche en amont à un seul ou plusieurs catalogues, et à terme aux enregistrements qui sont en ligne.¹⁷ La capacité à gérer des droits d'accès différenciés est également indispensable, puisque certaines informations concernant les informateurs ou des parties de leurs témoignages sont confidentielles. En ce qui concerne la recherche elle-même, la possibilité d'opérer au moins à deux niveaux (grand public et expert) paraît être un minimum, de même que la capacité à pouvoir consulter des listes associées à chaque champ à chaque fois que c'est possible (thésaurus, listes closes, formes d'autorités pour les personnes ou les lieux...). Pour la recherche en expert, il faudrait qu'il y ait la possibilité de faire des combinaisons de recherche à l'aide des outils standards : opérateurs booléens et troncature.

¹⁷ Voir en annexe 1.1 pour une suggestion plus détaillée de concordance des champs.

¹⁸ Une sélection par différents critères comme le genre de document (conte, chanson, etc.) a été souhaitée par certains centres. Malheureusement, même si la classification des centres de la FAMDT sur le sujet est homogène, en l'état actuel, la BnF n'offre pas ce type de distinction sur ses notices.

2. Les scénarios

2.1. Z39.50 : le poids de l'expérience

2.1.1. Définition technique

Z39.50 est une norme créée en 1984 par la Library of Congress qui spécifie la structure des données et les règles d'échange qui permettent d'interroger simultanément plusieurs catalogues, fussent-ils hétérogènes, et d'avoir accès à leurs ressources. A partir d'une interface d'interrogation, l'utilisateur peut ainsi lancer une même recherche sur plusieurs bases distantes, utilisant des formats et des logiciels documentaires différents, et récupérer de façon unitaire et centralisée des informations à l'origine éparées. Un tel catalogue collectif virtuel repose sur une architecture client-serveur. Le client Z39.50 est un logiciel qui adresse des requêtes à des serveurs Z39.50 qui eux-mêmes interrogent les bases sur lesquelles ils sont installés, récupèrent les données sélectionnées puis les transmettent au client Z39.50. A l'origine, les serveurs étaient interrogés tantôt directement par les clients, tantôt via le protocole TCP/IP (Internet). Désormais, il est de plus en plus fait appel aux technologies du Web qui reposent sur le protocole http, et Z39.50 se trouve, dans la plupart des cas, associé à des « passerelles web »¹⁹. Dans la dernière version de la norme, publiée en 2003²⁰, l'*eXtensible Markup Language*, plus connu sous le nom de XML, constitue l'enjeu majeur des révisions effectuées, tant au niveau de la structuration des données qu'au niveau du langage de requête.

2.1.2. Avantages

Dans le contexte précis qui est le nôtre, Z39.50 semble une solution tout à fait envisageable dans la mesure où le logiciel Alexandrie peut être équipé d'un serveur Z39.50. Or, sur les quatre centres, Dastum l'utilise déjà tandis que le UPCP-Metive, la MMSH et le Conservatoire occitan comptent l'acquérir sous

¹⁹ Voir schéma ci-après.

²⁰ NISO Z39.50-2003.

peu²¹. Dans une première étape, les quatre partenaires disposeraient donc du même logiciel de traitement documentaire équipé dudit serveur Z39.50, ce qui allégerait considérablement le travail de paramétrage et limiterait tout autant les risques d’erreurs. Cette solution présente en outre l’avantage de ne bouleverser en rien les pratiques de catalogage des uns et des autres et se place par là même dans la continuité des réflexions menées jusqu’à présent par les membres de la FAMDT.

2.1.3. Inconvénients

Reste en effet en suspens la question des fonds documentaires conservés à la Bibliothèque nationale de France : en l’état actuel des choses, les notices d’enregistrements sonores inédits manquent cruellement de visibilité dans le catalogue Bn-Opale plus. L’idéal serait par conséquent , dans une deuxième étape, de créer une base spécifique de notices d’archives sonores inédites et de la rendre accessible *via* un serveur Z39.50, dont le CCFr, qui mène une politique active de soutien à la création de catalogues collectifs, pourrait éventuellement être l’administrateur.

La troisième et dernière étape de ce scénario consisterait à définir l’hébergement du client Z39.50 qui serait assuré par la BnF ou par la FAMDT. Si l’implication financière de la BnF et son rôle fédérateur légitiment la première solution, la grande majorité des fonds est détenue par la FAMDT qui peut, à bon droit, revendiquer l’hébergement du client Z39.50, et ce, d’autant plus qu’une réflexion est précisément en cours sur la création d’un portail de la FAMDT qui donnerait accès aux catalogues des associations concernées.

Il apparaît *in fine* que la mise en place d’une telle architecture client-serveur pose un certain nombre de problèmes politiques, qui sont autant d’obstacles prévisibles au métacatalogue.

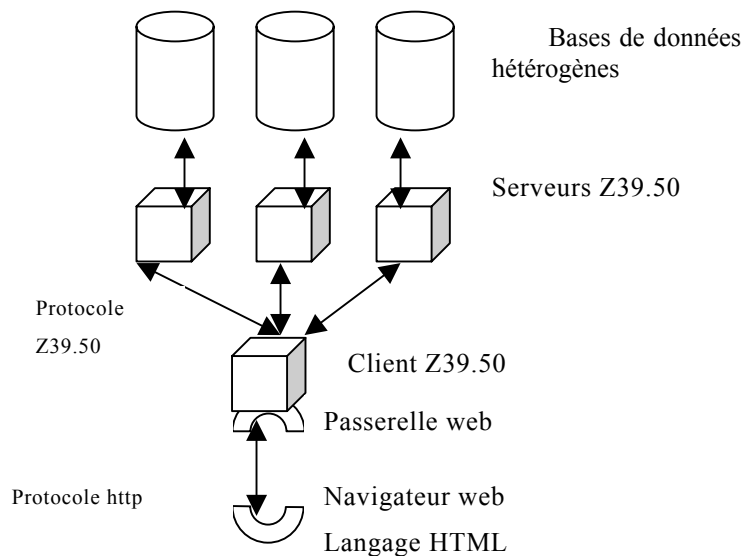
2.1.4. Investissements

Ces trois étapes – achat et installation de serveurs Z39.50, création d’une base BnF de notices d’archives sonores, achat et installation d’un client Z39.50 – exigent de fortes compétences informatiques (paramétrage²² et maintenance) et

²¹ Voir plus haut.

²² Le paramétrage du système consiste, en grande partie à collecter les profils des serveurs Z39.50 afin de les rendre accessibles par le client. Le profil d’un serveur Z39-50 est l’ensemble

représentent un investissement qui voisine les 100 000 euros²³. Malgré les évolutions considérables qu'a subies la norme depuis sa création au début des années quatre-vingt, on ne peut finalement s'empêcher de penser que l'avenir n'est pas devant elle, car désormais les technologies du Web offrent des services comparables pour un coût bien moindre.



2.2. OAI-PMH : un moissonnage balisé

2.2.1. Définition technique

A sa naissance en 2000, l'*open archive initiative* a d'emblée été présentée comme une alternative à Z39.50.

D'une installation plus simple, elle permet notamment de collecter des métadonnées de formats hétérogènes en les recueillant dans un type de description minimal, le Dublin Core.

des caractéristiques techniques qui doivent être communiquées au client Z39-50 pour que celui-ci puisse lui adresser des requêtes. Ces informations essentielles ont paradoxalement pu nuire à la diffusion de Z39-50 en raison de leur complexité. Une simplification s'avérait par conséquent nécessaire, qui a bel et bien eu lieu avec la création du « Bath profile » qui permet une utilisation plus souple de la norme en en proposant diverses utilisations.

²³ Cette estimation grossière prend en compte l'achat de serveurs Z39.50, d'un client Z39.50 du type de Book-Line de la société Archimed, ainsi que la main d'œuvre nécessaire à l'installation du système.

La version actuelle de ce protocole, l'OAI-PMH II (deuxième version de l'*Open Archive Initiative Protocol for Metadata Harvesting*) mis au point en juin 2002 articule trois concepts principaux : la ressource, l'item et l'enregistrement.

Les ressources (*resources*) ne sont rien d'autre que les objets ou les données qu'un musée, une bibliothèque, une base de données, ou un site web mettent à disposition du public.

L'item (*item*) est la partie d'un entrepôt qui peut fournir les métadonnées correspondant à une seule ressource dans un certain nombre de formats (au minimum dans le format XML comportant les 15 balises Dublin Core). Chaque item se voit attribuer un numéro identifiant.

L'enregistrement (*record*) est la représentation d'une ressource dans celui des formats proposés par l'item qui est demandé dans la requête.

OAI n'est pas un moteur de recherche. Ce système se borne à définir deux choses : comment l'échange entre un moissonneur et un entrepôt s'effectue et comment les métadonnées échangées sont représentées lors de cet échange, et seulement lors de cet échange. D'où la nécessité de lier OAI-PMH à un moteur de recherche (SDX ou Z39.50).

L'image qui a été choisie pour symboliser ce protocole est celle du moissonneur (*harvester*) qui rassemble le contenu de plusieurs silos dans lesquels chaque centre verse une partie définie à l'avance de ses métadonnées afin de le rendre disponible pour une recherche. Le schéma figurant en fin de chapitre peut donner une idée des opérations qui se déroulent lors d'un de ces transferts.

OAI sert en général à gérer des métadonnées qui sont plus hétérogènes que celles du pôle associé. Mais qui peut le plus peut le moins. Cette hétérogénéité qui est plus ou moins la règle dans le domaine culturel (coopération entre musées, bibliothèques, centres d'archives, laboratoires...) explique le recours au Dublin Core qui constitue la seule contrainte inhérente au fait de s'associer au protocole, lorsqu'on possède un serveur.

Si le choix des partenaires se porte sur l'OAI, il faudra que les serveurs des centres de collecte soient paramétrés de manière à devenir fournisseurs de

données, la BnF servant à la fois de fournisseur de données et de fournisseur de services.

La qualification minimale du format d'échange des métadonnées (Dublin Core) n'empêche aucunement les différents centres de la FAMDT et la BnF de cataloguer comme par le passé dans leurs bases locales -les catalogues locaux resteraient inchangés. Mais chacun doit assurer en vue du moissonnage la conversion de ses notices de l'UNIMARC ou l'INTERMARC au schéma XML de base augmenté des balises Dublin Core. Il faut donc que les partenaires s'entendent sur la définition des balises (mise en correspondance avec les champs utilisés dans les catalogues de la FAMDT) et la manière de les utiliser. La société qui devrait implanter le protocole fournirait alors un accompagnement technique dans les premiers temps dont le coût devrait s'ajouter à celui de l'installation du protocole elle-même.

Pour ce qui est de la programmation des serveurs, elle représente par centre à peu près une demi-journée de travail et ne pose pas de problèmes majeurs à un programmeur expérimenté.

L'idéal serait de brancher un module entrepôt OAI sur le logiciel de catalogage Alexandrie, dans la mesure où ces catalogues gérés par Alexandrie peuvent être directement ouverts sur Internet.

Il serait souhaitable que les différents centres et la BnF se mettent d'accord sur les formats de présentation des notices (autre que Dublin Core) et sur les restrictions possibles au moissonnage, même si techniquement chaque centre conserve le contrôle et la gestion de ses notices.

2.2.2. Avantages

La capacité d'OAI à s'étendre n'est pas un atout négligeable

En effet, le peu d'exigences de ce protocole, qui s'oppose à l'homogénéisation des métadonnées que comporte fatalement le protocole Z39.50, garantit un moindre travail d'installation en même temps qu'un accroissement rapide du nombre de fournisseurs de données et donc une mutualisation de plus en plus grande d'archives sonores de par le monde. Un centre comme Dastum verrait peut-être un intérêt à consulter à terme les notices de tel centre de collecte d'archives sonores en Irlande ou au pays de Galles.

Par ailleurs, s'il est d'actualité un jour d'étendre le moissonnage aux données elles-mêmes²⁴, le format XML permettrait de faire rapidement le lien entre la notice et le document sonore, via un outil de consultation qui irait puiser à partir de la notice la ressource présente dans le serveur du centre correspondant. Cet accès pourrait être restreint à l'envi par la technique de segmentation des entrepôts en plusieurs lots (lots « copyright » et lots « domaine public » par exemple ; ce qui est d'emblée possible pour les métadonnées).

Pour ces deux raisons principales, OAI semble être une solution d'avenir.

2.2.3. Inconvénients

Le principal obstacle que les centres sont susceptibles de rencontrer ne tiennent finalement pas tant au protocole OAI qu'aux transformations préliminaires à son installation. En effet, la conversion des notices de l'UNIMARC en XML, de même que la pose des balises Dublin Core nécessite des opérations manuelles qu'il est difficile de mesurer mais qui ne manqueront pas d'être nécessaire quelque soit l'état initial des notices.

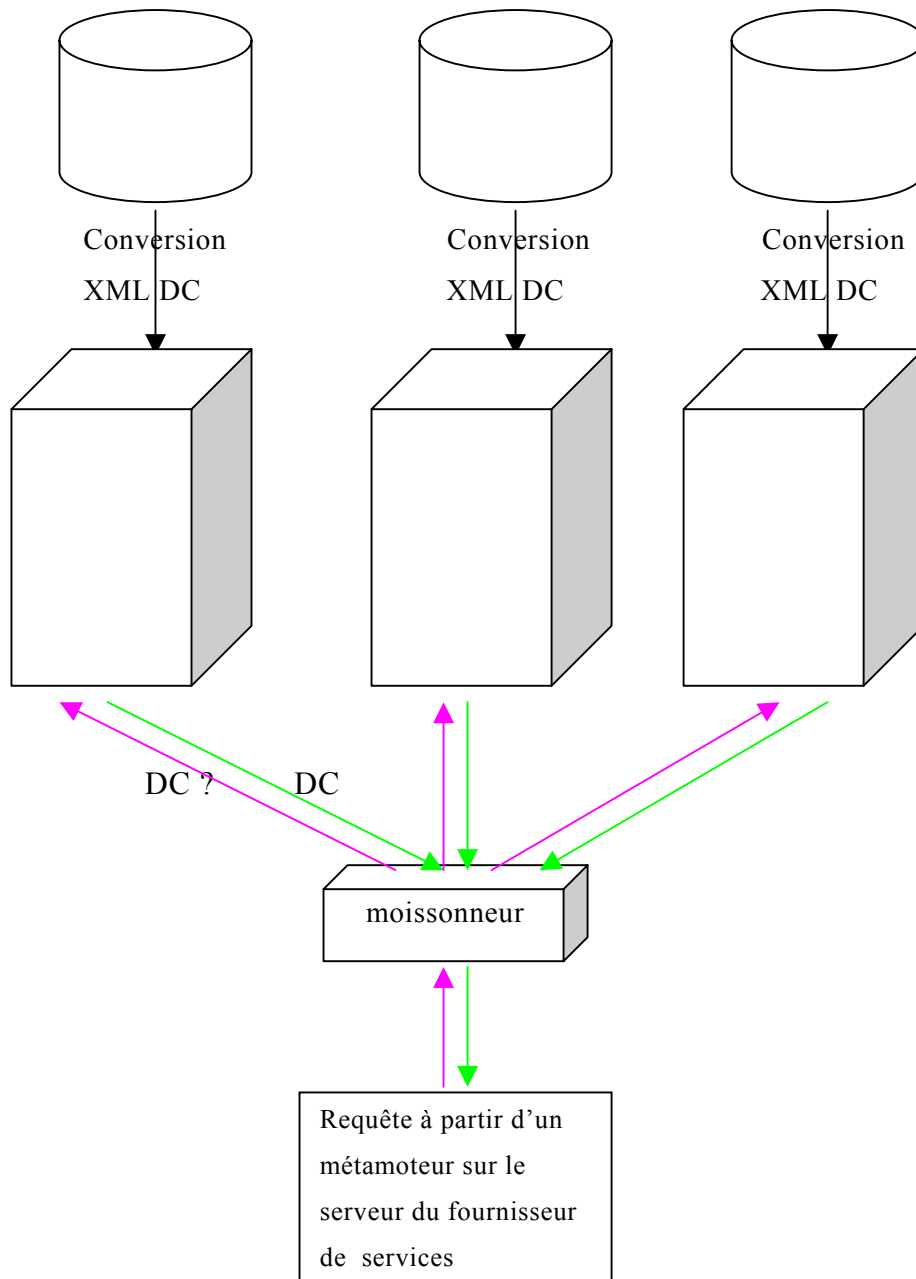
2.2.4. Investissements

Dans le cas où OAI-PMH serait la solution retenue, il faudrait doter chaque centre

- -nécessairement d'un module OAI par centre (toujours le même)
- -en option d'un outil de consultation (le même pour chaque centre mais personnalisable)
- -nécessairement d'un module moissonneur OAI pour le catalogue collectif
- -nécessairement d'un moteur de recherche et de consultation pour le catalogue collectif (adapter ce que la BnF peut déjà fournir à ces nouvelles tâches)

²⁴ Le site PubMed (www.ncbi.nlm.nih.gov) qui collecte des descriptions d'articles par le biais d'OAI permet aussi dans un grand nombre de cas (grâce à une convention passée avec les périodiques qui sont fournisseurs de données) d'accéder au texte intégral.

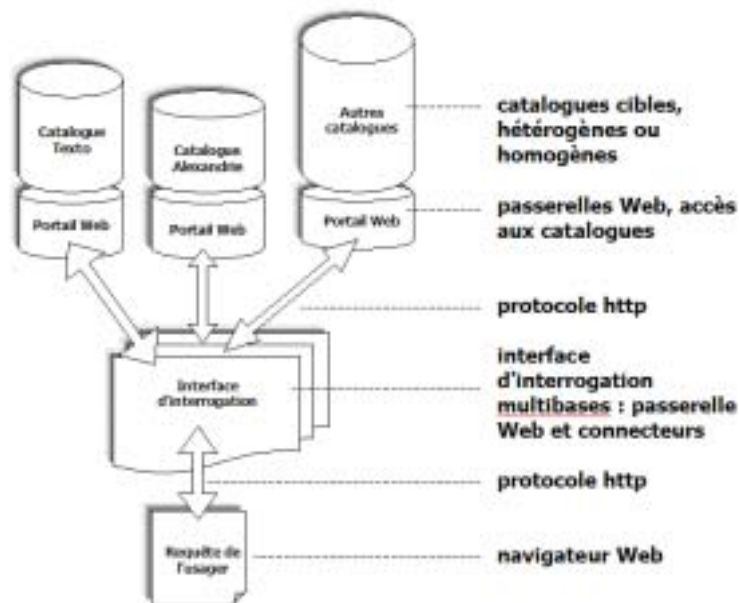
Le tout est évalué, selon une estimation de Martin Sévigny de la société AJSLM, à 30 000 euros (avec l'option outil de consultation). A quoi, il faudrait rajouter le coût de la formation.



2.3. Les connecteurs : autonomie, simplicité et fonctionnalité

2.3.1. Définition technique

Les catalogues sont fédérés au moyen d'une interface d'interrogation multibases couplée à une passerelle web et équipée de connecteurs²⁵. Les connecteurs traduisent à la volée la requête de l'utilisateur dans le langage de la cible et reformatent les réponses en langage HTML. Il y a autant de connecteurs qu'il y a de cibles hétérogènes²⁶. Les résultats de l'interrogation sont ensuite homogénéisés, s'affichent sur l'écran de l'utilisateur et permettent de consulter l'équivalent d'un « catalogue à la volée »²⁷.



2.3.2. Avantages

Le logiciel s'adapte aux bases existantes ; l'hétérogénéité des cibles n'est plus un obstacle. Par ce biais, des catalogues aussi différents les uns des autres que ceux du pôle associé pourront faire l'économie d'un travail de fond sur leurs

²⁵ Diverses sociétés développent de tels connecteurs, par exemple Xerox et G.B. Concept.

²⁶ L'interface d'interrogation peut viser simultanément un grand nombre de catalogues : catalogue sous Doris/Web, catalogue sous Alexandrie/Web, catalogue sous Loris/Web, etc. (*Documentaliste – Sciences de l'information*, 2004, vol.41, n°1, p.14-25)

²⁷ L'agglomération de Grenoble et la région Languedoc-Roussillon ont adopté le principe d'un catalogue à la volée pour la constitution d'un réseau de bibliothèques et de centres documentaires.

Voir : <http://www.bomlr.info/askonce.jsp?locale=FR>
et http://askonce.grenet.fr:2828/Metro_Interface.jsp (consulté 2004-04-29).

notices²⁸. De futurs partenaires rejoignant la fédération échapperaient aussi à un lourd travail d'homogénéisation.

L'intégration de nouveaux partenaires est simple. Elle se fait en interne si l'un des partenaires peut supporter, à titre d'investissement, la formation d'un informaticien à la technologie des connecteurs²⁹ ; elle se fait au moyen d'une intervention externe si les partenaires disposent de moyens plus modestes. Une telle souplesse peut convenir à l'hétérogénéité des acteurs du pôle associé.

2.3.3. Inconvénients

Une fédération de catalogues où chacun demeure attaché à ses pratiques habituelles risque de ne pas encourager la concertation. Le risque existe aussi que les partenaires soient à la merci du prestataire choisi faute d'avoir pu acquérir une compétence interne pour le développement de l'outil. En outre le coût de cette solution est élevé.

2.3.4. Investissements

Le coût que nous avançons comprend : la mise en ligne des deux catalogues qui ne le sont pas encore, le passage à un nouveau logiciel documentaire, les paramétrages (matériels et logiciels), la création d'une interface graphique, le développement d'un connecteur spécifique à la BnF ainsi que l'investissement matériel³⁰. Ce coût global s'élève à 48 000 euros³¹.

Les coûts humains comprennent des coûts de paramétrages initiaux (15 jours), des coûts de formation interne des personnels et du temps de travail au long cours (l'entretien du serveur et du réseau occupera entre 15 et 20% du temps de travail d'un informaticien).

²⁸ Cela ne dispense pas d'un effort initial d'établissement des correspondances entre les champs des notices et ceux du formulaire d'interrogation.

²⁹ Le coût optionnel d'une formation au développement de la technologie des connecteurs s'élève à 4000euros, correspondant à 5 jours de formation.

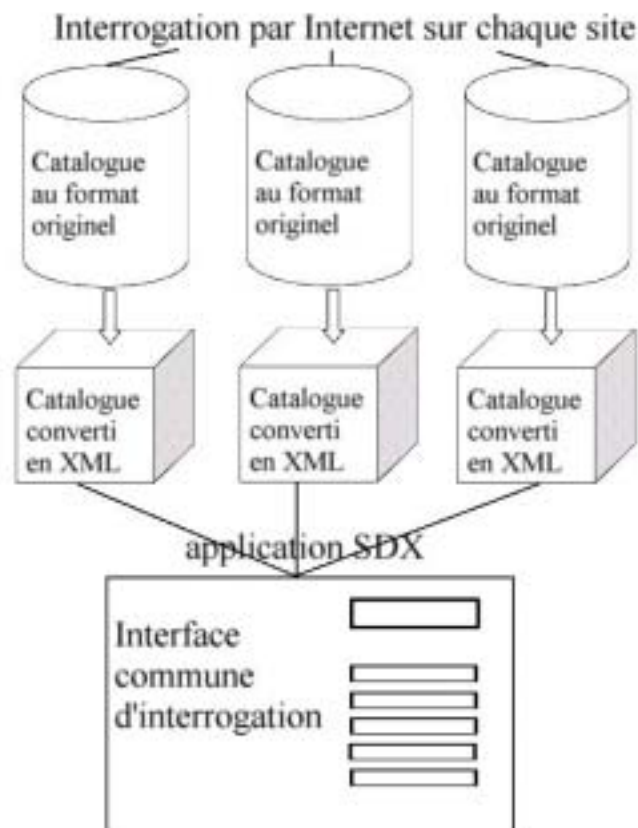
³⁰ Dont investissement matériel : 5000 euros.

³¹ Deux remarques au sujet du chiffrage des coûts, côté FAMDT puis côté BnF : 1. en adoptant un même logiciel documentaire (Alexandrie), les centres de la FAMDT réaliseront une économie d'échelle puisqu'il suffirait alors de ne développer qu'un seul connecteur pour toucher leurs différentes cibles ; 2. la BnF dispose d'un important service informatique. Les phases du projet peuvent alors relever soit d'une compétence interne, selon la disponibilité du service, soit de la compétence d'un prestataire. Pour évaluer les coûts, nous sommes partis du principe que la constitution de la base et sa mise en ligne se feraient en interne, mais que l'acquisition d'une

2.4. Le métamoteur XML : une collaboration active

2.4.1. Définition technique

La conception d'un métamoteur XML fonctionnerait de la façon suivante : les notices de documents sont converties en XML sous forme allégée (sans l'intégralité des renseignements), et chaque centre de documentation FAMDT disposerait de deux catalogues : l'un en format d'origine, sous Alexandrie, éventuellement interrogeable en ligne ; l'autre en XML.



Les modalités de la conversion en XML sont à définir ; elles dépendent en partie des capacités du module du logiciel Alexandrie proposant un export XML³². Il faut aussi déterminer la DTD choisie : BiblioML³³, ou EAD³⁴, ou encore un

application supportant la technologie des connecteurs et celle des connecteurs se feraient par le biais du prestataire pour la durée de la phase initiale.

³² Ce module étant optionnel, il est probable que certains centres ne le possèdent pas.

³³ <http://www.biblioml.org/>.

³⁴ <http://www.loc.gov/ead/>.

répertoire de balises choisi en interne, par exemple à partir du langage Dublin Core³⁵. Le métamoteur utilisé serait SDX, une plate-forme de logiciels libre et gratuite³⁶ élaborée par le Ministère de la Culture en collaboration avec la société AJLSM³⁷.

2.4.2. Avantages

L'intérêt majeur de cette solution réside dans l'image de modernité dont bénéficie aujourd'hui XML, image à laquelle ne sont pas insensibles les collectivités territoriales, non plus que le Ministère de la Culture. Par ailleurs, XML est d'une très grande souplesse d'emploi et permet de multiples utilisations, y compris pour d'autres projets. Enfin, son faible coût, du fait de l'emploi de logiciels gratuits, et son indépendance vis-à-vis de toute société informatique, lui garantissent une pérennité et une forte capacité d'évolution à peu de frais.

2.4.3. Inconvénients

En revanche, il est incontestable que l'avenir de SDX, de même que celui d'XML, sont encore complètement inconnus, ce qui peut représenter un gros risque — risque à relativiser puisque des notices en XML peuvent être aisément reconverties en tout autre langage structuré. Par ailleurs, l'adoption du métamoteur XML nécessite une très forte collaboration de tous les partenaires, une grande volonté de travailler en réseau cohérent et serré. Or, c'est un aspect dont l'opportunité est diversement appréciée. Enfin, SDX nécessite un réel effort d'adaptation de la part de tous les responsables des centres de documentation, qui doivent être formés au maniement du XML et de SDX (pour régler les problèmes mineurs qui ne manqueront pas de se poser). Ils doivent également être disponibles pour l'élaboration des notices XML : il faut s'entendre sur le nombre de champs à conserver et les simplifications inévitables à appliquer ; ensuite, vérifier l'exactitude des notices converties en XML ; enfin, alimenter continuellement la base avec les nouvelles notices.

2.4.4. Investissements

³⁵ Cf. le site officiel <http://dublincore.org/>, ainsi que des explications sur le site : <http://www.mutu-xml.org/xml-base/shared/KEY-DUBLINCORE.html>.

³⁶ <http://adnx.org/sdx/>.

³⁷ <http://www.ajlsm.com/>.

Le montant du coût ne portera pas sur les programmes eux-mêmes, mais sur leur paramétrage. Il peut donc varier énormément selon que l'on effectue ce paramétrage en interne, grâce aux informaticiens de la BnF par exemple, ou par une société prestataire de services. L'achat du matériel informatique, essentiellement des serveurs, constitue le seul coût fixe pour les centres de la FAMDT. En comptant très largement, l'ensemble des programmes d'export XML et de paramétrage d'une application SDX ne saurait dépasser 30 000 euros (n'est pas prise en compte le développement de l'interface d'interrogation).

Toutefois, il est évident que l'investissement qui prime dans ce scénario n'est pas financier, mais bel et bien humain.

3. Tableau de comparaison des

Critères	Degré d'innovation	Efficacité	Economie des moyens financiers	Economie des adaptations nécessaires (techniques et humaines)	Capacité d'évolution	Stabilité et sécurité du système
Solutions						
Connecteurs		+		+	—	+
Z39.50	—		—		—	+
Métamoteur XML	+	+	+			—
OAI	+			—	+	

— Faible

| Moyen

+ Fort

différentes solutions

Peut-être est-il utile de préciser ce que prennent en considération certains des critères indiqués dans le tableau ci-dessus.

Le degré d'innovation désigne le caractère récent d'une technique, ainsi que son adaptation à son environnement technique, celui d'aujourd'hui mais également, dans la mesure où cela est prévisible, celui de demain. Il reflète aussi l'opinion que les utilisateurs de ces différentes techniques peuvent avoir de celles-ci.

L'économie des adaptations nécessaires désigne l'effort à fournir (à l'exception de l'effort financier) pour créer le catalogue collectif, en considérant que l'économie est forte quand l'effort est faible, et que, dans le cas de l'investissement humain comme dans celui de l'investissement financier, c'est l'économie qu'il convient de privilégier.

La stabilité et sécurité du système rend compte de l'expérience déjà acquise ailleurs, par d'autres projets. Il est évident qu'une technique très neuve sera moins bien connue qu'une technique plus ancienne, et donc comportera plus de risques.

Compte-rendu décisionnel du 2^{ème} comité de pilotage

Partie 3 – Cahier des charges

Objet du présent document

Cette troisième partie développe les modalités du scénario retenu, à savoir un métamoteur interrogeant des catalogues éclatés au format XML, selon une application SDX. Le scénario tient compte de l'état actuel des catalogues existants, et de toutes les opérations à effectuer à partir d'eux.

Il rend également davantage dans le détail des coûts, opération par opération, en indiquant le mieux possible quels coûts incombent à chacun des partenaires.

Méthode suivie

Plusieurs arguments ont incité les acteurs du projet à porter leur choix sur une application SDX.

D'abord un aspect que l'on pourrait qualifier de politique, ou d'intellectuel, à savoir le choix du « free », c'est-à-dire de l'outil libre et gratuit. Ce type d'outils semble dans une voie d'expansion considérable, et les développements d'applications gratuites sont parfois beaucoup plus rapides. En outre, les versions ultérieures améliorées, lors d'une mise à jour du produit, coûteront moins cher en paramétrage que l'acquisition de la nouvelle version d'un produit payant.

Le fait que ce produit ait été développé par une société, AJLSM, à la demande du Ministère de la Culture, est un point extrêmement important pour comprendre les motifs d'un tel choix par la Bibliothèque nationale.

De leur côté, les associations se sont également révélées très favorables à l'adoption de cet outil, entre autres raisons dans la perspective d'exploitations parallèles à celle du seul catalogue collectif. SDX est adopté dans le monde des institutions de conservation patrimoniale, ce qui rendra progressivement, mais rapidement, de plus en plus aisé son développement pour ce projet, à mesure que les expériences extérieures se multiplieront.

Le choix de SDX n'est pas le choix du gratuit dans le sens où son exploitation serait gratuite : les coûts sont reportés non sur l'achat du produit, mais sur son développement. En revanche, d'une part le paramétrage lui-même est moins cher que l'achat d'un matériel commercialisé, et en outre l'entretien d'un tel outil se révélera sans nul doute moins coûteux, et plus facile à mettre à jour et à

développer. Son caractère évolutif, surtout dans le cadre d'un projet où il ne s'agit pas de tout mettre en place le plus vite possible, s'est révélé déterminant.

1 Les prérequis

1.1. Une phase préparatoire pour la FAMDT

1.1.1. La mise en ligne des catalogues

Deux des centres de la FAMDT, le Conservatoire occitan et la MMSH n'ont pas encore mis en ligne leurs catalogues. Or cette mise en ligne est prévue et utile pour la suite, parce qu'elle constituerait pour les partenaires une première étape dans la réalisation du projet, notamment l'acquisition d'un matériel (serveur, progiciels), et d'une expérience logistique utile (maintenance).

1.1.2. L'acquisition d'un progiciel adéquat

Le passage des quatre partenaires à une version de logiciel documentaire ou de SIGB récente est également un préalable important, car ce projet de catalogue collectif est fondé sur la conversion des notices en XML. Ce format de données ne s'étant généralisé que depuis peu de temps dans le monde des bibliothèques, seules les versions les plus récentes des progiciels offrent des applications liées à ce format, notamment des modules d'export permettant la conversion des données du format propre au SIGB en XML. Ce type d'application est important car l'une des conditions nécessaires à la bonne marche du projet sur le long terme est de permettre aux centres de garder une certaine autonomie.

La réalisation de ce travail préparatoire s'opérerait de la manière suivante :

- D'abord l'acquisition par le Conservatoire Occitan, UPCP-Metive et la MMSH d'un nouveau logiciel documentaire³⁸, et éventuellement pour Dastum d'une nouvelle version de son logiciel documentaire.
- En parallèle, la récupération des données du Conservatoire occitan, de UPCP-Metive pour l'intégrer au nouveau logiciel (la MMSH vient de passer sous Alexandrie).
- Enfin la mise en ligne des catalogues de ces trois centres³⁹.

³⁸ Ce changement semble être déjà prévu par les partenaires de la FAMDT qui projettent de passer sous le logiciel Alexandrie développé par la société GB Concept. Nous sommes partis de cette hypothèse pour le calcul des coûts qui sont présentés en annexe 2, ainsi que dans la partie 4 du présent document.

1.2. Le choix d'une DTD

1.2.1. Critères de choix

Il convient avant toute chose de récapituler quels sont les éléments que doit être capable d'intégrer la DTD, en fonction de ce qui est déjà contenu dans les notices existantes, et des nouvelles exigences liées à la nature du catalogue collectif virtuel.

Une notice XML doit contenir les catégories de renseignements suivantes :

- L'origine de la notice, et l'institution dépositaire ou propriétaire du document sonore. La BnF a rajouté l'exigence que puisse être indiquée l'origine de la description, qui peut être liée aux diverses strates de constitution d'une collection (pour la BnF : Archives de la Parole, par exemple).
- La description proprement dite du document, avec une structure qui peut être simplifiée par rapport à la notice originelle.
- Le lien éventuel vers le document numérisé, si celui-ci existe.
- Les indications quant aux droits d'accès, d'une part au document numérisé, d'autre part aux métadonnées.

1.2.2. Les formats envisageables

Il ne nous a pas été possible de trancher en faveur des DTDs existantes. Nous ne pouvons que nous contenter d'avancer les intérêts et inconvénients de chacune. Plusieurs possibilités s'offrent, qui remplissent toutes les exigences précisées ci-dessus. Mais les implications de chacune diffèrent fortement. Trois s'imposent comme réellement pertinentes, que nous avons classées de la plus simple à la plus élaborée – ou de la plus maniable à la plus complexe : Dublin Core, BiblioML et METS⁴⁰.

³⁹ Cette opération n'est pas nécessaire pour Dastum, dont le catalogue est déjà en ligne. Pour le centre de documentation de UCP-Metive, il s'agit de remettre le catalogue sur Internet avec le nouveau logiciel, puisqu'il l'était déjà avec l'ancien.

⁴⁰ On pourra noter qu'à proprement parler aucun de ces trois noms ne désigne précisément une DTD : Dublin Core est un format de métadonnées ; BiblioML est le programme de conversion d'UNIMARC en XML ; et METS est, comme Dublin Core, un format de métadonnées, mais beaucoup plus complexe et élaboré que ce dernier.

Il faut préciser au préalable que, quelle que soit la solution retenue, il est impossible de faire l'économie d'une table de concordance entre chacun des formats des centres d'une part, et UNIMARC d'autre part. En effet le passage d'UNIMARC à Dublin Core est établi de façon standard ; d'UNIMARC en BiblioML, c'est évident ; et il reste sans doute à faire la correspondance entre UNIMARC et METS. Cependant, il faut passer par un format intermédiaire, UNIMARC, donc, pour être sûr d'aboutir à des résultats identiques.

Le détail de chacune de ces solutions (définition, avantages et inconvénients) a été rejeté en annexe. L'essentiel du débat est le suivant :

Le choix porte sur deux critères à confronter : le sacrifice d'une partie de la rigueur intellectuelle d'un côté ; le sacrifice de plusieurs heures de travail pour l'élaboration des notices en XML d'autre part. Le choix du Dublin Core permettrait une mise en place plus aisée et plus rapide du catalogue collectif virtuel, et il serait plus facile de convaincre ultérieurement de nouveaux partenaires de s'intégrer, mais les notices y perdraient de leur richesse. Le choix de METS, à l'inverse, serait intellectuellement pleinement satisfaisant, mais à condition d'en venir à bout. Entre les deux se situe, BiblioML qui n'est pas forcément la meilleure solution sous prétexte qu'elle servirait de compromis.

1.3. La conversion des notices au format XML

1.3.1. Du côté de la FAMDT

1.3.1.1. Une phase de concertation

Le choix de la DTD utilisée reste à déterminer par le biais d'une consultation entre les partenaires des pôles associés. Comme on vient de le voir, le passage par le format UNIMARC s'avérera probablement nécessaire, et même si on souhaitait en faire l'économie, quelle que soit la DTD choisie, il faudrait établir une table de concordance entre les champs de la FAMDT et les différentes balises afin de rendre possible la conversion des données en XML par le biais d'un programme informatique. Cette opération de concertation préalable représente un travail non négligeable, mené conjointement par les Centres avec l'aide d'un spécialiste de ce format, ou d'UNIMARC le cas échéant.

Une seconde opération de concertation devra porter sur l'harmonisation des critères appliqués aux restrictions d'accès.

1.3.1.2. Les modalités et les étapes de la conversion

La conversion des notices au format XML peut s'opérer de diverses façons pour les centres de la FAMDT. Soit l'intégralité des notices de chaque centre est convertie au début du projet par une firme spécialisée, soit chaque centre acquiert son propre programme permettant de convertir ses notices en XML⁴¹. Cette dernière solution a l'avantage de permettre aux centres d'être autonomes et de gérer eux-même l'enrichissement du métacatalogue. Le risque existe, si tel n'était pas le cas, que la base de donnée au format XML soit figée car constituée uniquement des notices ayant été converties au moment de la mise en place du projet sans qu'aucun enrichissement ne se fasse par la suite.

Les centres ne conservent pas que des enregistrements sonores inédits, il faudra donc dans un deuxième temps isoler les notices concernées par le catalogue collectif⁴².

Une fois ce travail de repérage mené à bien, chaque Centre aurait à charge d'acquérir une « moulinette » assurant la conversion des notices en XML conformément au format XML choisi, ce qui permettrait soit de les envoyer vers le réservoir de données, soit de convertir les notices en XML à la volée.

L'effort pour former un informaticien sur place à XML est également un préalable indispensable à la bonne marche du projet.

1.3.2. Du côté de la BnF

La situation de la BnF est différente puisque le système informatique qui supporte le catalogue Bn-Opale plus ne propose pas d'export XML. Pour que la plate-forme SDX puisse avoir accès aux notices, il semble nécessaire de procéder à leur extraction, de les structurer en XML, et d'en constituer une base de donnée accessible sur un serveur.

1.3.2.1. Enrichir les notices pour permettre leur sélection

⁴¹ A supposer évidemment que les centres puissent devenir réellement les propriétaires du programme de conversion des notices en XML, ce qui est du domaine de la négociation.

Les documents des fonds de la BnF à signaler dans le métacatalogue représentent actuellement un lot évalué de 500 à 1000 notices. Il n'est pas encore possible de les isoler par une opération informatique du reste du catalogue Bn-Opale plus. Au préalable, il faut y intégrer les indications qui, une fois croisées, le permettent : des données codées renseignant sur le pays d'enregistrement et sur la nature du document (archive, collectage). Ce travail est facilité par le fait qu'il concerne des fonds définis, et qu'il peut être mené pour une part en coordination avec le programme de numérisation des supports d'origine des Archives de la Parole.

1.3.2.2. Renseignements des droits d'accès aux métadonnées

Ces données ne figurent pas actuellement dans les notices de la BnF. Elles peuvent être ajoutées *a posteriori* dans les notices XMLisées, par exemple au moment d'une relecture de celles-ci pour contrôle.

Il serait toutefois moins lourd de les intégrer aux notices préalablement à la conversion si cela est possible, par exemple au même moment que l'enrichissement évoqué ci-dessus. Cela n'est utile en effet qu'au moment du passage des notices en UNIMARC. Or la gestion des droits apparaîtrait dans les données locales, et il faut vérifier que ce type de données apparaisse dans la notice convertie en UNIMARC. On pourrait convenir d'utiliser un champ déjà existant (indications sur le public ou la confidentialité, ou bien une position inutilisée dans les zones fixes), à moins de demander la création d'une zone supplémentaire.

1.3.2.3. Montée en charge progressive

L'enrichissement des notices ne représente pas un travail considérable. Cependant, il est peu probable que le personnel qui en sera chargé s'y consacrera à plein temps. Ces opérations devraient se répartir sur plusieurs mois en fonction de la disponibilité des uns et des autres. On peut toutefois supposer qu'au bout de huit mois à un an, on disposera d'un échantillon suffisamment conséquent. Si le catalogue collectif est appelé à être opérationnel dans un délai assez court, il faut

⁴² Ce repérage des notices concernées ne devrait pas poser de problème, puisque les notices disposent d'un champ « type » dans lequel est mentionné « son inédit ».

donc prévoir que toutes les notices de la BnF n'y figureront pas d'emblée, mais qu'elles viendront enrichir la base de données par incrémentations successives.

1.3.2.4. Programmes à prévoir pour la conversion au format XML

Un certain nombre d'opérations ponctuelles doivent être programmées, qui paraissent inutilement coûteuses, voire compliquées à externaliser. On peut avantageusement les confier à du personnel de la BnF, éventuellement au sein du Département de l'Audiovisuel.

- Il faut d'abord procéder à l'extraction des notices à convertir, sur bande magnétique ou sur support numérique. On peut prévoir de renouveler l'opération si la conversion est divisée en lots successifs.
- La conversion se fait d'INTERMARC vers UNIMARC puis vers BiblioML. Il faut convertir préalablement l'ensemble des notices en UNIMARC. La BnF possède déjà le programme qui le permet, qu'elle utilise en interne pour la Bibliographie nationale.
- L'échantillon de notices en UNIMARC est ensuite converti en données XML structurées selon la DTD retenue. A noter que la BnF dispose du logiciel de conversion vers BiblioML⁴³ mis au point par Jouve. Elle n'aurait donc pas à acquitter des droits pour l'exploitation de cet outil, et les éventuelles mises au point seraient limitées, compte-tenu qu'il est déjà développé pour permettre la conversion des données des enregistrements sonores édités de la Bibliographie nationale. Rien n'interdit, même si BiblioML n'est pas employé pour les données de la FAMDT, de le retenir par commodité pour les notices de la BnF, puisque la plate-forme SDX peut être programmée pour interroger n'importe quelle DTD.

Au sein de la BnF, ni le Catalogue collectif de France, ni le Département de la bibliothèque numérique n'ont la possibilité d'héberger la base des notices XMLisées.

1.4. La question de l'hébergement des données

1.4.1. Stockage des données ou conversion à la volée ?

1.4.1.1. Le stockage des données au format XML

Les données peuvent être stockées en XML sur plusieurs serveurs de données (chaque centre hébergeant ses données), ou sur un seul serveur de données commun qui serait hébergé par un des partenaires.

Cette solution pourrait être mise en place par transfert régulier des notices cataloguées de manière « traditionnelle » vers la base de données XML par le biais d'un « script » macro transférant régulièrement les données de la base de données « classique » vers le serveur de données dédié à XML. Ce transfert régulier des notices permettrait de réduire le risque que les deux catalogues parallèles finissent par s'éloigner l'un de l'autre.

Le stockage « éclaté » des notices au format XML, est un peu plus coûteux, puisque chaque partenaire est dans l'obligation de posséder et de rendre accessible son propre serveur de données internet. Cependant si l'on part du principe que le projet de métacatalogue n'est que le début d'une plus vaste opération qui verra la mise en ligne d'une partie des enregistrements sonores numérisés, qui devront obligatoirement être stockés « en dur », cette dépense supplémentaire se justifie.

1.4.1.2. La conversion des données au format XML « à la volée »

La conversion des notices en XML peut aussi se faire « à la volée », les centres se servant alors de leur module d'export XML pour convertir les notices au moment où l'application de recherche commune viendra « piocher » dans le catalogue en ligne. Les notices seraient alors stockées sous le format propre au SIGB et converties en XML au moment de la requête, de manière éphémère, c'est-à-dire sans être stockées. Néanmoins cette opération, qui a l'avantage d'éviter aux centres de la FAMDT un stockage des notices, n'est pas possible dans le cas de la BnF qui devra de toute façon extraire ses notices de Bn-Opale plus. Cet

⁴³ Dans le cas du choix d'un autre format pour les données XML, la problématique demeure la même, pourvu que l'on dispose du tableau de concordance vers UNIMARC.

inconvenient n'est pas un problème insurmontable, et il n'en reste pas moins que la conversion à la volée demeure probablement la solution la plus adéquate et la plus opérationnelle pour les centres : une fois le travail préliminaire de concordance et de réglage établi, ceux-ci n'ont pas à se préoccuper de stocker leurs notices en XML. Cependant pour que cela fonctionne de manière satisfaisante, le paramétrage de cette conversion doit être très précis, ce qui nécessite un travail de réglage préliminaire particulièrement lourd.

1.4.2. Qui héberge les données ?

Dans tous les cas, l'application de recherche doit obligatoirement être hébergée pour avoir une adresse URL. Si les partenaires optent pour un stockage centralisé des notices, il faudra aussi héberger ce serveur de données.

Le Département de l'Audiovisuel de la BnF doit prendre en charge et abriter à la fois ses notices XMLisées, et l'application jusqu'à ce que son hébergement définitif soit fixé⁴⁴. L'hébergement par le CCFr d'un serveur de données centralisant les notices XML de tous les partenaires ne semble pas envisageable, ni même l'hébergement des seules notices de la BnF en XML extraites de Bn-Opale plus.

Ce problème reste donc en suspens, le CCFr n'étant pas en mesure d'héberger l'application et la base de données avant 2007.

L'idéal serait que l'un des partenaires – et c'est une des vocations de la BnF –, disposant de moyens logistiques suffisants pour pouvoir intervenir directement en cas de problèmes, se charge de l'hébergement et de la maintenance du réservoir de données centrales⁴⁵. Comme le stockage de données au format XML ne demande pas un trop gros investissement matériel, le critère discriminant pour choisir le site serait la présence d'un informaticien pouvant être formé à XML.

⁴⁴ Il est envisageable que ce type d'application soit à terme hébergé par le CCFr.

⁴⁵ Ou seulement du réservoir de notices de la BnF, si la majorité des centres opte pour une conversion « à la volée » des notices au format XML.

2. Mise en place de SDX

2.1. Paramétrage de l'application SDX

Comme l'a clairement fait savoir la BnF lors du second comité de pilotage, le paramétrage d'une application SDX pour l'exploitation des notices en XML sera confiée à un prestataire de services.

Plusieurs éléments doivent être nécessaires.

2.1.1. L'exploitation public/professionnel

Le moteur SDX doit disposer des adresses IP des postes détenus par les différents centres de conservation partenaires du catalogue virtuel. Pour les professionnels, il faut qu'ils puissent disposer de l'intégralité des renseignements fournis dans les notices, auxquelles, pour certaines d'entre elles, le public peut n'avoir qu'un accès limité. Les modalités quant à la consultation dans chaque centre des documents numérisés accessibles sous certaines conditions juridiques, restent à préciser par les responsables des fonds.

2.1.2. Les modes d'interrogation

L'application SDX doit servir de moteur de recherche dans les différentes bases existantes, qu'elle interroge en direct, selon les différents critères retenus. Elle se charge également de gérer l'indexation des notices.

2.1.3. La gestion des bases de données en XML : de SDX à Cocoon

Dans le cas d'un stockage des données XML, il n'est pas prudent de tout faire reposer sur les capacités de SDX. La gestion des bases de données elles-mêmes, sur place, ne peut se faire que temporairement avec cet outil, et il est plus judicieux de prévoir, à échéance assez brève, l'emploi de l'outil *Cocoon*⁴⁶ pour cela. Celui-ci n'est cependant indispensable que si les partenaires décident de

⁴⁶ Cf. <http://cocoon.apache.org/1.x/>, et en français : <http://www.epita.net/~cocoon/cocoon2.html>.

développer un site internet *dynamique*. Pour la seule recherche documentaire, SDX suffit.

2.1.4. La mise en place du portail

L'application SDX gère la page d'interrogation, ainsi que l'affichage des réponses aux requêtes, selon une feuille de style définie en accord avec l'ensemble du portail BnF⁴⁷.

Au sujet de la composition du portail rappelons les fonctionnalités qui ont été recensées dans le document de comité de pilotage précédent :

- -deux modes de consultation des catalogues : mode public et mode expert.
- -possibilité de limiter en amont la recherche à un ou plusieurs catalogues.
- -possibilité d'interroger en plein texte.
- -possibilité d'utiliser des opérateurs booléens et des troncatures.
- -interrogation des champs suivants : titre - mention de responsabilité – mot clef – lieu – date – langue. Il serait sans doute utile d'y adjoindre une interrogation par fonds de collecte.

A quoi l'on peut ajouter :

- -accès à un historique des recherches et un historique des résultats.
- -possibilité de limiter la recherche aux notices qui ont un lien vers un document numérisé.

Les résultats de recherche pourraient être représentés par groupes d'une dizaine par exemple pour les notices courtes avec un lien vers les notices plus détaillées. Le fonds d'origine du document est une information d'importance à mettre en valeur dans la liste des résultats.

⁴⁷ Dans l'hypothèse où c'est le choix d'un hébergement du portail d'interrogation sous l'adresse de la BnF qui est retenu.

3. Maintenance et perspectives

Dans la problématique concernant l'entretien, il convient de souligner ici que, à côté des avantages de l'emploi d'un logiciel libre, pèse l'inconvénient suivant : en cas de problème au moment de son utilisation, si le système bloque pour quelque raison, il faut impérativement un informaticien sur place disposant d'un minimum de connaissances, car il n'existe aucun interlocuteur privilégié pour la maintenance qui maîtrise parfaitement un produit qui n'est pas le sien.

3.1. L'alimentation des bases existantes

Les fonds d'enregistrements sonores inédits de la BnF, tout comme ceux des centres de la FAMDT, continuent de recevoir un certain nombre de dons et de dépôts de collectes. Dans un cas comme dans l'autre, une réflexion préalable sur les modalités d'enrichissement du réservoir de notices en XML s'impose. La BnF dispose des moyens logistiques nécessaires à l'enrichissement du catalogue en XML (programme de conversion, informaticiens...).

Dans le cas des centres de la FAMDT, qui disposent de moyens logistiques moins conséquents, tout en enrichissant encore leur catalogue à un rythme plus accéléré, il est important de prévoir une solution simple pour l'alimentation du réservoir de données centralisé, si cette solution est choisie. Doter les centres de la FAMDT d'un module d'export au format XML et peut-être d'un système assurant automatiquement l'alimentation de la base est un investissement important, mais qui sera profitable sur le long terme.

3.2. La numérisation des documents et leur accessibilité

3.2.1. Communication à distance des métadonnées

Pour des raisons pratiques, certaines données n'apparaissent pas toujours dans les notices actuelles. Ainsi, on soustrait les éléments suivants à la connaissance du public : noms des droits et des ayant-droits, notes confidentielles, informations portant sur la gestion des exemplaires, noms des personnes qui ont fait l'analyse et la copie. Lorsque pour des raisons de confidentialité, le nom de l'informateur est

supprimé de l’affichage, il est d’ores et déjà remplacé par un numéro. Toutefois il est important que le nom de l’informateur soit non seulement visible mais aussi interrogeable le plus souvent possible, car c’est la seule information de base que possèdent un grand nombre d’utilisateurs. Il en va de même pour les lieux d’enregistrement.

Comme il a déjà été dit plus haut, les pratiques des centres étant différentes en matière de publication des métadonnées, une harmonisation préalable devrait avoir lieu. A défaut, les pratiques les plus restrictives devraient être choisies au détriment des autres.

3.2.2. Communication à distance des données numérisées

En raison de la législation sur les droits d’auteur, la numérisation des documents sonores dans les différents centres est jusqu’à présent davantage dictée par des impératifs de conservation que par l’ouverture des archives au grand public.

La numérisation des nouvelles bandes est systématique à la MMSH depuis 2000 ; 1500 heures ont été digitalisées à ce jour. Dastum a déjà procédé à plus de 400 heures de numérisation⁴⁸ sur les 4500 heures d’enregistrement conservées (chiffre indiqué sur le site du ministère de la culture au 4 juin 2004). UPCP-Metive a réalisé en copies numérisées l’équivalent de 800 heures d’enregistrement. Le Conservatoire Occitan a quant à lui numérisé environ 35 % du son inédit qu’il conserve.

Le temps que dure la numérisation d’un enregistrement d’une heure varie entre 4 et 6 heures selon le contenu. L’encodage proprement dit n’excédant pas la durée de l’enregistrement, le gros du travail comprend le repérage des séquences (plus délicat à faire dans un récit de vie que dans un recueil de chants ou de contes), la création de la fiche d’enregistrement ou la réalisation du CD comprenant marqueur et son *wave*, l’ajout de métadonnées relatives à chaque séquence, le traitement du son le cas échéant, la duplication des copies.

Ces enregistrements, selon la norme définie dans le *Guide d’analyse documentaire du son inédit* (p. 45), se sont vu attribuer dès leur catalogage, avant

d'être numérisés, un code de consultation entre 0 (non consultable) et 3 (libre à la consultation et à la copie).

Il nous semble que la mise en ligne d'un catalogue collectif des archives sonores en France engendrerait beaucoup de frustration s'il ne permettait d'accéder à aucun des enregistrements dont l'utilisateur aurait pu consulter les notices. Il en va de la notoriété d'une telle entreprise et de la promotion des archives sonores dont on peut dire au mieux qu'elles peinent en France à trouver leur public. Cet aspect est un des enjeux majeurs du projet, qui doit orienter les solutions techniques et juridiques adoptées.

Donner accès à certains documents sonores à partir d'un lien créé dans leur notice nous apparaît infiniment préférable au fait de constituer une banque d'enregistrements indépendante.

Les obstacles à la diffusion de ces documents sur la toile sont avant tout de nature juridique, cela va sans dire (voir *Guide*, p. 46)

Leur mise en ligne nécessite un ajustement des codes, chaque code indiquant un type de consultation par rapport à leur diffusion sur Internet.

Il faut prévoir ainsi plusieurs types de consultation :

- par Internet. (Il serait souhaitable, dans les cas où cela se pratique, de passer d'un accès sur une préfiguration sonore de x sons à un accès complet aux séquences.)

- par intranet.

- par extranet. (Lecteurs autorisés en dehors des centres et munis de codes d'accès.)

- sur site.

La problématique commerciale n'entrant pas en ligne de compte (il ne s'agit que de confidentialité et de droit moral des acteurs sur leur création), la question du téléchargement n'est pas vraiment pertinente.

⁴⁸ Il semblerait que ce chiffre soit à réactualiser, puisque les responsables de Dastum parlent quant à eux de plus de 2500 heures d'enregistrement numérisées.

SDX a les capacités de gérer ces accès, tout comme les accès aux métadonnées, pour peu que ces codes soient traduits d'une manière ou d'une autre dans la DTD choisie.

Il est difficile d'évaluer la part des fonds de chaque centre qui demain pourrait figurer en libre accès sur le net. Il s'agirait principalement (il s'agit déjà à Metive...) de « formes esthétisantes telles que chansons, musiques, contes, légendes, formes brèves⁴⁹ ». Ce corpus pourrait constituer 30% des fonds de la MMSH selon l'estimation de Mme Ginouvès.

Les récits de vie d'une part et les témoignages thématiques d'autre part n'ont pas vocation à paraître en libre accès sur la toile. C'est du moins la politique qui a été définie par UPCP-Metive parce que les besoins en danse et en musique traditionnelles sont plus importants pour le grand public et les professionnels de la musique. Certains centres ne sont pas opposés à mettre en ligne en intranet cette deuxième catégorie de documents à condition d'en excepter certains dont le contenu aborde des sujets sensibles, comme la Résistance et la sorcellerie.

3.3. L'intégration de nouvelles bases

3.3.1. Tous les types de documents

Deux grandes exigences pour l'avenir gouvernent ce projet : son extension à d'autres centres patrimoniaux, et sa capacité à intégrer des documents numérisés. Sur ce dernier point, l'aptitude de SDX à gérer des documents sonores est attestée⁵⁰. S'il laissait de côté les questions liées à l'exploitation des documents d'archives sonores – questions essentielles qu'il faut très sérieusement prendre en charge –, le site expérimental dont nous avons observé le fonctionnement témoignait en faveur des larges possibilités techniques de SDX. Quant à l'ouverture du métacatalogue à d'autres musées ou centres de documentation, elle ne soulève aucune difficulté, comme propose de le constater ADNIX en dirigeant le

⁴⁹ Selon Madame Coulais, de UPCP-Metive.

⁵⁰ Michel Fingerhut, directeur de la médiathèque de l'IRCAM, a développé avec succès à titre expérimental un site intégrant de multiples sources documentaires et reposant sur des choix technologiques dont notre projet partage les options. www.planson.ircam.fr, non accessible.

visiteur de ses pages vers une liste de sites exploitant aujourd'hui la technologie SDX ou en voie de l'adopter⁵¹.

3.3.2. Indépendance

En-deçà des interventions des fournisseurs de SIGB et des développeurs de SDX, l'architecture du projet repose sur la mise en œuvre de logiciels libres. De tels logiciels ont la propriété de ne pas être liés à la pérennité d'une société, d'être lisibles par tous les informaticiens, de pouvoir être développés en vue de l'interrogation de toute source documentaire nouvelle. Sous ces différents aspects, l'emploi de logiciels libres paramétrés en interne est préférable à toute autre solution. Dans la perspective du choix de SDX, la consultation des documents sonores eux-mêmes peut être envisagée comme effective à moyen terme. Seuls l'équipement des chercheurs et les restrictions juridiques ou institutionnelles décidées en amont pourraient contrarier l'accès et la consultation de ces nouveaux documents.

3.3.3. D'un XML à l'autre

Le fonctionnement du catalogue reposera sur le fait que les documents possèdent une structure XML. On ne peut se cacher le fait que, à ce jour, il n'existe encore aucune DTD spécifique exclusivement appropriée aux enregistrements sonores inédits. La réflexion sur ce que devrait être cette structure de référence n'a pas encore abouti. Quoi qu'il en soit des différences potentielles entre la structuration choisie et celle qui fera l'objet d'un consensus à l'avenir, on sait néanmoins qu'il sera possible de convertir une structuration XML dans une autre structuration XML. Le développement des catalogues collectifs de sources documentaires passe par la voie de l'XML.

⁵¹ <http://vernier.gamsau.archi.fr:8080/JSPWiki/Wiki.jsp?page=SitesEnExploitation>

4. Table des coûts et calendrier

Les prix sont tous repris de différentes sociétés qui ont accepté de nous établir des devis sans engagement, avec une marge d'incertitude compte tenu des fluctuations, des alternatives et des inconnus.

Les chiffres qui précèdent les tâches reprennent la numérotation indiquée dans la feuille des lots et tâches, placée en annexe. La tâche 3.1. correspond à la tâche 1 du lot 3. Il faut également compléter ce calendrier avec les précisions fournies en annexe concernant la répartition des catalogues, avec le détail des coûts respectifs des solutions proposées⁵².

Il faut dans les pages qui suivent, placer trois calendriers à la suite l'un de l'autre, selon la solution retenue :

(1)	(2)	(3)
	↗ Base XML commune	↘
Mise en ligne des catalogues	→ Bases XML éclatées	→ installation du portail
	↘ XML généré à la volée	↗

On trouvera comme dernier tableau un calendrier synthétique reprenant les phases principales du projet avec leur inscription dans le temps.

⁵² Tous les prix sont données par centre (hormis la mise en place du métamoteur SDX, qui n'est à faire qu'une fois), et hors taxe.

4.1. Préalables : mise en ligne des catalogues qui ne le sont pas encore

A cette mise en ligne a été ajouté l'enrichissement des notices BnF, qui est indépendant du choix de la répartition des catalogues

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
1.1. Choix préalables (DTD, répartition des coûts, répartition des catalogues) 1 jour interne								
Si BiblioML est retenue Préalable : Enrichissement du vocabulaire de BiblioML pour les documents musicaux six mois 0								
1.2. Elaboration du tableau de concordance avec UNIMARC 2 semaines en interne + réunion avec un spécialiste d'UNIMARC interne								
1.3. Rédaction d'un cahier des charges 2 semaines Interne								

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
(Préalable) Partenaires mettent leurs catalogues en ligne Acquisition d'un logiciel de gestion documentaire permettant cette fonctionnalité 1 mois 2300								
Récupération des données et réglage du module d'export UNIMARC 2 jours de travail espacés de 10 jours pour les tests 3200								
Mise sur Internet 1 semaine 2000								
Contrat de maintenance 500								
Personnalisation de l'interface web 800 2 semaines								

TACHES			mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
(Dastum)	Frais matériels et logiciels	5000								
	Mise à jour de la version logicielle de Dastum (frais de logiciel et formation)	1200								
TOTAL par centre de la mise en ligne	hypothèse basse	2800								
	hypothèse haute	13800								
1.4. Enrichissement des notices BnF										
sept mois										
interne										

4.2. La question de la répartition des catalogues

4.2.1. Base XML commune

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
1.6. Mise en place d'un module d'export XML Paramétrage d'un module d'export XML pour chacun des centres 3 jours 1600								
Acquisition d'un "script" de transfert 2 jours 2600								
Formation d'un informaticien sur place 1 jour 1000								
Total par centre pour le module d'export 5200								

		hypothèse basse												hypothèse haute												
Total par centre (préalables + module)	hypothèse basse	8000																								
	hypothèse haute	19000																								

4.2.2. Bases XML éclatées

TACHES		mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
1.6. Mise en place d'un module d'export XML	Paramétrage d'un module d'export XML pour chacun des centres								
	3 jours								
	1600								
	Acquisition d'un "script" de transfert								
	2 jours								
	2600								
Formation d'un informaticien sur place									
	1 jour								
	1000								

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
1.6. Acquisition Matériel d'un serveur de données supplémentaire 1500 destiné à stocker les informations Installation au format XML 1/2 400 journée								
Coût total pour l'installation du catalogue XML 7100 Total par centre (préalables + module)								
hypothèse basse 9900 hypothèse haute 20900								

4.2.3. XMLisation à la volée (sans stockage des données XML)

TACHES		mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
1.6. Mise en place d'un module d'export XML	Paramétrage d'un module d'export XML pour chacun des centres								
	3 jours 2400								
	Formation d'un informaticien sur place								
	1 jour 1000								
Coût total pour la conversion XML									
	3400								
Total par centre (préalables + module)	hypothèse basse								
	6200								
	hypothèse haute								
	17200								

4.3. Mise en place du portail d'interrogation

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
3. Paramétrage de SDX 3.1. Rédaction d'un cahier des charges, traitant notamment de la question des droits 2 semaines interne								
3.2. Passation d'un marché avec une société prestataire de services 1 mois 6000								
3.3. Paramétrage de SDX et réalisation du portail d'intégration 18000 2 mois								
4. Enrichissement progressif BnF 4 h/mois FAMDT 2 h/mois								

[illegible]

⁵³ Le chiffre du coût total considère la mise en ligne de 2 catalogues, l'XMLisation de 5 catalogues, et le paramétrage d'une application SDX. Les coûts internes (le nombre d'heures passées par le personnel des différents partenaires), qui ont été indiqués comme durée de travail à chaque étape, n'ont pas été additionnés et ne sont pas pris en compte en termes budgétaires dans le total final.

4.4. Calendrier synthétique

TACHES	mois 1	mois 2	mois 3	mois 4	mois 5	mois 6	mois 7	mois 8
Choix préalables (DTD, répartition des coûts, répartition des catalogues)								
Elaboration du tableau de concordance avec UNIMARC								
Rédaction d'un cahier des charges								
(Préalable) Partenaires mettent leurs catalogues en ligne								
Enrichissement des notices BnF								
Mise en place d'un module d'export XML et accessibilité des notices XML en ligne								
Paramétrage de SDX								

Compte-rendu décisionnel du 3^{ème} comité de pilotage

Calendrier de la prestation

Semaines	Janvier				Février			
	1	2	3	4	1	2	3	4
Lancement du projet								
Choix du projet et constitution de l'équipe								
Rencontre avec le commanditaire à la Bibliothèque nationale de France								
Répartition des rôles								
Préparation du stage 1; étude lettre de mission; Plan d'action du stage								
Prise de contact avec les partenaires du projet								
Stage 1								
Familiarisation avec les questions techniques								

Semaines	Janvier				Février			
	1	2	3	4	1	2	3	4
Définition des contraintes liées au projet								
Elaboration du cahier des charges du projet								
Mise au point des questionnaires et préparation des entretiens avec les partenaires du projet								
Interrogation des catalogues des différents partenaires, et mise au point d'une bibliographie sur le sujet (aspects techniques et renseignements sur les associations)								
Rencontre avec des spécialistes des questions techniques								
Rencontre avec des prestataires de services								

Semaines	Mars				Avril			
	1	2	3	4	1	2	3	4
Interrogation des catalogues des différents partenaires, et mise au point d'une bibliographie sur le sujet (aspects techniques et renseignements sur les associations)								

Semaines	Mars				Avril			
	1	2	3	4	1	2	3	4
Réunion du premier comité de pilotage; Définition de la prestation et validation								
Préparation du plan d'action du stage 2								
Rencontre avec les partenaires du projet								
Rencontre avec des spécialistes des questions techniques								
Poser les problèmes juridiques								
Rencontre avec des prestataires de services								
Stage 2								
Elaboration des scénarios								

Semaines	Mai				Juin			
	1	2	3	4	1	2	3	4
Elaboration des scénarios								
Réunion du deuxième comité de pilotage; choix d'un scénario								
Approfondissement du scénario								
Réunion du troisième comité de pilotage; validation d'un scénario								

Budget de la prestation

1. Budget prévisionnel de la prestation

	Montant en Euros	Prise en charges
Personnel : 6 personnes, 160 h. à 9,36 €	8985	ENSSIB
Frais de déplacement : 27 allers-retours Paris-Lyon	2700	ENSSIB
Frais de déplacement : 4 allers-retours Lyon-centres FAMDT	400	
Frais de séjour : 2 semaines à Paris, 6 personnes	2380	ENSSIB
Téléphone	30	ENSSIB
Connexion Internet	80	ENSSIB
Photocopies, impressions	50	ENSSIB
Total	14625	ENSSIB

2. Budget réalisé de la prestation

	OBJET	DÉTAIL	MONTANT EN EUROS	PRISE EN CHARGE
FRAIS SALARIAUX		6 personnes Taux horaire à 9,36€ 6 x 4h/semaine x 15 sem. 6 x 35h x 2 sem. soit 780h	7300,80	ENSSIB
FRAIS DE STAGES ET COMITÉS DE PILOTAGE	Visite SNCF sur site RATP Indemnités	6 x AR Paris-Lyon (100,00€) 2 x 6 tickets, 1,40€ 6 x taux de base à 8,82€	600,00 16,80 52,92	ENSSIB ENSSIB ENSSIB
(les indemnités journalières s'élèvent à 3 taux de base par jour et par personne, le taux de base étant fixé à 8,82€)	Stages SNCF (2 stages) RATP Indemnités	6 x 2 x AR Paris-Lyon 6 x 2 x tarif hebdo à 15,40€ 6 x 2 x 5 x 3 Taux de base	1200,00 184,80 1587,60	ENSSIB ENSSIB ENSSIB
	Comité de pilotage (3 CP) Indemnités	□ 3 x 3 AR Paris-Lyon	900,00	ENSSIB
		3 x 3 x 2 tickets à 1,40€	25,20	ENSSIB
		3 x 3 taux de base	79,38	ENSSIB
FRAIS DE FONCTION- NEMENT		Téléphone Internet Photocopies et impressions	30,00 80,00 50,00	ENSSIB ENSSIB ENSSIB
FRAIS EXCEP- TIONNELS	Nuitée Repas	2 nuits 2 repas	106,72 30,50	ENSSIB ENSSIB
TOTAL			12244,72	ENSSIB

La somme totale indiquée dans ce budget réalisé diffère notablement de celle chiffrée dans le budget prévisionnel. Cela s'explique par les deux raisons suivantes.

D'une part, le groupe avait envisagé d'une façon large les heures de travail nécessaires à la conduite de ce projet. Dans ce second budget, le groupe est revenu vers un mode de calcul orthodoxe qui fixe le nombre des heures devant être chaque semaine consacrées à la conduite du projet. Ce nombre est considéré comme ne devant pas excéder une demi-journée par semaine. C'est pourquoi nous avons compté 15 semaines de travail à raison de 4 heures par semaine et par personne ainsi que deux semaines, les stages, à 35 hebdomadaire et par personne..

D'autre part, le groupe avait inclus dans son budget une somme forfaitaire destinée à couvrir d'éventuels déplacements en direction des centres partenaires du Pôle associé. Ces déplacements étaient considérés comme probables et utiles dans l'intérêt d'une meilleure connaissance des partenaires. Les circonstances n'ont pas voulu que de tels déplacements aient lieu, sinon sous la forme d'une unique et intéressante visite rendue à l'occasion d'un déplacement privé.

Les économies réalisées par rapport au budget prévisionnel ont par conséquent libéré une somme qui a largement couvert les frais exceptionnels occasionnés par des dépenses d'hébergement sur Paris. En raison d'un mouvement social qui hypothéquait un voyage en train le jour-même du Comité de pilotage, les délégués du groupe pour ce CP ont choisi de se rendre à Paris la veille du CP.

ANNEXES A : DOCUMENTS RELATIFS AUX COMITES DE PILOTAGES

Table des annexes A

ANNEXE 1 : LES FORMATS XML	IV
Annexe 1.1. Dublin Core	IV
a. Définition.....	IV
b. Inconvénients	V
c. Avantages.....	VI
Annexe 1.2. BiblioML	VI
a. Définition.....	VI
b. Avantages.....	VII
c. Inconvénients	VII
Annexe 1.3. METS	VIII
a. Définition.....	VIII
b. Avantages.....	IX
c. Inconvénients	IX
ANNEXE 2 : INVESTISSEMENTS À OPÉRER POUR LA FAMDT.....	X
Annexe 2.1. Dans le cas d'un réservoir de données en format XML centralisé	XI
a. Préalable	XI
b. Mise en place d'un module d'export XML	XII
Annexe 2.2. Dans le cas d'un réservoir de données XML pour chaque centre.....	XIII
a. Préalable	XIV
b. XMLisation des notices, et mise en place d'un réservoir de données au format XML	XIV
Annexe 2.3. Dans le cas d'une conversion des notices en XML « à la volée ».....	XV
a. Préalables.....	XVI
b. XMLisation des notices	XVI
ANNEXE 3 : FEUILLE DES LOTS ET TÂCHES.....	XVII
Lot 1 : XMLisation des données	XVII
Tâche 1. Choix préalables.....	XVII
Tâche 2. Etablissement du tableau de concordances.....	XVII
Tâche 3. Rédaction du cahier des charges	XVII
Tâche 4. BnF : Enrichissement des notices	XVII
Tâche 5. BnF : Passation d'un marché pour le programme d'XMLisation	XVIII
Tâche 6. FAMDT : Achat des programmes et du matériel pour la conversion et le stockage des données générées en XML	XVIII
Tâche 7. Vérification de la validité des notices converties par les responsables des fonds	XVIII

Lot 2 : le portail d'interrogation	XVIII
Lot 3 : paramétrage de SDX	XVIII
Tâche 1. Rédaction d'un cahier des charges, traitant notamment de la question des droits	XVIII
Tâche 2. Passation d'un marché avec une société prestataire de services.....	XIX
Tâche 3. Paramétrage de SDX et réalisation du portail d'intégration.....	XIX
Lot 4 : entretien du portail et des bases de données.....	XIX
Lot 5 : Documents numérisés et intégration de nouveaux partenaires.....	XIX
Tâche 1. Définition d'une politique d'extension	XIX
Tâche 2. Planification du projet de numérisation	XX

Annexe 1 : Les formats XML

Annexe 1.1. Dublin Core

a. Définition

Précisons d'abord que Dublin Core⁵⁴ n'est pas une DTD. Mais il fournit quinze zones qui permettent de décrire des documents, qu'ils soient sonores ou non. La sémantique est peu contraignante : les 15 zones sont facultatives, répétables et peuvent être saisies dans n'importe quel ordre.

On peut distinguer ces éléments en trois groupes, selon leur contenu intellectuel.

Contenu	Propriété intellectuelle	Matérialisation
Titre	Auteur	Date
Sujet	Éditeur	Type de ressource
Description	Contributeur	Format
Source	Gestion des droits	Identifiant
Langage		
Relation		
Couverture		

Voici un second tableau fournissant quelques précisions :

Élément	Traduction française	Précisions
Title	Titre	nom donné à la ressource par son auteur ou son éditeur
Subject	Sujet	description du contenu sémantique (mots clef ou

⁵⁴ Site officiel de Dublin Core : <http://www.dublincore.org>. Pour un guide d'utilisation du Dublin Core, cf. <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>.

		fichier d'autorité)
Description	Description	description textuelle du contenu
Source	Source	description d'une autre ressource, dont celle-ci peut être dérivée
Language	Langage	langue utilisée dans la ressource
Relation	Relation	mentions d'autres ressources en relation avec la première
Coverage	Couverture	caractéristiques spatiales et temporelles du contenu (région, époque, ...)
Creator	Auteur	responsable de la création du document
Publisher	Éditeur	entité responsable de l'édition de la ressource
Contributor	Contributeur	responsabilité secondaire par rapport à Auteur
Rights	Gestion des droits	mention des droits d'auteur, de représentation, etc.
Date	Date	date de création ou de publication de la ressource
Type	Type de ressource	catégorisation de la ressource (thèse, rapport, communication, etc.)
Format	Format	format de la ressource (doit identifier le logiciel permettant de la traiter)
Identifier	Identifiant	identificateur unique (ISBN, URL, etc.)

La notice originale, évidemment plus riche et structurée que ce qui est prévu par les zones Dublin Core, rentrerait donc dans un moule étroit où elle perdrait de sa structure, mais non son contenu. Si l'on prend par exemple une notice UPCP-Metive, qui compte 36 champs⁵⁵, plusieurs de ces champs se retrouveraient encadrés par une balise <dc:description>⁵⁶ : RES, TIC, TIRES, INC, DES, NOT1.

b. Inconvénients

⁵⁵Cf.

<http://195.101.116.236/belcerdoupload/metive/cerdo/bascerdo/fiche/son/docfiche/FDC00005.htm>

⁵⁶ On écrit parfois aussi : <META NAME="DC.Description">.

Les inconvénients du Dublin Core sautent aux yeux : la qualité de description est en grande partie perdue. Pour obtenir les renseignements que l'on pouvait jusque là atteindre par catégories, il est à présent incontournable de formuler une interrogation en plein texte.

c. Avantages

Si l'on ne peut y trouver qu'un seul inconvénient, il est de taille. Les avantages à l'inverse sont nombreux. La manipulation de telles notices serait beaucoup plus simple que si elles sont structurées par une DTD plus élaborée, et la conversion serait de ce fait plus aisée et plus rapide. Il ne faut en effet pas oublier que la conversion des notices d'un format structuré dans l'autre doit s'accompagner d'une relecture par échantillons de test pour s'assurer de la correction des nouvelles notices. Or si celles-ci sont trop complexes, cela peut demander aux relecteurs un temps excessif et donc, d'une certaine façon, entraîner la mort du projet.

Par ailleurs, l'emploi de notices structurées de façon simple favorise l'adhésion de nouveaux partenaires au métacatalogue d'archives sonores. Si l'on pense par exemple au Musée des Arts et Traditions Populaires, où l'informatisation du catalogue n'est pas encore faite, il est plus simple de créer des notices avec du Dublin Core que d'avoir affaire directement à une DTD complexe.

Enfin, le Dublin Core permet une bonne gestion des droits, et n'interdit pas un ajout de balises créées à usage interne, ce qui autorise certaines précisions non permises par ce langage, et que les partenaires auraient à cœur.

Annexe 1.2. BiblioML

a. Définition

BiblioML⁵⁷ a été conçu comme la transformation d'UNIMARC en XML, non pour pouvoir rajouter des informations supplémentaires ou une structuration plus

⁵⁷ BiblioML est développé et géré par l'ADNX, Association pour la documentation numérique en XML (<http://www.adnx.org>), qui gère également SDX. Cf. les outils et la documentation sur BiblioML à l'adresse : <http://adnx.org:8080/JSPWiki/Wiki.jsp?page=BiblioML>. On peut trouver comme applications : la Bibliographie nationale française (<http://bibliographienationale.bnf.fr/>), et deux catalogues de bibliothèques qui sont encore des prototypes : La Bibliothèque francophone pour l'Asie-Pacifique ou BEFAP (<http://brea.culture.fr/sdx/befap/index.xsp>) et le centre de documentation de la Direction du Livre et de la Lecture (<http://brea.culture.fr/sdx/biblioml/index.xsp>).

élaborée, mais pour permettre un traitement des notices plus souple, tel que le permet le XML – et que ne permet pas UNIMARC.

Le répertoire de balises produit par BiblioML⁵⁸ reproduit toute la richesse d'UNIMARC, donc d'une qualité intellectuelle nettement satisfaisante.

Une telle notice se structure généralement selon quatre parties, dont voici le modèle :

```
<BiblioRecord>
  <Meta>
  <Description>
  <Subjects>
  <Copies>
</BiblioRecord>
```

b. Avantages

Intellectuellement, la structuration reprise d'UNIMARC est parfaitement satisfaisante. En outre, la conversion en XML est déjà bien balisée, et nécessiterait a priori des relectures peu contraignantes.

BiblioML est tout à fait apte à gérer la question des droits sur l'accès aux documents, par l'utilisation de précisions dans les données locales, dont l'application SDX tiendrait compte. La gestion des droits pourrait se trouver en données locales, et serait indiquée de la sorte :

```
<Copies>
  <Copy>
    <Note Value="Access"
```

c. Inconvénients

Inconstestablement, on obtient avec BiblioML des notices plus complexes que les précédentes, et qui nécessitent une période de familiarisation, et éventuellement de formation.

⁵⁸ La liste de ces balises est fournie à l'adresse suivante : <http://adnx.org:8080/JSPWiki/Wiki.jsp?page=BiblioML> (« La documentation en HTML »).

Par ailleurs, BiblioML, qui reprend la structuration du format UNIMARC, obéit à une logique bibliothéconomique, qui peut ne pas correspondre aux critères de description documentaire en usage dans les musées ou autres centres de conservation patrimoniaux. Dans le cas d'intégrations de partenaires issus du monde des musées, l'emploi de BiblioML pourrait donc poser quelques problèmes quant à sa logique descriptive.

En outre, si de futurs partenaires, par exemple des musées, décident d'utiliser Dublin Core parce que l'adoption de BiblioML serait pour eux beaucoup trop lourde à mettre en place, l'emploi de BiblioML par les premiers partenaires du métacatalogue se révélerait un effort inutile : l'application SDX ne pourrait plus interroger ces notices en BiblioML qu'au même degré de précision que le Dublin Core, par une conversion à la volée des notices de BiblioML vers Dublin Core.

Annexe 1.3. METS

a. Définition

METS (Metadata Encoding and Transmission Standards)⁵⁹ est, comme Dublin Core, un format de métadonnées, beaucoup plus complexe et riche dans sa structure.

Il est composé de cinq parties :

- Descriptive metadata** (métadonnées de description du document), qui peuvent être internes au document, ou renvoyer par exemple à un document UNIMARC ou d'un autre format.

- Administrative metadata** (métadonnées comportant les informations sur la création, la provenance et le stockage du fichier, la gestion des droits de propriété intellectuelle, le document original dont est issu l'objet numérique considéré).

- File group** (liste de tous les fichiers contenant toutes les versions du document numérique) – cette partie, *a priori*, ne nous concerne pas.

⁵⁹ Site officiel : <http://www.loc.gov/standards/mets/>. Cf. aussi : <http://loc.gov/standards/mets/news031703.html> et une liste d'applications de METS à l'adresse : <http://sunsite.berkeley.edu/~rbeaubie/metsimpl/>.

-Structural map (la "carte structurelle" établit une structure hiérarchique pour l'objet numérique et construit les liens existants entre les éléments de cette structure et les métadonnées relatives à chacun des éléments, pour permettre à l'utilisateur de naviguer logiquement entre ces éléments).

-Behaviour (section permettant d'associer des programmes exécutables avec le contenu de l'élément METS).

METS correspond donc pleinement au projet de métacatalogue. En effet, BiblioML est conçu pour la description de métadonnées, mais ne prévoit pas qu'au catalogue en XML s'ajoute l'accès aux documents numériques. C'est précisément ce pour quoi est conçu METS, dans chacune de ces parties : pour décrire des documents accessibles en ligne, et leur associer par exemple des applications. La gestion des droits y est également très précise.

b. Avantages

Un avantage considérable réside dans le fait que ce format de métadonnées sait faire absolument tout ce qui est exigé par le présent projet, et peut même aller bien au-delà. Il offre donc pour l'avenir de considérables possibilités, si par exemple le métacatalogue est amené à s'étendre à des fonctions extérieures à la seule recherche documentaire.

c. Inconvénients

La manipulation de METS par des non-informaticiens est bien plus délicate que celle de Dublin Core, ou même de BiblioML. Sa logique, sa structure, sont tout orientées vers les possibilités offertes par Internet et l'accès en ligne, alors que les autres solutions proposées sont plus traditionnelles, moins risquées et surtout plus maniables.

Le choix d'adopter METS peut décourager, d'une part de futurs partenaires potentiels du métacatalogue, mais peut-être même les acteurs présents du projet.

Annexe 2 : Investissements à opérer pour la FAMDT

Note : Ces tableaux des investissements sont ceux qui ont été fournis au 3^{ème} comité de pilotage. Nous avons appris entre-temps que la MMSH était déjà passé sous le logiciel Alexandrie : il convient donc en réalité de retirer les coûts liés à l'achat de ce logiciel et de ne prendre en compte que la mise en ligne, l'acquisition d'un script macro et le réglage du module d'export au format XML pour ce Centre.

Annexe 2.1. Dans le cas d'un réservoir de données en format XML centralisé⁶⁰

La mise en place de cette solution s'effectue en trois étapes.

Chaque centre met en ligne son catalogue selon un modèle « classique » : les données sont stockées dans un format correspondant au SIGB que le centre utilise.

Un des partenaires héberge l'application de recherche de type SDX, ainsi qu'une base de données. Dans cette base de données seront stockées au format XML les notices des cinq partenaires. C'est dans cette base que l'application ira chercher les données.

Chaque centre envoie vers ce serveur de données ses notices au format XML par le biais d'un module d'export au format XML, au rythme de l'enrichissement de son catalogue.

Dans ce cas de figure, les centres de la FAMDT peuvent très bien se contenter d'un site internet « simple », sans catalogue en ligne, comportant un lien vers l'adresse de l'application de recherche centrale. SDX permettant de gérer les accès, une page d'accueil et une pré-sélection automatique de notices adaptées à chaque centre est envisageable.

a. *Préalable*

Changement de logiciel documentaire et mise en ligne des catalogues pour UPCP-Metive, le Conservatoire occitan et la MMSH : Acquisition d'un logiciel de gestion documentaire permettant cette fonctionnalité Récupération des données et réglage du module d'export UNIMARC (4 jours de travail par centre, en deux fois séparées par dix jours de tests)	2300 € 3200 € 2000 €
Coût de la mise en ligne sur Internet	500 €
Contrat de maintenance	8000 €
Total par centre	24 000 €

⁶⁰ Plusieurs réserves sont à émettre sur cette prévision des coûts comme pour les deux suivantes : tout d'abord nous avons intégré à ces tableaux des coûts optionnels, comme une conversion éventuelle des notices de la FAMDT en UNIMARC au cas où la DTD choisie le rendrait nécessaire, ainsi que l'acquisition et le réglage d'un script de transfert « macro » des notices au format XML dans le cas où on choisirait de stocker les notices. Ensuite, en raison du laps de temps imparti, il ne nous a pas été possible de vérifier certains chiffres, notamment ceux concernant les réglages et la conversion de notices.

Total pour les trois centres	
Personnalisation de l'interface web	800 € x 3 = 2400 €
Frais matériels et logiciels ⁶¹	Environ 5000 € par centre
Total pour les trois centres	15000 €
Mise à jour de la version logicielle de Dastum	
Frais de logiciel et formation	1200 € avec formation
Réglage du module d'export UNIMARC pour Dastum (2 x 1 jour de travail entrecoupé de 10 jours de test)	1600 €
Coût total	Environ 44200 €
Coût total par centre	Entre 2800 € et 13 800 €

b. Mise en place d'un module d'export XML

Paramétrage d'un module d'export XML pour chacun des centres en fonction de la DTD décidée (2 jours de travail)	1600 € x 4
Soit pour tous les centres	6400 €
Acquisition d'un « script » de transfert des données : Coût du programme (ou du module ?)	1800 € (HT) x 4 = 7200 €
Frais de personnalisation pour chaque centre	800 € x 4 = 3200 €
Soit au total	10400 €
Formation d'un informaticien sur place (1 journée)	1000 € x 4 = 4000 € pour tous les centres
Coût total	20 800 € soit 5200 € par centre
Coût total incluant les préalables	65000 € soit entre 8000 € et 19000 € par centre

⁶¹ Tous les centres de la FAMDT disposent déjà d'un site internet ou d'un catalogue en ligne : il est donc difficile d'estimer si les équipements en leur possession seront en mesure d'absorber les nouvelles applications ou s'ils devront s'équiper de matériel neuf. Le coût de cette nouvelle installation peut en effet varier du simple au double. Nous sommes partis du principe que le plus important des centres, qui dispose déjà d'un gros équipement, n'aura pas à acquérir un nouveau matériel.

Annexe 2.2. Dans le cas d'un réservoir de données XML pour chaque centre

Cette solution diffère de la précédente sur un point. La différence tient dans le fait que, si l'hébergement de l'application de recherche commune de type SDX est assuré par l'un des partenaires, comme dans la précédente solution, les notices restent la possession de chaque centre.

Parallèlement à son catalogue, chaque centre met donc en ligne un serveur de données spécifique⁶² dans lequel il entrepose ses notices au format XML. L'alimentation du serveur peut s'effectuer au rythme de l'enrichissement du catalogue par le biais d'un module d'export au format XML.

Comme la précédente, cette solution a le désavantage de faire coexister deux réservoirs de données. Là encore cependant, le transfert régulier des notices vers la base de données XML par le biais d'un « script » macro pourrait réduire le risque que les catalogues parallèles finissent par s'éloigner les uns des autres.

Ce cas de figure est également un peu plus coûteux que le précédent puisque chaque centre est dans l'obligation de posséder son propre serveur de données internet et de le rendre accessible. Ce surcoût doit cependant être mesuré sur le long terme. Si l'on part en effet du principe que le projet de métacatalogue n'est que le début d'une plus vaste opération, une opération qui verra la mise en ligne d'une partie des enregistrements sonores numérisés, alors, ceux-ci devant nécessairement être stockés « en dur », la dépense supplémentaire apparaît justifiée.

⁶² Théoriquement, les documents en format XML occupant très peu de mémoire, les centres devraient pouvoir stocker leurs notices au format du SIGB et celles en format XML sur le même serveur de données. Cependant, dans l'optique d'un accroissement des documents XMLisés (son numérisé par exemple), et pour des raisons de sécurité, avoir deux serveurs de données séparés constitue la configuration optimale.

a. *Préalable*

Changement de logiciel documentaire et mise en ligne des catalogues pour UCP-Metive, le Conservatoire occitan et la MMSH :	2300 €
Acquisition d'un logiciel de gestion documentaire permettant cette fonctionnalité	
Récupération des données et réglage du module d'export UNIMARC (4 jours de travail par centre, en deux fois séparées par dix jours de tests)	3200 €
Coût de la mise en ligne sur Internet	2000 €
Contrat de maintenance	500 €
	8000 €
Total par centre	24 000 €
Total pour les trois centres	
Personnalisation de l'interface web	800 € x 3 = 2400 €
Frais matériels et logiciels ⁶³	Environ 5000 € par centre
Total pour les trois centres	15000 €
Mise à jour de la version logicielle de Dastum	
Frais de logiciel et formation	1200 € avec formation
Réglage du module d'export UNIMARC pour Dastum (2 x 1 jour de travail entrecoupé de 10 jours de testage)	1600 €
Coût total	Environ 44200 €
Coût total par centre	Entre 2800 € et 13 800 €

b. *XMLisation des notices, et mise en place d'un réservoir de données au format XML*

Paramétrage d'un module d'export XML pour chacun des centres en fonction de la DTD décidée (2 jours de travail)	1600 € x 4
Soit pour tous les centres	6400 €
Acquisition d'un « script » de transfert des données : Coût du programme (ou du module ?)	1800 € (HT) x 4 = 7200 €
Frais de personnalisation pour chaque centre	800 € x 4 = 3200 €
Soit pour tous les centres	10400 €
Acquisition d'un serveur de données supplémentaire destiné à stocker les informations au format XML	
Matériel	1500 € x 4 = 6000 €
Installation (1/2 journée)	400 € x 4 = 1600 €
Coût total	1900 € par centre, soit 7600 € au total
Formation d'un informaticien sur place (1 journée)	1000 € x 4 = 4000 € au total
Coût total	28400 € soit environ 7100 € par centre
Coût total incluant les préalables	72600 € soit entre 9900 € et 20900 € par centre

⁶³ Tous les centres de la FAMDT disposent déjà d'un site internet ou d'un catalogue en ligne : il est donc difficile d'estimer si les équipements en leur possession seront en mesure d'absorber les nouvelles applications ou s'ils devront s'équiper de matériel neuf. Le coût de cette nouvelle installation peut en effet varier du simple au double. Nous sommes partis du principe que Dastum, qui dispose déjà d'un gros équipement, n'aura pas à acquérir un nouveau matériel.

Annexe 2.3. Dans le cas d'une conversion des notices en XML « à la volée »

Chaque centre met en ligne son catalogue selon un modèle « classique » avec des données stockées dans un format correspondant au SIGB qu'il utilise.

Un des partenaires héberge l'application de recherche commune de type SDX.

Parallèlement à son catalogue, chaque centre acquiert et fait paramétrer selon la DTD choisie un module d'export XML. Ce module permet de convertir « à la volée » les notices en XML au moment où l'application de recherche centrale vient « piocher » dans les catalogues.

Des trois solutions envisagées, cette solution-ci est probablement la plus immédiatement opérationnelle pour les centres. Une fois le travail préliminaire de concordance et de réglage établi, les centres n'ont plus en effet à se préoccuper de stocker leurs notices en XML. Il faut toutefois souligner que, pour que la solution fonctionne de manière satisfaisante, le paramétrage de cette conversion doit être très précis. Cela nécessite un travail de réglage préliminaire assez lourd.

a. *Préalables*

Changement de logiciel documentaire et mise en ligne des catalogues pour UPCP-Metive, le Conservatoire occitan et la MMSH	2300 €
Acquisition d'un logiciel de gestion documentaire permettant cette fonctionnalité	3200 €
Récupération des données et réglage du module d'export UNIMARC (4 jours de travail par centre, en deux fois séparées par dix jours de tests)	2000 €
Coût de la mise en ligne sur Internet :	500 €
Contrat de maintenance	8000 €
Total par centre	24 000 €
Total pour les trois centres	
Personnalisation de l'interface web	800 € x 3 = 2400 €
Frais matériels et logiciels ⁶⁴	Environ 5000 € par centre
Total pour les trois centres	15000 €
Mise à jour de la version logicielle de Dastum	
Frais de logiciel et formation	1200 € avec formation
Réglage du module d'export UNIMARC pour Dastum (2 jours de travail)	1600 €
Coût total	Environ 44200 €
Coût total par centre	Entre 2800 € et 13 800 €

b. *XMLisation des notices*

Paramétrage d'un module d'export XML pour chacun des centres en fonction de la DTD décidée (3 jours de travail)	2400 €
Soit pour tous les centres	9600 €
Formation d'un informaticien sur place (1 journée)	1000 € x 4 = 4000 € au total
Coût total	13600 € soit environ 3400 € par centre
Coût total incluant les préalables	57800 € soit entre 6200 € et 17200 € par centre

⁶⁴ Tous les centres de la FAMDT disposent déjà d'un site internet ou d'un catalogue en ligne : il est donc difficile d'estimer si les équipements en leur possession seront en mesure d'absorber les nouvelles applications ou s'ils devront s'équiper de matériel neuf. Le coût de cette nouvelle installation peut en effet varier du simple au double. Nous sommes partis du principe que Dastum, qui dispose déjà d'un gros équipement, n'aura pas à acquérir un nouveau matériel.

Annexe 3 : Feuille des lots et tâches

Préalable : mise en ligne des catalogues⁶⁵.

Lot 1 : XMLisation des données

Tâche 1. Choix préalables

1 jour

Choix de la DTD ou du format de métadonnées.

Choix de la répartition des coûts.

Choix de la répartition des catalogues (en base commune, stockés par chaque partenaire, ou XML créé à la volée).

Tâche 2. Etablissement du tableau de concordances

2 semaines

Conversion : Format des notices → UNIMARC → Format XML choisi

Tâche 3. Rédaction du cahier des charges

2 semaines

Tâche 4. BnF : Enrichissement des notices

7 mois

⁶⁵ Cf. *infra*, annexe 2.

Tâche 5. BnF : Passation d'un marché pour le programme d'XMLisation

2 mois

Tâche 6. FAMDT : Achat des programmes et du matériel pour la conversion et le stockage des données générées en XML

1 mois

Tâche 7. Vérification de la validité des notices converties par les responsables des fonds

3 semaines

Lot 2 : le portail d'interrogation

1 semaine

Définition des fonctionnalités du portail d'interrogation : modalités de recherche, fonctions autres que la seule recherche documentaire (historique, ...), affichage des résultats (XML, HTML).

Forme : déterminer son image, ce qu'il affiche à l'écran de l'utilisateur.

Contenu : choisir les champs d'interrogation.

Définir le contenu des menus déroulants (s'il y en a, ce qui est souhaitable parce que ça cadrerait la saisie des chercheurs).

Etablir des accès avec d'autres catalogues, ou bien avec le catalogue de chaque centre.

Choisir les options qui personnaliseront la fréquentation (panier, historique, profil...).

Anticiper l'intégration de nouvelles fonctionnalités : téléchargement de documents sonores par exemple.

Lot 3 : paramétrage de SDX

Tâche 1. Rédaction d'un cahier des charges, traitant notamment de la question des droits

2 semaines

Il s'agit d'adapter SDX et ses fonctionnalités de recherche (recherches par champs ou en plein texte ; présentation des résultats ; accès aux notices complètes à l'exception de quelques informations qu'il conviendra de tenir hors de la connaissance du public, comme l'identité de certains informateurs⁶⁶) au projet de catalogue collectif d'archives sonores.

Tâche 2. Passation d'un marché avec une société prestataire de services

1 mois

Tâche 3. Paramétrage de SDX et réalisation du portail d'intégration

2 mois

Lot 4 : entretien du portail et des bases de données

Enrichissement progressif

On peut estimer, selon des chiffres cependant assez arbitraires et sans doute hasardeux, le temps de travail suivant :

BnF : 4 heures/mois

FAMDT : 2 heures/mois.

Lot 5 : Documents numérisés et intégration de nouveaux partenaires

Tâche 1. Définition d'une politique d'extension

1 jour

Tous les partenaires actuels doivent s'entendre sur les critères permettant ou excluant l'accueil ou l'invitation de futurs candidats à l'adhésion au catalogue collectif virtuel.

Dans la même logique, il faut impérativement fournir un *cadre* à ce projet de métacatalogue, déterminer de quelle logique il participe : s'il est effectué

⁶⁶ Au sens de la FAMDT : personne contactée et enregistrée par un collecteur dans le cas d'une collecte patrimoniale.

uniquement dans le cadre du pôle associé (auquel cas la BnF a une forte prépondérance sur les autres partenaires) ou dans une autre optique.

En outre, il faut désigner un arbitre ou modérateur, en cas de problème dans l'existence du catalogue collectif virtuel, ainsi qu'une fréquence de réunion (annuelle, semestrielle, ...) pour s'assurer de régulièrement du bon fonctionnement du système.

Tâche 2. Planification du projet de numérisation

1 semaine

Il faut calculer le taux d'accroissement moyen pour chaque partenaire, en vue du calcul de la taille du serveur nécessaire pour héberger les documents numérisés à court, moyen et long terme.

ANNEXES B :

DOCUMENTS

DE TRAVAIL

INTERNE

Table des annexes B

ANNEXE 4 : GRILLE DU QUESTIONNAIRE À LA FAMDT	XXIII
ANNEXE 5 : JOURNAL DU PROJET	XXXIV

Annexe 4 : grille du questionnaire à la FAMDT

Note : dans le cadre de notre étude, il nous était nécessaire de rassembler un certain nombre de renseignements sur les centres de la FAMDT disséminés en province : état du catalogage des enregistrements sonores inédits, enrichissement des fonds, possibilités techniques offertes sur place... En raison de leur éloignement il ne nous a pas toujours été possible de nous rendre sur place. La présente grille de questionnaire a été réalisée pour s'assurer que le même type d'information serait recueilli dans tous les centres sans qu'un point soit oublié ou moins étudié d'un centre à l'autre.

Domaines abordés par le questionnaire :

I. les moyens

Les moyens humains

Les moyens matériels

Les moyens financiers

Origine des ressources

Subventions

Partenariats

II. le catalogue et le traitement documentaire

Le classement

Fonds catalogué/en attente de catalogage

Formats et niveaux de catalogage

Différents niveaux de catalogage adoptés dans votre Centre

La configuration de votre catalogue

III. Le fonds

Typologie des enregistrements sonores inédits

Accroissement des collections d'enregistrements sonores inédits

Repérage des enregistrements sonores inédits au sein du catalogue

Numérisation

IV. L'analyse des questions juridiques

V. Les publics

VI. Attentes relatives au projet

I. LES MOYENS

Les moyens humains

Dont personnel qualifié documentation.

Dont personnel bénévole.

Dont personnel avec compétence informatique (lesquelles, quel niveau).

Sur les questions informatiques, êtes-vous en contact avec une personne ressource ou une organisation qui vous aide à gérer votre parc informatique (responsable informatique d'une collectivité partenaire, service technique d'un prestataire de service)

Pouvez-vous évaluer approximativement la proportion de temps passé à cataloguer (combien de personnes s'en occupent ? Quelle proportion de leur temps cette tâche leur prend-elle ?)

Quel est le temps moyen nécessaire à l'analyse et au catalogage d'un enregistrement sonore ?

Les moyens matériels

Votre centre se situe-t-il sur un ou sur plusieurs sites ?

De quels éléments votre parc informatique se compose-t-il (nombre de postes et configuration client/serveur, pare-feu...) ?

Votre site internet

Disposez-vous d'un site web ?

Si oui :

Hébergez-vous le site vous-même ou est-il hébergé par un partenaire institutionnel ou un prestataire de service ?

Nom de la personne-contact pour des renseignements techniques :

La conception puis la réalisation du site ont-elles été faites en interne ou avez-vous fait appel à un prestataire extérieur ?

Dans ce dernier cas quel est le nom du prestataire de service auquel vous avez fait appel ?

De même, la gestion technique du site (mises à jours...) se fait-elle en interne ou est-elle déléguée ?

Dans ce dernier cas, quel est le nom du prestataire de service auquel vous avez fait appel ?

Connaissez-vous le nombre quotidien, hebdomadaire ou mensuel de connexions au site ?

Quels sont les services offerts sur votre site ?

Dans le cas d'un catalogue consultable sur le site, quelle interface utilisez-vous pour la mise en ligne du catalogue ?

Avez-vous des projets en cours dans ce domaine (création d'un site entier, développement de nouveaux services, mise en ligne de votre catalogue...) ?

Les logiciels

Quels sont les différents logiciels professionnels que vous utilisez dans votre centre de documentation (SGIB, SGBD...) ?

Le logiciel employé pour le catalogage inclut-il un module de catalogage assisté ou faites-vous du catalogage en format ?

Pourriez-vous nous apporter quelques précisions concernant les caractéristiques techniques de votre SIGB, notamment concernant :

Les possibilités qu'il offre en matière d'import et d'export de données (format XML, UNIMARC...), et si l'intégration ou la migration vers d'autres bases de données sont possibles.

Les possibilités d'évolutions de ce système (ajout de fonctionnalités...) ;

La compatibilité Z 39-50 : quel est l'obstacle principal s'il existe ?

Les moyens financiers

Il apparaît nécessaire de ne pas envisager le projet du métacatalogue en-dehors de toute considération du coût d'une telle opération. Il faut donc mesurer de quel budget annuel dont disposent les partenaires.

Origine des ressources

Subventions

Etat :

Région

Département :

Commune :

Fonds européens :

BnF :

Mécénat :

Cotisation des adhérents :

Recettes propres :

Autres :

Partenariats

En ce qui concerne les enregistrements sonores, avez-vous noué des partenariats particuliers (notamment dans le cadre du label de centre des musiques et danses traditionnelles) ?⁶⁷

II. LE CATALOGUE ET LE TRAITEMENT DOCUMENTAIRE

Le classement

Quel est le type de classement et de description adopté dans votre centre : par fonds ? Par collections ? Par supports ? Par genre ? Par items ?

Fonds catalogué/en attente de catalogage

Quel est la proportion du fonds des enregistrements sonores qui sont catalogués...

...sur la base informatique ?

...sur papier ?

À quelle part du fonds estimez-vous la quantité de notices restant à informatiser et à créer ?

Formats et niveaux de catalogage

Quel est le format du catalogage :

UNIMARC ?

Format développé conforme aux normes établies par la FAMDT dans le *Guide d'analyse documentaire du son inédit* ?

Format « maison » (par exemple parce que l'on trouve dans votre centre des types d'enregistrements sonores particuliers qui auraient nécessité une adaptation de ces normes ?)

Dans ces deux derniers cas, votre logiciel inclut-il un programme de conversion de votre format de catalogage vers le format UNIMARC ?

Différents niveaux de catalogage adoptés dans votre Centre

Notice générale de l'enregistrement sans description par item (proportion par rapport à l'ensemble des notices) :

Description et catalogage jusqu'à l'item (proportion par rapport à l'ensemble des notices) :

À ce sujet, quelle est la configuration qu'adopte votre logiciel pour lier les notices générales aux notices d'items ?

- notices séparées dans un même fichier avec hiérarchie notice mère/notices filles.

- fichiers d'items indépendants incluant la référence du document dont ils sont issus.

Y a-t-il un lien permettant de retomber sur la notice générale à partir de la notice de l'item et vice versa ?

La configuration de votre catalogue

Y a-t-il un unique format professionnel de saisie ou le logiciel gère-t-il un affichage et un module de recherche simplifié pour le public ?

Dans ce cas, quelles sont les différences entre les deux (champs apparaissant dans l'un et pas dans l'autre, modalités de recherche...) ?

Sur quels champs le lecteur peut-il opérer sa recherche ?

Quels champs vous semblent-ils indispensables à un outil de recherche efficace pour le public ?

Champs de la notice générale :

Identification	Nom du centre
	Localisation de l'original

Mode de consultation : Niveau de consultation possible (utilisation d'un code ?)

Désignation	Type de document
	Genre du document
	Nature de l'enregistrement

Auteur(s)	Date de collecte et d'enregistrement
Enquêteur(s)	
Interprète(s)	

Titres	Titre du document sonore
--------	--------------------------

Description technique

Analyse	Descripteur thématique
---------	------------------------

Champs de l'item :

Identification	Numéro de référence d'où est issu l'item
----------------	--

Mode de consultation Niveau de consultation possible (utilisation d'un code ?)

Désignation	Genre de l'item Nature de l'item
Auteur(s) Enquêteur(s) Interprète(s)	Date de collecte et d'enregistrement
Titres	Titre de l'item
Description technique	
Analyse	Descripteur thématique

Pour les champs faisant l'objet d'une description par vocabulaire contrôlé (dans le cas d'une recherche par descripteur thématique, ou pour le genre et la nature de l'enregistrement), le chercheur peut-il faire apparaître un menu déroulant dans certains champs ?

Sinon, le logiciel gère-t-il une liste de vocabulaire contrôlé pour chaque champ ?

Utilisez-vous un thésaurus (pour les descripteurs thématiques par exemple) et si oui, pour quels champs ? (*merci de fournir quelques exemples du thésaurus*)

Utilisez-vous un thésaurus propre à l'établissement ? Avez-vous été amenés à ajouter des termes au vocabulaire contrôlé préconisé dans le *Guide d'analyse documentaire du son inédit* ?

III. LE FONDS

Typologie des enregistrements sonores inédits

Témoignage

Histoires de vie

Chansons

Contes

Musique

Danse

Langue(s) utilisée(s)

Domaine géographique couvert

Quels sont les différents types de supports d'enregistrements sonores inédits que vous conservez (*exemples*) ?

Accroissement des collections d'enregistrements sonores inédits

Si vos collections d'enregistrements sonores inédits continuent à s'enrichir, pouvez-vous évaluer le rythme d'accroissement ?

Par quels moyens l'accroissement se fait-il (dons, achats, campagnes de collectes...) ?

Le temps pour le catalogage des fonds nouveaux pose-t-il problème (quel est le délai moyen d'intégration au catalogue) ?

Repérage des enregistrements sonores inédits au sein du catalogue

Y a-t-il possibilité d'isoler les enregistrements sonores inédits au sein de votre catalogue (recherche restreinte, indicateur codé) ?

Ce type de répertoire existe-t-il sous forme imprimée ?

Sinon, comment évaluez-vous les possibilités de réaliser un tel inventaire dans votre centre (en terme de coût, de temps du personnel...) ?

Numérisation

Quels supports utilisez-vous pour la conservation des documents ?

Faites-vous systématiquement des copies numériques des bandes magnétiques ?

Quel est le pourcentage des enregistrements numérisés, par type de documents ?

Quelle est l'accessibilité de ces enregistrements par rapport à celle des originaux (sur place, intranet, extranet, internet) ?

IV. L'ANALYSE DES QUESTIONS JURIDIQUES

Si le dépôt des fonds sonores a fait l'objet d'une convention entre le collecteur et l'informateur et/ou entre le donateur et l'association, quelle part des fonds cela représente-t-il ?

Avez-vous établi une politique précise en ce qui concerne la communication des documents du point de vue juridique (droit d'interprète) ?

Quelle est votre opinion concernant l'idée de mettre en ligne, à terme, en extranet ou même sur internet, certains enregistrements sonores (implications juridiques, conditions nécessaires à cette mise en place) ?

V. LES PUBLICS

Quelles sont les conditions d'accès aux enregistrements sonores inédits que vous conservez ?

Avez-vous effectué une enquête sur vos publics ?

Quelle part du public les chercheurs⁶⁸ représentent-ils ? L'origine de ce public est-elle locale ou plus lointaine ?

Comment se déroule l'identification du document à communiquer ?

... les personnes prennent rendez-vous en vous signalant leur domaine de recherche et vous effectuez une recherche préalable à leur visite.

... chacun effectue une recherche autonome sur le catalogue.

... le chercheur et vous-même cherchez ensemble.

Votre travail d'accompagnement est-il important plutôt parce que les chercheurs demandent un conseil, ou plutôt parce que l'outil de recherche ne leur permet pas une recherche autonome ?

Les chercheurs ont-ils exprimé certains besoins ?

D'une manière générale, votre public est-il familiarisé avec l'outil informatique (par exemple : vous parle-t-il de votre site internet) ?

VI. ATTENTES RELATIVES AU PROJET

Qu'espérez-vous d'un catalogue mutualisé ?

Quelle importance accordez-vous au fait de mutualiser les catalogues d'enregistrements sonores de différentes régions ?

Dans quel domaine ce projet vous inspire-t-il des craintes ?

Comment imaginez-vous le catalogue collectif ?

Quelle fonctionnalité voudriez-vous trouver dans un tel catalogue (recherche en modes expert ou grand public, possibilité de choisir entre plusieurs modes de recherche...) ?

⁶⁸ Le terme "chercheur" est employé ici dans son acception la plus large.

Annexe 5 : journal du projet