



CAHIER DES CHARGES

pour la mise en œuvre d'un pilote pour une plate-forme
d'archivage électronique dans la sphère publique



TABLE DES MATIERES

1.	Objet du cahier des charges	5
2.	La direction des Archives de France	7
3.	Le contexte.....	9
3.1	La gestion des archives publiques en France.....	9
3.1.1	L'application aux archives papier	10
3.1.2	La problématique des archives électroniques	11
3.2	La direction générale pour la modernisation de l'Etat (DGME) et le plan stratégique pour l'administration électronique (PSAE).....	13
3.2.1	La DGME	13
3.2.2	L'action ADELE 103 : Archivage électronique	14
3.3	Le département du système d'information (DSI) du ministère de la culture et de la communication et le schéma directeur du système d'information (SDSI)	15
3.3.1	Le DSI (source : présentation du DSI sur Sémaphore).....	15
3.3.2	Le Schéma directeur du système d'information (SDSI) du ministère de la Culture	16
3.3.3	Le schéma directeur de la sécurité des systèmes d'information (SDSSI) du ministère de la culture	17
4.	Référentiels, normes et standards à prendre en compte pour le pilote	19
4.1	Documents de référence	19
4.1.1	Norme ISO 14721 (modèle OAIS)	19
4.1.2	La DTD EAD (Encoded Archival Description)	20
4.1.3	La politique d'archivage sécurisé dans la sphère publique	21
4.2	Les documents applicables	22
4.2.1	Le standard d'échange de données pour l'archivage électronique (versement, communication, élimination)	22
4.2.2	Guide UML XML élaboré par la DGME	24
4.2.3	PRIS version 1	24
4.2.4	Le cadre commun d'interopérabilité version 2.1	24
5.	Périmètre.....	26
5.1	Présentation générale	26
5.2	Sources de données à prendre en compte	27
5.3	Les acteurs du pilote	30
5.3.1	La DGME	30
5.3.2	Les missions des archives nationales auprès des ministères (Premier ministre, Education et Equipement).....	31
5.3.3	Achatpublic.com	31
5.3.4	Le Conseil général des Yvelines – la Direction des Archives Départementales et la DSI	31
5.3.5	Le centre des archives contemporaines de Fontainebleau (service Constance)	31
5.3.6	Le département des systèmes d'information (DSI) du ministère de la Culture	31
6.	Présentation des fonctionnalités du pilote	32
6.1	Le modèle OAIS	32
6.1.1	Contenu d'information	32
6.1.2	Information de pérennisation	33
6.1.3	Définition d'un paquet d'information (ou lot)	34
6.1.4	Information de description.....	34
6.2	Présentation des fonctionnalités	35

7.	Détail fonctionnel	37
7.1	F1. Versement	37
7.1.1	P0 : Préparation du transfert et transfert	39
7.1.1.1	Sous-processus P0-1 : préparation du transfert	39
7.1.1.2	Sous-processus P0-1a : Transmission manuelle du lot sur support amovible	40
7.1.1.3	Sous-processus P0-1b : Transmission manuelle du lot au service d'archives par réseau informatique	41
7.1.1.4	Sous-processus P0-1c : Transmission automatique du lot au service d'archives par réseau informatique (actuellement hors périmètre)	41
7.1.2	P1.Recevoir versement	42
7.1.3	P2.Vérifier transmission	42
7.1.4	P3.Contrôler la conformité	44
7.1.5	P4.Journaliser.....	45
7.1.6	P5.Consulter.....	45
7.1.7	P6.Convertir.....	45
7.1.8	P7.Générer PIA.....	46
7.1.9	P8.Coordonner les mises à jour	47
7.2	F2. Stockage	47
7.2.1	P1.Recevoir.....	49
7.2.2	P2.Mode de stockage	49
7.2.3	P3.Détruire.....	50
7.2.3.1	Sous-processus P3-1 : Demande d'accord pour l'élimination	50
7.2.3.2	6.4.2 Sous-processus P3-2 : Accord pour l'élimination/Refus pour l'élimination	51
7.2.3.3	6.4.3 Sous-processus P3-3 : Notification d'élimination.....	51
7.2.4	P4.Garantir l'intégrité.....	51
7.2.5	P5.Gérer les migrations	52
7.2.5.1	Sous-processus P5-1 : Régénération des supports (tâche entièrement prise en charge par la plate-forme).....	52
7.2.5.2	Sous-processus P5-2 : Migration de formats (hors périmètre)	52
7.2.5.3	Sous-processus P5-3 : Restauration du contenu d'un support.....	53
7.2.6	P6.Préparer.....	54
7.2.7	P7.Fournir les statistiques	54
7.3	F3. Gestion des données descriptives	54
7.3.1	P1.Assurer lien.....	55
7.3.2	P2.Mettre à jour	55
7.3.3	P3.Garantir l'intégrité.....	56
7.3.4	P4.Administrer.....	56
7.4	F4. Communication / Consultation des Archives	56
7.4.1	P1.Vérifier les accès	58
7.4.2	P2.Gérer les demandes (Consultation de la base Archives)	58
7.4.3	P3.Exécuter requêtes (Sortie d'une commande).....	59
7.4.4	P4.Mettre en forme	60
7.4.5	P5.Communiquer	60
7.4.5.1	Sous-processus P5-1 : Transmission par support amovible	60
7.4.5.2	Sous-processus P5-2 : Transmission par réseau	61
7.5	F5 Moteur de recherche	62
7.5.1	Fonctionnalités.....	62
7.5.2	Paramétrage	62
7.5.3	Présentation des résultats de la recherche.....	63
8.	Architecture générale.....	66

8.1	Schéma général.....	66
8.2	Composants applicatifs.....	66
8.3	Remarques et limites concernant l'architecture	71
8.3.1	Nombre de sites	71
8.3.2	Distinction des serveurs.....	Erreur! Signet non défini.
8.3.3	Interopérabilité du système.....	71
8.3.4	Evolutivité de la solution proposée.....	Erreur! Signet non défini.
8.3.5	Gestion de la duplication des informations. Résumé des configurations possibles et comparaison.....	72
9.	Contraintes et exigences techniques	73
9.1	Pré-requis technique : Standards pour exploitation, réseau et postes de travail	73
	Suivi des anomalies	73
	Messagerie électronique	73
	Référentiels	73
9.2	L'environnement informatique du centre des archives contemporaines (le service Constance)	73
9.3	Standards de développement	74
9.4	La plate-forme logicielle et matérielle.....	74
10.	Politique de sécurité à mettre en œuvre	78
10.1	Identification/authentification.....	79
10.2	Protection des accès logiques	82
10.3	Sécurité logique des systèmes informatiques	Erreur! Signet non défini.
10.4	Journalisation et procédures de constitution des données de traçabilité.....	83
10.5	Mesures de sécurité liées à l'intégrité des objets archivés	83
10.6	Mesures de sécurité liées à l'horodatage des opérations d'archivage.....	83
11.	Déroulement du projet	84
11.1	Pilotage du projet.....	84
11.2	Conduite du projet : Les structures de pilotage	85
12.	Définitions	87

1. Objet du cahier des charges

La prestation qui fait l'objet du présent marché vise la réalisation d'une plate-forme d'archivage électronique, c'est-à-dire une infrastructure matérielle et logicielle de préservation à long terme de données électroniques comportant un ensemble complet de fonctionnalités de transfert, de réception, de contrôle, de stockage réparti de communication et d'élimination des données.

Cette infrastructure, opérationnelle et exploitable, constituera un **pilote**.

Le présent marché concerne la mise en œuvre d'un **pilote** pour cette plate-forme. Les prestations auront lieu **au centre des archives contemporaines de Fontainebleau (Archives nationales) et au département des systèmes d'information du ministère de la Culture et de la Communication à Saint-Quentin en Yvelines**.

Il est demandé aux soumissionnaires de proposer une ou des solutions pour la mise en place complète de l'application, depuis les spécifications à la production, en passant par le transfert de compétences et la formation des utilisateurs. Des précisions supplémentaires sur les prestations attendues sont fournies plus bas.

La plate-forme pilote est alimentée par diverses sources de données (ministères, collectivités territoriales...) en nombre et volume limités. L'ensemble des fonctions décrites dans ce document doit être couvert dans le cadre de ce marché (sauf mention particulière "Hors de la prestation du pilote").

Cette plate-forme peut résulter soit d'un développement spécifique, soit d'un progiciel à paramétrer par intégration de briques logicielles en vue de réaliser le pilote. La solution pourra également reposer sur des briques construites à partir de logiciels libres, dans le cadre des contraintes techniques du schéma directeur informatique du ministère de la culture.

L'ensemble des services est réalisé au profit de :

**Direction des archives de France
56, rue des Francs-Bourgeois
75141 Paris Cedex 03**

2. La direction des Archives de France

La direction des Archives de France est l'une des directions du ministère de la Culture. Elle est chargée de proposer les choix stratégiques à opérer en matière d'archives et de les mettre en oeuvre. Elle exerce un rôle de conseil, d'incitation, de réglementation, d'évaluation et de contrôle en ce qui concerne la collecte, le tri, le classement, la description, la conservation et la communication tant aux chercheurs qu'à l'ensemble des citoyens, des archives publiques autres que celles des ministères des Affaires étrangères et de la Défense.

Afin de garantir le respect de la légalité et l'application de normes scientifiques et techniques uniformes, elle exerce son contrôle sur les Archives nationales, sur les services d'archives régionaux, départementaux et communaux ainsi que sur les services d'archives des organismes autorisés, à titre dérogatoire, à gérer leurs archives définitives.

Au sein de la direction des archives de France, le département de l'innovation technologique et de la normalisation (DITN) :

- définit les normes professionnelles en matière de traitement des archives (notamment le classement, la description, la rédaction des instruments de recherche), de conservation, notamment la conservation préventive, la restauration, le transfert sur des supports de substitution, **la sauvegarde et l'accessibilité des archives électroniques** ;
- **suit et valide la mise en oeuvre de ces normes** ;
- **assure la veille technologique dans ces domaines**, en concertation avec les organismes spécialisés ;
- suit les projets d'aménagement et de construction de bâtiments d'archives, accorde le visa technique de la direction sur ces projets et conseille les services d'archives en la matière ;
- assure une fonction de **conseil et de veille en matière d'informatisation des services d'archives** ;
- **coordonne la politique de recherche de la direction.**

Les Archives nationales, placées sous l'autorité de la direction des Archives de France, accueillent les archives des ministères et des services nationaux en général. Elles sont constituées de cinq centres :

- Centre historique des Archives nationales (Paris) : documents antérieurs à 1958 et archives des chefs de l'Etat ;
- Centre des archives contemporaines (Fontainebleau) : documents majoritairement postérieurs à 1958 ;
- Centre des archives d'Outre Mer (Aix-en-Provence) : documents sur les anciennes possessions françaises outre-mer ;
- Centre des archives du monde du travail (Roubaix) : fonds d'entreprises, de syndicats, d'associations, d'architectes ;
- Centre national du microfilm (Espeyran) : microformes originales des documents conservés dans les autres centres (nationaux ou territoriaux).

Le centre des archives contemporaines de Fontainebleau abrite **le service Constance, chargé depuis une vingtaine d'années de la conservation des archives électroniques des ministères**, au premier rang desquelles des bases de données statistiques.

La construction d'un nouveau Centre a été annoncée par le Président de la République le 9 mars 2004. L'ouverture de ce nouveau bâtiment, situé à Pierrefitte-sur-Seine (Seine-Saint-Denis), est prévu pour 2010 ; le projet est suivi, sous tout ses aspects par un directeur de projet placé auprès de la directrice des Archives de France. Destiné à conserver les archives depuis 1790 et pour les trente ans à venir, le bâtiment de Pierrefitte accueillera une partie des fonds actuellement conservés au Centre historique des Archives nationales, à Paris, et au Centre des archives contemporaines, à Fontainebleau, et fonctionnera en lien avec ces deux sites.

Les centres des archives nationales ont vocation à devenir des services à compétence nationale à compter du 1^{er} janvier 2007, avec la constitution de trois pôles dont le pôle francilien rassemblant les sites des actuels centre historique des archives nationales, centre des archives contemporaines ainsi que celui de Pierrefitte-sur-Seine.

Au-delà du réseau des Archives nationales, fonctionne le réseau des services publics en France : archives départementales, archives municipales, archives régionales, l'ensemble de ces services étant des services décentralisés.

- 3. Toutefois, le contrôle scientifique et technique de l'Etat (direction des Archives de France, Inspection générale des archives de France, directeurs des services d'archives départementales qui sont des agents de l'Etat mis à disposition des collectivités territoriales), à savoir les conditions de versement, gestion, traitement et élimination des archives publiques, continue à s'exercer sur ce réseau.**

Le contexte

3.1 La gestion des archives publiques en France

Le code du patrimoine définit les archives comme "l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité" (article L. 211-1) et les archives publiques comme "les documents qui procèdent de l'activité de l'Etat, des collectivités territoriales, des établissements et entreprises publics ; les documents qui procèdent de l'activité des organismes de droit privé chargés de la gestion des services publics ou d'une mission de service public ; les minutes et répertoires des officiers publics ou ministériels" (article L. 211-4).

Le décret n° 79-1037 du 3 décembre 1979 relatif à la compétence des services d'archives publics et à la coopération entre les administrations pour la collecte, la conservation et la communication des archives publiques définit les règles de gestion des documents d'archives publiques au long de leur cycle de vie.

Ces documents sont d'abord conservés par les services qui les ont produits tant qu'ils leur servent régulièrement. Ils sont alors qualifiés d' "archives courantes".

Ensuite, lorsque leur utilisation devient exceptionnelle mais qu'ils gardent une utilité de preuve, ils restent conservés suivant des modalités diverses soit dans le service producteur, soit dans un dépôt dit de préarchivage, ou enfin dans un service public d'archives. Ils sont alors qualifiés d' "archives intermédiaires".

A l'issue de cette période, un tri est effectué entre les documents présentant un intérêt pour l'histoire ("archives définitives"), qui sont versés dans les services publics d'archives (Archives nationales, archives régionales, archives départementales, archives municipales), et les autres, qui sont détruits.

Pour chaque type de documents, la durée de ces périodes est définie par accord entre l'administration productrice et l'administration des archives (article 15 du décret n° 79-1037).

A chaque étape de leur cycle de vie, la conservation des documents d'archives publiques est contrôlée par l'administration des archives (article 2 du décret n° 79-1037).

En particulier, l'élimination de documents par un service producteur ne peut se faire sans le visa de l'administration des archives (article 16 du décret n° 79-1037 et article R. 1421-3 du code général des collectivités territoriales). De même, l'élimination de documents par un service public d'archives ne peut se faire sans le visa de l'administration productrice (article 16 du décret n° 79-1037 et article L. 212-14 du code du patrimoine).

Les services de l'Etat peuvent, dans des cas particuliers et sous certaines conditions, confier à des sociétés privées d'archivage la conservation d'archives intermédiaires qui seront détruites à terme. Cette faculté n'existe pas pour les collectivités territoriales (article L. 212-6 du code du patrimoine). Les établissements de santé peuvent faire héberger leurs dossiers médicaux électroniques dans des conditions prévues par décret.

Lors du transfert des documents d'archives dans un dépôt de préarchivage ou dans un service public d'archives, il est établi un bordereau descriptif par les soins du service qui effectue le versement (article 18 du décret n° 79-1037).

Par la suite, le service versant peut avoir à tout moment accès aux documents qu'il a versés, sauf s'il s'agit de bases de données nominatives. La consultation par le public est également possible, selon des délais définis notamment par la loi n° 78-753 du 17 juillet 1978 sur l'accès aux documents administratifs et par le code du patrimoine (articles L. 213-1 à L. 213-4).

Les archives étant définies sans distinction de date, de forme et de support, l'ensemble des règles qui précèdent s'appliquent aussi bien aux documents "traditionnels" papier qu'aux données électroniques (bases de données, documents bureautiques, documents numérisés gérés dans des systèmes de GED, documents échangés dans le cadre de téléservices, messages électroniques, etc.). Leur mise en oeuvre peut cependant différer.

3.1.1 L'application aux archives papier

Les relations entre services producteurs et services publics d'archives se déroulent actuellement, la plupart du temps, de la manière suivante.

- Les archives courantes et intermédiaires

La conservation des archives courantes et des archives intermédiaires se fait dans les locaux du service producteur (exceptionnellement dans une société privée d'archivage) ou, par anticipation, dans un service public d'archives.

Les services publics d'archives effectuent des visites régulières pour des conseils sur la tenue des dossiers et les conditions de stockage.

Pour faciliter la gestion des archives, les services producteurs et les services d'archives mettent au point des documents, appelés tableaux de tri ou tableaux de gestion (cf. exemple en annexe), qui précisent, pour chaque type de document, la durée de sa période courante, la durée de sa période intermédiaire et son sort à l'issue de la période intermédiaire (conservation ou destruction ou conservation partielle). Ces documents peuvent prendre la forme d'instructions interministérielles nationales ou/et d'accords locaux.

- Les archives définitives

A la fin de la période intermédiaire ou dans certains cas avant la fin de cette période, il est effectué le versement dans le service public d'archives des documents à conserver de manière définitive ; et la destruction, après visa du service d'archives (article 16 du décret n° 79-1037 et article R. 1421-3 du code général des collectivités territoriales), des documents sans intérêt historique.

Le versement est accompagné d'un bordereau (cf. exemple en annexe), normalisé par l'administration des archives (cf. circulaire AD 93-3 du 10 mars 1993), comportant notamment :

- l'identité du service versant (service qui transmet des documents au service d'archives)
- l'identité de l'agent responsable du versement (nom, numéro de téléphone)
- le nom du service d'archives destinataire
- la signature du chef du service versant
- le nom du service producteur (service qui a reçu ou créé les documents transmis, parfois différent du service versant)

- le volume
- les dates extrêmes
- le sort final (durée de conservation)
- une description sommaire de l'ensemble
- pour chaque boîte : numéro d'ordre, description sommaire, dates extrêmes.

Le versement s'effectue à une date convenue entre le service versant et le service d'archives.

Une fois le versement arrivé, le service d'archives effectue un contrôle du contenu des boîtes et l'archiviste appose sa signature pour indiquer qu'il prend en charge le versement. Un exemplaire du bordereau signé est adressé au service versant.

Le service d'archives intègre ensuite le contenu du bordereau de versement (informations de gestion, informations sur le contenu des documents), après l'avoir éventuellement complété (dates de communicabilité, indexation, normalisation des descriptions...), dans son système d'information.

3.1.2 La problématique des archives électroniques

L'utilisation croissante de l'informatique dans les administrations et l'évolution du cadre légal sur la valeur probante des documents électroniques conduit à une augmentation forte de la production d'archives électroniques.

En particulier, dans le cadre du développement de l'e-administration (programme ADELE coordonné par l'Agence pour le Développement de l'Administration Electronique), la plupart des informations circulant entre les administrations ou entre les citoyens et l'administration doivent être progressivement dématérialisées sous la forme de flux de données transitant par des espaces sécurisés.

Cela concerne par exemple la dématérialisation du contrôle de légalité, actuellement testée dans le département des Yvelines au moyen de la plate-forme FAST mise en œuvre par la Caisse des dépôts et consignations, la dématérialisation des marchés publics, rendue obligatoire par l'article 56 du code des marchés publics (annexé au décret n° 2004-15 du 7 janvier 2004 portant CMP), la dématérialisation de la comptabilité publique locale (programme Hélios), etc.

Mais il existe d'autres types d'archives électroniques :

- bases de données ;
- systèmes de gestion électronique de documents (GED) intégrant des images (par exemple des documents papier numérisés), des documents bureautiques ou des courriers électroniques ;
- sites intranet ;
- images, documents bureautiques ou courriers électroniques non intégrés dans un système de GED
- etc.

Même si les règles qui s'appliquent aux archives électroniques sont les mêmes que celles qui s'appliquent aux archives papier (cf. code du patrimoine, article L. 211-1), leur mise en œuvre doit naturellement être adaptée pour tenir compte des caractéristiques des données électroniques.

Les points suivants doivent notamment être pris en compte.

- La sécurité du stockage

A la différence des documents papier, les documents électroniques sont stockés sur des supports numériques dont la durée de vie est limitée et où les risques d'altération volontaire ou accidentelle sont accrus.

Une surveillance particulière doit donc être mise en oeuvre, à toutes les étapes du cycle de vie des documents.

Des migrations - copies des données numériques d'un support vers un autre de même type ou non - doivent être effectuées lorsque les supports deviennent obsolètes tout en permettant une vérification de l'intégrité et un maintien continu de la preuve de cette intégrité.

- La normalisation des formats de documents

A la différence des documents papier, lisibles immédiatement, les documents électroniques se présentent sous la forme de fichiers composés de bits, selon des formats divers, dont la lecture requiert des logiciels particuliers, qui peuvent devenir obsolètes.

Il est donc indispensable de choisir dès l'origine des formats considérés comme pérennes et d'effectuer, en temps voulu, les conversions nécessaires pour maintenir la lisibilité des données.

- La normalisation des métadonnées nécessaires à l'archivage

Comme les documents papier, les documents électroniques ne peuvent pas être conservés s'ils ne sont pas accompagnés, au moment de leur versement aux archives, d'informations descriptives, autrement appelées métadonnées.

Ces métadonnées comprennent les mêmes informations que les bordereaux de versement de documents papier, et notamment : administration versante, date de versement, description du contenu des documents, dates des documents (métadonnées fonctionnelles), communicabilité, durée de conservation, traçabilité (métadonnées de suivi).

Mais, à la différence des documents papier, il est capital de donner en outre des informations sur le format des documents versés et des indications sur l'environnement logiciel voire matériel nécessaire à la lecture et à la présentation des bits d'information (métadonnées techniques).

Par ailleurs, il est souhaitable que ce bordereau de versement sous forme électronique se présente de manière très normalisée. Il pourra ainsi accompagner les archives électroniques qui seront versées par réseau et pourra faire l'objet de traitements automatiques, notamment en vue d'être intégré dans le système d'information des archives.

Enfin, il apparaît de plus en plus que les relations qu'ont les services d'archives avec leurs partenaires (services producteurs, services versants, plates-formes de transmission, plates-formes de production, utilisateurs et demandeurs d'archives) entraînent, notamment pour les archives électroniques, la nécessité de

mettre en œuvre des formats d'échanges entre les systèmes d'information de ces différents partenaires, afin de permettre une meilleure interopérabilité entre ces différents systèmes.

3.2 La direction générale pour la modernisation de l'Etat (DGME) et le plan stratégique pour l'administration électronique (PSAE)

3.2.1 La DGME

La Direction Générale pour la Modernisation de l'Etat est rattachée au Ministère de l'Economie, des Finances et de l'Industrie. Toutefois, de par ses missions, elle a une vocation interministérielle.

Cette direction a été créée par le décret n° 2005-1792 du 30 décembre 2005, à partir du regroupement de quatre directions dédiées à la réforme de l'Etat : la délégation aux usagers et aux simplifications administratives (DUSA), la délégation à la modernisation de la gestion publique et des structures de l'Etat (DMGPSE), l'agence pour le développement de l'administration électronique (ADAE) et la direction de la réforme budgétaire (DRB).

La DGME a pour mission de « coordonner, aider et inciter, au niveau interministériel, les administrations en vue de moderniser les modes de fonctionnement et de gestion de l'Etat pour améliorer le service rendu aux usagers, contribuer à une utilisation plus performante des deniers publics et mobiliser les agents publics ». Elle pilote en particulier les audits de modernisation, le plan ADELE d'administration électronique, les lois de simplification et la politique de qualité au sein de l'Etat. Elle anime le réseau des secrétaires généraux.

La DGME est organisée en trois services : le service de la qualité et de la simplification, le service de la modernisation de la gestion publique et le service du développement de l'administration électronique (SDAE). Elle comprend également un secrétariat général (affaires générales, relations internationales, synthèse et coordination), ainsi que des départements rattachés au directeur général (communication, formation et accompagnement du changement, normes comptables).

Le SDAE a pour objectif premier de favoriser un développement rapide, cohérent et efficace de l'administration électronique au service de la modernisation de l'Etat.

Il formalise et suit les orientations de l'administration électronique. A ce titre il conçoit et pilote le Schéma Directeur et le programme ADELE, ainsi que les référentiels généraux d'interopérabilité et de sécurité des systèmes d'information. Il anime et outille le réseau des acteurs l'administration électronique pour faciliter l'expression des besoins, la circulation d'information et la mutualisation des pratiques et des moyens.

Ce département contribue ainsi globalement à la vocation de la DGME.

⇒ Pour « une modernisation qui apporte des résultats concrets » :

- Il anime des réseaux d'usagers pour rester au plus proche de leurs besoins dans le développement de l'administration électronique et favorise la mise en place de nouveaux services ;
- Il identifie et promeut de nouveaux projets notamment en cohérence avec les audits de modernisation ;

- Il renforce l'efficacité publique par la mise en cohérence et la mutualisation des investissements et du fonctionnement des systèmes d'information ;
 - Il garantit la pérennité des systèmes d'informations en assurant leur interopérabilité ;
 - Il valorise les agents par le développement des démarches collaboratives ;
- ⇒ Pour « une modernisation au service du gouvernement » :
- Il structure les orientations de développement de l'administration électronique et en assure le suivi ;
 - Il propose des nouveaux modes de travail en particulier en capitalisant et mutualisant ;
- ⇒ Pour « une approche interministérielle » et avec l'ensemble des acteurs :
- Il anime des réseaux non seulement ministériels mais de l'ensemble des acteurs sur les questions de l'administration électronique pour rester au plus proche de leur besoin, et pour les impliquer dans une approche globale ;
 - Il définit en mode collégial les cadres de référence et les modes de travail dans ces domaines.

L'archivage électronique est un thème inscrit dans le plan ADELE. Le pilote objet de ce cahier des charges est un volet de ce thème et identifié sous le l'action 103.

3.2.2 L'action ADELE 103 : Archivage électronique

Le comité interministériel de la Réforme de l'Etat du 2 février 2004 a inscrit dans le plan stratégique de l'administration électronique (projet ADELE), le point 103. Il s'agit d'une part de sensibiliser l'ensemble des acteurs de l'administration électronique à la question de l'archivage, d'autre part de réaliser avec la DGME, un référentiel normatif et enfin de renforcer la plate-forme d'archivage pour les Archives nationales et mettre à disposition des outils pour les collectivités territoriales.

La mise en œuvre d'un service d'archivage devra se faire dans un cadre normatif décliné selon cinq points :

- ⇒ le cadre juridique :
 - la valeur légale des documents électroniques
 - la législation et la réglementation propre aux archives (Code du Patrimoine, Code général des collectivités territoriales)
 - les jurisprudences en matière d'obligations d'archivage
- ⇒ le processus d'archivage :
 - les normes d'archivage
 - les règles de bonnes pratiques
- ⇒ le dispositif organisationnel supportant le processus :
 - le rôle et les responsabilités des acteurs

- les contrats ou conventions les reliant
 - leurs statuts juridiques
- ⇒ les métadonnées accompagnant le document pour donner à l'archivage le rôle qu'il est censé assurer (juridique, administratif, historique)
 - les métadonnées utiles à la consultation ultérieure
 - les métadonnées utiles à la gestion technique du document
 - les métadonnées utiles à la traçabilité du document
- ⇒ les moyens informatiques de mise en œuvre jugés sur leurs qualités techniques au regard des exigences juridiques (il s'agit de l'imputabilité, l'intégrité, la traçabilité, la fiabilité) et fonctionnelles (il s'agit de la durée de la conservation et la facilité de la consultation) de l'archivage:
 - les formats des données
 - les formats des métadonnées
 - les technologies de sécurisation
 - les supports et matériels
 - l'architecture technique
 - le centre d'exploitation ou data center (attention au lieu géographique du stockage)

L'objectif de l'action Adèle 103, « Archivage et cycle de vie du document électronique », est double :

- ⇒ produire un référentiel reprenant les points cités auparavant
- ⇒ réaliser une plate-forme pilote d'archivage électronique, dont le cahier des charges est le présent document.

3.3 Le département du système d'information (DSI) du ministère de la culture et de la communication et le schéma directeur du système d'information (SDSI)

Le DSI et le SDSI (source : SDSI sur Sémaphore)

3.3.1 Le DSI (source : présentation du DSI sur Sémaphore)

Relevant de la Direction de l'administration générale du Ministère de la Culture et de la Communication, le Département des Systèmes d'Information a la responsabilité des actions menées par le ministère dans les domaines de l'informatique et des télécommunications, qu'il s'agisse de l'infrastructure, de l'équipement matériel et logiciel (pour les services d'administration centrale), de la gestion et de l'administration des bases de données, de la réalisation d'applications métier.

Missions du Département des Systèmes d'Information

- Assurer le suivi du schéma directeur informatique du Ministère

- Organiser la fonction informatique du Ministère
- Renforcer le dialogue avec les maîtrises d'ouvrage des Directions
- Harmoniser les pratiques par une animation régulière des réseaux, et le partage de connaissances entre les régions et les départements, les correspondants informatiques
- Participer activement en tant que coordinateur aux projets inter ministériels
- Veiller à l'intégrité et la sécurité du système d'information du Ministère par la mise en œuvre du schéma directeur de la sécurité approuvé par le Cabinet tout début 2004.

Accompagner le Ministère dans ses projets

- Mise en place de la LOLF ("loi organique relative aux lois de finances") dans le système d'information du Ministère, et cela dans les délais exigés (1er janvier 2006)
- Diffusion des données patrimoniales étant une des missions essentielles du MCC, réfléchir à la nouvelle architecture des systèmes documentaires de façon à la rendre cohérente, harmonisée et capable, au delà de ses aspects scientifiques, et fournir une aide efficace à la décision lors de l'évaluation des politiques culturelles
- Proposer, de concert avec les Directions sectorielles, des modes de fonctionnement novateurs, largement basés sur les outils informatiques, permettant à la décentralisation et à la déconcentration de s'opérer pleinement
- Tenir compte et tirer partie de toutes les actions interministérielles qui visent à rationaliser les moyens que l'État consacre à ses systèmes d'information, le DSI devra participer activement au programme ADELE lancé par la DGME.

3.3.2 Le Schéma directeur du système d'information (SDSI) du ministère de la Culture

Le ministère de la Culture s'est doté, à la fin de l'année 2005, d'un schéma directeur des systèmes d'information pour la période 2006-2008.

Le ministère s'est fixé trois orientations stratégiques qui ont présidé à l'ensemble des travaux sur le schéma directeur :

- mettre en oeuvre la nouvelle Loi Organique sur les Lois de Finance et évoluer d'une culture de moyens vers une logique de résultats ;
- rendre l'administration plus accessible aux usagers ainsi qu'aux agents en tirant le meilleur parti possible des apports des nouvelles technologies ;
- améliorer notre offre culturelle et le porter à connaissance de nos richesses patrimoniales, en adaptant notre offre aux diverses catégories de publics.

La déclinaison de ces orientations stratégiques en des objectifs tangibles et mesurables a permis de dégager 12 objectifs prioritaires.

De plus, des projets directionnels qui affecteront l'ensemble du ministère ont donc donné lieu à 5 objectifs structurants.

L'archivage électronique est un de ces objectifs structurants (ST1 : "pérenniser les données gérées sous forme électronique pour l'ensemble des ministères").

Le projet ST1.04 concerne spécifiquement la mise en place d'une plate-forme d'archivage électronique.

Cet objectif porte sur la mise en oeuvre des recommandations de la DGME - SDAE (ADELE 103) relatives à l'archivage électronique des Ministères. Le CAC de Fontainebleau a pour mission de collecter les archives des administrations et en particulier des MAN (Missions des Archives Nationales) de tous les Ministères. Celles-ci ont pour rôle de collecter tous les documents relevant des archives administratives contemporaines, et a fortiori des documents dématérialisés électroniques provenant de tous les services.

A l'heure où la production de documents devrait être (ce n'est pas encore une réalité hormis pour les marchés publics ou le JO électronique) en très grande partie nativement électronique, il devient essentiel de pouvoir pérenniser ces données, en organisant la collecte, le stockage et la restitution de ces archives entre les services producteurs, les MAN et le CAC (à terme, le centre de Pierrefitte).

Cet objectif opérationnel est **prioritaire** à deux égards :

- il entre dans le cadre des recommandations de la DGME - SDAE,
- il est la clé de voûte des futurs développements des systèmes des Archives Nationales pour l'archivage électronique.

Il se traduit concrètement par la mise en place d'une solution informatisée appropriée et la cible proposée est donc la suivante : **Une plate-forme d'archivage électronique est en place pour gérer les archives électroniques du MCC et une plate-forme interministérielle permet d'accueillir les archives électroniques de tous les Ministères. "**

Il est prévu une étude préalable jusqu'à l'automne 2006 puis une mise en oeuvre à partir de l'automne 2006 et jusqu'à fin 2008.

Une plate-forme opérationnelle devra être prête à l'ouverture du nouveau centre des Archives nationales.

3.3.3 Le schéma directeur de la sécurité des systèmes d'information (SDSSI) du ministère de la culture

En matière de sécurité des systèmes d'information le SDSSI a été adopté le 16 mars 2004.

En résumé :

- les impacts ont été hiérarchisés (page 16 du schéma directeur),
- des niveaux de sécurité ont été définis (pages 22 du schéma directeur),

- par grands types d'informations, un niveau maximal de sécurité a été attribué (page 23 du schéma directeur),
- des objectifs de sécurité en ont dérivé (pages 24 à 28 du schéma directeur) ; ils concernent le personnel, l'organisation et la technique,

Il n'existe actuellement aucune politique de la sécurité des systèmes d'information (PSSI). La rédaction d'un tel document prévue pour fin 2007.

4. Référentiels, normes et standards à prendre en compte pour le pilote

4.1 Documents de référence

4.1.1 Norme ISO 14721 (modèle OAIS)

La norme ISO 14721:2003 (Systèmes de transfert des informations et données spatiales -- Système ouvert d'archivage de l'information -- Modèle de référence), plus connue sous le nom de modèle OAIS (Open Archival Information System) est consultable à l'adresse suivante :

["http://www.ccsds.org/CCSDS/documents/650x0b1.pdf"](http://www.ccsds.org/CCSDS/documents/650x0b1.pdf) <http://www.ccsds.org/CCSDS/documents/650x0b1.pdf>.

Une traduction française, en cours de normalisation, est accessible à l'adresse suivante :

["http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf"](http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf)http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf.

Cette norme conceptuelle, mise au point par les principaux centres d'études spatiales du monde dont le CNES (Centre National d'Etudes Spatiales), définit les objets d'information, les métadonnées nécessaires à leur préservation et l'organisation à mettre en place pour leur archivage, leur conservation et leur communication.

Tout versement d'information à un OAIS (service d'archives) par un Producteur, toute diffusion d'information auprès d'un Utilisateur, a lieu sous forme d'une ou de plusieurs sessions de transmissions distinctes. Il est donc utile de définir le concept de Paquet d'informations (Information Package).

Un Paquet d'informations est un conteneur conceptuel de deux types d'informations appelés Contenu d'information (Content Information) et Information de pérennisation (Preservation Description Information ou PDI).

Le Contenu d'information et le PDI sont identifiés et encapsulés par une Information d'emballage (Packaging Information). Le paquet qui en résulte peut être retrouvé grâce à l'Information de description (Descriptive Information).

Le Contenu d'information est l'information qui constitue la cible originale de la pérennisation. Il est constitué par l'Objet-contenu de données (Objet physique ou Objet numérique, c'est-à-dire les bits) (Content Data Object) et son Information de représentation (Representation Information), nécessaire à la compréhension de cet objet. L'Information de pérennisation s'applique au Contenu d'information. Elle est requise pour conserver le Contenu d'information, assurer qu'il est clairement identifié, et appréhender l'environnement de création du Contenu d'information. L'Information de pérennisation se subdivise en quatre catégories d'informations : provenance, contexte, identification, et intégrité.

L'Information d'emballage est l'information qui, réellement ou de façon logique, assemble, identifie et met en relation Contenu d'information et PDI.

L'Information de description est l'information qui est utilisée pour identifier le paquet dont le Contenu d'information est intéressant. En fonction du contexte, il peut s'agir d'un simple titre descriptif du Paquet d'informations apparaissant dans un libellé, ou bien d'un jeu complet d'attributs pour effectuer une recherche dans un catalogue. " Délibérations 2004 de la commune de Versailles " est un exemple.

Exemple : Délibérations transmises par les collectivités aux préfetures pour le contrôle de légalité

- objet-contenu : fichiers PDF¹ correspondant aux délibérations transmises et les informations de signature éventuelle associées
- information de représentation : indication du format PDF ou base64 (la documentation technique d'utilisation et de compréhension de ces formats se trouve dans la base de connaissance)
- information de pérennisation : informations générales relatives aux délibérations archivées (collectivité émettrice, nom de l'utilisateur ayant réalisé la transmission, référence de la délibération transmise, ...)
- information d'empaquetage : informations sur la transmission (empreinte de transmission, date, référence de la transmission, ...)
- information descriptive : données utilisées pour identifier une délibération (date de la délibération, objet, collectivité émettrice, ...)

L'empreinte du fichier fait partie de l'information d'intégrité (l'empreinte est le résultat d'un algorithme informatique de hachage ; la recalculer une deuxième fois permet de vérifier l'égalité des deux empreintes et atteste de la non altération du document).

4.1.2 La DTD EAD (Encoded Archival Description)

La DTD EAD version 2002 est consultable à l'adresse ["http://www.loc.gov/ead/"](http://www.loc.gov/ead/) <http://www.loc.gov/ead/>. Une version française du dictionnaire des balises est accessible à l'adresse suivante : http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/EAD%202002_Complet_20040930.pdf.

La DTD EAD offre un cadre pour la description des documents d'archives.

Un fichier XML suivant la DTD EAD est organisé en trois sections :

- un en-tête (<eadheader>) ;
- une page de titre (<frontmatter>), facultative ;
- une section de description des archives (<archdesc>).

La section <archdesc>, qui constitue le corps du fichier, peut elle-même être subdivisée en plusieurs niveaux de descriptions (fonds, série, dossier, pièce).

Par principe, les informations d'un niveau supérieur valent pour tous les niveaux inférieurs et ne doivent y être répétées.

¹ En cas d'encapsulation dans un fichier XML, le document PDF est encodé en base 64.

La DTD EAD permet une mise en oeuvre de la norme conceptuelle ISAD(G) - Norme générale et internationale de description archivistique
["http://www.ica.org/biblio/isad_g_2f.pdf"](http://www.ica.org/biblio/isad_g_2f.pdf)http://www.ica.org/biblio/isad_g_2f.pdf

La DTD EAD est connue de tous les services publics d'archives et de plus en plus utilisée pour la production et la diffusion des instruments de recherche.

Elle a été publiée dans le répertoire des schémas des administrations, accessible à l'adresse suivante
["http://www.adae.gouv.fr/IMG/rtf/repertoire_schemas_xml_version_1_juin.rtf"](http://www.adae.gouv.fr/IMG/rtf/repertoire_schemas_xml_version_1_juin.rtf)
http://www.adae.gouv.fr/IMG/rtf/repertoire_schemas_xml_version_1_juin.rtf.

4.1.3 La politique d'archivage sécurisé dans la sphère publique

Une politique d'archivage type (PA) pour le secteur public a été ainsi élaboré : l'archivage électronique, objet de cette Politique d'archivage type (PA), tend à conserver l'information en la restituant de manière intégrée et conforme à l'information d'origine. Cette opération visant à conserver des Archives ayant une force probante et des effets juridiques concerne toutes les personnes juridiques sans exception, qu'elles soient physiques, morales, privées ou publiques.

La PA fixe par conséquent les obligations que doivent remplir les autorités d'archivage (AA). Il s'agit par exemple de entités qui prennent la responsabilité du processus d'archivage que ce soit dans les administrations centrales, les administrations déconcentrées, collectivités territoriales, collectivités locales, personnes privées chargées d'une mission de service public. Ces responsables changent suivant le cycle de vie de l'archive. Il peut s'agir par exemple d'un service producteur tant que l'archive est courante, puis d'un service d'archives intermédiaires dès lors que le producteur lui verse ses archives et que le service les prend en charge, puis d'un service d'archives public qui, à son tour, prend en charge les archives définitives transférées par le service d'archives intermédiaires.

La PA Type définit les exigences minimales, en termes juridiques, fonctionnels, opérationnels, techniques et de sécurité, qu'une autorité d'archivage doit respecter afin que l'archivage électronique mis en place puisse être regardé comme fiable. Cette PA Type repose sur des contraintes " standard " à mettre en place. Il en est ainsi :

- des contraintes en matière d'identification/authentification de l'origine de l'Archive ;
- de l'intégrité des Archives, des Paquets d'informations et des Objets d'archives ;
- de l'intelligibilité / lisibilité des Archives ;
- de la durée de conservation de l'Objet d'archives ;
- de la traçabilité des différentes opérations (notamment versement, consultation, élimination)
- de la disponibilité et de l'accessibilité des Archives.

La PA Type constitue donc un référentiel de la sécurité de l'archivage électronique pour qu'il puisse être qualifié de « fiable ». Une grille d'audit constituée à partir de ses différents chapitres permet par ailleurs à un auditeur de contrôler la fiabilité d'un service d'archivage électronique. La PA s'accompagne également d'un modèle de cahier des charges pour la mise en œuvre d'un service d'archivage électronique.

4.2 Les documents applicables

4.2.1 Le standard d'échange de données pour l'archivage électronique (versement, communication, élimination)

Ce standard est actuellement publié sur le site de la direction générale de la modernisation de l'Etat (DGME).

Il est annexé au présent cahier des charges. **Il devra être implémenté pour un certain nombre des échanges qu'il couvre, dans le cadre de ce pilote.**

Contexte

Les services d'archives entretiennent de nombreux échanges :

- avec les services producteurs, qui transfèrent des documents et des données devant être conservés ;
- avec des demandeurs, qui souhaitent accéder aux archives ;
- avec d'autres services d'archives, selon le cycle de vie des documents ou en vue de changer de lieu de stockage ;
- avec des services de contrôle, qui sont amenés à donner des autorisations de communication ou de destruction.

Ces échanges concernent encore le plus souvent des documents papier et font généralement l'objet de formulaires traditionnels.

Or, les documents et les données échangés sont de plus en plus souvent électroniques. Les producteurs référencent leurs dossiers papier dans des bases de données, voire gèrent ces dossiers sous forme entièrement dématérialisée. Les services d'archives, de leur côté, disposent de plus en plus souvent d'outils de gestion informatisée pour enregistrer, stocker et communiquer les dossiers qu'ils conservent. Les demandeurs, enfin, souhaitent pouvoir accéder en réseau aux informations qui les intéressent.

Il existe un besoin très fort de connecter entre eux ces divers systèmes d'information, afin d'éviter les ruptures de chaîne, qui occasionnent des coûts supplémentaires et une perte de qualité de l'information transmise. Par exemple, il serait souhaitable que les données descriptives fournies par les services producteurs à l'occasion d'un versement puissent être intégrées automatiquement dans le système d'information du service d'archives.

Le standard d'échange de données pour l'archivage, proposé ici, vise à répondre à cet enjeu.

Le standard d'échange de données pour l'archivage

Le standard d'échange de données a été mis au point par la direction des archives de France (ministère de la Culture et de la communication) et la direction générale de la modernisation de l'Etat (ministère de l'Economie, des finances et de l'industrie).

Le standard d'échange de données pour l'archivage fournit un modèle pour les différentes transactions spécifiques qui interviennent entre un service d'archives et ses partenaires :

- demande de transfert ;
- transfert ;
- communication ;
- avis de modification ;
- élimination à la demande du service producteur ;
- élimination à la demande du service d'archives ;
- restitution.

Chaque transaction fait l'objet de plusieurs messages. Par exemple, pour le transfert, interviennent successivement le transfert des données proprement dit (composé d'un en-tête, d'une description des données et des données elles-mêmes), un accusé de réception, une notification d'acceptation ou un avis d'anomalie, et enfin, si nécessaire, un accusé de réception d'avis d'anomalie.

Chaque message prend la forme d'un flux XML, conforme aux recommandations de l'UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business – <http://www.unece.org/cefact/>).

Utilisation

Le standard d'échange de données pour l'archivage est destiné à être mis en oeuvre par :

- les producteurs d'archives ou entités versantes (dont les plateformes d'échanges telles que FAST ou achatpublic.com) ;
- les services d'archives, publics ou privés en liaison avec leurs services informatiques
- les tiers-archivistes ;
- les éditeurs de logiciels de gestion d'archives, dont les outils doivent pouvoir accueillir automatiquement des données versées conformes au standard ;
- les éditeurs de logiciels sectoriels, dont les outils doivent pouvoir exporter automatiquement des données à archiver conformes au standard.

Il est conçu pour s'adapter aux archives électroniques ou aux archives papier.

Il devrait également pouvoir être utilisé par le secteur public et le secteur privé.

D'ores et déjà, plusieurs projets de mise en oeuvre ont débuté : pour le versement de délibérations soumises au contrôle de légalité dans un service d'archives départementales, pour le versement de dossiers de marchés publics dans un autre service d'archives départementales.

4.2.2 Guide UML XML élaboré par la DGME

Ce guide méthodologique concerne la génération de schémas XML d'échanges de données entre plusieurs systèmes d'information à partir des modèles UML appropriés (Version 1.1, Juillet 2005). Ce guide est en cours de révision afin de le rendre compatible avec les dernières recommandations internationales de standardisation.

Il est disponible à l'adresse suivante :

http://www.adae.gouv.fr/article.php3?id_article=901

4.2.3 PRIS version 1

Afin de fédérer et d'homogénéiser ces travaux, avec pour objectif de permettre à l'ensemble des acteurs publics ou privés d'en bénéficier, de donner aux acteurs économiques une vision claire des attentes de l'Administration et de proposer aux usagers des outils de confiance, l'État a décidé de définir et de rendre public un ensemble documentaire unique, baptisé PRI, « Politique de Référencement Intersectorielle », destiné à fournir un cadre aux autorités de certification (partie « offre ») et aux promoteurs d'applications (partie « demande »), publics ou privés. Cette première version de la PRI a été élaborée conjointement par le Ministère de l'Économie, des Finances et de l'Industrie, le GIP-MDS, représentant la sphère sociale, la DCSSI et la DGME.

La PRIS est composée de deux sections, la section « entreprises » (A) et la section « individu » (B). Trois niveaux de sécurité sont définis pour chaque section : les niveaux moyen, fort et qualifié (également dénommés * ; ** et ***).

Cette première version de la PRIS inclut la politique de référencement du MINEFI (PC Entreprise V3.1) actuellement en vigueur, en tant que niveau fort de la section « entreprises », elle comporte également les niveaux fort (**) et qualifié (***) pour la section « individus ». Elle sera complétée progressivement, et notamment dès sa version 2004 avec le niveau moyen (*) pour la section « individu ».

La PRIS version 1 est disponible à l'adresse suivante : http://www.adae.gouv.fr/article.php3?id_article=220

4.2.4 Le cadre commun d'interopérabilité version 2.1

Une deuxième version du cadre commun d'interopérabilité a été élaborée et publiée en février 2003 pour répondre à la nécessité d'une interopérabilité accrue entre les systèmes d'information publics, en consolidant les bases nécessaires pour garantir une collaboration efficace au sein des collectivités publiques. L'objectif est de renforcer la cohérence entre les systèmes d'information, et de favoriser le partenariat entre l'Etat et les collectivités territoriales, notamment dans le domaine des services en ligne offerts notamment aux citoyens ou aux entreprises.

Ce document est disponible à l'adresse suivante : http://www.adae.gouv.fr/article.php3?id_article=219

5. Périmètre

5.1 Présentation générale

La prestation qui fait l'objet du présent cahier des charges vise la réalisation d'une plate-forme d'archivage électronique, c'est-à-dire une infrastructure matérielle et logicielle comportant un ensemble complet de fonctionnalités de transfert, de réception, de contrôle, de stockage réparti, de communication et d'élimination des données.

Les objectifs de la plate-forme pilote qui sera mise en place sont de :

- Réaliser le noyau générique de la plate-forme d'archivage électronique, modèle réutilisable pour les spécifications d'une plate-forme de production réelle (à usage des services producteurs Etat et collectivités territoriales, et du réseau des archives ainsi que du nouveau centre de Pierrefitte-sur-Seine) : réception des archives, contrôle de ces archives, écriture sur plusieurs sites distincts, extraction des métadonnées vers une base de données descriptives consultables en ligne (réseau étendu) , communication des Archives tant aux producteurs que, plus généralement, dans le respect des délais légaux de communicabilité, aux usagers, élimination éventuelle des archives ;
- Mettre en oeuvre les principes énoncés dans les référentiels, en conformité avec le SDSI du MCC ;
- Établir une base de connaissance sur la conduite de ce type de projet (difficultés rencontrées, solutions trouvées) ;
- Prouver la pertinence de la solution en faisant fonctionner des situations réelles ;
- Profiter du pilote pour renforcer l'équipement technique du CAC.

La plate-forme d'archivage sera alimentée par des données en provenance de divers services versants, mais en nombre et volume limités.

L'extraction de ces données à partir des applications sources est hors du champ de la prestation qui fait l'objet du présent cahier des charges.

La mise au format spécifié par le standard d'échange sera faite par l'intermédiaire du pilote à l'exception du contrôle de légalité conservé en amont sur la plate-forme FAST et des dossiers de marchés conservés sur la plate-forme achatpublic.com (déjà formatés au moment du transfert).

En revanche, le transfert de ces données à la plate-forme d'archivage électronique, via le standard d'échange de données, fait partie de la prestation.

Le pilote est une expérimentation à portée générale, il couvre donc les services de l'Etat mais aussi les collectivités territoriales. Il permet de recevoir, gérer et communiquer des archives tant intermédiaires que définitives.

Le pilote sera administré sur le site du centre des archives contemporaines de Fontainebleau et exploité par le DSI du ministère de la culture.

Conformément aux préconisations des référentiels, un test de réplique et de sauvegarde sera réalisé avec le centre des archives contemporaines (CAC).

Le Conseil général des Yvelines a prévu de développer de son côté, dans des délais brefs, une application de réception et de consultation des délibérations soumises au contrôle de légalité et du sommier de l'Etablissement public d'aménagement de la ville nouvelle de Saint-Quentin-en-Yvelines (base de données de gestion foncière). Les enseignements de cette expérience seront utiles pour le projet.

En revanche, de par sa vocation de pilote, tous les composants développés doivent pouvoir être réutilisables ultérieurement dans le système d'information opérationnel de la plate-forme d'archivage électronique qui sera développée dans le nouveau centre de Pierrefitte-sur-Seine (2008-2010), qu'il s'agisse des composants fonctionnels ou des composants technologiques.

En effet, à terme, la plate-forme d'archivage électronique qui sera développée pour le centre de Pierrefitte-sur-Seine devra être compatible avec le système d'information archivistique qui est en cours de conception, dans le cadre de ce projet, pour assurer les tâches archivistiques dans les trois sites franciliens des Archives nationales.

Constitué d'un outil de gestion de la chaîne archivistique, d'un outil de description des fonds et de production d'instruments de recherche et d'une salle des inventaires virtuelle, le système devra permettre l'échange d'informations à la fois avec le réseau interministériel des Missions des Archives nationales et des services d'archives des ministères, et avec le grand public. Il sera accessible par un portail Internet.

Les archives électroniques relevant de la compétence des Archives nationales devront pouvoir être référencées dans ce système d'information.

Le pilote devra implémenter le standard d'échange des données pour l'archivage, pour un certain nombre de messages et transactions : demandes de transfert et transfert, communication (à l'exception de l'intervention des services de contrôle), éliminations à la demande du service d'archives (à l'exception de l'intervention des services de contrôle), l'avis de modification. Ne sont pas pris en compte les messages du standards d'échange suivants : éliminations à la demande du service producteur, restitution d'archives.

5.2 Sources de données à prendre en compte

Le pilote devra permettre de tester l'archivage de plusieurs types de données, dont les durées de conservation sont variables (de quelques années de conservation à une durée de conservation illimitée) :

⇒ Documents issus d'applications métier :

- Marchés publics (provenant de la plate-forme achatpublic.com) : 1 000 consultations

[Pour l'année 2005, il y a eu 7000 consultations sur la plate-forme marchés publics.fr (pour seulement 100 réponses électroniques soit 100 Go environ].

La nature des fichiers sont dans le tableau ci-dessous :

Nature	Type ou extension
AVIS	PDF

DAC, DCE	ZIP
RECOMMANDES	ZIP
PREUVES	PEP ; PDR ; PER ; PCR ; POP ; PDE ; PDP
REGISTRES	CSV ; PDF
SIGNATURES	SIG
JOURNAL DES EVENEMENTS	PDF

- Délibérations transmises par le contrôle de légalité (provenant de la plate-forme FAST expérimentée au Conseil général des Yvelines) : 1 000 transmissions (900 Ko en moyenne par transmission) soit 900 MO
- Allocations personnalisées à l'autonomie (APA) (provenant de la plate-forme FAST expérimentée au Conseil général des Yvelines) : 500 dossiers (environ 10 pages de formulaires par dossier)

Bases de données :

Le service central des études économiques et statistiques du ministère de l'agriculture (SCEES)

Trois enquêtes agricoles seront retenues.

Chaque enquête comporte un fichier de données et les fichiers de métadonnées techniques et de contexte liées au fichier et à l'enquête.

Un fichier de données moyen : 30 MO

Métadonnées liées au fichier : 15 MO

Documents liés à l'enquête (le contexte) : 15 MO

On peut estimer de 50 à 100 MO l'ensemble des fichiers d'une enquête de taille moyenne.

Le versement (ou transfert) sera composé de 12 à 20 fichiers, avec des formats différents (txt, pdf, tiff, ascii).

Ministère de l'Education Nationale : base nationale des élèves du premier degré.

Il s'agit ici d'une base de données vivante ; les versements seront annuels et concerneront les élèves sortant du premier cycle.

Le premier versement sera quasiment similaire à ceux du SCEES dans sa structuration et ses composants : un fichier de données, des fichiers de métadonnées techniques liées au fichier, des fichiers de métadonnées liées à l'application (contexte). Le versement de l'année suivante pourrait ne comporter que le fichier de données si le contexte technique n'a pas bougé.

La première année, on peut estimer la volumétrie du transfert à :

Les données : 25 MO

Les métadonnées : entre 30 et 100 MO, selon la forme de certaines métadonnées nativement numériques, numérisées ou papier.

Soit une volumétrie de 60 à 130 MO.

Sommier de l'Etablissement public d'aménagement de la ville nouvelle de Saint-Quentin-en-Yvelines :

22 MO (22214 KO précisément). Il est découpé en 4 fichiers (parcelle, affectation, sommier, lot)

Messageries

de l'Agence française de la sécurité alimentaire des aliments (AFSSA)

2 services concernés, soit six personnes (dont un de cinq personnes).

Après filtrage par les personnes, l'estimation est la suivante :

4 messages / jour/ personne

soit 25 messages / semaine/personne

soit 150 messages/ semaine

soit 150 messages X 50 semaines = 7500 messages/an

si l'on compte le poids moyen d'un message à 4 ko, on obtient :

$7500 \times 4 \text{ ko} = 30 \text{ Mo / an}$

Avec les annexes (documents joints)

Si on considère que l'on a une annexe pour 5 messages, en moyenne, on obtient 1200 pièces / an

Une pièce variera entre 5 KO et 1 MO

Si on prend une moyenne raisonnable de 100 KO / pièce, le poids des pièces jointes est de :

$100 \text{ KO} \times 1200 = 120 \text{ MO / an}$

Soit au total et approximativement :

$120 \text{ MO} + 30 \text{ MO} = 150 \text{ MO / an}$

- du Cabinet du Premier Ministre

⇒ Gestion électronique de documents/bureautiques

- Rapports du conseil général des ponts et chaussées

La production annuelle ne dépasse pas 300 rapports.

Un rapport à une volumétrie variant entre 30 KO et 150 KO ;

Si l'on prend une valeur moyenne de 90 KO par rapport, on obtient :

$$90 \times 300 = 27 \text{ MO / an}$$

S'y ajoute les données du répertoire détaillé qui forment un fichier distinct des autres, soit un volume annuel probable de 30 MO

Les volumétries indiquées ci-dessus le sont à titre indicatif et pourront dans le cadre du pilote, évoluer. Toutefois, les volumes seront réduits et limités à une capacité de **2 TO sur chacun des sites** (voir partie stockage) et par conséquent, les performances du pilote en terme de réception et de stockage seront limitées. Ceci étant, il doit pouvoir, ultérieurement, évoluer, afin de supporter une montée en charge importante (transmissions automatiques par réseaux, en flux tendus ; stockage de volumétries fortes). Le pilote sert ici à tester les fonctionnalités d'une plate-forme d'archivage électronique, hors capacités de performance liées à l'architecture réseaux.

Les données ne seront pas à extraire des applications et bases de données d'origine mais auront déjà faits l'objet d'un export.

Leur mise au format dans celui spécifié dans le standard d'échange sera à faire par le pilote soit en amont soit en aval du versement. Le paramétrage de ce format, sur la base du standard d'échange, notamment pour la partie " Description ", sera précisément déterminé et finalisé, application par application, lors de la phase de conception détaillée.

5.3 Les acteurs du pilote

La DAF (DITN) en collaboration avec la DGME assure la coordination entre les différents acteurs.

Ils seront, pour le pilote, en tant qu'acteurs et utilisateurs du pilote, entre 20 et 25 personnes.

5.3.1 La DGME

Concernant les marchés publics versés depuis la plate-forme achatpublic.com, c'est la DGME en tant que maître d'ouvrage de la dématérialisation des marchés publics, qui participe au contrôle des dossiers archivés suivant le format spécifié par le standard de versement des archives électroniques, qui teste les fonctionnalités liées à leur consultation/communication.

5.3.2 Les missions des archives nationales auprès des ministères (Premier ministre, Education et Equipement)

Elles reçoivent et gèrent, avant prise en charge par le pilote, les objets à archiver suivants : messageries Cabinet Premier ministre, messageries de l'Agence française de la sécurité alimentaire des aliments (AFSSA), rapports du CGPC

Elles mettent au format spécifié par le standard de versement des archives électroniques à travers le pilote, et testent les fonctionnalités liées à leur consultation/communication.

5.3.3 Achatpublic.com

La plate-forme achatpublic.com permettra l'export des données de dossiers de marchés suivant le standard de versement des archives électroniques. Elle joue par conséquent le rôle de service versant.

5.3.4 Le Conseil général des Yvelines – la Direction des Archives Départementales et la DSI

La direction des archives départementales reçoit, sur sa plate-forme d'archivage électronique gérée par la Direction des systèmes d'information, depuis la plate-forme FAST et déjà formaté suivant le standard de versement des archives électroniques, les objets à archiver pour le contrôle de légalité et l'allocation personnalisée à l'autonomie. Les transferts depuis FAST vers cette plate-forme seront également repris par le pilote. De même, le pilote intégrera le versement déjà effectué aux archives départementales des Yvelines, du sommier de l'Etablissement public d'aménagement de la ville nouvelle de Saint-Quentin-en-Yvelines .

5.3.5 Le centre des archives contemporaines de Fontainebleau (service Constance)

En tant que service participant à l'administration du pilote, le service Constance participe au contrôle des objets transférés, contrôle l'extraction des métadonnées vers la base de données descriptives, contrôle l'écriture des objets versés sur les supports de conservation et plus généralement l'infrastructure de stockage ainsi mise en place, teste les fonctions de consultation/communication des objets archivés

5.3.6 Le département des systèmes d'information (DSI) du ministère de la Culture

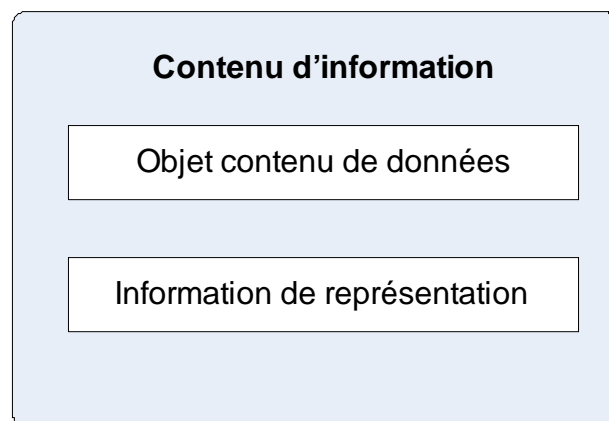
Le DSI assure l'exploitation technique du pilote réalisé.

6. Présentation des fonctionnalités du pilote

6.1 Le modèle OAIS

6.1.1 Contenu d'information

Si l'on se réfère au modèle OAIS (voir plus haut), un contenu d'information (Content Information) est un ensemble d'informations constituant l'objet principal de la pérennisation dévolue au SAE. Il est composé d'un objet contenu de données (Content Data Object) et de son information de représentation (Representation Information).



Un objet contenu de données peut être un objet physique ou un objet numérique sachant que ne sont traités ici que des objets numériques. Un objet numérique (Digital Object) est un objet constitué d'une suite de bits qui prend la forme d'un fichier électronique généré dans un format donné (par exemple un format image ou un format texte).

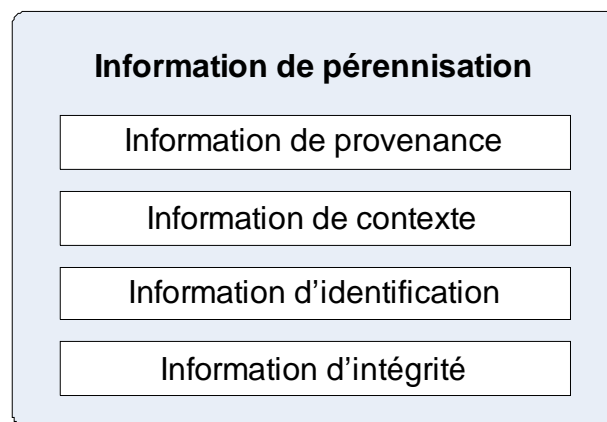
L'information de représentation (Représentation Information) est l'information qui traduit un objet contenu en des concepts plus explicites. Il pourra s'agir par exemple de la définition et de la description du format image dans lequel a été généré le fichier et qui permettra de convertir la séquence de bits dont il se compose sous une forme intelligible par l'utilisateur. Cette information de représentation peut soit être fournie par le service versant avec l'objet contenu de

données, soit être gérée séparément par le service d'archives dans une base de connaissances. Dans ce dernier cas le service d'archives a la charge de contrôler, lors des versements, l'existence de la documentation correspondante dans sa base de connaissances.

6.1.2 Information de pérennisation

Afin qu'un contenu d'information puisse être correctement conservé, il doit être accompagné d'une information de pérennisation (Preservation Description Information - PDI) qui se décompose de la façon suivante :

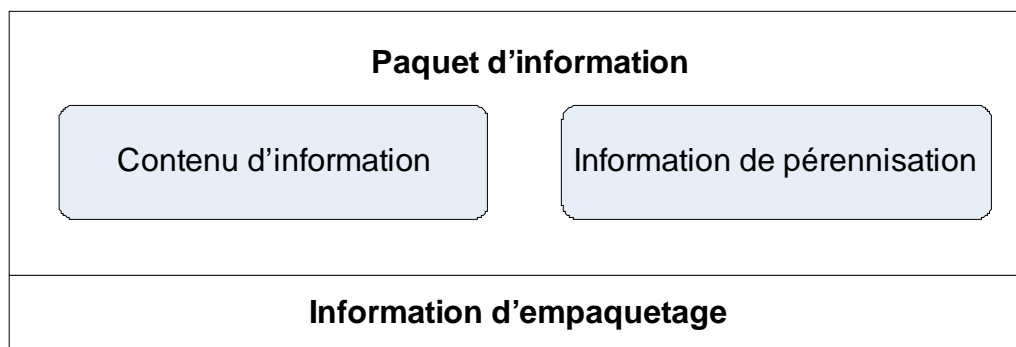
- Information de provenance (Provenance Information) : information qui documente l'historique du contenu d'information. Cette information renseigne sur l'origine ou la source du contenu d'information, sur toute modification intervenue depuis sa création et sur ceux qui en ont eu la responsabilité. Exemple : nom du principal responsable de l'enregistrement des données, informations relatives au stockage, à la manipulation et à la migration des données.
- Information de contexte (Context Information) : information qui décrit les liens entre un contenu d'information et son environnement. Elle inclut entre autres les raisons de la création de ce contenu d'information et son rapport avec d'autres objets contenu de données.
- Information d'identification (Reference Information) : information qui identifie et si nécessaire décrit le ou les mécanismes d'attribution des identificateurs au contenu d'information. Elle inclut aussi les identificateurs qui permettent à un système externe de se référer sans équivoque à un contenu d'information particulier. Exemple : un ISBN (International Standard Book Number).
- Information d'intégrité (Fixity Information) : description des mécanismes et des clés d'authentification garantissant que le contenu d'information n'a pas subi de modification sans que celle-ci ait été tracée. Par exemple, le code CRC (contrôle de redondance cyclique) pour un fichier ou mieux le calcul d'empreinte.



6.1.3 Définition d'un paquet d'information (ou lot)

D'après l'OAIS, l'ensemble des échanges d'informations effectués entre le système d'archivage et l'extérieur s'effectue par l'intermédiaire de paquets d'informations.

Un paquet d'informations (Information Package IP) est l'association du Contenu d'information et de son Information de pérennisation (PDI). A ce paquet d'informations est aussi associée une Information d'empaquetage qui permet de relier et d'identifier les composants d'un Paquet d'informations.



On distingue ainsi trois types de paquets :

- Les paquets d'informations à verser (Submission Information Package - SIP) : Paquet d'informations livré par le service producteur ou service versant au système d'archivage pour l'élaboration d'un ou plusieurs Paquets d'informations archivés (AIP).
- Les paquets d'informations archivés (Archival Information Package - AIP) : Paquet d'informations conservé dans le système d'archivage et constitué d'un Contenu d'information et de l'Information de pérennisation (PDI) associée.
- Les paquets d'informations diffusés (Dissemination Information Package - DIP) : Paquet d'informations reçu par l'Utilisateur en réponse à sa requête au système d'archivage. Ce paquet provient d'un ou de plusieurs Paquets d'informations archivés (AIP).

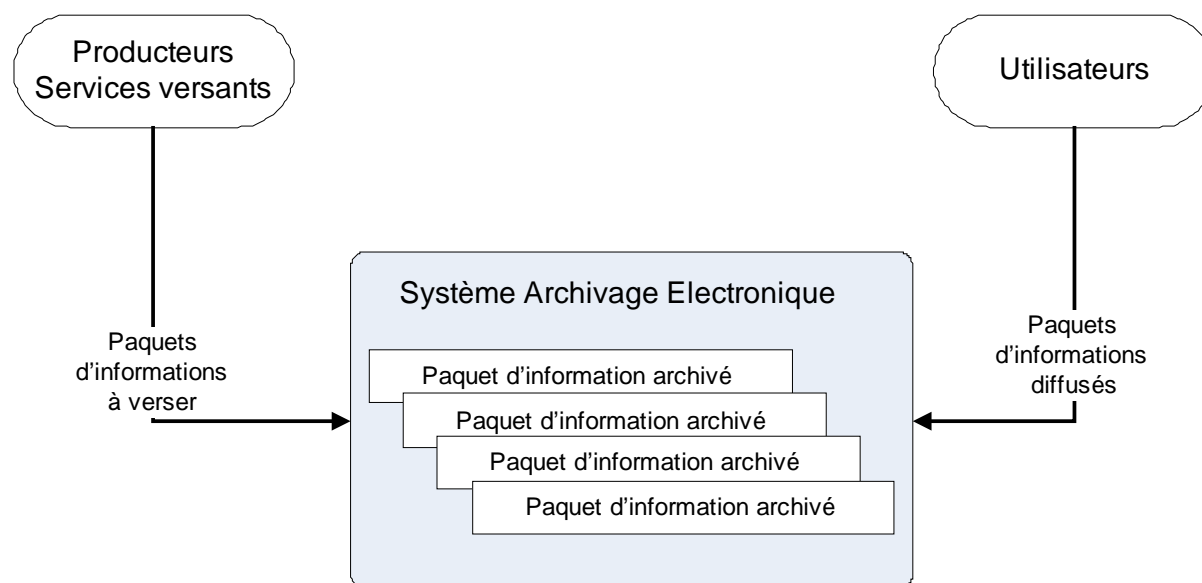
6.1.4 Information de description

Enfin, l'Information de description (Descriptive Information) est un ensemble d'informations, extraites de l'information de représentation et des informations de pérennisation, constitué principalement de descriptions de paquets et permettant aux utilisateurs de rechercher, commander et récupérer des informations du système d'archivage.

Par exemple dans le cadre du contrôle de la légalité cette information descriptive, destinée à identifier une délibération, pourrait être la date de la délibération et le nom de la collectivité émettrice.

6.2 Présentation des fonctionnalités

En terme de flux d'information, le schéma ci-après fournit le fonctionnement général des échanges avec les services versants et les utilisateurs.



Les fonctionnalités générales sont les suivantes.

F1. Versement : permet le traitement des paquets d'informations en provenance des Services versants dans son ensemble. Cette fonction inclut tous les mécanismes de préparation, transmission, contrôle, rejet, complément d'information ainsi que tous les traitements de ces informations pour une intégration dans le dispositif de Stockage des contenus et celui de gestion des données descriptives.

F2. Stockage : gère l'ensemble des services liés à la conservation des paquets d'informations archivés à partir du moment où ils sont mis à sa disposition par la fonction de Versement jusqu'à leur destruction/élimination s'il y a lieu tout en garantissant leur intégrité. Cette fonction prend entre autres en compte les aspects de choix de supports, de gestion de l'ensemble des migrations et restaurations.

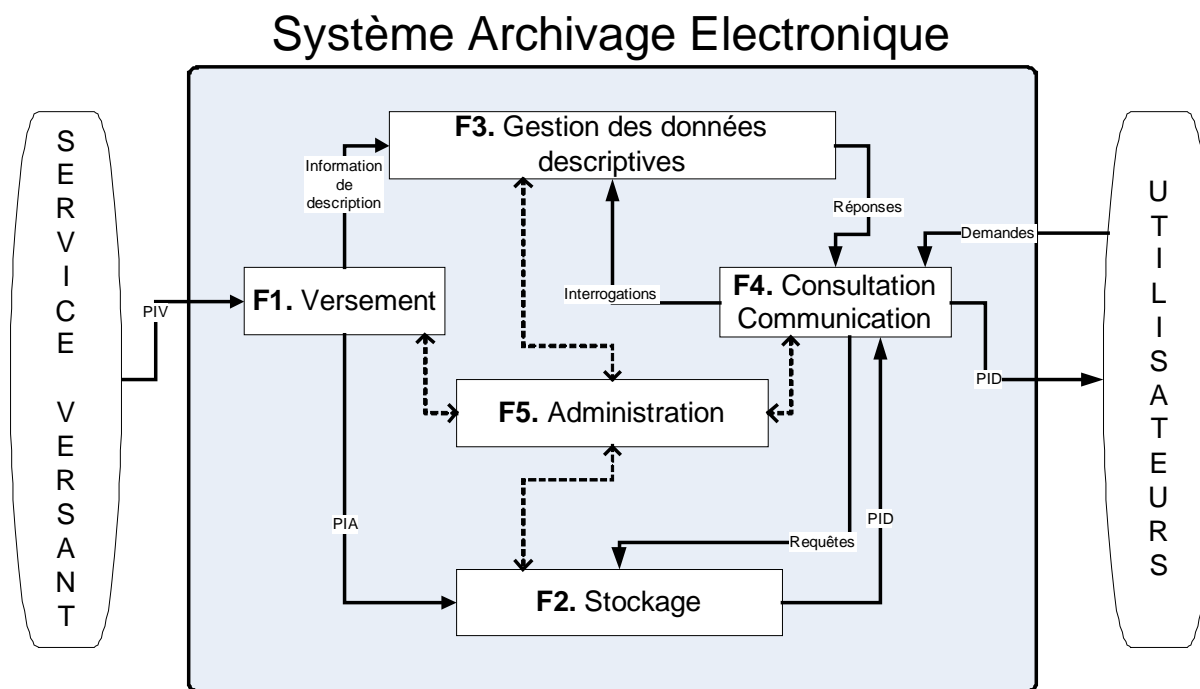
F3. Gestion des données descriptives : assure la conservation, la mise à disposition et la mise à jour des informations descriptives associées aux contenus d'informations, conservés par la fonction Stockage. Ces informations doivent servir aux utilisateurs comme point d'entrée au Système d'archivage électronique (SAE) et permettre de retrouver les données qu'ils recherchent en assurant le lien avec leur identification de localisation dans le système de stockage.

F4. Consultation et communication : prévoit l'ensemble des mécanismes permettant d'accéder, de consulter et de livrer les informations disponibles dans le SAE, qu'il s'agisse des données descriptives ou du contenu lui-même. Elle comprend la mise à disposition d'une interface de consultation, un système de recherche effectuée à partir des données descriptives, un principe de visualisation du résultat, la sélection de contenus à communiquer et la livraison effective de ces contenus sous forme de paquets d'informations diffusés. Dans la mesure où la communication du contenu peut être différée par rapport au moment de l'interrogation, cette fonction doit également prévoir un mécanisme de commandes à destination des utilisateurs, le suivi étant assuré par la fonction Administration.

F5. Administration : permet d'assurer l'exploitation d'ensemble du Système d'archivage électronique et sa pérennisation ainsi que la gestion des utilisateurs du SAE au sens de leurs droits d'accès.

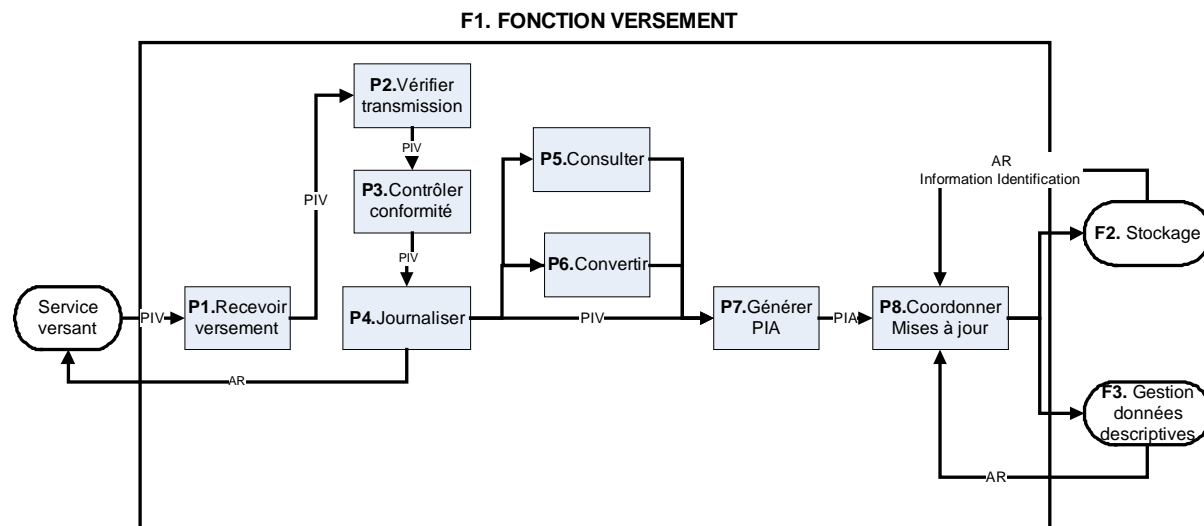
7. Détail fonctionnel

Le schéma ci-après précise les liens entre les différentes fonctions auxquelles doit répondre le SAE envisagé.



7.1 F1. Versement

La finalité du versement est de transformer les paquets d'informations versés en un ou plusieurs paquets d'informations archivés.



La fonction Versement est prise en charge par le système de gestion des versements qui se compose :

- d'une application de gestion des versements. Cette application se compose elle-même :
 - d'une base de données recensant l'ensemble des transmissions effectuées par les services producteurs
 - d'une interface de type web permettant aux services producteurs de référencer manuellement les transmissions qu'ils s'approprient à effectuer
 - d'une interface non interactive permettant de traiter par lot les transmissions.
- d'un service de transmission par réseau informatique. Ce service assure :
 - la transmission des données entre le service producteur et le service versant
 - la confidentialité et l'intégrité des données durant cette transmission
- d'un annuaire utilisateurs permettant de répertorier les utilisateurs, leurs droits et d'entreposer les informations de vérification pour leur authentification
- d'un ou plusieurs postes de travail équipés des matériels et logiciels nécessaires à :
 - la prise en charge des transmissions reçues sur support amovible
 - le contrôle et le traitement des versements transmis avant archivage
- d'un espace de stockage temporaire des fichiers versés en cours de traitement et avant leur transmission sur la plate-forme de stockage.

La fonction versement peut être décomposée en neuf processus.

7.1.1 P0 : Préparation du transfert et transfert

Remarque :

La mise au format spécifié dans le standard de versement des archives électroniques pour le transfert, se fera soit en amont soit après le versement par les archivistes qui testeront le pilote (voir les acteurs du projet). Cette mise au format se fera à travers une des fonctions du pilote (préparation du versement), à travers une interface de gestion des versements permettant, suivant les cas, une plus ou moins grande automatisation des tâches.

7.1.1.1 Sous-processus P0-1 : préparation du transfert

Acteur : service versant

Tâche « Extraction du contenu d'information à archiver et des métadonnées associées et mise en forme de ces éléments »

Le déroulement de cette tâche se fait comme suit.

- Le responsable concerné (en collaboration avec le service producteur de l'objet à archiver) spécifie dans l'application source les informations devant faire l'objet de l'archivage
(hors du périmètre)
- Il extrait les fichiers objet de l'archivage de l'application source
(hors du périmètre)
- Il convertit si nécessaire ces fichiers dans un format répondant aux exigences de pérennité définies par le service d'archives et stocke ces fichiers dans un espace temporaire
(hors du périmètre)
- Il extrait les métadonnées métier correspondantes de l'application source
(hors du périmètre)
- Le service versant constitue un bordereau de transfert au format XML conforme au standard d'échange de données pour l'archivage, comprenant les métadonnées décrivant le transfert et les fichiers transférés, soit encapsulés dans le bordereau XML, soit sous forme de fichiers joints.

Tâche « Contrôle du lot »

L'acteur de cette tâche est le service versant.

Son déroulement se fait comme suit :

- le service versant vérifie que les fichiers objet de l'archivage générés sont produits dans les bons formats et exploitables par le service d'archives
- il vérifie par sondage que les fichiers objet de l'archivage correspondent bien aux informations devant faire l'objet de l'archivage (en contrôlant par exemple le nombre de fichiers produits)
- il vérifie par sondage que les fichiers XML transférés sont conformes aux exigences du service d'archives

Tâche " Référencement de la transmission dans le système de gestion des versements " (messages se rapportant à la "demande de transfert d'archives", dans le standard d'échange)

Cette tâche se déroule comme suit.

- Le responsable accède à l'application de gestion des versements et s'authentifie.
- Il crée un nouveau formulaire de transmission et y saisit les informations suivantes :
 - intitulé du service versant et coordonnées associées (calculé à partir des informations d'authentification)
 - agent responsable de la transmission et coordonnées associées dont adresse mél (calculé à partir des informations d'authentification)
 - mode de transmission : envoi d'un support amovible ou transmission réseau
 - Si la transmission s'effectue par support amovible : nombre de lots transmis, nombre de supports, liste des lots inscrits sur chaque support.
- Il enregistre le formulaire
- Le système fournit à l'utilisateur un numéro provisoire affecté à chaque lot.

7.1.1.2 Sous-processus P0-1a : Transmission manuelle du lot sur support amovible

Acteur : service versant

Les pré-requis à cette opération sont :

- ⇒ Les lots sont disponibles
- ⇒ Un formulaire de transmission a été saisi dans l'application de gestion des versements
- ⇒ Le service versant dispose d'un équipement de gravure et des médias correspondants compatibles avec les équipements de lecture à la disposition du service d'archives

Tâche « Préparation et envoi du support » ("transfert d'archives" dans le standard d'échange)

Cette tâche se déroule comme suit.

- L'agent génère éventuellement (dépend des sources de données) un condensat associé au lot et stocke celle-ci dans un répertoire temporaire
- Il sélectionne l'ensemble des lots concernés par la transmission - Dans le cas où les fichiers archivés seraient trop volumineux pour être contenus sur un seul support, il les répartit dans différents sous-répertoires de façon à ce que la taille des sous-répertoires soit inférieure à la capacité du support utilisé
- Il inscrit les fichiers sur un ou plusieurs supports amovibles
- Il étiquette le ou les supports en reportant sur ces étiquettes le nom du service producteur, la date du versement et le(s) numéro(s) provisoire(s) affecté(s) au(x) versement(s).
- Il dépose le support au service d'archives ou envoie celui-ci par courrier (courrier recommandé).

7.1.1.3 Sous-processus P0-1b : Transmission manuelle du lot au service d'archives par réseau informatique

Acteur : service versant

Tâche « Transmission des fichiers » ("transfert d'archives" dans le standard d'échange)

Cette tâche se déroule comme suit.

- l'agent accède à l'application de gestion des versements et s'authentifie
- il recherche le formulaire de transmission associé au lot à verser et consulte celui-ci
- le transfert, suivant le type de sources de données, peut être signé et une empreinte générée par le service versant pour certains objets archivés
- il valide ces informations et commande au système de transmettre les fichiers spécifiés
- le système sélectionne les fichiers et opère la transmission des données vers l'espace de stockage temporaire destiné aux fichiers versés

7.1.1.4 Sous-processus P0-1c : Transmission automatique du lot au service d'archives par réseau informatique (actuellement hors périmètre)

Acteur : service versant

Tâche " Transmission des fichiers " ("transfert d'archives" dans le standard d'échange)

Cette tâche se déroule comme suit.

- l'agent du service d'archives valide, dans l'application source, la réalisation d'une transmission d'archives
- l'application source commande au système de transmettre les fichiers concernés par la transmission
- le système sélectionne les fichiers, génère l'empreinte accompagnant la transmission et opère cette transmission des données vers l'espace de stockage temporaire destiné aux fichiers versés.

7.1.2 P1.Recevoir versement

Ce processus consiste à effectivement réceptionner dans un espace de stockage tampon, les Paquets d'informations versés (PIV) en provenance du Service versant.

Cas du support amovible

Tâche « Réception du support par le service d'archives » ("accusé de réception de transfert d'archives" du standard d'échange)

Cette tâche se déroule comme suit.

- l'agent du service réceptionne le ou les supports
- il verse les fichiers contenus sur les supports dans un répertoire temporaire
- selon les cas, il entrepose les fichiers dans l'espace de stockage temporaire dédié aux fichiers versés en cours de traitement

Dans le cas du réseau, les fichiers se retrouvent directement dans l'espace de stockage tampon.

7.1.3 P2.Vérifier transmission

Ce processus vérifie que le Paquet d'informations transmis par le Service versant a bien été réceptionné dans son intégralité et sans altération. Il conviendra par exemple de vérifier que les fichiers transférés ne sont pas porteurs de codes malveillants (virus, chevaux de Troie...).

Suivant les différentes sources de données, l'intégrité globale de l'envoi ainsi que l'intégrité des différents Paquets d'informations transmis et reçus devront pouvoir être contrôlées (signature électronique pour l'envoi et/ou empreintes des objets archivés).

Le pilote devra pouvoir par conséquent supporter différents algorithmes ainsi que différents formats de signatures.

Deux tâches peuvent s'opérer, l'une ou l'autre suivant la nature de la transmission (support amovible ou réseau).

Cas du support amovible

Tâche « Réception du support par le service d'archives » ("accusé de réception de transfert d'archives" du standard d'échange)

Cette tâche se déroule comme suit.

- L'agent du service d'archives accède à l'application de gestion des versements et s'authentifie
- il contrôle l'empreinte associée aux fichiers
- il recherche le formulaire de transmission associé au lot réceptionné et consulte le formulaire
- il complète le formulaire en renseignant :
 - la date de réception des supports
 - le résultat du contrôle d'empreinte : succès ou échec
- le système transmet au service versant un courriel de notification de la transmission ou de rejet suite à une erreur d'intégrité. Cette notification doit être journalisée.

Cas du réseau.

Tâche « Réception du lot par le service d'archives ("accusé de réception de transfert d'archives" du standard d'échange)

Cette tâche se déroule comme suit.

- Une fois transmis, le système procède à une vérification de signature/ contrôle d'empreinte. Le système met à jour le formulaire relatif au transfert en précisant :
 - la date de réception du transfert
 - le résultat de la vérification de signature et des contrôles d'empreinte : succès ou échec
- Le système envoie un message à l'agent du service d'archives en charge des versements pour l'alerter de la transmission de nouveaux transferts
- Le système transmet au service versant un courriel de notification de transmission ou de rejet suite à une erreur d'intégrité. Ce message doit être journalisé.
- L'agent du service d'archives en charge des versements reçoit le message d'alerte reçu par le système de versement

7.1.4 P3. Contrôler la conformité

("avis d'anomalie de transferts d'archives" et "accusé de réception d'avis d'anomalie")

Ce processus contrôle que le paquet d'informations versé est conforme et respecte bien les conditions définies entre le service versant et le service d'archives, entre autres en matière de structuration de l'ensemble des données et de leur complétude, en matière de format de description, en matière de respect des formats d'encodage des objets versés et de leurs sous-objets. Le pilote devra par conséquent être capable d'ouvrir les différents fichiers, de contrôler si leur nombre est bien conforme à ce qui est inscrit dans les métadonnées des objets à archiver, qui accompagnent le transfert ; de contrôler si les formats annoncés sont bien ceux qu'ils prétendent être ; contrôler si la structure des objets versés et de leurs métadonnées est bien conforme...

Acteur : service d'archives

Les pré-requis de cette opération sont :

- ❑ Les lots transmis sont disponibles sur l'espace de stockage temporaire du système
- ❑ Un formulaire de transmission associé aux fichiers versés a été validé par le système
- ❑ Le service d'archives dispose d'un logiciel lui permettant :
 1. De contrôler automatiquement la conformité des lots transmis (conformité au schéma du standard d'échange de données pour l'archivage, contrôle des empreintes, contrôle du format des fichiers, contrôle de la taille des fichiers, contrôle des dates...)
 2. De consulter si nécessaire les bordereaux XML et les fichiers joints ou encapsulés. La consultation du bordereau de transfert XML doit être possible à la fois sous forme native et via une feuille de style (livrée par le titulaire).

L'opération se déroule comme suit.

- le responsable du service d'archives accède à l'application de gestion des versements et s'authentifie
- il recherche le formulaire de transmission associé au lot à contrôler et consulte celui-ci en vérifiant si tous les renseignements prévus dans le standard d'échange (informations de gestion quant aux droits d'accès, aux délais de conservation, aux délais de communicabilité..., informations techniques sur les formats, informations descriptives), figurent bien et les complète si besoin. L'application devra ainsi permettre de contrôler automatiquement la présence ou l'absence de ces informations et de pointer les informations manquantes, afin de guider le responsable dans son travail. Ainsi, par rapport aux formats, le système devra pouvoir repérer les objets dont le format ne correspond pas à un de ceux listés dans la table des formats et l'archiviste décidera alors de rejeter ou non cet objet.

Concernant l'identification du service versant, le responsable pourra compléter certaines informations. L'application de gestion des versements pourra notamment s'interfacer avec une application permettant de gérer et de tracer l'historique des services producteurs (et le rattachement de tel service à sa hiérarchie par exemple) et leur lien avec les services versants, traçabilité qui permettra notamment de pouvoir mettre à jour les droits d'accès aux objets archivés, en cas d'évolutions et de changements de structures (organigrammes) des services producteurs.

Cette application permettant l'historisation des producteurs est hors du périmètre mais l'interface doit être possible ainsi que la récupération de données depuis cette application vers l'application de gestion des versements.

- il procède au contrôle de conformité (structuration et formats) des lots (opération réalisée automatiquement par le système sur l'ensemble des fichiers transmis).
- il contrôle l'empreinte de l'objet archivé (si celle-ci a été générée par le producteur)
- il consulte, par sondage, un ou plusieurs lots
- il contrôle que les documentations nécessaires à l'exploitation du contenu versé existent en fonction des résultats de ces contrôles, il modifie le formulaire de transmission en indiquant s'il valide ou rejette l'ensemble des versements associés ainsi que les éventuelles raisons du rejet. Le rejet est notifié au service versant et le message envoyé est journalisé.

Le cas échéant, le service d'archives peut procéder à une conversion de format des données transférées. Dans ce cas, il en informe le service versant (avis de modification).

7.1.5 P4.Journaliser

Ce processus répond à un impératif consistant à enregistrer dans un journal l'intégralité des opérations effectuées et des événements. En parallèle, ce processus envoie, comme indiqué dans le standard d'échange, des notifications et accusés de réception (ou un rapport d'anomalie en cas de contrôles négatifs) au Service versant précisant le résultat de l'opération, suite aux différents contrôles effectués.

7.1.6 P5.Consulter

Ce processus doit permettre si nécessaire, aux personnes habilitées du service d'archive, de consulter le contenu du paquet d'information versé.

7.1.7 P6.Convertir

Ce processus est optionnel et répond au besoin résultant du cas où le service producteur n'est pas en mesure de produire le paquet d'information versé en respectant les spécifications attendues. On peut ainsi envisager que, dans certains cas, le SAE opère une migration de formats soit à l'arrivée soit au terme d'un délai dépendant de l'obsolescence du format d'origine.

7.1.8 P7.Générer PIA

Ce processus revient à constituer un Paquet d'informations archivé conforme aux normes de documentation et de formatage des données. Ce processus prend également en compte les conditions d'archivage spécifiques aux paquets d'informations versés : conditions de préservation, de communication et éventuellement de destruction.

Tâche « Extraction des objets d'archives »

Acteur : le service d'archives

Les pré-requis de cette opération sont :

- ☐ Les bordereaux XML ainsi que les fichiers de données joints ou encapsulés sont disponibles sur l'espace de stockage temporaire du système
- ☐ Un formulaire de transmission associé au lot versé a été validé par le système puis validé par le service d'archives après contrôles

L'opération se déroule comme suit.

- le *service d'archives* accède à l'application de gestion des versements et s'authentifie
- il recherche le formulaire de transmission associé aux versements à archiver et consulte celui-ci
- il valide l'archivage effectif des versements
- le cas échéant, il extrait de chaque fichier XML correspondant à l'objet-contenu le contenu éventuellement codé (par exemple en Base64) et génère un ou plusieurs nouveau(x) fichier(s) dans son(leur) format original
- il met les fichiers XML et les fichiers au format d'origine à la disposition du système de stockage

Tâche « Extraction des informations descriptives »

Acteur : le service d'archives

Les pré-requis de cette opération sont :

- ☐ Les lots sont disponibles sur l'espace de stockage temporaire du système
- ☐ Un formulaire de transmission associé aux fichiers versés a été validé par le système puis validé par le service d'archives après contrôles, le statut du formulaire de versement est " Archivage effectué ", il contient la ou les références fournies par le logiciel de gestion du stockage après l'archivage

- ❑ Le service d'archives dispose d'un logiciel lui permettant d'extraire automatiquement les métadonnées contenues dans les bordereaux XML conformes au standard d'échange de données pour l'archivage

L'opération se déroule comme suit.

- L'agent valide la transmission vers la base archives des métadonnées comprises dans les bordereaux XML associées à une transmission
- le système de versement extrait automatiquement des fichiers concernés, les métadonnées (dont les références fournies par le logiciel de gestion du stockage) et met ces informations à disposition de la base archives sur un répertoire temporaire
- le système de versement modifie le formulaire de transmission correspondant en renseignant la date de transmission à la base archives
- le système de versement avertit la base archives de l'existence de nouvelles données à importer

7.1.9 P8.Coordonner les mises à jour

Ce processus consiste d'une part à transmettre à la fonction stockage le contenu d'information à conserver et d'autre part à transmettre à la fonction gestion des données descriptives les métadonnées accompagnant les objets à archiver. Le processus attend ensuite en retour l'accusé de réception du résultat de l'opération. Dans le cas du stockage, l'accusé de réception doit contenir l'information d'identification de l'espace de stockage. Cette dernière donnée est ensuite envoyée en complément des informations précédentes à la fonction gestion des données descriptives.

Ce processus établit un tableau de suivi des « entrées-sorties » vers les bases descriptives et de stockage.

7.2 F2. Stockage

L'objectif du stockage dans un SAE est absolument essentiel dans la mesure où il consiste à garantir la pérennité et l'intégrité de l'ensemble des informations qui y sont conservées. Par souci de clarté le terme de paquet d'informations archivés est utilisé par la suite alors que dans la réalité et si l'on se réfère au standard d'échange, il s'agirait plutôt du contenu d'information au sens OAIS.

Le pilote devra disposer d'un système de stockage possédant les caractéristiques suivantes : fiabilité, disponibilité, maintenabilité, sécurité.

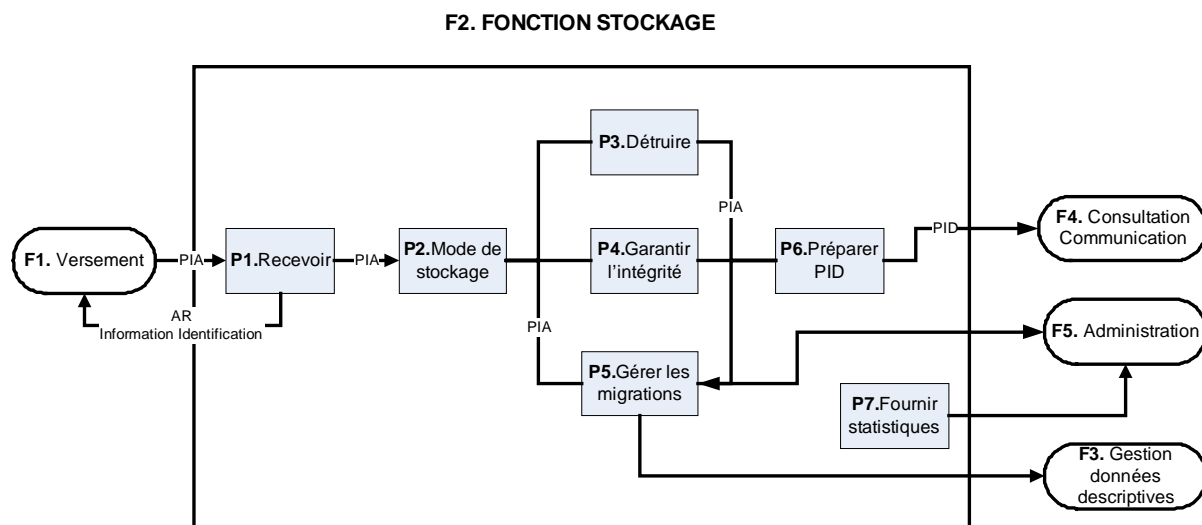
Le système doit également permettre l'abstraction de la plate-forme matérielle (pouvoir évoluer sans impact sur l'organisation logique des archives), être extensible, interopérable et évolutif (capacité de stockage, puissance de traitement et nombre d'utilisateurs)

Concernant le choix des typologies de stockage et des types de supports, ce sera au soumissionnaire de faire des propositions, sachant que le pilote doit permettre une capacité de 2 TO (sur chacun des deux sites distants : voir partie 8) ; et que, par ailleurs, les CD et DVD ne sont pas préconisés (volumétries trop réduites, pérennité pour les DVD non prouvée et normalisation insuffisante).

Les supports devront posséder les qualités suivantes :

- stabilité intrinsèque du support et robustesse,
- large diffusion de la technologie et offre multi-constructeurs ou reposant sur des normes publiques,
- existence d'outils de contrôle des supports,
- chemin d'accès aux données protégé,
- simplicité des opérations de recopie,
- protection contre l'effacement accidentel.

Voir la partie architecture pour la plate-forme de stockage



Le stockage comprend les sept processus décrits ci-après.

7.2.1 P1.Recevoir

Ce processus revient à réceptionner les paquets d'informations archivés en provenance du processus de Versement et à les transférer physiquement vers le volume de stockage le mieux approprié et correspondant aux conditions d'archivage (durée, fréquence de consultation, communication en ligne ou différée, destruction in fine, ...) indiquées au moment du versement.

Lorsque les paquets d'informations archivés sont effectivement écrits sur le support de stockage adapté, il y a transmission au processus de Versement du résultat de l'opération comprenant l'information d'identification correspondant à l'espace de stockage où se trouve physiquement les paquets d'informations archivés qui viennent d'être traités. L'accusé de réception du système doit être envoyé seulement lorsque l'écriture est véritablement effective sur le support en question et non pas en attente de traitement dans un espace mémoire tampon et plus précisément ne parvient qu'après écriture effective sur les deux sites (voir plus bas, paragraphe sur les types de sauvegardes et réplication).

Tâche " Importation des objets d'archives dans la base de stockage " ("notification d'acceptation de transfert d'archives" du standard d'échange)

La tâche se déroule comme suit.

- le *service d'archives* met les fichiers XML et les fichiers au format d'origine à la disposition du système de stockage
- le système de stockage prend en charge le versement
- l'application de versement reçoit du système de stockage les avis de réception comme quoi les fichiers ont bien été pris en charge (après écriture effective sur les deux sites distants : voir partie sur l'architecture globale du pilote)
- l'application de versement modifie le statut du formulaire de transmission (Nouveau statut : archivage effectué– import dans la base de stockage effectué »), affecte des numéros de versement à chacun des fichiers versés lors de la transmission et enregistre la date de l'archivage ainsi que les références de localisation associées fournies par le système de stockage

7.2.2 P2.Mode de stockage

Voir également la partie Architecture

Acteur : le service d'archives

Ce processus consiste à conserver effectivement les paquets d'informations archivés et à choisir le support adéquat en fonction d'un certain nombre de critères dont les principaux sont l'accessibilité et la durée. Il est demandé d'envisager de mettre en place un système de **HSM (hierarchical storage management)** afin d'aider à cette gestion des différents supports, notamment afin de pouvoir passer d'un accès en ligne vers un accès « hors ligne » (sur rayonnages).

Cette opération assure les tâches suivantes.

- Préparation des supports d'archivage
- Archivage des contenus sur une plate-forme automatisée qui se déroule comme suit.
 - le système de stockage prend en charge l'information à archiver et procède à son inscription sur son support d'archivage dans deux sites distincts (voir partie sur l'architecture globale du pilote)
 - le logiciel de gestion du stockage associe au versement une référence permettant d'en assurer la localisation, ainsi que celle de tous ses composants. Le niveau de rattachement (par fichier archivé, par lot, etc.) sera précisé dans les spécifications détaillées.
 - Il renvoie au système de versement un accusé de prise en charge du versement accompagné de la ou des références associées au lot versé

7.2.3 P3.Détruire

Ce processus est destiné à traiter la destruction si nécessaire des paquets d'informations archivés de façon manuelle ou automatique.

Le pilote devra posséder une telle fonction comportant au minimum un dispositif de suppression des accès aux contenus d'informations par suppression des index et mieux un véritable dispositif d'effacement des contenus d'information. Ce dispositif devra par ailleurs être conçu de telle sorte à ne laisser aucune trace sur le support d'origine, due entre autre au phénomène physique de rémanence des supports magnétiques.

En ce qui concerne les supports amovibles type bande ou CD, la destruction sera opérée sur l'ensemble du contenu et du contenant.

7.2.3.1 Sous-processus P3-1 : Demande d'accord pour l'élimination

("demande d'élimination d'archives", "accusé de réception de la demande d'élimination d'archives" du standard d'échange)

Acteur : le service d'archives

Cette opération se déroule comme suit.

- Le *service d'archives* accède à l'application de gestion des commandes et s'authentifie
- Il enregistre une demande d'accord pour élimination, (génération d'un formulaire d'élimination) avec enregistrement d'un identifiant pour la demande, de sa date, de l'identifiant des objets archivés concernés, et message en texte libre
- Cette demande est suivie d'une notification avec l'enregistrement de sa date

- Suivant le même processus que pour la communication, le système transmet l'objet concerné au service à qui la demande est adressée, afin que celui-ci puisse prendre connaissance du contenu de l'objet et procéder à sa vérification d'empreinte.

7.2.3.2 6.4.2 Sous-processus P3-2 : Accord pour l'élimination/Refus pour l'élimination

("acceptation de demande d'élimination d'archives", "rejet de demande d'élimination d'archives", "accusé de réception de rejet" du standard d'échange)

Acteur : le service à qui la demande a été faite

Cette opération se déroule comme suit.

Le service peut ou non accepter l'élimination. Dans tous les cas, l'application permettra d'enregistrer, en complétant le formulaire d'élimination, la date de l'accord ou du refus (un message sera également rempli par le service en cas de refus, afin d'explicitier celui-ci) ;

7.2.3.3 6.4.3 Sous-processus P3-3 : Notification d'élimination

("notification d'élimination d'archives" du standard d'échange)

Acteur : le service d'archives

Cette opération se déroule comme suit.

Elle est enregistrée par le *service d'archives* (avec sa date) et transmise au service, une fois l'élimination effectivement effectuée.

7.2.4 P4.Garantir l'intégrité

Ce processus est extrêmement important dans la mesure où il doit garantir l'intégrité de l'ensemble des paquets d'informations archivés et en conséquence, la vérifier systématiquement. Cette tâche peut être entièrement prise en charge par la plate-forme sans intervention humaine.

Il est en effet nécessaire de contrôler régulièrement les paquets d'informations archivés sur les différents supports afin d'anticiper d'éventuelles erreurs et surtout de prévoir des dispositifs d'avertissement d'une part et de correction d'autre part, voire de remplacement.

En cas de détection d'une erreur d'intégrité la seule façon de la corriger est de remplacer les données concernées par un jeu de données identiques non corrompues dont on disposera grâce à un système de duplication adapté de l'ensemble des données. Par conséquent, le contrôle devra pouvoir être ponctuel (au moment de l'écriture sur les supports de conservation dans les deux sites, ou encore au moment d'une communication par exemple), mais également

régulier par sondage. Ces contrôles doivent également pouvoir être paramétrés en fonction du type de supports et de leurs âges respectifs, voire du type d'objets archivés que l'on souhaiterait contrôler plus fréquemment.

7.2.5 P5.Gérer les migrations

Il s'agit de maîtriser l'ensemble des migrations requises par le système tant des supports que des formats.

Ces migrations interviennent soit de façon planifiée (voir fonction Administration) soit par exemple pour corriger des erreurs détectées sur tel ou tel support.

7.2.5.1 Sous-processus P5-1 : Régénération des supports (tâche entièrement prise en charge par la plate-forme)

Il s'agit d'un **changement de supports**.

Ce premier type de migration consiste à permettre de remplacer, renouveler des supports sur lesquels les données sont conservées (soit régénération quand on recopie l'information sur le même type de support, soit migrations quand le type de support change). Ces changements pourront faire suite à des erreurs répétitives sur un support ou tout simplement être programmés au préalable en fonction du type de support et de leur âge. Les erreurs dont il est ici fait mention sont essentiellement de deux types : erreurs de lecture du support ou erreur d'intégrité. Dans les deux cas il est nécessaire et indispensable de disposer d'un dispositif de correction automatique de ces erreurs dont la conséquence principale revient justement à changer de support.

7.2.5.2 Sous-processus P5-2 : Migration de formats (**hors périmètre**)

La migration de format concerne la migration au niveau des formats d'encodage des données.

Il pourra ainsi s'agir de supports fonctionnant sous un nouveau système d'exploitation disposant d'une gestion de fichiers spécifique. La migration de formats pourra également être rendue nécessaire en raison d'une obsolescence technologique des formats de données archivés, en raison d'une veille technologique anticipant la disparition de tels formats ou au contraire de l'apparition sur le marché, d'un nouveau format plus approprié à la pérennisation (par exemple, des fichiers au format PDF vers le format PDF/A normalisé ISO 19005). On privilégiera par conséquent les technologies qui se prêtent bien aux migrations (logique de cellules de stockage indépendantes par exemple).

Le projet de migration se déroulera en deux phases qui seront journalisées :

Phase 1 : préparation

- Recensement, à l'aide du logiciel de gestion du stockage, de l'ensemble des fichiers concernés par l'opération de migration
- Spécification du nouveau format de conservation
- Spécification d'une procédure de migration de l'ancien vers le nouveau format.
- Développement ou acquisition d'une application permettant de convertir les fichiers dans le format souhaité
- Tests et mise au point de la procédure et de l'application
- Planification des opérations de migration dans le logiciel de gestion du stockage

Phase 2 : migration

- Suivi des opérations de migration réalisées automatiquement par le logiciel de gestion du stockage et l'application de conversion
 - Mise à jour des informations dans la base de gestion des données descriptives : extension du fichier (liens)
 - Mise à jour de la base de gestion du stockage : localisation

7.2.5.3 Sous-processus P5-3 : Restauration du contenu d'un support

La restauration du contenu d'un support intervient lorsque un support s'avère inutilisable et qu'il est donc nécessaire de procéder à la restauration des fichiers concernés sur un nouveau support à partir d'autres exemplaires conservés soit sur un autre support, soit sur l'autre plate-forme de stockage.

Cette opération s'effectue dans l'un ou l'autre cas comme selon les étapes suivantes (qui seront journalisées) :

- identification des fichiers contenus sur le ou les supports concernés
- identification des autres supports contenant les autres exemplaires des fichiers concernés
- extraction des fichiers des supports
- inscription des fichiers sur un ou plusieurs nouveau(x) support(s)
- implémentation du (des) support(s) dans la plate-forme

- référence du support et des fichiers dans le logiciel de gestion du stockage
- contrôle et enregistrement du résultat support par support

7.2.6 P6.Préparer

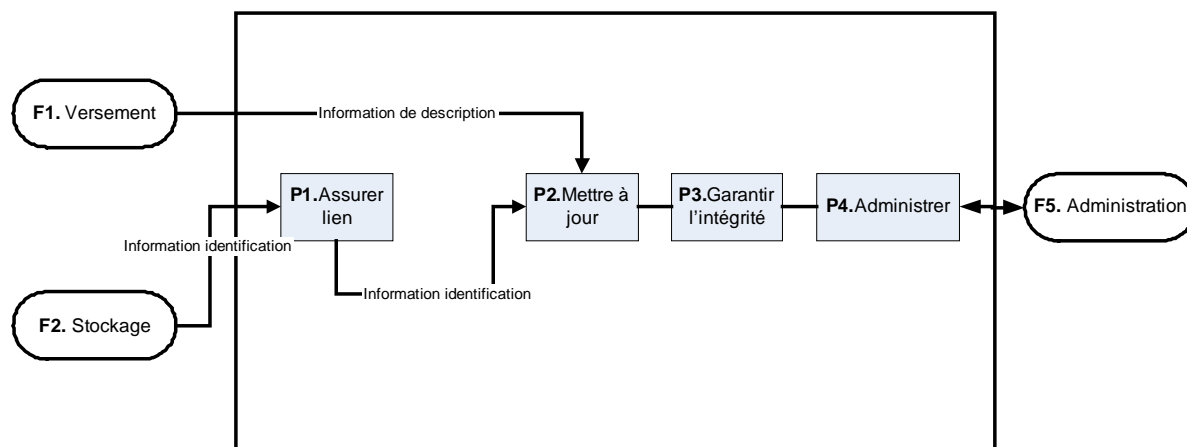
Ce processus est destiné à transmettre les paquets d'informations diffusés suite à une sollicitation du processus de communication.

7.2.7 P7.Fournir les statistiques

Il s'agit de bâtir des statistiques d'exploitation relatives d'une part aux capacités utilisées par rapport aux différents supports et espaces de stockage, ainsi que sur l'état des supports et d'autre part en matière de communication des paquets d'informations archivés, en compléments aux statistiques de consultation, sans oublier l'évolution des paquets d'informations versés.

7.3 F3. Gestion des données descriptives

La finalité de cette fonction est d'assurer la gestion des informations descriptives disponibles relatives aux contenus d'informations conservés par la fonction Stockage. Au moment de la prise en charge par le pilote, les métadonnées accompagnant les objets à archiver suivant le standard d'échange sont à extraire et à intégrer dans la base de données descriptive, qui peut par ailleurs être enrichie et complétée. De façon globale la gestion des données descriptives doit ainsi permettre l'enrichissement et/ou l'ajout de métadonnées destinées à décrire les objets archivés par rapport à leur contexte, leur contenu et leur structure.

F3. FONCTION GESTION DES DONNEES DESCRIPTIVES

La fonction gestion des données descriptives est composée des quatre processus suivants.

7.3.1 P1.Assurer lien

Ce processus consiste à maintenir le lien entre les informations descriptives et la localisation physique ou électronique des contenus d'informations.

7.3.2 P2.Mettre à jour

Le processus doit permettre la mise à jour des données correspondantes et au besoin en enregistrer de nouvelles suite à un nouveau transfert, une élimination ou suite à une opération de migration ("avis de modification" du standard d'échange).

Tâche " Importation des informations descriptives dans la base archives " ("notification d'acceptation de transfert d'archives" du standard d'échange)

La tâche se déroule comme suit.

- la base archives réceptionne la notification du système de versement
- elle procède à l'importation des données correspondantes au sein de la base
- elle produit un rapport d'importation qu'elle transmet au système de versement
- le système de gestion de la base archives transmet au système de versement un accusé de prise en charge
- le responsable valide le nouvel enregistrement dans la base et le complète éventuellement (compléments pouvant être apportés tout au long du cycle de vie du lot archivé)
- le système de versement modifie le statut du formulaire de versement correspondant (nouveau statut : Archivage effectué - indexation dans la base archives effectuée)
- le système de versement supprime les fichiers associés au lot de l'espace temporaire destiné aux versements en cours de traitement
- le système transmet au service versant un accusé de réception l'informant que l'archivage du lot versé a été effectué. Cet accusé de réception comprend notamment les numéros de versement associés aux différents fichiers transmis. Il doit être journalisé. Cet accusé de réception ne doit intervenir que lorsque l'écriture sur les deux sites est effective et l'importation des métadonnées descriptives effectuée.

7.3.3 P3.Garantir l'intégrité

Ce processus revient à garantir l'intégrité de l'ensemble des données gérées et à la vérifier régulièrement. Il est ainsi nécessaire de contrôler d'éventuelles erreurs à l'aide de fonctionnalités appropriées complétées par des systèmes d'avertissement et si possible de correction.

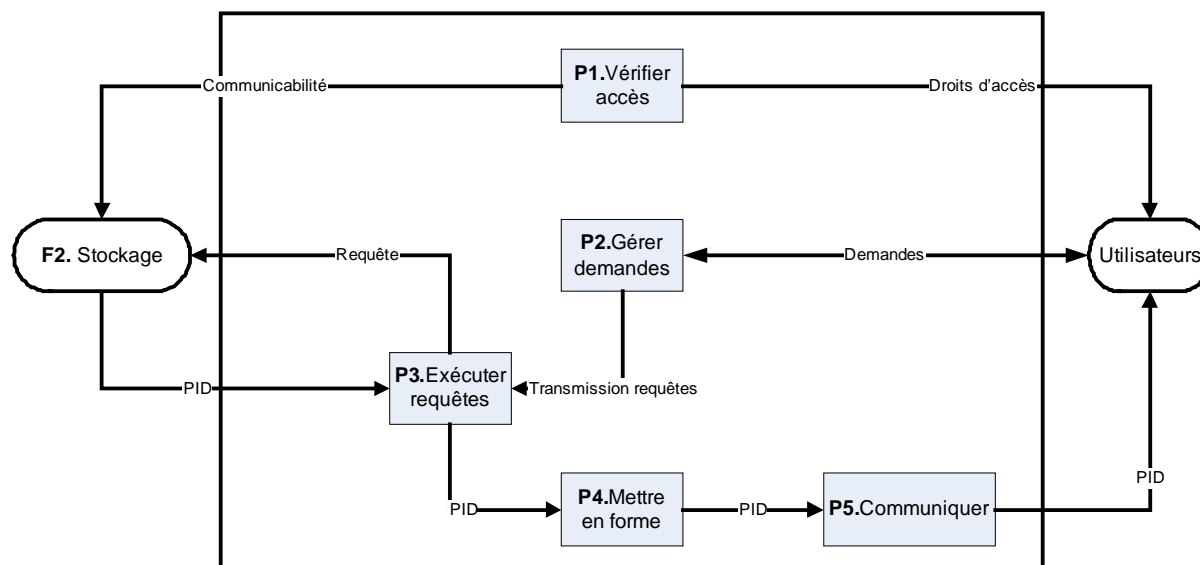
7.3.4 P4.Administrer

De façon spécifique par rapport à la fonction d'administration d'ensemble du pilote, ce processus doit administrer les fonctions de la base de données le cas échéant, à savoir conserver et tenir à jour les schémas des tables utilisées, les définitions des vues et autres états ainsi que garantir son intégrité référentielle.

7.4 F4. Communication / Consultation des Archives

Cette fonction constitue l'objectif principal de tout SAE à savoir offrir à l'utilisateur la possibilité de retrouver une information. Au-delà de cette vérification d'existence (données descriptives consultables en ligne), cette fonction processus permet également à l'utilisateur de demander à recevoir le ou les contenus d'information, accompagnés d'informations complémentaires constituant au final un ou plusieurs paquets d'informations diffusés.

F4. FONCTION CONSULTATION / COMMUNICATION



La fonction Communication / Consultation des archives est prise en charge par le système de consultation et de communication qui est composé de :

- une interface de consultation, de recherche et de commande de fichiers archivés exploitant les informations conservées dans la Base archives ;
- un logiciel de gestion des commandes passées par les utilisateurs ;
- un espace de stockage temporaire des fichiers communiqués par la plate-forme. Cet espace permet au service d'archives :
 - de contrôler les fichiers communiqués avant mise à disposition du demandeur
 - de donner accès aux fichiers au demandeur dans le cas d'une consultation à partir du réseau local du service d'archives
 - de préparer l'inscription des fichiers sur un support amovible dans le cas d'une communication sur support amovible
 - de préparer la transmission des fichiers par réseau dans le cas d'une communication par réseau
- un ou plusieurs postes de travail équipés des matériels et logiciels nécessaires :
 - au contrôle des fichiers restitués avant mise à disposition du demandeur
 - à la production des supports amovibles dans le cas d'une restitution sur support amovible

- un service de transmission sécurisé par réseau informatique. Ce service assure :
 - la transmission des données associées aux commandes entre le service d'archives et les services producteurs ayant commandé la restitution d'archives
 - l'intégrité des données durant la transmission

Les cinq processus composant cette fonction sont décrits ci-après.

7.4.1 P1.Vérifier les accès

Ce processus revêt un double objectif, tout d'abord vérifier les autorisations d'accès des utilisateurs et d'autre part vérifier la communicabilité des paquets d'informations archivés.

7.4.2 P2.Gérer les demandes (Consultation de la base Archives)

("demande de communication d'archives", "accusé de réception de demande de communication d'archives" et "rejet de demande de communication d'archives" du standard d'échange)

Acteur : utilisateurs

Ce processus permet aux utilisateurs d'enregistrer des demandes sous forme de commandes. Il assure également l'information des utilisateurs quant à l'avancement du traitement de leurs commandes. Pour effectuer ces demandes le processus devra mettre à disposition des utilisateurs un système de consultation accessible en ligne s'appuyant sur les données descriptives.

Les pré-requis pour cette opération sont :

- ❑ Un annuaire permettant d'authentifier les utilisateurs
- ❑ Une interface de consultation permettant d'interroger les informations descriptives contenues dans la base archives
- ❑ Un logiciel de gestion des commandes permettant d'enregistrer les commandes de restitutions

Elle se déroule comme suit.

- L'utilisateur accède à l'interface de consultation de la base archives et s'authentifie. Si l'utilisateur ne figure pas déjà dans un annuaire métier (au titre des services producteurs), il sera par défaut dans la catégorie "grand public". Il ouvre le moteur de recherche et saisit ses critères de recherche. Il peut soit :
 - effectuer une recherche par mot clé qui portera sur l'ensemble des informations descriptives disponibles dans la base archives : recherche plein texte ;
 - effectuer une recherche par champ : il sélectionne un ou plusieurs champs et y associe ses critères de recherche.
- Il lance la recherche.
- Le système produit un résultat sous forme de liste et présente ce résultat à l'utilisateur.
- L'utilisateur peut consulter le détail des informations descriptives associées aux différents fonds proposés.
- Il sélectionne les résultats correspondant à ses besoins et le système indique alors à l'utilisateur le nombre de fichiers correspondants.
- L'utilisateur valide son panier et commande la communication des documents correspondants. Un numéro d'identifiant de la demande est généré par l'application et la date de la demande enregistrée. Il précise le mode de restitution choisi (ou proposé par le système) : consultation en ligne, support amovible ou transmission réseau.

7.4.3 P3.Exécuter requêtes (Sortie d'une commande)

Acteur : le service d'archives

Ce processus lance les requêtes destinées à rechercher les éléments réclamés par l'utilisateur et assure le lien avec la fonction stockage afin d'obtenir les contenus d'information désirés. Ce processus devra également contrôler l'intégrité de l'information obtenue en retour avant de la transmettre à l'utilisateur.

Les pré-requis à l'opération sont :

Une commande a été passée par un utilisateur et enregistrée dans le logiciel de gestion des commandes.

Elle se déroule comme suit.

- A la suite de la commande, le système de gestion des commandes génère une notification de la demande de communication ainsi qu'un identifiant de la demande de communication ; la date de la notification est enregistrée ; il génère également un formulaire de suivi de la commande dans lequel il indique les références des fichiers à communiquer ainsi que les informations relatives à leur localisation
- Il transmet au logiciel de gestion de la plate-forme de stockage la référence des fichiers à restituer.
- Le logiciel de gestion de la plate-forme de stockage identifie les fichiers et en stocke une copie dans un répertoire spécifique de l'espace de stockage temporaire.

- Il contrôle l'intégrité des fichiers et produit un fichier conforme au standard d'échange des données électroniques, contenant l'ensemble des fichiers à restituer et des métadonnées correspondantes.
- Il modifie le formulaire de suivi en indiquant que la commande est prête à être transmise.
- Le système de gestion des commandes alerte l'agent en charge des communications.

7.4.4 P4.Mettre en forme

Ce processus consiste à préparer les paquets d'information diffusés, résultat de la recherche, avant leur communication.

7.4.5 P5.Communiquer

Acteur : le service d'archives

Comme son nom l'indique ce processus revient à communiquer les paquets d'informations diffusés aux utilisateurs. En fonction du type de demande, la communication des résultats de la recherche pourra être obtenue soit directement en ligne, soit être transmise sur tout autre support. Le pilote devra permettre ces deux possibilités suivant le type d'objets archivés.

Concernant le pilote, il n'est pas pertinent de fixer le nombre de consultations envisagées de façon globale ou simultanée, ou encore le nombre a priori de communications à prévoir en fréquence, en ligne ou asynchrone. Seules les fonctionnalités ici seront testées, la possibilité d'avoir un double régime de consultation (en ligne, asynchrone) et des tests sur plusieurs utilisateurs simultanés (5 à 10).

Les communications sur supports amovibles pourront se faire sur support papier (imprimantes laser noir et blanc), sur support optique non réinscriptible du type WORM (format, densité, label, ...) tels les CD-R, DVD-R, sur support magnétique (disque dur externe, clé USB...).

Les communications par télétransmission se feront par la mise en place de serveurs Web.

7.4.5.1 Sous-processus P5-1 : Transmission par support amovible

("communication d'archives", "accusé de réception de communication d'archives" et "avis d'anomalie de réception" du standard d'échange)

Les pré-requis pour cette opération sont :

Les fichiers à restituer ont été extraits de la plate-forme de stockage.

Elle se déroule comme suit.

- Le service génère un condensat associé au fichier contenant les fichiers à restituer et stocke celui-ci dans le répertoire.
- Il sélectionne les fichiers et le condensat.
- Dans le cas où les fichiers seraient trop volumineux pour être contenu sur un seul support, il répartit les fichiers en plusieurs lots de tailles compatibles avec le support utilisé.
- Il inscrit les fichiers sur un ou plusieurs supports amovibles.
- Il étiquette le ou les supports en reportant sur ces étiquettes le nom du destinataire, le numéro et la date de la commande.
- Il envoie le ou les supports par courrier au demandeur ou lui remet en main propre.
- Il modifie le formulaire de suivi de la commande en indiquant que la commande a été transmise.
- La date de la communication est enregistrée
- Le système alerte le demandeur de l'envoi de la commande
- Lors de la réception des fichiers, le service demandeur génère un accusé de réception qui est enregistré, ainsi que sa date, par l'application (standard).

7.4.5.2 Sous-processus P5-2 : Transmission par réseau

("communication d'archives", "accusé de réception de communication d'archives" et "avis d'anomalie de réception" du standard d'échange)

Les pré-requis à l'opération sont :

Les fichiers à communiquer ont été extraits de la plate-forme de stockage.

Elle se déroule comme suit.

- Le service d'archives accède à l'application de gestion des commandes et s'authentifie.
- Il ouvre le formulaire de suivi de la commande et le complète en renseignant l'emplacement des fichiers à restituer.
- Il valide ces informations et commande au système d'alerter le demandeur de la mise à disposition des fichiers commandés.
- L'utilisateur reçoit l'alerte et consulte le formulaire de suivi.

- Si l'utilisateur se trouve hors du réseau local du service d'archives, il lance le téléchargement des fichiers.
- Si l'utilisateur se trouve sur le réseau local du service d'archives, il choisit soit d'ouvrir le fichier depuis sa localisation actuelle, soit de le télécharger.
- Une fois téléchargé, le système modifie le formulaire de suivi en précisant que la communication a été effectuée et la date de la communication.
- Lors de la réception des fichiers, le service demandeur génère un accusé de réception qui est enregistré, ainsi que sa date, par l'application (standard).

7.4.6 Moteur de recherche

7.4.6.1 Fonctionnalités

Le pilote doit être doté d'un outil d'investigation s'apparentant aux moteurs de recherches du WEB.

La recherche, par l'utilisateur, des dossiers et des documents situés tant en zone de transit qu'en zone d'archivage peut se faire :

- par saisie de critères de recherche, sur l'ensemble des métadonnées ;
- par recherche en texte intégral.

Plus précisément, le moteur de recherche doit offrir les fonctionnalités suivantes :

- ;
- recherche booléenne avec proximité, sensibilité typographique, lemmatisation ;
- recherche " floue " : cette technologie permet d'obtenir le résultat voulu même si le terme est mal orthographié dans la requête (utile par exemple pour la recherche de noms propres) ;
- tris multiples (date, taille, scoring, etc...).

7.4.6.2 Paramétrage

Deux niveaux de recherche sont souhaités, répartis dans autant d'onglets :

- ✎ **recherche simple :**
cet écran permet de renseigner des critères simples de recherche
- ✎ **recherche avancée :**
cet écran permet de combiner, avec des opérateurs booléens, de proximité, plusieurs critères de recherche
- ✎ **historique :**
des recherches effectuées dans la session utilisateur

7.4.6.3 Présentation des résultats de la recherche

L'affichage de la liste de résultats comporte les fonctionnalités suivantes :

- ✎ Affichage en mode liste, triée par critères multiples (date, pertinence, taille, etc...)
- ✎ Liste paginée
- ✎ Affichage du nombre de résultats trouvés
- ✎ Mise en surbrillance des termes correspondant aux critères de recherche
- ✎ Possibilité de sélectionner un ou plusieurs éléments de la liste (cases à cocher)
- ✎
- ✎ Un lien sur le résultat de la recherche donnera un accès direct au document (dans la mesure où cet accès sera autorisé)

7.5 F5. Administration du Système d'archivage électronique

L'objectif de cette fonction est d'assurer l'exploitation d'ensemble du pilote tant au niveau des utilisateurs qu'en ce qui concerne son fonctionnement interne. De façon plus détaillée, sont fournis ci-après les principaux processus à prendre en considération classés par thèmes.

Exploitation

- Gérer la configuration du matériel et des logiciels du pilote consistant à en assurer la maîtrise technique destinée à surveiller en permanence son fonctionnement global ;
- Contrôler l'exploitation du pilote, de son fonctionnement et de ses performances en fonction de l'utilisation qui en est faite en fournissant entre autres des statistiques détaillées. Par ailleurs dès qu'une anomalie qu'elle quelle soit est détectée une alerte doit automatiquement être générée et transmise pour information et traitement.
- Exploitation (utiles pour la gestion du stockage):
 - Cartographie des espaces de stockage disponibles, pourcentage d'occupation et accroissement, état des supports
 - Temps d'accès moyen aux Archives
 - Fréquences d'accès par plage horaire, par jour, par mois, par an
 - Nombre d'incidents classés par type et fréquence
 - Nombre et types de migrations effectuées par mois,...

Ces statistiques devront pouvoir être exportées vers des outils bureautiques.

Sécurité

- Contrôler l'accès physique au pilote en fonction des règles de sécurité définies et des dispositifs de sécurité adoptés en conséquence
- Assurer la protection de l'ensemble des données gérées par le pilote dont certaines sont confidentielles : contenus d'informations, informations descriptives, données de gestion. Ce processus devra assurer la sauvegarde globale de l'ensemble des informations. **Le titulaire explicitera la mise en œuvre de ces sauvegardes (fréquences, types, supports utilisés).**
- Permettre la restauration totale ou partielle des données suite à un sinistre
- Assurer la traçabilité complète de tout ce qui se passe dans le pilote au travers de la gestion d'un journal d'évènement y compris le suivi de résolution des incidents rencontrés quelle qu'en soit l'origine. Ce processus devra également permettre l'enregistrement des tentatives d'accès par des utilisateurs non autorisés

Gestion

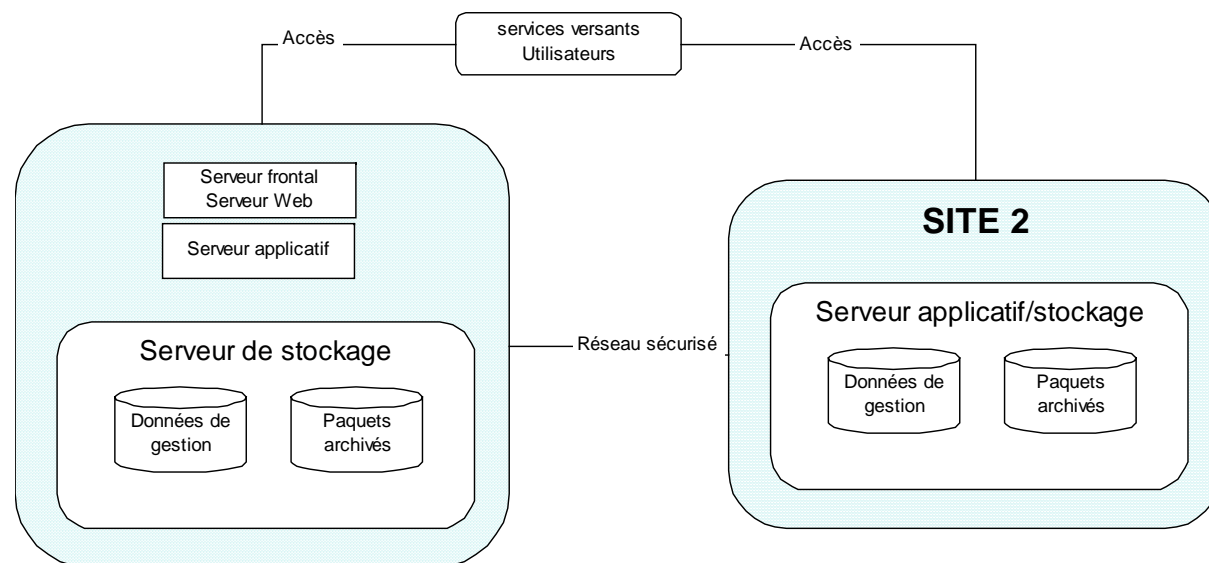
- Permettre un suivi des commandes de communication en cours afin de pouvoir renseigner les utilisateurs sur l'avancement des traitements
- Gérer les données administratives comme celles relatives aux utilisateurs (dont suppression de comptes) afin d'en assurer le suivi et permettre le cas échéant la production des éléments de facturation résultants des commandes effectuées ;

- Vérifier et garantir l'intégrité de l'ensemble des données administratives directement liées à l'exploitation vis-à-vis des utilisateurs mais aussi en interne

Pour chaque transaction, suivant le standard d'échange, les messages contiennent un code permettant de signaler toutes sortes d'anomalies au niveau des transferts, communications, éliminations, restitutions... Ces messages seront journalisés.

8. Architecture générale

8.1 Schéma général



Dans le cadre du pilote, tous les échanges de données devront pouvoir se faire par réseau local, à l'exception de la réplication entre sites par réseau étendu. De même, les fonctions de consultation ou de recherche devront, elles, pouvoir se faire par réseau étendu sur des sites distants.

L'échange d'archives par réseau étendu est hors périmètre.

8.2 Composants applicatifs

La plate-forme doit se composer des éléments suivants :

- un annuaire utilisateur,
- une application de gestion de versement

- une application de conversion des formats
- une base archives,
- une application de gestion des commandes,
- une base de connaissance,

Annuaire utilisateurs (voir partie 10.1)

Toutes les applications utilisées par le système reposeront sur le même annuaire utilisateurs.

L'annuaire utilisateurs devra pouvoir être alimenté de manière mixte :

- par récupération des annuaires grâce à des API ou des commandes en lignes permettant l'écriture de scripts
- par synchronisation entre annuaires, basée sur des protocoles ouverts (LDAP, LDUP)
- par importation de fichiers dans un format standard, le plus courant est le format de fichier LDIF; le format pourra être également du CSV
- par saisie manuelle (interface web).

Dans le cas d'une solution intégrée autour d'un seul progiciel du marché, cet annuaire pourra éventuellement être un annuaire interne à cette solution. Il devra offrir une interface LDAP.

On distinguera différentes catégories d'utilisateurs :

- les utilisateurs des services d'archives type grand public,
- les utilisateurs des différents services producteurs habilités à procéder à des versements ou à commander des communications pour le compte de leur entité d'appartenance.
- les archivistes : droits sur les archives versées et leurs métadonnées, pour l'élimination et l'ajout de données
- les administrateurs : ensemble des droits plus l'administration de la base et des comptes des utilisateurs (création et suppression de comptes, modifications de profil)

L'annuaire permettra par conséquent de gérer des groupes en fonction desquels seront gérés les accès aux différentes fonctions de la plate-forme ainsi qu'aux informations accessibles dans la base archives.

Application de gestion de versement

L'application de gestion des versements doit se composer :

- d'une base de données recensant l'ensemble des versements effectués et en cours de traitement,
- d'une interface utilisateur de type " web " permettant aux services versants de :
 - constituer si nécessaire le lot à archiver (fichier XML) en saisissant les métadonnées requises par le standard d'échange et en sélectionnant les fichiers à transférer,
 - référencer leur transmission depuis un site distant,
 - rattacher à la transmission l'ensemble des objets versés (Fichiers XML),
- d'une interface utilisateurs à destination des services d'archives permettant de consulter et de contrôler les versements. Cette interface devra intégrer un parser XML permettant de lire (la consultation du bordereau de transfert XML doit être possible à la fois sous forme native et via une feuille de style fournie par le titulaire) et de contrôler les fichiers transmis par les services versants², [détailler les contrôles nécessaires : conformité au schéma du standard d'échange de données pour l'archivage, visualisation des métadonnées, contrôle des empreintes, contrôle du format des fichiers, contrôle de la taille des fichiers, contrôle des dates...]
- d'une fonction permettant d'exporter les données descriptives issues d'un lot et de convertir si nécessaire les fichiers en Base64 (en cas d'encapsulation dans un fichier XML) vers leur format d'origine,
- d'une fonction de journalisation des messages échangés entre le service d'archives et le service versant (notifications, accusés de réception, etc.).

L'application de gestion des versements devra proposer, à l'usage du service versant comme du service d'archives, des **fonctions de conversion de formats** permettant par exemple de convertir des lots de documents encodés selon des formats fermés, vers des formats ouverts ou standards (par exemple pour le texte TXT, PDF/A, un fichier structuré en XML, ou pour l'image, les formats TIFF, PNG ou JPEG). Le format originel après conversion sera conservé au même titre que le document dans son nouveau format. Toute l'opération de conversion doit être tracée et journalisée.

L'application de gestion de versements devra permettre de paramétrer les différentes tables : table des services versants, table des services d'archives, tables des différents types d'identifiants, tables des différents types de dates, tables des formats, tables des codes et nomenclatures métier, table des tailles, table des algorithmes, table des méthodes de mise sous forme canonique, tables des durées de conservation, tables des délais de communicabilité ...

L'accès à cette application s'effectue exclusivement après authentification. Cette authentification s'appuie sur la gestion des utilisateurs effectuée dans l'annuaire présenté ci-dessus.

Cette application est interfacée :

² Suivant le niveau d'automatisation des tâches souhaitées, le parser XML pourra soit être intégré dans l'application, soit être installé localement sur les postes des utilisateurs concernés.

- d'une part avec le logiciel de gestion du stockage afin de transmettre à celui-ci les fichiers à conserver et d'obtenir en échange les références nécessaires à la localisation physique ou électronique des fichiers une fois archivés ;
- d'autre part avec la base de gestion des données descriptives et la base de connaissances afin de transmettre à celle-ci les données descriptives associées aux lots versés ainsi que les références fournies par le logiciel de gestion du stockage.

Toutes les tables doivent pouvoir se croiser pour récupérer un objet archivé. Elles peuvent toutes, en tant que de besoin, être modifiées dans leur contenu par l'administrateur et le personnel des archives habilitées.

Les différents formulaires pour les opérations de versement/élimination/communication, doivent pouvoir être éditables et exportables dans un format bureautique.

Base de gestion des données descriptives

La base doit se composer :

- d'une base de données gérant les informations descriptives relatives à l'ensemble des contenus archivés. Cette base de données pourra soit être une base de données relationnelle, soit une base de données XML
- d'une fonction d'importation des données issues des bordereaux XML mis à disposition par le système de versement
- d'une interface utilisateurs permettant aux services d'archives de gérer et de modifier les données durant le cycle d'archivage des contenus
- d'une interface utilisateurs de type " web " permettant aux services d'archives, aux services producteurs et au public d'effectuer des recherches, de consulter les données descriptives disponibles sur chaque lot et fichier archivé, et de passer commande de restitution d'un ou plusieurs fichiers

La fonction d'importation des données issues des bordereaux XML consistera à alimenter la base automatiquement avec à la fois les données descriptives et les références fournies par le système de stockage concernant la localisation des contenus associés.

L'interface de " gestion " permettra au service d'archives de modifier et ajouter certaines méta-données associées aux contenus archivés. Un historique devra permettre de tracer l'ensemble des modifications effectuées.

L'accès aux interfaces utilisateurs n'est possible qu'après authentification (avec une valeur par défaut "grand public").. Quels que soient les délais de communicabilité associés aux fonds archivés, les informations descriptives sont publiques par défaut. Le service d'archives pourra néanmoins décider, pour l'ensemble des fichiers décrits ou pour un fichier donné, de rendre une ou plusieurs méta-données associées non consultables par le public avant l'expiration du délai de communicabilité. Dans ce cas, le service producteur concerné disposera toujours d'un accès à l'ensemble des méta-données. La base archives devra également gérer les droits de restitution associés aux différents contenus archivés. Ces droits dépendent du délai de communicabilité et du service versant. La base archives devra donc gérer une table de correspondance entre le nom du service versant fourni lors du versement et le groupe utilisateurs correspondant géré dans l'annuaire.

Cette fonction se matérialisera par la possibilité ou l'impossibilité pour l'utilisateur d'effectuer une commande pour un fichier donné lorsque celui-ci consulte les informations descriptives correspondantes dans la base archives.

L'interface de " recherche et consultation " permettra d'effectuer une recherche pouvant si nécessaire être affinée au fur et à mesure.

L'interface de recherche permettra à l'utilisateur d'effectuer :

- soit une recherche " texte libre " à l'aide d'un ou plusieurs mots clés et éventuellement d'opérateurs booléens (ET, OU, NON, SAUF, ...). Dans ce cas, l'ensemble des informations descriptives est pris en compte par la recherche ;
- soit une recherche ciblée sur une ou plusieurs métadonnées. Dans ce cas, l'utilisateur spécifie les métadonnées sur lesquels il souhaite effectuer sa recherche et renseigne les critères associés (voir la partie **moteur de recherche** dans les fonctionnalités de la plate-forme).

Il doit être possible de faire des requêtes précises sur les formats des fichiers stockés, en vue d'opérations de migration de ces formats par exemple.

Application de gestion des commandes

Le logiciel de gestion des commandes se compose :

- d'une interface utilisateur permettant la création d'une commande avec la liste des archives à commander,
- d'une base de données recensant l'ensemble des commandes reçues,
- d'une interface utilisateurs permettant au service d'archives :
 - de consulter et de modifier le bon de commande,
 - de contrôler et préparer les lots de restitution correspondants en fonction du mode de communication choisi³

Base de connaissances

C'est une base de données qui se compose de l'ensemble des informations de référence permettant d'explicitier, caractériser, contrôler les données archivées : table des formats de données, codes et nomenclatures (notamment pour les données extraites de bases de données). Il s'agit en définitive de l'ensemble des informations qui n'intègrent pas la base de données descriptives.

³ En fonction du niveau d'automatisation des tâches prévu, les outils associés à cette fonction peuvent soit être intégrés dans l'application de gestion des commandes soit être implémentés localement sur le poste de l'utilisateur concerné

8.3 Sécurité des données

8.3.1 Nombre de sites

Les données archivées seront répliquées sur 1 site : le centre des archives contemporaines de Fontainebleau pour le pilote .

Il est à noter que la duplication ne dispense pas d'une sauvegarde périodique dont le contenu devra être garanti.

Les deux sites présentés ci-dessus doivent être géographiquement distincts. Les paquets d'information archivés existeront ainsi sur deux supports différents dans deux endroits différents. Ceci ne tient évidemment pas compte des possibilités de sécurité offerte directement par les systèmes de stockage eux-mêmes comme par exemple la technologie RAID qui duplique automatiquement les données. De ce fait et en supposant que l'on dispose d'un tel système sur les site 1 et 2, outre la copie de sauvegarde, les paquets d'informations archivés seront disponibles en quatre exemplaires (2 sur le site 1, 2 sur le site 2,).

Par conséquent, le système proposé devra permettre :

- la mise à disposition d'un matériel équivalent en secours sur le site numéro 1 qui pourrait indifféremment servir à remplacer l'un des serveurs identifiés dans le schéma précédent ;
- la mise en place d'un système totalement redondant sur le site numéro 1, tant pour le stockage qu'au niveau des accès assurant ainsi une continuité de service sauf incident au niveau global du site ;
- la mise en place d'un accès pour les utilisateurs sur le site numéro 2 avec un système de basculement automatique d'un site à l'autre. Il sera dès lors également possible de répartir la charge des accès des utilisateurs sur les deux sites. Un dimensionnement fin devra être opéré afin de tenir compte de ces trois éléments que sont :
 - le nombre d'accès simultanés au niveau des producteurs, services versants ;
 - le nombre d'accès simultanés au niveau des utilisateurs ;
 - les coûts générés par l'ouverture des accès sur le site numéro 2.

8.3.2 Interopérabilité du système

Une attention particulière sera portée sur l'interopérabilité des systèmes à installer. Un changement de plate-forme de stockage doit en effet être transparent pour les utilisateurs et l'organisation logique des archives. De même il doit être possible de changer une brique du système sans devoir tout remettre en cause.

La conception globale de ce dernier doit ainsi prévoir l'interopérabilité par une organisation si possible en couches logiques, quasiment indépendantes les unes des autres, ce qui autorise leur changement sans perturber l'ensemble. On pourra ainsi proposer une **solution modulaire** intégrant des briques logicielles spécialisées, chacune de ces briques étant intégrée via une **couche**. Cette **fédération de composants** facilitera l'intégration au fil des évolutions technologiques, de briques nouvelles répondant aux mêmes contrats de service offrant ainsi la possibilité de remplacer une brique.

8.3.3 Gestion de la duplication des informations

L'objectif principal est d'assurer la non-perte de données (redondance locale et réplication sur un site distant). Le titulaire précisera le taux de disponibilité des données suivant l'architecture proposée (réplication avec ou sans accès doublé par exemple), soit le temps nécessaire pour pouvoir accéder à nouveau au système et aux objets archivés :

- Type de stockage, simple ou doublé. Il s'agit en fait de l'utilisation ou non de technologies de type RAID ou autres, permettant d'avoir l'équivalent d'une réplication systématique et locale de l'information stockée : ainsi la technique RAID1 correspondant à la notion de miroir elle sera privilégiée pour les objets archivés et les applicatifs
- Accès vis-à-vis de l'extérieur, simple ou doublé. Ceci revient quasiment à doubler les équipements correspondant aux accès
- Serveur applicatif de secours, présence ou non
- Type de réplication (en continu, synchrone).

9. Contraintes et exigences techniques

9.1 Pré-requis technique : Standards pour exploitation, réseau et postes de travail

Suivi des anomalies

Le suivi des anomalies de l'application pendant les phases de développement, test, recette et maintenance sera effectué à l'aide d'un outil fourni par le ministère. Cet outil générique est accessible via une interface web. Il est hébergé et administré par le ministère, et est accessible :

- en interne sur le réseau culture via <http://mantis.culture.fr>
- sur internet via <https://mantis.culture.fr>

L'accès à l'outil pour les intervenants extérieurs nécessite la création d'un compte invité sur l'intranet du ministère. L'accès à l'intranet et à l'outil de suivi des anomalies est bien entendu soumis à la cause de confidentialité décrite ci-après.

Messagerie électronique

Si l'application doit envoyer des courriers électroniques, elle doit le faire via le protocole SMTP, en s'interfaçant avec les serveurs de messagerie du ministère (type sendmail ou postfix).

L'envoi de courriers électroniques peut être sujet à des restrictions, en particulier :

- ⇒ l'envoi de grandes quantités de courrier est limité, typiquement si on souhaite distribuer un courrier à de nombreux destinataires, il faut prévoir de lancer les envois par lots,
 - ⇒ la gestion des listes de diffusion est centralisée dans un outil spécialisé Sympa (<http://www.sympa.org/>).
- Une application gérant des listes de diffusion doit donc s'interfacer avec cette gestion centralisée.

Référentiels

Lorsque l'application fait référence à des éléments identifiables par un code INSEE, utiliser ce code en tant que clé. Voir le tableau ci-après qui liste des éléments géographiques qui doivent impérativement être référencés selon la nomenclature INSEE.

Code	Type	Commentaire
Région	2 chiffres	
Département	Alphanumérique, longueur 3	Alphanumérique à cause de la Corse (2A, 2B) et de longueur 3 pour les DOM.
Arrondissement	1 chiffre	Groupement de cantons
Canton	2 chiffres	Groupement de communes
Commune	5 chiffres	Attention, le département est codé sur les 2 premiers chiffres pour les départements métropolitains, mais sur 3 chiffres pour les DOM.

9.2 L'environnement informatique du centre des archives contemporaines (le service Constance)

Protocole : TCP / IP

Transfert de fichiers par ftp via ftp.culture.fr sur comptes spécifiques soumis à déclaration

Système d'exploitation serveur : LINUX

Navigateur pour l'INTRANET (dont les outils de contribution) : mozilla firefox et si nécessaire IE6
Sur les postes utilisateurs : Windows XP, suite bureautique Office 97 ou Open Office
liaison extérieure transfix 1920K
réseau local ethernet suivant la spécification IEEE 802.3 en 10BaseT seulement

9.3 Standards de développement

- Outils de développement : Java ou éventuellement PHP
- Pour les outils Java, framework de développement conforme à J2EE et architecture 3 niveaux selon J2EE
- Modélisation en UML

Le Ministère de la culture et de la communication n'impose pas d'outil de développement particulier. En revanche, il est important que le code source de l'application puisse être manipulé sans environnement de développement intégré (IDE) particulier. Cette exigence ne pose généralement pas de problèmes particulier concernant le code lui-même car dans la plupart des technologies utilisées au ministère (PHP, Java, Zope...) le code est stocké dans des fichiers textes lisibles avec un simple éditeur (Notepad, vi...).

En revanche les fichiers de « projet » qui permettent d'agréger les sources et stocker des méta-informations, des procédures de compilation, sont plus problématiques. Quel que soit l'outil de développement les fonctionnalités de type compilation ou déploiement doivent être accessibles sans disposer d'aucun environnement de développement intégré particulier.

Pour ce qui est de la compilation par exemple, il faut fournir des Makefile, des tâches Ant, ou tout autre moyen standard adapté permettant de lancer la compilation de l'intégralité des binaires. La livraison d'un fichier de « projet » associé à un environnement intégré, quel qu'il soit, ne convient pas.

La génération de code automatique est proscrite si elle dépend d'un outil de développement intégré. En effet, tout code ainsi généré serait non manipulable par un opérateur qui ne disposerait pas de l'environnement en question.

Le ministère doit disposer de l'intégralité du code source de l'application, ainsi que de l'intégralité des éléments permettant de générer les binaires.

L'emploi de briques logicielles opaques, notamment de fichiers .class Java pré compilés, dont le code source ne serait pas disponible, doit être motivé et justifié, car ce type de composant peut s'avérer bloquant lors des inévitables mises à jour du système.

Les composants binaires dont le source est indisponible (bibliothèques client Oracle par exemple) doivent être signalés.

9.4 La plate-forme logicielle et matérielle

Le Ministère de la Culture et de la Communication supporte les bases de données suivantes :

- MySQL 4.0 ou 4.1, sous Mandrake Linux
- PostgreSQL 7.4 (distribution) ou 8.0 (PostGIS 1.0 pour le stockage d'informations géographiques)
- Oracle v9, sous AIX

Il est possible de développer l'application en utilisant l'une ou l'autre de ces bases de données. Le choix doit être argumenté, et le serveur choisi en fonction de l'adéquation des possibilités du SGBDR avec l'application.

Les coûts induits par la mise en place de ces serveurs (licences, paramétrages spécifiques nécessitant une expertise) seront pris en compte par le ministère.

Dans le cas d'Oracle, la base de données est installée sur un serveur tiers, sous AIX, dédié à l'hébergement de bases de données.

L'application doit s'intégrer à la plate-forme d'exploitation du ministère.

Pendant la phase de réalisation de l'application, les développements devront rester compatibles avec les mises à jour de sécurité des logiciels utilisés (système d'exploitation, serveur web, moteur de servlets, base de données...).

Cette compatibilité devra aussi être conservée lors de la phase de garantie.

Pour les applications Web, l'accès à l'application se fera via les protocoles HTTP ou HTTPS exclusivement, à l'aide d'un navigateur qui répond aux standards HTML 4.01 et XHTML 1.0, avec des feuilles de style CSS1 ou CSS2.

Les postes utilisateurs utilisent potentiellement n'importe quel système d'exploitation (Windows, MacOS, Linux,...), et ce même pour les opérations de saisie, car la saisie peut être effectuée en dehors du réseau du Ministère, par des établissements publics par exemple, dont on ne maîtrise pas la plate-forme.

L'intérêt d'une application web étant entre autres de minimiser les procédures de déploiement, l'application ne devra pas, sauf mention contraire et explicite, nécessiter la présence de logiciels particuliers sur le poste client, hormis un navigateur web. Ainsi, par exemple, on ne peut pas supposer arbitrairement que tous les postes clients disposent d'une suite bureautique. En particulier, le ministère insiste sur le fait que les formats .doc, .xls., .ppt (Microsoft Office) ne font pas partie des standards du W3C, et ne peuvent pas être directement interprétés par un navigateur web.

Tout système équipé d'un navigateur compatible HTML 4.01 / XHTML 1.0 doit pouvoir se connecter à l'application, et exploiter la totalité des fonctionnalités disponibles.

Concernant le serveur d'application :

Plate-forme J2EE. Le ministère supporte le sous-ensemble de la plate-forme J2EE connu sous le nom de container de servlet. La plate-forme d'exécution des servlets choisie est Tomcat (versions 4.1.X, 5.0.X supportées), avec un JDK 1.4.2 ainsi que 5.5 avec JDK 1.5.

Zope 2.7. Zope 2.7 est utilisé avec CPS3 dans le cadre de la gestion de contenu et de la publication sur internet.

PHP 4.3.

Autres :

Apache 2.0
OpenLDAP 2.1 ou 2.2
Squid 2.4
Pound 1.6

Contexte d'exploitation : il est impératif que l'applicatif soit :

"scalable" dans le sens où ses performances doivent pouvoir être améliorées en ajoutant de nouvelles machines

utilisable en environnement mutualisé. Ceci implique en particulier :

- l'absence totale de références à des chemins "en dur" de type "/opt/application_xyz/bin/programme" dans le code, sauf dans des fichiers de configuration clairement identifiés
- la possibilité de placer un "reverse proxy" (Pound, Squid, Apache) de manière transparente en amont, avec éventuellement une réécriture des URLs, un changement de port IP, un passage en SSL. Il est important de prendre en compte ces contraintes dès le début du développement. Penser que la même application sera potentiellement accessible par plusieurs URLs, par exemple on peut avoir les 4 combinaisons:
 - ⇒ <http://appli.culture.fr/>
 - ⇒ <https://appli.culture.fr/>
 - ⇒ <http://saisie-appli.culture.fr/>
 - ⇒ <https://saisie.culture.fr/appli/>

Cette liste n'est pas exhaustive. D'une manière générale, soit la «racine» du site n'a aucune incidence sur le code, soit elle doit être totalement paramétrable et multivaluée.

Normes de réalisation liées au Web

Le code HTML généré doit respecter la norme XHTML 1.0 du W3C. Ce code devra passer les tests de validation du W3C (<http://validator.w3.org>). Les feuilles de style (CSS1 ou CSS2) seront utilisées pour la mise en forme des pages. Les tableaux (tag « <table> ») ne doivent pas être utilisés pour la mise en forme générale des pages, mais seulement pour la présentation de données sous forme de tableau. De même

l'utilisation d'images transparentes ou autres astuces similaires pour la mise en forme est à proscrire. Il faut utiliser les feuilles de style. Les frames sont interdites.

Le code HTML généré doit au moins respecter les règles d'accessibilité de niveau1 énoncées par la WAI (<http://www.w3.org/WAI/>). Le Ministère insiste sur le fait que ces règles n'interdisent pas mais limitent fortement les possibilités offertes par JavaScript. Le respect des règles niveau 2 et niveau 3 est souhaitable. Se référer aux exigences détaillées en terme d'accessibilité.

Les divers entêtes HTTP - en particulier les délais d'expiration - doivent être correctement renseignés, afin d'éviter les problèmes avec les caches, et aussi prévenir les incohérences liées à l'utilisation de la fonction « retour » du navigateur Web.

On privilégiera une mise en page à base d'onglets. L'utilisation de fenêtres de type popup est à réserver à des cas exceptionnels, par exemple l'aide à la saisie (ce type de fenêtres est de toutes façons souvent incompatible avec les critères d'accessibilité).

Les vérifications de conformité de champs (format, valeurs limites, cohérence,...) doivent être faites systématiquement côté serveur. Les éventuels contrôles côté client (par JavaScript ou autre) ne sont que des aides à la saisie qui ne sauraient se substituer à une vérification forte et systématique côté serveur.

Le poids total des pages HTML ne devra pas excéder 50ko, sauf si le dépassement est justifié par l'affichage de données particulièrement volumineuses.

Formats de fichiers

Concernant les éditions et les exports de documents, on se limitera à des formats ouverts : RTF ou PDF pour les documents texte, CSV pour les tableaux.

Pour les impressions simples, privilégier la mise en place de feuilles de styles spécifiques à l'impression (media="print") à toute autre solution. Pour les impressions complexes, passer par un fichier PDF.

Gestion des accès concurrents

Concernant les ressources qui ne doivent pas être modifiées simultanément par 2 utilisateurs différents, un mécanisme de verrou exclusif doit être mis en place pour réserver l'accès à un document pendant sa modification. Ce mécanisme doit permettre de gérer des scénarios du type :

- l'utilisateur prend la main sur le document, et le réserve de manière exclusive. A partir de cet instant, aucun autre utilisateur ne peut prendre la main et modifier ce document, mais le document est toujours consultable.
- l'utilisateur modifie des données en exécutant plusieurs requêtes HTTP (il voit plusieurs pages donc), sans que la base de données soit modifiée (ou bien seulement dans le cadre de tables de travail temporaires). Pendant cette phase, le document reste consultable par les autres utilisateurs, qui voient « l'ancienne » version.
- l'utilisateur valide ses modifications, qui sont alors intégrées définitivement dans la base de données. Le verrou est alors levé et un autre utilisateur peut travailler sur ce document.

Lorsque les modifications peuvent être effectuées en une seule requête HTTP, avec une seule transaction atomique, ce mécanisme n'a évidemment pas besoin d'être mis en place.

Sécurité

L'application devra prendre en compte le fait qu'on ne peut pas faire confiance aux données provenant du poste client. En particulier, veiller à respecter les recommandations énoncées dans le document « CERTA-2004-INF-001 » publié par le CERTA :

<http://www.certa.ssi.gouv.fr/site/CERTA-2004-INF-001/index.html>

Ces recommandations couvrent les attaques suivantes :

- utilisation des méta-caractères
- failles de codage HTML, PHP, ASP
- attaques de type « SQL injection »

D'une manière générale, quelles que soient les requêtes HTTP envoyées au serveur, celui-ci ne devra pas renvoyer plus d'informations que celles auxquelles le client a droit, ni offrir plus de fonctionnalités.

Aucune information technique dévoilant le fonctionnement interne de l'application, son paramétrage, ne doit être affichée au client. Ainsi les messages d'erreur par défaut affichés par PHP, Tomcat ou Zope, bien qu'ils restent utiles en développement, doivent être désactivés en production, et masqués par une page d'erreur propre à l'application. Les pages ou répertoires par défaut (page d'accueil Apache, Zope ou Tomcat par exemple) doivent aussi être rendus inaccessibles.

Les événements pertinents en terme de sécurité, par exemple une connexion en tant qu'administrateur global de l'application, une suppression de données non réversible, doivent être inscrits dans le fichier syslog du système. L'idée n'est pas de noyer syslog avec tous les événements, ni de remplacer les fichiers de logs propres à l'application, mais de faire remonter les actions critiques uniquement.

Temps de réponse

L'application doit respecter tous les critères de performances décrits ci-après. Le respect des temps de réponse en local ne se substitue pas à celui des temps de réponse en situation réelle, et vice-versa.

En local :

Sur une machine de type :

- Processeur x86 2GHz
- 1Go de RAM

et dans les conditions suivantes :

- fonctionnement purement local, sans passer par le réseau
- la machine ne traitant qu'une et une seule requête à la fois
- aucun cache web de type Squid

les temps de réponse doivent être inférieurs au 1/10ème de seconde (temps mesurés au niveau du serveur sans tenir compte des temps d'affichage liés au navigateur ou au chargement des images, avec un outil de mesure de type ApacheBench)

Dans le cas particulier d'opérations complexes (éditions, statistiques, ...) des temps de traitement supérieurs sont acceptables, sans toutefois jamais dépasser 3 secondes.

En fonctionnement normal : Sur une machine de type :

- ⇒ Processeur x86 2GHz
- ⇒ 1Go de RAM

et dans les conditions suivantes :

- sur le réseau du Ministère (Paris)
- avec plusieurs utilisateurs simultanés
- cache web de type Squid autorisé

le système doit être capable de:

- traiter 10 requêtes en parallèle, c'est-à-dire ne pas sérialiser les opérations mais avoir au moins 10 files d'attente de traitement distinctes
- absorber une charge moyenne d'au moins 2 requêtes par seconde en régime permanent
- maintenir des temps de réponse systématiquement inférieurs à 3 secondes
- être capable d'absorber une charge supérieure moyennant l'ajout de nouvelles machines

Les temps donnés ici sont mesurés au niveau du client, et prennent en compte le chargement des images. Ils peuvent être mesurés avec un outil de type OpenSTA par exemple.

Le terme requête désigne ici un chargement de page complet, c'est à dire la requête HTTP principale qui renvoie le code HTML de la page plus toutes les requêtes HTTP associées (images par exemple) ainsi que les requêtes exécutées par la base de données. Une requête peut donc englober plusieurs ordres HTTP et SQL, selon la manière dont est codée l'application.

10. Politique de sécurité à mettre en œuvre

La politique de sécurité à mettre en œuvre vise à :

- prévenir les atteintes à l'intégrité et à la confidentialité des données archivées, des informations de gestion et des données relatives aux personnes (utilisateurs, usagers),
- garantir un haut degré de disponibilité et la continuité de service,
- offrir des points d'accès aux informations et aux fonctions nécessaires aux contrôles de sécurité et aux réactions en cas d'incident de sécurité.

Pour ce faire, le prestataire devra intégrer dans le pilote les éléments techniques nécessaires au respect des clauses décrites dans le présent chapitre.

10.1

Organisation de la sécurité

Le pilote doit comporter une interface permettant de gérer les droits d'accès des utilisateurs et des accès aux différents processus et sous-processus. Cette gestion des droits est possible à au moins deux administrateurs des utilisateurs et des droits.

Le pilote confronte le degré de communicabilité et le profil du demandeur avant toute diffusion d'information. L'accès aux journaux d'événements (système et applicatifs) est autorisé aux responsables de la sécurité du système (AQSSI du centre, RSSI, FSSI), en particulier lors de contrôles, après incidents ou en période de crise.

Les coordonnées (URL, listes de diffusions, etc.) des sources d'informations permettant de suivre l'évolution de la sécurité (annonce de failles, correctifs...) des produits intégrés dans le pilote sont fournies par le prestataire. Si un abonnement ou une inscription est nécessaire, le prestataire fait la démarche initiale et donne au service d'archives et au DSI les moyens de poursuivre l'abonnement ou de proroger l'inscription.

10.2

Cloisonnement

L'architecture du pilote permet d'isoler le système d'accès pour les usagers des éléments internes (archives, modules de gestion), même dans l'hypothèse d'une salle de lecture publique dans les locaux du SAE.

L'architecture et l'écriture du pilote permettent l'interposition de relais mandataire ("*reverse proxy*") entre l'utilisateur ou l'utilisateur et le pilote. Cette interposition permet un éventuel filtrage amont des accès, masque à l'utilisateur l'architecture physique et facilite l'évolution de cette architecture.

Les flux ne dérogent pas à l'architecture du SI :

- les flux de messagerie utilisent le circuit défini par l'architecture de messagerie (en 2006 au MCC, flux SMTP et POP3 vers/depuis le serveur de messagerie du site),
- les flux HTTP, FTP, Telnet, SSH transitent par le relais mandataire (proxy) en usage pour chaque flux, et sont soumis aux règles de filtrage communes au ministère.

La liste des flux est reportée dans la documentation technique du pilote. La matrice des flux entre les éléments du pilote et entre le pilote et les autres composantes du SI figure également dans cette documentation.

Lorsque des mécanismes cryptographiques sont nécessaires, le choix des algorithmes et des paramètres (ex. : longueur des clés) obéit aux recommandations de la DCSSI.

10.3

Développement et vie du pilote

Les standards de développement et de codage du prestataire sont intégrés ou annexés au plan assurance qualité.

Pendant les phases de développement, de recette et de garantie, le pilote reste fonctionnel lors des applications des correctifs de sécurité publiés par le constructeur. Si une incompatibilité apparaît, le titulaire dispose de 72 heures pour adapter le pilote ou trouver une solution de contournement interne au pilote apportant le même degré de sécurité que l'application du correctif.

La construction du pilote permet systématiquement :

- de contrôler les entrées syntaxiquement (type, longueur, grammaire des expressions logiques...) et, lorsque les valeurs possibles sont définies, de contrôler la validité, et de gérer les entrées incorrectes,

- de contrer les dépassements de capacités (" *buffer overflow* "),
- de traiter de manière adéquate les caractères spéciaux (parenthèses, apostrophe et apostrophes inversées...),
- pour les sites web, de se rendre indépendant de la configuration du navigateur client acceptant HTML 4 et CSS 2, en particulier sur l'acceptation ou non des *cookies*, des scripts et des codes distants (applettes ou *activex*),
- de masquer les technologies employées, sur les serveurs en production, par suppression des informations techniques sur les bannières de connexion et sur les pages d'erreur, par la suppression des pages par défaut, par l'inaccessibilité des pages d'exemples, de la documentation en ligne, etc.
- de détecter les portes dérobées,
- de détecter la présence de codes malveillant dans les informations traitées (ex. : antivirus analysant les documents soumis à l'archivage),
- la confidentialité des éléments secrets (ex. : stockage d'un mot de passe sous forme de condensé, effacement sécurisé d'une clé secrète dès qu'elle n'est plus utilisée, même en mémoire...),
- de produire des traces (journal d'évènements) dans un format standard (CLF, ELF), et, si besoin, de les transférer sur un serveur dédié).

La télémaintenance du pilote n'est possible que depuis le même domaine de sécurité (que depuis l'intérieur du réseau du ministère).

La conception du pilote et les logiciels intégrés doivent pouvoir être maintenus par un tiers au marché de réalisation.

10.4

Continuité de service

Le pilote doit assurer une continuité de service dans les conditions définies précédemment.

Le prestataire évalue l'impact technique et fonctionnel de la défaillance des composantes du pilote (matériels, logiciels, interruption d'un flux réseau, coupure d'un lien réseau) compte tenu des technologies employées.

Il rédige et teste avec les exploitants, selon l'architecture technique et fonctionnelle retenue, les procédures d'arrêt, de redémarrage, de repli (exemple arrêt des seuls accès usagers), de bascule, de sauvegarde, de restauration, de reprise d'activité.

10.5

Exploitation

Le prestataire rédige les procédures d'exploitation en y intégrant les aspects SSI : gestion des éléments de sécurité (mots de passe, clés, sauvegardes traces...) et les procédures de mise à jour.

La téléadministration et la télémaintenance du pilote ne sont possibles que depuis le même domaine de sécurité (que depuis l'intérieur du réseau du ministère).

La configuration met en œuvre la défense en profondeur : filtrage, contrôle, traces au niveau du réseau, des systèmes d'exploitation et des applications.

Le pilote intègre les outils nécessaires à la destruction sécurisée des données, à la réutilisation de supports amovibles sans compromission des informations précédemment contenues sur lesdits supports.

10.6 Identification/authentification

Identification

L'identification des utilisateurs et des usagers est systématique, que la connexion ait lieu par une application, par le système d'exploitation ou par un logiciel du pilote. Pour le portail grand public, une option "grand public" sera défini par défaut.

A l'exception de cet usager, toutes les identifications s'accompagnent d'une authentification :

- au minimum " faible " (identifiant, mot de passe), de préférence " renforcée " pour les consultations,

- impérativement " forte " (cryptosystème à clé publique) pour les opérations sensibles (administration, destruction d'archives...) à définir lors des spécifications détaillées.
- Les connexions (authentification réussie) et les tentatives (échec de l'authentification) sont journalisées, ainsi que les opérations de gestion des comptes : création, blocage, déblocage, changement de mot de passe, clôture.

Le pilote prévient au moins les attaques élémentaires par recherche exhaustive :

- un nombre seuil, paramétrable et valant 3 par défaut, d'échecs de connexion sur un compte bloque provisoirement l'accès au compte et envoie sans délai une alerte à l'administrateur du pilote,
- un nombre seuil, paramétrable et valant 3 par défaut, d'échecs de connexion depuis une même source (adresse IP, machine), ou avec le même mot de passe, envoie sans délai une alerte à l'administrateur du pilote.

Le pilote assure que les mots de passe répondent aux critères suivants :

- longueur minimale paramétrable, par défaut 8 caractères,
- complexité minimale : au moins un chiffre, au moins une majuscule, au moins une minuscule,
- non présence dans un dictionnaire classique en la matière,
- dissemblance avec l'identifiant auquel il est associé,
- possibilité de mettre des caractères spéciaux dont la liste est précisée,
- durée vie minimale paramétrable, par défaut d'un jour.

Pour les utilisateurs, le pilote assure également que les mots de passe répondent aux critères suivants :

- durée de vie maximale paramétrable, par défaut 90 jours,
- différent des derniers mots de passe précédemment utilisés, par défaut des 4 derniers.

Le prestataire peut proposer dans sa réponse, pour les utilisateurs agents du ministère de la culture et de la communication, l'utilisation du mot de passe contenu dans l'annuaire du ministère et les modalités techniques qui la conditionnent.

Lorsqu'un utilisateur ou un usager choisit un mot de passe, celui-ci doit être confirmé. En cas de divergence entre la première saisie et la saisie de confirmation, le processus est repris depuis le début.

Le pilote propose une génération de mot de passe aléatoire avec l'interface de renouvellement de mot de passe.

En cas de perte de son mot de passe, l'utilisateur devra entreprendre les mêmes démarches que pour l'ouverture de son compte. Il n'y a pas de fonction " J'ai oublié mon mot de passe ".

L'utilisateur ou l'usager est averti de la proximité de l'expiration de son mot de passe quand il se connecte à partir d'un délai paramétrable qui est de 14 jours par défaut.

Le pilote vérifie que ce délai est compatible avec les paramètres de durée de vie minimale et maximale. Dans la négative, une erreur à l'installation ou au rechargement de la configuration alerte l'administrateur du pilote et l'installation ou la modification de configuration n'est pas prise en compte.

Les certificats utilisés pour l'authentification forte sont conforme à le PRIS v1, niveau ** (certificats des porteurs).

Les certificats de même qualité sont utilisés par les services pour identifier et authentifier l'origine des transferts (service versant vers SAE, par exemple) et pour l'authentification de serveurs.

L'autorité de certification pour ces certificats est à terme soit celle du SAE, soit celle du ministère de la culture, soit une autorité interministérielle.

Authentification. Principe

Les mécanismes d'authentification s'appuieront sur le protocole LDAP V3. Très concrètement, la liste des utilisateurs du système ainsi que leurs mots de passes seront stockés dans un annuaire LDAP, commun à plusieurs applications.

A minima, l'application devra utiliser LDAP pour authentifier les personnes.

Le Ministère fournira le détail du schéma LDAP à utiliser, qui est relativement standard. En effet les utilisateurs sont de type MCCPerson, dérivé de inetOrgPerson, et l'authentification se fait à l'aide des champs uid et userPassword. Tous les utilisateurs sont stocké dans un même « bag ».

Ils doivent impérativement correspondre à des personnes réelles. Ainsi, l'utilisation de comptes fonctionnels du type « responsable du service X » est proscrite. Pour gérer ce type de problématique, il faut utiliser une gestion des rôles appropriée, qui associe les personnes et les fonctionnalités. Voir le paragraphe « gestion des droits ».

Le serveur LDAP est accessible en lecture seule.

Informations techniques

Paramètre	Valeur
Serveur LDAP	OpenLDAP 2.1 ou 2.2 avec backend BDB
DN de base pour les personnes	ou=personnes,o=gouv,c=fr
Champ utilisé pour l'identification	uid
Champ utilisé pour le mot de passe	userPassword
Exemple du dn d'une personne	uid=jean.dupont.culture.gouv.fr,ou=personnes,o=gouv,c=fr

Extrait du schéma LDAP :

```
objectclass ( 1.3.6.1.4.1.10696.5.2.1 NAME 'MCCPerson'
  DESC 'MCC Person'
  SUP inetOrgPerson
  STRUCTURAL
  MAY ( gender
    $ MCCcivility
    $ MCCIntranetLogin
    $ MCCDisplayRank1
    $ MCCSiteRef
    $ MCCOuManager
    $ MCCIsGroupManager
    $ MCCOuBis
    $ MCCTitleBis
    $ MCCIsGuest
```

Gestion des droits

Lorsqu'une application nécessite plus qu'une simple authentification, et qu'une gestion des droits doit être mise en place, un composant spécifique est fourni par le ministère pour gérer les habilitations.

Le schéma d'habilitation repose sur la notion de rôle. Des rôles sont définis pour une application donnée, et à chaque utilisateur, on associe un ou plusieurs rôles. La mise en place d'utilisateurs fonctionnels ou de conventions « en dur » qui associent un utilisateur à une fonctionnalité est proscrite.

Dans ce contexte, il n'est plus nécessaire d'appeler directement l'API LDAP. Une API intermédiaire, reposant sur les notions d'utilisateur, de rôle, et d'application, est utilisée. Le fait que techniquement un annuaire LDAP soit utilisé est complètement masqué, le ministère souhaite découpler l'annuaire technique LDAP d'une part, et les applications d'autre part, se réservant ainsi la possibilité de faire évoluer cet élément technique ultérieurement.

En contrepartie de l'utilisation de cette interface, la gestion des utilisateurs et des rôles n'a pas besoin d'être gérée au sein de l'application

Le ministère dispose d'une interface d'administration, autonome, qui permet de gérer les utilisateurs, les rôles et les applications (au sens « entrée application qui définit un domaine dans le système de gestion de droit »). L'affectation des droits se fait à ce niveau.

Lorsque le modèle « application / rôle / utilisateur » ne permet pas de gérer tous les cas rencontrés dans l'application, par exemple si on souhaite attribuer des droits en fonction d'une localisation précise dont l'application seule a connaissance, ou encore si on souhaite attribuer des droits avec une granularité très fine au document près, la gestion des droits devra faire l'objet d'une étude fonctionnelle et technique en début de projet.

La documentation détaillée concernant la brique de gestion des droits sera fournie au lancement du projet. Le ministère dispose d'une implémentation en langage Java de ce composant, qui sera mise à disposition avec son code source.

Dans le cas d'application non développées en langage Java, le composant n'est pas fourni par le ministère. Un composant similaire (API identique) devra donc être développé dans le langage cible. A titre indicatif, l'implémentation en Java fait environ 4000 lignes (commentaires inclus), définit 6 interfaces et comporte 17 classes.

Exemple de code Java permettant une authentification :

```
CerbereData userData = null;
String utilisateur, password;
utilisateur = request.getParameter("utilisateur");
password = request.getParameter("password");    if( utilisateur == "" || utilisateur == null || password
== "" || password == null) {          throw new ServletException("Utilisateur ou mot de passe non renseigné
!");
}
try {

    userData = authService.authentifier(utilisateur, password);
} catch (CerbereBadUserIdException ex) {
    throw new ServletException("<H1>Utilisateur invalide !</H1>");

} catch (CerbereExpiredAccountException ex) {
    ex.printStackTrace();
} catch (CerbereBadPasswordException ex) {
    throw new ServletException("<H1>Mot de passe invalide!</H1>");
} catch (CerbereBlockedAccountException ex) {
    ex.printStackTrace();
} catch (CerbereConnexionException ex) {
    throw new ServletException("<H1>Problème de connexion</H1>");
}
}
```

10.7 Protection des accès logiques

Le pilote contrôle les accès par l'application, par le système d'exploitation et par les logiciels intermédiaires (base de données, progiciels...).

A partir du moment où un Utilisateur est identifié et authentifié, le pilote dispose d'un système contrôlant et limitant ses accès par rapport à ses droits. L'ensemble des Utilisateurs et des droits correspondants devra être contenu dans un annuaire régulièrement mis à jour en fonction des évolutions.

L'utilisation de rôles et de groupes d'utilisateurs est encouragée pour faciliter la gestion des droits.

Les allocations et les modifications de droits, réussies ou non, sont journalisées. Toute modification réussie est répercutée sans délai. La session peut s'en trouver interrompue. Dans ce cas, trace en est gardée dans le journal d'évènements.

Les accès inactifs sont inhibés (deconnexion temporisée, avec une durée paramétrable et égale, par défaut à 20 minutes). La trace de la deconnexion est conservée, avec indication du compte qui a ouvert la connexion.

10.8 Journalisation et procédures de constitution des données de traçabilité

Afin de constituer un ensemble de données à la fois suffisantes et cohérentes en matière de traçabilité, les opérations suivantes doivent être effectuées et validées avant la mise en place du pilote :

- définir une fréquence a priori minimum de traitements des journaux d'événements même si toute latitude doit être laissée à ce niveau.
- par rapport à ce dernier point, définir une période de conservation des journaux d'événements ;
- vérifier les dispositifs de sécurité mis en place et destinés à assurer la protection des journaux d'événements ;
- vérifier en particulier, en complément au point précédent, la procédure de sauvegarde des journaux d'événements.

10.9 Mesures de sécurité liées à l'intégrité des objets archivés

Sur la base du standard d'échange, des prises d'empreinte ou hachages, voire des signatures sont prévues au niveau des transferts, permettant au pilote de s'assurer que les objets transférés n'ont pas été altérés durant leur transmission.

Les certificats nécessaires pour la signature électronique seront acquis par le titulaire auprès d'un prestataire de certification. Ils devront obéir aux prescriptions de la PRIS version 1 (niveau 2).

Par ailleurs, les objets peuvent également faire l'objet d'un hachage par le service versant, qui sera ensuite vérifié par le pilote à plusieurs reprises : lors de la prise en charge, lors d'une communication, lors d'une demande d'accord pour élimination, avant une opération de migration de format, et plus généralement à intervalles réguliers par le système (voir les fonctionnalités du pilote).

Après migration du format d'un ou plusieurs objets, le pilote doit pouvoir générer une nouvelle empreinte du nouvel (nouveaux) objet(s) après migration et de les vérifier suivant les mêmes modalités.

10.10 Mesures de sécurité liées à l'horodatage des opérations d'archivage

Afin d'éviter toute remise en cause de l'horodatage réalisé par le pilote, ce dernier aura recours à deux sources de temps distinctes afin de dater les différentes opérations réalisées (Versement, Communication, Consultation, Elimination).

11. Déroulement du projet

11.1 Pilotage du projet

La maîtrise d'ouvrage est assurée par la direction des archives de France (chef de projet : Françoise Banat-Berger, chef du département de l'innovation technologique et de la normalisation).

La maîtrise d'oeuvre est assurée par le DSI et la DGME (chef de projet : Gabriel Ramanantsoavina).

Le titulaire assurera une assistance à la maîtrise d'oeuvre.

Trois instances de suivi sont prévues.

Le comité de pilotage

Il est constitué des différents acteurs du projet, ainsi que d'experts extérieurs, assistés par un représentant du titulaire du marché. Le comité de pilotage est présidé par la directrice des Archives de France ou son représentant. Il valide et arbitre les grandes orientations du projet. Le comité de pilotage se réunira selon un rythme fixé et révisable d'un commun accord. Sa composition, pour la maîtrise d'ouvrage est la suivante :

- Françoise Banat-Berger, DAF
- Olivier de Solan, DAF
- Marc Meyer, DGME – SDAE
- Gabriel Ramanantsoavina, DGME - SDAE
- Serge Novaretti, DSI ministère de la culture
- Christine Pétilat, CAC de Fontainebleau
- Carole Gragez, CAC de Fontainebleau
- Jean-Pierre Teil, CAC de Fontainebleau
- La mission des archives nationales auprès des services du Premier ministre
- Isabelle Neuschwander, directrice du projet du nouveau centre des archives nationales de Pierrefitte-sur-Seine
- Christophe Alviset, DSI Minefi
- Thierry Ehret—Franck, DSI conseil général des Yvelines
- Elisabeth Gautier-Desvaux, Archives départementales des Yvelines
- Gaëlle Mignot, Archives départementales des Yvelines
- Patrice Guérin, Archives départementales des Yvelines

L'équipe de projet

Elle est constituée de représentants d'une partie des différents acteurs du projet assistés par un représentant du titulaire du marché. Elle a pour but de coordonner les acteurs du marché, d'étudier les aspects techniques et de soumettre les problèmes. La DAF se réserve la possibilité d'organiser certaines

réunions dans les locaux du titulaire pendant la période de développement de l'application. Sa composition pour la maîtrise d'ouvrage, est la suivante :

- Françoise Banat-Berger
- Olivier de Solan
- Gabriel Ramanantsoavina
- Carole Gragez
- Jean-Pierre Teil
- Gaëlle Mignot
- Luc Vallée, chargé de la sécurité des systèmes d'information (DSI du MCC)

Le comité des utilisateurs

Pour le pilote, il est constitué par les personnes participant au transfert des données dans le pilote : producteurs, missions des archives, DGME.

11.2 Conduite du projet : Les structures de pilotage

Il appartiendra au titulaire d'assurer la conduite du projet. Dans ce cadre, il sera nécessaire de :

- ☐ Préparer les réunions techniques et les différents comités de pilotage
- ☐ Animer ces réunions et en assurer les comptes rendus
- ☐ Gérer les indicateurs de progression, les risques, les plannings et les tableaux de bord du projet

12. Définitions

- **Archive** : Paquet d'informations reçu, conservé et communiqué par un Service d'archives (cette définition issue du standard d'échange est la définition de référence dans le présent Cahier des charges).
- **archives** : documents sous forme électronique, quels que soient leur date et leur support, produits ou reçus par tout service ou organisme public ou privé, dans l'exercice de leur activité (définition issue du code du patrimoine).
- **Archive courante** : les Archives qui sont d'utilisation habituelle pour l'activité des services, établissements et organismes qui les ont produites ou reçues.
- **Archive définitive** : les Archives qui ont subi les tris et éliminations définis aux articles 15 et 16 du décret n° 79-1037 du 3 décembre 1979.
- **Archive intermédiaire** : les Archives qui ont cessé d'être considérées comme des Archives courantes et les Archives qui ne peuvent encore, en raison de leur intérêt administratif, faire l'objet de tri et d'élimination conformément à l'article 16 du décret n° 79-1037 du 3 décembre 1979.
- **Authentification** : procédé visant à vérifier l'identification d'une personne physique par des moyens techniques, tels que mot ou phrase de passe, un code secret, une réponse à un défi ou encore une sécurisation numérique (Certificat).
- **Autorité d'archivage** : entité responsable de la gestion du service d'archive et du système d'archivage.
- **Certificat** : document sous forme électronique attestant du lien entre l'identité du titulaire et les données de vérification de signature électronique.
- **Communication** : fait de porter l'Archive ou toute information relative à l'Archive à la connaissance d'une personne déterminée ou d'un groupe d'intéressés ou des Usagers.
- **Conservation** : opération(s) juridique(s) ou (et) matérielle(s) destinées à assurer la sauvegarde d'un droit, d'une chose, d'un patrimoine, etc.
- **Consultation** : interrogation du Système d'archivage électronique destinée à vérifier l'existence ou non d'un Objet d'archives.
- **Contenu d'information** : ensemble d'informations constituant l'objet principal de la pérennisation.
- **Elimination** (ou Destruction) : opération autorisée par un visa d'élimination consistant, après tri, à détruire l'Objet d'archive.
- **Empreinte (empreinte numérique ou condensat ou hash)** : Résultat d'une fonction de hachage appliquée sur une chaîne de caractères de longueur quelconque visant à réduire celle-ci en une donnée de longueur fixe représentative de cette chaîne de caractères. L'empreinte est l'un des éléments permettant de vérifier l'intégrité d'un document, d'un flux, d'un lot, d'une transmission,... (comparaison d'empreintes).
- **Information de pérennisation** : se décomposant en information de provenance, information d'identification, information d'intégrité et information de contexte, l'information de pérennisation accompagne le Contenu d'information afin qu'il puisse être correctement conservé.
- **Journaux d'évènement** : Enregistrement d'un ensemble de données relatives aux différentes opérations effectuées ou anomalies survenues au sein du SAE et destiné à assurer la traçabilité du

service. Par ailleurs ces journaux doivent être conservés pendant une période à définir et donc faire l'objet d'une procédure de sauvegarde particulière.

- **Métadonnées** : données encapsulées dans le Paquet d'Information avec l'Archive et/ou l'Objet d'archives.
- **Migration de formats** : opération qui consiste à migrer le contenu de certains types de formats vers d'autres types afin que le format de fichier utilisé pour la conservation des Archives reste adapté compte tenu de l'évolution des technologies.
- **Migration de supports** : opération qui consiste à migrer le contenu de certains types de supports vers d'autres types, notamment afin d'anticiper l'obsolescence du support concerné.
- **Module de sécurité** : système de confiance basé sur une ressource cryptographique éprouvée. Une ressource sera considérée comme éprouvée si elle a subi une évaluation selon des critères d'évaluation de la sécurité des systèmes d'information en vigueur, avec une cible de sécurité et un niveau d'assurance et de résistance suffisant.
- **Objet d'archives** : Données qui font l'objet de l'archivage (définition issue de du Standard d'échange)
- **Opérateur d'archivage** : entité qui fournit les services, liés au Service d'archivage, demandés et spécifiés par l'Autorité d'archivage au bénéfice de cette dernière, opérant dans un cadre hiérarchique, réglementaire ou contractuel.
- **Paquet d'Informations** : association du Contenu d'information et de son Information de pérennisation. A ce Paquet d'informations est associée une information d'emballage qui permet de relier et d'identifier les composants d'un Paquet d'informations.

On distingue trois types de paquets :

- o les Paquets d'information à verser : Paquet d'informations livrés par le Service producteur au Système d'archivage électronique pour l'élaboration d'un ou plusieurs Paquets d'informations archivés.
- o Les Paquets d'information archivés : Paquets d'informations conservé dans le Système d'archivage électronique et constitué d'un Contenu d'information et de l'Information de pérennisation associée.
- o Les Paquets d'informations diffusés : Paquets d'informations reçu par l'Utilisateur en réponse à sa requête au Système d'archivage électronique. Ce paquet provient d'un ou plusieurs Paquets d'informations archivés.
- **Politique d'archivage** : ensemble de règles portant un nom qui indiquent les exigences relatives à un archivage électronique sécurisé pour une communauté particulière et/ou une classe d'application avec des exigences de sécurité communes.
- **Politique de sécurité** : ensemble de règles portant un nom qui définit les exigences physiques, techniques et logiques afin de garantir un niveau de sécurité déterminé pour une communauté particulière et/ou une classe d'applications.
- **Titulaire** : le fournisseur du Système d'archivage électronique.
- **Service d'archives** : entité destinataire du Versement et assurant la gestion des Archives, des Paquets d'informations et des Objets d'archives pour le compte d'un Service producteur. Le Service d'archives et le Service producteur peuvent être assurés par une même personne juridique.
- **Service producteur** : entité qui a initialement reçu ou produit l'Archive et qui en est propriétaire. Le Service producteur et le Service d'archives peuvent être assurés par une même personne juridique.

- **Service versant** : entité qui verse un Paquet d'informations à un Service d'archives.
- **Signature électronique** : donnée sous forme électronique, jointe ou liée logiquement à d'autres données électroniques et qui sert de méthode d'authentification de l'origine des informations et garantit leur intégrité.
- **Support** : tout instrument permettant à l'Utilisateur de stocker des informations, de telle sorte que celles-ci puissent être consultées ultérieurement pendant une période adaptée à l'objectif de ces informations, et permettant la reproduction exacte des informations stockées.
- **Stockage** : opération consistant à garder des Archives sur un Support pendant une durée déterminée et dans un format pérenne.
- **Système d'archivage électronique** : système consistant à recevoir, conserver, traiter, restituer des Archives, des Paquets d'informations, des Objets d'archives, et qui s'appuie sur une plate-forme informatique.
- **Usager** : personne physique ou morale autorisée à consulter les Archives conservées sur le Système d'archivage électronique dans le respect de la législation applicable en matière de communication des Archives.
- **Utilisateur** : toute personne physique ou morale autorisée à utiliser un Système d'archivage électronique.
- **Versement** : transmission par un Service versant d'un Paquet d'informations à un Service d'archives.