

Normalisation du transfert informatique des index et répertoires des archives historiques

Rapport final d'étude

Archives d'Etat de Genève

Version	1.0
Statut	Pour information
Date	20.12.02
Référence	NM-RF-AEG-2012-02
Auteur	N. Marquet

Rapport final d'étude

Entête de document			
Titre du Document / Sous-titre	Normalisation du transfert informatique des index et répertoires des archives historiques / Rapport final d'étude		
Client	Archives d'Etat de Genève		
Type de Document	Rapport de synthèse		
Référence du Document	NM-RF-AEG-2012-02		
Numéro de Révision	1.0		
Statut	Pour information		
Dernière mise à jour	20.12.02		
Dernière mise à jour	N. Marquet		
Nom du Fichier	RapportSyntheseAEG_Norme.doc		
Liste de distribution			
Société	Nom	Objet de la distribution (Information, Approbation)	
Archives d'Etat de Genève	Roger Rosset	Information	
Systèmes d'Information et Géomatique	François Mumenthaler	Information	
Centre des Technologies de l'Information	Patrick Morand	Information	
Altran	Xavier Mirabaud	Information	
Liste d'approbation			
Titre	Nom	Date	Signature
Chef de Projet AEG	Roger Rosset		
Consultant ATS	Nathalie Marquet	20.12.2002	
Historique de révisions			
Version	Date	Nom(s)	Description
1.0	20.12.2002	Nathalie Marquet	Création du document

Rapport final d'étude

Table de référence des documents				
Code	Titre du document	Référence	Vers.	Date
NM-CDC-AEG-2709-02	Cahier des Charges		1.1	27.09.2002
NM-PP-AEG-0310-02	Plan de projet		1.0	03.10.2002
NM-EP-AEG-1110-02	Rapport d'étude préliminaire		1.0	11.10.2002
NM-PJ-AEG-2810-02	Rapport d'analyse pour le Pouvoir Judiciaire		1.0	28.10.2002
NM-CE-AEG-2810-02	Rapport d'analyse pour la Chancellerie		1.0	28.10.2002
NM-AAEG-AEG-0511-02	Rapport d'analyse pour les Archives d'Etat		1.1	05.11.2002
NM-DAEL-AEG-0611-02	Rapport d'analyse pour le DAEL		1.0	06.11.2002
NM-DIP-AEG-1211-02	Rapport d'analyse pour le DIP		1.0	12.11.2002
NM-HUG-AEG-2511-02	Rapport d'analyse pour le DASS		1.0	25.11.2002
NM-DIAE-AEG-2811-02	Rapport d'analyse pour le DIAE		1.0	28.11.2002
NM-N-AEG-1212-02	Norme		1.1	12.12.2002

Table des abréviations	
Abréviation	Signification
AEG	Archives d'Etat de Genève
CIA	Conseil International des Archives
DAEL	Département de l'Aménagement, de l'Equipeement et du Logement
DASS	Département de l'Action Sociale et de la Santé
DIAE	Département de l'Intérieur, de l'Agriculture et de l'Environnement
DIP	Département de l'Instruction Publique
DTD	Définition de type de document
EAD	Encoded Archival Description
HUG	Hôpitaux Universitaires de Genève
HVU	Module « High Volume Update » du logiciel BASIS
ISAAR	Norme internationale du CIA sur les notices d'autorité archivistiques relatives aux collectivités, aux personnes et aux familles
ISAD(G)	Norme générale et internationale de description archivistique du CIA
LAP	Loi sur les archives publiques, 01.12.2000
LIPAD	Loi sur l'information du public et l'accès aux documents, 05.10.2001
LITAO	Loi sur les informations traitées automatiquement par ordinateur
OCA	Organismes Créateurs d'Archives
SGML	Standard Generalized Markup Language
W3C	Consortium du Web
XML	Extensible Markup Language

Table des matières

1. RESUME DU DOCUMENT.....	6
2. BUT DU DOCUMENT	8
3. BUT DE L'ETUDE	8
3.1 TRAITEMENT DES ARCHIVES AUX ARCHIVES D'ETAT.....	8
3.1.1 <i>Généralités.....</i>	8
3.1.2 <i>Classification des fonds.....</i>	9
3.1.3 <i>Réalisation des inventaires</i>	9
3.2 LES MOTIVATIONS DU PROJET	10
3.3 CONSEQUENCES POUR LES SERVICES PRODUCTEURS D'ARCHIVES	10
4. PRINCIPE DE LA NORME.....	12
5. RAPPEL DU DEROULEMENT DE L'ETUDE.....	13
5.1 INTRODUCTION.....	13
5.2 LES ACTEURS DE L'ETUDE	13
5.3 LES RESULTATS ET LIVRABLES INTERMEDIAIRES	14
5.3.1 <i>Etude préliminaire.....</i>	14
5.3.2 <i>Analyse pour la définition du standard du point de vue des Archives d'Etat.....</i>	15
5.3.3 <i>Etude de la gestion informatique des archives dans huit services producteurs d'archives...</i>	16
5.3.4 <i>Récapitulatif des livrables intermédiaires.....</i>	17
5.4 CONSTRUCTION ET REDACTION DU STANDARD	18
6. CONCLUSION GENERALE DES ANALYSES INTERMEDIAIRES	19
6.1 ETUDE MENEES AUX ARCHIVES D'ETAT	19
6.2 ETUDES MENEES DANS LES SERVICES VERSANTS	20
6.3 CONCLUSION QUANT A LA FAISABILITE DE LA NORME	21
7. CONSTRUCTION DE LA NORME	23
7.1 PREAMBULE.....	23
7.2 INTRODUCTION.....	23
7.3 LE CHAMP D'APPLICATION DE LA NORME	24
7.4 LA DESCRIPTION DES DOCUMENTS D'ARCHIVES.....	24
7.4.1 <i>Typologie des documents à indexer.....</i>	24
7.4.2 <i>Les niveaux de description</i>	25
7.5 LA DEFINITION DES INDEX	28
7.5.1 <i>Les données</i>	28
7.5.1.1 Définition des données	28
7.5.1.2 Correspondance entre les index définis dans la norme et les données dans BASIS	29
7.5.2 <i>La structure des données</i>	31
7.5.2.1 Définition des types de données	31
7.5.2.2 Définition des tailles maximale des données.....	31
7.5.2.3 Définition du nombre d'occurrences	31
7.6 LE CONTENU DES DONNEES	32
7.6.1 <i>Contenu des index.....</i>	32
7.6.2 <i>Règles et formalisme pour la saisie des index</i>	33
7.7 LE FICHIER DE TRANSFERT	34
7.7.1 <i>Dénomination du fichier</i>	35
7.7.2 <i>Structure générale du fichier</i>	35
7.7.3 <i>Format du fichier.....</i>	35

Rapport final d'étude

7.7.3.1	La norme XML	36
7.7.3.2	La norme archivistique EAD.....	36
7.8	ELEMENTS D'ORGANISATION.....	37
8.	IMPACT DE LA NORME SUR LES ARCHIVES D'ETAT	38
8.1	TRAITEMENT DES FICHIERS D'IMPORT	38
8.1.1	<i>Validation du fichier</i>	39
8.1.2	<i>Import des données dans BASIS</i>	39
8.1.2.1	Traitement du fichier.....	39
8.1.2.2	Autres traitements	40
8.1.3	<i>Contrôle des données</i>	42
8.2	CONSEQUENCES ORGANISATIONNELLES	42
9.	CONCLUSION DE L'ETUDE.....	44
10.	PROCHAINES ETAPES	46
11.	AXES DE TRAVAIL AUX ARCHIVES D'ETAT	47
11.1	DEVELOPPEMENT DE THESAURUS	47
11.2	DEVELOPPEMENT D'UN MODULE DE SAISIE DES INDEX « AU RAYON » POUR LES ARCHIVISTES D'ÉTAT.....	47
11.3	DECLARATION DE LA BASE DONNEES BASIS DANS LE CADRE DE LA LITAO	48
11.4	MISE A DISPOSITION D'UN OUTIL POUR LA MISE A JOUR DES FICHES DE VERSEMENTS D'ARCHIVES	48
11.5	COMPARATIF D'OUTILS OU BASES DE DONNEES POUR LA GESTION DOCUMENTAIRE	48

1. Résumé du document

Ce rapport final présente l'étude menée pour les Archives d'Etat entre septembre et décembre 2002 afin d'informatiser la transmission des instruments de recherche (les index) des archives versées par les services de l'administration cantonale aux Archives d'Etat, ainsi que les résultats de cette étude.

Les chapitres d'introduction rappellent le contexte de réalisation de l'étude, les besoins et contraintes des archives d'Etat, et le principe d'une transmission des index normalisée entre les systèmes d'information des services producteurs d'archives et le logiciel documentaire BASIS utilisé aux Archives d'Etat pour la gestion des fonds d'archives.

Différentes analyses ont été menées afin de produire la « norme » décrivant la structure, le contenu et le format des fichiers d'index à transmettre. Les services de l'administration genevoise ayant collaboré à cette étude sont :

- Le service des Archives d'Etat
- Le secrétariat général du département de l'Instruction Publique
- Le secrétariat général du département de l'Intérieur, de l'Agriculture et de l'Environnement
- La Direction cantonale de la Mensuration officielle
- Le Registre foncier
- Le service des archives centrales des Hôpitaux Universitaires de Genève
- Le service de la DOSID pour la Chancellerie d'Etat
- Les services Gestion des archives et Système d'information du Palais de Justice
- La Direction de la Police des Constructions pour le département de l'Aménagement, de l'Equipement et du Logement.

Pour chacune de ces analyses, un rapport descriptif présentant le contexte, les fonds d'archives gérés, les instruments de recherche papier et/ou informatiques produits par le service, et les systèmes d'information et bases de données utilisés a été produit.

A partir de la confrontation des différents systèmes, la faisabilité de la normalisation des processus et de l'outil informatique a été établie, et la rédaction d'une norme a été entreprise. Celle-ci a pour but de constituer la documentation de référence des services versants pour la transmission de fichiers d'index, et en tant que telle traite de chacun des points suivants :

- Le champ d'application de la norme
- La typologie des documents d'archives à indexer
- Les niveaux d'indexation (de description) possibles
- Les types d'index à produire en fonction du niveau d'indexation
- La structure des données (des index)
- Le contenu de l'information à renseigner dans chaque index
- Les règles et formalismes applicables à chaque index
- La dénomination, la structure et le format du fichier de transfert
- Les éléments organisationnels relatifs à la transmission des fichiers aux Archives d'Etat.

Rapport final d'étude

La norme a été établie sur la base des normes métiers et des pratiques mises en place aux Archives d'Etat : normes de description des archives (ISAD(G) et ISAAR) édictées par le Conseil International des Archives, outil de type DTD (Document Type Definition) et son dictionnaire de balises basé sur XML pour la description des instruments de recherche archivistiques, développé par la société des archivistes américains et utilisé dans le monde entier et notamment par les archives de France. Les pratiques et recommandations spécifiques à la Suisse (niveau fédéral, association des archivistes suisses, archives vaudoises) ont été mises en évidence lors de l'étude préliminaire et ont également servi de base pour établir la norme.

Les derniers chapitres du rapport présentent les conclusions générales de l'étude, les conséquences pour les Archives d'Etat et les prochains travaux à mettre en œuvre pour une mise en place opérationnelle de la norme et des bénéfices attendus.

2. But du document

Ce document constitue la synthèse de l'étude menée pour les Archives d'Etat de Genève, ci-après les AEG, dont l'objectif est détaillé dans le document intitulé cahier des charges et daté du 27 septembre 2002. Le but principal de cette étude est la réalisation d'une norme pour le transfert des index, répertoires ou inventaires des archives historiques entre les organismes publics genevois producteurs d'archives et les AEG.

Au terme de l'étude et en parallèle de la production de la norme, le présent rapport a pour but de présenter les éléments qui ont conduit à la mise en place de la norme, les contraintes induites par le développement de cette norme et les prochaines étapes du projet.

3. But de l'étude

Ce paragraphe a pour but de replacer l'étude dans son contexte, et de rappeler les principaux éléments qui ont conduit à la mise en place de la norme.

3.1 Traitement des archives aux Archives d'Etat

3.1.1 Généralités

D'une manière générale, les AEG constituent, gèrent et conservent les fonds d'archives publiques.

Dans ce mandat, le service des AEG a une double mission : assurer la préservation à long terme des archives publiques du canton de Genève et mettre à disposition du public une partie de ces archives.

Dans le cadre de sa première mission, les AEG prennent le relais des autres services de l'Administration genevoise dans la conservation des archives. Chaque juridiction, département ou service peut verser tout ou partie de ses archives historiques aux AEG, qui sont alors chargées de procéder à l'inventaire des documents transmis, à leur cotation et à l'archivage physique des archives dans des rayonnages.

Les fonds d'archives versés aux AEG se composent généralement :

- D'un ensemble de documents de même type (*Série*), regroupés en dossiers, classeurs ou registres.

Rapport final d'étude

- D'instruments de recherche de type Répertoires, Index, Inventaires ou Rôles, généralement sous forme de fichiers papier ou de livre relié, mais qui peuvent exister sous format électronique, et qui décrivent l'ensemble des documents transmis et constituent le point d'entrée vers ces documents.

Afin d'enregistrer les versements, de saisir l'inventaire des archives et d'effectuer des recherches, les collaborateurs des AEG utilisent le logiciel de gestion documentaire **Basis**. L'architecture de Basis comprend une base de données propriétaire et des outils de développement et d'administration.

3.1.2 Classification des fonds

Les archives arrivant aux AEG sont classées en fonction de leur provenance. Pour les archives publiques, la provenance des fonds est établie sur la base du modèle hiérarchique recommandé par la norme archivistique internationale ou norme ISAD(G). Les six niveaux de classement des archives aux AEG sont :

Niveau 1 : les Départements de l'Administration genevoise

Niveau 2 : les Subdivisions Administratives de chaque Département

Niveau 3 : les services de chaque subdivision, appelés Organismes Créateurs d'Archives ou OCA

Niveau 4 : la Série organique

Niveau 5 : le Dossier

Niveau 6 : la Pièce.

Les trois premiers niveaux identifient la provenance du fonds, les trois niveaux suivants identifient son contenu.

3.1.3 Réalisation des inventaires

L'une des tâches principales des AEG est la réalisation de l'inventaire des fonds d'archives, qui consiste à décrire les dossiers (ou boîtes ou classeurs ou registres) et leurs pièces, suivant l'ordre de classement physique des fonds. L'inventaire est réalisé :

- Soit par l'analyse de chaque document d'archives ;
- Soit à partir des instruments de recherche papiers remis par le service versant (répertoires, rôles, etc.) ;
- Soit à partir d'une copie papier de fichiers informatiques, ce qui oblige à la ressaisie d'informations.

Une autre possibilité consisterait à alimenter automatiquement la base de données BASIS à partir de données informatiques transmises par les services versants. Pour cela, une documentation doit être produite à l'attention des services versants décrivant les données à transmettre ainsi que le format d'échange. La norme mise en place constitue cette documentation.

3.2 Les motivations du projet

Les AEG visent plusieurs objectifs avec la mise en place de cette étude :

1 – Prendre en compte l'informatisation de la gestion documentaire

Les répertoires ou inventaires ou rôles actuellement transmis aux AEG sont sur support papier. Cette situation est amenée à évoluer dans les années à venir si l'on tient de l'informatisation croissante et généralisée de toutes les activités administratives. Aujourd'hui, les services de l'Administration cantonale ont pour la plupart mis en place des outils informatiques qui leur ont permis de se constituer des bases de données répertoriant les documents et contenant les éléments de description de ces documents. C'est la réutilisation des données saisies dans ces bases qui est visée dans ce projet.

2 – Agir de manière proactive dans la gestion des archives

L'entrée en vigueur en mars 2002 de la Loi sur l'Information du public et l'accès aux Documents (LIPAD) accroît la nécessité pour les institutions de mettre en place les moyens et les outils utiles à la structuration et à l'identification de leurs archives. Cette nécessité va certainement se traduire par le développement d'applications informatiques d'indexation des documents, index qui entreront dans le cadre des données à récupérer. La mise en place d'un standard serait donc souhaitable avant le développement « anarchique » de ces nouvelles applications.

3 – Améliorer, accélérer et faciliter les tâches d'archivage

Les archivistes des AEG ont pour mission d'inventorier les fonds qui leur sont versés. Cet inventaire est fastidieux en raison du volume d'archives entrant chaque année. Les AEG emploient régulièrement des stagiaires pour effectuer les tâches de saisie. Même ainsi, les objectifs fixés annuellement pour la saisie des inventaires ne sont pas atteints. En automatisant une partie de la saisie, il semble envisageable de réduire ou de supprimer la part non productive du travail qui consiste à ressaisir des données qui existent déjà au format électronique dans les applications utilisées en amont de la chaîne d'archivage.

4 - Respecter le règlement d'application de la loi sur les archives publiques

La nouvelle loi sur la gestion des archives publiques donne notamment pour mission aux AEG de s'investir dans la problématique de l'archivage des données informatiques. Cet investissement doit passer par la production d'outils (de directives) à l'attention des autres services d'Archives de l'Administration. La norme produite constituerait donc une de ces directives.

3.3 Conséquences pour les services producteurs d'archives

Pour les services versants, la mise en place d'une norme de transmission informatique de données aux AEG ne devrait pas induire d'inconvénients majeurs.

Rapport final d'étude

D'une part la norme a été établie en prenant en compte les contraintes de huit services versants de l'Administration cantonale, échantillon représentatif des contextes et modes de gestion archivistique en place dans l'institution publique. Le standard d'indexation (procédure, données, structure, outil) a été construit en tenant compte des exigences et pratiques de ces services.

Deuxièmement, l'un des buts premiers du standard est de fournir les spécifications techniques à respecter lors de l'informatisation future d'un service ou de la migration vers un nouveau système. La norme ne vise pas à demander une adaptation immédiate des systèmes existants, mais à guider les services lors des projets futurs d'évolution des systèmes d'information. Les directives pourront donc être appliquées à court, moyen ou à long terme.

Dans la majorité des services interviewés durant l'étude, et qui disposent déjà des systèmes d'information permettant d'extraire les index depuis une base de données, les moyens informatiques et organisationnels sont généralement disponibles pour permettre la réalisation des traitements d'export conformément à la norme. L'utilisation croissante de technologies ouvertes et récentes facilite généralement les traitements informatiques (modules d'export intégrés, compatibilité avec le langage XML, etc.).

Enfin, pour les services versants qui ne disposent pas aujourd'hui des outils informatiques utiles pour répertorier leurs fonds d'archives historiques, une étude devrait être menée suite au présent projet. Elle consisterait à définir puis à mettre à disposition de ces services le(s) outil(s) informatique(s) adéquat(s) leur permettant de produire les index informatiques conformes au standard défini par les AEG. Les AEG accompagneraient donc les services versants qui le désirent en mettant à leur disposition les outils pour respecter la norme d'indexation.

4. Principe de la norme

Il s'agit de mettre en place un échange de données standardisé entre tout service producteur d'archives d'une part (Système A), et les Archives d'Etat de Genève d'autre part (Système B).

Les données échangées doivent permettre d'identifier et de décrire les fonds d'archives. Ces données sont généralement contenues dans les instruments de recherche qui accompagnent les archives. Dans la suite du document, nous utiliserons un vocabulaire plus générique. La locution « **fichier d'index** » désignera tout instrument de recherche informatisé décrivant un ensemble d'unités documentaires, les index constituant les éléments de description (titre, date, type, etc.) des documents.

Le schéma qui suit présente le concept général du projet.

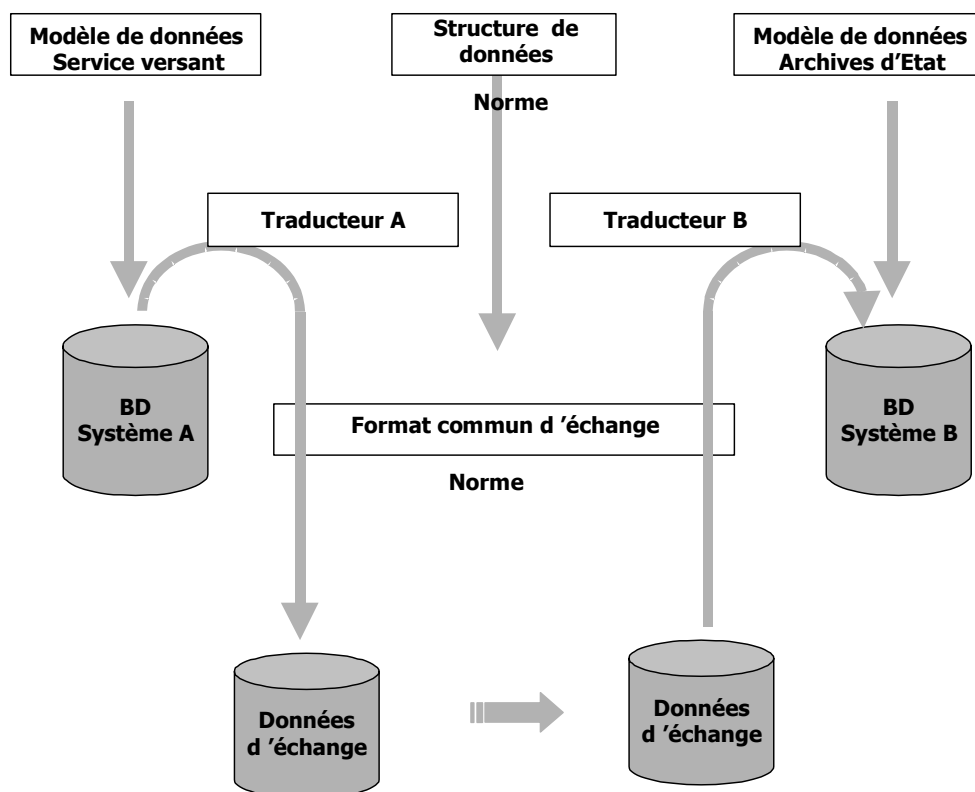


Figure 1: Transfert de données entre deux bases de données à l'aide d'une structure commune de données décrite avec un langage commun de description des données.

Le projet vise à mettre en place les paramètres de transfert réglant l'échange de données des services versants vers les Archives d'Etat. Le principe du transfert est le suivant :

Le service versant produit le fichier d'index à partir d'un export depuis sa base de données source (A).

Rapport final d'étude

Le fichier d'index contient les données d'échange. La structure des données et le format du fichier sont spécifiés par la norme d'indexation.

A la réception du fichier d'index, le service des Archives d'Etat réalise l'import des données dans la base de données cible (B).

Le principe ainsi énoncé part évidemment de l'hypothèse que le service versant dispose d'un outil informatique (base de données A) stockant les données utiles à l'indexation des fonds d'archives.

5. Rappel du déroulement de l'étude

5.1 Introduction

Le besoin identifié dans le cahier des charges consistait à étudier les possibilités de transfert informatique des instruments de recherche associés aux archives versées aux AEG, les directives produites étant basées sur les outils informatiques existants, des adaptations à ces outils et éventuellement de nouveaux outils informatiques.

La construction du nouveau standard devait également respecter un certain nombre d'exigences métiers et normatives, identifiées tout au long de l'étude.

Le résultat produit au terme de l'étude détaillée consiste en un **standard normalisé** à l'attention de l'ensemble des organismes producteurs d'archives dans l'institution publique genevoise. Ce standard a été défini en étudiant les besoins et exigences du service des AEG et de services producteurs d'archives sélectionnés par les AEG.

Le responsable du projet désigné par les Archives d'Etat, et notre principal interlocuteur pour la réalisation de l'étude, a été Monsieur Roger Rosset.

5.2 Les acteurs de l'étude

Les départements, juridictions ou services de l'administration genevoise ayant été interviewés durant cette étude sont :

- Les Archives d'Etat
- La Chancellerie d'Etat
- Le Pouvoir judiciaire
- Le secrétariat général du département de l'Instruction Publique
- Le secrétariat général du département de l'Intérieur, de l'Agriculture et de l'Environnement
- Le service du Registre Foncier
- La Direction cantonale de la Mensuration Officielle
- Le service des Systèmes d'Information et de Géomatique
- Les archives des Hôpitaux Universitaires de Genève
- La Direction de la Police des Constructions (DAEL).

Rapport final d'étude

Les autres organismes, services ou sociétés ayant apporté leur contribution à la réalisation de cette étude sont :

- Les archives cantonales vaudoises
- Les archives de la ville de Lausanne
- La société DOMAIN SA, fournisseur du logiciel de gestion documentaire BASIS
- L'Observatoire technologique de l'Etat de Genève.

5.3 Les résultats et livrables intermédiaires

La finalité du projet est d'arriver à la mise en place, dans le canton de Genève, d'un transfert automatique des éléments de description (ou *instruments de recherche*) des archives historiques. Différentes étapes ont été nécessaires pour aboutir à la réalisation de la première phase du projet, à savoir la production d'une norme pour le transfert de données, destinée aux différents services publics producteurs d'archives. Ces étapes intermédiaires ont permis de recueillir les éléments d'information nécessaires à la construction du standard.

Les paragraphes qui suivent présentent brièvement chaque étude intermédiaire menée. Le détail de chaque étude et de ses résultats est disponible dans les livrables intermédiaires dont la liste est fournie dans le paragraphe 5.3.4.

5.3.1 Etude préliminaire

Objectif :

L'analyse préliminaire a eu pour but de positionner l'étude dans son contexte et d'analyser les différentes exigences et contraintes inhérentes à la mise en place d'un standard de transmission des données entre les services versants et les AEG. Elle a consisté à étudier les normes ou études en relation avec la gestion des archives au niveau national ou international. Les éléments et recommandations contenus dans le rapport d'étude préliminaire ont servi de base de travail pour la définition du standard.

Travaux réalisés :

- étude du contexte législatif en vigueur (LAP, LIPAD).
- Etude du contexte normatif (norme archivistique ISAD(G)).
- Consultation des travaux de l'Association des Archivistes Suisses, et notamment de l'étude fédérale intitulée « Stratégie globale pour la conservation à long terme des documents électroniques en Suisse », réalisée entre 2000 et 2002.
- Consultation des archives vaudoises (archives cantonales et archives de la ville de Lausanne).

Rapport final d'étudeRésultats :

A partir des différentes sources d'information consultées, des éléments importants pour la mise en place de ce projet ont pu être identifiés :

- Des règles de description archivistique pour les éléments de description à inclure dans le standard,
- Des orientations techniques pour l'extraction des données sources,
- Des points-clés sur lesquels une réflexion doit être menée (adéquation, contrôle et traitement des données dans le cadre d'un transfert automatisé).

5.3.2 Analyse pour la définition du standard du point de vue des Archives d'Etat

Objectif :

Cette analyse a eu pour but de recenser l'ensemble des informations utiles à la constitution du standard du point de vue des Archives d'Etat de Genève, à savoir :

- La description de l'outil de gestion documentaire Basis, utilisé pour la saisie des inventaires des documents d'archives,
- La description des processus organisationnels et informatiques d'identification des fonds versés aux Archives d'Etat,
- La description du modèle de données pour la gestion documentaire,
- L'identification des données utiles dans le cadre d'une alimentation automatisée de la base de données,
- La description des contraintes liées à la mise en place du standard du point de vue des AEG.

Travaux réalisés :

- Etude du système d'information BASIS (architecture, base de données)
- Définition des éléments de description « standards » requis aux Archives d'Etat (données, type de données, règles de saisie des données).
- Définition des supports informatiques possibles (type de fichiers).

Résultats :

Les éléments importants qui ont pu être définis lors de cette analyse sont :

- Les caractéristiques de la base de données BASIS et la structure des données gérées,
- Les modes et formats d'import de données dans BASIS,
- Les niveaux de classification d'un fonds dans BASIS,
- La définition des contraintes techniques et fonctionnelles liées à l'intégration automatisée de données dans BASIS,
- Une description générale des tables DOCUMENT et CONTENU, tables cibles pour l'import de données dans le cadre du présent projet,
- L'identification et la description détaillée des champs susceptibles d'être inclus dans la norme.

5.3.3 Etude de la gestion informatique des archives dans huit services producteurs d'archives

Objectif :

Ces analyses ont eu pour but de recenser l'ensemble des informations utiles à la constitution du standard du point de vue de chaque service interviewé dans le cadre de ce projet. Les huit entités administratives (service ou juridiction) ont été choisies par les Archives d'Etat pour l'intérêt que peut représenter l'informatisation du transfert des données décrivant les archives versées par ces entités aux Archives d'Etat. Les archives historiques étudiées sont les suivantes :

- Les procédures judiciaires du Palais de Justice
- Les Registres décennaux du Conseil d'Etat, pour la Chancellerie
- Les dossiers de requêtes en autorisation de construire, pour le département de l'Aménagement, de l'Equipeement et du Logement
- Les dossiers médicaux pour les Hôpitaux Universitaires de Genève
- Les dossiers du secrétariat général du département de l'Intérieur, de l'Agriculture et de l'Environnement
- Les inventaires des dossiers et fonds d'archives gérés par le secrétariat général du département de l'Instruction Publique
- Les archives cadastrales de la Direction cantonale de la Mensuration Officielle
- Les archives du Registre Foncier.

Travaux réalisés :

Pour chaque service responsable de la création ou de l'archivage de ces documents, les informations utiles à la constitution du standard ont été rassemblées, à savoir :

- La description des fonds d'archives et des instruments de recherche relatifs à ces fonds,
- La description des instruments de recherche informatisés disponibles dans les services concernés,
- La description des bases de données en place permettant de produire un index des fonds d'archives historiques à verser à court, moyen ou long terme aux Archives d'Etat,
- La description des tables et des données susceptibles d'être extraites de ces bases de données,
- La description des contraintes fonctionnelles ou techniques dans les services pour s'adapter à la future norme produite.

Résultats :

Le détail des résultats de chacune de ces études est fourni dans les différents rapports d'analyse. Les principaux éléments utiles à la construction d'un standard et identifiés dans ces rapports sont :

- L'existence ou la non-existence des éléments de descriptions (index) des documents d'archives au format informatique ;

Rapport final d'étude

- Les caractéristiques (nom, base de données, système d'exploitation) du système à partir duquel les données pourront être extraites ;
- Les données informatiques permettant d'identifier un document ou un dossier et leur structure ;
- Les contraintes, remarques, restrictions particulières relatives à la transmission informatique de données dans chaque service.

5.3.4 Récapitulatif des livrables intermédiaires

L'ensemble des livrables transmis aux membres de l'équipe de projet et au comité de pilotage est donné ci-après, selon le formalisme suivant : nom du document, référence du document, date de la dernière version, auteur.

1. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'étude préliminaire**, NM-EP-AEG-1110-02, 11.10.2002, N. Marquet
2. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour les AEG**, NM-AAEG-0511-02, 05.11.2002, N. Marquet
3. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour la Chancellerie d'Etat**, NM-CE-AEG-2810-02, 28.10.2002, N. Marquet
4. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour le Pouvoir Judiciaire**, NM-PJ-AEG-2810-02, 29.10.2002, N. Marquet
5. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour le DAEL**, NM-DAEL-AEG-0611-02, 05.11.2002, N. Marquet
6. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour le DIP**, NM-DIP-AEG-1211-02, 12.11.2002, N. Marquet
7. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour le DASS**, NM-HUG-AEG-2511-02, 25.11.2002, N. Marquet
8. Normalisation du transfert informatique des index et répertoires des archives historiques / **Rapport d'analyse pour le DIAE**, NM-DIAE-AEG-2811-02, 28.11.2002, N. Marquet.

L'ensemble des analyses et résultats contenus dans ces rapports d'étude intermédiaires fait partie intégrante du présent rapport final d'étude, en ce sens que le détail n'est pas repris dans le présent document.

5.4 Construction et rédaction du standard

Objectif :

A partir des résultats des analyses précédentes, l'étape suivante a consisté à définir une norme issue de la synthèse et de la confrontation des systèmes d'information en place. Cette norme, archivistique et surtout informatique, s'appuie sur les caractéristiques et les contraintes des systèmes existants et sur les normes métiers en vigueur.

Travaux réalisés :

- Confrontation des différents systèmes.
- Définition du standard informatique (champs, structure des données, support informatique).
- Définition du standard archivistique (contenu et règles de saisie des index).
- Définition de la procédure de transfert (méthode, organisation).
- Rédaction de la norme de transfert informatique des index des archives historiques entre services producteurs et Archives d'Etat.

Résultats :

Les résultats de ces travaux sont décrits dans les paragraphes suivants :

- Confrontation des analyses et faisabilité de la norme
- Choix informatiques
- Choix archivistiques
- Choix organisationnels
- Conséquences pour les Archives d'Etat
- Etapes nécessaires à la mise en œuvre opérationnelle de la norme.

6. Conclusion générale des analyses intermédiaires

Ce paragraphe présente les conclusions obtenues au terme des différentes analyses intermédiaires, avant la construction de la norme proprement dite.

6.1 Etude menée aux Archives d'Etat

Cette étude a permis de mettre en évidence les besoins des Archives d'Etat, les contraintes liées au système d'information actuellement utilisé pour la gestion des fonds d'archives, et les contraintes liées au mode de travail des collaborateurs.

Le principal travail a consisté à décrire le logiciel documentaire BASIS et sa base de données relationnelle. Les tables et données utiles dans le cadre de la mise en place de la norme, c'est-à-dire les tables et données à alimenter lors du traitement d'import, ont été identifiées et décrites. Les règles de saisie et les formalismes utilisés ont été également donnés lorsque cela était nécessaire.

Un certain nombre de points-clés, liés à des contraintes ou à des choix techniques ou organisationnels, ont été mis en évidence et ont été utilisés pour la construction de la norme.

Les procédures, outils et technologies pour l'import de données dans BASIS ont également été décrits. Suite à la première séance d'avancement de projet qui s'est déroulée le 6 novembre 2002, il a notamment été décidé que le choix du format d'import se porterait sur un standard ouvert, le XML.

Au terme de cette étude, les premières conclusions ont pu être dégagées :

- Un besoin d'automatisation réel, pour faire face à des problématiques légales et organisationnelles
- Un système d'information (BASIS) apte à l'import de données en masse
- Un système d'information pérenne, qui devrait être conservé dans les années à venir
- Des besoins « limités » en terme de nombre de données à importer
- Un mode de gestion respectant les standards archivistiques internationaux, donc sur lequel les directives émises pourront s'appuyer
- Une nécessaire adaptation future face à la mise en place de la norme, d'un point de vue informatique, organisationnel ou fonctionnel.

6.2 Etudes menées dans les services versants

Les études menées dans huit services producteurs d'archives ont constitué une part importante des travaux réalisés. Ces études visaient plusieurs objectifs :

- Informer l'administration cantonale du projet des Archives d'Etat.
- Etudier la typologie des documents d'archives versés aux Archives d'Etat.
- Etudier le niveau d'informatisation des services pour les tâches liées à la gestion documentaire.
- Caractériser les bases de données sources, à partir desquelles les fichiers d'index pourraient être créés.
- Identifier les données potentiellement intéressantes pour la description archivistique, donner leur structure et les informations renseignées dans les services versants (contenu et formalisme des index potentiels).

Plusieurs conclusions générales ont été tirées suite à ces huit études.

1. **Le contexte général est propice** à la mise en place de ce type de projet : les services interviewés sont favorables, voire parfois demandeurs, de directives émanant des archives d'Etat face à l'informatisation de leurs outils de gestion et aux problématiques d'archivage des données.

2. La gestion documentaire s'est informatisée au cours des deux dernières décennies, et sauf exceptions, la **production d'index informatiques est envisageable**, au cas par cas pour les Archives produites entre 1980 et 1990, de manière généralisée depuis les années 1990-1995.

3. Il existe un **réel besoin**, à court ou moyen terme, **de directives** concernant la constitution d'index informatiques des documents d'archives dans certains services.

- Certains services ont déjà identifié la nécessité d'extraire des données à partir de bases de données « anciennes » (Basis pour le Palais de Justice, Oracle pour la Chancellerie), soit pour épurer la base de données, soit pour produire informatiquement des répertoires et index en remplacement des anciennes versions papier (respectivement les rôles et les registres décennaux).
- Certains services envisagent la migration de leur système d'information (Police des Constructions) ou la création d'une application de saisie d'index (Mensuration officielle) et sont demandeurs de directives archivistiques impactant sur la définition de ces nouveaux systèmes.

4. **Les besoins** tels qu'ils ont été identifiés lors de l'analyse aux AEG **pourraient être couverts à partir de la majorité des systèmes sources** étudiés : disponibilité des index dans les bases de données, extraction possible des données au format XML, etc.

Rapport final d'étude

5. Il existe quelques services pour lesquels le **développement d'un applicatif dédié à la saisie des index est nécessaire**, soit parce que l'outil utilisé actuellement ne permet pas l'export de données (Ex : Microsoft Word pour le département de l'Instruction Publique), soit parce que le service envisage l'indexation informatique d'un fonds d'archives historiques avant son versement aux AEG (Ex : indexation des archives cadastrales par la Direction de la Mensuration officielle). Deux options peuvent être envisagées pour ce développement :

- Soit chaque service non informatisé met en place ses propres outils informatiques conformément aux spécifications énoncées dans la norme ;
- Soit une application « universelle » est proposée par les Archives d'Etat et déployée de manière transversale dans l'ensemble de l'administration publique cantonale.

6. **Deux niveaux d'indexation** peuvent être envisagés en fonction des outils et des procédures de gestion mises en place dans les services versants :

- Indexation des Articles (dossier, classeur, registre) uniquement : Palais de Justice, Hôpitaux Universitaires, Secrétariat général du DIAE et Police des Constructions ;
- Indexation au niveau des pièces constituant les Articles : Chancellerie d'Etat, Secrétariat général du DIAE et Police des Constructions (après migration des outils informatiques actuels pour le DIAE et le DAEL, projets en cours).

6.3 Conclusion quant à la faisabilité de la norme

Au terme des analyses précédentes, et au moment de confronter les différents résultats obtenus, il s'avère que la mise en place de recommandations en terme de production de données informatiques est sinon nécessaire, en tout cas souhaitable, pour chacune des parties concernées.

Pour les services versants, les principales raisons ont été évoquées dans le paragraphe ci-dessus et détaillées dans chacun des rapports d'analyse.

Pour les Archives d'Etat, le besoin est identifié depuis quelques années. A l'origine, il avait été envisagé de concentrer l'étude sur quelques cas particuliers « prioritaires » dans la transmission d'index (Palais de Justice et Chancellerie d'Etat notamment). Lors du lancement effectif de l'étude au mois de septembre 2002, le comité de pilotage, constitué de M. Roger Rosset pour les Archives d'Etat, de M. François Mumenthaler pour le département DIAE et de M. Patrick Morand représentant le CTI, a jugé plus judicieux d'élargir le cadre de l'étude de manière à définir des bases communes pour l'ensemble des services versants et à proposer des outils normés à l'ensemble de l'administration publique. L'homogénéisation des pratiques d'une part, et un souci de rentabilisation de l'étude menée d'autre part, ont été à l'origine de la globalisation de l'étude.

Pour conclure quant à la faisabilité de la norme, deux points doivent être pris en considération.

Rapport final d'étude**1. La faisabilité technique**

Après étude des technologies, applications et bases de données utilisées dans l'échantillon des services de l'Administration, il apparaît qu'aucun obstacle majeur n'entrave la faisabilité technique du présent projet. La transmission d'index est possible à court terme dans les services versants qui disposent d'une gestion documentaire informatisée supportée par une base de données relationnelle (éventuellement d'autres outils gérant des données structurées).

2. La faisabilité fonctionnelle et organisationnelle

D'un point de vue fonctionnel, la question de la faisabilité d'une norme UNIQUE, utilisable par l'ensemble des services versants, s'est posée en cours d'étude. Les analyses des huit services « pilotes » ont en effet montré les spécificités de chaque service en terme de contenu et de typologie des archives produites, en terme de données gérées d'un service à l'autre, en terme de mode de gestion des mots-clés, etc. A partir de cette constatation, il s'est avéré que les directives énoncées, pour pouvoir être appliquées par tous, devaient rester assez générales (Ex : désignation d'une unité documentaire). En respectant cette contrainte, l'énoncé de règles standardisées est possible.

D'un point de vue organisationnel, les services qui disposent des systèmes d'information adéquats pour produire les fichiers d'index disposent généralement en support des moyens organisationnels (service informatique) et des compétences permettant de se conformer aux directives des Archives d'Etat. Dans les autres services, il faudra déterminer les moyens / outils à instaurer pour mettre en application les directives des Archives d'Etat.

7. Construction de la norme

7.1 Préambule

1. Afin de faciliter la compréhension de la norme et étant donné le large public visé, une terminologie neutre a été utilisée. Les termes ou locutions tels que « Article », « Pièce » ou « instrument de recherche » ont remplacé des appellations plus spécifiques utilisées dans le monde archivistique. De même, le terme « Index » a été choisi pour désigner tout élément de description (titre, date, mot-clé) documentaire alors qu'au sens archivistique pur, le terme « Index » a une autre signification.
2. Un glossaire a été inclus dans la norme afin de faciliter la compréhension des concepts archivistiques et informatiques utilisés dans le document.

7.2 Introduction

A partir de la confrontation des différentes analyses précédentes, et après avoir conclu quant à la faisabilité d'une norme de transmission des fichiers d'index des archives historiques aux Archives d'Etat, l'étape suivante a consisté à mettre en place les différents éléments de la norme.

Afin de permettre aux différents services producteurs d'archives de l'administration publique du canton de Genève de mettre en place les traitements informatiques permettant d'extraire les données adéquates de leurs systèmes d'information en terme de contenu, de structure et de format, le résultat de l'étude doit fournir une documentation décrivant de manière exhaustive les éléments suivants :

- Le champ d'application de la norme
- La typologie des documents d'archives à indexer
- Les niveaux d'indexation (de description) possibles
- Les types d'index à produire en fonction du niveau d'indexation
- La structure de ces données (index)
- Le contenu de l'information à renseigner dans chaque index
- Les règles et formalismes applicables à chaque index
- La dénomination, la structure et le format du fichier de transfert
- Les éléments organisationnels relatifs à la transmission des fichiers aux Archives d'Etat.

Les paragraphes qui suivent présentent pour chacun de ces éléments la démarche d'analyse adoptée, les références utilisées (procédures, outils), les choix réalisés pour la définition de la norme.

7.3 Le champ d'application de la norme

Le champ d'application de la norme a été défini dès la rédaction du cahier des charges de l'étude. Les Archives d'Etat sont chargées de la gestion de tous les documents officiels issus des trois pouvoirs de l'Etat et des grands établissements rattachés à l'Etat (hôpitaux Universitaires de Genève, Services Industriels de Genève, Transports publics genevois, etc.). La norme s'adresse donc aux services issus des départements et juridictions de l'administration publique genevoise et aux différents services d'établissements de droit public.

Dans un premier temps, d'autres organismes créateurs d'archives (ex : études de notaire) et tous les fonds d'origine privée sont exclus de cette norme.

Le champ d'application est stipulé dans les chapitres Introduction (§2) et Principes généraux (§4.1) de la norme.

7.4 La description des documents d'archives

Afin de préciser le périmètre de la norme, deux éléments essentiels doivent être spécifiés :

- Une définition de la notion d'INDEX
- Une présentation des niveaux de description (c'est-à-dire des niveaux d'indexation) possibles.

7.4.1 Typologie des documents à indexer

Lorsque des archives historiques (ou fonds d'archives) sont versées aux Archives d'Etat, un instrument de recherche accompagne généralement le fonds. Cet instrument de recherche est élaboré par le service versant de manière logique (alphabétique ou chronologique) afin de décrire brièvement chaque document d'archives. Chaque document est généralement doté d'une cote, mentionné sur l'instrument de recherche, qui renvoie à la localisation physique du document à des fins de consultation. Dans le langage archivistique, on distingue différents types d'instruments de recherche en fonction de leur provenance, de leur contenu ou de l'ordonnancement des documents au sein de l'instrument de recherche. Dans la norme, tout instrument de recherche quel qu'il soit, permettant de décrire une liste d'unités documentaires, peut constituer un index.

La norme fournit les lignes directrices pour la transmission des instruments de recherche informatisés ou qui peuvent être informatisés. Ces index informatiques peuvent lister tout type de pièce produite par un service dans le cadre de ses activités :

- des registres
- des documents (pièce comptable, procès-verbal, mémoire, rapport, loi, etc.)
- des correspondances
- des livres ou des publications
- des procédures
- des dossiers
- des annexes, etc.

Rapport final d'étude

La norme ne spécifiant pas les types de documents à indexer, tout document d'archives, quel que soit sa nature, peut entrer dans le champ d'application de la présente norme.

Une autre notion, celle du *support* du document d'archives est également traitée dans la norme. Tous les documents d'archives, quel que soit leur support physique (photo, papier, microfiche, microfilm, etc.), peuvent faire l'objet d'une indexation informatique.

7.4.2 Les niveaux de description

La référence : la norme ISAD(G)

La norme générale et internationale de description archivistique ISA(G), proposée par le CIA (Comité International des Archives), fournit les lignes directrices pour la description archivistique. Elle doit être utilisée comme point de départ pour le développement de normes nationales.

La description archivistique est basée sur un modèle hiérarchique, allant du général au particulier (du groupe de document vers chaque unité documentaire). Il existe des niveaux de description avec différents degrés de détail appropriés à chaque niveau de classement. Un fonds peut être décrit dans sa globalité dans une description simple, ou représenté dans ses différentes parties à différents niveaux de description. Ainsi, il peut y avoir une description au niveau du fonds, une description au niveau de la série organique, une description au niveau du dossier et une description au niveau de chaque pièce. Toutes ces notions archivistiques sont définies dans le glossaire de la norme (§3).

Le mode de classification adopté aux Archives d'Etat est basé sur les principes de l'ISAG(G). Les quatre règles fondamentales fournies par la même norme et qui doivent être respectées lors de la description à plusieurs niveaux sont :

- Description du général au particulier
- Informations adaptées au niveau de description
- Lien entre les niveaux de description
- Non – répétition des informations entre plusieurs niveaux.

Ces références et règles archivistiques sont rappelées dans la norme (§4.2).

Les niveaux de description : pratique des Archives d'Etat et résultat de la confrontation des analyses

Les archives d'Etat ont mis en place un système de description hiérarchique basé sur 6 niveaux :

Niveau 1 : Département	<i>Provenance du fonds</i>
Niveau 2 : Subdivision administrative	
Niveau 3 : Service versant	
Niveau 4 : Série organique	<i>Description du fonds</i>
Niveau 5 : Document	
Niveau 6 : Contenu	

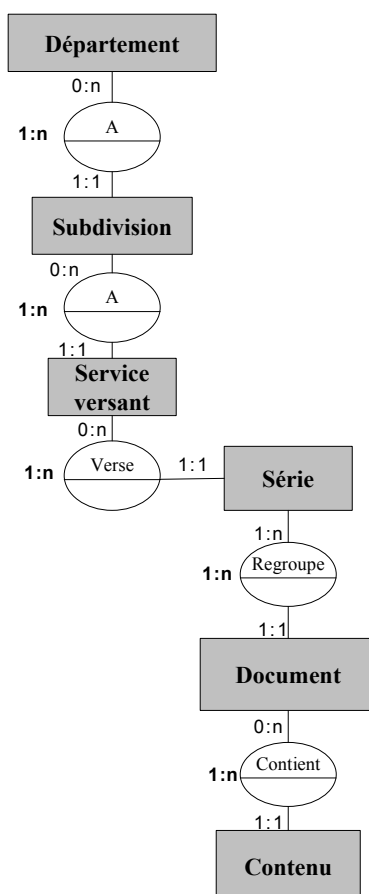
Rapport final d'étude

Les 3 premiers niveaux, qui permettent de décrire la provenance du fonds, sont hors étude.

Les 3 derniers niveaux qui permettent de décrire le contenu du fonds, sont ceux qui sont potentiellement intéressants pour l'étude.

- Le niveau Série permet de décrire un ensemble de documents de même provenance, de même support ou de même type.
- Le niveau Document permet de décrire un Article (dossier, registre, classeur), qui peut généralement se décomposer lui-même en sous unités documentaires.
- Le niveau Contenu permet de décrire une Pièce (lettre, rapport, page, etc.) ; il s'agit de la plus petite unité documentaire pouvant être décrite.

Dans le logiciel de gestion documentaire BASIS, le modèle de données relationnel qui a été défini pour supporter cette description à plusieurs niveaux est le suivant :



Lors de l'analyse menée aux Archives d'Etat, les hypothèses suivantes ont été validées :

- Les niveaux de description permettant de décrire la provenance du fonds (Département, Subdivision et Service versant) sont supposés renseignés dans BASIS avant l'import du fichier d'index.

Rapport final d'étude

- Un fichier transmis doit se référer à une et à une seule série. L'entité SERIE doit être définie avant l'import des données dans BASIS, via l'un des deux traitements suivants :
 1. L'enregistrement correspondant dans la table SERIE est saisi préalablement par les AEG.
 2. Lors de la procédure d'import, un traitement permet de créer la série concernée.
- Chaque élément de CONTENU doit se référer à un et un seul document. Les données transmises doivent permettre d'identifier les différents documents et leurs contenus respectifs (différenciation par nom ou numéro).

A partir de ces hypothèses, on en conclut que les fichiers transmis contiendront uniquement des données pouvant alimenter les niveaux 5 et 6 (Document et Contenu), les niveaux supérieurs étant alimentés préalablement par les Archives d'Etat via un traitement à définir ultérieurement.

Cette hypothèse a été confirmée lors de l'étude des différents services versants. Les différentes analyses menées ont permis de mettre en évidence deux niveaux d'indexation possibles :

- une indexation au niveau d'un Dossier ou d'un Registre, relatif à une même affaire ou à un même sujet, chaque dossier ou registre étant constitué de plusieurs pièces.

Ex : dossier relatif à une affaire judiciaire
Ex : dossier relatif à une requête en autorisation de construire
Ex : dossier relatif à une affaire traitée par le Secrétariat général du DIAE

- une indexation au niveau d'une Pièce, correspondant à une sous unité de description d'un dossier ou d'un registre.

Ex : page d'un registre des Affaires du Conseil d'Etat
Ex : pièce d'un dossier relatif à une requête en autorisation de construire
Ex : Lettre d'un dossier relatif à une affaire traitée par le Secrétariat général du DIAE.

A partir de ces éléments, les bases de construction de la norme sont définies :

- Les données contenues dans le fichier d'index pourront alimenter la table DOCUMENT seule (indexation au niveau d'un article), ou bien la table DOCUMENT et la table CONTENU (indexation des pièces constituant les articles avec indexation des articles concernés) de la base de données utilisée aux Archives d'Etat.

Les niveaux de description et donc d'indexation possibles sont donnés dans le chapitre Indexation de la norme (§4.3) et sont repris dans le chapitre Structure générale du fichier (§5.1.1).

7.5 La définition des index

L'un des éléments importants dans le processus de normalisation de la transmission informatique de fichiers d'index est la présentation des éléments de description archivistique d'un Article ou d'une Pièce. Si l'on envisage d'automatiser la récupération d'index pour alimenter la base de données de saisie des inventaires aux Archives d'Etat, il faut s'assurer que la description des documents réalisée par les services versants sera en parfaite adéquation avec les règles de description appliquées par les archivistes des AEG. Pour cela, il faut donner aux services versants les éléments d'information nécessaires à la compréhension du processus de description d'un fonds d'archives et décrire les règles qui s'appliquent dans la profession.

Les principales règles de description archivistiques sont fournies par la norme ISAD(G). Les règles générales qui régissent la description hiérarchique d'un fonds ont déjà été évoquées et sont reprises dans les Principes généraux dans la norme. La norme ISAD(G) va plus loin en définissant 26 éléments de description qui peuvent être combinés entre eux pour constituer la description d'une entité archivistique. La structure et le contenu de l'information dans chacun de ces éléments peuvent ensuite être formulés en accord avec les règles nationales applicables.

Ce paragraphe a pour but de présenter les éléments de description retenus (parmi les 26 de la norme ISAD(G)) pour constituer les index ainsi que la structure de ces index.

Le prochain paragraphe constituera le complément de ce chapitre en décrivant pour chaque élément de description les règles relatives à la saisie de l'information.

7.5.1 Les données

Une vue d'ensemble des données définies pour chaque niveau de description est présentée dans les tableaux des chapitres 5.2.2 et 5.2.3. de la norme.

7.5.1.1 Définition des données

Le paragraphe précédent a déterminé les deux niveaux d'indexation possible. La détermination des index à fournir pour chaque niveau d'indexation a été réalisée principalement à partir de l'étude menée aux Archives d'Etat, en suivant la démarche suivante :

1. Description générale des tables DOCUMENT et CONTENU correspondant aux deux niveaux de description retenus. Cette description générale a permis de décrire l'ensemble des données définies dans chaque table ainsi que leur structure (taille, type, donnée obligatoire, nombre d'occurrences).
2. Tri des données décrites précédemment, qui a permis de mettre en évidence les champs susceptibles d'être inclus dans la norme. Ce tri a été effectué par élimination des champs spécifiques à la gestion interne aux Archives d'Etat ou spécifiques au système d'information BASIS.

Rapport final d'étude

3. Les données conservées après ce filtrage ont subi deux analyses :
 - a. Tri des données par source d'information (données pouvant être renseignées par le service versant, données créées automatiquement par le système, données à traiter en lot par les Archives d'Etat lors de l'import du fichier de données).
 - b. Description pour chacune des données des règles de saisie et du formalisme en vigueur aux Archives d'Etat.
4. La confrontation de cette structure de données et des « index » disponibles dans les bases de données des services versants interviewés a permis de finaliser la définition des index à inclure dans la norme.

Un minimum de données sont requises afin d'encourager la production des fichiers d'index informatiques sans multiplier les contraintes pour les services versants. Dans la norme telle qu'elle est définie aujourd'hui, seuls un intitulé et une date de création peuvent suffire à l'indexation d'un ensemble documentaire. Il est cependant demandé de fournir le maximum d'information en fonction des données disponibles dans les bases de données sources. Les index tels qu'ils sont requis dans la norme seront produits par les services sources :

- Soit directement par extraction de la donnée concernée
- Soit après extraction et traitement de(s) donnée(s) concernée(s) (ex : concaténation de données, conversion du format, etc.).

7.5.1.2 Correspondance entre les index définis dans la norme et les données dans BASIS

Les noms des index ont été normalisés dans le document destiné aux services producteurs d'archives. Les tableaux qui suivent donnent la concordance avec les données telles qu'elles sont définies dans BASIS, afin de faciliter les développements ultérieurs nécessaires au contrôle et à l'import des données dans la base.

Tableau de concordance ARTICLE (terminologie norme) – DOCUMENT (terminologie Basis) :

Nom des éléments descripteurs définis dans la norme	Nom des champs définis dans la table Document de BASIS
ART_CODE	SAGCC
ART_COTE	COTE
ART_DATECOMM	DATCOMM
ART_DATECREATION	DATDEB
ART_DATEFIN	DATFIN
ART_DATEMICROFILM	DATMICR
ART_DATEXTREME	DATEXT

Rapport final d'étude

ART_DESIGNATION	DESI
ART_DESSOMMAIRE	DESCRSOM
ART_LANGUE	LANGUE
ART_MCLIEU	MCLIEU
ART_MCNOM	MCNOM
ART_THEME	MCCOM
ART_NOTE	OBS
ART_PIECE	PIECES
ART_RECHERCHE	INSTRECH
ART_SUPPORT	SUPPORT

Tableau de concordance PIECE (terminologie norme) - CONTENU (terminologie Basis) :

Nom des éléments descripteurs définis dans la norme	Nom des champs définis dans la table Contenu de BASIS
PI_CODE	NPLIBRE
PI_COTE	COTE
PI_DATECREATION	DATDEB
PI_DATEFIN	DATFIN
PI_DESIGNATION	CONTENU
PI_MCLIEU	MCLIEU
PI_MCNOM	MCNOM
PI_MCTHEME	MCCOM
PI_NOTE	OBS
PI_NUMEROREG	NUMREG

Les lignes surlignées en rouge correspondent aux index qui ont été rendus obligatoire dans la norme.

7.5.2 La structure des données

La définition de la structure des données a consisté à indiquer pour chaque index les éléments suivants :

- Type de donnée
- Taille maximale
- Nombre d'occurrences.

Ces éléments ont principalement été définis à partir de la structure des champs établie dans la base de données des Archives d'Etat. Un contrôle de faisabilité a été réalisé en confrontant cette définition avec la structure des données dans les systèmes d'information sources, et des éventuelles adaptations ont été apportées afin de prendre en compte des spécificités récurrentes.

7.5.2.1 Définition des types de données

Deux principaux types de données ont été retenus :

- Type Caractère
- Type Date (entier sur 8 digits).

L'utilisation de types de données tels que les données numériques a été évitée pour mettre en place une norme la plus ouverte possible. Par exemple, en typant l'index « Code de l'article » ou « code de la pièce » en Caractère et non en Numérique, on peut prendre en compte à la fois les codages numériques et alphanumériques réalisés dans les services versants.

7.5.2.2 Définition des tailles maximale des données

Les tailles maximales définies dans la norme correspondent aux caractéristiques des données définies dans BASIS. Ces tailles maximales sont suffisamment importantes pour permettre aux services versants de décrire les différents index de manière suffisamment complète, explicite et détaillée.

7.5.2.3 Définition du nombre d'occurrences

La base de données BASIS offre la possibilité de définir pour chaque donnée un nombre variable d'occurrences :

- 1 occurrence : une seule valeur possible
- n occurrences : plusieurs valeurs possibles pour un index, les occurrences étant séparées par un « ; » lors du stockage des données dans BASIS.

Des champs multi-valués de type « Caractère » ont été définis dans la norme. Pour faciliter la compréhension des types et des tailles de chaque champ, des types spécifiques ont été définis :

- type « Caractère » : donnée mono-valuée, taille variable
- type « Mot-clé » : 150 données « Caractère » * 60 caractères maximum par donnée
- type « Mot-clé personnes » : 150 données « Caractère » * 60 caractères maximum par donnée
- type « Langue » : 5 données « Caractère » * 25 caractères maximum par donnée
- type « Pièce » : 10 données « Caractère » * 150 caractères maximum par donnée.

Rapport final d'étude

Les nombres d'occurrences définis correspondent à la structure de données actuelles. Il nous a semblé que les nombres d'occurrences possibles stipulés dans la norme sont suffisamment importants pour ne pas créer de contraintes supplémentaires dans la sélection des données par les services versants.

Exemple d'occurrences admissibles pour un article : 5 langues, 150 mots-clés de Lieux, 150 mots-clés de Noms propres, 150 mots-clés de Matières, 10 Pièces constitutives (si le détail des pièces est connu, on tâchera d'indexer chaque pièce, sans limitation de taille. L'index ART_PIECE n'est utilisé que pour renseigner une description sommaire d'un contenu d'article non indexé).

7.6 Le contenu des données

Si la règle ISAD(G) ne décrit pas le contenu de l'information à mentionner dans chaque élément de description (ou index), elle fournit pour chacun des éléments les informations suivantes :

- Nom de l'élément de description régi par la règle
- Enoncé sur l'utilité de l'élément dans la description
- Enoncé de la règle ou des règles générales applicable(s) à cet élément
- Exemples illustrant l'application de la (ou des) règle(s).

Dans la norme édictée par les Archives d'Etat, les règles décrivant le contenu de l'information à donner pour chaque index ont été énoncées sur la base du modèle ISAD(G). Pour chaque index, on trouvera généralement les informations suivantes :

- Objectif de l'index, correspondant à l'énoncé sur l'utilité de l'élément donné par l'ISAD(G)
- Enoncé de la règle : règle(s) applicable(s) à l'élément en complément de l'énoncé
- Exemple : illustration par un ou plusieurs exemples types
- Donnée(s) concernée(s) : nom des données telles qu'elles ont été définies dans la structure de données
- Eventuellement formalisme (détaillé dans le chapitre suivant).

La description du contenu de l'information pour chaque index obligatoire ou optionnel de la norme est réalisée dans le chapitre Nomenclature des données (§5.3).

7.6.1 Contenu des index

Chaque index a été traité de manière à permettre aux services versants :

- D'en comprendre l'utilité
- D'en cibler le périmètre
- De trouver les éléments d'information nécessaires pour y faire mention de l'information adéquate en accord avec les règles archivistiques.

Les « règles » présentées dans la norme ont été définies sur la base des recommandations de l'ISAD(G) et des pratiques archivistiques des Archives d'Etat.

Rapport final d'étude

Les index communs à plusieurs niveaux de description ont été regroupés et traités de manière unique.

Pour certains éléments de description, en particulier pour la Désignation d'une unité documentaire, les règles énoncées restent assez générales, pour permettre leur mise en application par l'ensemble des services versants de l'administration publique cantonale. Etant donné les spécificités de chaque service cible auquel la norme peut s'adresser, il est nécessaire de fournir des directives générales tout en laissant la responsabilité à chaque service versant d'adapter les directives à son contexte, à ses processus et à ses archives. L'un des rôles des Archives d'Etat lors de l'intégration de ces données dans BASIS sera de procéder au contrôle de la qualité des descriptions externalisées et le cas échéant de procéder aux adaptations nécessaires.

7.6.2 Règles et formalisme pour la saisie des index

Il est nécessaire que certaines données respectent un formalisme dans la saisie de l'information. Ce formalisme a été établi à partir :

- Des formalismes adoptés par les Archives d'Etat pour la saisie des données dans BASIS, et identifiés lors de l'analyse menée aux Archives d'Etat
- De formalismes prônés par la profession (norme ISAAR sur les notices d'autorité archivistiques édictée par le CIA)
- De formalismes utilisés dans l'administration cantonale (registre des entreprises du canton de Genève, services versants).

Trois types d'index sont particulièrement concernés :

- Les mots-clés (lieux, personnes, matières)
- Les dates (formalisme donné par la structure des données)
- Les dates extrêmes d'un article.

La mise en place d'un formalisme strict répond à une problématique précise. Les deux exemples ci-dessous illustrent deux problématiques distinctes.

3. Formalisme des mots-clés de personnes, de lieux et de matières

Les noms de personnes, lieux et matières sont indexés dans la base de données et utilisés pour la recherche documentaire informatique. Ces index vont enrichir des bases d'index utilisées lors de la recherche documentaire. L'utilisateur peut ainsi interroger la base de données soit sur un élément de description, soit en combinant plusieurs éléments de description (ex : recherche sur un lieu et/ou sur un nom de personne).

Il est nécessaire de respecter un formalisme strict lorsqu'il n'existe pas de thésaurus ou de liste de mots-clés permettant de contrôler les termes d'indexation choisis. Ces deux types d'instrument utilisent une terminologie normalisée et contribue à aider l'utilisateur à sélectionner de manière logique des occurrences dans la base de données. Par rapport à la liste de mots-clés, le thésaurus permet d'améliorer la qualité de la sélection et d'éliminer les données récurrentes par l'utilisation de sa structure hiérarchique et relationnelle.

Rapport final d'étude

Lorsqu'il n'existe pas de vocabulaire de termes contrôlés d'indexation, il est nécessaire de respecter des règles afin d'éviter la multiplication des termes équivalents ou renvoyant à un même concept, ou de termes similaires orthographiés différemment. Des entrées multiples diminuent en effet la pertinence et la performance de la recherche. En formalisant la saisie des index, on met en place un premier contrôle du vocabulaire d'indexation.

Les principales règles édictées dans la norme concernant la saisie des noms propres concernent :

- Le formalisme MAJUSCULES - MINUSCULES
- La définition de(s) élément(s) d'entrée et de(s) élément(s) rejeté(s)
- L'emploi de qualificatifs pour distinguer des homonymes
- Les noms propres à particules
- L'utilisation des articles
- L'utilisation des abréviations.

2. Formalisme des dates extrêmes

Les dates extrêmes doivent être indiquées selon le formalisme AAAA-AAAA (années extrêmes séparées par un tiret).

Le champ « Date extrême de l'article » n'est pas indexé dans la base de données BASIS. Par contre, les dates de création et de clôture sont indexées, obligatoires et constituent des critères de recherche essentiels. Le formalisme requis vise à faciliter le retraitement de l'information avec pour objectif final de faciliter la recherche documentaire.

Les dates extrêmes d'un article doivent être indiquées si la date de création (éventuellement la date de clôture) ne sont pas connues de manière précise (jour mois année). Dans ce cas et pour permettre la création du fichier d'index par le service versant, un format « libre » est accepté, et les dates de création et de clôture ne sont pas requises. Ces dates sont cependant des champs obligatoires dans la base de données BASIS. C'est pourquoi, lors de la procédure d'import des données, le champ au format libre va devoir subir un traitement informatique permettant de renseigner les dates obligatoires. Seule une structure de données standard permettra d'automatiser le retraitement de l'index date extrêmes :

- 4 premiers digits : utilisés pour créer la date de création
- 4 derniers digits : utilisés pour créer la date de clôture.

Exemple :

Dates extrêmes fournie par le service versant : 1950-1951

Traitement des données dans BASIS :

Dates extrêmes renseignée : 1950-1951

Date de création créée : 19500101

Date de clôture créée : 19511231

7.7 Le fichier de transfert

Un fichier est un ensemble d'information plus ou moins structurées, pouvant contenir tous les types d'information possibles et interprété par le programme d'application.

Rapport final d'étude

Lorsque l'on parle d'échange de données informatiques, l'une des questions sous-jacentes concerne le format utilisé pour transmettre les données d'une base de données à une autre base de données. La définition du format du fichier de transfert doit comporter la définition de la structure du fichier et des données échangées, ainsi que le langage informatique requis.

L'un des tâches importantes durant l'étude a consisté à identifier et étudier les différents formats de fichiers possibles. D'autres éléments tels que la dénomination du fichier ont également été étudiés. Ce chapitre décrit les résultats des différentes études menées.

7.7.1 Dénomination du fichier

La normalisation des noms de fichiers qui vont être créés dans le cadre des directives émises suite à cette étude a été approuvée par le comité de pilotage du projet. Il s'agit clairement d'éviter toute confusion dans la compréhension des fichiers transmis :

- Identification de la provenance du fichier
- Identification du contenu du fichier
- Identification du type de fichier
- Utilisation d'intitulés longs et explicites.

Les directives pour la dénomination des fichiers informatiques transmis aux Archives d'Etat sont décrites dans le paragraphe 5.1.2 de la norme.

7.7.2 Structure générale du fichier

Tous les éléments généraux relatifs au contenu du fichier et à sa structure générale sont décrits dans le paragraphe 5.1.1 de la norme. Cette présentation générale a pour but de fournir au service versant une vue d'ensemble du fichier d'index, certains éléments étant repris et détaillés lors de la description du format proprement dit.

La présentation générale fournit également les éléments d'information requis pour la description du fichier lui-même. Cette information, appelée méta-information, doit accompagner tout document électronique et documenter le contexte de création du fichier et de traitement du fichier. Cette information est généralement donnée dans l'en-tête d'un fichier.

Il n'est pas prévu pour l'instant de récupérer informatiquement ces méta-données pour les intégrer dans la base de données BASIS. Cependant, la possibilité de mettre en place un traitement informatique de ces méta-données a été abordée lors de l'étude. Ce développement pourrait être envisagé dans les prochaines étapes du projet.

Enfin, les éléments généraux permettent de délimiter le contenu possible d'un fichier d'index :

- Unicité de la provenance des données
- Unicité de la série de documents indexés.

7.7.3 Format du fichier

La description du format de fichier retenu pour le transfert des fichiers d'index est réalisée dans le chapitre Format de transfert (§6) de la norme.

Rapport final d'étude**7.7.3.1 La norme XML**

Plusieurs formats de fichiers ont été envisagés lors de cette étude :

- Des formulaires d'import spécifiques au logiciel BASIS
- Des formats universels incluant des séparateurs de données et d'enregistrements (ex : format CSV)
- La norme XML dédiée à l'échange de documents informatiques. XML est un format texte utilisant des balises pour délimiter les données.

Le choix retenu est le XML, pour les raisons suivantes :

- La norme XML est particulièrement adaptée à l'échange de données et de documents structurés.
- La norme XML permet de décrire à la fois la structure des données et les données elles-mêmes.
- Un même fichier XML peut alimenter plusieurs tables (donc plusieurs niveaux de description).
- Le XML est une norme ouverte, indépendante des plates-formes logicielles ou matérielles.
- La norme XML est une norme établie, recommandée par le W3C depuis 1998, et basée sur un métalangage devenu norme ISO (le SGML).
- La norme XML est recommandée par le CTI dans la mise en œuvre des projets informatiques de l'Etat de Genève.

7.7.3.2 La norme archivistique EAD

Les recherches effectuées lors de l'étude ont montré qu'il existait une possibilité pour baser la norme des Archives d'Etat sur un outil archivistique fondé sur XML : l'Encoded Archival Description plus communément appelé l'EAD.

Pour construire un fichier XML, on peut écrire directement des documents en XML, utiliser des DTD (Document Type Definition) ou des Schémas XML qui aident les développeurs à définir leurs propres formats basés sur XML.

Or il s'avère qu'il existe une DTD pour les instruments de recherche archivistiques (répertoires et inventaires) appelée EAD. La DTD formalise la structure générale du fichier XML, c'est-à-dire les entités, éléments et notations utilisés dans le document. Un dictionnaire de balises EAD présente en détail les 145 éléments et leurs attributs. La plupart des éléments sont facultatifs, ce qui facilite l'encodage. L'EAD permet de décrire le nombre de niveaux hiérarchiques que l'on souhaite, en allant du fonds à la pièce.

L'EAD a été élaborée aux Etats-Unis dans le cadre d'un partenariat entre la société américaine des archivistes, propriétaire de l'EAD, et des groupements de grandes bibliothèques de recherche pour leurs collections spécialisées.

En Europe, quelques sites participent à l'évolution de la DTD, en Angleterre, en Espagne et en Italie. L'EAD a été testée depuis 1998 par la Direction des Archives de France. Il s'avère qu'elle est parfaitement compatible avec le standard de description international ISAD(G), que c'est un bon outil pour les répertoires et les inventaires, et que XML et l'EAD peuvent être utilisés en surcouche pour exporter des données. La Direction des Archives de France a d'ailleurs entamé des travaux de traduction et de mise en ligne de plusieurs instruments de recherche sous XML en EAD.

Rapport final d'étude

La version 1 de l'EAD est terminée depuis septembre 1998 et la version 2002 vient de paraître. La documentation est distribuée sur le site Internet de la Library of Congress.

La norme produite par les Archives d'Etat fournit les éléments d'information permettant d'encoder le fichier d'index selon la norme EAD : contenu du fichier, structure du fichier, définition de quelques balises, documentation de référence à consulter. Elle fournit également une illustration d'un inventaire à deux niveaux de description encodé sous XML suivant l'EAD. L'ensemble de ces informations a pour but de donner aux services versants les éléments permettant à chacun de comprendre le principe d'encodage, de se référer aux documentations existantes et de mettre en place les outils d'export adéquats.

7.8 Eléments d'organisation

Le dernier chapitre de la norme (§7) spécifie quelques éléments organisationnels relatifs à la production et au transfert des fichiers d'index :

- Contexte de transmission du fichier
- Cas particulier des échantillonnages
- Support de transmission des fichiers
- Périodicité de transmission des fichiers.

A ce stade du projet, la définition de procédures détaillées apparaît précoce. Cependant, ce chapitre pourra être étoffé en fonction des remarques des services versants lors de la mise en consultation de la norme et au fur et à mesure de sa mise en application.

8. Impact de la norme sur les Archives d'Etat

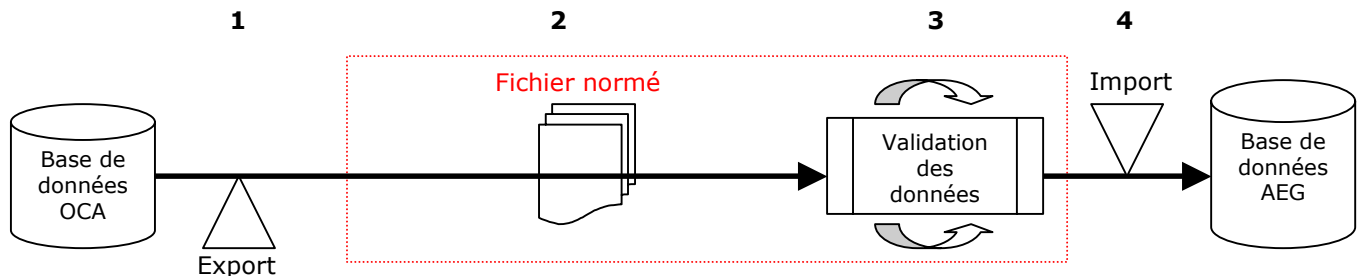
La mise en place de la norme ne constitue pas la fin du projet pour les Archives d'Etat. Les moyens techniques et organisationnels permettant de rendre opérationnel le traitement des fichiers d'index qui pourraient être transmis aux Archives doivent être mis en œuvre.

Les paragraphes qui suivent présentent :

- Les développements informatiques à réaliser pour automatiser l'alimentation des données dans BASIS
- Les conséquences organisationnelles pour les Archives d'Etat.

8.1 Traitement des fichiers d'import

Le schéma suivant présente la chaîne de production et de traitement des fichiers d'index depuis leur création jusqu'à leur import dans la base de données BASIS.



- Etape 1 : les index sont exportés de la base de données source et alimentent un fichier conforme à la norme des Archives d'Etat.
- Etape 2 : le fichier d'index est transmis aux Archives d'Etat.
- Etape 3 : la structure et le contenu du fichier d'index sera contrôlé via un outil de validation à développer.
- Etape 4 : un programme d'import des données dans BASIS sera développé afin d'automatiser l'alimentation de la base de données.

Les Archives d'Etat doivent donc mettre en place les outils nécessaires **pour automatiser l'alimentation de la base de données** BASIS. Deux niveaux de traitements sont nécessaires :

- Niveau 1 : mise en place d'un outil de validation du fichier informatique, permettant de contrôler et de valider le format du fichier ;

Rapport final d'étude

- Niveau 2 : développement de l'interface entre le fichier d'index et Basis
- Niveau 3 : contrôle de la qualité des données transmises. Ce contrôle pourra être réalisé avant ou après l'import dans BASIS.

Les paragraphes qui suivent décrivent de manière plus détaillée chacun des traitements informatiques à mettre en œuvre. L'ensemble de ces traitements pourra être intégré dans un module général facilitant le traitement des fichiers par les collaborateurs des Archives d'Etat.

8.1.1 Validation du fichier

Un traitement informatique doit permettre de contrôler formellement le fichier de transfert et de vérifier sa conformité aux exigences de la norme. Une analyse devra permettre de définir les contrôles nécessaires à mettre en œuvre. On peut cependant déjà envisager les traitements suivants :

- Contrôle de la syntaxe du fichier (conformité aux normes XML et EAD)
- Traitement des erreurs (génération d'un fichier d'erreurs à traiter)
- Contrôle de la structure du fichier et du nombre de niveaux de description.

8.1.2 Import des données dans BASIS

Une application informatique doit être développée pour permettre l'alimentation automatique de la base de données. Une analyse devra être menée pour spécifier les différents traitements à mettre en œuvre et définir le mode et le format d'import des données dans BASIS. Ces traitements peuvent être découpés en deux parties :

4. Conversion du fichier d'index au format choisi pour l'import dans la base de donnée, et import des données dans les tables BASIS
5. Mise en place des traitements annexes pour créer les données annexes non contenues dans le fichier transmis.

8.1.2.1 Traitement du fichier

Lors de l'analyse menée aux Archives d'Etat, les deux modules d'import de données disponibles dans BASIS ont été étudiés. Cette analyse a mis en évidence qu'entre les deux modes d'alimentation possible, à savoir l'alimentation d'un fichier texte via le module HVU (High Volume Update) et l'alimentation depuis un outil externe via le pilote ODBC, le choix se porterait de manière préférentielle sur l'alimentation par le module HVU pour les raisons suivantes :

- Il est adapté aux traitements de données en masse ;
- Il offre l'avantage d'un chargement rapide de la base de données ;
- Il contient une procédure de validation de la structure des données durant l'import ;
- Il permet la création d'un fichier des enregistrements rejetés qui facilite leur retraitement ;
- Il permet un retraitement plus aisé du lot de données (ajout de données spécifiques aux AEG par exemple).

Rapport final d'étude

Le module HVU propose plusieurs formats propriétaires de « formulaires » d'import de données. Si ce choix est confirmé, le fichier XML devra être converti dans un de ces formats.

La possibilité d'importer directement un fichier au format XML pourra être envisagée en fonction des nouveaux outils proposés avec le logiciel BASIS.

Le programme de transfert des données dans les tables et les champs de la base de données des archives d'Etat sera développé à partir de l'étude de la structure d'un fichier EAD et des informations contenues dans la norme et dans le présent rapport, en suivant la démarche ci-après :

- Table de correspondance entre la structure de données EAD et la structure de données BASIS
- Conversion du fichier au format d'import retenu
- Import des données dans la base de données.

8.1.2.2 Autres traitements

En parallèle de l'import des données contenues dans le fichier d'index transmis par le service versant, l'application informatique de traitement des fichiers doit prévoir un certain nombre de traitements annexes. Ces traitements sont rendus nécessaires par l'automatisation du processus de création de données dans BASIS. Lors d'une saisie manuelle d'information dans la base de données, l'archiviste peut ou doit renseigner certaines données et tables qui n'ont pas été intégrées à la norme. Il s'agit d'automatiser la saisie de ces données afin de respecter les procédures de saisie des inventaires en vigueur aux Archives d'Etat.

Les différentes études menées ont mis en évidence un certain nombre de ces traitements informatiques à automatiser (ex : données obligatoires pour qu'un enregistrement soit valide et qui ne peuvent être fournies par le service versant). On peut également envisager qu'une partie des traitements décrits ne soit pas automatisée, mais dans ce cas une « procédure » devra être mise en place aux Archives d'Etat pour décrire les données à renseigner avant (éventuellement après) l'import d'un fichier d'index. Une étude complémentaire devra être menée par les Archives d'Etat pour compléter ou modifier les premiers traitements identifiés ci-dessous.

1. Identification du service versant (table OCA)

Lors de la réception d'un fichier d'index, les Archives d'Etat vérifient que l'organisme versant est identifié dans le système, et au besoin le créent ou le modifient.

2. Identification de la série organique (table SERIE)

Lors de la réception d'un fichier d'index, les Archives d'Etat vérifient que la Série Organique est identifiée dans le système, et au besoin la créent ou la modifient. Les éléments permettant d'identifier la série pourront être en partie récupérés dans le fichier d'index.

Rapport final d'étude**3. Identification du versement (table ENTREE)**

Lorsqu'un fond d'archives est versé aux AEG, la première étape consiste généralement à attribuer un numéro d'entrée (« cote ») et à effectuer une description du versement dans une table dédiée (table ENTREE). Afin de maintenir les relations informatiques entre la table ENTREE et la table DOCUMENT, le numéro d'entrée attribué à la série de documents indexés pourra être saisi et alimenter automatiquement le champ « Numéro d'entrée » de chaque nouvel enregistrement créé dans la table DOCUMENT lors de l'import de données.

4. Données créées automatiquement par le système

Lors de l'étude menée aux AEG, les champs des tables DOCUMENT et CONTENU qui sont créés automatiquement par BASIS lors de la saisie d'un nouvel enregistrement ont été identifiés (date création de la fiche, créateur de la fiche, clé du document, etc.). Lors de la procédure d'import des index dans BASIS à partir de fichiers externes, un traitement devra prévoir l'alimentation automatique de ces champs.

5. Données complémentaires à renseigner en lot

Lors de l'étude menée aux AEG, les champs des tables DOCUMENT et CONTENU pour lesquels l'information ne peut pas être fournie par le service versant mais qui doivent pouvoir être renseignés en lot lors de l'import des données ont été identifiés.

Lors de l'import de données dans la table DOCUMENT, les champs « Code département », « Code OCA », « Code Série », « Nom OCA », « Numéro d'entrée » doivent être renseignés dans chaque enregistrement. Pour un fichier donné, le même code OCA, code Série, etc. sera valable pour tous les enregistrements dans la table DOCUMENT.

Lors de l'import de données dans la table DOCUMENT, une « cote AEG » devra être attribuée à chaque nouvel enregistrement.

Lors de l'import de données dans la table CONTENU, les champs « Code document », « Cote AEG », « Nom OCA » doivent être renseignés dans chaque enregistrement. Contrairement au « Nom OCA » valable pour tout le fichier transmis, les données renvoyant au niveau de description supérieur (Cote et Code Document) seront identiques pour un ensemble de Pièces d'un même document mais différeront d'un document à un autre.

6. Traitement des méta-données du fichier

Les fichiers d'index transmis contiendront les méta-données décrivant le fichier lui-même (créateur, mode de création, etc.) Deux options peuvent être envisagées pour traiter cette information :

- soit prévoir l'extraction des méta-données dans un format permettant leur impression dans le but d'être retraitées manuellement.
- soit prévoir l'alimentation des méta-données dans BASIS et dans ce cas mettre en place les développements nécessaires permettant de stocker cette information.

8.1.3 Contrôle des données

Un contrôle qualitatif des index transmis devra être réalisé avant, pendant ou après l'import des données dans la base de données BASIS. Ce traitement devra permettre :

- De contrôler la qualité des index transmis (terminologie, pertinence, information explicite, etc.)
- D'apporter les modifications jugées nécessaires par le service des Archives d'Etat
- De compléter les enregistrements en renseignant les champs qui n'ont pas été inclus dans la norme, soit parce qu'ils sont propres à la gestion archivistique en vigueur dans le service, soit parce qu'ils résultent de travaux de recherche et d'analyse complémentaires menés aux AEG, soit enfin parce qu'ils concernent la conservation à long terme des documents (urgence de restauration par exemple).

Les Archives d'Etat devront définir la procédure à mettre en place pour le contrôle qualitatif des données transmises afin d'assurer la cohérence et le respect des pratiques archivistiques du service, et de s'assurer de la pertinence des données informatiques mises à disposition du public.

8.2 Conséquences organisationnelles

L'ensemble des travaux menés durant cette étude, la norme produite au terme de cette étude, et les développements informatiques complémentaires rendus nécessaires pour permettre la mise en œuvre opérationnelle du projet ont des conséquences organisationnelles pour le service des Archives d'Etat.

Il est nécessaire de rappeler quelques-uns des buts visés en instaurant une norme pour la transmission informatique des index des archives historiques.

Les archivistes des AEG ont pour mission d'inventorier les fonds qui leur sont versés. Cet inventaire est fastidieux en raison du volume d'archives entrant chaque année. Les AEG emploient régulièrement des stagiaires pour effectuer les tâches de saisie. Même ainsi, les objectifs fixés annuellement pour la saisie des inventaires ne sont pas atteints. Les AEG ont deux possibilités pour faire en sorte d'atteindre leurs objectifs : embaucher plus de personnel ou diminuer le volume de travail par l'automatisation d'une partie de la saisie.

La seconde option est celle qui a été retenue. Il semble envisageable de réduire ou de supprimer la part non productive du travail qui consiste à ressaisir des données qui existent déjà au format électronique dans les applications utilisées en amont de la chaîne d'archivage. Les AEG souhaitent gagner en efficacité et en productivité, permettre une mise à jour au plus vite de la base de production pour mettre prochainement à disposition du public une base de consultation dont le contenu soit représentatif des archives historiques réellement disponibles.

Or cela nécessite la mobilisation de ressources pour faire face à ce nouveau mode de fonctionnement.

Rapport final d'étude

Dans un premier temps, il va falloir mettre en œuvre les analyses et les projets informatiques permettant de mettre en place les outils de contrôles et d'import de données. Les archives d'Etat devront consacrer du temps afin de définir avec leurs partenaires les modalités des traitements informatiques à déployer.

Les archives d'Etat vont également devoir définir en interne des procédures de travail relatives au traitement des fichiers informatiques transmis. L'élaboration de procédures de travail basées sur les nouveaux outils développés est essentielle pour assurer l'application de pratiques standards et maîtrisées, pour assurer la diffusion de l'information et la pérennité des compétences dans le service, le cas échéant pour définir la répartition des tâches entre les collaborateurs. L'élaboration de ces procédures nécessitera également du temps et de l'investissement humain.

Suite à la mise en consultation de la norme et à la mise en application qui devrait en découler, il est possible que le service des Archives d'Etat soit sollicité par un certain nombre de services versants pour apporter un support dans l'élaboration des fichiers d'index. Des moyens devront être dégagés pour faire face à cette demande.

L'informatisation de l'alimentation de la base de données, si elle permet d'accélérer la mise à jour des inventaires, ne peut être mise en place sans un contrôle archivistique et une mise à jour des descriptions transmises. Etant donné les volumes importants de documents d'archives qui pourraient être ainsi indexés et reçus aux AEG (ex : 20'000 dossiers médicaux annuels pour les HUG seulement), le temps nécessaire pour le traitement et le contrôle du contenu des fichiers d'index transmis nécessitera du temps et donc des ressources humaines dédiées à ce travail. On peut penser que la transmission de fichiers d'index par les services versants va se mettre en place de manière progressive, mais étant donné les données informatiques potentiellement disponibles dans les services interviewés et dans l'administration cantonale en général, la périodicité des versements devrait s'accélérer dans les années à venir. On peut envisager qu'un ou plusieurs postes de travail dédiés au traitement de ces fichiers doivent être créés. Le risque si les moyens humains ne sont pas mis en place serait de retomber dans le décalage qui existe actuellement entre la réception des archives et la mise à jour de la base de données de production.

Enfin, dans un contexte d'informatisation croissante des outils de travail et en l'occurrence des index transmis, le service des Archives devra envisager la possibilité de se doter de compétences informatiques. Il s'agit pour les Archives d'Etat de disposer des compétences nécessaires pour pouvoir émettre en connaissance de cause les directives informatiques recommandées par la loi sur les archives publiques. Il s'agit également de jouer un rôle de conseil et d'information auprès des services versants, de mettre en place des projets tels que l'étude qui vient d'être menée avec pour objectif de jouer un rôle proactif dans la gestion informatique des archives cantonales au lieu de subir les conséquences de mode d'informatisation, de gestion et de stockage de données trop disparates. Il s'agit enfin de disposer en interne des compétences nécessaires pour développer ou du moins utiliser efficacement les systèmes informatiques de gestion documentaire.

9. Conclusion de l'étude

En conclusion de l'étude et au vu des résultats obtenus, le bilan général est positif puisque les objectifs fixés dans le cahier des charges ont été atteints.

La norme établie s'inscrit dans un cadre légal en réponse au règlement d'application de la nouvelle loi sur les archives publiques et aux obligations qui y sont données pour les Archives d'Etat.

La faisabilité de l'étude et de la rédaction d'une norme standardisée pour l'ensemble des services versants n'était pas établie au démarrage des analyses, en raison de la multiplicité des services versants (quelques 1'200 OCA identifiés dans la base de données des Archives d'Etat) et des spécificités de chacun d'entre eux, en terme de documents d'archives produits, d'instruments de recherche utilisés et de systèmes informatiques déployés. Il s'est avéré au terme de l'étude que la rédaction de la norme était envisageable, chaque service ayant la responsabilité d'adapter les standards fournis en fonction de ses propres documents et outils.

L'un des résultats de l'étude a été de constater et même de démontrer qu'il existait un réel besoin, dans l'échantillon de services producteurs d'archives interviewés, de recommandations (directives ou normes) ayant trait à l'archivage de données électroniques. En choisissant de mettre en place cette étude et de fournir aux services versants des éléments d'information sur les données qui doivent être prioritairement créées pour décrire des fonds d'archives, ainsi que des recommandations sur les technologies à utiliser pour stocker ces données en vue d'un échange informatique, les Archives d'Etat ont répondu en premier lieu à leur propre besoin mais également aux questions posées dans un certain nombre de services administratifs. Il faudra maintenant attendre les réactions de ces services quant au contenu des recommandations émises, tout en sachant qu'il est de toute façon nécessaire de réfléchir à ces problématiques et de mettre en place les outils permettant de les résoudre.

Ce projet de récupération des index informatiques permettant de décrire les archives constitue une avancée importante dans la stratégie relative à la gestion électronique des archives. L'étude a permis de mettre en évidence que cette problématique n'est pas isolée, et qu'au niveau fédéral, des réflexions sont également menées pour prendre en compte le devenir des archives électroniques. Les Archives d'Etat de Genève ne prévoient pas à l'heure actuelle de proposer des solutions globales pour le stockage des documents d'archives au format électronique, mais ici comme ailleurs, la question devra être débattue étant donné l'explosion des outils de gestion électronique des documents (ex : AIGLE dans l'administration cantonale).

Un autre résultat de l'étude a été d'étudier les solutions archivistiques utilisées et préconisées dans les autres pays. L'étude de l'outil EAD, développé par les américains et repris par de nombreux pays dont la France, a démontré qu'il s'agissait d'un standard en XML adapté à l'encodage des instruments de recherche archivistiques, que l'outil était ouvert, pérenne et évolutif et pouvait constituer la référence technique et métier à adopter dans la norme.

Rapport final d'étude

La définition de la norme s'est en effet appuyée sur les outils, standards, processus existants, le but étant de fournir un outil ouvert, respectant les standards de la profession. Les principales références utilisées ont été les normes internationales archivistiques du CIA, l'outil EAD de la société des archivistes américains, les recommandations de l'association des archivistes suisses et des archives du canton de Vaud, et bien sûr les outils et pratiques des archives d'Etat de Genève.

La norme a été produite dans l'optique de guider les services versants lors de la construction de nouveaux systèmes d'information en fournissant les spécifications à prendre en compte d'un point de vue archivistique. Elle s'inscrit donc dans une vision à moyen et long terme, ce qui n'empêche pas les services versants qui pourraient dans les mois à venir produire des index conformes à la norme de mettre en place les moyens nécessaires. Afin de faciliter ce travail et de motiver la création des fichiers d'index, il a été décidé de ne rendre obligatoire qu'un minimum de données.

Enfin, il est nécessaire de préciser qu'en tant que document de référence, il sera nécessaire que les Archives d'Etat prévoient de maintenir la norme produite et de procéder à ses révisions périodiques, ceci afin de tenir compte des évolutions techniques ou fonctionnelles de leur service, du monde archivistique ou de l'administration cantonale dans son ensemble.

10. Prochaines étapes

La norme réalisée au terme de l'étude a été validée par les membres de l'équipe de projet et remise aux Archives d'Etat à la fin du mois de décembre 2002.

Dès le début de l'année 2003, une procédure d'information de l'administration genevoise sera mise en place. Un premier niveau d'information sera effectué au sein du service des Archives d'Etat. Une deuxième information permettra de présenter l'étude et ses résultats à la Direction du Département de l'Intérieur, de l'Agriculture et de l'Environnement. Enfin, une information générale sera donnée dans le cadre d'une présentation ouverte à tous les participants à l'étude ainsi qu'aux services de l'Etat intéressés ou concernés par le projet. Ces présentations auront un but d'information et également de retour d'information sur le standard produit.

Une procédure de mise en consultation de la norme sera mise en place, afin d'aboutir à un consensus sur le résultat de l'étude. Si l'accueil est favorable, les Archives d'Etat devront mettre en œuvre les étapes nécessaires afin de rendre applicable leurs nouvelles directives.

Nous avons évoqué dans le paragraphe décrivant l'impact du projet pour les Archives d'Etat les traitements informatiques qui doivent être mis en place afin d'importer les données transmises dans leur base de données. L'année 2003 devra donc voir la définition et le développement des outils informatiques décrits.

Enfin, la possibilité de développer un outil de saisie des index permettant de créer le fichier d'index au format donné dans la norme a été évoquée durant l'étude. Cet outil pourrait être mis à disposition des services versants qui ne disposent pas d'outils informatiques pour répertorier leurs dossiers et/ou qui doivent mettre à disposition des AEG de faibles volumes d'archives. Les Archives d'Etat seraient responsables de la définition de cette application. Une interface de saisie disponible sur le Web pourrait par exemple être développée et utilisée directement par chaque service versant qui le souhaite pour alimenter la base de données.

11. Axes de travail aux Archives d'Etat

Pour clore ce rapport final d'étude, un certain nombre d'axes de travail ou de réflexions à mener ont été identifiées lors notamment des séances avec Monsieur Roger Rosset. Il s'agit d'améliorations à apporter au système de gestion documentaire tel qu'il existe aujourd'hui, dans le but de faciliter le travail des collaborateurs des Archives d'Etat ou de combler des lacunes dans les outils informatiques existants. Les différents « projets » identifiés sont listés ci-après.

11.1 Développement de thésaurus

L'une des faiblesses de la base de données utilisée pour la saisie des inventaires est l'absence de liste de mots-clés et de thésaurus permettant de contrôler les termes d'indexation saisis dans les tables. Cette lacune a dû être contournée dans la norme qui a été rédigée, soit en donnant des consignes strictes concernant le formalisme de saisie des index, soit en donnant des consignes très générales et en laissant la responsabilité aux services versants de choisir « judicieusement » les index (pour les mots-clés matières notamment).

Plusieurs thésaurus seraient utiles dans le système de gestion documentaire :

- Thésaurus pour identifier la provenance d'un fonds
- Thésaurus de noms géographiques
- Thésaurus de mots-clés matières.

Le développement d'un thésaurus est un travail long et complexe, qui n'a pas été mis en œuvre jusqu'à ce jour aux Archives d'Etat. La possibilité « d'acquérir » des thésaurus déjà établis a été envisagée sans succès.

11.2 Développement d'un module de saisie des index « au rayon » pour les archivistes d'Etat

Les Archives d'Etat ont un besoin interne de saisie des inventaires depuis un ordinateur portable. Cet inventaire pourrait être réalisé depuis une application bureautique et les données importées dans la base (table DOCUMENT) via le pilote ODBC driver de BASIS qui devrait être prochainement acquis par les Archives d'Etat. Le développement de l'outil de saisie pourrait être rapide en fonction du logiciel choisi (Excel, Access) et le retour sur investissement important en terme de temps gagné pour la saisie des inventaires d'archives déjà stockées aux AEG.

11.3 Déclaration de la base données BASIS dans le cadre de la LITAO

La base de données des Archives d'Etat a vocation de contenir de plus en plus de données protégées par la loi et relatives à des personnes. Elle entre donc dans le cadre de la LITAO qui s'applique à tous les fichiers de l'Etat relatifs à des personnes et aux résultats du traitement de ces fichiers. Le Conseil d'Etat doit donner son autorisation pour la création de tels fichiers ou banques de données ainsi que pour la transmission de données entre services.

Une demande devrait donc être soumise au Conseil d'Etat concernant le traitement des données personnelles dans BASIS. Le dossier porté au catalogue des fichiers géré par la Commission de contrôle de l'informatique de l'Etat devra être constitué sur la base des spécifications données dans le règlement d'exécution de la LITAO.

11.4 Mise à disposition d'un outil pour la mise à jour des fiches de versements d'archives

En parallèle de l'informatisation de la transmission des instruments de recherche des archives historiques, un projet d'informatisation au niveau cantonal de la saisie des fiches de versements pourrait constituer une avancée pour les Archives d'Etat. Ces fiches de versement permettent de décrire sommairement un fonds d'archives lors de son versement. Deux possibilités peuvent être envisagées pour atteindre ce but :

- Fournit un accès à distance au système BASIS pour les services versants pour la création des fiches d'enregistrement des versements (table ENTREE)
- Réaliser et mettre à disposition un outil compatible ODBC (Excel, Access) pour la saisie par les services versants des informations relatives au versement, suivant le même principe que celui évoqué pour la saisie des inventaires au rayon (§11.2).

11.5 Comparatif d'outils ou bases de données pour la gestion documentaire

L'opportunité de mener une étude permettant d'établir un comparatif entre le logiciel de gestion documentaire BASIS, et d'autres outils documentaires basés sur des technologies recommandées de manière transversale dans l'administration cantonale a été envisagée durant l'étude. Si le système BASIS semble pérenne aujourd'hui et qu'il n'est pas envisagé de le migrer vers une autre application dans les années à venir, la possibilité d'une réorientation informatique ne peut être complètement écartée. Dans cette optique, il serait intéressant de comparer les fonctionnalités de différents outils et bases de données, soit de solutions packagées utilisées dans le monde des archives, soit de solutions développées à partir d'une base de données ORACLE.