



Modèle OAIS et logiciel VITAM

Journée archivage Grenoble

GRICAD – 20/11/2019

Olivier Rouchon



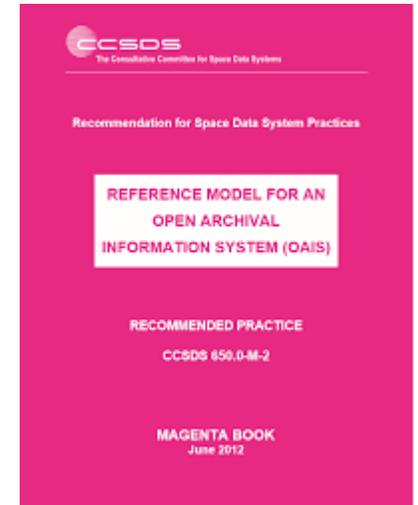
Le modèle

- Élaboré à l'initiative des agences spatiales (CNES, ESA, NASA)
- Avec le concours d'un certain nombre de représentants d'archives nationales et de grandes bibliothèques
 - Norme internationale ISO 14721 depuis Mars 2003
 - <http://public.ccsds.org/publications/archive/650x0b1.pdf>
 - Version française également normalisée à l'ISO :
 - <http://public.ccsds.org/publications/archive/650x0b1%28F%29.pdf>
 - Version révisée publiée en Septembre 2012, sans changements fondamentaux ;
 - Nouvelle version en cours de finalisation.



Les objectifs

- Proposer un cadre normatif général pour :
 - Définir les concepts indispensables à la compréhension de l'archivage numérique
 - Permettre la comparaison de l'organisation, de l'architecture et du mode de fonctionnement des archives existantes et futures
 - Constituer un guide pour l'identification et la production de normes dans ce domaine

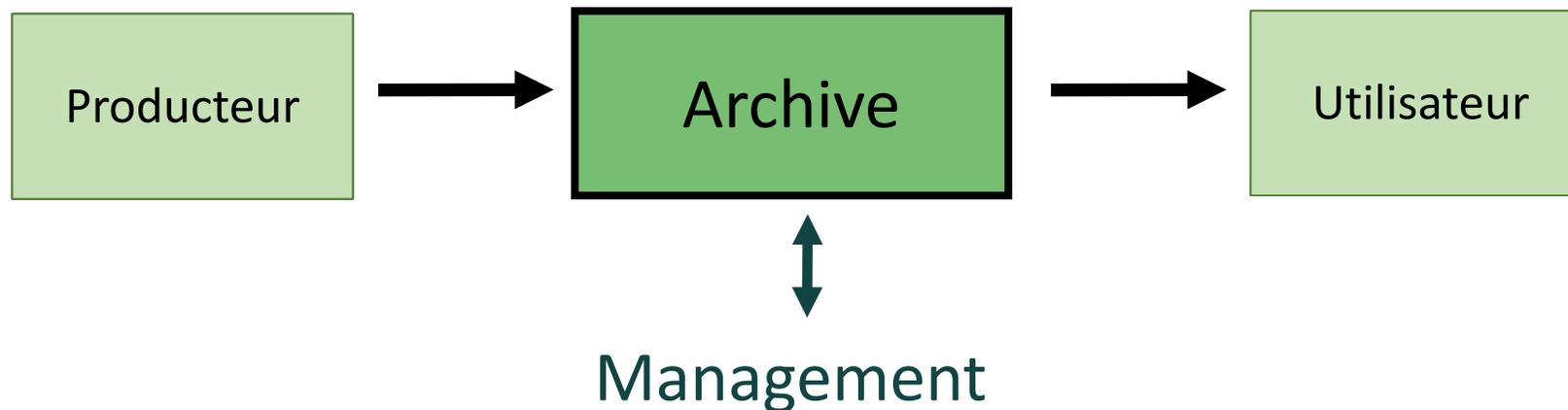


Le périmètre

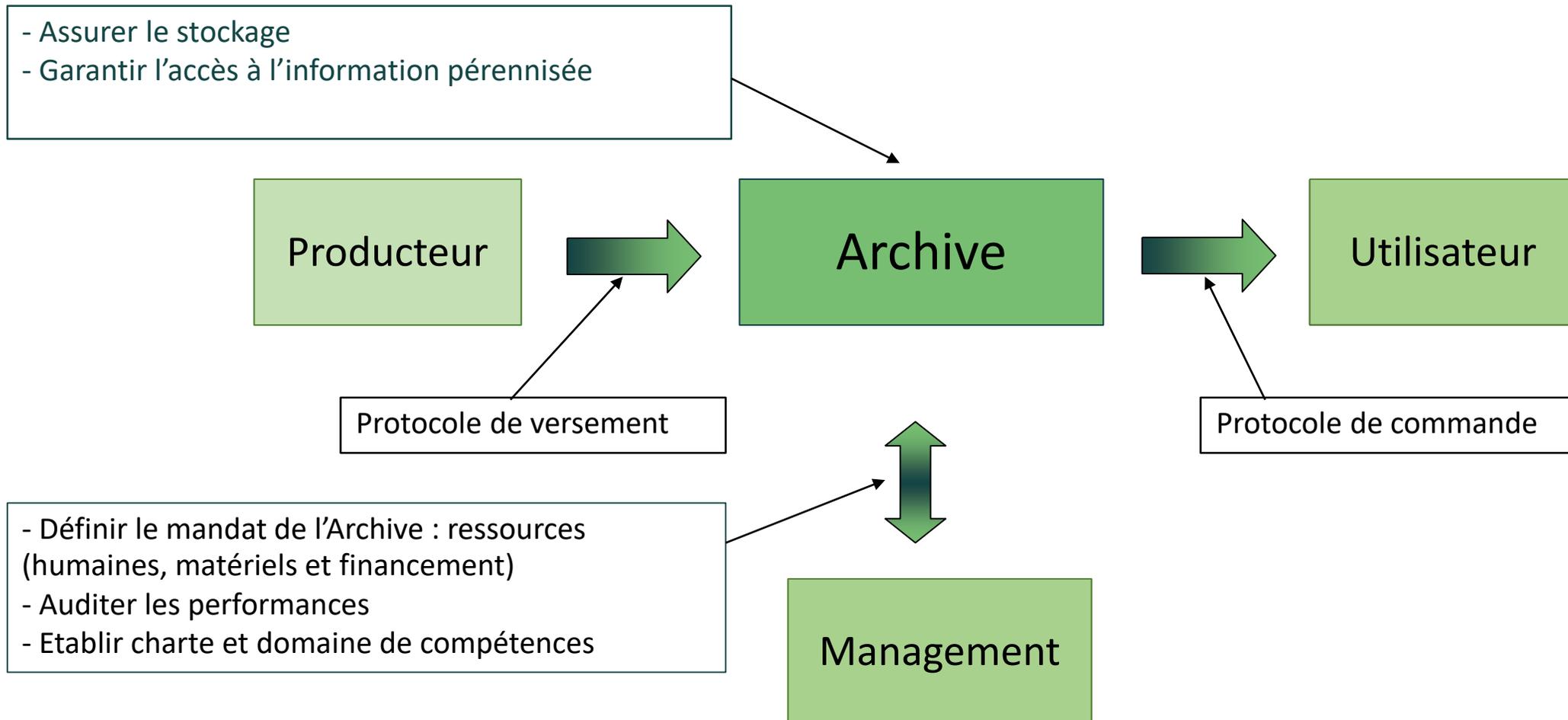
- Ce que la norme définit :
 - Un vocabulaire
 - Un modèle d'information définissant une typologie des catégories d'information
 - Les entités fonctionnelles concernées par la problématique d'archivage
 - Les interactions entre ces entités
 - Des analyse complémentaires relatives à la problématique de pérennisation : migrations, coopérations entre archives...
- Ce qu'elle NE définit PAS
 - Méthode pour la mise en œuvre
 - Modèle non applicable directement
 - Directives de méthodes de préservations (formats de données, de supports)

L'environnement de l'archive

- Le **producteur** crée des objets numériques qui sont transférés à l'archive
- L'**utilisateur** recherche, sélectionne, commande et récupère les objets numériques qui l'intéressent
- Le **management** : décideurs de la politique globale de l'archive en cohérence avec la politique globale de l'organisme ou de l'entreprise.



L'environnement de l'archive



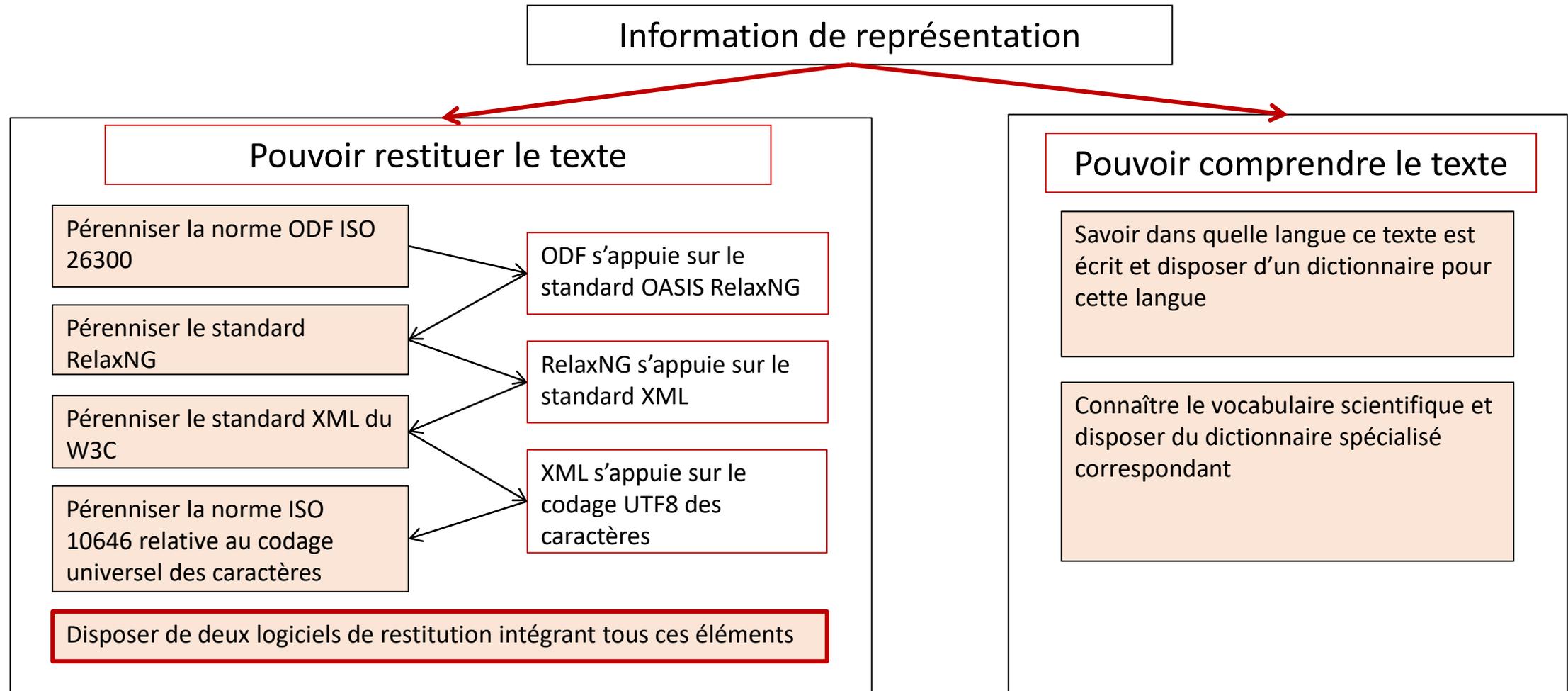
L'information de représentation



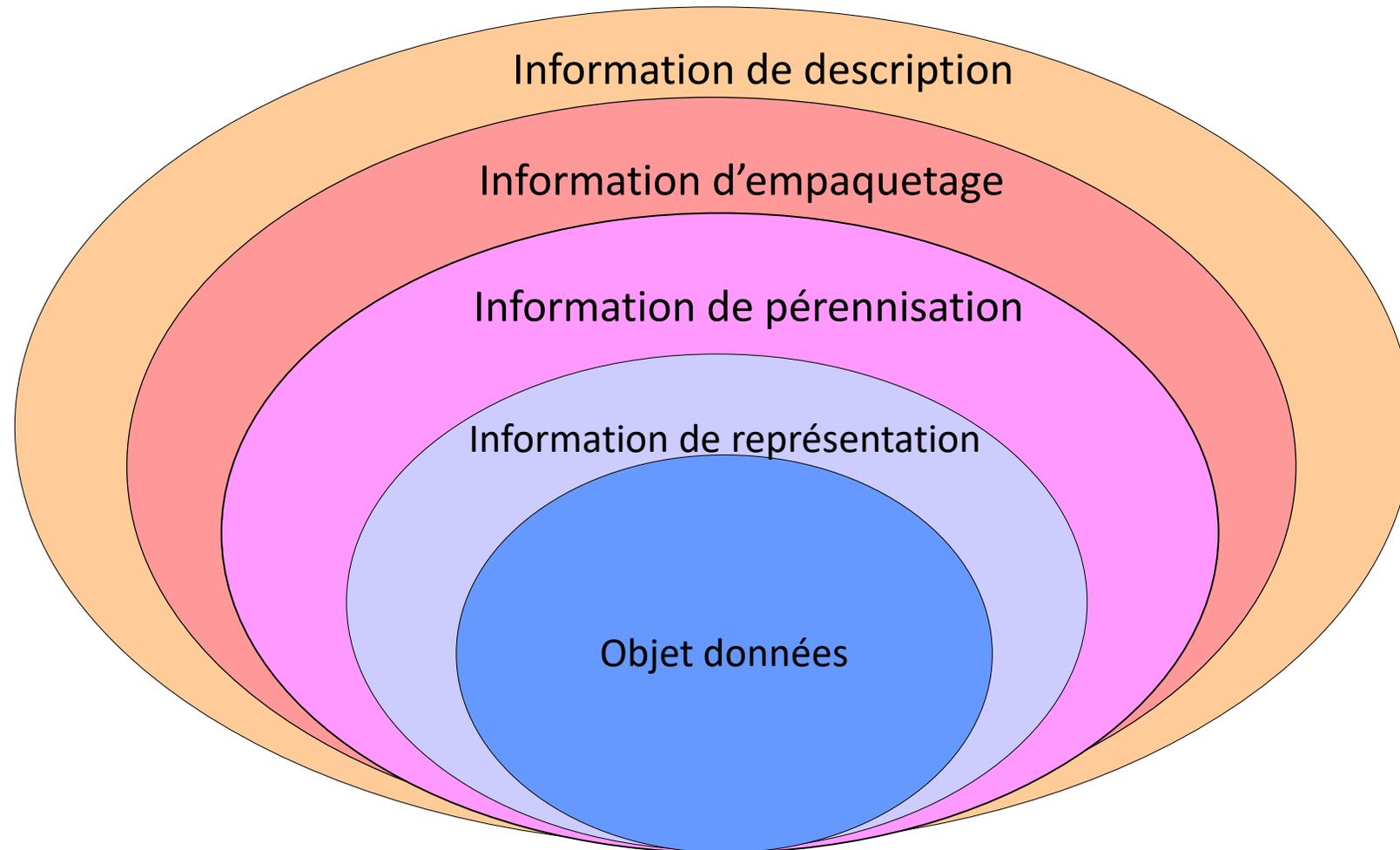
Pour que la pérennité d'un **document numérique** soit garantie, il est indispensable que l'archive préserve ce **document numérique** et **l'Information de représentation** qui lui est associée

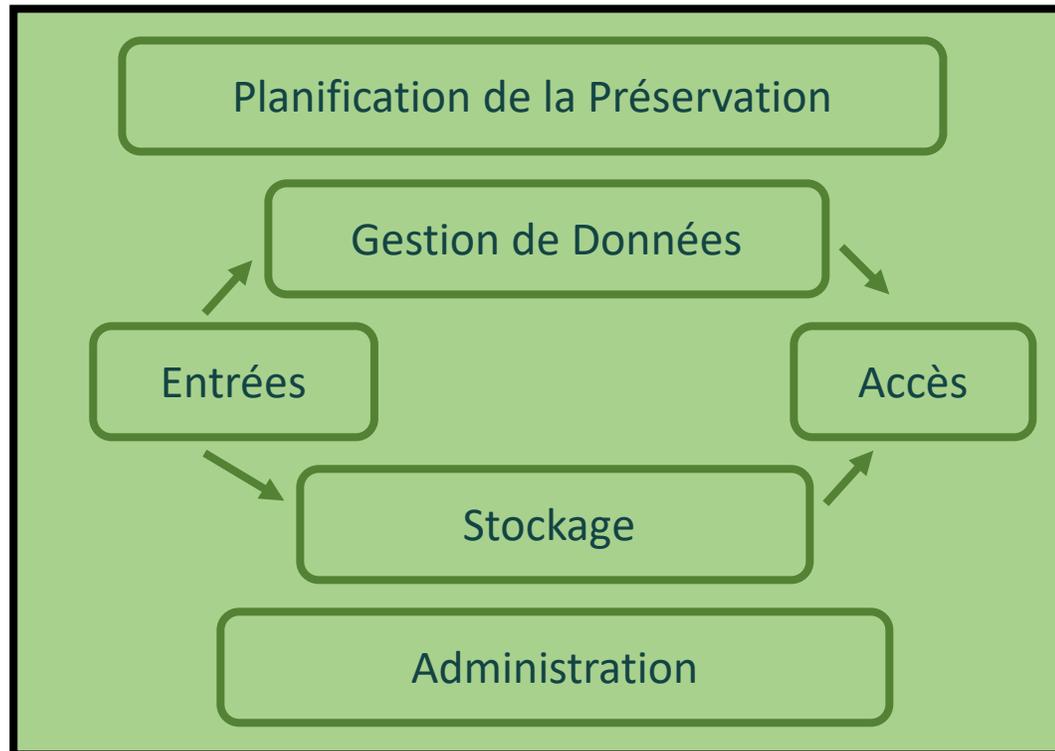
- Concept de base du Modèle OAIS.
- Différents types d'IR
 - Information de structure
 - Information sémantique.
- Extrême précision de la description de l'IR
 - Bases de connaissance lié à l'Objet données.
 - Nécessité de conserver des dictionnaires du langage utilisé
- L'IR est un objet d'information
 - Récursivité

L'information de représentation



Le modèle d'information





- L'entité **Entrées** insère dans l'OAIS les objets numériques provenant du producteur
- L'entité **Accès** transmet les objets numériques aux utilisateurs
- L'entité **Gestion de Données** gère le contenu des fonds de l'OAIS
- L'entité **Stockage** préserve les AIP sur le long terme

Les paquets d'information

- Conteneurs utilisés dans le processus d'archivage
 - le Paquet d'information à verser (SIP)
 - fourni par les producteurs de données
 - possibilité d'avoir plusieurs SIP pour obtenir l'ensemble du Contenu d'Information et l'Information Complémentaire à la Pérennisation (PDI)
 - relation SIP / AIP non univoque
 - le Paquet d'information archivé (AIP)
 - paquet de pérennisation
 - respecte rigoureusement des concepts de l'OAIS,
 - le Paquet d'information diffusé (DIP)
 - fourni aux consommateurs de données.





SEDA

Standard d'Échange de Données pour l'Archivage



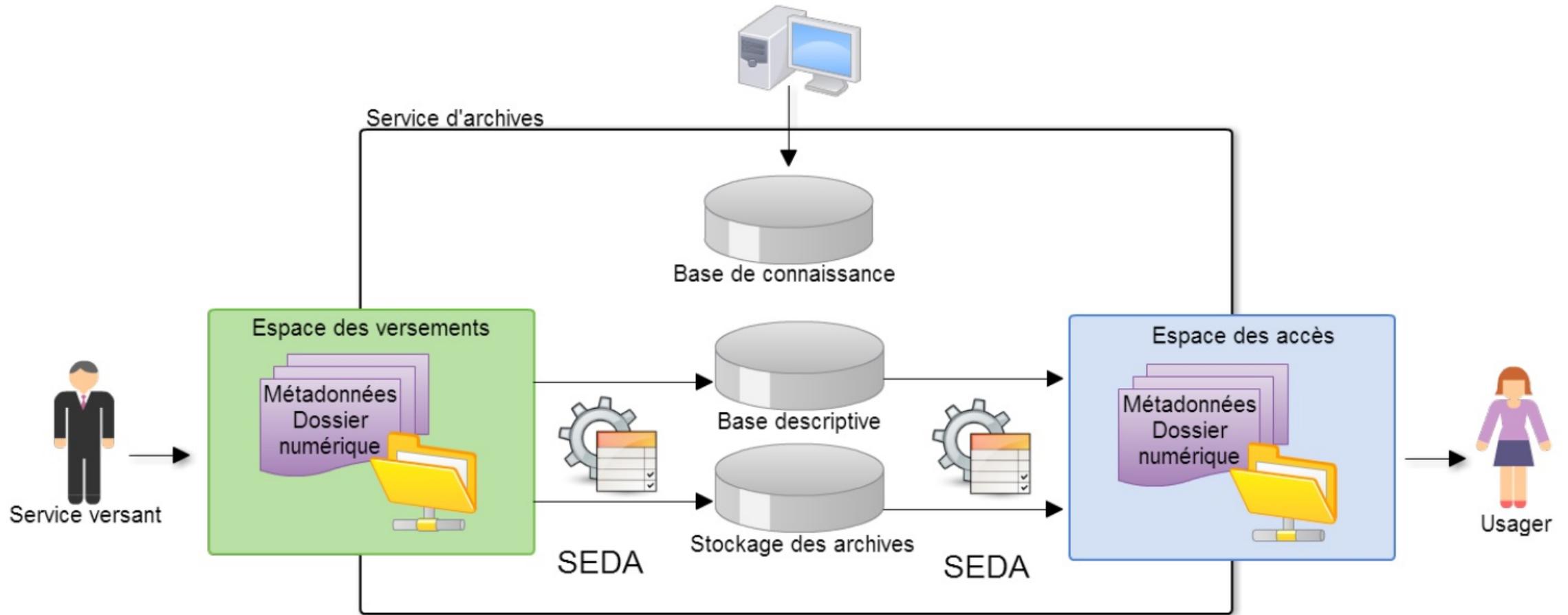
Le périmètre

- Modélisation des transactions entre acteurs dans le cadre de l'archivage de données
 - Transfert / Communication / Modification / Élimination / Restitution
 - Service d'archives / Service versant / Service producteur / Service de contrôle / Opérateur de versement / Demandeur
- Structure et un vocabulaire pour représenter ces échanges
 - Schémas XML pour chaque type de messages
 - Métadonnées pour décrire les archives (inspirées de l'EAD)
- Automatisation des procédures d'échanges et des contrôles
 - Profils d'archivage pour définir les règles qui régissent le versement, le contenu des messages et la prise en charge d'un paquet d'information
 - Grammaire XML pour contrôler automatiquement le bon respect de ces règles (est égal à, ne peut pas être différent de, n'existe qu'en un seul exemplaire, provient de, etc...)

L'historique

- Collaboration DGME / DAF dans le cadre de l'action 103 du programme ADELE
- Appel à commentaires jusqu'au 23 décembre 2005, suivi d'ateliers (décembre et janvier 2006)
- Les versions
 - Mars 2006, publication de la version 0.1 suivie d'une instruction DITN/RES/2006/001 au 8 mars 2006
 - Janvier 2010 publication de la version 0.2
 - Septembre 2012 publication de la version 1.0
 - Décembre 2015: SEDA 2.0 (spécification de MEDONA)
 - Juin 2018: SEDA 2.1

La place dans l'archive

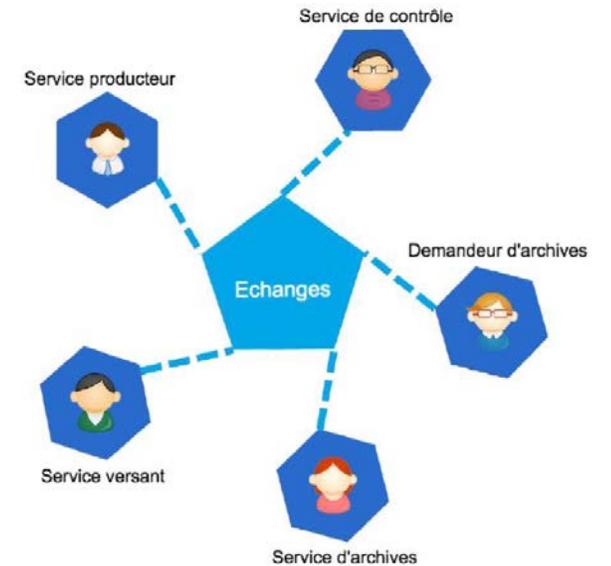


Les inspirations normatives

- Norme conceptuelle ISO 14721 (modèle OAIS)
- Norme AFNOR Z42-013
- DTD EAD (Encoded Archival Description)
- Méthodologie de l'UN/CEFACT (Centre for Trade Facilitation and Electronic Business) organisme des Nations Unies qui promeut le développement et la simplification des échanges électroniques professionnels, du commerce électronique et des procédures administratives
- Norme AFNOR Z44-0252 ou MEDONA (Modélisation des Echanges de DONnées pour l'Archivage)
- Norme ISO 20614 ou Depip (Data Exchange Protocol for Interoperability and Preservation)
- Langage XML, recommandation du W3C

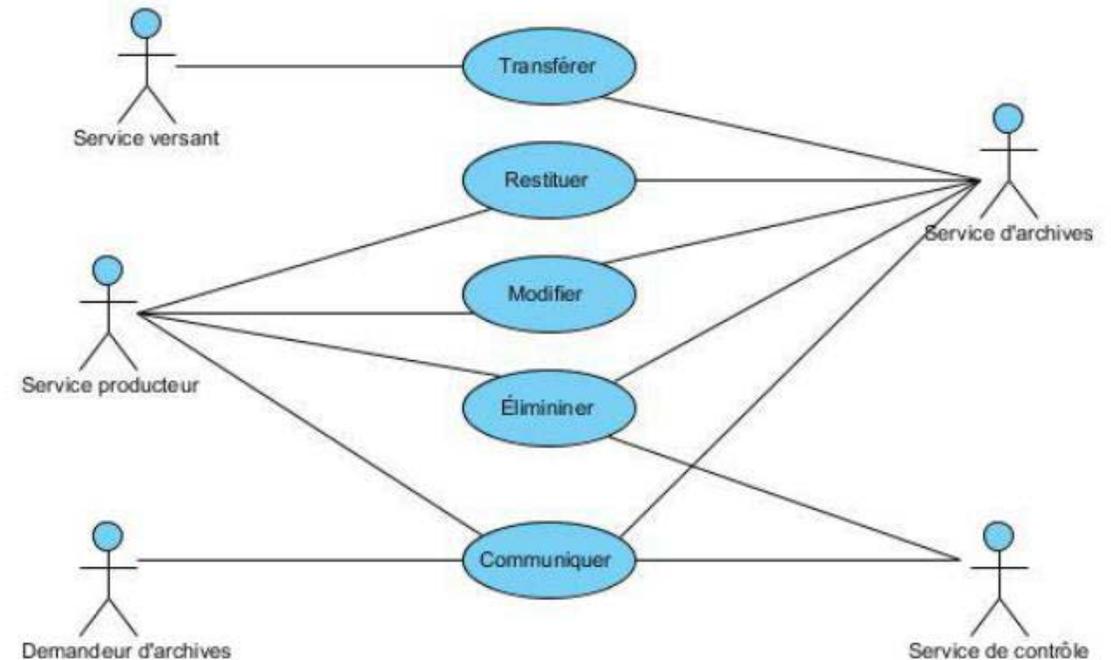
Les acteurs

1. Service d'archives (Archival Agency) : entité destinataire du transfert et assurant la gestion des informations transférées par les Services versants et destinées à être communiquées aux Demandeurs dans le respect des conditions légales, réglementaires ou contractuelles ;
2. Service versant (Transferring Agency) : entité qui transfère des archives dont il est responsable à un Service d'archives, qu'il les ait produites ou qu'il les ait héritées d'un Service producteur dont il a repris les attributions. ;
3. Service producteur (Originating Agency) : entité qui a produit les informations, c'est-à-dire qui les a créées ou reçues dans le cadre de son activité ;
4. Service de contrôle (Control Authority) : entité qui, le cas échéant, autorise ou non la Communication et l'Élimination ;
5. Opérateur de versement : entité chargée des opérations techniques de transfert des archives demandées par un Service versant, mais qui n'est ni le Service producteur ni le Service versant (par exemple : une DSI ou un tiers-archiveur) ;
6. Demandeur (Requester) : toute personne physique ou morale qui souhaite consulter les informations conservées par le Service d'archives dans le respect des conditions légales, réglementaires ou contractuelles en vigueur.



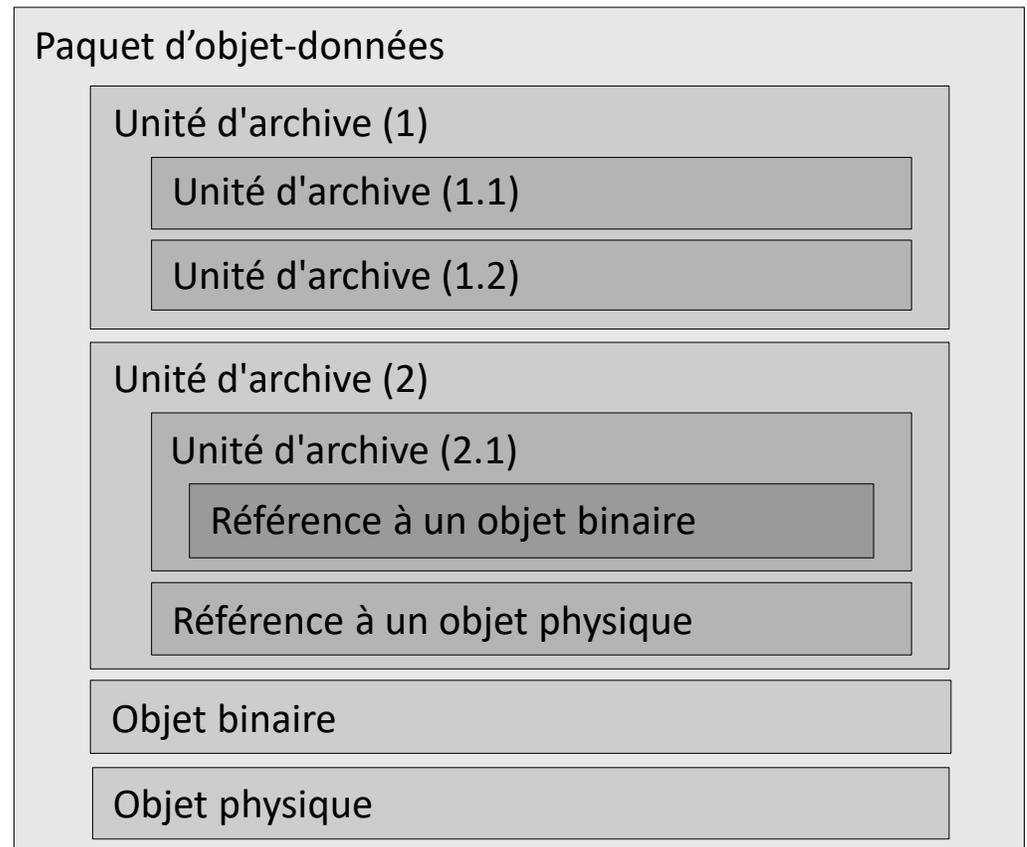
Les transactions

- Transfert d'archives (Archive Transfer)
- Communication d'archives (Archive Delivery)
- Modification d'archives (Archive Modification Notification)
- Élimination d'archives (Archive Destruction Notification)
- Restitution d'archives (Archive Restitution)



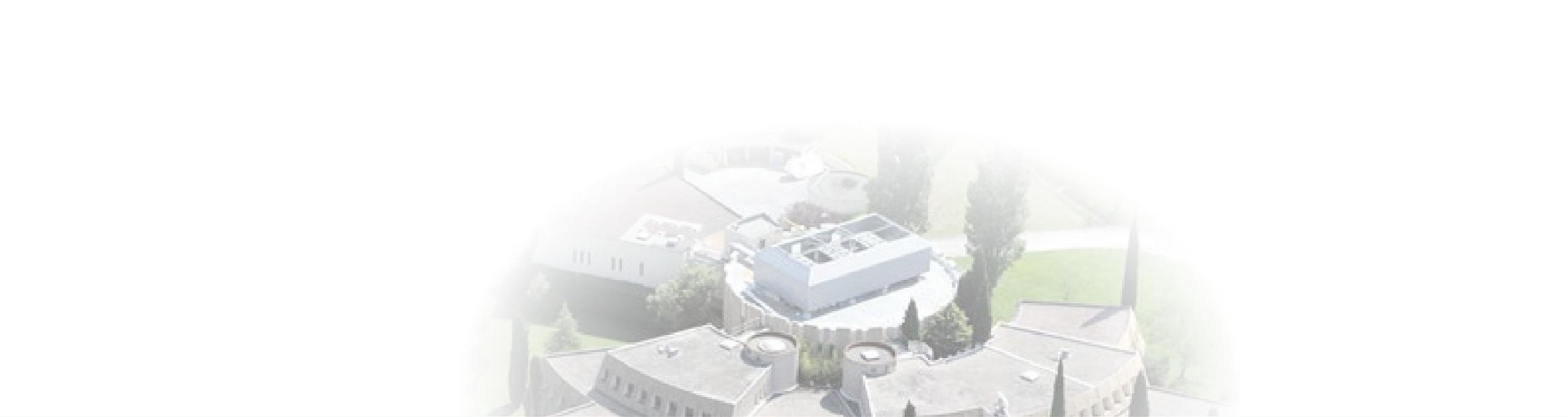
Les objets échangés

- Paquet d'objet-données
 - Ensemble des données transférées
 - Composé de métadonnées (informations de représentation/pérennisation) et de documents
 - Représente en quelque sorte un versement.
- Unité d'archive
 - Subdivision intellectuelle du paquet d'objet-données qui possède des caractéristiques propres de conservation.
 - Peut elle-même être subdivisée en sous-unités.
 - Peut être comparée à un niveau de description d'un instrument de recherche.
- Objet binaire ou physique
 - Transport des contenus de données (fichiers ou documents physiques) associés à leurs métadonnées.
 - Représente la pièce.



La normalisation

- Début 2013, démarche auprès de l'AFNOR pour une normalisation du SEDA
 - Besoin de stabilité, de généricité et d'indépendance exprimé par :
 - Les utilisateurs actuels du SEDA (éditeurs, tiers archiveurs, services d'archives)
 - D'autres communautés (banques, éditeurs de coffre-forts numériques)
 - Les tutelles : le délégué interministériel aux archives de France (DIAC), la Direction interministérielle des systèmes d'information et de communication (DINSIC)
- La normalisation
 - Dans le cadre de la commission de normalisation 4 « interopérabilité technique »
 - Enquête publique du 21 juin au 28 août 2013
 - Les commentaires étaient tous favorables mais nombreux (environ 200)
 - Janvier 2014, sortie de la norme AFNOR Z44-022 (« Modélisation des Échanges de DONnées pour l'Archivage » - MEDONA)
 - Mars 2018, publication de la norme ISO 20614 (« Protocole d'échange de données pour l'interopérabilité et la préservation » - DEPIP)



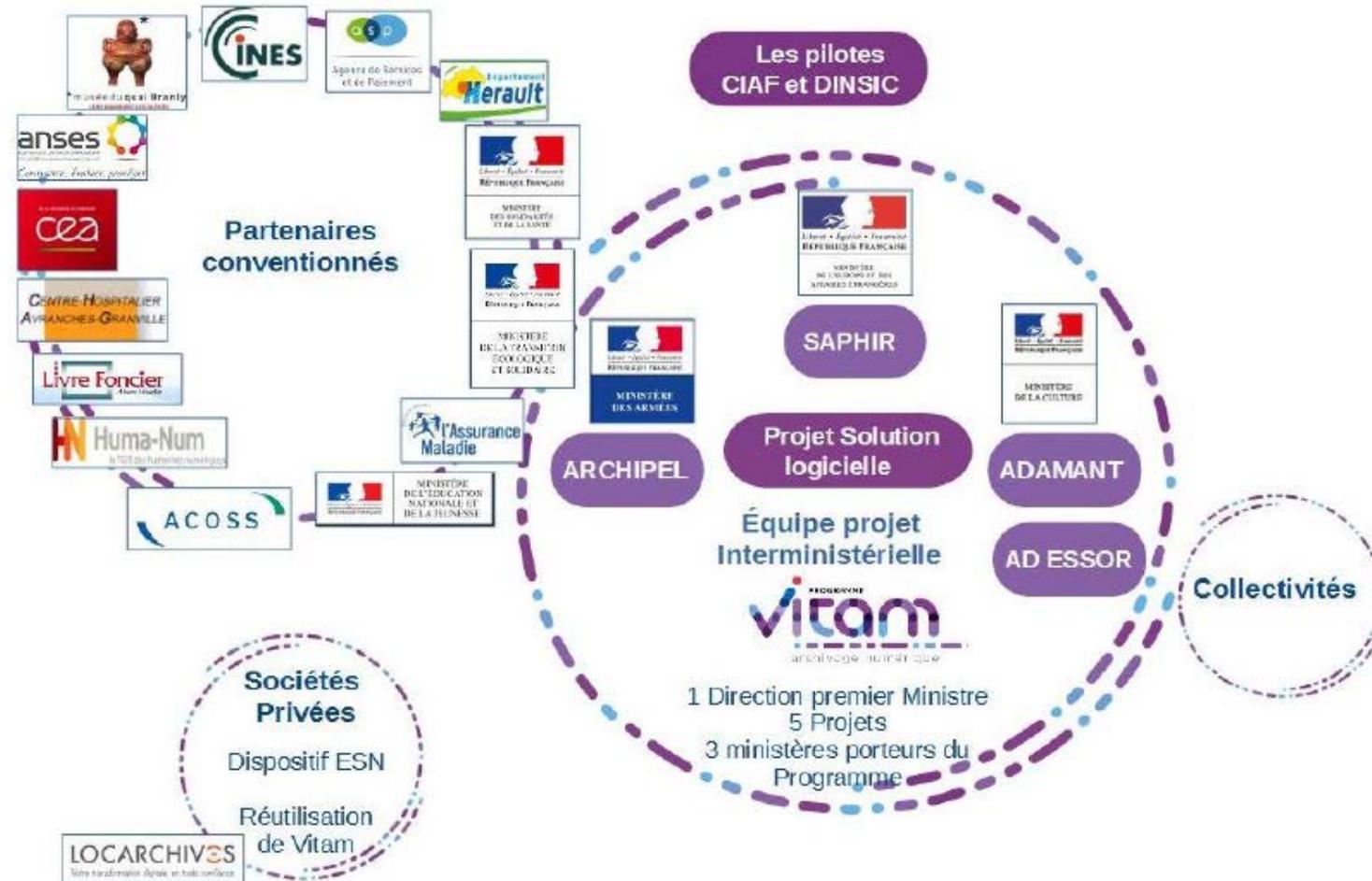
Le programme VITAM

Le contexte

- Programme interministériel soutenu et financé par les ministères
 - de l'Europe et des Affaires étrangères,
 - des Armées
 - et de la Culture.
- Les ministères porteurs du Programme sont représentés par leurs services informatiques et leurs services d'archives (opérateurs et services de tutelle).
- Trois objectifs :
 - Développer un logiciel libre d'archivage numérique capable de gérer, préserver et donner un accès à long terme aux archives numériques qui lui sont versées (back office).
 - Assurer que ce logiciel sera utilisé, a minima, par les 3 ministères porteurs du Programme afin de construire leur propre plate-forme d'archivage numérique.
 - Permettre une réutilisation la plus large possible du logiciel via les partenariats avec d'autres entités publiques.



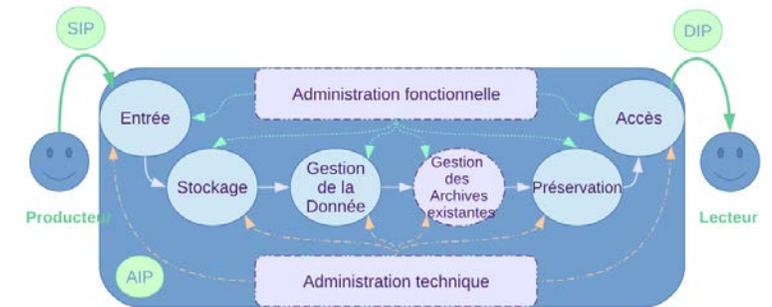
La démarche



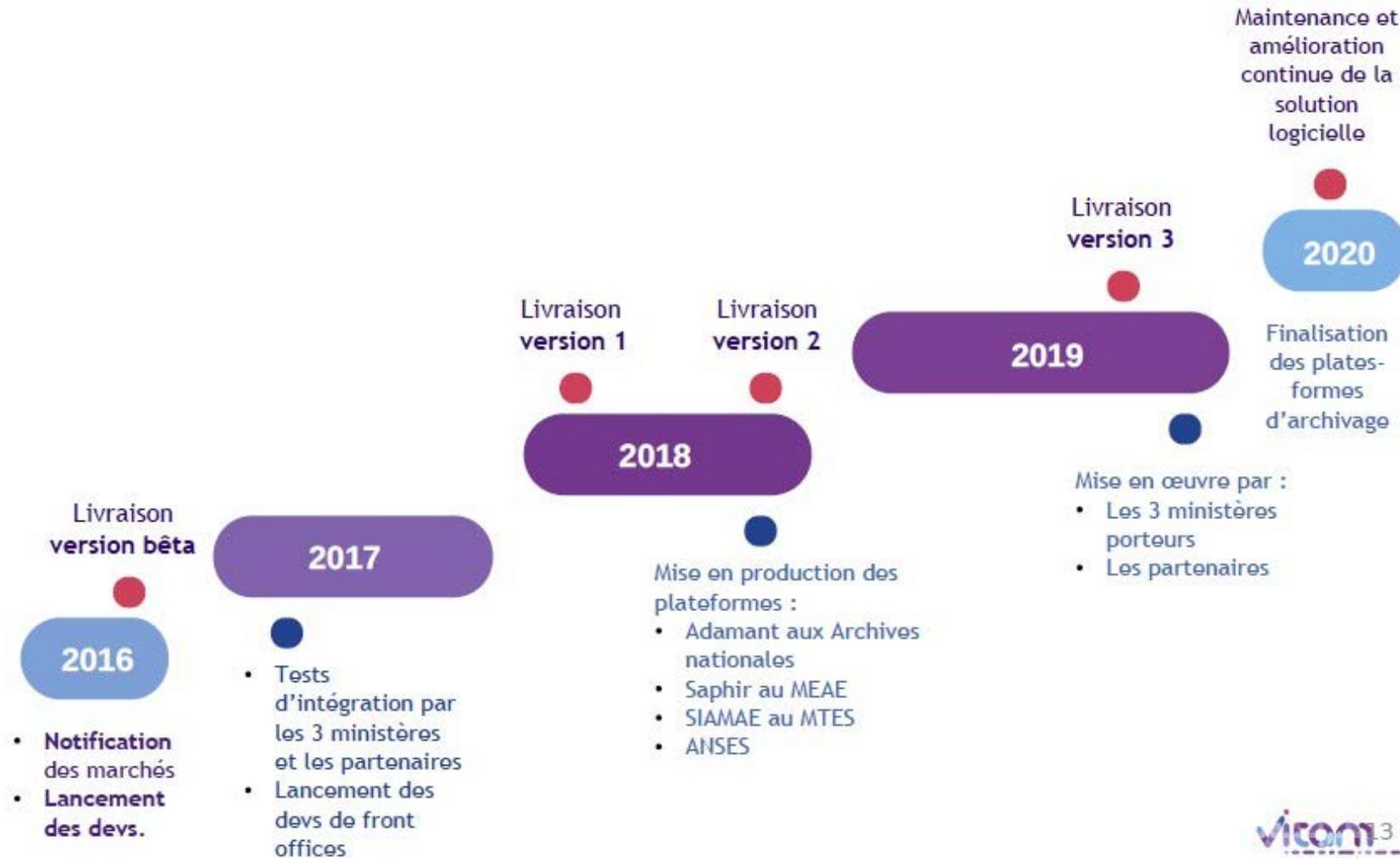
Les objectifs

Permettre :

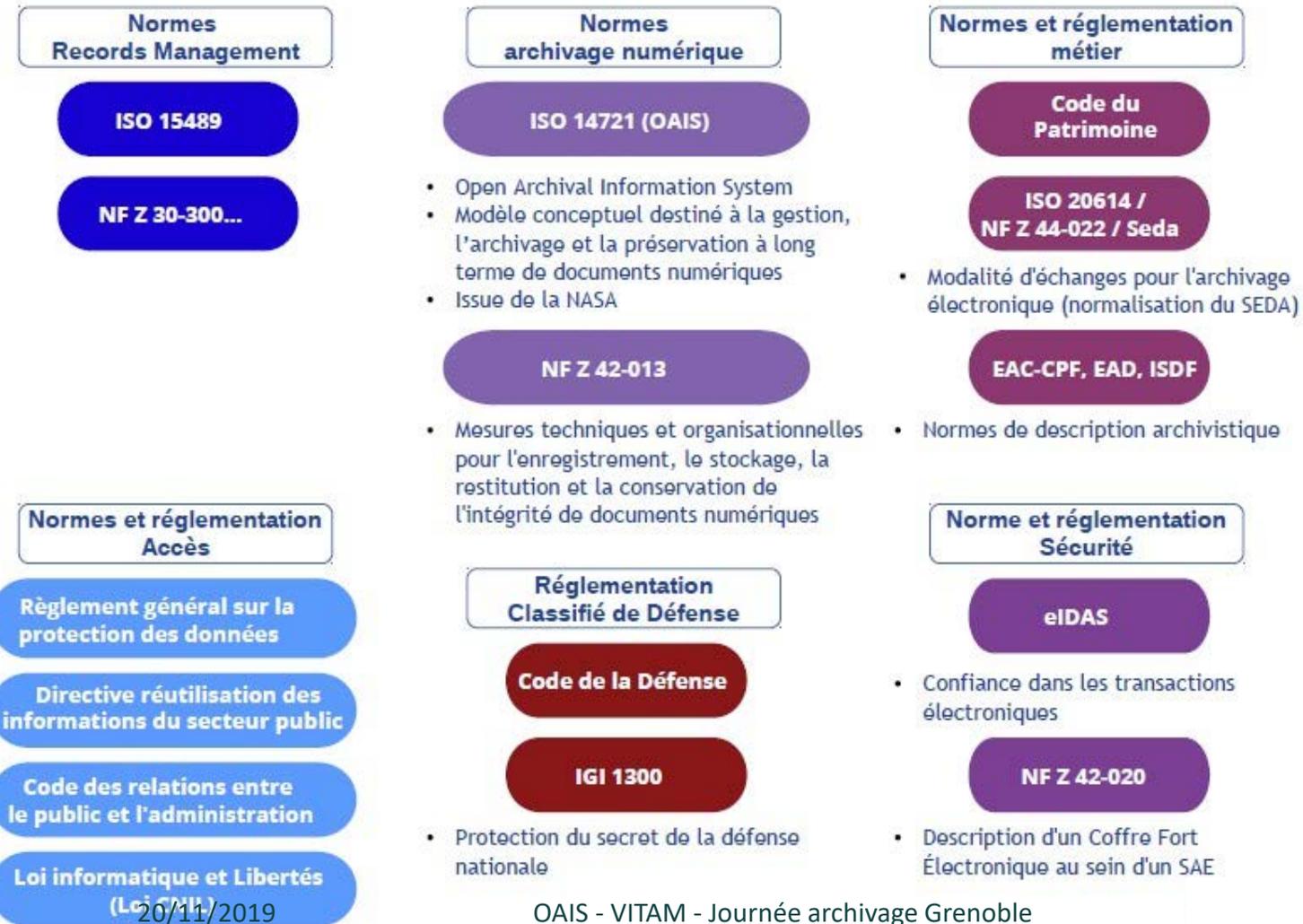
- d'archiver un très gros volume d'archives et de méta-données (objectif de plusieurs milliards d'objets)
 - Tant pour l'archivage courant, intermédiaire que définitif : fréquences de versement et d'accès variables
 - Y compris dans un contexte mutualisé (multi-tenant)
 - Quel que soit le type d'archives numériques : texte, audio, vidéo, image, plan, base de données, etc.
- d'assurer les fonctionnalités archivistes traditionnelles : entrées, indexation, gestion des métadonnées, récolement, recherche et consultation, gestion du cycle de vie
 - Et plus spécifiques au numérique : gestion des infrastructures de stockage, gestion de la valeur probante, pérennisation des fichiers (Modèle OAIS)



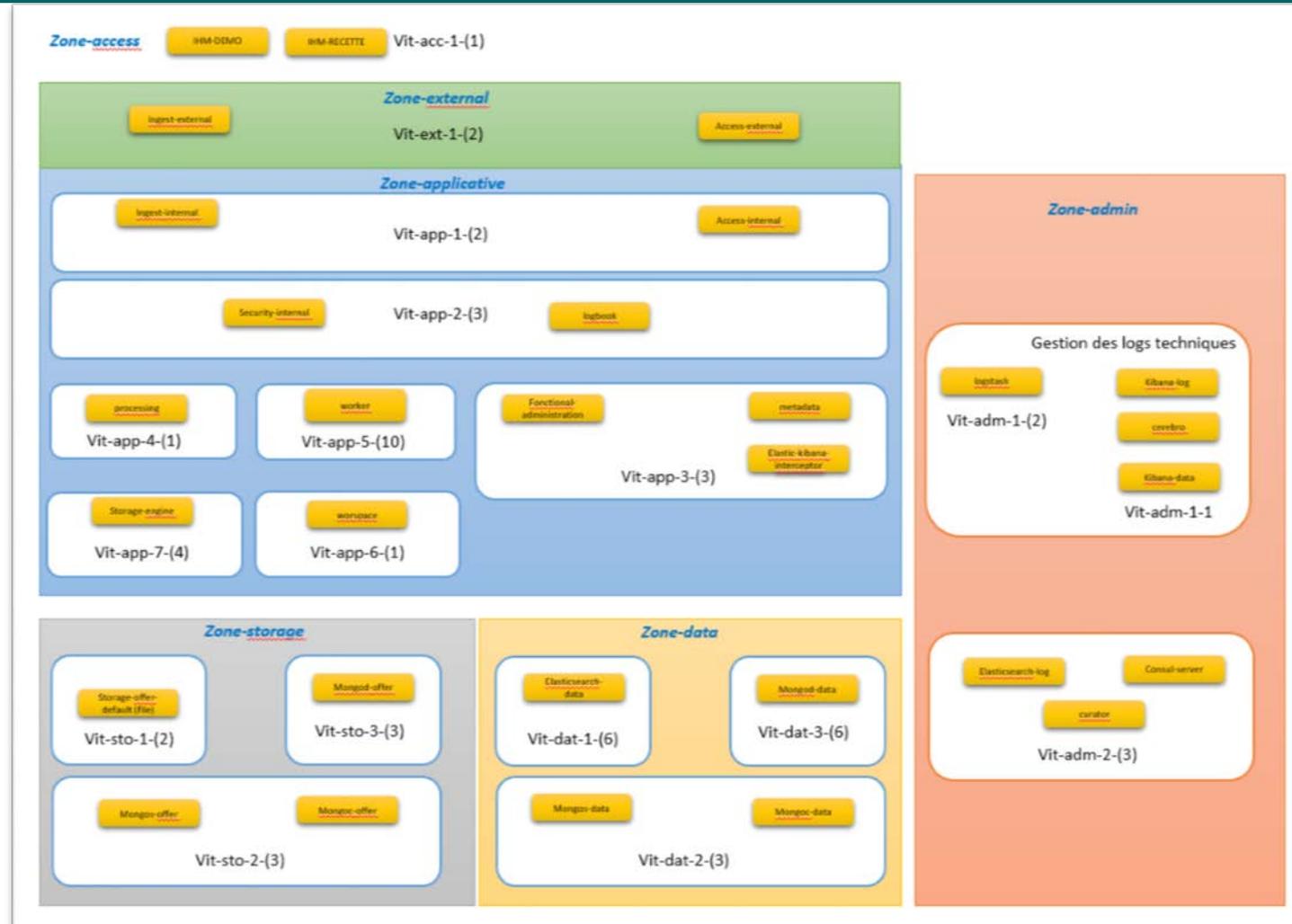
L'historique



Les aspects normatifs



L'architecture



Merci de votre attention

