

# ***Comité Consultatif pour les Systèmes de Données Spatiales***

**RECOMMANDATION POUR LES NORMES SUR  
LES SYSTEMES DE DONNEES SPATIALES**

## **Modèle de référence pour un Système ouvert d'archivage d'information (OAIS)**

**CCSDS 650.0B-1 (F)**

**Livre Bleu**

Mars 2005





## AUTORITE COMPETENTE

Edition	Livre bleu, Edition N°1
Date:	Mars 2005
Location:	Non Applicable

Cette publication a été approuvée par le Conseil de Management du Comité Consultatif pour les Systèmes de Données Spatiales (CCSDS). Le présent document constitue l'accord technique des agences membres qui participent au CCSDS. La procédure de revue et d'approbation des recommandations du CCSDS est détaillée dans le document '*Procedures Manual for the Consultative Committee for Space Data Systems*'. Les informations relatives à l'enregistrement de la participation des agences dans le processus d'approbation de ce document peuvent être demandées au secrétariat du CCSDS dont l'adresse est donnée ci-dessous.

Cette recommandation est publiée et mise à jour par :

CCSDS Secretariat  
Program Integration Division (Code M-3)  
National Aeronautics and Space Administration  
Washington, DC 20546, USA

## DECLARATION D'INTENTION

Le Comité Consultatif pour les Systèmes de Données Spatiales (CCSDS) est une organisation créée officiellement par les agences spatiales membres. Le Comité se réunit périodiquement afin de traiter des problèmes de systèmes de données qui sont communs à tous les participants, et de formuler des solutions techniques adaptées à ces problèmes. Attendu que la participation au sein du CCSDS n'est pas obligatoire, les décisions du Comité sont appelées **RECOMMANDATIONS** et n'engagent pas les agences.

La présente Recommandation est émise par l'Assemblée Plénière du CCSDS et en constitue l'accord général. L'adhésion d'une agence à cette Recommandation n'est pas obligatoire, mais elle suppose l'acceptation des conditions suivantes :

- Chaque fois qu'une agence crée une Norme qui est du ressort du CCSDS, cette Norme doit être conforme à la Recommandation applicable. La création d'une Norme n'exclut pas d'autres dispositions qu'une agence pourrait entreprendre par ailleurs.
- Chaque fois qu'une agence crée une Norme qui est du ressort du CCSDS, l'agence doit fournir aux autres agences membres du CCSDS les informations suivantes :
  - La norme elle-même.
  - La date prévue pour l'entrée en vigueur.
  - La durée prévue du service opérationnel.
- Les dispositions pour un service spécifique doivent être prises par l'intermédiaire de protocoles d'accord. Ni la présente Recommandation, ni aucune Norme en résultant, ne peut se substituer à un protocole d'accord.

Cinq ans après sa date d'émission au plus tard, cette recommandation sera révisée par le CCSDS afin de déterminer : (1) si elle peut rester en vigueur sans aucune modification, (2) si elle doit être modifiée afin de refléter l'influence de nouvelles technologies, de nouveaux besoins ou de nouvelles orientations ; ou (3) si elle doit être retirée ou annulée.

Dans ce contexte, lorsqu'une nouvelle version d'une Recommandation est émise, les standards et mises en oeuvre d'agences existants liés au CCSDS ne sont pas invalidés ou déclarés incompatibles avec les Recommandations du CCSDS. Il est de la responsabilité de chaque agence de déterminer à quel moment tel standard ou mise en oeuvre doit être modifié. Cependant, il est fortement recommandé à chaque agence d'adapter ses nouveaux standards et mises en oeuvre aux dernières versions de la Recommandation.

## AVANT-PROPOS

Ce document est une Recommandation technique destinée à favoriser un large consensus sur les exigences applicables à une Archive chargée de conserver définitivement ou sur le long terme de l'information sous forme numérique.

Cette Recommandation définit une base commune de termes et de concepts s'appliquant à un Système ouvert d'archivage d'information (OAIS - Open Archival Information System). Elle permet une véritable comparaison et une mise en évidence des différences entre les Archives aussi bien actuelles que futures. Elle fournit une base pour le développement des futurs standards dans le domaine de l'archivage. Elle devrait permettre une plus grande prise de conscience et prise en compte par les fournisseurs, des besoins spécifiques à l'archivage.

Dans le cours d'évolution de ce document, on pourra procéder à des extensions, des suppressions ou des modifications. Cette Recommandation est donc soumise aux procédures de contrôle et de modification des documents CCSDS qui sont définies dans '*Procedures Manual for the Consultative Committee for Space Data Systems.*' Les versions actuelles des documents CCSDS sont disponibles sur le site Web du CCSDS :

<http://www.ccsds.org/>

Toute question concernant le contenu ou le statut de ce document doit être adressée au Secrétariat du CCSDS dont l'adresse est donnée page i.

Au moment de la publication, les agences membres actifs et observateurs du CCSDS étaient les suivants :

Agences Membres

- Agenzia Spaziale Italiana (ASI)/Italy.
- British National Space Centre (BNSC)/United Kingdom.
- Canadian Space Agency (CSA)/Canada.
- Centre National d'Etudes Spatiales (CNES)/France.
- Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)/Germany.
- European Space Agency (ESA)/Europe.
- Instituto Nacional de Pesquisas Espaciais (INPE)/Brazil.
- National Aeronautics and Space Administration (NASA)/USA.
- National Space Development Agency of Japan (NASDA)/Japan.
- Russian Space Agency (RSA)/Russian Federation.

Agences Observateur

- Austrian Space Agency (ASA)/Austria.
- Central Research Institute of Machine Building (TsNIIMash)/Russian Federation.
- Centro Tecnico Aeroespacial (CTA)/Brazil.
- Chinese Academy of Space Technology (CAST)/China.
- Commonwealth Scientific and Industrial Research Organization (CSIRO)/Australia.
- Communications Research Centre (CRC)/Canada.
- Communications Research Laboratory (CRL)/Japan.
- Danish Space Research Institute (DSRI)/Denmark.
- European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/Europe.
- European Telecommunications Satellite Organization (EUTELSAT)/Europe.
- Federal Service of Scientific, Technical & Cultural Affairs (FSST&CA)/Belgium.
- Hellenic National Space Committee (HNSC)/Greece.
- Indian Space Research Organization (ISRO)/India.
- Institute of Space and Astronautical Science (ISAS)/Japan.
- Institute of Space Research (IKI)/Russian Federation.
- KFKI Research Institute for Particle & Nuclear Physics (KFKI)/Hungary.
- MIKOMTEK: CSIR (CSIR)/Republic of South Africa.
- Korea Aerospace Research Institute (KARI)/Korea.
- Ministry of Communications (MOC)/Israel.
- National Oceanic & Atmospheric Administration (NOAA)/USA.
- National Space Program Office (NSPO)/Taipei.
- Swedish Space Corporation (SSC)/Sweden.
- United States Geological Survey (USGS)/USA.

## GESTION DE CONFIGURATION DU DOCUMENT

<b>Document</b>	<b>Titre</b>	<b>Date</b>	<b>Statut/Observations</b>
CCSDS 650.0-B-1	Reference Model for an Open Archival Information System (OAIS)	Janvier 2002	Edition originale en langue anglaise
CCSDS 650.0B-1 (F)	Modèle de référence pour un Système Ouvert d'Archivage (OAIS)	Mars 2005	Traduction en français de l'édition originale.  Les annexes A et F de la version originale ne font pas partie de la Recommandation. Elles n'ont pas été insérées dans la version française.  Un Index des mots clé a été inséré en annexe F.

## TABLE DES MATIERES

<u>Section</u>	<u>Page</u>
<b>1 INTRODUCTION.....</b>	<b>1-1</b>
1.1 OBJET ET DOMAINE D'APPLICATION.....	1-1
1.2 APPLICABILITE.....	1-2
1.3 JUSTIFICATIONS .....	1-3
1.4 CONFORMITÉ.....	1-3
1.5 GUIDE POUR LE DEVELOPPEMENT DE NORMES ASSOCIEES .....	1-4
1.6 STRUCTURE DU DOCUMENT.....	1-4
1.7 DEFINITIONS .....	1-6
<b>2 CONCEPTS OAIS.....</b>	<b>2-1</b>
2.1 ENVIRONNEMENT OAIS.....	2-2
2.2 INFORMATIONS DE L'OAIS .....	2-3
2.3 INTERACTIONS EXTERNES DE HAUT NIVEAU D'UN OAIS .....	2-8
<b>3 RESPONSABILITES D'UN OAIS .....</b>	<b>3-1</b>
3.1 RESPONSABILITES OBLIGATOIRES.....	3-1
3.2 EXEMPLES DE MECANISMES POUR ASSUMER LES RESPONSABILITES.....	3-1
<b>4 MODELES DETAILLES.....</b>	<b>4-1</b>
4.1 MODÈLE FONCTIONNEL .....	4-1
4.2 MODÈLE D'INFORMATION.....	4-19
4.3 TRANSFORMATIONS DE PAQUETS D'INFORMATIONS .....	4-50
<b>5 PERSPECTIVES DE LA PERENNISATION.....</b>	<b>5-1</b>
5.1 PERENNISATION DE L'INFORMATION .....	5-1
5.2 PRESERVATION DES SERVICES D'ACCÈS.....	5-10
<b>6 INTEROPERABILITE DES ARCHIVES .....</b>	<b>6-1</b>
6.1 NIVEAUX TECHNIQUES D'INTERACTION ENTRE ARCHIVES OAIS .....	6-2
6.2 LE MANAGEMENT DES GROUPEMENTS D'ARCHIVES.....	6-8



**TABLE DES MATIERES (SUITE)**

<u>Section</u>	<u>Page</u>
<b>ANNEXE A : RELATIONS AVEC D'AUTRES NORMES OU EFFORTS .....</b>	<b>A-1</b>
<b>ANNEXE B : INTRODUCTION RAPIDE AU LANGAGE DE MODÉLISATION UNIFIÉ (UML).....</b>	<b>B-1</b>
<b>ANNEXE C : MODÈLE POUR L'UTILISATION DE LOGICIELS DANS L'INFORMATION DE REPRÉSENTATION .....</b>	<b>C-1</b>
<b>ANNEXE D : REFERENCES INFORMATIVES .....</b>	<b>D-1</b>
<b>ANNEXE E : INDEX.....</b>	<b>E-1</b>

<u>Schémas</u>	<u>Page</u>
2-1 Modèle d'environnement d'un OAIS .....	2-2
2-2 Obtention d'Information à partir de Données .....	2-4
2-3 Paquet d'informations : concepts et relations .....	2-6
2-4 Flux de données externes à une Archive OAIS .....	2-9
4-1 Entités fonctionnelles OAIS .....	4-1
4-2 Fonctions de l'Entité « Entrées » .....	4-6
4-3 Fonctions de l'Entité « Stockage ».....	4-8
4-4 Fonctions de l'Entité « Gestion de données ».....	4-10
4-5 Fonctions de l'Entité « Administration ».....	4-11
4-6 Fonctions de l'Entité « Planification de la pérennisation ».....	4-14
4-7 Fonctions de l'Entité « Accès » .....	4-16
4-8 Diagramme de flux de Données OAIS .....	4-18
4-9 Diagramme de contexte de l'Entité « Administration ».....	4-19
4-10 Objet-information .....	4-21
4-11 Objet-information de représentation .....	4-23
4-12 Taxinomie de l'Objet-information.....	4-25
4-13 Contenu d'un Paquet d'informations.....	4-33
4-14 Taxinomie du Paquet d'informations .....	4-33
4-15 Paquet d'informations archivé (AIP).....	4-35
4-16 Information de pérennisation (PDI) .....	4-36
4-17 Description de paquet .....	4-37
4-18 Paquet d'informations archivé (Vue détaillée) .....	4-39
4-19 Spécialisations de l'AIP .....	4-40
4-20 Spécialisation de la Description de paquet .....	4-41
4-21 Unité d'informations archivée (AIU).....	4-42
4-22 Description d'unité .....	4-42
4-23 Vue logique d'une Collection d'informations archivée .....	4-44

## TABLE DES MATIERES (SUITE)

4-24	Description de collection .....	4-46
4-25	Information de l'Entité « Gestion de données ».....	4-49
4-26	Vue de haut niveau sur les Flux de données dans un OAIS .....	4-51
5-1	Vue conceptuelle des relations entre noms et composants d'AIP .....	5-3
6-1	Archives coopérantes ayant un accord d'échange mutuel .....	6-4
6-2	Archives coopérantes avec des méthodes standard de Versement et d'Accès .....	6-4
6-3	Groupement d'OAIS utilisant un catalogue commun .....	6-5
6-4	Archives avec stockage partagé .....	6-8
B-1	Les éléments clés dans les diagrammes UML .....	B-1
C-1	Modèle d'information en couches .....	C-1

### Table

4-1	Exemples de types de PDI .....	4-31
-----	--------------------------------	------

# 1 INTRODUCTION

## 1.1 OBJET ET DOMAINE D'APPLICATION

L'objet du présent document est de définir le Modèle de référence ISO (Organisation Internationale de Normalisation) pour un **Système ouvert d'archivage d'information** (OAIS). Un OAIS est une Archive constituée de personnes et de systèmes, dont la responsabilité est de conserver des informations et de les rendre accessibles à une **Communauté d'utilisateurs cible**. L'OAIS exerce un ensemble de responsabilités qui sont définies dans le présent document, et qui le différencient d'autres utilisations du terme « archive ». Le terme « ouvert » de l'OAIS signifie que la présente recommandation ainsi que les futures recommandations et normes associées seront développées au sein de forums ouverts, mais ne sous-entend pas que l'accès à l'Archive s'opère sans restrictions.

On estime que l'information traitée doit être pérennisée, même si l'OAIS n'a pas une durée illimitée. Le **Long terme** est défini comme étant suffisamment long pour être soumis à l'impact des changements technologiques, y compris à la prise en compte de nouveaux supports et nouveaux formats de données ou à des changements de la communauté d'utilisateurs. Le Long terme peut se poursuivre indéfiniment. Dans le présent Modèle de référence, l'accent est mis sur l'information numérique, d'une part comme forme principale de l'information pérennisée, d'autre part comme information complémentaire de ce qui est archivé aussi bien sous forme numérique que sous forme physique. Par conséquent, le modèle s'adapte aux informations qui sont par nature non numériques (par exemple, un échantillon physique), mais la modélisation et la conservation de ce type d'information ne sont pas traitées en détail. Le présent Modèle de référence :

- fournit un cadre pour une meilleure compréhension et une prise de conscience croissante des concepts d'archivage requis pour la Pérennisation des informations numériques, et pour l'accès à ces informations,
- fournit les concepts nécessaires aux organisations non spécialisées dans l'archivage afin de leur permettre de participer efficacement au processus de conservation,
- fournit un cadre, y compris terminologique et conceptuel, pour la description et la comparaison des architectures et du fonctionnement des Archives existantes ou à venir,
- fournit un cadre pour la description et la comparaison des différentes stratégies et techniques de Pérennisation,
- fournit une base pour la comparaison de modèles de données d'informations numériques conservées par des Archives, et pour l'analyse de la manière dont les modèles de données et l'information sous-jacente peuvent évoluer dans le temps,
- fournit une base qui pourra être étendue à d'autres travaux afin de couvrir la Pérennisation de l'information qui N'EST PAS sous forme numérique (par exemple, des supports ou des échantillons physiques),

- élargit le consensus sur les éléments et les processus de conservation et d'accès à l'information numérique sur le long terme, et permet l'extension d'un marché que les fournisseurs peuvent couvrir,
- est un guide pour l'identification et la production des normes reliées à l'OAIS.

Le Modèle de référence aborde toute une série de fonctions liées à la conservation d'informations archivées comprenant le versement des données, le stockage, la gestion des données, l'accès et la diffusion. Il aborde également la migration d'informations numériques vers des supports et des formats nouveaux, les modèles de données utilisés pour représenter l'information, le rôle du logiciel dans la conservation de l'information, ainsi que l'échange d'informations numériques entre Archives. Il identifie les interfaces internes et externes des fonctions d'archivage ainsi qu'un certain nombre de services de haut niveau relatifs à ces interfaces. Le présent Modèle de référence présente divers exemples concrets, ainsi que quelques recommandations de « bonnes pratiques ». Il définit les responsabilités minimales auxquelles une Archive doit souscrire pour être qualifiée d'OAIS, ainsi qu'un profil maximal, en s'appuyant sur un large ensemble de termes et de concepts utiles.

## 1.2 APPLICABILITE

Le modèle d'OAIS exposé dans le présent document peut s'appliquer à toutes les Archives. Il s'applique plus particulièrement aux organisations chargées de rendre l'information disponible sur le Long terme. Ces organisations peuvent avoir d'autres responsabilités, telles que le traitement et la diffusion de données en réponse à des obligations programmatiques.

Ce modèle présente aussi un intérêt pour les organisations et les personnes qui produisent des informations dont la Pérennisation peut être nécessaire, et celles qui peuvent avoir besoin d'acquérir des informations auprès de telles Archives.

Le modèle, avec ses concepts de modélisation de l'information et de modélisation fonctionnelle, permet notamment de comparer et de concevoir des systèmes d'information non pérennes, et ceci pour deux raisons :

- Compte tenu de la rapidité des changements technologiques ou des changements possibles au sein d'une Communauté d'utilisateurs cible, il n'est pas exclu que des services censés conserver l'information à titre temporaire réalisent que tout ou partie de l'information qu'ils détiennent doit être pérennisée,
- Bien que certains services détenteurs d'information puissent être eux-mêmes temporaires, la conservation sur une durée indéfinie de tout ou partie des informations détenues peut être nécessaire. Ces services doivent être des acteurs dans l'effort de Pérennisation.

Les personnes chargées de l'élaboration des normes sont censées utiliser le présent modèle comme une base pour poursuivre la normalisation dans ce domaine. Un grand nombre de

normes associées peuvent être élaborées. Un guide pour le développement de ces normes est présenté en 1.5.

**Le présent Modèle de référence ne spécifie ni une conception ni une réalisation. Les fonctionnalités peuvent être regroupées ou décomposées de manière différente selon chaque réalisation.**

### 1.3 JUSTIFICATIONS

L'augmentation considérable de la puissance des ordinateurs, de la bande passante du réseau et de la connectivité, a donné lieu à une multiplication des organisations spécialisées dans la mise à disposition de l'information numérique. Les transactions entre les différents types d'organisations s'effectuent sous des formats numériques qui remplacent les supports traditionnels tels que le papier.

La conservation de l'information sous forme numérique est beaucoup plus complexe que la conservation de l'information sur supports papier ou film. Ceci n'est pas seulement un problème pour les Archives traditionnelles, mais également pour de nombreux organismes qui, jusque-là, n'avaient jamais eu conscience d'assurer une fonction d'archivage. En définissant un certain nombre d'exigences minimales pour une Archive OAIS, ainsi qu'un ensemble de concepts d'archivage, le présent Modèle de référence est censé offrir un cadre commun à partir duquel pourront être relevés les défis d'archivage, et en particulier ceux qui concernent l'information numérique. Il devrait permettre à un plus grand nombre d'organismes de comprendre les problèmes posés et de prendre les mesures adéquates pour garantir la Pérennisation de l'information. Il devrait également servir de base à la poursuite de la normalisation et, par conséquent, à l'élargissement du marché dès lors que les fournisseurs satisfont aux exigences en matière d'archivage.

### 1.4 CONFORMITE

La réalisation d'une Archive conforme à l'OAIS doit s'appuyer sur le modèle d'information décrit en 2.2. Le Modèle de référence OAIS ne définit ni n'exige aucune méthode spécifique de mise en œuvre de ces concepts.

Une Archive OAIS conforme doit assumer les responsabilités énumérées en 3.1. La sous-section 3.2 donne des exemples de mécanismes utilisables pour effectivement s'acquitter des responsabilités identifiées en 3.1. Ces mécanismes ne sont pas requis pour la mise en conformité.

On suppose que les personnes chargées d'une réalisation spécifique utiliseront le présent Modèle de référence comme guide, afin de fournir des services et un contenu clairement identifiés. Le présent document ne préjuge d'aucune plate-forme informatique, environnement système, paradigme de conception système, méthodologie de développement de système, système de gestion de base de données, paradigme de conception de base de

données, langage de définition des données, langage de commande, interface système, interface utilisateur, technologie ou supports, et n'en cautionne aucune.

Une Archive OAIS conforme peut fournir aux utilisateurs des services supplémentaires qui vont au-delà de ceux qu'exige un OAIS.

Le Modèle de référence OAIS est conçu comme un cadre conceptuel au sein duquel il est possible d'analyser et de comparer des Archives. En tant que tel, il tente de couvrir toutes les activités principales d'une Archive qui conserve de l'information afin de définir un ensemble utile et cohérent de termes et de concepts. Toute norme ou document qui prétend être conforme au Modèle de référence OAIS doit utiliser les termes et les concepts définis dans le Modèle de référence OAIS, selon les mêmes modalités.

## **1.5 GUIDE POUR LE DEVELOPPEMENT DE NORMES ASSOCIEES**

Le présent Modèle de référence, mis au point par le Panel 2 du CCSDS en réponse à l'ISO TC20/SC13, sert à identifier les domaines qui se prêtent au développement de normes associées à l'OAIS. Certaines de ces normes peuvent être développées par le Panel 2, d'autres peuvent être développées par d'autres organismes de normalisation. Toutefois, tout travail de ce type entrepris par d'autres organismes doit être coordonné afin de minimiser les incompatibilités et les efforts. Parmi les domaines pour des normes associées OAIS potentielles on trouve :

- norme(s) pour les interfaces entre Archives de type OAIS,
- norme(s) pour la méthodologie de versement utilisée par une Archive,
- norme(s) pour le versement de données numériques à l'Archive,
- norme(s) pour la livraison de données numériques par l'Archive,
- norme(s) pour le versement à l'Archive de métadonnées numériques, concernant des données numériques ou physiques,
- norme(s) pour l'identification de données numériques dans l'Archive,
- norme(s) de protocole pour la recherche et l'extraction d'informations concernant des données numériques et physiques,
- norme(s) pour l'accès aux supports permettant le remplacement des systèmes de gestion des supports sans devoir réécrire les supports,
- norme(s) pour des supports physiques spécifiques,
- norme(s) pour la migration de l'information vers de nouveaux supports et formats,
- norme(s) pour les pratiques d'archivage recommandées,
- norme(s) pour l'agrément des Archives.

## **1.6 STRUCTURE DU DOCUMENT**

### **1.6.1 RECOMMANDATIONS POUR LA LECTURE DU DOCUMENT**

Tous les lecteurs devraient lire les sous-sections « Objet et domaine d'application » (1.1), « Applicabilité » (1.2), et « Conformité » (1.4) pour avoir une idée des objectifs et de l'applicabilité du présent document.

Les lecteurs ne souhaitant qu'un aperçu des principaux concepts, doivent lire également les sections « Concepts d'un OAIS » (section 2) et « Responsabilités d'un OAIS » (section 3).

Les lecteurs qui sont appelés à réaliser ou à administrer quotidiennement des Archives de type OAIS doivent lire la totalité du document.

## 1.6.2 ORGANISATION PAR SECTIONS

La section 1 comprend les sous-sections concernant l'objet, le domaine d'application, l'applicabilité et les définitions, communes à de nombreuses normes. Elle comprend également la justification de l'effort, les exigences de conformité et un guide pour le développement de normes associées.

La section 2 comporte une présentation de haut niveau des principaux concepts utilisés dans une Archive OAIS. Elle fournit une présentation de l'environnement d'une Archive OAIS et des rôles joués par les acteurs en relation avec l'Archive. Elle précise ce que signifie le terme "information" et ce qui est nécessaire pour la conserver à Long terme. Elle définit les concepts essentiels pertinents pour la réalisation d'Archives conformes à l'OAIS.

La sous-section 3.1 définit les responsabilités obligatoires qu'une Archive OAIS doit assumer pour conserver l'information, et la sous-section 3.2 précise les types d'activités nécessaires dans de nombreuses Archives pour assumer ces responsabilités.

La section 4 présente les modèles nécessaires à une compréhension détaillée d'une Archive OAIS. Elle décompose l'OAIS en un certain nombre de domaines fonctionnels et identifie certains services élaborés au niveau de leurs interfaces. De même, elle présente des modèles de données détaillés d'informations s'appuyant sur des schémas réalisés en UML (*Unified Modeling Language*).

La section 5 fournit quelques perspectives sur les problèmes posés par la conservation de l'information lorsqu'on utilise la Migration numérique entre supports et entre nouveaux formats ou nouvelles représentations. Elle fournit également des perspectives sur les questions concernant la conservation des services d'accès à l'information numérique par portage de logiciel, par encapsulation, et par émulation de matériel.

La section 6 est une introduction aux différentes possibilités permettant, grâce aux partenariats entre Archives, de fournir des services plus étendus et plus rentables.

Les annexes ne font pas partie de la Recommandation et sont fournies pour faciliter la lecture :

- l'annexe A met en relation certaines parties du présent Modèle de référence avec d'autres normes,
- l'annexe B fournit un bref tutoriel sur le Langage UML,
- l'annexe C fournit un modèle d'information en couches,

- l'annexe D fournit une liste des références informatives,
- l'annexe E fournit un index croisé des principaux mots clés et de leur utilisation dans le document

## 1.7 DEFINITIONS

### 1.7.1 ACRONYMES ET ABREVIATIONS

Les acronymes et abréviations sont identiques à ceux de la version anglaise de la présente Recommandation.

<b>AIC</b>	Archival Information Collection (Collection d'informations archivée)
<b>AIP</b>	Archival Information Package (Paquet d'informations archivé)
<b>AIU</b>	Archival Information Unit (Unité d'informations archivée)
<b>ASCII</b>	American Standard Code for Information Interchange
<b>CAD</b>	Computer-Automated Design
<b>CCSDS</b>	Consultative Committee for Space Data Systems
<b>CDO</b>	Content Data Object
<b>CD-ROM</b>	Compact Disk - Read Only Memory
<b>CEOS</b>	Committee on Earth Observation Satellites
<b>CRC</b>	Cyclic Redundancy Check
<b>DED</b>	Data Entity Dictionary
<b>DBMS</b>	Data Base Management System
<b>DDL</b>	Data Description Language
<b>DIP</b>	Dissemination Information Package (Paquet d'informations diffusé)
<b>DVD</b>	Digital Versatile Disk
<b>EBCDIC</b>	Extended Binary Coded Decimal Interchange Code
<b>ECS</b>	EOSDIS Core System
<b>EOSDIS</b>	Earth Observing System Data and Information System
<b>FITS</b>	Flexible Image Transfer System
<b>GIF</b>	Graphics Interchange Format
<b>HFMS</b>	Hierarchical File Management System
<b>ICS</b>	Interoperable Catalogue System
<b>IEEE</b>	Institute of Electrical and Electronic Engineers
<b>IMS</b>	Information Management System
<b>IP</b>	Information Packaue (Paquet d'informations)
<b>ISBN</b>	International Standard Book Number
<b>ISO</b>	International Organization for Standardization
<b>LSDA</b>	Life Sciences Data Archive
<b>NARA</b>	National Archives and Records Administration
<b>NASA</b>	National Aeronautics and Space Administration
<b>NSSDC</b>	National Space Science Data Center
<b>OAIS</b>	Open Archival Information System (Système ouvert d'archivage d'information)



<b>ODL</b>	Object Description Language
<b>PDI</b>	Preservation Description Information (Information de pérennisation)
<b>PDMP</b>	Project Data Management Plan
<b>PDS</b>	Planetary Data System
<b>PSDD</b>	Planetary Science Data Dictionary
<b>SIP</b>	Submission Information Package (Paquet d'informations à verser)
<b>UML</b>	Unified Modeling Language
<b>UNICODE</b>	Universal Code
<b>WWW</b>	World-Wide Web

### 1.7.2 TERMINOLOGIE

Un bon nombre des termes utilisés dans le présent Modèle de référence ont une signification bien définie qui est présentée dans cette sous-section. Lors de leur première utilisation dans le texte, ces termes figurent en gras, avec la première lettre en majuscule. Dans la suite du texte, ils sont utilisés uniquement avec la première lettre en majuscule.

Afin de faciliter la lecture de ce document, les conventions suivantes sont spécifiques de la version française et ont été appliquées aux termes du glossaire :

- le nom des entités fonctionnelles apparaît entre guillemets,
- le type d'un objet apparaît séparé d'un tiret, par exemple Objet-données,
- dans le glossaire, le terme français est suivi entre parenthèses du terme d'origine en anglais, ainsi que, s'il existe, de l'acronyme d'origine.

Le présent Modèle de référence s'applique à toutes les disciplines et organisations qui conservent et fournissent de l'information sous forme numérique (ou prévoient de le faire). Par conséquent, les termes définis ci-dessous peuvent ne pas correspondre à tous les termes d'une discipline particulière (par exemple, les archives traditionnelles, les bibliothèques numériques, les centres de données scientifiques). L'approche retenue est celle d'utiliser des termes non encore chargés de signification, afin de réduire le risque de véhiculer des acceptions involontaires. Il sera donc nécessaire pour toutes les disciplines et toutes les organisations de trouver la correspondance entre certains de leurs termes propres les plus courants, et les termes du présent Modèle de référence OAIS. Ceci ne devrait pas être difficile et doit être perçu comme une contribution à la réussite du Modèle de référence et non comme une dissuasion. Par exemple, l'archivistique met l'accent sur la conservation de « Records ». Ce terme n'est pas utilisé dans le Modèle de référence OAIS, mais peut être rattaché globalement à « Contenu d'information à l'intérieur d'un Paquet d'informations archivé » (voir les définitions ci-dessous, ainsi qu'en 2.2 et 4.2 pour le contexte). Quelques correspondances sont fournies à l'Annexe A.

**AIP dérivé** (Derived AIP) : Paquet d'informations archivé (AIP) produit par l'extraction ou l'agrégation d'informations à partir d'un ou plusieurs AIP.

**Archive** (Archive) : organisation chargée de conserver l'information pour permettre à une Communauté d'utilisateurs cible d'y accéder et de l'utiliser.

**Archives coopérantes** (Co-operating Archives) : Archives dont les Communautés d'utilisateurs cibles partagent des centres d'intérêt. Elles peuvent se commander et se verser des données entre elles. Pour les requêtes inter-archives, ces Archives coopératives doivent se mettre d'accord sur au moins un Paquet d'informations à verser (SIP) et un Paquet d'informations diffusé (DIP) communs.

**Base de connaissance** (Knowledge Base) : ensemble d'informations préalablement acquises par une personne ou un système lui permettant de comprendre l'information reçue.

**Client** (Client) : application qui échange de l'information avec une autre application (voir aussi Utilisateur).

**Collection d'informations archivée** (Archival Information Collection - AIC) : Paquet d'informations archivé (AIP) dont le Contenu d'information est un agrégat d'autres AIP.

**Collection virtuelle** (Access Collection) : collection de Paquets d'informations archivés (AIP), définie par une Description de collection, mais dépourvue d'une Information d'empaquetage dans l'Entité « Stockage ».

**Commande** (Adhoc Order) : requête formulée par un Utilisateur pour une information déclarée disponible par l'OAIS.

**Communauté d'utilisateurs cible** (Designated Community) : groupe identifié d'utilisateurs potentiels, susceptibles de comprendre un ensemble donné d'informations. La Communauté d'utilisateurs cible peut être constituée de plusieurs communautés d'utilisateurs.

**Communauté élargie** (Global Community) : Communauté d'utilisateurs cible étendue, dans le cadre d'un Groupement d'archives, ayant accès aux collections de plusieurs archives via un ou plusieurs Outils de recherche communs.

**Communauté locale** (Local Community) : dans le cadre d'un Groupement d'archives, Communauté d'utilisateurs cible d'un des membres du groupement.

**Contenu d'information** (Content Information) : ensemble d'informations constituant l'objet principal de la pérennisation. Il s'agit d'un Objet-information constitué d'un Objet-contenu de données et de son Information de représentation. Le Contenu d'information sera par exemple un simple tableau de nombres représentant des températures et interprétables comme telles, mais sans la documentation associée décrivant son historique et son origine, son lien avec d'autres observations, etc.

**Copie de substitution** (Replication) : Migration numérique sans modification de l'Information d'empaquetage ni du Contenu d'information ni de l'Information de pérennisation (PDI). Les bits utilisés pour représenter ces Objets-information sont conservés lors du transfert, que le support soit le même ou soit nouveau.

**Demande d'abonnement** (Event Based Order) : requête émise par un Utilisateur pour obtenir une information périodique produite sur la base d'un événement ou d'une série d'événements.

**Description de collection** (Collection Description) : type de Description de paquet dédié à la fourniture d'informations sur une Collection d'informations archivée (AIC). Cette information est destinée aux Outils d'accès.

**Description d'élément** (Member Description) : Notice descriptive d'une partie de collection.

**Description de paquet** (Package Description) : information destinée aux Outils d'accès.

**Description d'unité** (Unit Description) : type de Description de paquet dédié à la fourniture d'informations sur une Unité d'informations archivée (AIU). Cette information est destinée aux Outils d'accès.

**Description globale** (Overview Description) : spécialisation de la Description de collection, prenant en compte la collection dans sa globalité.

**Dictionnaire de données** (Data dictionary) : référentiel formel de termes utilisés pour décrire des données.

**Données** (Data) : une représentation formalisée de l'information, adaptée à la communication, l'interprétation ou le traitement. Exemple : une séquence de bits, un tableau de nombres, les caractères d'une page, un enregistrement de paroles ou un échantillon de roche lunaire.

**Données de gestion** (Data Management Data) : données créées et stockées dans l'espace permanent de stockage de l'Entité « Gestion de données » et qui sont relatives au fonctionnement de l'archive. Citons par exemple : données de facturation, données d'habilitation des Utilisateurs, réglementation interne, Demande d'abonnement, historique de la conservation et données statistiques pour le rapport d'activité.

**Edition** (Edition) : lorsque le contenu d'un Paquet d'informations archivé (AIP) a été enrichi ou amélioré, et que le nouvel AIP ainsi créé est susceptible de remplacer l'AIP d'origine, on tracera ce changement à l'aide d'un attribut de l'AIP appelé Edition.

**Entité « Accès »** (Access) : entité de l'OAIS regroupant les fonctions et services de mise à disposition des utilisateurs des collections archivées.

**Entité « Administration »** (Administration) : entité de l'OAIS regroupant les fonctions et services requis pour une supervision continue du fonctionnement des autres entités de l'OAIS.

**Entité « Entrées »** (Ingest) : entité de l'OAIS regroupant les fonctions et services qui :

- prennent en charge les Paquets d'informations à verser (SIP) livrés par les Producteurs,
- préparent les Paquets d'informations archivés (AIP) en vue de leur stockage,
- et assurent la bonne intégration dans l'OAIS de ces AIP et de leur Information de description.

**Entité « Gestion de données »** (Data Management) : entité de l'OAIS regroupant les fonctions et services dédiés à l'alimentation, la maintenance et la mise à disposition d'informations aussi variées que les catalogues et inventaires de ce qu'il est possible de se procurer auprès de l'Entité « Stockage », les algorithmes de traitement applicables aux données récupérées, les statistiques de consultation, les éléments de facturation, les Demandes d'abonnement, les contrôles de sécurité, les plannings, la politique et les procédures propres à l'OAIS.

**Entité « Stockage »** (Archival Storage) : entité de l'OAIS regroupant les fonctions et services utilisés pour le stockage et la récupération des Paquets d'informations archivés (AIP).

**Groupement d'Archives** (Federated Archives) : groupe d'Archives qui ont accepté de mutualiser l'accès à leurs fonds via un ou plusieurs Outils de recherche communs.

**Immédiatement compréhensible** (Independently Understandable) : qualité propre à une information suffisamment documentée pour être comprise et utilisée par la Communauté d'utilisateurs cible sans recourir à des ressources particulières difficiles d'accès, comme par exemple des personnes physiques.

**Information** (Information) : toute connaissance pouvant être échangée. Lors de l'échange, elle est représentée par des données. Exemple : une chaîne de bits (les données) accompagnée d'une description permettant d'interpréter cette chaîne de bits comme des nombres représentant des mesures de températures en degrés Celsius (Information de représentation).

**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 d'information.

**Information de description** (Descriptive Information) : ensemble d'informations, constitué principalement de Descriptions de paquet, et fourni à l'Entité « Gestion de données » pour aider les Utilisateurs à rechercher, commander et récupérer des informations de l'OAIS.

**Information d'emballage** (Packaging Information) : information permettant de relier et identifier les composants d'un Paquet d'informations. Par exemple les informations de volume et de répertoire dans un CD-ROM conforme à la norme ISO 9660, permettant d'accéder aux fichiers supports du Contenu d'information et de l'Information de pérennisation (PDI).

**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.

**Information de pérennisation** (Preservation Description Information - PDI) : information nécessaire à une bonne conservation du Contenu d'information, et qui peut être décomposée en Informations de provenance, d'identification, d'intégrité et de contexte.

**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 représentation** (Representation Information) : information qui traduit un Objet-données en des concepts plus explicites. Par exemple, la définition du code ASCII décrit comment une séquence de bits (un Objet-données) est convertie en caractères.

**Information de structure** (Structure Information) : information qui explique la façon dont d'autres informations sont organisées. Elle établit par exemple une correspondance entre les trains de bits et les types de données courants sur ordinateurs (tels que caractères, nombres, pixels ou agrégats de ces types tels que chaînes de caractères et tableaux).

**Jeu de résultats** (Result Set) : ensemble des éléments descriptifs des Paquets d'informations archivés (AIP) d'un OAIS répondant aux critères de recherche d'un Utilisateur, ou encore, résultats issus d'une recherche soumise à l'Entité « Gestion de données ».

**Logiciel d'accès** (Access Software) : logiciel à même de présenter tout ou partie de l'information sous une forme compréhensible par l'homme ou la machine.

**Logiciel de restitution de la représentation** (Representation Rendering Software) : logiciel permettant d'afficher l'Information de représentation d'un Objet-information sous forme compréhensible par l'homme.

**Long terme** (Long Term) : période suffisamment longue pour qu'il soit nécessaire de prendre en compte les changements technologiques, et notamment la gestion des nouveaux supports et formats de données ainsi que l'évolution de la communauté d'utilisateurs. Cette période n'est pas limitée dans le temps.

**Management** (Management) : décideurs de la politique globale de l'OAIS en cohérence avec la politique générale de l'organisme ou de l'entreprise.

**Métadonnées** (Metadata) : données sur d'autres données.

**Méthode d'accès** (Access Method) : méthode de récupération d'un Paquet d'informations archivé (AIP) à partir de son nom ou de son identificateur, accessible aux utilisateurs autorisés.

**Migration numérique** (Digital migration) : transfert d'informations numériques au sein de l'OAIS dans le but de les pérenniser. La Migration numérique se distingue des autres types de transfert sur trois points :

- on vise la pérennisation de l'intégralité du contenu,
- on présume que l'information résultant de ce transfert est destinée à remplacer l'ancienne,
- on admet que l'OAIS dispose de la pleine maîtrise et responsabilité de tous les aspects du transfert.

**Modèle de référence** (Reference Model) : cadre permettant la compréhension des relations essentielles entre entités d'un environnement donné, ainsi que le développement de normes ou de spécifications cohérentes liées à cet environnement. Un Modèle de référence s'appuie sur un nombre restreint de concepts unificateurs et peut servir de support à la formation et à la sensibilisation des non-spécialistes à ces normes.

**Notice descriptive** (Associated Description) : information décrivant le contenu d'un Paquet d'informations pour un Outil d'accès particulier.

**Objet-contenu de données** (Content Data Object) : Objet-données qui, avec l'Information de représentation associée, est l'objet principal de la pérennisation.

**Objet-données** (Data Object) : Objet physique ou Objet numérique.

**Objet-information** (Information Object) : Objet-données avec son Information de représentation.

**Objet numérique** (Digital Object) : objet constitué de séquences de bits.

**Objet physique** (Physical Object) : objet (roche lunaire, spécimen biologique, préparation pour microscope, etc) présentant des propriétés physiques observables qui sont autant d'informations qu'il convient de documenter pour la pérennisation, la diffusion et l'usage indépendant.

**Opération de diffusion de données** (Data Dissemination Session) : livraison d'un lot de supports matériels ou télétransmission, destinée à fournir des Données à un Utilisateur. Le mode de transmission et le contenu de l'Opération de diffusion de données sont basés sur un modèle de données négocié entre l'OAIS et l'Utilisateur dans un Protocole de commande. Ce

modèle de données spécifie les structures logiques utilisées par l'OAIS et leur mode de représentation pour chaque livraison de supports matériels ou lors de chaque télétransmission.

**Opération de versement** (Data Submission Session) : livraison d'un lot de supports matériels ou télétransmission, destinée à fournir des Données à un OAIS. Le mode de transmission et le contenu de l'Opération de versement sont basés sur un modèle de données négocié entre l'OAIS et le Producteur dans un Protocole de versement. Ce modèle de données spécifie les structures logiques utilisées par le Producteur et leur mode de représentation pour chaque livraison de supports matériels ou lors de chaque télétransmission.

**Outil d'accès** (Access Aid) : logiciel ou document qui permet aux Utilisateurs de localiser, analyser et commander les Paquets d'informations archivés (AIP) qui les intéressent.

**Outil de commande** (Ordering Aid) : application qui permet à l'Utilisateur de connaître le coût des Paquets d'informations archivés (AIP) qui l'intéressent, et de les commander.

**Outil de recherche** (Finding Aid) : Outil d'accès qui permet à un Utilisateur de rechercher et d'identifier les Paquets d'informations archivés (AIP) pertinents pour sa recherche.

**Outil de récupération** (Retrieval Aid) : application qui permet aux utilisateurs habilités de récupérer le Contenu d'information et les Informations de pérennisation (PDI) décrites dans la Description de paquet.

**Paquet d'informations** (Information Package – IP) : association du Contenu d'information et de son Information de pérennisation (PDI) destinée à faciliter la conservation du Contenu d'information. A ce Paquet d'informations est aussi associée une Information d'emballage utilisée pour circonscrire et identifier le Contenu d'information et son PDI.

**Paquet d'informations archivé** (Archival Information Package - AIP) : Paquet d'informations conservé dans un OAIS et constitué d'un Contenu d'information et de l'Information de pérennisation (PDI) associée.

**Paquet d'informations à verser** (Submission Information Package - SIP) : Paquet d'informations livré par le Producteur à l'OAIS pour l'élaboration d'un ou plusieurs Paquets d'informations archivés (AIP).

**Paquet d'informations diffusé** (Dissemination Information Package - DIP) : Paquet d'informations reçu par l'Utilisateur en réponse à sa requête à l'OAIS. Ce paquet provient d'un ou de plusieurs Paquets d'informations archivés (AIP).

**Pérennisation** (Long Term Preservation) : fait de conserver à Long terme l'information sous une forme correcte et Immédiatement compréhensible.

**Producteur** (Producer) : toute personne ou système client qui fournit des informations à pérenniser. Il peut s'agir d'autres OAIS, ou de personnes ou systèmes internes à l'OAIS.

**Protocole de commande** (Order Agreement) : accord entre l'Archive et l'Utilisateur spécifiant les détails physiques de la livraison, tels que le type de support et le format des Données.

**Protocole de versement** (Submission Agreement) : accord négocié entre l'OAIS et le Producteur, spécifiant un modèle de données pour l'Opération de versement. Ce modèle identifie le format, le contenu et la structure logique utilisés par le Producteur et leur mode de représentation pour chaque support de livraison ou lors d'une télétransmission.

**Rafraîchissement de support** (Refreshment) : Migration numérique visant à remplacer un support par une copie suffisamment exacte pour que l'Entité « Stockage » (matériel et logiciel) continue à fonctionner comme auparavant.

**Ré-empaquetage** (Repackaging) : Migration numérique au cours de laquelle l'Information d'empaquetage du Paquet d'information archivé (AIP) est modifiée.

**Réseau d'informations de représentation** (Representation Network) : ensemble d'Informations de représentation qui décrit entièrement la signification d'un Objet-données. L'Information de représentation étant elle-même sous forme numérique, elle exige une Information de représentation supplémentaire pour que sa forme numérique soit compréhensible sur le Long terme.

**Services de base** (Common Services) : services supports tels que la communication interprocessus, les services de nommage, l'affectation d'espace de stockage provisoire, le traitement d'exceptions, la sécurité et les services d'annuaires nécessaires au fonctionnement de l'OAIS.

**Session de recherche** (Search Session) : session de travail d'un Utilisateur, au cours de laquelle il utilise des Outils de recherche pour identifier et repérer les collections pertinentes pour sa recherche.

**Système ouvert d'archivage d'information** (Open Archival Information System - OAIS) : Archive, avec son équipe et ses systèmes, officiellement chargée de conserver l'information et de la mettre à disposition d'une Communauté d'utilisateurs cible. L'OAIS assume un ensemble de responsabilités, définies à la section 3.1, qui permettent de le distinguer d'une autre Archive. Le terme « ouvert » dans OAIS signifie que cette recommandation et les recommandations et normes ultérieures liées à l'OAIS sont élaborées dans des forums ouverts, et en aucun cas que l'accès à l'Archive est libre.

**Transformation** (Transformation) : Migration numérique au cours de laquelle le Contenu d'information ou l'Information de pérennisation (PDI) d'un Paquet d'informations archivé (AIP) est modifié. Par exemple, la conversion des codes ASCII en UNICODE dans un document textuel est une transformation.

**Unité d'informations archivée** (Archival Information Unit - AIU) : Paquet d'informations archivé (AIP) dont le Contenu d'information ne peut plus être décomposé en d'autres



éléments de Contenu d'information, doté chacun de sa propre Information de pérennisation (PDI). Il peut être considéré comme un AIP « élémentaire ». Exemple d'AIU : un tableau de nombres représentant des températures dans une région donnée, auquel est associée toute la documentation expliquant où et avec quelle méthode les mesures ont été effectuées, avec quels instruments, par qui, pour quels motifs, quel traitement leur a été appliqué, qui a été responsable de la conservation de ces mesures depuis leur établissement, quels liens unissent ces mesures à d'autres informations, comment elles peuvent être référencées de manière unique, etc.

**Utilisateur** (Consumer) : toute personne ou système client en relation avec les services de l'OAIS pour trouver des informations archivées présentant un intérêt, et pour accéder au détail de ces informations. Il peut s'agir aussi d'autres OAIS, ou de personnes ou systèmes internes à l'OAIS.

**Version** (Version) : lorsqu'un Paquet d'informations archivé (AIP) a subi une transformation, et que le nouvel AIP ainsi créé a vocation à remplacer l'AIP d'origine, on tracera ce changement à l'aide d'un attribut de l'AIP appelé Version.



## 2 CONCEPTS OAIS

L'objet de cette section est de justifier et de décrire les concepts clés d'un OAIS. Une présentation plus complète, ainsi qu'une modélisation formelle de ces concepts sont fournies section 4.

Aujourd'hui, le terme « Archive » sert à désigner une grande variété de fonctions et de systèmes de stockage et de pérennisation. On entend par Archive traditionnelle un système ou une organisation qui conserve des documents d'archives créés à l'origine par ou pour un organisme d'Etat, une institution ou une société privée, et qui permet à des communautés publiques ou privées d'y accéder. L'Archive, dès lors qu'elle est détentrice des documents d'archives, s'assure qu'ils sont compréhensibles pour les communautés susceptibles d'y accéder, et les gère de façon à conserver leur contenu d'information et leur authenticité. Historiquement, ces documents d'archive se présentaient sous forme de livres, de supports papier, de cartes, de photographies et de films, lisibles directement par l'homme ou à l'aide de simples moyens d'agrandissement optique ou d'outils de visualisation. Le principal objectif de la conservation de ces informations était de s'assurer qu'elles étaient stockées sur des supports stables à long terme, et que l'accès à ces supports était soigneusement contrôlé.

Le développement massif et rapide de l'information sous forme numérique a constitué un sérieux défi pour les Archives traditionnelles et leurs fournisseurs d'information, mais aussi pour de nombreux autres organismes d'Etat, le secteur privé et les organismes à but non lucratif. Toutes ces entités réalisent, ou vont réaliser, qu'elles devront prendre en charge les fonctions de conservation de l'information qui sont habituellement dévolues aux Archives traditionnelles, car les données numériques sont facilement perdues ou corrompues. La rapidité de l'évolution technologique rend obsolètes les matériels et les logiciels au bout de quelques années. Cette évolution peut remettre fortement en cause la capacité des formats ou des structures de données à représenter correctement toute l'information. En effet, la majeure partie des méta-informations nécessaires à la conservation de l'information n'est facilement disponible, ou exclusivement disponible, qu'au moment de la production de l'information originale, de sorte que ces organisations doivent s'investir dans l'effort de Pérennisation et suivre les principes retenus dans le présent Modèle de référence OAIS afin de garantir la Pérennisation de l'information. Leur participation à ces efforts minimisera les coûts liés à la gestion du cycle de vie du document et permettra effectivement cette Pérennisation.

L'augmentation fulgurante de la puissance des ordinateurs et de la capacité des supports numériques a donné lieu au développement de nombreux systèmes pour lesquels le rôle du Producteur et le rôle de l'Archive sont de la responsabilité d'une même entité. Ces systèmes, appelés parfois Archives actives, devraient adhérer aux objectifs de la Pérennisation abordés dans le présent document. La conception de tels systèmes doit prendre en compte le fait que certaines activités liées à la pérennisation peuvent être en conflit avec les objectifs de production et de diffusion rapides de produits pour les Utilisateurs. Les concepteurs et les architectes de ces systèmes devraient donc documenter les solutions adoptées.

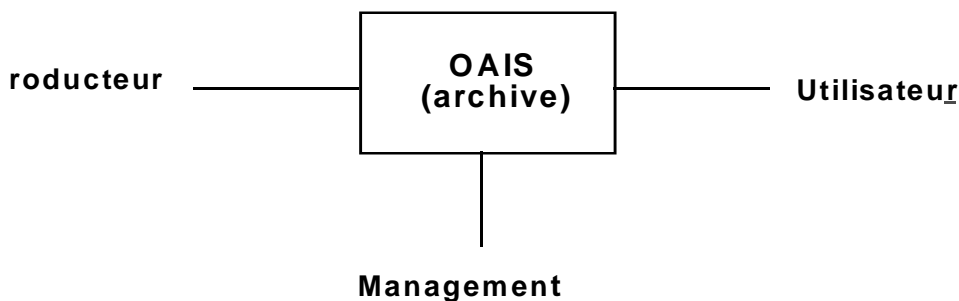
L'un des objectifs principaux du présent Modèle de référence est de permettre une compréhension plus large des éléments nécessaires à la pérennisation et à l'accès à cette

information sur le Long terme. Pour éviter toute confusion avec de simples fonctions de "stockage de bits", le Modèle de référence définit un Système ouvert d'archivage d'information (OAIS) qui assure la Pérennisation de l'information et l'accès à cette information. La spécificité d'une Archive OAIS est de conserver l'information afin qu'une Communauté d'utilisateurs cible puisse y accéder et l'utiliser, et d'être conforme aux exigences énoncées dans la section 3. Il peut s'agir d'Archives qui doivent faire face à des flux d'informations entrants réguliers ou d'autres dont les entrées sont essentiellement ponctuelles. Il peut s'agir également d'Archives fournissant une grande variété de services d'accès évolués, aussi bien que de celles qui n'acceptent que les types de demandes les plus simples. Dans le reste du présent document, les termes Archive et OAIS seront utilisés de façons synonymes et feront référence à une Archive OAIS, sauf si le contexte indique clairement que ce n'est pas le cas (par exemple : Archives traditionnelles).

Le modèle OAIS prend acte du caractère déjà très répandu des fonds d'informations numériques et du besoin pour les mises en oeuvre locales de disposer de règles et de procédures efficaces pour la conservation des informations. Ceci peut donner lieu à une grande diversité de schémas organisationnels, et à des rôles variés pour les Archives traditionnelles, dans cet objectif de conservation. Les organisations qui souhaitent pérenniser de l'information devraient trouver dans l'utilisation des termes et des concepts de l'OAIS une assistance dans la poursuite de cet objectif.

## 2.1 ENVIRONNEMENT OAIS

Le modèle simple du schéma 2-1 décrit l'environnement d'un OAIS.



**Schéma 2-1 : Modèle d'environnement d'un OAIS**

L'OAIS s'interface avec les **Producteurs**, les **Utilisateurs**, et le **Management** :

- Le rôle de Producteur est joué par les personnes, ou les systèmes clients, qui fournissent l'information à conserver.
- Le rôle de Management est joué par les décideurs qui définissent la politique globale de l'OAIS composante d'une politique plus vaste. En d'autres termes, la conduite du Management de l'OAIS n'est qu'une des responsabilités du Management. Le

Management n'est pas concerné par le fonctionnement quotidien de l'Archive. La responsabilité de la gestion quotidienne de l'OAIS fait partie de l'Entité « Administration » du modèle OAIS, qui sera décrite en 4.1.

- Le rôle d'Utilisateur est joué par les personnes, ou les systèmes clients, qui interagissent avec les services de l'OAIS pour chercher et récupérer les informations conservées qui les intéressent. La Communauté d'utilisateurs cible est une classe particulière d'Utilisateurs : c'est l'ensemble des Utilisateurs en mesure de comprendre les informations conservées.

Certaines Archives OAIS peuvent ne pas être visibles par la Communauté d'utilisateurs cible. Ces Archives peuvent établir des accords particuliers, cohérents avec le Management et les besoins de l'OAIS. Elles peuvent, sur la base d'accords préalables et pour des raisons diverses, interagir de façon plus ou moins formalisée avec une autre Archive. Tel OAIS peut assumer le rôle de Producteur pour un autre OAIS : ceci se produit par exemple en cas de transfert de la responsabilité de conservation d'un type d'information d'une Archive vers une autre. Tel autre peut jouer le rôle d'Utilisateur d'un autre OAIS : ceci se produit notamment lorsque le premier OAIS décide de confier à un autre OAIS un type d'information dont il a rarement besoin et qu'il a choisi de ne pas conserver localement. Un accord formel portant notamment sur les changements de réglementation pouvant affecter cette relation doit être établi. Un tel accord doit prévoir les besoins de communication entre Archives pour tout changement de cette nature. L'étendue des interactions possibles entre Archives OAIS est exposée à la section 6, « Interopérabilité des Archives ».

## 2.2 INFORMATIONS DE L'OAIS

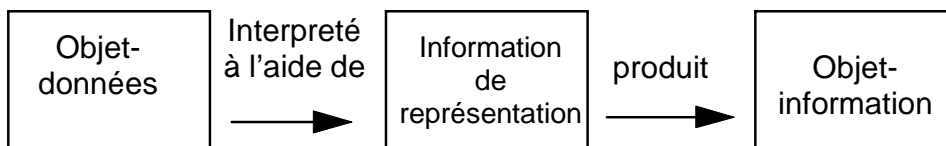
### 2.2.1 DEFINITION DE L'INFORMATION

Pour qu'un OAIS puisse conserver une information, il est capital qu'elle soit clairement définie. Quelques concepts essentiels sont présentés dans cette sous-section et une modélisation formelle de l'information est fournie à la section 4.

On peut dire d'une personne ou d'un système qu'il a une **Base de connaissance** lui permettant de comprendre l'information reçue. Par exemple, une personne qui a une Base de connaissance incluant la compréhension de l'anglais sera en mesure de lire et de comprendre un texte en anglais.

L'information est définie comme tout type de connaissance pouvant être échangé, et cette information est toujours exprimée (c'est-à-dire représentée) par un certain type de données. Par exemple, l'information contenue dans un livre traditionnel est habituellement exprimée par des caractères observables (les données) qui, lorsqu'ils sont combinés à la connaissance de la langue utilisée (la Base de connaissance), sont convertis en information plus intelligible. Si la base de connaissance du lecteur n'inclut pas l'anglais, le texte en anglais (les données) devra être accompagné d'un dictionnaire anglais et de l'information grammaticale (c'est-à-dire l'**Information de représentation**), ces derniers devant être compréhensibles par la base de connaissance de ce lecteur.

De la même façon, l'information stockée dans un fichier sur un CD-ROM est exprimée par les bits (les données) qu'il contient. Ces bits, une fois combinés avec leur Information de représentation, sont convertis en information plus intelligible à condition que l'Information de représentation soit compréhensible par la Base de connaissance du destinataire. Prenons le cas où les bits représentent un tableau ASCII de nombres qui indiquent les coordonnées d'un lieu sur la Terre, exprimées en degrés de latitude et longitude Est. Dans ce cas, l'Information de représentation inclura la définition des caractères ASCII ainsi que la description du format des nombres et leur emplacement dans le fichier, leur définition en tant que latitude et longitude, et la définition de leurs unités en degrés. Elle peut également inclure des éléments sémantiques supplémentaires attribués à ce tableau. En général, on peut dire que les « données interprétées en utilisant l'Information de représentation produisent l'Information ». C'est ce que montre le schéma 2-2.



**Schéma 2-2 : Obtention d'Information à partir de Données**

Afin que cet **Objet-information** soit conservé avec succès, il est primordial pour un OAIS d'identifier clairement et de comprendre l'**Objet-données** et son Information de représentation. Pour des informations numériques, ceci signifie que l'OAIS doit identifier clairement les bits et l'Information de représentation correspondante. Cette exigence de transparence jusqu'au niveau des bits est une des caractéristiques de la conservation de l'information numérique. Elle va à l'encontre des concepts orientés objet, qui tendent à masquer toutes les questions de mise en oeuvre. Ceci représente un défi significatif pour la conservation de l'information numérique.

Une complexité supplémentaire provient de la nature récursive de l'Information de représentation. En effet, cette dernière se compose de données propres et d'autres Informations de représentation, ce qui conduit à un réseau d'Objets-information de représentation. Étant donné qu'un des objectifs principaux d'un OAIS est de conserver des informations pour une Communauté d'utilisateurs cible, l'OAIS doit connaître la Base de connaissance de cette communauté afin d'établir quelle est l'Information de représentation minimale à conserver. A ce stade, l'OAIS doit choisir entre maintenir l'Information de représentation minimale requise pour sa Communauté d'utilisateurs cible ou maintenir une quantité plus importante d'Informations de représentation pouvant permettre la compréhension par une communauté plus large mais dotée d'une Base de connaissance moins spécialisée. Avec le temps, l'évolution de la Base de connaissance de la Communauté d'utilisateurs cible peut nécessiter des mises à jour de l'Information de représentation, pour assurer la pérennité de la compréhension.

D'un point de vue pratique, un logiciel est utilisé pour accéder à un Objet-information, ce logiciel est doté d'une certaine compréhension du réseau des Objets-information de représentation concernés. Toutefois, ce logiciel ne doit pas dispenser d'identifier et de rassembler les Informations de représentation compréhensibles qui caractérisent cet Objet-information. Il est en effet plus difficile de conserver un logiciel en état de fonctionnement que de conserver de l'information sous forme numérique ou sous forme papier.

Le Modèle de référence OAIS met l'accent sur la conservation du contenu d'information. L'évolution de la technologie numérique, de la technologie multimédia et plus particulièrement l'interdépendance complexe entre données et techniques de présentation de ces données, vont conduire certains organismes à exiger la conservation du "look and feel" de la présentation originale de l'information. Ce type d'exigence peut impliquer la conservation des logiciels et des interfaces utilisés pour accéder aux données. La nature propriétaire de certains logiciels peut rendre ce problème compliqué dans le futur. Différentes techniques pour conserver le "look and feel" de l'accès à l'information font actuellement l'objet de recherches et de prototypage. Elles comprennent l'émulation de matériel, l'émulation de diverses couches logicielles applicatives d'interfaces de programmation (API), et le développement de machines virtuelles. Ces techniques étudient la conservation du train de bits originel et des logiciels au travers des évolutions de la technologie. Bien que le Modèle de référence OAIS ne mette pas l'accent sur ces techniques émergentes, il doit néanmoins fournir une base d'architecture pour le prototypage et la comparaison de ces techniques. Une analyse plus détaillée des problèmes relatifs à la conservation du "look and feel" de l'accès à l'information est présentée à la section 5.2 du présent document.

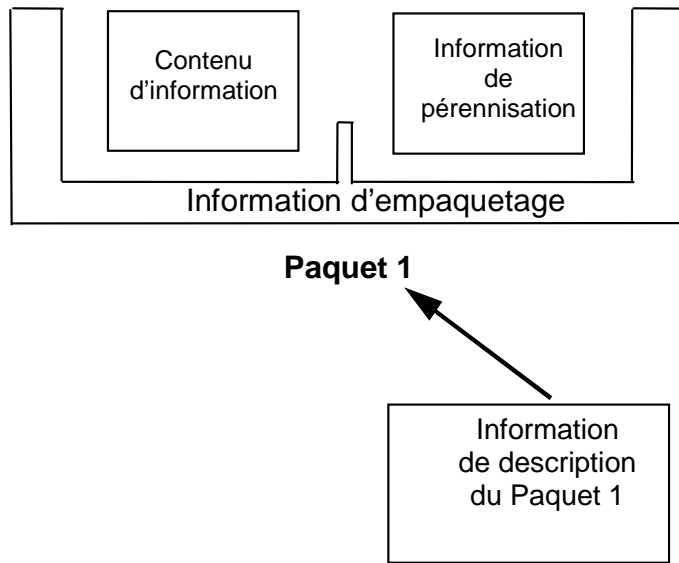
### 2.2.2 DEFINITION D'UN PAQUET D'INFORMATIONS

La définition d'un Objet-information s'applique à tous les types d'informations analysés dans la présente sous-section, et les sous-sections suivantes. En d'autres termes, ils sont tous dotés d'une Information de représentation même si cela n'est pas explicite.

Tout versement d'information à un OAIS 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**.

Un Paquet d'informations est un conteneur conceptuel de deux types d'informations appelés **Contenu d'information** et **Information de pérennisation (PDI)**. Le Contenu d'information et le PDI sont identifiés et encapsulés par une **Information d'empaquetage**. Le paquet qui en résulte peut être retrouvé grâce à l'**Information de description**.

Ces relations au sein du Paquet d'informations sont présentées dans le schéma 2-3.



**Schéma 2-3 : concepts et relations d'un Paquet d'informations**

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) et son Information de représentation, nécessaire à la compréhension de cet objet par la Communauté d'utilisateurs cible. Par exemple, l'Objet-contenu de données peut être une image fournie sous forme d'un train de bits dans un fichier sur un CD-ROM, ce CD-ROM renfermant d'autres fichiers contenant l'Information de représentation.

C'est uniquement après que le Contenu d'information ait été clairement défini qu'une évaluation de l'Information de pérennisation peut être effectuée. 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, dont voici une brève description :

- La provenance décrit l'origine du Contenu d'information, qui en a eu la charge depuis sa création, ainsi que son historique (y compris l'historique des traitements subis).
- Le contexte décrit les relations existant entre le Contenu d'information et d'autres informations situées hors du Paquet d'informations. Par exemple, il peut expliquer pourquoi le Contenu d'information a été produit, et inclure une description de la façon dont ce Contenu est relié à un autre Objet-contenu d'information existant.



- L'identification fournit un ou plusieurs identificateurs, ou systèmes d'identificateurs, grâce auxquels le Contenu d'information peut être identifié de façon unique. Parmi les exemples, on peut citer le numéro ISBN pour un livre, ou un ensemble d'attributs permettant de différencier les Contenus d'information entre eux.
- L'intégrité fournit un mécanisme ou un dispositif protecteur pour prémunir le Contenu d'information contre toute altération non documentée. Par exemple, il peut s'agir d'un checksum sur le Contenu d'information d'un Paquet d'informations numérique.

L'Information d'empaquetage est l'information qui, réellement ou de façon logique, assemble, identifie et met en relation Contenu d'information et PDI. Par exemple, si le Contenu d'information et le PDI correspondent au contenu de certains fichiers sur un CD-ROM, alors l'Information d'empaquetage inclura la norme ISO 9660 relative aux structures de volumes et de fichiers sur CD-ROM, ainsi que les informations concernant les noms des fichiers et des répertoires sur ce CD-ROM.

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.

### 2.2.3 VARIANTES DE PAQUETS D'INFORMATION

Il est nécessaire de distinguer entre un Paquet d'informations conservé par un OAIS et les Paquets d'informations versés à un OAIS ou diffusés par un OAIS. Ces distinctions traduisent le fait que certains versements disposent d'une Information de représentation ou d'un PDI insuffisants pour satisfaire aux exigences finales de conservation de l'OAIS. De plus, le mode d'organisation de ces paquets peut être très différent de celui de l'information conservée par un OAIS. Enfin, les Paquets d'informations fournis aux Utilisateurs par l'OAIS peuvent ne contenir qu'une partie de l'Information de représentation ou du PDI associés au Contenu d'information diffusé. On distinguera donc les appellations : Paquet d'informations à verser (SIP), Paquet d'informations archivé (AIP), et Paquet d'informations diffusé (DIP).

Le **Paquet d'informations à verser (SIP)** est le paquet envoyé à l'OAIS par un Producteur. Sa forme et son contenu détaillés sont négociés entre le Producteur et l'OAIS. La plupart des SIP contiennent du Contenu d'information et du PDI, mais la constitution de l'AIP complet avec son Contenu d'information et son PDI peut nécessiter plusieurs SIP. Un seul SIP peut contenir de l'information destinée à être incluse dans plusieurs AIP. L'Information d'empaquetage sera toujours présente sous une forme ou sous une autre.

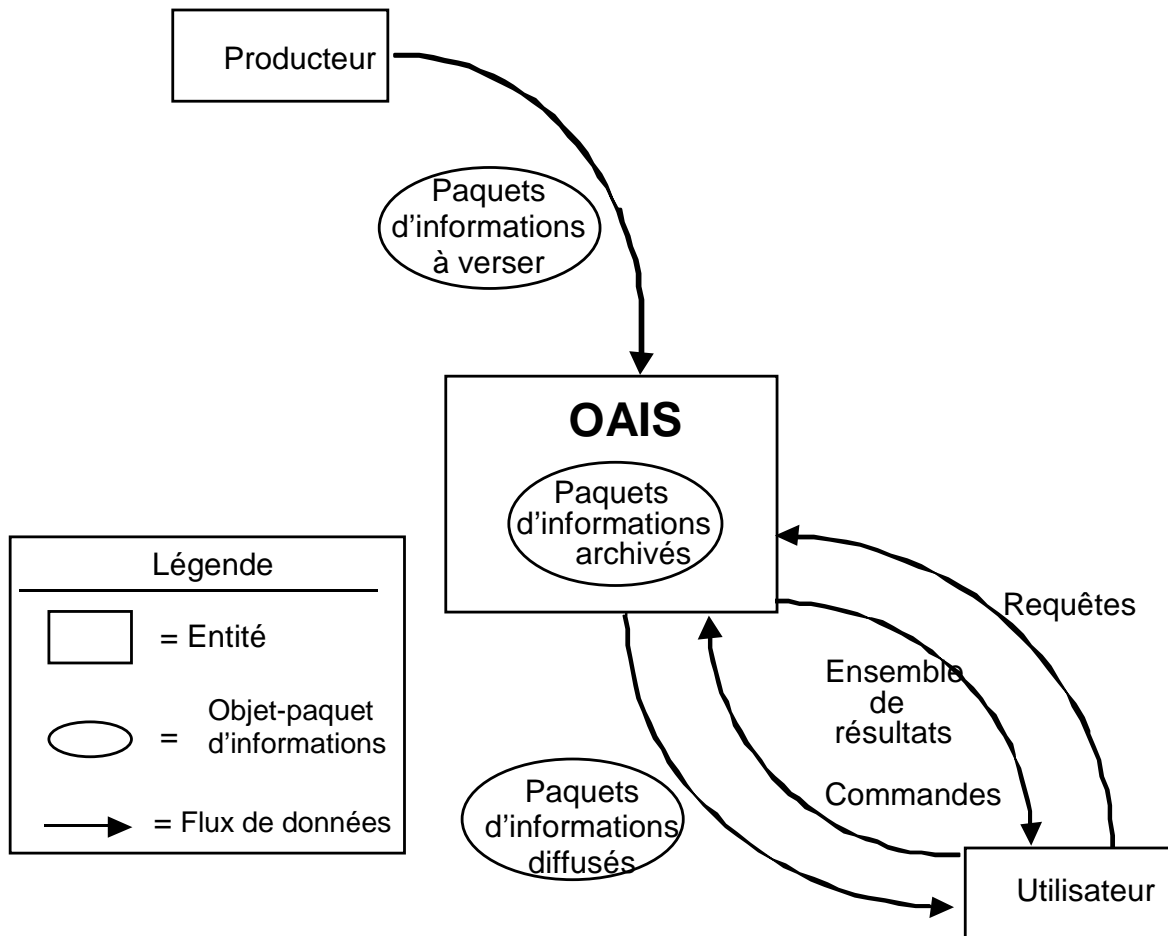
Au sein de l'OAIS, un ou plusieurs SIP sont transformés en un ou plusieurs **Paquets d'informations archivés (AIP)** afin d'être conservés. L'AIP inclut un jeu complet de PDI relatif au Contenu d'information. L'AIP peut également contenir une collection d'autres AIP.

Ceci est analysé et modélisé à la section 4. L'Information d'emballage de l'AIP doit être conforme aux standards internes de l'OAIS, et peut varier dans le cadre de sa gestion par l'OAIS.

En réponse à une requête, l'OAIS fournit tout ou partie d'un AIP à un Utilisateur sous la forme d'un **Paquet d'informations diffusé (DIP)**. Le DIP peut également inclure des collections d'AIP, et contenir un PDI complet ou non. L'Information d'emballage doit nécessairement être présente sous une forme telle que l'Utilisateur puisse clairement distinguer l'information recherchée. En fonction du support de diffusion et des besoins de l'Utilisateur, l'Information d'emballage peut prendre des formes variées.

### **2.3 INTERACTIONS EXTERNES DE HAUT NIVEAU D'UN OAIS**

Les sous-sections suivantes présentent une vue d'ensemble des interactions entre les entités présentes dans un environnement OAIS. Le schéma 2-4 montre un diagramme des flux de données externes d'une Archive OAIS en fonctionnement. Ce diagramme se focalise sur le flux d'information entre les Producteurs, les Utilisateurs et l'OAIS. Il n'inclut pas les flux concernant le Management qui sont traités plus loin, à la section 4.



**Schéma 2-4 : Flux de données externes à une Archive OAIS**

### 2.3.1 INTERACTIONS AVEC LE MANAGEMENT

Le Management fournit à l'OAIS sa charte et son domaine d'application. La charte doit être mise en œuvre par l'Archive, mais il reste important que le Management approuve formellement les activités de l'Archive. Le domaine d'application détermine la composition des groupes Producteurs et Utilisateurs en relation avec l'Archive.

Voici quelques exemples d'interactions entre l'OAIS et le Management :

- Le Management est souvent la source principale de financement pour un OAIS, et peut donner des directives pour l'utilisation des ressources (personnel, équipement, installations).

- Le Management procède en principe à des revues régulières afin d'évaluer les performances de l'OAIS et sa progression vers des objectifs à long terme.
- Le cas échéant, le Management définit, ou du moins approuve, la politique tarifaire des services de l'OAIS.
- Le Management intervient dans la résolution des conflits entre les Producteurs, les Utilisateurs ou l'administration interne d'un OAIS.

Un Management efficace se doit également de soutenir l'OAIS par l'élaboration de procédures qui garantissent l'utilisation de l'OAIS dans son domaine de compétence. Par exemple, la politique managériale doit imposer que toutes les activités financées au sein de ce domaine de compétence conduisent au versement de données à l'Archive et obéissent aux standards et procédures de l'Archive.

### 2.3.2 INTERACTIONS AVEC LE PRODUCTEUR

Le premier contact entre l'OAIS et le Producteur est une demande faite à l'OAIS afin qu'il conserve les données créées par le Producteur. Ce contact peut être initié par l'OAIS, le Producteur ou le Management. Le Producteur établit un **Protocole de versement** avec l'OAIS. Ce protocole identifie les SIP à verser ; il peut couvrir toute la période nécessaire à ce versement. Certains Protocoles de versement refléteront le caractère obligatoire de la fourniture d'information à l'OAIS, alors que d'autres seront basés sur un cadre de versements libres. Même dans le cas où il n'existe aucun Protocole de versement formel, comme cela peut être le cas pour l'archivage d'un site *World Wide Web* (WWW), un Protocole de versement virtuel peut exister, spécifiant le format des fichiers et le thème général de ce site.

Une ou plusieurs **Opérations de versement** sont spécifiées au sein du Protocole de versement. Des intervalles de temps significatifs peuvent séparer ces Opérations de versement. Une Opération de versement peut se traduire par la livraison d'un ou plusieurs SIP sur un ensemble de supports ou par une télétransmission unique. Le contenu de l'Opération de versement est basé sur un modèle de données négocié entre l'OAIS et le Producteur dans le Protocole de versement. Ce modèle de données identifie les composants logiques du SIP (par exemple le Contenu d'information, le PDI, l'Information d'empaquetage, l'Information de description) à fournir et comment ils sont représentés dans chaque Opération de versement où ils figurent. Dans un Protocole de versement, toutes les livraisons de données sont acceptées comme répondant à ce protocole, et se réfèrent en général à un modèle de données cohérent spécifié dans ce protocole. Prenons l'exemple d'une Opération de versement comprenant un ensemble de Contenus d'informations correspondant à un ensemble d'observations, organisés en fichiers sur un CD-ROM. L'Information de pérennisation est elle-même répartie en deux autres fichiers. Tous ces fichiers nécessitent une Information de représentation qui doit être fournie d'une manière ou d'une autre. Le CD-ROM et sa structure de répertoires/fichiers constituent l'Information d'empaquetage qui définit l'encapsulation et l'identification du Contenu d'information et du PDI dans l'Opération de versement. Le Protocole de versement indique comment fournir l'Information de représentation pour chaque fichier, comment reconnaître le CD-ROM, comment utiliser l'Information d'empaquetage pour identifier et encapsuler le Contenu d'information et le PDI du SIP, et spécifie la

fréquence des Opérations de versement (par exemple : une fois par mois pendant deux ans). Il fournit également d'autres informations nécessaires telles que les restrictions d'accès aux données.

Pour être complets, tous les SIP dans une Opération de versement doivent satisfaire aux exigences minimales de complétude de l'OAIS. Toutefois, dans certains cas, il peut être nécessaire de disposer de plusieurs SIP avant de pouvoir constituer un AIP acceptable et le verser intégralement dans l'OAIS. Dans d'autres cas, un seul SIP peut contenir des données à inclure dans plusieurs AIP. Un Protocole de versement contient ou référence également les procédures et les protocoles selon lesquels un OAIS vérifiera vis-à-vis du Producteur la bonne réception et la complétude d'une Opération de versement, ou bien interrogera le Producteur sur le contenu de l'Opération de versement.

### 2.3.3 INTERACTIONS AVEC L'UTILISATEUR

Il existe de nombreux types d'interactions entre l'Utilisateur et l'OAIS. Il peut s'agir de questions posées à un service d'assistance aux utilisateurs, de demandes de documentation, de recherches dans un catalogue, de commandes et de demandes sur l'état d'avancement de la commande. Le processus de commande présente un intérêt particulier pour le Modèle de référence OAIS, dans la mesure où il traite du flux des fonds de l'Archive entre l'OAIS et l'Utilisateur.

L'Utilisateur établit avec l'OAIS un **Protocole de commande** pour recevoir des informations. Ces informations peuvent d'ores et déjà exister dans l'Archive, ou y être versées plus tard. Le Protocole de commande peut porter sur une durée quelconque, et peut impliquer une ou plusieurs **Opérations de diffusion de données**. Une Opération de diffusion de données peut se traduire par le transfert d'un ensemble de supports ou par une télétransmission unique. Le Protocole de commande spécifie le ou les AIP intéressants, la façon dont ces AIP doivent être transformés et structurés en Paquets d'informations diffusés (DIP), et comment ces DIP seront empaquetés pour l'Opération de diffusion de données. Le Protocole de commande va également spécifier d'autres informations utiles telles que des informations sur la livraison (par exemple le nom ou l'adresse postale) et éventuellement les tarifs applicables. Il existe deux types courants de commandes émises par les Utilisateurs : la **Demande d'abonnement** et la **Commande**.

Dans le cas d'une Commande, un Protocole de commande portant sur l'information disponible dans l'Archive est établi entre l'Utilisateur et l'OAIS. Si l'Utilisateur ne connaît pas *a priori* quels sont les fonds spécifiques de l'OAIS qui l'intéressent, il va ouvrir une **Session de recherche** avec l'OAIS. Au cours de cette session, l'Utilisateur va se servir des **Outils de recherche** de l'OAIS qui s'appuient sur l'**Information de description**, ou dans certains cas, sur les AIP eux-mêmes, pour identifier et examiner des fonds potentiellement intéressants. Ceci peut se traduire par la soumission de requêtes de la part de l'Utilisateur et le renvoi à ce dernier d'ensembles de résultats. Ce processus de recherche est le plus souvent itératif, un Utilisateur identifiant dans un premier temps des critères assez larges, et affinant ces critères en fonction des résultats de la recherche précédente. Une fois que l'Utilisateur a spécifié les AIP de l'OAIS qui l'intéressent, il peut établir avec l'OAIS un Protocole de

commande. Cet accord spécifie les identificateurs des AIP que l'Utilisateur souhaite recevoir, et comment les DIP seront récupérés depuis l'OAIS. Si les AIP sont disponibles, une Commande sera passée. Toutefois, si les AIP requis ne sont pas encore disponibles, une Demande d'abonnement peut être passée.

Dans le cas d'une Demande d'abonnement, un Protocole de commande - portant sur l'information que l'Utilisateur désire recevoir sur la base d'un événement déclencheur - est établi entre lui-même et l'OAIS. Cet événement peut être périodique, comme par exemple la distribution mensuelle de tous les AIP versés dans l'OAIS par un Producteur donné, ou alors cet événement peut être unique comme par exemple le versement d'un AIP donné. Le Protocole de commande spécifiera également d'autres informations nécessaires, telles que l'événement déclencheur de nouvelles Opérations de diffusion de données et les critères de sélection des fonds de l'OAIS à inclure dans chaque nouvelle Opération de diffusion de données.

Le Protocole de commande peut ne pas être un document formel. En général, un OAIS aura une politique de tarification globale et disposera d'une base actualisée d'information des adresses électroniques et physiques de ses utilisateurs. Dans ce cas, l'établissement du Protocole de commande pourra se limiter à remplir un formulaire sur Internet pour spécifier les AIP intéressants.

### 3 RESPONSABILITES D'UN OAIS

La sous-section 3.1 identifie les responsabilités minimales qu'un OAIS doit assumer. La sous-section 3.2 donne quelques exemples de mécanismes pour assumer ces responsabilités, bien que ces exemples ne soient pas tous applicables à tous les OAIS.

#### 3.1 RESPONSABILITES OBLIGATOIRES

Cette sous-section décrit les responsabilités obligatoires qu'une organisation doit assumer afin de mettre en œuvre une Archive OAIS.

L'OAIS doit :

- Négocier avec les Producteurs d'information et accepter les informations appropriées de leur part.
- Acquérir une maîtrise suffisante de l'information fournie, au niveau requis pour pouvoir en garantir la Pérennisation.
- Déterminer, soit par lui-même, soit en collaboration avec d'autres, quelles communautés doivent constituer la Communauté d'utilisateurs cible en mesure de comprendre l'information fournie.
- Assurer que l'information à conserver est **Immédiatement compréhensible** pour la Communauté d'utilisateurs cible. En d'autres termes, la communauté doit être en mesure de comprendre les informations sans l'assistance des experts ayant produit ces informations.
- Appliquer une stratégie et des procédures documentées garantissant la conservation de l'information contre tout imprévu dans les limites du raisonnable, et permettant la diffusion d'une information, copie authentifiée de l'original ou permettant de remonter à l'original.
- Rendre l'information conservée disponible pour la Communauté d'utilisateurs cible.

#### 3.2 EXEMPLES DE MECANISMES POUR ASSUMER LES RESPONSABILITES

Cette sous-section fournit des exemples de mécanismes qui permettent d'assumer les responsabilités décrites en 3.1. Ces mécanismes ne sont pas tous applicables à tous les OAIS.

##### 3.2.1 NEGOCIER ET ACCEPTER LES INFORMATIONS

Une organisation qui met en œuvre un OAIS doit avoir établi des critères qui aident à déterminer le type d'information qu'elle est prête à accepter ou qu'on lui impose. Ces critères peuvent inclure, entre autres, le sujet, la source d'information, le caractère unique ou original, ainsi que la nature des techniques utilisées pour représenter l'information (par exemple, le support physique, le support numérique, le format). En général, l'information peut être versée

sous une grande variété de formes plus ou moins usuelles telles que des livres, des documents, des cartes, des ensembles de données, des roches lunaires, en utilisant différents moyens de communication tels que les réseaux, le courrier et les livraisons particulières.

L'OAIS négocie avec le Producteur pour s'assurer que les Contenus d'information et PDI associés qu'il va recevoir correspondent bien à sa mission et aux besoins de la Communauté d'utilisateurs cible. Ceci peut nécessiter plusieurs allers-retours avant de s'accorder sur l'information à verser et la forme sous laquelle l'OAIS peut l'accepter. Par exemple, on peut convenir de numériser des supports audio ou vidéo, ou de scanner du texte.

L'OAIS doit extraire - ou obtenir par d'autres moyens - une Information de description suffisante pour que la Communauté d'utilisateurs cible puisse trouver le Contenu d'information qui l'intéresse. L'OAIS doit également garantir que l'information respecte tous ses standards internes.

### **3.2.2 ACQUERIR UNE MAITRISE SUFFISANTE POUR GARANTIR LA CONSERVATION**

Il est important qu'un OAIS reconnaisse la différence qui peut exister entre, d'une part, la propriété physique ou la détention de Contenus d'information et, d'autre part, la propriété intellectuelle de cette information. S'il a créé l'information et qu'il est le propriétaire légal du Contenu d'information, l'OAIS peut *de facto* appliquer les règles requises pour conserver l'information et la rendre disponible. Lorsqu'il acquiert le Contenu d'information d'un autre Producteur ou d'une autre entité, l'OAIS doit s'assurer qu'il existe un protocole de transfert, valide au regard de la loi, qui transfère les droits de propriété intellectuelle à l'OAIS ou spécifie clairement les droits accordés à l'OAIS et toutes les restrictions imposées par les ayants droit. L'OAIS doit garantir que ses actions ultérieures de conservation et de mise à disposition de l'information seront conformes à ces droits et restrictions. Lorsque l'OAIS n'acquiert pas les droits de propriété intellectuelle, le protocole doit spécifier en quoi le(s) titulaire(s) des droits sera/seront impliqué(s) dans la conservation, le Management ou la mise à disposition de l'information. En général, il est préférable pour l'OAIS de négocier un protocole qui spécifie les exigences du(des) titulaire(s) des droits et autorise l'OAIS à agir dans le respect de ces exigences sans intervention directe du titulaire pour chaque cas particulier.

L'OAIS doit avoir une maîtrise suffisante du Contenu d'information et de l'Information de pérennisation afin d'être en mesure de les pérenniser. Cette question ne se pose pas avec l'Information d'empaquetage des AIP puisque, par définition, elle est sous le contrôle interne de l'OAIS. Vu la nature essentiellement numérique du Contenu d'information et de l'Information de pérennisation, les problèmes liés à leur maîtrise peuvent être classés dans les trois catégories suivantes :

- gestion du copyright, de la propriété intellectuelle ou d'autres restrictions légales en vigueur,
- autorisation de modification de l'Information de représentation,
- protocoles avec des organismes externes.



**Gestion du copyright, de la propriété intellectuelle ou d'autres restrictions légales en vigueur :** une Archive respecte toutes les restrictions légales en vigueur. Ces questions se posent lorsqu'un OAIS agit en tant que gardien de l'information. Un OAIS doit connaître le copyright et les lois en vigueur avant d'accepter des informations soumises à copyright. Il peut établir des directives pour le versement d'informations et, le cas échéant, des règles pour la diffusion et la duplication de l'information. Le présent document n'a pas vocation à développer la question de la législation nationale et internationale du copyright.

**Autorisation de modification de l'Information de représentation :** bien que l'Information d'intégrité au sein de l'Information de pérennisation d'un AIP garantisse que les bits associés à un Contenu d'information n'ont pas été altérés, il arrive un moment où ces bits ne sont plus sous une forme qui convient aux Utilisateurs. En effet, si ces bits sont entièrement documentés sous forme de documents papier, l'information n'est pas à proprement parler perdue mais, dans la pratique, elle est devenue inaccessible. L'OAIS doit être autorisé à migrer le Contenu d'information vers de nouvelles formes de représentation. S'il agit en simple gardien de l'information, il peut avoir besoin d'une autorisation supplémentaire pour effectuer de telles migrations. Si l'information est soumise à copyright, l'OAIS doit avoir au préalable négocié l'autorisation d'effectuer les migrations nécessaires pour satisfaire aux objectifs de conservation. Il peut avoir recours à des experts du domaine, extérieurs à l'OAIS, pour garantir que l'information ne sera pas perdue. Dans l'idéal, lorsque cette situation se présente, les AIP originaux (entièrement décrits) et les nouveaux AIP seront conservés. Les problèmes liés à la Migration numérique sont analysés plus en profondeur à la section 5.1.

**Protocoles avec des organismes externes :** un OAIS peut établir une série de protocoles avec d'autres organismes qui l'aideront à atteindre ses objectifs de conservation. Par exemple, il peut établir un protocole avec un autre OAIS afin de ne pas avoir à conserver tous les Objets-information de représentation communs qui peuvent être associés à ses propres Objets-contenu d'information. Les protocoles avec d'autres organismes doivent faire l'objet d'un suivi afin de s'assurer qu'ils sont respectés et qu'ils restent utiles.

### 3.2.3 DETERMINER UNE COMMUNAUTE D'UTILISATEURS CIBLE

Le versement ou le projet de versement de Contenus d'information avec leur PDI requiert une identification des Utilisateurs ou de la Communauté d'utilisateurs cible potentiels de cette information. Ceci est nécessaire afin de déterminer si l'information, telle qu'elle est représentée, sera compréhensible par cette communauté. Par exemple, une Archive peut décider que certains Contenus d'information doivent être compréhensibles par le grand public qui devient de ce fait la Communauté d'utilisateurs cible.

Pour certaines informations scientifiques, la Communauté d'utilisateurs cible pourrait être définie par les individus ayant le niveau de première année d'études supérieures dans la discipline scientifique concernée. Ceci est une question complexe, car il est difficile de déterminer le degré de spécialisation de la terminologie scientifique qui est effectivement acceptable. Les Producteurs de telles informations spécialisées utilisent souvent une terminologie connue des seuls experts du domaine. Il est donc particulièrement crucial de

définir clairement la Communauté d'utilisateurs cible pour ces informations et de faire le nécessaire pour que cette communauté puisse les comprendre.

Il convient également de prendre en considération l'évolution potentielle de la définition de la Communauté d'utilisateurs cible. Il peut se révéler un jour nécessaire de donner un plus large accès à une information destinée primitivement à une communauté restreinte. Par exemple, une information destinée à l'origine à une communauté scientifique spécifique peut devoir être diffusée au grand public. Ceci implique qu'il faudra vraisemblablement ajouter des informations explicatives à l'Information de représentation et à l'Information de pérennisation. Du reste, ces informations sont sans doute plus difficiles à obtenir avec le temps. Lorsqu'une information est proposée la première fois à la pérennisation, le choix dès le départ d'une Communauté d'utilisateurs cible plus large (par exemple, le grand public) minimise la difficulté et augmente la probabilité que cette information soit comprise par tous dans la communauté d'origine.

### **3.2.4 GARANTIR QUE LES INFORMATIONS SONT IMMEDIATEMENT COMPREHENSIBLES**

Le niveau de compréhension de l'information véhiculée par un Contenu d'information et son PDI associé par une Communauté d'utilisateurs cible, est généralement assez subjectif. Néanmoins, il est essentiel qu'une Archive le détermine, afin d'optimiser la conservation de l'information. Le Contenu d'information numérique et son PDI doivent être accompagnés des Informations de représentation appropriées pour être Immédiatement compréhensibles par la Communauté d'utilisateurs cible. Ces Objets-information de représentation sont généralement diversifiés (voir l'analyse en section 4.2).

Par exemple, considérons le Contenu d'information issu d'un ensemble numérique de mesures de précipitations, de température, de pression atmosphérique, de vitesse du vent et d'autres paramètres recueillis partout dans le monde au cours d'une année. Ce Contenu d'information est très diversifié. Sa forme n'est généralement pas adaptée au feuilletage ou à la lecture directe par l'homme, mais plutôt à la recherche et à la manipulation par des logiciels d'application. Un tel contenu risque de n'être compréhensible que par les Producteurs d'origine, à moins que la signification des différents champs et leurs interrelations ainsi que les relations entre les valeurs et les instruments de mesure ne soient dûment documentées. Dans de tels domaines spécialisés, un effort supplémentaire est nécessaire pour garantir que le Contenu d'information et l'Information de pérennisation sont compréhensibles par la Communauté d'utilisateurs cible. Si l'Archive n'a pas le niveau d'expertise requis, elle peut avoir recours à des représentants de la communauté extérieure qui analysent l'information pour garantir sa compréhension à long terme. Dans le cas contraire, une partie de l'information risque de n'être comprise que par une poignée de spécialistes et donc d'être perdue si ceux-ci ne sont plus disponibles.

Même si, à un moment donné, il a été acquis qu'un ensemble d'informations était compréhensible par une certaine Communauté d'utilisateurs cible, il peut s'avérer avec le temps que la Base de connaissance de cette communauté évolue à un point tel que certains aspects importants de l'information peuvent ne plus être facilement compréhensibles. A ce

stade, il est nécessaire qu'un OAIS enrichisse l'Information de représentation afin qu'elle redevienne facilement compréhensible à la Communauté d'utilisateurs cible.

Un autre exemple : le Contenu d'information d'un manuscrit est écrit en anglais et son contenu peut donc être généralement compris par un large public. Toutefois, si le motif de sa création n'est pas clairement documenté, une grande partie de sa signification peut être perdue. Cette information sur « le motif » fait partie du Contexte et doit figurer dans l'Information de pérennisation.

Un accès performant au Contenu d'information numérique implique l'usage de logiciels. Toutefois, assurer la maintenance de logiciels spécifiques pour accéder aux Contenus d'information sur le Long terme reste coûteux en raison du caractère restreint de leur domaine d'application. Lorsque de tels logiciels sont indispensables à la pérennisation et à la compréhension des informations, sachant que des changements mineurs de l'environnement matériel et logiciel peuvent suffire à bloquer leur fonctionnement, le risque de perte d'information est grand. Il se peut que ce risque ne soit pas identifié, à moins qu'il existe un programme de test et de validation solide et permanent.

### **3.2.5 APPLIQUER LES METHODES ET LES PROCEDURES DE PRESERVATION MISES EN PLACE**

Il est essentiel qu'un OAIS se dote de méthodes et procédures documentées pour la conservation de ses AIP et qu'il applique ces procédures. La pertinence des méthodes et des procédures dépend, au minimum, de la nature des AIP et du rôle de sauvegarde que les Archives peuvent assurer entre elles. Par exemple, les migrations qui modifient un Contenu d'information ou un PDI doivent être soigneusement contrôlées et le PDI doit être mis à jour. Cette exigence de précision, qui est une garantie contre les erreurs de traitement, nécessite que des procédures et des méthodes rigoureuses soient définies et appliquées.

Il est nécessaire de transmettre aux communautés de Producteurs et d'Utilisateurs les standards, méthodes et procédures de versement et de diffusion destinés à la pérennisation. Un suivi de la Communauté d'utilisateurs cible doit permettre de s'assurer qu'elle est toujours en mesure de comprendre le Contenu d'information. Il peut arriver que la Communauté d'utilisateurs cible voie sa connaissance de la terminologie diminuer ou que cette communauté soit élargie à d'autres membres de formation différente. C'est pourquoi, une révision périodique avec des représentants de la Communauté peut faciliter le suivi.

Il est essentiel d'établir un plan d'utilisation des technologies à long terme et de le mettre à jour au fur et à mesure des évolutions technologiques, afin d'éviter les pièges d'une maintenance système très onéreuse, de changements de système dans l'urgence et de coûteuses transformations de représentation des données.

### **3.2.6 RENDRE L'INFORMATION DISPONIBLE**

Par définition, un OAIS rend ses AIP visibles et disponibles à ses Communautés d'utilisateurs cibles. Il peut présenter différentes facettes de ses fonds, grâce à divers Outils de recherche capables de balayer les collections d'AIP. Certains AIP n'ont d'existence qu'en tant que sorties d'algorithmes fonctionnant à partir d'autres AIP. Ils apparaissent sous forme de DIP qui, pour la diffusion, doivent inclure une description de la manière dont ils ont été dérivés d'autres AIP. Les attentes des Utilisateurs d'un OAIS vis à vis des services d'accès peuvent varier considérablement d'une Archive à l'autre ainsi que dans le temps, en fonction des évolutions technologiques. Les demandes visant un accès plus efficace doivent être évaluées en fonction des exigences de conservation en tenant compte des contraintes de ressources.

Certains AIP peuvent faire l'objet de restrictions d'Accès et donc n'être diffusés qu'auprès des Utilisateurs autorisés. L'OAIS doit rendre publique sa politique d'accès et les restrictions induites, afin de protéger les droits de chacune des parties.

La distribution des DIP peut utiliser les différents moyens de communication existants, y compris les réseaux et les supports physiques.

## 4 MODELES DETAILLES

Le but de cette section est de donner une vision plus détaillée du modèle des entités fonctionnelles de l'OAIS et des informations manipulées par l'OAIS. Elle constitue une aide aux concepteurs de futurs systèmes OAIS et fournit un jeu plus précis de termes et de concepts pour aborder les systèmes actuels.

### 4.1 MODELE FONCTIONNEL

Le modèle OAIS présenté dans le schéma 2-1 est éclaté dans le schéma 4-1 en six entités fonctionnelles avec leurs interfaces correspondantes. Seuls les principaux flux d'information sont représentés. Les lignes reliant les entités indiquent les voies de communication par lesquelles les informations circulent dans les deux sens. Les lignes arrivant aux **Entités** « **Administration** » et « **Planification de la pérennisation** » sont tracées en tirets dans le seul but de rendre le diagramme plus lisible.

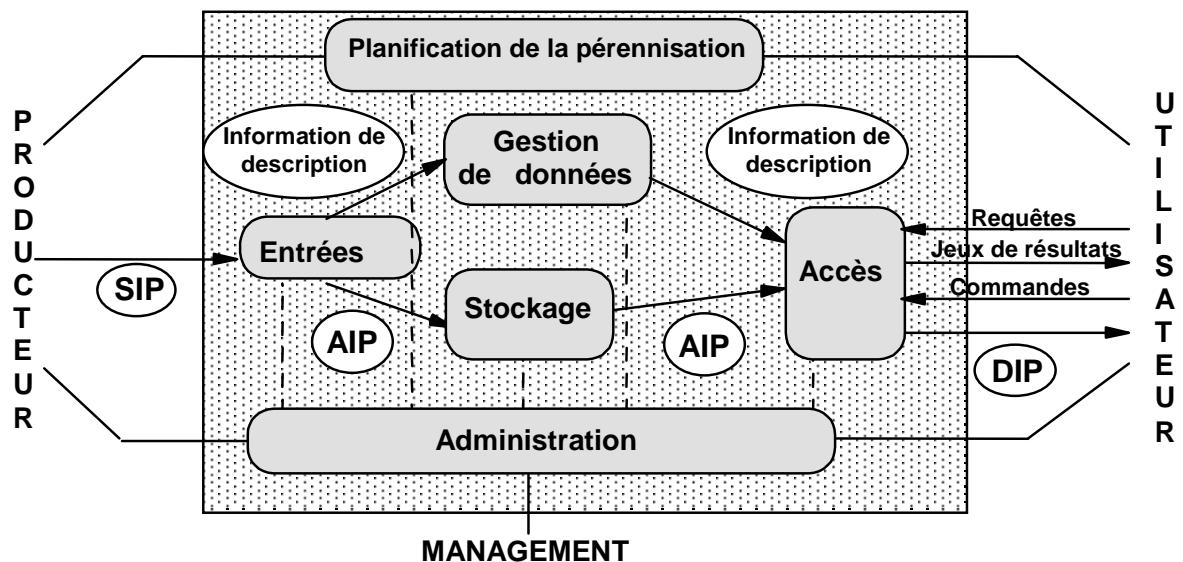


Schéma 4-1 : Entités fonctionnelles OAIS

Le rôle joué par chacune des entités du schéma 4-1 est décrit brièvement comme suit :

**Entité « Entrées »** : cette entité assure les fonctions et services relatifs à l'acceptation des Paquets d'informations à verser (SIP) provenant des Producteurs (ou d'éléments internes sous le contrôle de l'Entité « Administration »), et à la préparation de leur contenu en vue du stockage et de la gestion des données au sein de l'Archive. Les fonctions de l'Entité « Entrées » comprennent : la réception des SIP, le contrôle d'Assurance Qualité sur ces SIP, la génération d'un Paquet d'informations archivé (AIP) conforme aux normes de documentation et de formatage des données de l'Archive, l'extraction de l'Information de description des

AIP pour l'inclure dans la base de données de l'Archive et la coordination des mises à jour à effectuer au niveau des Entités « Stockage » et « Gestion de données ».

**Entité « Stockage »** : cette entité assure les fonctions et services relatifs au stockage, à la maintenance et à la récupération des AIP. Les fonctions de l'Entité « Stockage » comprennent notamment : la réception des AIP en provenance de l'Entité « Entrées » et leur insertion dans l'espace de stockage permanent, la gestion de la hiérarchie du stockage, le renouvellement des supports sur lesquels les fonds de l'Archive sont stockés, les contrôles d'erreurs spécifiques et de routine, la fourniture des moyens de sauvegarde et la mise en œuvre des plans de reprise d'activité, et la transmission des AIP à l'Entité « Accès » en réponse aux commandes.

**Entité « Gestion de données »** : cette entité assure les fonctions et services relatifs à l'enrichissement, la conservation et l'accès à l'Information de description (qui identifie et documente les fonds de l'Archive) et aux données administratives utilisées pour gérer l'Archive. Les fonctions de l'Entité « Gestion de données » comprennent : l'administration des fonctions de la base de données de l'Archive (conserver et tenir à jour les schémas, les définitions de vues et l'intégrité référentielle), les mises à jour de la base de données (chargement de nouvelles Informations de description ou de nouvelles données administratives de l'Archive), la recherche d'éléments de l'Entité « Gestion de données » pour générer des jeux de résultats, et la production de rapports à partir de ces résultats.

**Entité « Administration »** : cette entité assure les fonctions et services relatifs à l'exploitation d'ensemble du système d'archivage. Les fonctions de l'Entité « Administration » incluent : la proposition et la négociation des Protocoles de versement avec les Producteurs, la vérification des versements pour s'assurer de leur conformité aux normes d'archivage ainsi que la gestion de la configuration du matériel et des logiciels du système. Elle fournit aussi les moyens techniques pour contrôler et améliorer l'exploitation de l'Archive, ainsi que pour inventorier, rendre compte et migrer/mettre à jour les contenus de l'Archive. Elle est également responsable de l'établissement et du maintien des normes et règles applicables à l'Archive, de l'aide à l'Utilisateur et de l'activation des requêtes enregistrées.

**Entité « Planification de la pérennisation »** : cette entité assure les fonctions et services relatifs à la surveillance de l'environnement de l'OAIS et à la production de recommandations visant à ce que les informations stockées dans l'OAIS restent accessibles sur le long terme à la Communauté d'utilisateurs cible, même si l'environnement informatique d'origine devient obsolète. Les fonctions de l'Entité « Planification de la pérennisation » incluent l'évaluation du contenu de l'Archive et la recommandation périodique de mises à jour de l'information archivée pour migrer les fonds courants, le développement de recommandations dans le domaine des normes et règles d'archivage, ainsi que la surveillance des évolutions à la fois de l'environnement technologique et des exigences de service de la Communauté d'utilisateurs cible, et enfin de sa Base de connaissance. Cette entité conçoit aussi des modèles de Paquets d'informations et accompagne la conception et la validation de ces modèles appliqués à des SIP ou des AIP, pour les adapter à des versements spécifiques. L'Entité « Planification de la pérennisation » développe aussi des plans détaillés de migration, des prototypes de logiciels et des plans de test pour répondre aux objectifs de migration de l'Entité « Administration ».

**Entité « Accès »** : cette entité assure les fonctions et services qui aident l'Utilisateur à déterminer si une information existe ou non dans un OAIS, à trouver sa description, son emplacement si elle est disponible, et à demander et recevoir des produits d'information. Les fonctions de l'Entité « Accès » incluent : la communication avec les Utilisateurs pour recevoir leurs demandes, les contrôles d'accès à l'information bénéficiant d'une protection particulière, la coordination du traitement des demandes jusqu'à leur exécution finale, la génération des réponses (Paquets d'informations diffusés, Jeux de résultats, rapports) et leur transmission aux Utilisateurs.

En plus des entités décrites ci-dessus, différents **Services de base** sont censés être disponibles. On considère que ces services constituent une autre entité fonctionnelle dans ce modèle. Mais cette entité est tellement omniprésente que, par souci de clarté, elle n'est pas représentée dans le schéma 4-1.

#### 4.1.1 DESCRIPTION DÉTAILLÉE DES ENTITÉS FONCTIONNELLES

Dans les sous-sections suivantes, les flux spécifiques d'information entre les entités sont indiqués *en italique* lors de leur première apparition dans le texte. La description détaillée des fonctions dans les sous-sections est accompagnée de diagrammes (schémas 4-2 à 4-7) qui décrivent seulement les principaux flux de données dans les entités et entre elles. Par souci de clarté, les flux de données de moindre importance, tels que les avis d'acquiescement, ont été omis.

##### 4.1.1.1 Services de base

Les applications informatiques modernes, distribuées, sous-entendent un certain nombre de services tels que la communication entre processus, les services de nommage, l'allocation d'espace de stockage temporaire, le traitement d'exceptions, la sécurité, les sauvegardes et la gestion des répertoires. D'excellents travaux ont déjà été réalisés dans le domaine des modèles de référence dans un environnement de systèmes ouverts. Les services décrits dans cette sous-section sont basés en partie sur les services décrits dans le Modèle de référence IEEE POSIX OSE (référence [5]).

Les **services du système d'exploitation** fournissent les services fondamentaux nécessaires pour exploiter et administrer la plate-forme d'application et fournir une interface entre le logiciel applicatif et la plate-forme. Au nombre de ces services :

- Le noyau assure les services de bas niveau nécessaires pour créer et gérer les processus, exécuter les programmes, définir et communiquer des signaux, définir et prendre en charge les opérations de l'horloge système, gérer les fichiers et répertoires et piloter le traitement des entrées/sorties à destination et en provenance de l'environnement extérieur.
- Les commandes et les utilitaires comprennent des mécanismes pour les tâches suivantes au niveau de l'opérateur : comparaison, impression et visualisation du contenu de fichiers, édition de fichiers, recherche de chaînes de bits, évaluation

d'expressions, enregistrement de messages, déplacement de fichiers d'un répertoire à l'autre, tri de données, exécution de scripts de commandes, et accès à l'information d'environnement.

- L'extension « temps réel » comprend les interfaces des applications et du système d'exploitation nécessaires aux domaines applicatifs qui requièrent une exécution, un traitement et une réactivité déterministes. L'extension définit l'interface entre les applications et les services de base du système en matière d'entrées/sorties, d'accès au système de gestion de fichiers et de gestion des tâches.
- La gestion du système inclut la possibilité de définir et de gérer l'affectation des ressources utilisateur et l'accès à ces ressources (c'est-à-dire, quelles sont les ressources gérées et quelles sont les classes d'accès définies), la gestion de la configuration et du fonctionnement des périphériques, les systèmes de gestion des fichiers, les processus d'administration (comptabilisation des tâches), les files d'attente, les profils de machine/plate-forme, l'autorisation d'utiliser les ressources, et la sauvegarde du système.
- Les services de sécurité du système d'exploitation définissent le contrôle d'accès aux données système, aux fonctions, au matériel et aux ressources logicielles par les utilisateurs et les processus utilisateurs.

**Les services réseau** fournissent les moyens et les mécanismes pour les applications distribuées qui exigent l'accès aux données et l'interopérabilité des applications dans des environnements hétérogènes en réseau. Au nombre de ces services :

- La communication de données comprend l'API (interface programmatique) et les spécifications du protocole pour une transmission de données de bout en bout, fiable et transparente, sur les réseaux de communication.
- L'accès transparent à des fichiers disponibles, situés n'importe où dans un environnement réseau hétérogène.
- La prise en charge d'ordinateurs (ou micro-ordinateurs) personnels pour l'interopérabilité avec des systèmes basés sur d'autres systèmes d'exploitation, notamment ceux des micro-ordinateurs, qui risquent de ne pas être formellement spécifiés dans une norme nationale ou internationale.
- Les services de protocole d'appel à distance comportent des spécifications pour étendre l'appel de procédure local à un environnement distribué.
- Les services de sécurité réseau couvrent les contrôles de l'accès, de l'authentification, de la confidentialité, de l'intégrité, et de la non-répudiation, ainsi que la gestion des communications entre expéditeurs et destinataires de l'information dans un réseau.

**Les services de sécurité** fournissent des moyens et des mécanismes de protection de l'information sensible, et des traitements dans le système d'information. La détermination du



niveau de protection approprié se base sur la valeur des informations pour les utilisateurs finaux de l'application et sur la perception des risques. Au nombre de ces services :

- Le service d'identification/authentification confirme l'identité des personnes qui demandent à utiliser les ressources du système d'information. L'authentification peut également s'appliquer aux fournisseurs de données. Le service d'authentification peut intervenir au début ou au cours d'une session.
- Le service de contrôle d'accès empêche l'usage non autorisé des ressources du système d'information. Ce service empêche également que des ressources soient utilisées de manière non autorisée. Il peut être appliqué à différents aspects de l'accès à une ressource (par exemple accès aux communications avec la ressource, accès en lecture, droit d'écriture ou de suppression d'une ressource d'information/de données, exécution d'une ressource de traitement) ou à tous les accès à une ressource.
- Le service d'intégrité des données s'assure que les données ne sont pas modifiées ou détruites sans autorisation. Ce service s'applique aux données figurant dans des espaces permanents de stockage et aux données présentes dans des messages de communication.
- Le service de confidentialité des données s'assure que les données ne sont pas mises à la disposition ou communiquées à des personnes ou des traitements informatiques non autorisés. Ce service est appliqué aux périphériques qui permettent une interaction entre l'homme et le système d'information. En outre, ce service rend impossible tout espionnage des modalités d'utilisation des ressources de communication.
- Le service de non-répudiation s'assure que les entités qui s'engagent dans un échange d'information ne peuvent pas contester leur implication. Ce service peut revêtir l'une des deux formes suivantes ou les deux : premièrement, le destinataire des données reçoit une preuve de l'origine des données, ce qui prévient toute tentative de la part de l'expéditeur de nier avoir transmis les données ou leur contenu ; deuxièmement, l'expéditeur des données reçoit une preuve de la livraison des données, ce qui prévient toute tentative ultérieure de la part du destinataire de nier avoir reçu les données ou leur contenu.

#### **4.1.1.2 Entité « Entrées »**

Les fonctions de l'Entité « Entrées » sont décrites dans le schéma 4-2.

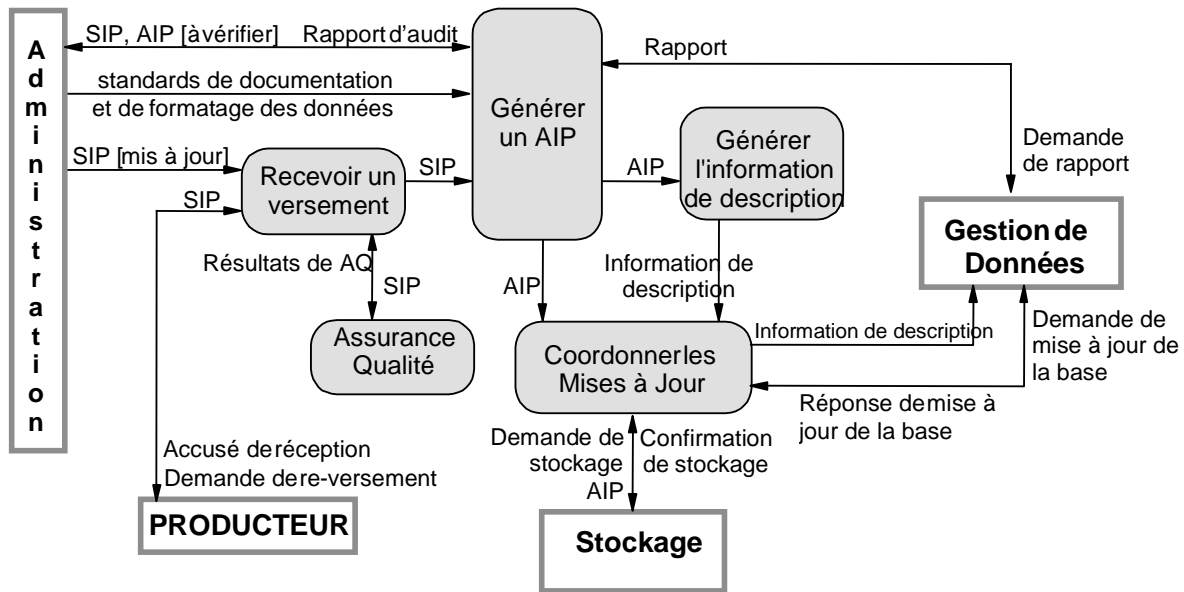


Schéma 4-2 : Fonctions de l'Entité « Entrées »

La fonction « **recevoir un versement** » fournit la capacité de stockage ou les périphériques appropriés pour recevoir un SIP venant du Producteur (ou de l'Entité « Administration »). Des SIP numériques peuvent être livrés par transfert électronique (par exemple, FTP), chargés à partir de supports versés à l'Archive, ou directement installés (par exemple, CD-ROM) sur le système de fichiers de l'Archive. Des SIP non numériques sont susceptibles d'être livrés selon des procédures d'envoi conventionnelles. Cette fonction peut constituer un transfert légal de propriété du Contenu d'information du SIP, et peut nécessiter d'effectuer des contrôles d'accès spéciaux sur les contenus. Elle fournit un *accusé de réception* d'un SIP vers le Producteur, accusé qui peut inclure une *demande de re-versement de SIP* en cas d'erreur intervenue lors du versement initial de ce SIP.

La fonction « **Assurance Qualité** » valide (*Résultats de AQ*) le transfert réussi du SIP vers un espace provisoire de stockage. Pour les versements numériques, ces mécanismes pourraient inclure des Contrôles de Redondance Cycliques (CRC) ou des checksums associés à chaque fichier de données ou encore l'utilisation de journaux de bord pour enregistrer et identifier tout transfert de fichier ou toute erreur de lecture/écriture du support.

La fonction « **générer un AIP** » transforme un ou plusieurs SIP en un ou plusieurs AIP conformes aux *standards de documentation et de formatage des données* de l'Archive. Elle peut impliquer des conversions de format de fichier, des conversions de représentation des données ou une réorganisation du Contenu d'information des SIP. La fonction « générer un AIP » peut émettre des *demandes de rapport* à l'intention de l'Entité « Gestion de Données » pour obtenir l'information dont elle a besoin pour produire l'Information de description qui complète l'AIP. Cette fonction envoie des *SIP ou des AIP à vérifier* à la fonction « auditer le versement » de l'Entité « Administration », et reçoit en retour un *rapport d'audit*.

La fonction « **générer l'Information de description** » extrait l'Information de description des *AIP* et recueille l'*Information de description* d'autres sources pour les fournir à la fonction « coordonner les mises à jour » et, finalement, à l'Entité « Gestion de données ». Cela inclut les métadonnées d'aide à la recherche et à la récupération d'*AIP* (par exemple, qui, quoi, quand, où, pourquoi ?) et pourrait également comprendre des produits spécifiques de feuilletage (vignettes, images) destinés aux Outils de recherche.

La fonction « **coordonner les mises à jour** » est responsable du transfert des *AIP* vers l'Entité « Stockage » et du transfert de l'*Information de description* vers l'entité « Gestion de données ». La transmission des *AIP* comprend une *demande de stockage* qui peut être faite sous forme électronique, physique, ou virtuelle (c'est-à-dire les données restant en place). Une fois le transfert terminé et vérifié, l'Entité « Stockage » renvoie une *confirmation de stockage* indiquant (ou vérifiant) l'information d'identification du stockage de l'*AIP*. La fonction « coordonner les mises à jour » incorpore aussi l'information d'identification du stockage dans l'Information de description de l'*AIP* et la transmet à l'Entité « Gestion de données », en même temps qu'une *demande de mise à jour de la base de données*. En retour, l'Entité « Gestion de données » fournit une *réponse de mise à jour de la base de données* indiquant l'état de la mise à jour. Ces mises à jour peuvent être effectuées sans transfert correspondant vers l'Entité « Stockage » lorsque le SIP contient l'Information de description d'un *AIP* y figurant déjà.

#### 4.1.1.3 Entité « Stockage »

Les fonctions de l'Entité « Stockage » sont décrites dans le schéma 4-3.

La fonction « **recevoir les Données** » reçoit une *demande de stockage* et un *AIP* en provenance de l'Entité « Entrées » et envoie l'*AIP* vers la zone de stockage permanent de l'Archive. La demande de transfert peut avoir à indiquer la fréquence prévue d'utilisation des Objets-données composant l'*AIP*, afin de préparer les moyens ou supports de stockage adéquats pour l'*AIP* concerné. Cette fonction sélectionne le type de support, prépare les périphériques ou les volumes et effectue le transfert physique des données vers les volumes de l'Entité « Stockage ». Une fois le transfert effectué, cette fonction envoie un message de *confirmation du stockage* à l'Entité « Entrées » incluant l'identification du stockage des *AIP*.

La fonction « **gérer la hiérarchie de stockage** » inscrit, via *des commandes*, les contenus des *AIP* sur les supports appropriés en fonction de la *politique de gestion* en matière de stockage, en fonction des statistiques d'exploitation, ou des directives données par l'Entité « Entrées » via la demande de stockage. Elle respecte également tous les niveaux spécifiques de service exigés pour l'*AIP* ainsi que toutes les mesures de sécurité requises et s'assure du niveau de protection approprié de l'*AIP*. Cela comprend le stockage en ligne, en différé ou en léger différé, le débit exigé, le taux d'erreurs de bits maximal autorisé, ou les procédures de traitement ou de sauvegarde spécifiques. Elle surveille les *comptes-rendus d'erreurs* pour s'assurer que les *AIP* n'ont pas été altérés durant le transfert. Cette fonction fournit aussi à l'Entité « Administration » des *statistiques d'exploitation* qui font l'inventaire des supports à disposition, de la capacité de stockage disponible dans les différentes couches de la hiérarchie de stockage, et des statistiques d'utilisation.

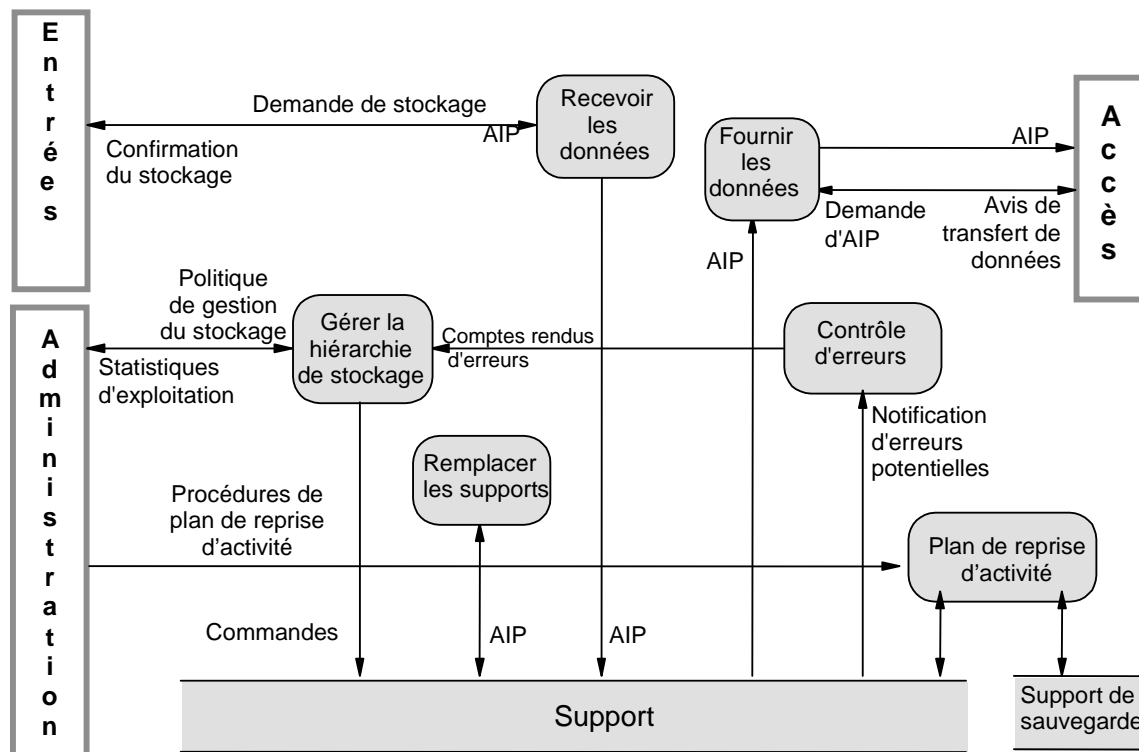


Schéma 4-3 : Fonctions de l'Entité « Stockage »

La fonction « **remplacer les supports** » offre la possibilité de reproduire les *AIP* dans le temps. Au sein de cette fonction, le Contenu d'information et l'Information de pérennisation (PDI) ne doivent pas subir de modification. Néanmoins, les données constituant l'Information d'emballage peuvent être changées, pour autant qu'elles continuent d'assurer la même fonction et qu'une mise en œuvre simple et directe, sans entraîner de perte d'information, est possible. La stratégie de migration doit consister à choisir un support de stockage en tenant compte des taux d'erreurs réels et attendus caractérisant les différents types de supports, de leurs performances, et de leur coût d'acquisition. Si des attributs liés aux supports (par exemple, tailles des blocs sur la bande, information de volume de CD-ROM) ont été inclus dans le Contenu d'information, il faut trouver un moyen de conserver ces données lors de la migration vers des supports de plus grande capacité avec une architecture de stockage différente. Pour anticiper sur la terminologie de la section 5, cette fonction peut effectuer un "Rafraîchissement de support", une "Duplication", et un "Ré-emballage" qui soient simples et directs. Le Ré-emballage consiste par exemple en la migration vers de nouveaux supports sous un nouveau système d'exploitation et de gestion de fichiers où le Contenu d'information et son PDI sont indépendants des systèmes de gestion de fichiers. Toutefois, un Ré-emballage complexe, de même que toute Transformation, est exécuté sous le contrôle de l'Entité « Administration » par la fonction « mise à jour de l'information archivée » pour s'assurer de la conservation des informations. Voir la section 5 pour la description détaillée des questions liées à la migration.

La fonction « **contrôle d'erreurs** » garantit, avec une probabilité statistiquement acceptable, qu'aucun composant de l'AIP n'a été corrompu lors d'un quelconque transfert interne des données de l'Entité « Stockage ». Cette fonction nécessite que tous les matériels et tous les logiciels au sein de l'Archive fournissent une *notification d'erreurs potentielles*, et que ces erreurs soient redirigées vers des *comptes-rendus* vérifiés par l'équipe de l'Entité « Stockage ». L'Information d'intégrité du PDI assure dans une certaine mesure que le Contenu d'information n'a pas subi de modification lors du déplacement de l'AIP et de l'accès à ce dernier. Une information similaire est nécessaire pour protéger le PDI proprement dit. Un mécanisme standard de suivi et de vérification de la validité de tous les Objets-données au sein de l'Archive peut également être utilisé. Par exemple, des contrôles de redondance cycliques (CRC) pourraient être appliqués à chaque fichier de données. Un niveau supérieur de service, tel que le codage Reed-Solomon qui permet à la fois la détection et la correction d'erreurs, pourrait aussi être assuré. Les procédures de gestion de l'espace de stockage devraient prévoir la vérification par sondage aléatoire de l'intégrité des Objets-données à l'aide des CRC ou autres systèmes de contrôle d'erreurs.

La fonction de « **plan de reprise d'activité** » fournit un mécanisme pour dupliquer les contenus numériques de la collection d'archives et stocker la copie dans une installation physiquement séparée. Cette fonction s'effectue normalement en copiant le contenu archivé sur certains types de support de stockage amovibles (par exemple, DLT, CD...), mais peut aussi être réalisée par transferts du matériel ou via le réseau. Les détails *des procédures de plan de reprise d'activité* » sont spécifiés par l'Entité « Administration ».

La fonction « **fournir les Données** » transmet à l'Entité « Accès » les copies des AIP stockés. Cette fonction reçoit une *demande d'AIP* précisant les AIP demandés, et les livre sur le type de support demandé ou les transfère vers un espace de stockage provisoire. Cette fonction envoie également un *avis de transfert de données* à l'Entité « Accès » à titre de réponse à la demande.

#### 4.1.1.4 Entité « Gestion de données »

Les différentes fonctions de l'Entité « Gestion de données » sont représentées dans le schéma 4-4.

La fonction « **administrer la base de données** » est responsable de l'intégrité de la base de données de l'Entité « Gestion de Données », elle contient à la fois les Informations de description et les informations système. Les Informations de description identifient et décrivent les fonds d'archives tandis que les informations système sont utilisées pour permettre les opérations sur l'Archive. La fonction « administrer la base de données » est chargée de créer tous schémas ou définitions de tables nécessaires à l'exécution des fonctions de l'Entité « Gestion de données » ; elle doit aussi fournir les moyens de créer, gérer et accéder à des vues Utilisateur personnalisées du contenu des données stockées ; elle est aussi responsable de la validation interne (par exemple, intégrité référentielle) du contenu de la base de données. La fonction « administrer la base de données » est mise en œuvre conformément aux *règles* édictées par l'Entité « Administration ».

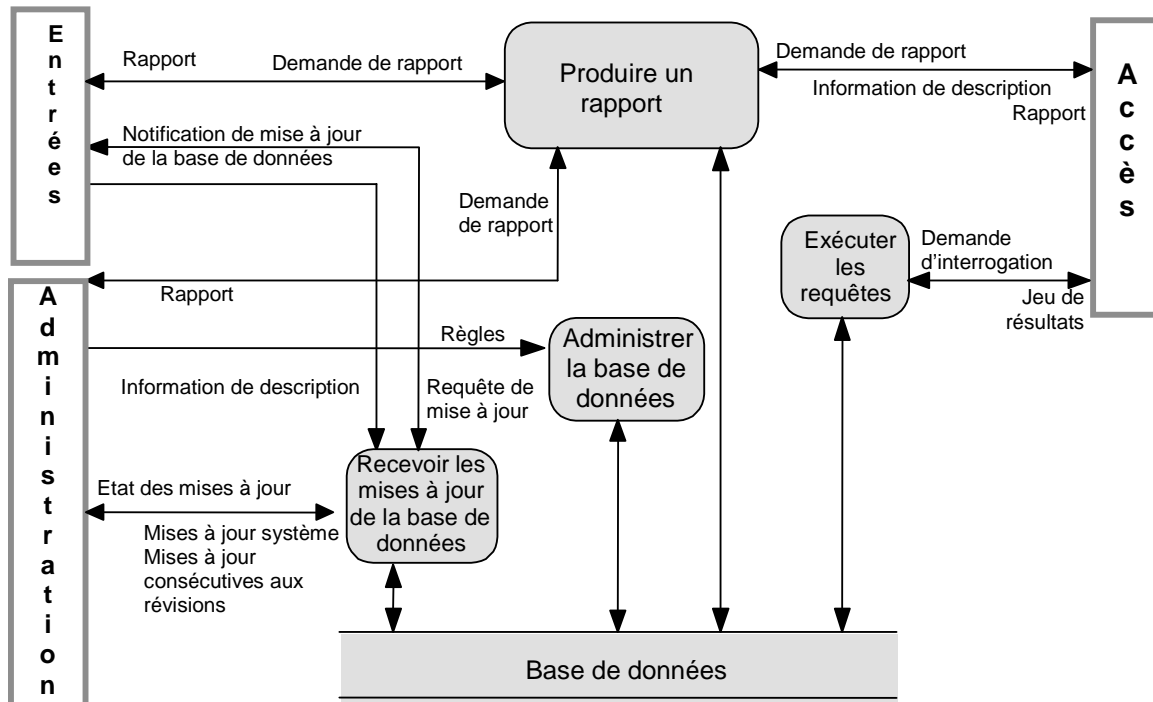


Schéma 4-4 : Fonctions de l'Entité « Gestion de données »

La fonction « **exécuter les requêtes** » reçoit une *demande d'interrogation* en provenance de l'Entité « Accès » et l'exécute pour produire un *Jeu de résultats* transmis au demandeur.

La fonction « **produire un rapport** » reçoit une *demande de rapport* en provenance des Entités « Entrées », « Accès » ou « Administration » et exécute toute requête ou tout autre processus nécessaire à la production du *rapport* destiné au demandeur. Généralement ces rapports incluent des résumés de fonds d'archives par catégories ou des statistiques d'utilisation concernant les accès aux fonds. Cette fonction peut également recevoir de l'Entité « Accès » une demande de rapport et elle fournit aussi l'*Information de description* d'un AIP spécifique.

La fonction « **recevoir les mises à jour de la base de données** » ajoute, modifie ou supprime des informations dans le stockage permanent de l'Entité « Gestion de données ». Les principales sources de mises à jour sont l'Entité « Entrées », qui fournit l'*Information de description* pour les nouveaux AIP, et l'Entité « Administration », qui fournit des *mises à jour système* et des *mises à jour consécutives aux révisions*. Les transactions de l'Entité « Entrées » concernent les Informations de description qui identifient les nouveaux AIP stockés dans l'Archive. Les mises à jour système incluent toute information concernant le système (statistiques d'exploitation, informations sur l'Utilisateur et état de la requête). Les mises à jour consécutives aux révisions sont déclenchées par les révisions et mises à jour périodiques des données d'information (par exemple, noms et adresses des contacts). La fonction « recevoir les mises à jour de la base de données » fournit régulièrement des rapports

à l'Entité « Administration » récapitulant l'état des mises à jour de la base de données et envoi aussi une notification de mise à jour de la base de données à l'Entité « Entrées ».

#### 4.1.1.5 Entité « Administration »

Les fonctions de l'Entité « Administration » sont représentées dans le schéma 4-5.

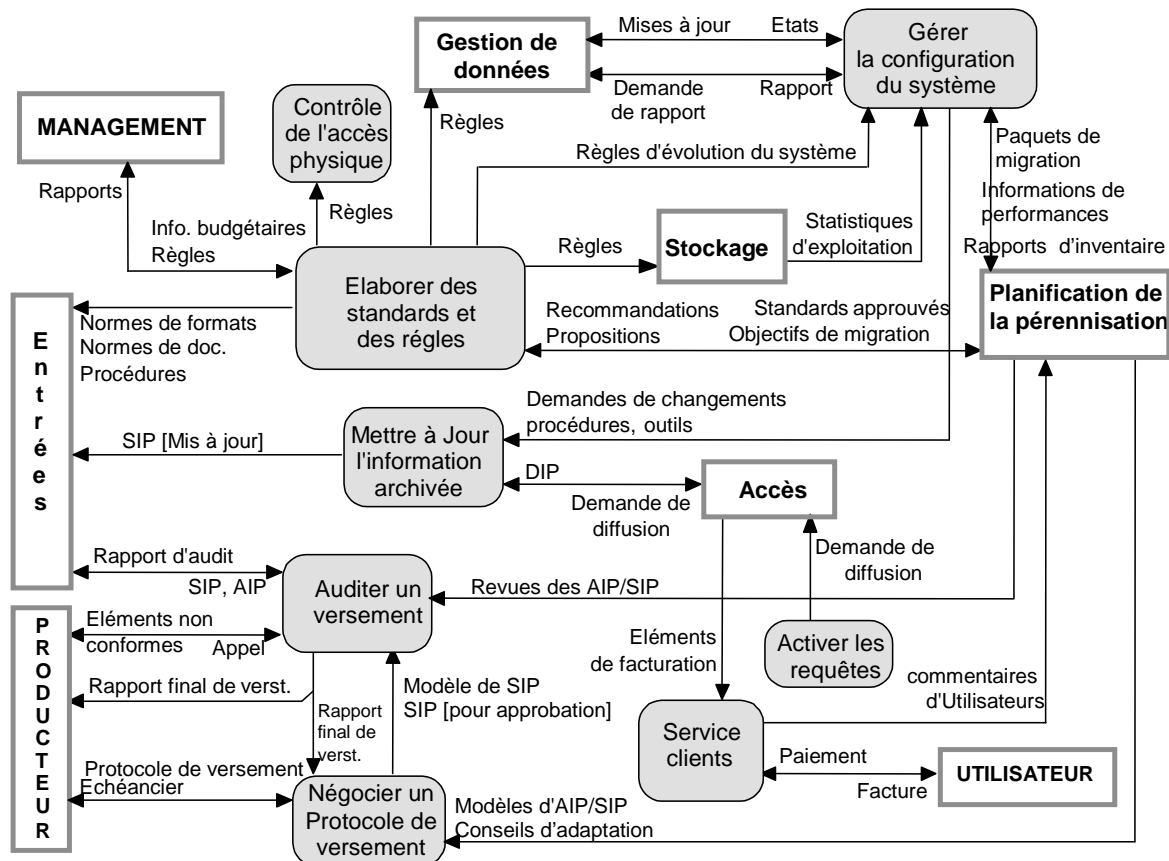


Schéma 4-5 : Fonctions de l'Entité « Administration »

La fonction « **négoier un Protocole de versement** » sollicite l'information qu'il serait souhaitable d'archiver et négocie les *Protocoles de versement* avec les Producteurs. Cette fonction négocie également un *échancier de versement des données* avec le Producteur. Elle tient à jour un calendrier des Opérations de versement attendues, nécessaires pour transférer un ou plusieurs SIP vers l'OAIS, ainsi qu'un état des besoins en ressources pour leur versement. Cette fonction reçoit des *modèles d'AIP/SIP* et des *conseils d'adaptation* de la part de l'Entité « Planification de la pérennisation » et envoie les *projets de SIP* et les *SIP* à la fonction « auditer un versement » en tant que partie prenante du processus d'approbation du versement. Les formats et procédures de versement des données doivent être clairement

documentés selon les règles de versement de l'Archive, et les fournitures doivent être identifiées par le Producteur dans le Protocole de versement.

La fonction « **gérer la configuration du système** » assure la maîtrise technique du système d'archivage pour surveiller en permanence son fonctionnement global et contrôler systématiquement les modifications de sa configuration. Cette fonction veille à l'intégrité et à la traçabilité de la configuration durant toutes les phases du cycle de vie du système. Elle effectue également des contrôles sur l'exploitation du système, ses performances, et son utilisation. Elle envoie à l'Entité « Gestion de données » des *demandes de rapport* sur les informations système et reçoit les *rapports* en retour. Elle reçoit des *statistiques d'exploitation* provenant de l'Entité « Stockage ». Elle synthétise ces rapports et fournit périodiquement à l'Entité « Planification de la pérennisation » des *informations de performances* de l'OAIS ainsi que des *rapports d'inventaire* des fonds. Elle envoie des informations de performances à la fonction « élaborer des standards et des règles ». Elle reçoit des *paquets de migration* de l'Entité « Planification de la pérennisation ». Elle reçoit des *règles d'évolution du système* de la part de la fonction « élaborer des standards et des règles ». A partir de ces entrées elle développe et implémente des plans d'évolution du système. Elle envoie les *demandes de changements*, les *procédures* et les *outils* à la fonction « mettre à jour l'information archivée ».

La fonction « **mettre à jour l'information archivée** » fournit un mécanisme de mise à jour des contenus de l'Archive. Elle reçoit les *demandes de changement*, les *procédures* et les *outils* de la fonction « gérer la configuration du système ». Elle fournit des mises à jour en envoyant une *demande de diffusion* à l'Entité « Accès », en mettant à jour le contenu des *DIP* résultants et en les soumettant à nouveau en tant que *SIP* à l'Entité « Entrées ».

La fonction « **contrôle de l'accès physique** » fournit les mécanismes de restriction ou d'autorisation de l'accès physique (portes, verrous, protections) aux éléments de l'Archive en fonction des règles adoptées par l'Archive.

La fonction « **élaborer des standards et des règles** » a la responsabilité d'établir et de tenir à jour les standards et règles du système de l'Archive. Elle reçoit du Management les *informations budgétaires* et les *règles* telles que la charte, le domaine, les consignes d'utilisation des ressources, et la politique tarifaire de l'OAIS. Elle fournit au Management des *rapports* périodiques. Elle reçoit de l'Entité « Planification de la pérennisation » des *recommandations* pour l'amélioration du système d'archivage et des *propositions* de nouveaux standards concernant les données de l'Archive. Elle reçoit également de la fonction « gérer la configuration du système » des informations de performances et des inventaires de fonds d'archives. A partir de ces entrées, les standards et règles de l'Archive sont élaborés et transmis pour mise en œuvre aux autres fonctions de l'Entité « Administration » et aux autres entités fonctionnelles. Ces standards comprennent les *normes sur les formats et sur la documentation, ainsi que les procédures à suivre pendant l'opération de versement*. Elle fournit les *standards approuvés* et les *objectifs de migration* à l'Entité « Planification de la pérennisation ». Elle conçoit également des *règles en matière de gestion de stockage* (pour la hiérarchie de stockage), y compris des règles en matière de migration pour s'assurer que les formats de stockage de l'Archive ne deviennent pas obsolètes. Elle conçoit aussi les règles



d'administration des fonctions de la base de données ainsi que *les règles de « plan de reprise d'activité »*. Cette fonction définit aussi des règles de sécurité pour protéger les données contenues dans l'Archive, y compris celles qui ont trait au « contrôle de l'accès physique » et aux techniques de contrôle d'erreurs utilisées dans toutes les fonctions de l'Archive.

La fonction « **auditer un versement** » vérifie que les versements (de *SIP* ou d'*AIP*) sont conformes aux spécifications du Protocole de versement. Cette fonction reçoit de l'Entité « Planification de la pérennisation » des comptes-rendus de *revue* d'*AIP* ou de *SIP* et peut également impliquer une commission extérieure (par exemple, révision scientifique et technique). L'audit doit vérifier que la qualité des données satisfait aux exigences de l'Archive et de la commission de revue. Il doit s'assurer que les Informations de représentation et les PDI sont appropriés à une compréhension et une utilisation immédiate du Contenu d'information par la Communauté d'utilisateurs cible. Le fonctionnement des révisions variera selon les règles internes de l'Archive. L'audit peut décider que certaines parties du *SIP* ne sont pas appropriées à une inclusion dans l'Archive et qu'elles doivent faire l'objet d'un nouveau versement ou être exclues. Un *rapport d'audit* est fourni à l'Entité « Entrées ». Une fois l'audit terminé, tous les *éléments non conformes* sont signalés au Producteur qui devra ensuite soumettre à nouveau le *SIP* à l'Entité « Entrées » ou *faire appel* de la décision auprès de l'Entité « Administration ». Une fois l'audit terminé, un *rapport final de versement* est préparé et fourni au Producteur et à la fonction « négocier un Protocole de versement ». Les méthodes d'audit peuvent comprendre des contrôles par échantillonnage, des révisions périodiques et des révisions d'experts.

La fonction « **activer les requêtes** » tient à jour un registre des demandes liées à des événements et le compare périodiquement au contenu de l'Archive pour déterminer si toutes les données requises sont disponibles. Si c'est le cas, la fonction génère une *demande de diffusion*, transmise à l'Entité « Accès ». Elle peut aussi générer des commandes à échéances régulières dont la périodicité est fixée par les Utilisateurs ou par l'occurrence d'un événement (par exemple, une mise à jour de la base de données).

La fonction « **service clients** » crée, tient à jour et clôture les comptes des Utilisateurs. Elle rassemble les *éléments de facturation* à partir de l'Entité « Accès », envoie les *factures* et encaisse le *paiement* des Utilisateurs pour l'utilisation des ressources du système d'archivage. Elle répond aux demandes d'informations générales. Cette fonction assure également la collecte et le traitement des retours à propos des services de l'Entité « Accès » et des produits. Le service clients se charge de synthétiser ces *commentaires* et de les mettre à disposition.

#### 4.1.1.6 Entité « Planification de la pérennisation »

Les fonctions de l'Entité « Planification de la pérennisation » sont illustrées dans le schéma 4-6.

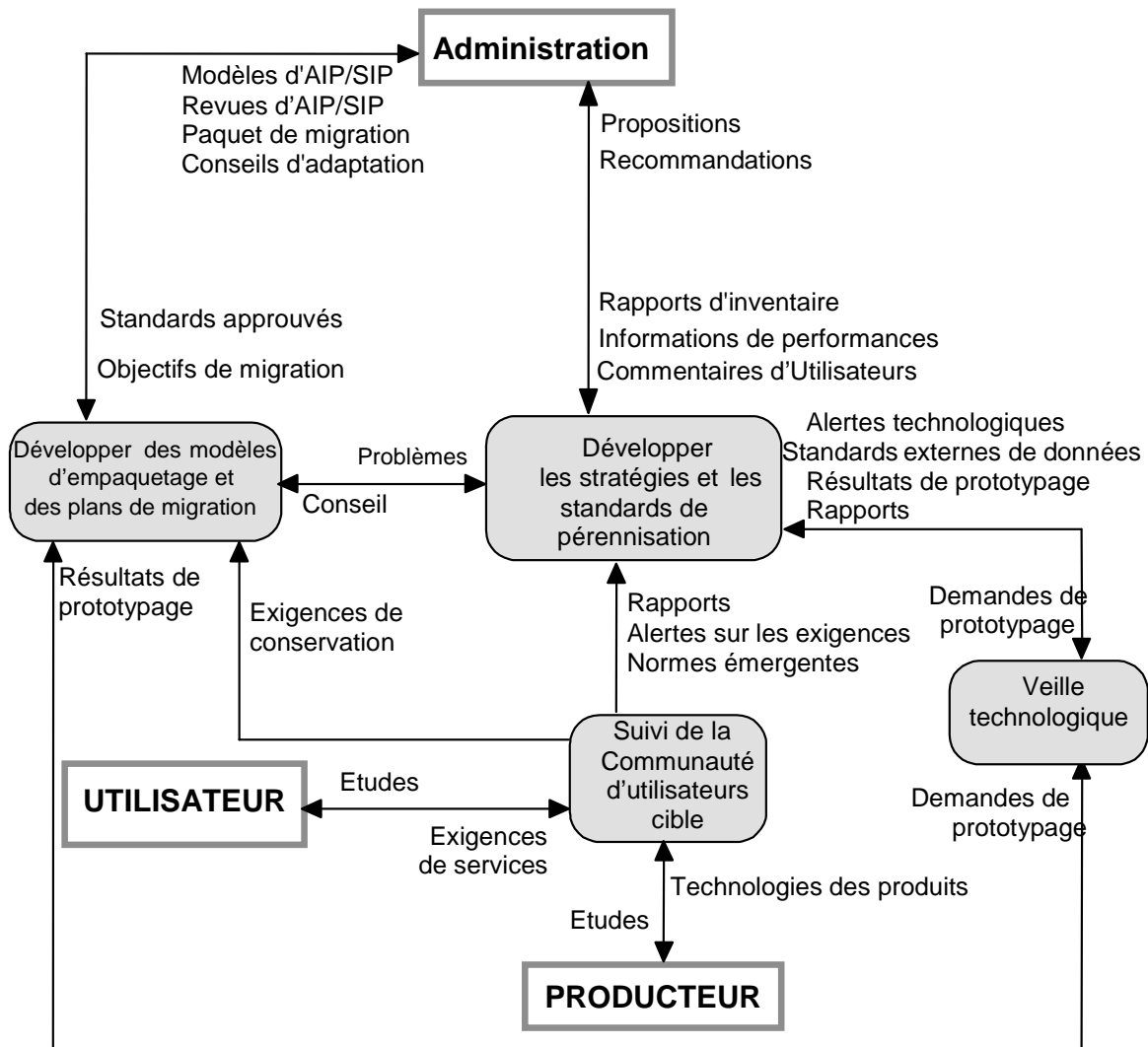


Schéma 4-6 : Fonctions de l'Entité « Planification de la pérennisation »

La fonction « suivi de la Communauté d'utilisateurs cible » interagit avec les Utilisateurs et Producteurs de l'Archive pour suivre les évolutions de leurs *exigences de services* et des *technologies* disponibles. Ces exigences peuvent porter sur les formats de données, le choix de supports, les préférences pour certains progiciels, les nouvelles plates-formes informatiques et les mécanismes de communication avec l'Archive. Cette fonction peut être réalisée à travers des *études*, un processus formel de révision périodique, des ateliers avec les communautés d'utilisateurs au cours desquels on sollicite un retour d'information des participants ou enfin par le biais d'échanges de personne à personne. Elle transmet des *rapports*, des *alertes sur les évolutions d'exigences* et des *normes émergentes* à la fonction « développer les stratégies et les standards de pérennisation ». Elle envoie des *exigences de conservation* à la fonction « développer des modèles d'empaquetage et des plans de migration ».

La fonction « **veille technologique** » est chargée de suivre les technologies numériques émergentes, les standards d'information et les plates-formes informatiques (c. à d. matérielles et logicielles) pour détecter les technologies qui pourraient conduire à une obsolescence de l'environnement informatique de l'Archive et empêcher l'accès à certains fonds courants de l'Archive. Cette fonction peut inclure une possibilité de prototypage, afin de mieux évaluer les technologies émergentes, et recevoir des *demandes de prototypage* en provenance de la fonction « développer les stratégies et les standards de pérennisation » et de la fonction « développer des modèles d'empaquetage et des plans de migration ». Cette fonction fournit des *rapports*, des *standards externes de données*, des *résultats de prototypage* et des *alertes technologiques* à la fonction « développer les stratégies et les standards de pérennisation ». Elle envoie également des résultats de prototypage à la fonction « développer des modèles d'empaquetage et des plans de migration ».

La fonction « **développer les stratégies et les standards de pérennisation** » est responsable du développement et des recommandations en matière de stratégies et de standards pour permettre à l'Archive de mieux anticiper les changements à venir sur les exigences de services venant de la Communauté d'utilisateurs cible ou sur les évolutions technologiques qui pourraient nécessiter soit la migration de fonds courants de l'Archive soit de nouveaux versements. Cette fonction reçoit des *rapports* des fonctions « suivi de la Communauté d'utilisateurs cible » et « veille technologique ». Elle reçoit aussi de la part de l'Entité « Administration » des *informations de performances*, des *rapports d'inventaire* et des synthèses de *commentaires d'utilisateurs*. Elle envoie des *recommandations* sur l'évolution du système à l'Entité « Administration ». Elle reçoit aussi des *standards externes de données* de la « veille technologique » et produit des profils de ces standards qui sont envoyés à l'Entité « Administration » en tant que *propositions* d'utilisation potentielle. Cette fonction traite aussi les *problèmes* qui lui sont signalés par la fonction « développer des modèles d'empaquetage et des plans de migration » en cas d'exigences imprévues de versement et lui *suggère* en retour la prise en compte des nouvelles exigences.

La fonction « **développer des modèles d'empaquetage et des plans de migration** » développe de nouveaux modèles d'IP et des plans de migration détaillés ainsi que des prototypes, afin de mettre en œuvre les règles et les directives de l'Entité « Administration ». Elle prodigue également des conseils pour l'application de ces modèles d'IP et fournit des plans de migration à des fonds d'archives et des versements spécifiques. Cette fonction reçoit de l'Entité « Administration » des *standards approuvés* d'archivage et des *objectifs de migration*. Ces standards, qui visent les formats, les métadonnées et la documentation, sont appliqués aux exigences de pérennisation. La fonction fournit à l'Entité « Administration » des *modèles d'AIP et de SIP* ainsi que des *conseils d'adaptation* et des comptes-rendus de *revue d'AIP/SIP* en vue de l'application de ces modèles. Si la fonction est confrontée à des versements qui ne correspondent pas aux standards et procédures existants, elle peut soumettre les *problèmes* à la fonction « développer les stratégies et les standards de pérennisation » et en recevoir un *conseil*, par exemple de nouveaux standards, pour aider à satisfaire aux nouvelles exigences du versement.

Les objectifs de migration reçus par cette fonction tendent à induire des transformations de l'AIP, y compris des transformations du Contenu d'information, pour éviter une perte d'accès due à l'obsolescence technologique. La réponse aux objectifs de migration peut induire le développement de nouveaux modèles d'AIP, de logiciels prototypes, de plans de tests, de plans de révisions par la communauté et de plans d'implémentation pour l'introduction progressive des nouveaux AIP. Le processus peut recourir à l'expertise ou aux ressources d'autres fonctions dans l'Entité « Planification de la pérennisation », telles que le développement de prototype par la « veille technologique ». Cette tâche va aussi requérir la consultation des autres entités fonctionnelles et de la Communauté d'utilisateurs cible. Une fois que le plan de migration, les modèles d'AIP associés et le logiciel auront été testés et validés, cette fonction enverra la totalité du *paquet de migration* à l'Entité « Administration », qui va organiser et réaliser la migration proprement dite.

#### 4.1.1.7 Entité « Accès »

Les fonctions de l'Entité « Accès » sont illustrées dans le schéma 4-7.

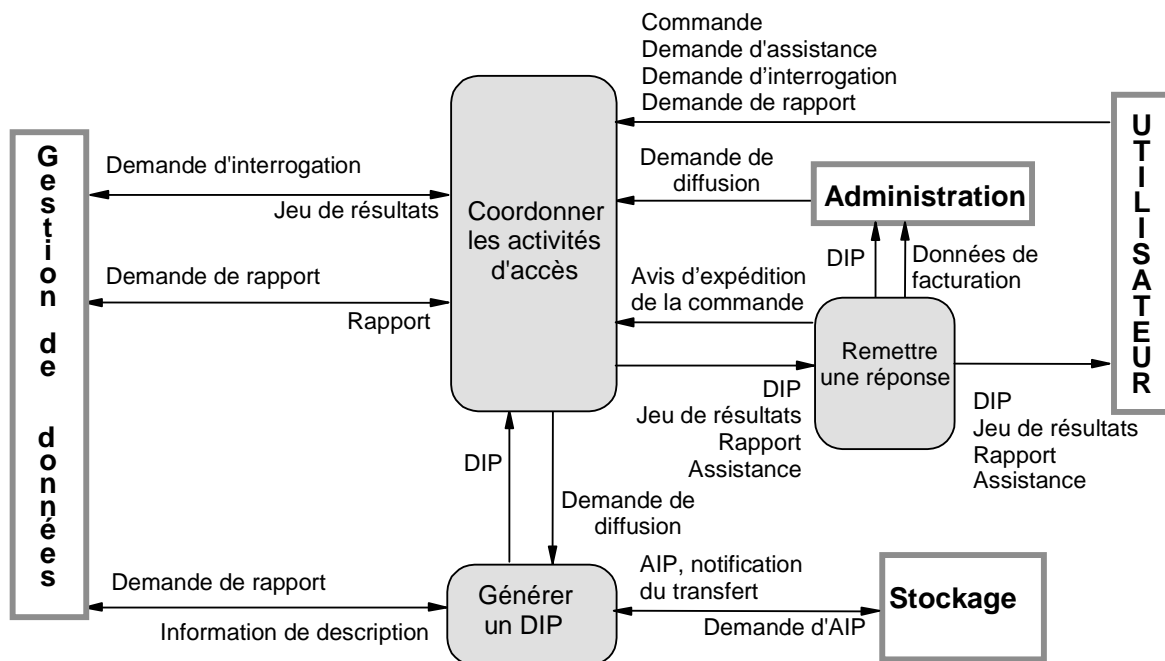


Schéma 4-7 : Fonctions de l'Entité « Accès »

La fonction « **coordonner les activités d'accès** » assure une interface unique entre l'Utilisateur et les fonds d'archives. Cette interface est normalement réalisée via un réseau informatique ou une liaison téléphonique vers un service en ligne, mais aussi par un accès physique aux locaux de l'Archive, par un service de commande sur catalogue imprimé ou par télécopie. On peut distinguer trois catégories de demandes des Utilisateurs : des *demandes d'interrogation* qui sont traitées au niveau de l'Entité « Gestion de données » et qui renvoient immédiatement des *Jeux de résultats* pour présentation à l'Utilisateur, des *demandes de rapport* qui peuvent nécessiter un certain nombre de requêtes et qui produisent des *rapports*

mis en forme pour livraison à l'Utilisateur, des *commandes* qui sont adressées à l'Entité « Gestion de données » et/ou à l'Entité « Stockage » en vue de préparer un *Paquet d'informations diffusé (DIP)* formel qui sera transmis en ligne ou en différé. La commande peut être une Commande exécutée une seule fois, ou bien une Demande d'abonnement qui sera suivie par la fonction « activer les requêtes » de l'Entité « Administration » et initialisée par une *demande de diffusion* qui pourra donner lieu à la livraison périodique des produits demandés. La fonction « mettre à jour l'information archivée » dans l'Entité « Administration » soumet aussi des *demandes de diffusion* pour obtenir les *DIP* requis pour faire ces mises à jour. D'autres types de demandes spéciales sont autorisés mais ne sont pas décrits en détail dans le présent document. Cette fonction détermine si les ressources pour traiter une demande sont disponibles, s'assure que l'Utilisateur est habilité à commander et à recevoir les produits demandés, prévient l'Utilisateur de l'acceptation ou du refus d'une demande (avec si possible une estimation du coût et une option d'annulation). Elle transmet ensuite la demande à l'Entité « Gestion de données » ou à la fonction « générer un DIP » pour exécution. Cette fonction procure également une *assistance* aux Utilisateurs de l'OAIS, comprenant la fourniture de l'état d'avancement des commandes et d'autres types de support utilisateur en réponse à une *demande d'assistance*.

La fonction « **générer un DIP** » accepte une *demande de diffusion*, récupère l'AIP à partir de l'Entité « Stockage » et envoie une copie des données vers le stockage provisoire pour traitement ultérieur. Cette fonction transmet aussi une *demande de rapport* à l'Entité « Gestion de données » pour obtenir l'*Information de description* nécessaire au DIP. Si un traitement spécifique est nécessaire, la fonction « générer un DIP » accède aux Objets-données dans la zone de stockage provisoire et effectue les traitements nécessaires. Les types d'opérations exécutées sont : les fonctions statistiques, le sous-échantillonnage en termes de temps ou d'espace, les conversions entre différents types de données ou formats de sortie et d'autres traitements spécialisés (par exemple le traitement d'images). Cette fonction renvoie le DIP réalisé dans la zone de stockage provisoire et notifie à la fonction « coordonner les activités d'accès » que le DIP est prêt pour la livraison.

La fonction « **remettre une réponse** » gère, tant en ligne qu'en différé, les livraisons des réponses (*DIP, Jeux de résultats, rapports et assistance*) aux Utilisateurs. Pour la livraison en ligne, la fonction reçoit une réponse de la fonction « coordonner les activités d'accès » et la prépare pour la distribution en ligne, en temps réel, via le réseau de communication : elle identifie le destinataire prévu, détermine la procédure de transmission requise, place la réponse dans la zone de stockage provisoire en vue de sa transmission puis assure sa transmission en ligne. Pour la livraison en différé, elle récupère la réponse de la fonction « coordonner les activités d'accès », prépare les listes de paquets et les autres dossiers d'expédition et enfin expédie la réponse. Quand la réponse a été expédiée, un *avis d'expédition de la commande* est envoyé à la fonction « coordonner les activités d'accès » et les *données de facturation* sont transmises à l'Entité « Administration ».

### 4.1.2 DIAGRAMMES DE FLUX DE DONNEES

Cette sous-section présente le flux d'éléments de données entre entités fonctionnelles OAIS sous forme de diagrammes. Le schéma 4-8 montre les flux de données les plus significatifs. Pour simplifier ce schéma, les flux de données de l'Entité « Administration », qui constituent généralement des tâches de fond, sont isolés dans un diagramme de contexte de l'Entité « Administration » - schéma 4-9. Les flux de données associés aux services de base sont implicites dans les fonctions présentées et n'y figurent donc pas.

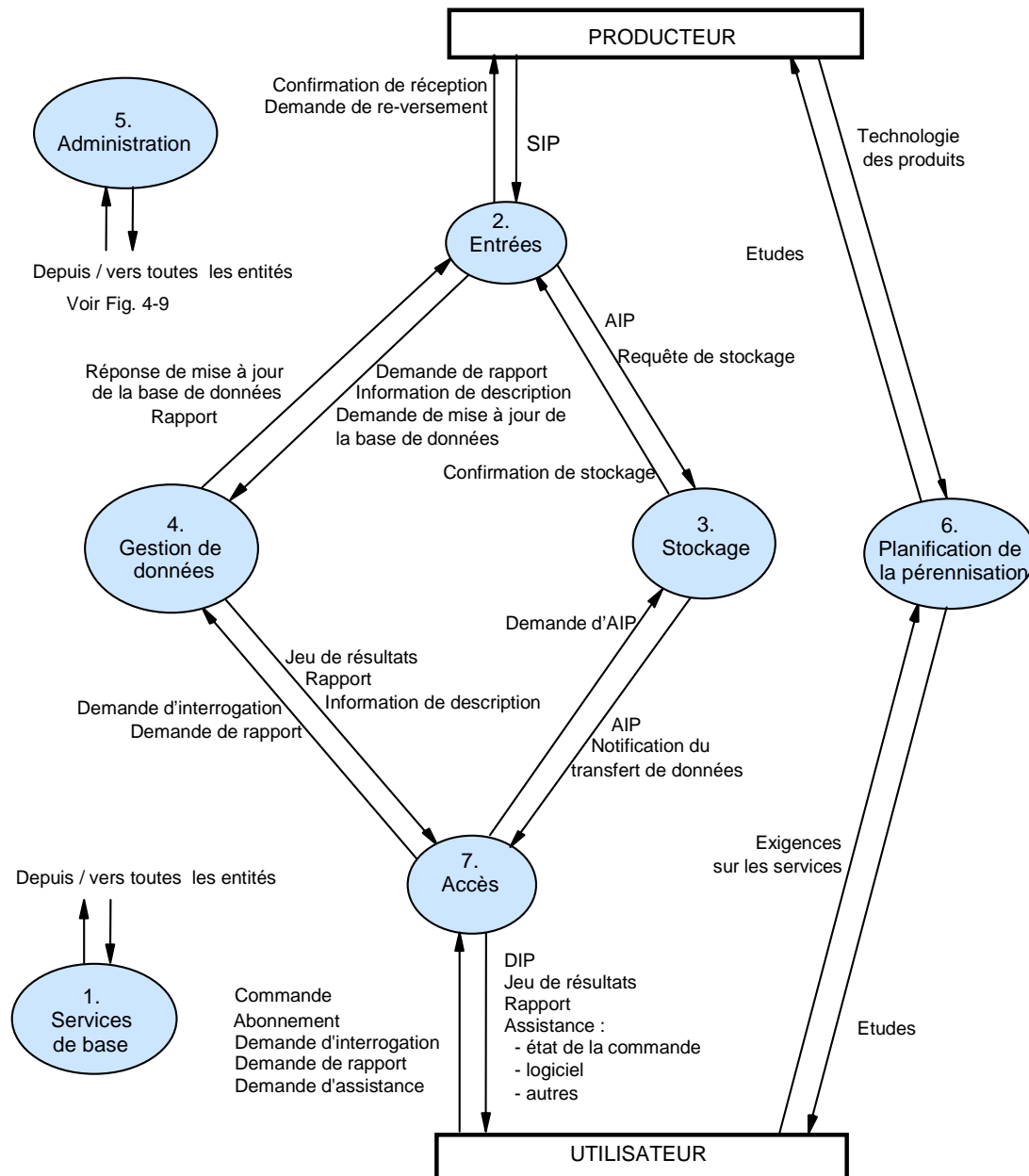
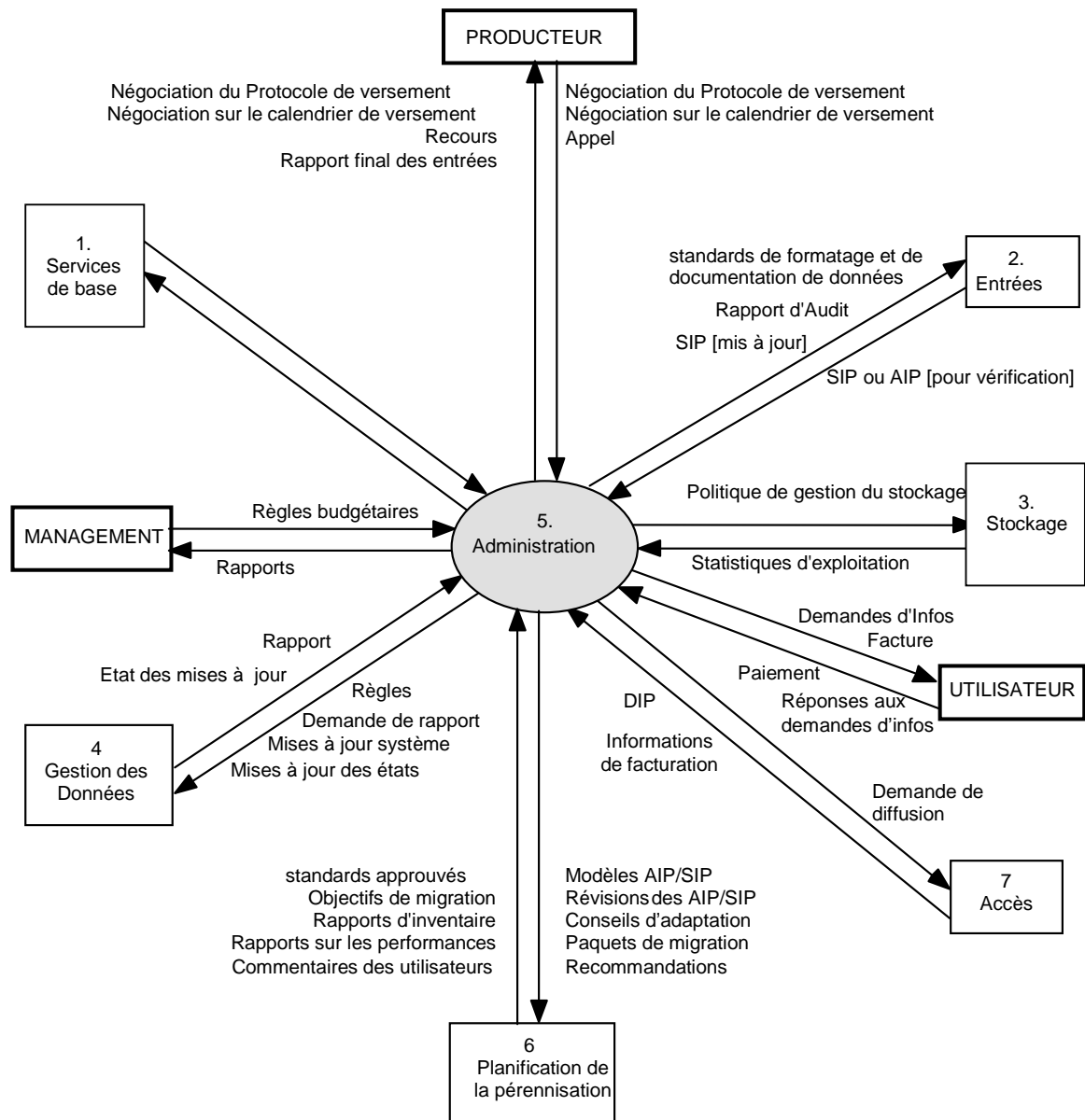


Schéma 4-8 : Diagramme de flux de Données OAIS



**Schéma 4-9 : Diagramme de contexte de l'Entité « Administration »**

## 4.2 MODELE D'INFORMATION

Cette sous-section s'appuie sur les concepts présentés dans la section 2 pour décrire plus précisément les catégories d'informations échangées et gérées dans l'OAIS. Elle définit aussi les Objets-information spécifiques utilisés dans l'OAIS pour pérenniser l'information confiée à l'Archive et y donner accès. Ce modèle plus détaillé d'Objets-information liés à l'OAIS est destiné à aider l'architecte ou le concepteur système de futurs OAIS. Les objets présentés dans cette sous-section sont conceptuels et ne prétendent pas définir une mise en oeuvre.

Comme cela a été présenté dans la section 2, l'objectif premier d'un OAIS est de pérenniser l'information pour une Communauté cible sur une période de temps indéfinie. Pour pérenniser cette information, un OAIS doit stocker beaucoup plus que le simple contenu de l'objet à conserver. Cette sous-section analyse les exigences requises pour les informations utilisées pour décrire les classes d'Objets-données associées à un OAIS. Elle utilise des diagrammes de modèle objet du Langage Universel de Modélisation (UML) afin d'illustrer les concepts décrits dans le texte. Une vue d'ensemble de la notation utilisée et des concepts clefs de la modélisation objet est présentée dans l'annexe B de ce document. La connaissance de cette notation est indispensable pour une compréhension complète des concepts présentés.

La sous-section 4.2.1 présente le modèle d'information indispensable pour une Pérennisation optimale. La sous-section 4.2.2 décrit les objets et conteneurs conceptuels qui constituent les contenus d'un OAIS.

## **4.2.1 MODELE LOGIQUE POUR UNE INFORMATION ARCHIVEE**

### **4.2.1.1. Objet-information**

Le concept d'information, en tant que combinaison de données et d'Information de représentation est un concept de base, du Modèle de référence OAIS. Le diagramme UML dans le schéma 4-10 illustre ce concept. L'**Objet-information** est composé d'un **Objet-données**, physique ou numérique, et de l'**Information de représentation** qui permet de traduire les données en information signifiante. Ce modèle est valable pour toutes les catégories d'informations d'un OAIS.



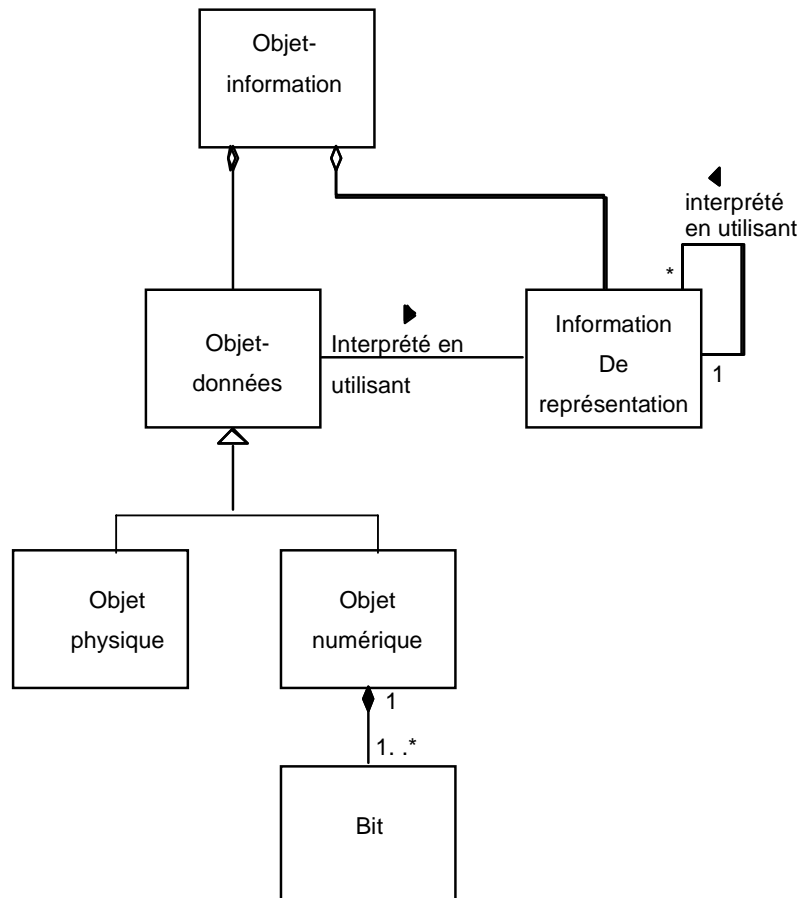


Schéma 4-10 : Objet-information

#### 4.2.1.2. Objet-Données

L'Objet-données peut aussi bien prendre la forme d'un objet physique (par exemple, une roche lunaire) accompagné d'Information de représentation, que celle d'un objet numérique (c'est-à-dire une séquence de bits) accompagné de l'Information de représentation qui confère une signification à ces bits.

#### 4.2.1.3. Information de représentation

L'Information de représentation accompagnant un objet physique telle qu'une roche lunaire peut apporter un complément d'information aux attributs physiquement observables de la roche, comme par exemple des résultats d'analyse. Cette information peut avoir été élaborée au fil du temps et les résultats, s'ils sont disponibles, feront partie de l'Objet-information.

L'Information de représentation qui accompagne un objet numérique, ou une séquence de bits sert à donner une signification supplémentaire à cet objet. Son rôle est d'établir la correspondance entre les bits et des types de données généralement reconnus comme un caractère, un nombre entier, un nombre réel, ou des groupes de ces types de données. Elle

relie ces types de données aux concepts signifiants de plus haut niveau, eux-mêmes à des relations complexes décrites par ailleurs.

La suite de cette sous-section traite de l'Objet-information de représentation quand l'Objet-données est spécialisé en Objet numérique.

#### 4.2.1.3.1. Types d'Information de représentation

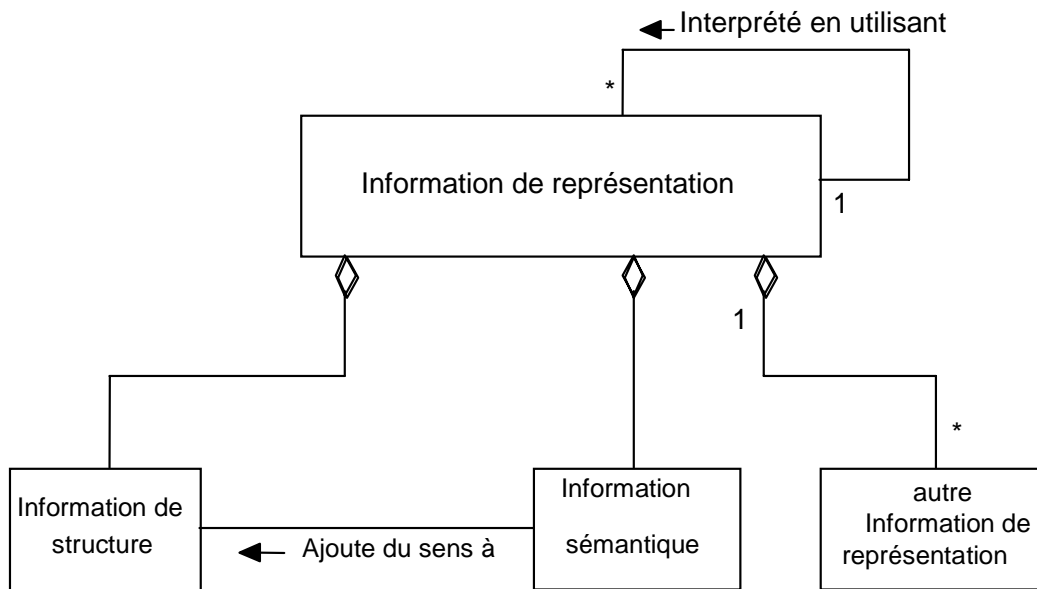
L'Objet numérique, comme cela est montré dans le schéma 4-10, est composé d'une ou plusieurs séquences de bits. La fonction de l'Objet-information de représentation est de traduire les séquences de bits en information intelligible. Pour cela, il décrit le format ou les concepts de structure de données applicables aux séquences de bits, ce qui aboutit à des éléments plus signifiants comme des caractères, des nombres, des pixels, des tableaux, des tables, etc. Ces types de données courants en informatique, les associations de ces types de données et les règles qui établissent les correspondances entre les types de données sous-jacents et des concepts de plus haut niveau nécessaires à la compréhension de l'Objet numérique sont appelés **Information de structure** de l'Objet-information de représentation. Ces structures sont communément identifiées par leur nom ou par leur position relative au sein des séquences de bits correspondantes.

L'Information de représentation fournie par l'Information de structure est rarement suffisante. Même dans le cas où l'Objet numérique est interprété comme une suite de caractères (texte) et décrit comme tel dans l'Information de structure, il faut fournir une information complémentaire sur la langue dans laquelle l'objet est exprimé. Ce type d'informations complémentaires nécessaire est appelé **Information sémantique**. S'agissant de données scientifiques par exemple, le contenu de l'Information sémantique peut être très diversifié et complexe. Cela peut comprendre la signification particulière associée à chacun des éléments de l'Information de structure, les opérations réalisables sur chaque type de données et leurs corrélations. Le schéma 4-11 souligne le fait que l'Information de représentation contient tout à la fois l'Information de structure et l'Information sémantique bien que, dans certaines mises en œuvre, la distinction entre les deux soit parfois subjective.

Ce schéma montre aussi que l'Information de représentation peut contenir des références à d'autres Informations de représentation. Couplé avec le fait que l'Information de représentation est un Objet-information possédant son propre Objet numérique et l'Information de représentation associée à la compréhension de cet Objet (comme cela est indiqué de façon concise par l'association : "interprété en utilisant"), le jeu résultant d'objets est appelé **Réseau de représentation**.

A titre d'exemple, la norme ISO 9660 (référence [13]) décrit le texte comme devant se conformer à la norme ASCII, mais en réalité elle ne décrit pas comment l'ASCII doit être mis en œuvre. Elle fait simplement référence à la norme ASCII, qui constitue une Information de représentation complémentaire indispensable pour une compréhension complète de la norme ISO 9660. La norme ASCII constitue donc une partie du Réseau de représentation associé à

l'ISO 9660. Il s'ensuit que la norme ASCII doit être présente dans l'OAIS d'une manière ou d'une autre, ou alors que l'OAIS vérifie la disponibilité permanente de cette norme afin qu'il puisse prendre à l'avenir les dispositions qui garantissent que l'Information de représentation ISO 9660 reste entièrement compréhensible.



**Schéma 4-11 : Objet-information de représentation**

#### 4.2.1.3.2 Réseaux de représentation

L'Information de représentation, qui est elle-même un Objet-information, peut être exprimée sous une forme physique (par exemple, un document papier) ou sous une forme numérique. Quand l'Information de représentation est sous forme numérique, une Information de représentation complémentaire est nécessaire pour comprendre les bits des Informations de représentation, comme nous l'avons décrit dans la sous-section précédente. En principe, ce processus de récursivité se poursuit jusqu'à ce que l'on aboutisse à des formes physiques. Par exemple, l'Information de représentation exprimée en ASCII nécessite une Information de représentation complémentaire pour l'ASCII, qui peut consister en un document physique présentant le standard ASCII. Chaque entité d'Information de représentation peut être constituée de multiples composants ou inclure de multiples références à des composants, chacun de ces composants ayant lui-même sa propre Information de représentation.

Pour conserver la signification d'un Objet-information, son Information de représentation doit être également conservée. Ceci est plus facile à réaliser quand les Objets-information de représentation sont exprimés dans une forme aisément compréhensible, par exemple des descriptions textuelles utilisant les normes en vigueur les plus répandues comme les caractères ASCII sous forme numérique. Un problème posé par l'usage exclusif de descriptions textuelles est que ces descriptions peuvent être ambiguës. Le problème peut être

résolu par l'utilisation de langages de description formels standardisés, contenant des syntaxes bien définies pour la description des structures de données. Ces langages doivent pouvoir être complétés par des descriptions textuelles pour transmettre pleinement la sémantique de l'Information de représentation.

Il existe deux types particuliers d'Informations de représentation : **le Logiciel de restitution de la représentation** et **le Logiciel d'accès**. Le Logiciel de restitution de la représentation est capable d'afficher les Informations de représentation sous forme compréhensible. Par exemple, la structure de fichiers et de répertoires de nombreux CD-ROM se conforme à la norme ISO 9660. Cette norme fournit de l'Information de représentation décrivant comment la plupart des structures de fichier de CD-ROM doivent être mises en œuvre et peut être obtenue sous forme papier. Cependant cette Information de représentation peut aussi être obtenue sous forme numérique, c'est-à-dire implicitement comme un objet PDF. Plutôt que d'obtenir effectivement la documentation du format PDF et d'écrire le logiciel pour interpréter la norme ISO 9660, un OAIS peut utiliser un logiciel de visualisation du PDF disponible afin de rendre la documentation ISO 9660 visible et lisible par un être humain. Dans ce cas le logiciel d'affichage PDF est considéré comme un Logiciel de restitution de la représentation parce qu'il est utilisé pour restituer l'Information de représentation. Si l'OAIS n'obtient pas également la description associée du format PDF, il doit en garder la trace et en assurer le suivi car, dès lors que l'accès et l'affichage des objets PDF deviendront trop coûteux, la documentation de la norme ISO 9660 exprimée sous forme d'objet PDF devra être migrée vers un nouveau format.

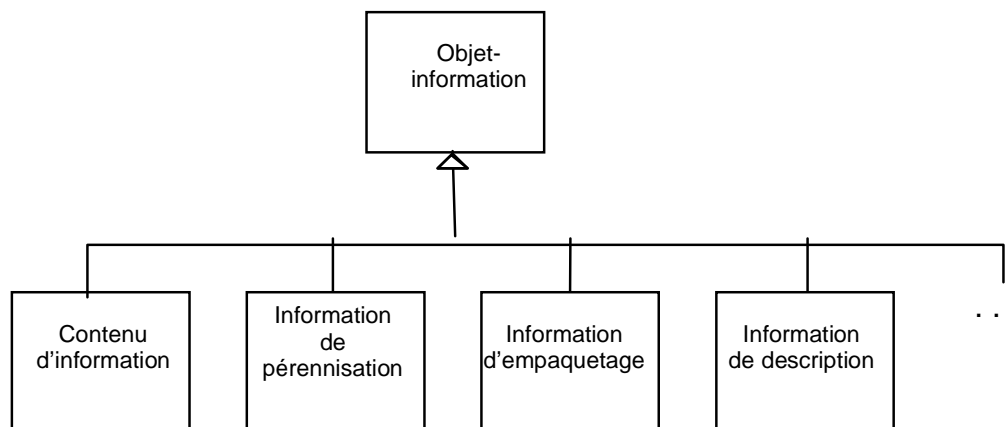
Le Logiciel d'accès présente tout ou partie du contenu de l'information d'un Objet-information sous des formes compréhensibles aux humains ou aux machines. Il peut aussi fournir différents types de service d'accès à un Objet-information, comme l'affichage, la manipulation, le traitement, ou l'extraction de sous-ensembles. Pour certains types d'Objets numériques, ce type de logiciel peut être très répandu ; il n'est pas nécessaire dès lors que l'OAIS le fournisse ou le maintienne. Cependant l'OAIS peut décider de le faire pour des types d'Objets numériques plus spécifiques.

Puisque les Logiciels d'accès ont tendance à inclure une part de sémantique des Informations de représentation, certaines Archives peuvent être tentées d'utiliser le Logiciel d'accès comme substitut à une Information de représentation complète. Le code source du Logiciel d'accès, qui permet une certaine compréhension des Informations de représentation associées à l'Objet-données, peut être utilisé comme documentation exprimant de telles Informations de représentation. Cette approche est problématique en ce sens que l'Information de représentation souhaitable peut ne pas être clairement identifiée si elle est mélangée avec des traitements divers et des algorithmes d'affichage ou s'avérer incomplète si le code source présuppose un environnement d'exploitation. Il peut alors être difficile de dire, à partir du code source du logiciel, quelles sont les Informations de représentation qui manquent. L'utilisation de Logiciels d'accès sous forme d'exécutables, sans accès au code source, comme dans le cas des formats propriétaires, présente un risque beaucoup plus élevé de perte d'informations parce qu'il est plus difficile de maintenir un environnement d'exploitation pour un logiciel que de migrer de la documentation au cours du temps. Il n'est pas prouvé que l'utilisation pratique de techniques d'émulation pour conserver le logiciel en état de

fonctionnement est plus rentable. C'est une question cruciale pour ceux qui désirent conserver le *'look and feel'* de l'accès à l'information. La migration ainsi que la conservation des logiciels sont traitées plus précisément à la section 5.

#### 4.2.1.4 Taxinomie des classes d'Objets-Information utilisées par l'OAIS

De nombreuses catégories d'informations sont mises en jeu pour la conservation d'informations à long terme dans un OAIS. Chacune de ces catégories peut être vue comme un Objet-information complet en ce qu'il contient un Objet-données et une Information de représentation adéquate pour comprendre ces données. Cette sous-section s'appuie sur les catégories d'information requises pour permettre la Pérennisation définies dans la section 2.2 d'une part, et l'examen du rôle de l'Information de représentation dans la sous-section précédente d'autre part. La modélisation d'informations dans cette sous-section traite de plusieurs catégories d'Objets-information qui sont utilisés dans l'OAIS. Les objets sont classés en fonction de leur contenu et de leur rôle dans le fonctionnement d'un OAIS, dans les catégories suivantes : Objets-contenu d'information, Objets-information de pérennisation, Objets-information d'empaquetage et Objets-information de description. Les sous-sections suivantes traitent du contenu de chacune des catégories d'Objets-information. Le schéma 4-12 montre une taxinomie de ces Objets-information utilisés dans l'OAIS.



**Schéma 4-12 : Taxinomie de l'Objet-information**

##### 4.2.1.4.1 Contenu d'information

Le Contenu d'information est l'ensemble d'informations qui constitue le premier objectif de conservation par l'OAIS. Décider de ce qu'est un Contenu d'information ne va pas de soi et peut nécessiter une discussion avec le Producteur. Le Contenu d'information, qui est un Objet-information tel qu'il est présenté sur le schéma 4-10, est l'Objet-contenu de données accompagné de son Information de représentation. L'Objet-contenu de données dans les Contents d'information peut être un Objet numérique ou un Objet physique (par exemple, un

échantillon physique, un microfilm). Tout Objet-information peut être utilisé comme Contenu d'information.

L'Information de représentation (tant sémantique que syntaxique) d'un Objet-contenu de données numérique est nécessaire pour convertir entièrement les bits en Contenu d'information. Cette nécessité va jusqu'à l'insertion de définitions (par exemple, le dictionnaire et la grammaire) de n'importe quel langage naturel (par exemple, l'anglais) utilisé dans le Contenu d'information. Sur de longues périodes de temps, la signification des expressions du langage naturel peut évoluer notablement tant dans leur usage général que pour des disciplines spécifiques.

En pratique, l'OAIS doit posséder une Information de représentation des bits de l'Objet-contenu de données dans le Contenu d'information qui soit suffisamment complète pour qu'il soit sûr que les membres de la Communauté cible prennent en compte le Réseau de représentation avec assez de connaissance pour interpréter correctement l'Information de représentation. C'est un domaine à fort risque pour un OAIS, particulièrement lorsqu'il est lié à une Communauté cible experte, le jargon et les termes apparemment largement compris pouvant avoir une durée de vie limitée. Dans de tels cas il est essentiel de s'assurer que l'évolution naturelle de la Base de connaissance de la Communauté cible n'entraîne pas effectivement une perte d'information dans les Contenus d'information.

Comme pour les Objets-information (voir ci-dessus), on peut considérer que les Informations de représentation sont enrichies par le Logiciel d'accès qui présente les Contenus d'information à l'Utilisateur. Citons comme exemple de ce type les logiciels de traitements de texte qui acceptent des représentations complexes de format de document et les systèmes de visualisation scientifique prenant en compte les représentations de Contenu d'information en tant que série temporelle ou tableau multidimensionnel. Le logiciel utilise sa propre connaissance de l'Information de représentation pour fournir de tels services.

Il arrivera souvent que l'information nécessaire soit intégrée dans les progiciels utilisés par la Communauté d'utilisateurs cible pour présenter et analyser les Contenus d'information. Conserver le Logiciel d'accès en état de fonctionnement résulte d'une analyse de commodité. En pratique, même avec un jeu complet d'Informations de représentation, l'accès à tout ou partie d'un Objet-contenu de données numérique exige l'utilisation du Logiciel d'accès. C'est pourquoi un module logiciel qui fournit un accès utile à un Objet-contenu de données numérique peut être préservé en état de fonctionnement parce que cela est commode.

Ceci ne présente pas de difficulté tant que l'environnement qui supporte le module logiciel est facilement disponible. Cet environnement est constitué d'une plate-forme matérielle et d'un système d'exploitation, de différents utilitaires qui complètent le système d'exploitation, de moyens de stockage, ainsi que de dispositifs d'affichage accompagnés de leurs pilotes. Une modification d'un de ces éléments peut rendre le module logiciel inopérant ou son fonctionnement incorrect, ou encore le rendre incapable de fournir des résultats à l'application ou à un utilisateur humain. C'est en général la complexité de ces interactions qui rend si ardue le maintien des logiciels en état de fonctionnement.

En résumé, l'utilisation d'un Logiciel d'accès pour remplacer les Réseaux de représentation est séduisante dans l'optique d'une réduction maximale des ressources requises pour verser les données à l'OAIS et offrir l'accès aux données aux utilisateurs. Cependant, le fait de s'en remettre au fonctionnement du logiciel peut entraîner des problèmes majeurs pour la Pérennisation quand ce logiciel cesse de fonctionner. La Pérennisation de l'information exige une description complète et compréhensible de l'Information de représentation. La sous-section 5.2 (Préservation des services d'accès) traite de quelques techniques pouvant être utilisées pour préserver le logiciel dans la durée et des risques induits par cette approche.

Distinguer la part Objet-contenu de données et la part Information de représentation dans un Contenu d'information est une fonction importante de l'OAIS. Cet aspect est déterminant pour une appréhension claire de ce qui sera conservé. L'identification du Contenu d'information numérique accompagné de ses Objets-information de représentation peut être effectuée selon les étapes suivantes :

- 1) Identifier les bits constituant l'Objet-contenu de données du Contenu d'information.
- 2) Identifier un Objet-information de représentation qui, d'une certaine façon, prend en compte tous les bits de l'Objet-contenu de données pour en faire une information plus signifiante.
- 3) Examiner le contenu de l'Objet-information de représentation identifié, afin de déterminer s'il est nécessaire d'ajouter d'autres Objets-information de représentation. Dans l'affirmative, obtenir les Objets-information de représentation nécessaires. Répéter cette étape jusqu'à ce que plus aucun Objet-information de représentation complémentaire ne soit identifié.
- 4) Pour chaque Objet-information de représentation de l'étape 3, qui est conservé sous la forme d'un Objet numérique, identifier tout Objet-information de représentation requis et répéter les étapes 3 et 4 jusqu'à ce qu'aucun nouvel Objet-information de représentation ne soit identifié.
- 5) Le Contenu d'information est constitué de l'Objet-contenu de données et de chacun des Objets-information de représentation identifiés dans les étapes 2 à 4.

Comme exemple de cette pratique, considérons un fichier électronique contenant une séquence de valeurs obtenues à partir d'un capteur observant l'environnement terrestre. Un deuxième fichier, encodé en ASCII, fournit l'information sur la façon d'interpréter le premier fichier. Il décrit comment interpréter les bits du premier fichier pour obtenir des nombres significatifs. Il explique ce que ces nombres signifient du point de vue de la Physique pour l'observation conduite. Il détermine la date et la période à laquelle les observations ont été effectuées, il donne une valeur moyenne pour les valeurs observées, et indique qui a réalisé les observations. Ces deux fichiers sont versés à un OAIS pour leur conservation.

Supposons que l'OAIS détermine que le Contenu d'information à conserver se compose des bits de l'observation, associés à leurs valeurs sous forme de valeurs numériques et à la signification physique de ces nombres. Cette information est transmise par la séquence de

bits du premier fichier associée à l'Information de représentation du deuxième fichier laquelle est nécessaire pour convertir les bits du premier fichier en des valeurs physiques significatives. Notons que, dans cet exemple, ni les supports physiques sous-jacents du premier fichier ni le système de gestion spécifique de fichiers qui gère les bits ne font partie du Contenu d'information. Seule une partie du contenu du deuxième fichier est considérée comme partie du Contenu d'information et c'est la partie qui permet la conversion des bits du premier fichier en valeurs physiques significatives. En fait ce second fichier ne véhicule pas toute l'Information de représentation requise pour effectuer cette transformation ; en effet, des informations complémentaires sont nécessaires, à savoir :

- l'information selon laquelle le deuxième fichier est encodé en ASCII de façon à ce qu'il soit lu en tant que caractères significatifs,
- l'information sur la manière dont les caractères sont utilisés pour en extraire les conversions de bits en des nombres en valeurs physiques significatives.

Cette information, qui se rapporte de façon classique à une combinaison d'informations sur le format et d'informations relevant du Dictionnaire de données, peut aussi inclure des valeurs d'étalonnage des instruments ainsi que l'information décrivant comment les étalonnages doivent être effectués. Toute cette information peut être aisément compréhensible dès que les caractères ASCII sont visibles parce qu'elle a été exprimée entièrement en anglais (ou tout autre langage naturel). Une partie de cette information peut aussi se présenter sous des formes plus structurées et nécessiter, pour être comprise, une Information de représentation complémentaire.

Donc, l'Information de représentation du deuxième fichier nécessite une Information de représentation complémentaire qui peut, à son tour, nécessiter aussi une Information de représentation complémentaire, etc. formant une chaîne de représentations de représentations. C'est un bon exemple de Réseau de représentation complexe.

Rappelons que dans l'exemple ci-dessus, on avait défini le Contenu d'information comme les valeurs relevées par les capteurs avec leurs significations. Ce n'est pas le seul choix possible : on aurait aussi bien pu définir l'Objet-contenu de données du Contenu d'information comme les séquences de bits du premier fichier plus toutes les séquences de bits du deuxième fichier. Le fait que certaines de ces dernières séquences de bits soient utilisées pour interpréter les premières séquences de bits du fichier est juste un exemple d'un jeu de bits en quelque sorte auto-descriptif. Le fait que certains des bits du deuxième fichier constituent l'information de base pour la date et la durée de la phase d'observation, la valeur moyenne de ces observations et leur auteur est sans importance. Une fois qu'il a été décidé que tous ces bits constituent l'Objet-contenu de données du Contenu d'information, l'Information de représentation est l'information nécessaire pour transformer ces bits en information signifiante. Le degré de détail de cette signification et le degré de développement du Réseau de représentation sont des questions internes à l'OAIS, à ses Producteurs et à ses communautés d'Utilisateurs.

Autre exemple : considérons un fichier électronique contenant un document issu du traitement de texte. Cet objet de données binaire a un format complexe et ne peut être appréhendé comme un document qu'à l'aide de son Information de représentation. En



pratique, c'est le Logiciel d'accès, intégrant une connaissance de l'Information de représentation, qui fournira cette vue. Le contenu de la séquence de bits du fichier électronique est le candidat le plus probable à la définition de l'Objet-contenu de données. L'Information de représentation, quant à elle, est au minimum une description du format du traitement de texte, et peut inclure aussi toute information jugée nécessaire à la bonne compréhension du sens du document visualisé. Si le format de traitement de texte est un format propriétaire et qu'il ne peut pas être décrit, même au niveau de la visualisation du document, il peut être nécessaire de migrer le document vers un format non propriétaire pour assurer sa Pérennisation.

A titre de variante de l'exemple précédent, on peut décider que le Contenu d'information à conserver n'est pas la vue complète du document donnée par le traitement de texte mais simplement une suite de paragraphes textuels qui peuvent être représentés correctement par des caractères ASCII. Dans ce cas, l'OAIS peut décider d'extraire ces textes en mode caractère et de les sauvegarder en tant que fichier texte. L'Objet-contenu de données serait alors probablement défini comme le flux de bits constitué par ces caractères. Les Informations de représentation consisteraient alors en une description de la manière d'interpréter ce flux de bits comme des caractères, enrichie de toute information complémentaire jugée nécessaire à la bonne compréhension du sens du texte.

#### 4.2.1.4.2 Information de pérennisation

En complément du Contenu d'information, les Informations archivées doivent inclure les informations qui permettront la compréhension de ce Contenu d'information sur une période de temps indéfinie. Le jeu particulier des Objets-information, requis pour cette fonction, est globalement appelé Information de pérennisation (PDI). Le PDI doit inclure les informations nécessaires à une bonne conservation des Contenus d'informations particuliers auxquels il se rapporte. Il vise particulièrement la description des états passés et présents des Contenus d'information, en assurant qu'il est identifiable de manière unique et qu'il n'a pas été altéré accidentellement.

Ces informations sont classiques pour tous les types d'archives et ont déjà fait l'objet d'une classification pour les archives traditionnelles. Pour les archives numériques, ces définitions de classe doivent cependant être élargies. Les définitions suivantes sont basées sur les catégories discutées dans l'article 'Preserving Digital Information' (référence [2]). Le rapport entre les concepts définis dans le Modèle de référence OAIS et l'article 'Preserving Digital Information' est abordé dans l'annexe A de ce document. La table 4-1 fournit des exemples illustrant cette information dans divers types de Contenu d'information courants.

- **Information d'identification** : elle identifie et, si nécessaire, décrit un ou plusieurs mécanismes utilisés pour fournir des identificateurs affectés aux Contenus d'information. Elle fournit aussi les identificateurs qui permettent à des systèmes externes de se référer, sans ambiguïté, à ces Contenus d'information spécifiques. Par exemple : les systèmes taxinomiques, les systèmes de référence et les systèmes d'enregistrement. Dans le Modèle de référence OAIS, la plus grande partie (si ce n'est

la totalité) de ces informations est dupliquée dans des Descriptions de paquet, ce qui permet à l'Utilisateur d'accéder aux Contenus d'information qui l'intéressent.

- **Information de contexte :** elle documente les relations entre le Contenu d'information et son environnement, y compris les motifs de création du Contenu d'information et les relations qu'il entretient avec d'autres Objets-contenu d'information existant par ailleurs.
- **Information de provenance :** elle documente l'historique des Contenus d'information. Elle témoigne de l'origine ou de la source des Contenus d'information, de tous changements intervenus depuis leur création et de qui en a été dépositaire depuis lors. Cela donne aux utilisateurs futurs une certaine assurance quant à la fiabilité des Contenus d'information. La provenance peut être vue comme un type particulier d'Information de contexte.
- **Information d'intégrité :** elle fournit des moyens de contrôles de l'intégrité des Données ou des clefs de validation/vérification utilisées pour assurer qu'aucun Objet-contenu d'information n'a été modifié de façon non documentée. Les Informations d'intégrité incluent le codage particulier et les modules de détection d'erreur qui sont spécifiques aux différents Objets-contenus. Les Informations d'intégrité ne comprennent pas les mécanismes de conservation d'intégrité fournis par les services de base de l'OAIS : mécanismes de correction d'erreur fournis par les supports physiques et pilotes de périphériques utilisés par l'Entité « Stockage ». L'Information d'intégrité peut cependant spécifier un niveau minimal de qualité pour les services offerts par ces mécanismes.

L'OAIS doit explicitement décider quelle est la définition exacte du Contenu d'information afin de pouvoir assurer qu'il détient aussi le PDI requis pour conserver ce Contenu d'information. Une fois le Contenu d'information défini, il est alors possible d'évaluer l'Information de pérennisation.

#### 4.2.1.4.3 Information d'empaquetage

L'Information d'empaquetage est celle qui, de façon réelle ou logique, lie ou met en relation les composants du paquet au sein d'une entité identifiable sur des supports spécifiques. Par exemple, si le Contenu d'information et son PDI sont définis comme le contenu de fichiers spécifiques sur un CD-ROM, alors l'Information d'empaquetage peut mentionner la norme ISO 9660 (structure de volume et de fichier des disques optiques compacts à mémoire fixe (CD-ROM) destinés à l'échange d'information). Ces choix font l'objet de définitions ou de conventions des Archives au niveau local. Dans la mesure où l'Information d'empaquetage ne contribue pas au Contenu d'information ou au PDI, elle n'a pas à être nécessairement conservée par l'OAIS. Cependant, il y a des cas où l'on peut exiger de l'OAIS qu'il reproduise exactement le versement original. Dans ce cas le Contenu d'information est défini de façon à ce qu'il contienne tous les bits versés.

**Table 4-1: Exemples de Types de PDI**

Type de Contenu d'information	Identification	Provenance	Contexte	intégrité
<b>Données scientifiques spatiales</b>	<ul style="list-style-type: none"> <li>Identificateur de l'Objet</li> <li>Référence du Journal</li> <li>Mission, instrument, titre et ensemble d'attributs</li> </ul>	<ul style="list-style-type: none"> <li>Description d'instrument</li> <li>Historique du traitement</li> <li>Description du capteur</li> <li>Instrument</li> <li>Mode Instrument</li> <li>Carte de décommutation</li> <li>Spécification d'Interface logiciel</li> </ul>	<ul style="list-style-type: none"> <li>Historique de l'étalonnage</li> <li>Jeux de données connexes</li> <li>Mission</li> <li>Historique du financement</li> </ul>	<ul style="list-style-type: none"> <li>CRC</li> <li>Checksum</li> <li>Codage Reed-Solomon</li> </ul>
<b>Information bibliographique</b>	<ul style="list-style-type: none"> <li>ISBN</li> <li>Titre</li> <li>Auteur</li> </ul>	<ul style="list-style-type: none"> <li>Historique de l'impression</li> <li>Droits d'auteur</li> <li>Position dans la série</li> <li>Références</li> </ul>	<ul style="list-style-type: none"> <li>Références associées</li> <li>Système Dewey</li> <li>Date de publication</li> <li>Editeur</li> </ul>	<ul style="list-style-type: none"> <li>Auteur</li> <li>Signature électronique</li> <li>Couverture</li> </ul>
<b>Progiciel</b>	<ul style="list-style-type: none"> <li>Nom</li> <li>Auteur/origine</li> <li>Numéro de version</li> <li>Numéro de série</li> </ul>	<ul style="list-style-type: none"> <li>Historique de la révision</li> <li>Titulaire de la licence</li> <li>Enregistrement</li> <li>Droits d'auteur</li> </ul>	<ul style="list-style-type: none"> <li>Fichier d'aide</li> <li>Guide utilisateur</li> <li>Logiciel associé</li> <li>Langage</li> </ul>	<ul style="list-style-type: none"> <li>Certificat</li> <li>Checksum</li> <li>Encryptage</li> <li>CRC</li> </ul>

Il faut également éviter que l'OAIS conserve le PDI ou le Contenu d'information uniquement en utilisant des conventions de structure des noms des fichiers ou des répertoires. Ces structures sont plutôt utilisées comme Informations d'empaquetage. Les Informations d'empaquetage ne sont pas conservées en cas de migration. Une information sauvegardée dans les structures de noms de fichiers ou de répertoires peut donc être perdue quand l'Information d'empaquetage est modifiée. La question de l'Information d'empaquetage est un point important à prendre en considération dans un OAIS lors de migration d'informations vers des supports plus récents. Ce sujet est traité en détail dans la section 5 de ce document.

#### 4.2.1.4.4 Information de description

Les Objets-information décrits précédemment dans cette section fournissent les informations nécessaires à la fonction de « Pérennisation » de l'Archive. En complément de la

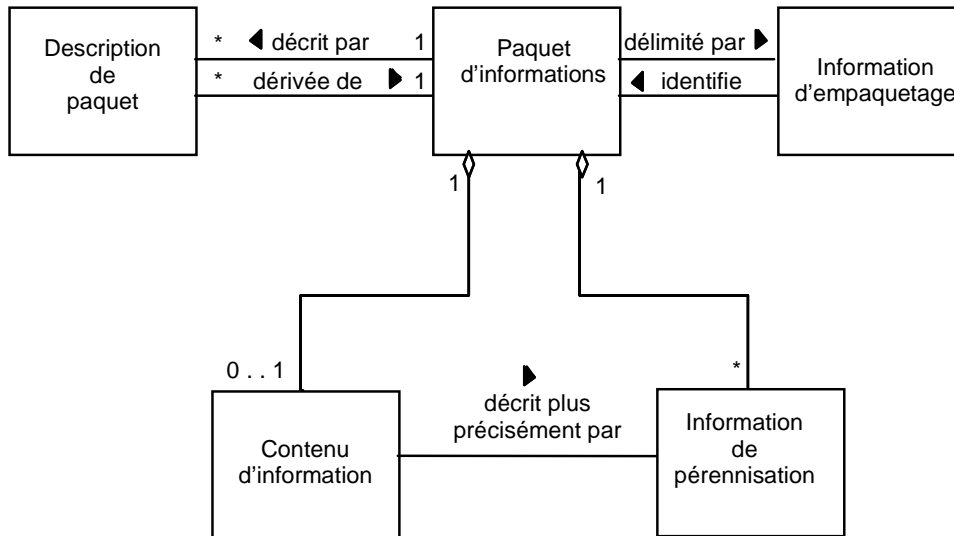
conservation d'information, l'OAIS doit fournir des dispositifs adéquats pour permettre aux Utilisateurs d'identifier les informations qui peuvent les intéresser, de les analyser et de les commander. Ceci est possible grâce à un Objet-information spécifique nommé Information de description, et qui contient les données d'accès aux documents ou aux applications appelées **Outils d'accès**. Les Informations de description sont en général extraites du Contenu d'information et du PDI. L'Information de description peut être vue comme un index permettant un accès efficace au Paquet d'informations associé via des Outils d'accès. Les Outils d'accès sont des documents ou des applications qui peuvent être utilisées pour localiser, analyser, récupérer ou commander de l'information à partir d'un OAIS.

#### **4.2.2 MODELE LOGIQUE D'INFORMATION DANS UN SYSTEME OUVERT D'ARCHIVAGE D'INFORMATION (OAIS)**

La sous-section précédente définit les types d'Objets-information nécessaires à un OAIS pour permettre la Pérennisation de l'information et un accès efficace aux informations conservées par la Communauté cible. Cette sous-section utilise les descriptions de l'Objet-information pour modéliser les structures conceptuelles de l'information requises pour ces fonctions. Les modèles présentés dans cette sous-section ne sont pas destinés à une mise en œuvre, mais plutôt à souligner la relation entre les différents types d'informations requis dans un processus d'archivage.

##### **4.2.2.1 Paquet d'informations**

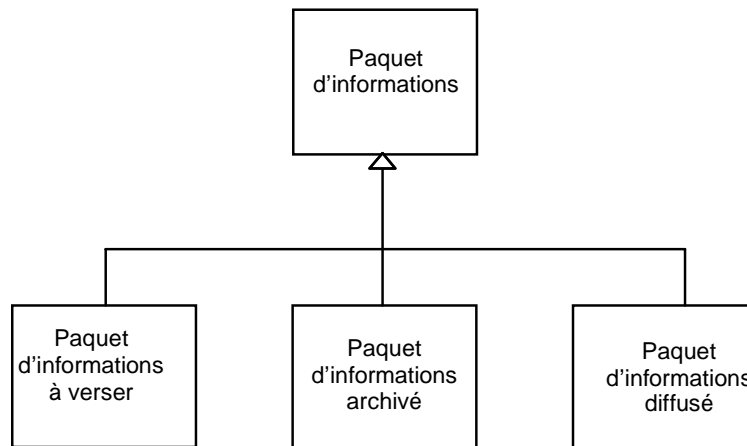
La structure conceptuelle qui permet la Pérennisation d'information est le Paquet d'informations. Un Paquet d'informations est une enveloppe qui contient deux types d'Objet-information : le Contenu d'information et l'Information de pérennisation (PDI). Le Paquet d'informations peut être associé à deux autres types d'Objet-information, l'Information d'empaquetage et les Descriptions de paquet. Il existe plusieurs types de Paquet d'informations utilisés dans le processus d'archivage. Ces Paquets d'informations peuvent être utilisés pour structurer et stocker les fonds de l'OAIS, pour transporter l'information requise du Producteur vers l'OAIS, ou pour transporter l'information demandée par les Utilisateurs à l'OAIS. Pour chacune de ces fonctions, l'information doit se conformer à des exigences distinctes. Le diagramme UML dans le schéma 4-13 illustre la vue conceptuelle d'un Paquet d'informations. Ce diagramme UML montre qu'un Paquet d'informations contient 0 ou 1 Objet-contenu d'information, 0 ou plus objets PDI et est associé exactement à une Information d'empaquetage, qui identifie et délimite le Paquet d'informations. Le Paquet d'informations est aussi associé à 0 ou plus Descriptions de paquet qui décrivent l'Objet-contenu pour permettre d'y accéder efficacement.



**Schéma 4-13 : Contenu d'un Paquet d'informations**

#### 4.2.2.2 Types de Paquets d'Informations

Il existe trois catégories de Paquets d'informations identifiées en 2.2 : le Paquet d'informations à verser (SIP), le Paquet d'informations archivé (AIP) et le Paquet d'informations diffusé (DIP). Les définitions de ces types de paquets en section 2 sont basées sur le processus fonctionnel de l'archivage au cours duquel le paquet et le passage du paquet d'un état à un autre sont pris en compte. Cette taxinomie des types de Paquet d'informations est illustrée par le schéma 4-14.



**Schéma 4-14 : Taxinomie du Paquet d'informations**

Il est nécessaire de distinguer entre les Paquets d'informations conservés par un OAIS et les Paquets d'informations à verser ou les Paquets d'informations diffusés par un OAIS. Ces différents paquets sont nécessaires afin de traduire le fait que certains versements à un OAIS

auront une Information de représentation ou un PDI insuffisant pour satisfaire aux exigences ultimes de conservation. De plus, l'organisation des paquets peut être très différente de l'organisation des informations au sein de l'OAIS. Au final, l'OAIS risquerait de diffuser aux Utilisateurs des Contenus d'information avec une Information de représentation incomplète ou un PDI incomplet. On distingue donc : le Paquet d'informations à verser (SIP), le Paquet d'informations archivé (AIP) et le Paquet d'informations diffusé (DIP). Bien que ceux-ci soient tous des Paquets d'Information, le contenu obligatoire et la multiplicité des associations entre les classes d'information qu'ils contiennent diffèrent selon les catégories.

**Le Paquet d'informations à verser (SIP)** est le paquet envoyé à un OAIS par un Producteur. Sa forme et son contenu détaillé sont généralement négociés entre le Producteur et l'OAIS. La plupart des SIP auront un Contenu d'information et un PDI, mais il faut parfois plusieurs SIP pour constituer un jeu complet de Contenu d'information et un PDI associé. Le Contenu d'information et le PDI ont tous deux une Information de représentation qui leur est associée, et, s'il y a plusieurs SIP qui utilisent les mêmes Informations de représentation, il est probable que l'on fournira une seule fois de telles Informations de représentation à l'OAIS. Autre variante, puisque certaines catégories de PDI s'appliqueront à des ensembles de SIP de même origine, on peut fournir ce PDI dans un SIP séparé qui ne contiendra pas de Contenu d'information. L'Information d'emballage sera toujours présente sous une forme quelconque.

L'Information de description associée à un SIP sera probablement fournie avant le versement du SIP à l'OAIS, mais elle peut en pratique être fournie à tout moment. Ce peut être une simple description textuelle avec un nom ou un titre, insérée dans l'Information d'emballage, qui permet de reconnaître le SIP.

Au sein de l'OAIS, un ou plusieurs SIP sont transformés en un ou plusieurs Paquets d'informations archivés (AIP) pour la conservation. L'AIP contient un jeu complet de PDI pour le Contenu d'information auquel il se rapporte. L'AIP peut contenir une collection d'autres AIP ; ce point est traité et modélisé plus loin dans cette sous-section. L'Information d'emballage de l'AIP doit se conformer aux standards internes de l'OAIS et peut varier dans le cadre de sa gestion par l'OAIS. L'Information de description d'un AIP, qui peut être conséquente, sera gérée par l'OAIS de telle manière que l'Utilisateur puisse trouver et commander le Contenu d'information qui l'intéresse.

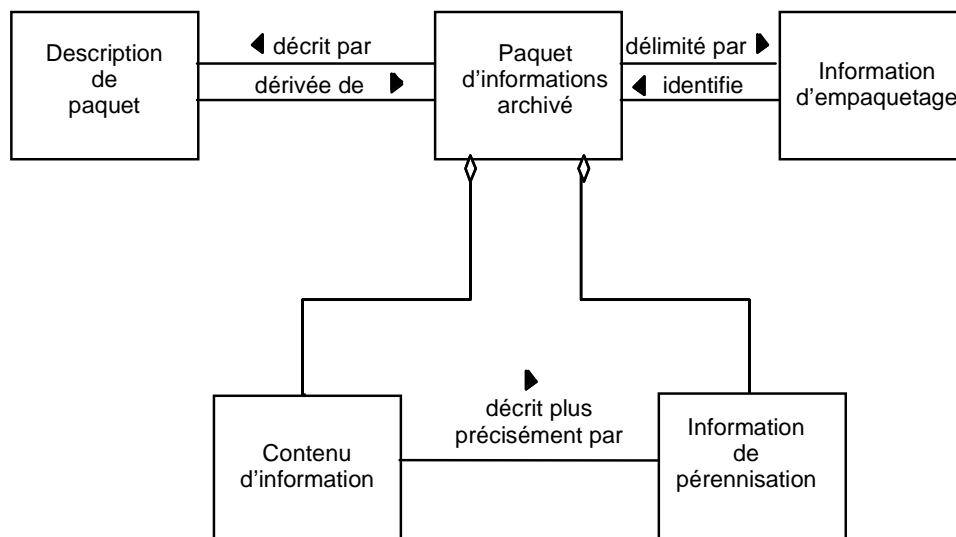
En réponse à une Commande, l'OAIS fournit tout ou partie d'un AIP à un Utilisateur sous la forme d'un Paquet d'informations diffusé (DIP). Le DIP peut comporter des collections d'AIP et avoir ou non un PDI complet. L'Information d'emballage sera toujours présente d'une façon ou d'une autre, de sorte que l'Utilisateur puisse clairement retrouver les informations qu'il a commandées. Les Informations d'emballage peuvent prendre plusieurs formes selon les supports de diffusion et les exigences de l'Utilisateur. On peut fournir l'Information de description d'un DIP en même temps que le SIP, mais aussi avant ou après le transfert. Son rôle est de donner à l'Utilisateur l'information suffisante pour reconnaître le DIP parmi d'éventuels paquets semblables. Cela peut se résumer à une description textuelle avec un nom ou un titre au sein des Informations d'emballage, par laquelle le DIP peut être reconnu.

Bien que la manière de construire l'AIP puisse varier d'une Archive à l'autre, la spécification de l'AIP en tant que conteneur rassemblant toute l'information nécessaire à la Pérennisation et à l'accès aux fonds archivés demeure valable. Le modèle d'information de l'AIP présenté en 4.2.2.3 devrait être utilisé comme référence pour établir les types d'informations exigées aux fins de Pérennisation et d'accès.

Les contenus exacts d'information du SIP et du DIP et leurs relations avec l'AIP correspondant dépendent des accords entre l'Archive, ses Producteurs et ses Utilisateurs. Le modèle pour ces deux paquets est le même que celui du Paquet d'informations décrit dans le schéma 4-13, tant pour le contenu obligatoire que pour la multiplicité des associations entre les classes d'informations qu'il contient. Les transformations entre SIP et AIP et entre AIP et DIP sont traitées plus en détail en 4.3.

#### 4.2.2.3 Le Paquet d'informations archivé

Le Paquet d'informations archivé (AIP), modélisé dans le schéma 4-15, est un Paquet d'informations spécifique. L'AIP est conçu pour fournir un moyen concis de référence à un ensemble d'informations disposant, en principe, de toutes les qualités requises pour la Pérennisation, permanente ou sur une période indéfinie, d'un Objet-information cible. L'AIP est lui-même un Objet-information, conteneur d'autres Objets-information. Dans l'AIP réside l'Objet-information cible et il est appelé Contenu d'information.

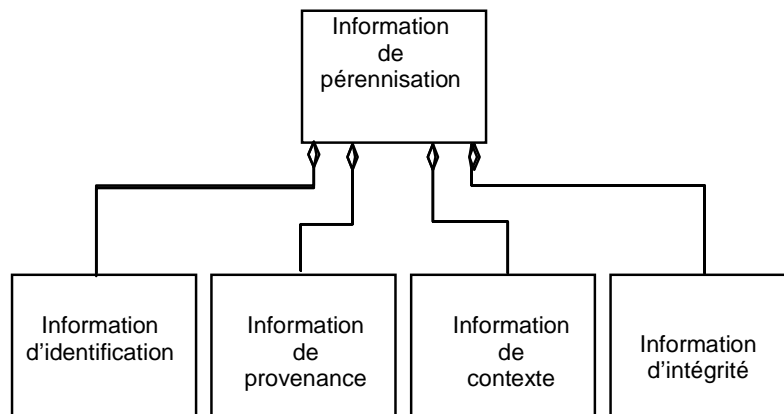


**Schéma 4-15 : Paquet d'informations archivé (AIP)**

L'AIP contient également un Objet-information appelé Information de pérennisation (PDI). Le PDI renferme des informations complémentaires sur le Contenu d'information : il est indispensable pour donner tout son sens au Contenu d'information dans l'optique de la Pérennisation sur une durée indéfinie.

Les exigences applicables à l'Information de pérennisation dans un AIP sont beaucoup plus contraignantes que les exigences applicables à l'Information de pérennisation d'un Paquet d'informations général. Tandis qu'aucun objet PDI n'est obligatoire dans un Paquet d'informations, toutes les classes d'informations PDI doivent être présentes dans un AIP. Ceci est illustré par le schéma 4-16. Le contenu de chaque type de PDI est laissé à la discrétion de chaque Archive.

Par exemple, dans certains fonds OAIS un texte mentionnant que le créateur du Contenu d'information est inconnu peut constituer une Information de provenance suffisante tandis que dans d'autres fonds OAIS il peut être obligatoire que la provenance soit être précisément renseignée.



**Schéma 4-16 : Information de pérennisation (PDI)**

L'AIP est délimité et identifié par les Informations d'empaquetage. Les Informations d'empaquetage peuvent être réellement présentes dans la structure du support qui contient l'AIP ou virtuelles, en ce sens qu'elles sont contenues dans l'Entité « Stockage » de l'OAIS. Cependant, l'OAIS doit définir clairement les fonctions de délimitation et d'identification interne.

Chaque AIP est associé à une forme structurée d'Information de description appelée **Description de paquet**, qui permet à l'Utilisateur de localiser l'information qui peut l'intéresser, d'analyser cette information et de commander l'information souhaitée. L'information nécessaire pour un Outil d'accès est appelée **Notice descriptive**. Une Description de paquet simple peut contenir plusieurs Notices descriptives selon le nombre d'Outils d'accès différents pour la localisation, la visualisation, la récupération ou la commande d'un Contenu d'information et de son PDI. Le schéma 4-17 présente un diagramme UML qui modélise la Description de paquet et l'Outil d'accès.



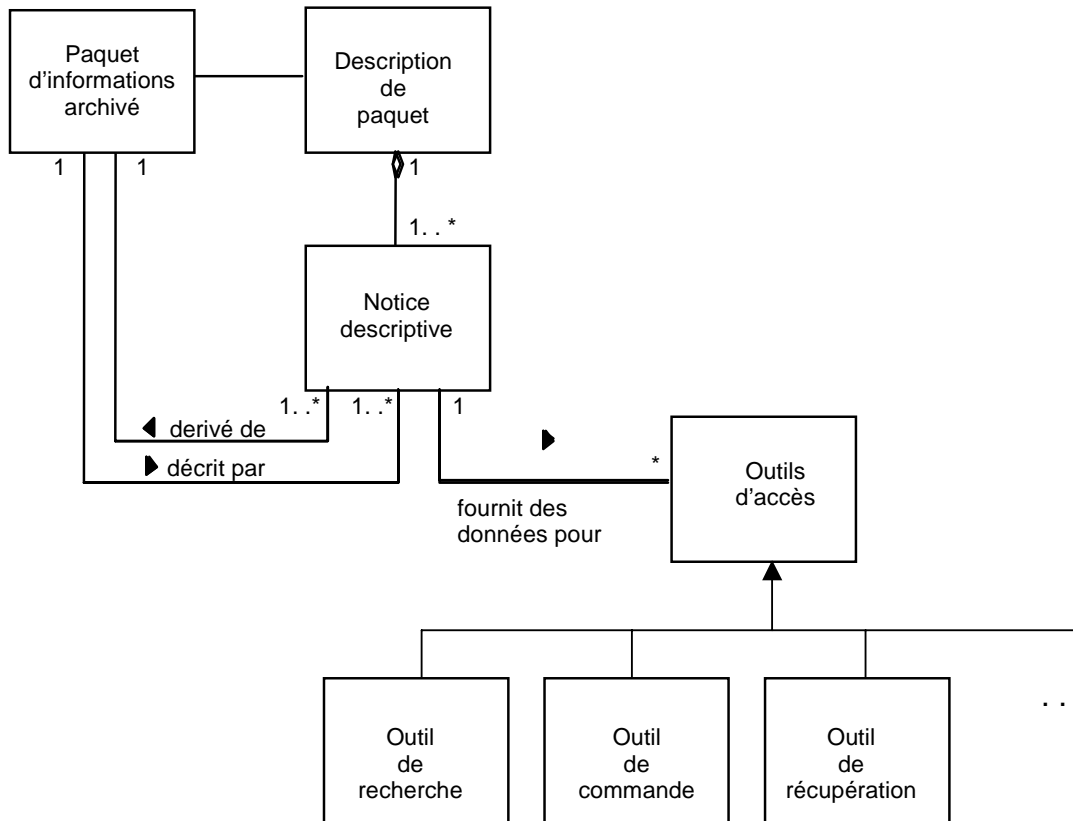


Schéma 4-17 : Description de paquet

La Description de paquet doit contenir une Notice descriptive qui fournit des données à un Outil de récupération permettant aux utilisateurs autorisés de récupérer le Contenu d'information et le PDI décrit dans la Description de paquet en question. Cet Outil de récupération fait généralement partie du domaine fonctionnel de l'Entité « Stockage ». Il utilise l'identificateur unique attribué à l'AIP par l'OAIS pour définir un ensemble d'opérations et de noms de fichiers nécessaires pour extraire l'AIP du système de gestion de fichier utilisé dans l'Entité « Stockage », puis restitue le Contenu d'information et le PDI pour l'AIP demandé. Dans la plupart des Archives actuelles, seuls les personnels et fonctions en charge du fonctionnement et des traitements internes à l'Archive sont autorisés à utiliser cet Outil d'accès. Cependant, comme les avancées technologiques augmentent la puissance de traitement des Archives et la bande passante entre Archives et utilisateurs, des méthodes d'accès telles que « les requêtes basées sur le contenu » et « la fouille de données » peuvent fournir à l'Utilisateur un accès direct en lecture seule au Contenu d'information.

La Description de paquet peut également contenir un certain nombre de Notices descriptives, chacune contenant des données pour un ou plusieurs Outils d'accès. L'**Outil de recherche** et l'**Outil de commande** constituent deux catégories d'outil d'accès.

Un Outil de recherche est une application qui aide l'Utilisateur à localiser une information intéressante. Un AIP unique peut disposer d'un certain nombre de Notices descriptives qui décrivent le Contenu d'information à l'aide de différentes technologies.

Un Outil de commande est une application qui renseigne l'Utilisateur sur les tarifs et lui permet de passer commande des AIP qui l'intéressent. L'Outil de commande permet aussi aux utilisateurs de spécifier des modifications de l'AIP avant diffusion. Parmi ces transformations, on trouve des transformations d'Objets-données comme l'extraction de sous-ensembles, le sous-échantillonnage ou la conversion de format. Les transformations peuvent aussi induire la modification du PDI dans l'AIP avant diffusion.

La Description de paquet n'est pas exigée pour la Pérennisation du Contenu d'information, mais est nécessaire pour donner une visibilité et un accès aux contenus d'une Archive. Le contenu de la Description de paquet dépend fortement de la structure du Contenu d'information et du PDI qu'il décrit. Les utilisations et les types de Descriptions de paquet dans un OAIS sont définis plus en détail en 4.2.2.4.

Le schéma 4-18 présente une vue détaillée du Paquet d'informations archivé en développant le PDI et le Contenu d'information. Toutes les relations du type "contient" abordées dans cette sous-section sont des relations de contenu logique. Ce type de relation faisant référence à ce qui est contenu peut être physique ou réalisée via un pointeur vers un autre objet dans le stockage.

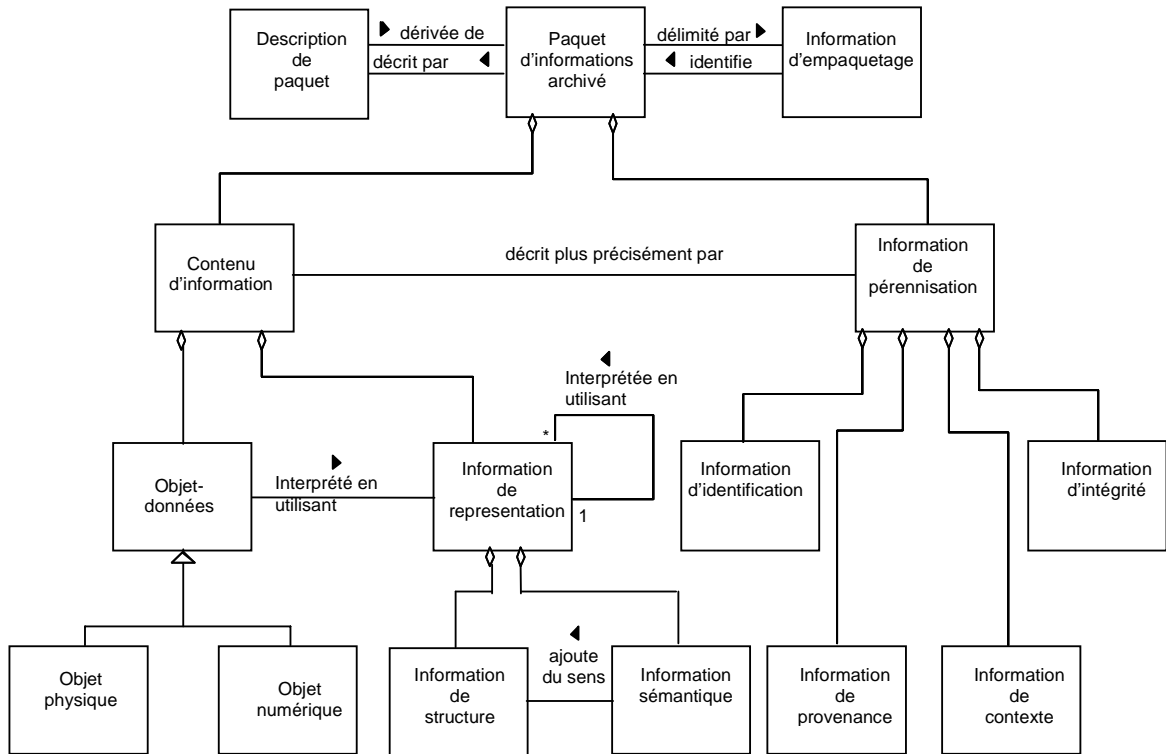


Schéma 4-18 : Paquet d'informations archivé (Vue détaillée)

#### 4.2.2.4 Spécialisation de l'AIP et de la Description de paquet

Cette sous-section aborde deux spécialisations de l'AIP qui sont l'**Unité d'informations archivée (AIU)** et la **Collection d'informations archivée (AIC)**. Le schéma 4-19 est un diagramme UML illustrant cette spécialisation. L'AIU et l'AIC sont des catégories de l'AIP et comme telles contiennent un ensemble structuré d'éléments permettant tout à la fois la Pérennisation et l'accès des Utilisateurs. L'AIU est le type utilisé pour la fonction de conservation d'un Objet-contenu d'information élémentaire. L'AIC organise un ensemble d'AIP (AIU et autres AIC) selon une hiérarchie thématique, qui permet un accès souple et efficace par la Communauté d'utilisateurs. Conceptuellement, tous les AIP organisés par un AIC sont englobés dans le Contenu d'information de cet AIC. Les différences entre AIU et AIC résident dans la complexité de leur Contenu d'information, de leurs Descriptions de paquets et de leurs Informations d'empaquetage. Le présent Modèle de référence considère que les différences au niveau du Contenu d'information et des fonctions associées d'Empaquetage et de Description entre AIU et AIC sont raisonnablement complexes et articulées pour justifier la définition de classes séparées.

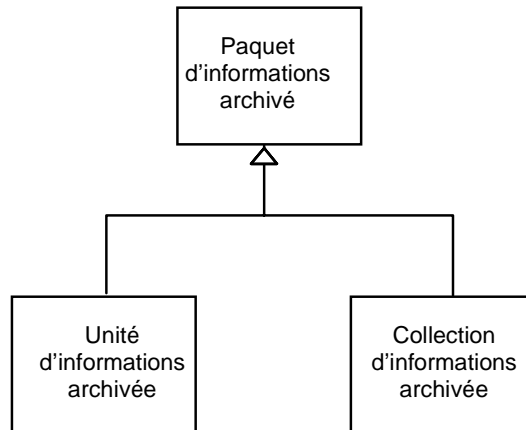
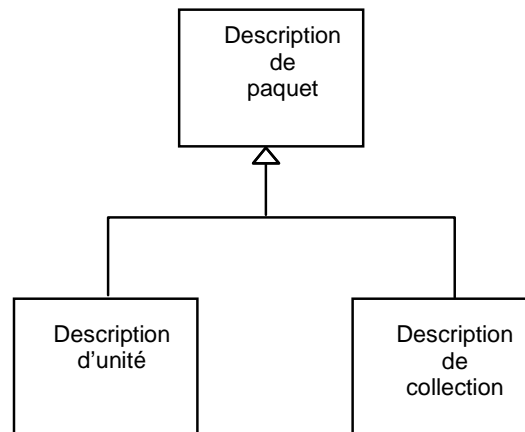


Schéma 4-19 : Spécialisations de l'AIP

Du point de vue de l'Entité « Accès », les nouvelles capacités d'extraction de sous-ensembles et de manipulation des données ont tendance à brouiller la distinction entre AIC et AIU. Les Objets-contenu qui sont habituellement vus comme élémentaires peuvent maintenant être vus comme contenant une grande variété de contenus basés sur les paramètres choisis pour définir ces sous-ensembles. A la limite, le Contenu d'information d'un AIU peut ne pas exister en tant qu'entité physique. Le Contenu d'information pourrait consister en plusieurs fichiers d'entrée (ou pointeurs vers les AIP contenant ces fichiers de données) et un algorithme qui utilise ces fichiers pour créer l'Objet-données intéressant.

Du point de vue de la conservation de l'information, la distinction entre AIU et AIC reste claire. Un AIU est vu comme ayant un seul Objet-contenu d'information qui est décrit par un seul jeu de PDI. Le Contenu d'information d'un AIC est vu comme une collection d'autres AIC et AIU, dont chacun a son propre PDI. De plus, l'AIC a lui-même son propre PDI qui décrit les critères et les traitements applicables à la Collection.

Il y a deux spécialisations de la Description de paquet : la **Description d'unité** et la **Description de collection**. Le schéma 4-20 est un diagramme UML illustrant cette spécialisation. La distinction entre ces deux classes est basée sur la différence entre les fonctions nécessaires pour accéder au contenu d'un AIU élémentaire et les fonctions nécessaires pour accéder aux AIP qui sont contenus dans un AIC.



**Schéma 4-20 : Spécialisation de la Description de paquet**

Pour faciliter la compréhension de ces classes, les deux sous-sections suivantes s'appuient sur l'exemple d'une société fondant un OAIS pour l'archivage numérique de films. Cet exemple est centré sur le Contenu d'information au sein des structures de l'AIP. La sous-section 4.3 illustre plus en détail les transformations de l'information et les flux de données dans un OAIS.

#### 4.2.2.5 Unité d'Informations Archivée

Les AIU peuvent être vus comme les « atomes » d'information que les Archives ont vocation à stocker. Un AIU contient un et un seul Objet-contenu d'information (lequel peut consister en plusieurs fichiers) et un seul jeu de PDI. Quand un Objet-information est versé dans l'OAIS, une Description d'unité, qui est une catégorie de la Description de paquet, est créée en extrayant les informations du Contenu d'information et du PDI et en y ajoutant des informations spécifiques définies par l'OAIS comme un identificateur unique. L'AIU est illustré par le schéma 4-21.

Dans l'exemple de l'OAIS qui archive des films numériques, l'AIU d'un film donné peut être vu comme trois objets, l'un contenant un codage numérique du film dans un format propriétaire, un autre contenant l'Information de représentation nécessaire à la compréhension du format propriétaire (ces deux objets forment le Contenu d'information) et le dernier contenant des données factuelles à propos du film comme la date de création, les acteurs, le réalisateur, le producteur, le montage, le studio de tournage et une clef de contrôle pour assurer l'intégrité du film numérisé (le PDI).

Dans la mesure où le Modèle de référence OAIS est indépendant des mises en œuvre, chacun de ces objets pourrait prendre la forme d'un seul ou de plusieurs fichiers. Cette information liée à la mise en œuvre est contenue dans l'Information d'empaquetage. Quand un film est versé à l'OAIS, une Description d'unité peut être créée pour un Outil de commande en extrayant des informations du Contenu d'information et du PDI et en constituant ainsi une information de commande unique pour ce film.

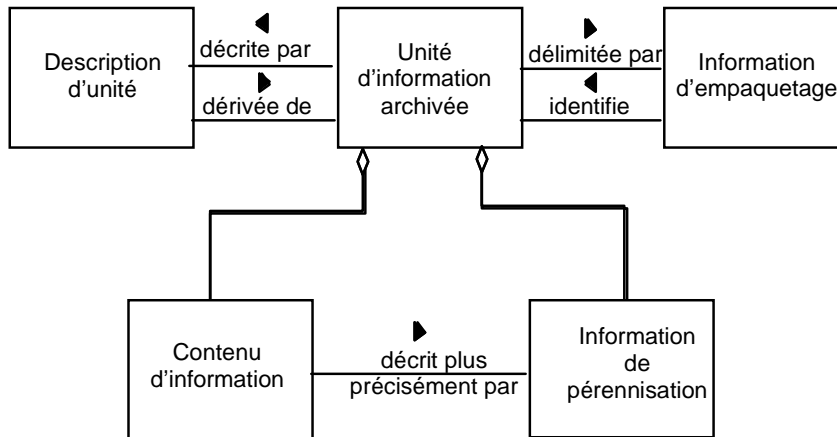


Schéma 4-21 : Unité d'informations archivée (AIU)

#### 4.2.2.6 Description d'Unité

La Description d'unité est une spécialisation de la Description de paquet. Elle contient un ensemble particulier de Notices. Le schéma 4-22 est un diagramme UML qui illustre le contenu de la Description d'unité.

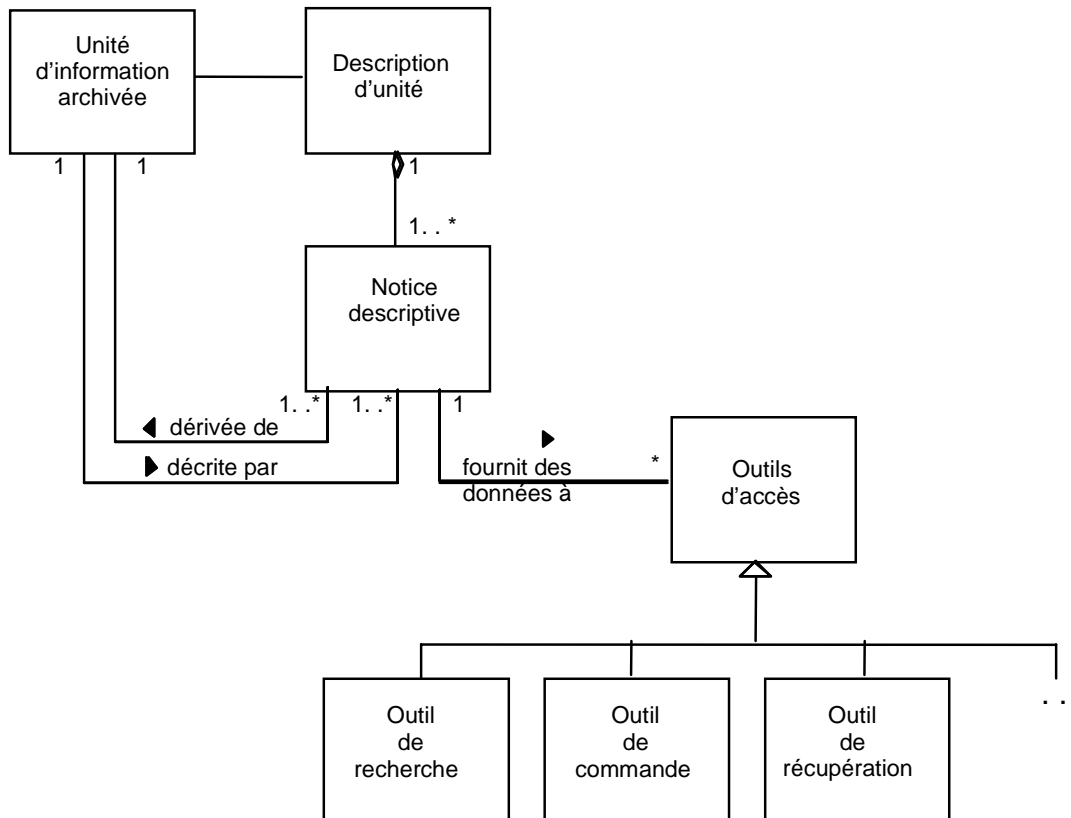


Schéma 4-22 : Description d'unité

Toutes les Descriptions d'unité doivent offrir une Notice descriptive destinée à un Outil de récupération. Cet outil permet aux utilisateurs autorisés de récupérer, depuis l'Entité « Stockage », l'AIU décrit par la Description d'unité. Cette description contient l'identificateur unique affecté à l'AIP par l'Entité « Stockage » pendant le processus de versement.

L'Outil de recherche constitue un type important d'Outil d'accès. Cet outil est une application qui aide l'Utilisateur à repérer les informations qui l'intéressent. Un AIU unique peut comporter un certain nombre de Notices descriptives décrivant le Contenu d'information en recourant à diverses technologies. De plus, avec l'émergence de nouvelles technologies d'extraction, de description et d'affichage, une Archive peut vouloir mettre à jour la Description d'unité associée à chacun de ses AIU, afin d'ajouter une nouvelle Notice descriptive qui utilise la nouvelle technologie permettant de mieux décrire les AIU.

Dans l'exemple de l'OAIS de films numériques, admettons qu'il y ait au départ trois Notices descriptives : une description en texte libre d'un film, un clip de cinq minutes et un enregistrement dans une base de données relationnelle utilisée par des collectionneurs de films pour rechercher des films. Après une certaine période de fonctionnement de l'Archive, une technique de compression des films numériques basée sur l'enregistrement d'une image sur dix a pu être développée. L'archiviste peut alors décider de créer un autre type de Notice descriptive en utilisant cette nouvelle technique. S'il le juge utile, il peut parcourir tous les AIU conservés dans l'Archive avec cette technique de compression et créer une nouvelle Notice descriptive pour chaque film, ou plus simplement inclure cette Notice descriptive pour les nouveaux AIU versés dans l'OAIS.

Une autre classe importante de Notices descriptives fournit des données aux Outils de commande, qui renseignent l'Utilisateur sur le tarif des AIU qui l'intéressent et lui permettent de les commander. L'Outil de commande permet aussi aux utilisateurs de spécifier des transformations à appliquer aux AIU avant diffusion. Parmi ces transformations, on trouve des transformations d'Objet-données comme la constitution de sous-ensembles, le sous-échantillonnage ou des transformations de format. Les transformations peuvent aussi impliquer la modification du PDI dans l'AIU avant diffusion.

Par exemple, l'OAIS pourrait permettre à un Utilisateur de commander un film numérique sous forme de cassette VHS, de disque laser ou d'objet MPEG diffusé en ligne. Chacun d'entre eux impliquerait une transformation de format et, en théorie, une mise à jour du PDI au sein de l'AIP pour créer des PDI dédiés à chaque DIP.

#### **4.2.2.7 Collection d'Informations Archivée**

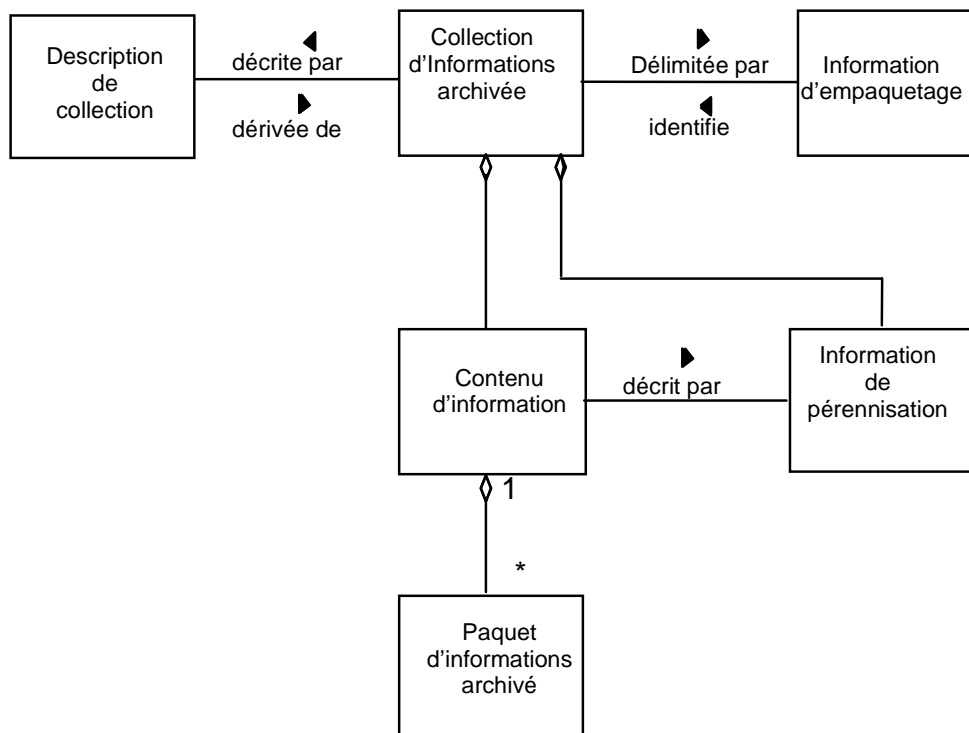
L'AIU et la Description d'unité qui lui est associée fournissent l'information nécessaire à un Utilisateur pour localiser et commander les AIU qui l'intéressent. Toutefois, il peut s'avérer impossible pour un Utilisateur de faire un tri parmi les millions de Descriptions d'unité qui sont contenues dans de grandes Archives. C'est le problème traité ici.

Le Contenu d'information d'un AIC est composé d'AIP complets ayant chacun son propre Contenu d'information, son PDI et l'Information d'empaquetage associée ainsi que les

Descriptions de paquets. Ces AIP sont alors agrégés en Collections d'informations archivées (AIC) sur la base de critères définis par l'archiviste. Le plus souvent les AIC sont basés sur des AIU intéressants relevant de la même thématique ou ayant des origines et un jeu commun de Notices descriptives. L'OAIS tout entier peut être vu comme le détenteur d'au moins un AIC qui contiendrait tous les AIP de l'OAIS.

Dans l'exemple des films numériques, l'OAIS peut définir des AIC basés sur la classification des films : le mystère, la science-fiction, l'horreur, etc. L'Archive peut aussi définir des AIC basés sur d'autres critères tels que le réalisateur ou l'acteur principal.

Un modèle logique d'AIC est présenté sur le schéma 4-23. Comme dans les sous-sections précédentes, toutes les relations de contenu sont logiques et peuvent être physiques ou réalisées via un pointeur vers un autre objet dans le stockage. Par exemple, le Contenu d'information d'un AIC peut être obtenu soit en créant les collections physiques des AIP soit en pointant vers les AIP. Un AIP unique peut appartenir à un nombre indéterminé d'AIC.



**Schéma 4-23 : Vue logique d'une Collection d'informations archivée**

Une technique de reconnaissance de formes pourrait être créée pour les films numériques et dans ce cas l'OAIS offrirait un service de recherche dans ses Archives permettant d'identifier de grandes structures comme des pyramides ou un profil de New York. Notons que ce type de



service exige une grande puissance de calcul et implique le transfert éventuel d'un grand nombre d'AIU de l'Entité « Stockage » vers l'Entité « Accès », afin de pouvoir piloter de là le processus d'analyse du Contenu d'information de chaque AIU. Si cette approche donne de bons résultats, l'archiviste peut alors synthétiser les résultats de cette « requête basée sur un contenu » dans la Notice descriptive d'un nouvel AIC qui contiendra des films montrant de grandes structures. Cette technique est généralement connue sous le nom de « fouille de données ».

Le fait qu'un AIC est un AIP complet avec son PDI est une caractéristique importante de l'AIC, comme on le voit dans le schéma 4-23. Le PDI fournit des informations globales sur l'AIC telles que la provenance (quand et pourquoi il a été créé), le contexte des AIC avec lesquels il est en relation et le niveau souhaité d'Information d'intégrité et de sécurité. Ce PDI de l'AIC vient donc en complément du PDI contenu dans les AIP membres. Ce type d'informations est souvent nécessaire pour rassurer l'Utilisateur sur la fiabilité d'un AIC. Dans l'exemple ci-dessus, l'utilité de l'AIC pour des films traitant de grandes structures est dans une certaine mesure basée sur l'algorithme utilisé mais aussi sur l'Information de provenance (date de création et de la dernière mise à jour).

#### **4.2.2.8 Description de Collection**

La Description de collection est une sous-catégorie de la Description de paquet, enrichie d'éléments facilitant la prise en compte du Contenu d'information complexe d'un AIC. La Description de collection est modélisée dans le schéma 4-24, elle contient les classes d'informations contenues dans la Description d'unité.

Il existe deux types de Notice descriptive dans une Description de collection :

- Une Description globale qui décrit la collection comme un tout.
- Aucune, une ou plusieurs Descriptions d'éléments qui décrivent séparément chaque élément de la collection.

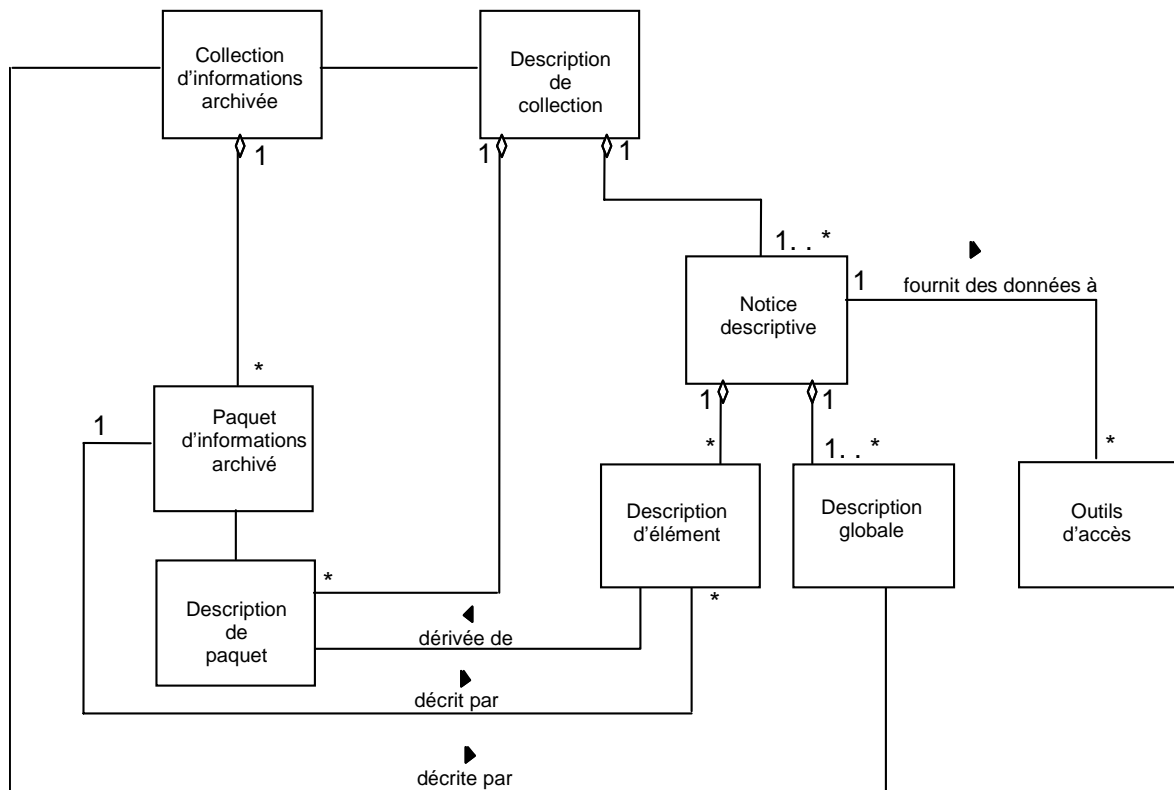


Schéma 4-24 : Description de collection

La Notice descriptive de la Description de collection fournit des informations aux Outils de commande qui donnent à un Utilisateur l'accès au Contenu d'information de l'AIC dans sa totalité ainsi qu'au PDI de cet AIC, mais pas nécessairement à chacun des AIP contenus dans l'AIC. La Description de collection peut, quant à elle, contenir les Descriptions de paquets des AIP contenus dans l'AIC. Cette relation de contenu est logique en cela que l'AIC peut effectivement contenir les Descriptions de paquets des Paquets d'informations élémentaires ou, ce qui est plus fréquent, utiliser des pointeurs vers les Descriptions de paquets de ces Paquets d'informations élémentaires. La liste des Descriptions de paquets des AIP contenus dans un AIC peut fournir aux Outils d'accès une méthode pour la récupération ou la commande des éléments de l'AIC.

Cela rend également possible la mise en œuvre d'Outils de recherche alternatifs qui permettent à l'Utilisateur d'identifier des AIP qui l'intéressent au sein d'un AIC. Les Notices descriptives qui fournissent des données pour ces Outils de recherche peuvent être mises en œuvre de manière centralisée (au moyen de la Notice descriptive dans la Description de collection) ou d'une manière distribuée (à l'aide de la Notice descriptive de chaque Description de paquet élémentaire).

Un autre avantage important de la Description de collection réside dans la possibilité de définir de nouvelles **Collections virtuelles**. Une Collection virtuelle peut être basée sur de nouveaux résultats de fouille de données ; elle peut également refléter des événements d'actualité ou des domaines d'intérêt temporaires. Dans un OAIS de films numériques, une Collection virtuelle seraient par exemple une collection constituée par les nouveautés ou une collection intitulée « Les vingt titres les plus populaires », mise à jour périodiquement. Un autre exemple de Collection virtuelle serait une collection basée sur les résultats d'un algorithme de reconnaissance de formes qui n'a pas encore été vérifié.

Pour créer une Collection virtuelle, les Archives peuvent créer une Description de collection qui n'est associée à aucun AIC. La Description de collection peut alors intégrer une Notice descriptive du Contenu nouvellement enrichi de chaque AIP membre. Un Outil de recherche spécifique pourrait utiliser cette nouvelle Notice descriptive d'élément en liaison avec les Notices descriptives présentes dans les informations de Description de paquet de chaque AIP élémentaire afin de localiser les AIP intéressants pour l'Utilisateur. Les Descriptions de paquets des AIP contenus pourraient aussi fournir des données pour un Outil de commande qui permettrait à l'Utilisateur de commander le Paquet d'informations qui l'intéresse.

Si un OAIS décide qu'une Collection virtuelle possède une valeur suffisante pour justifier sa Pérennisation, il peut stocker le Contenu d'information et le PDI requis par l'Entité « Stockage », créant ainsi un nouvel AIC.

Une autre application importante des Collections virtuelles consiste à localiser des éléments d'une collection dont le versement est prévu dans le futur. Dans ce cas, les Notices descriptives prises en compte par un Outil de recherche permettent la localisation des futurs AIP. Cependant, la Notice descriptive pour l'Outil de commande et/ou l'Outil de récupération indique que ce produit n'est pas encore disponible et permet à l'Utilisateur de saisir une Demande d'abonnement qui sera déclenchée lorsque l'AIP en question sera disponible.

### **4.2.3 INFORMATION DE GESTION DES DONNEES**

Généralement les Descriptions de paquets sont archivées dans un système de stockage permanent comme par exemple un système de gestion de base de données afin de permettre un accès et une mise à jour aisés et souples des Notices descriptives qui s'y trouvent. Outre les Descriptions de paquets traitées dans les précédentes sous-sections, toute l'information nécessaire pour le fonctionnement d'une Archive peut être stockée dans des bases de données, en tant que classes de données permanentes. Le schéma 4-25 illustre les différentes catégories "d'informations de gestion des données" au sein d'un OAIS. L'information de l'Entité « Administration » de l'Archive couvre l'ensemble de l'information nécessaire au fonctionnement quotidien de l'Archive. Cette information comprend :

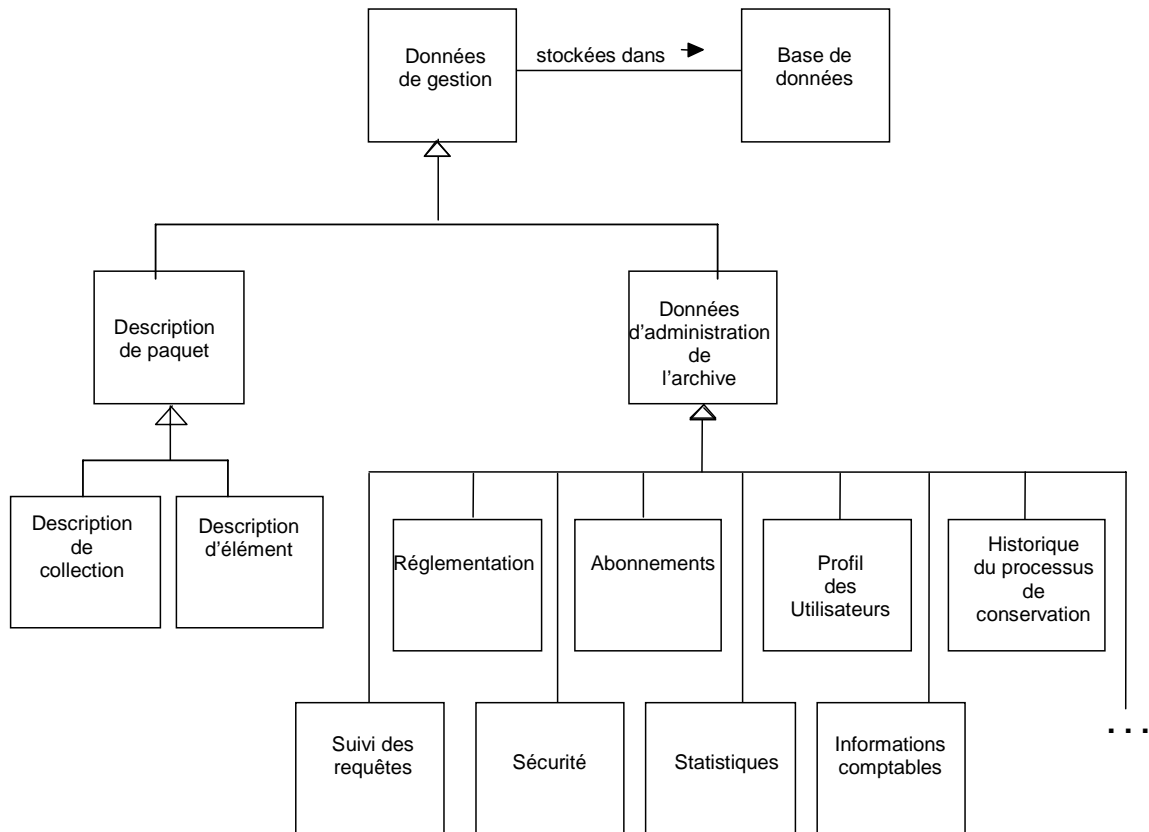
- les informations relatives à la politique tarifaire et aux contraintes de disponibilité pour la commande d'informations archivées,
- le suivi des requêtes qui enregistre l'état d'avancement de chaque transaction entre l'Utilisateur et l'Archive. Le processus de suivi des demandes peut être assez

compliqué, et impliquer des événements et des déclencheurs au sein de la base de données. Ce processus peut aussi prendre la forme d'un simple fichier qui trace les requêtes de commandes,

- les informations de sécurité dont les noms des utilisateurs, les mots de passe ou autres mécanismes nécessaires pour établir l'authenticité de l'identité des utilisateurs et valider leurs privilèges,
- les informations relatives aux Demandes d'abonnement nécessaires pour permettre les commandes répétitives ou anticipées,
- les informations statistiques utiles au Management et à l'Entité « Administration » de l'Archive pour définir les procédures et optimiser les performances avec l'objectif de fonctionnement plus efficace de l'Archive. Ces informations statistiques sont par exemple la comptabilisation du nombre de commandes d'un AIP donné sur une période déterminée, ou la durée moyenne entre la réception d'une commande et l'expédition du produit demandé,
- l'historique du processus de conservation qui trace les migrations d'AIP, y compris les remplacements de supports et les transformations d'AIP,
- les profils d'Utilisateurs qui permettent à l'Archive de conserver des renseignements comme le nom et l'adresse de l'Utilisateur afin de lui éviter d'avoir à ressaisir ces données à chaque fois qu'il interroge l'Archive,
- les informations comptables et notamment les données nécessaires à l'exploitation commerciale de l'Archive, à savoir les données de registre du personnel, les données de comptes fournisseurs et les données des comptes de recette.

Ces classes sont données à titre d'exemples et non comme une liste exhaustive des données requises pour administrer l'Archive ; elles sont conceptuelles et peuvent varier de manière significative d'une Archive OAIS à l'autre lors de la mise en œuvre. Par exemple, un OAIS donné peut choisir de combiner les types d'information liés à l'Utilisateur, tels que sécurité et profil de l'Utilisateur, au sein d'une seule base de données.

Recommandation CCSDS pour un Modèle de référence OAIS

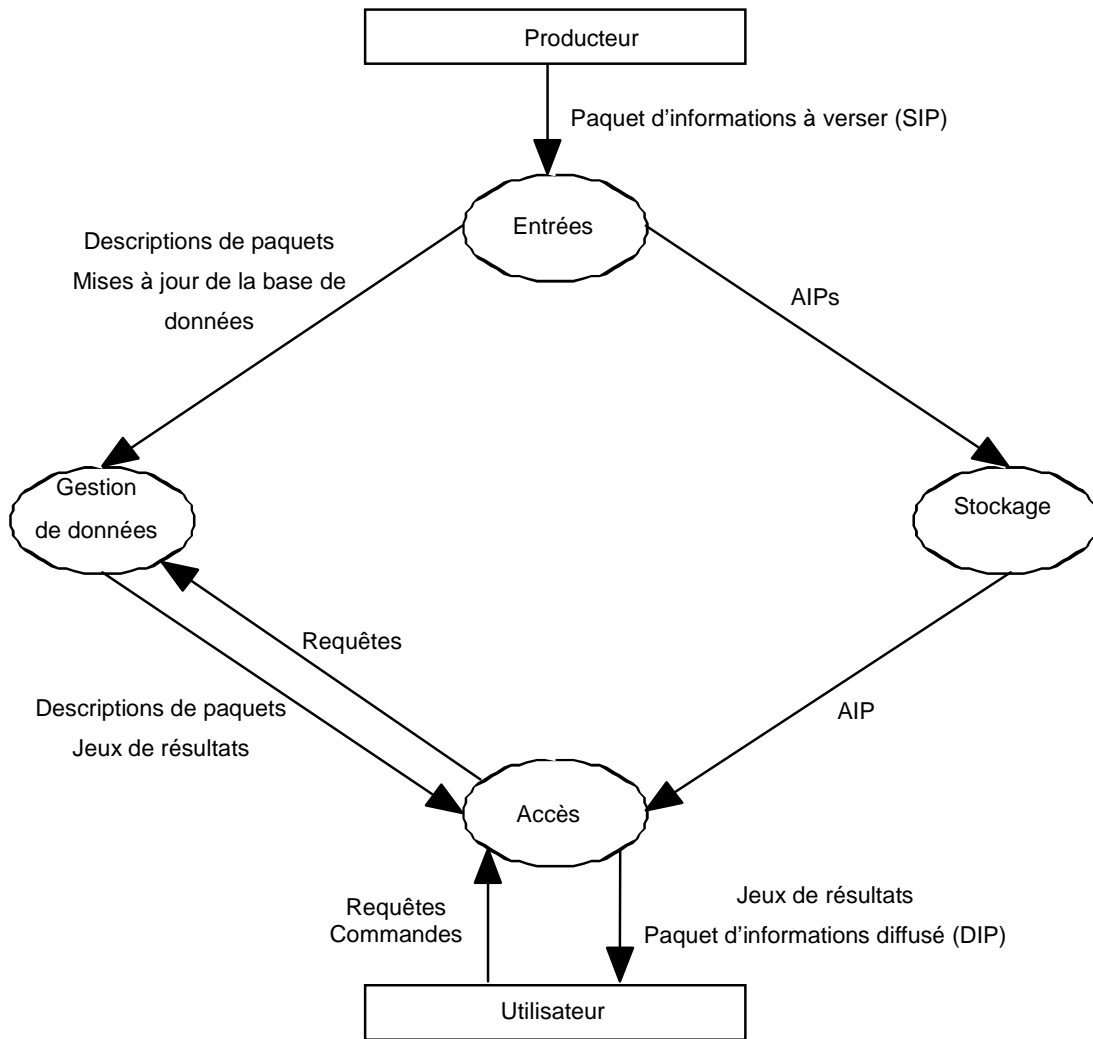


**Schéma 4-25 : Information de l'Entité « Gestion de données »**

### **4.3 TRANSFORMATIONS DE PAQUETS D'INFORMATIONS**

Les parties précédentes de cette section ont traité de l'architecture fonctionnelle d'un OAIS et d'une architecture de l'information permettant de représenter les Paquets d'informations et les Descriptions de paquets associées, ainsi que les Informations d'empaquetage. Cette sous-section aborde les transformations, tant logiques que physiques, du Paquet d'informations et des objets qui lui sont associés au cours du cycle de vie qui va du Producteur à l'OAIS et de l'OAIS à l'Utilisateur.

Le schéma 4-26 présente un diagramme de haut niveau qui décrit les principaux flux de données mis en jeu dans des opérations de l'OAIS. Ces flux n'incluent pas les flux administratifs tels que la comptabilité et la facturation.



**Schéma 4-26: Vue de haut niveau sur les flux de données dans un OAIS**

#### 4.3.1 TRANSFORMATIONS DE DONNÉES CHEZ LE PRODUCTEUR

Les données du Producteur sont privées et peuvent se présenter dans tout format souhaité par le Producteur. Cependant, quand la décision de stocker les données dans un OAIS est prise, le Producteur responsable des données rencontre les archivistes pour « négocier un Protocole de versement » comme cela a été abordé en 2.3.2. Ce protocole définit des informations telles que le contenu, le format et la date d'arrivée prévue du Paquet d'informations à verser (SIP). Un SIP est un Paquet d'informations fourni à l'OAIS par le Producteur qui consiste en un Contenu d'information auquel s'ajoutent les données nécessaires à l'OAIS pour garantir que ces Contents pourront être maintenus, interprétés dans le futur et utilisés par les Utilisateurs de l'OAIS.

Ces SIP sont périodiquement transférés vers l'OAIS par une Opération de versement. Le nombre de sessions de versement de Données entre un OAIS et un Producteur peut aller d'une simple session pour le transfert d'un produit de données final, à plusieurs sessions journalières, dans le cas d'un OAIS actif qui stocke des données pour des expériences toujours en cours. L'Opération de versement peut être logiquement vue comme des jeux d'Objets-contenu de données et d'objets de description, bien que physiquement la description - les métadonnées -, puisse figurer dans les Objets numériques eux-mêmes (objets auto décrits) ou être répartie au sein de plusieurs éléments descriptifs séparés. En complément de la vue logique de données (le SIP), la spécification d'une session de livraison de données doit aussi décrire la mise en correspondance des objets avec les supports sur lesquels ils sont livrés. Cette correspondance indique l'encodage de l'objet et de la description ainsi que l'affectation d'objets logiques aux fichiers.

#### **4.3.2 TRANSFORMATIONS DE DONNÉES DANS L'ENTITE FONCTIONNELLE « ENTREES »**

Une fois que le SIP est dans l'OAIS, sa forme et son contenu peuvent changer. Un OAIS n'a pas toujours l'obligation de conserver l'information versée exactement dans le même format que dans le SIP. En effet, conserver les informations originales exactement telles qu'elles ont été versées peut ne pas être souhaitable. Par exemple le support informatique sur lequel les images versées sont enregistrées peut devenir obsolète ; il s'avère alors nécessaire de recopier ces images sur un support plus moderne. En outre, certaines catégories d'informations, comme l'identificateur unique utilisé pour localiser le Paquet d'informations dans l'OAIS, ne seront pas accessibles au Producteur et doivent être ajoutées au moment du versement dans l'OAIS.

La relation entre SIP et AIP n'est pas une relation bijective de 1 pour 1. Voici quelques exemples :

- Un SIP – un AIP : une administration est prête à archiver ses documents électroniques de l'exercice budgétaire précédent. Tous les documents de l'année sont stockés sur bandes magnétiques et sont versés comme un seul SIP. Les Archives stockent les bandes ensemble comme un seul AIP.
- Plusieurs SIP - un AIP : un instrument spatial fait des observations de la Terre sur une durée d'un an. Chaque semaine toutes les dernières données de l'instrument sont versées à l'Archive sous la forme d'un SIP. L'Archive détient un AIP unique contenant toutes les observations du détecteur pendant l'année. L'Entité « Entrées » fusionne le Contenu d'information de chaque SIP hebdomadaire dans le ou les fichiers spécifiés au sein d'un moyen de stockage propre au système de versement. Les données du PDI pour l'AIP sont envoyées après réception des dernières données du détecteur pour l'année. Lorsque tous les SIP hebdomadaires et le SIP contenant le PDI sont arrivés, l'Entité « Entrées » traite l'AIP.
- Un SIP - plusieurs AIP : une société soumet des rapports financiers à l'Archive sous la forme d'un seul SIP. L'Archive choisit de stocker ces informations en deux AIP : l'un



contient des informations publiques et l'autre contient des informations confidentielles. Pour l'Archive, cela facilite la gestion de l'accès aux informations.

- Plusieurs SIP - plusieurs AIP : une société pétrolière et gazière collecte des informations sur ses puits. Chaque année elle verse à une Archive des SIP contenant toutes les informations relatives à l'état des puits. L'Archive maintient un AIP pour chaque champ pétrolier ou gazier et éclate les informations sur chacun des puits dans un AIP spécifique en fonction de ses coordonnées géographiques.
- Un SIP - aucun AIP : un enquêteur, ou un personnel de l'Archive, crée un nouvel algorithme pour détecter les ouragans dans des images. Il exécute cet algorithme pour toutes les images contenues dans l'Archive. Ces données sont combinées dans une nouvelle Notice descriptive ou dans un jeu de mises à jour de Description de paquet qui est versé en tant que SIP.

Le processus de versement transforme les SIP reçus durant l'Opération de versement en un jeu d'AIP et de Descriptions de paquets qui peuvent être stockés et acceptés par les Entités « Stockage » et « Gestion de données ». La complexité de ce processus de versement peut varier énormément d'un OAIS à l'autre ou d'un Producteur à l'autre au sein du même OAIS. La forme la plus simple de ce processus consiste à extraire le Contenu d'information, le PDI et les Descriptions de paquets des supports de transfert du Producteur pour les placer dans une file d'attente en vue de leur prise en charge par les Entités « Stockage » et « Gestion de données ». Dans des cas plus complexes, le PDI et les Descriptions de paquets doivent être extraits du Contenu d'information ou saisis par le personnel de l'OAIS pendant le déroulement de l'Opération de versement. L'encodage des Objets-information ou leur affectation à des fichiers peuvent devoir être modifiés. Dans les cas extrêmes, la granularité du Contenu d'information peut être changée et l'OAIS doit produire de nouveaux PDI et des Descriptions de paquets traduisant les Objets-information nouvellement produits. Si la création d'un AIP requiert de nombreux SIP, l'Entité « Entrées » organise le stockage temporaire de ces SIP jusqu'à ce que tous les SIP requis pour l'AIP soient arrivés.

Par ailleurs, l'Entité « Entrées » classe les Objets-information à leur arrivée, détermine à quelle(s) collection(s) existante(s) chaque objet appartient et produit des messages permettant la mise à jour des Descriptions de collections appropriées après stockage de l'AIP par l'Entité « Stockage ». L'OAIS et les organismes externes peuvent fournir des Notices descriptives supplémentaires et les Outils de recherche qui fournissent des moyens d'accès alternatifs aux Objets-information recherchés. Les chercheurs développeront des modèles d'accès aux Objets-information tout à fait nouveaux et fondamentalement différents des modèles existants. Il est important que les modèles de versement de l'OAIS et les modèles de données internes soient suffisamment souples pour pouvoir être enrichis de ces nouvelles descriptions afin que la communauté d'utilisateurs puisse bénéficier de ces efforts de recherche. Un bon exemple de ce type de nouvelle description associée est celui d'une base de données de phénomènes dans l'Observation de la Terre, permettant aux utilisateurs d'obtenir les données souhaitées pour un événement tel qu'un ouragan, ou l'éruption d'un volcan, données provenant de multiples instruments d'observations, à partir d'une simple requête. Il est

important de noter que de tels Outils de recherche peuvent devenir obsolètes sauf si les données qu'ils nécessitent sont conservées en tant que parties des AIP auxquels ils accèdent.

On suppose que l'Entité « Entrées » coordonne les mises à jour entre l'Entité « Gestion de données » et l'Entité « Stockage » et dispose de moyens de synchronisation et de correction d'erreur adaptés. L'AIP devra d'abord être stocké dans l'Entité « Stockage ». La confirmation de cette opération inclura un identificateur unique pour récupérer l'AIP dans l'Entité « Stockage ». Cet identificateur devra être inséré dans la Description de paquet avant que cette Description ne soit elle-même ajoutée à la Description de collection destinée à l'Entité « Gestion de données ».

#### **4.3.3. TRANSFORMATIONS DE DONNÉES DANS LES ENTITES FONCTIONNELLES « STOCKAGE » ET « GESTION DE DONNÉES »**

L'Entité « Stockage » récupère les AIP issus du processus de versement et les intègre aux fonds d'archives permanents. L'Entité « Gestion de données » récupère les Descriptions de paquets produites par l'Entité « Entrées » et enrichit ainsi les Descriptions de collections existantes. Le modèle logique des données versées devra avoir été mis en correspondance avec le modèle logique des collections d'archives. Ainsi la principale transformation qui se produit à ce stade est la mise en correspondance de la session d'acquisition avec le modèle de versement des données physiques. Cela se fera généralement depuis le stockage temporaire, jusqu'au stockage permanent de l'OAIS via des systèmes de gestion de base de données (DBMS - Database Management Systems) ou des systèmes de gestion hiérarchique de fichiers (HFMS - Hierarchical File Management Systems), ou par toute combinaison des deux.

La vue interne de l'OAIS est la représentation permanente des données archivées, dans laquelle l'encodage et les correspondances doivent être bien documentés et compris. Le processus de transfert des objets versés consiste fréquemment en un processus logiciel tel qu'un pilote HFMS ou un DBMS. Dans ce cas, il revient à l'OAIS de maintenir une version du logiciel en fonctionnement ou de documenter soigneusement les formats internes de façon à ce qu'à l'avenir, les données puissent être transférées vers d'autres systèmes sans perte d'information.

#### **4.3.4 FLUX DE DONNEES ET TRANSFORMATIONS DANS L'ENTITE FONCTIONNELLE « ACCES »**

Quand un Utilisateur souhaite utiliser les données de l'OAIS, un Outil de recherche peut être utilisé pour localiser les informations intéressantes. Ces Outils de recherche présentent à l'Utilisateur une vue logique des fonds de l'OAIS de telle manière qu'il puisse déterminer quels AIP acquérir. La vue d'accès dont il dispose sera au minimum la vue logique de haut niveau des Descriptions de collections traitées en 4.2.2.8. L'OAIS peut devoir investir beaucoup de temps et de nombreux efforts pour développer des Notices descriptives et des Outils de recherche tels que des catalogues qui constitueront une aide pour l'Utilisateur dans

la localisation des AIP ou AIC qui l'intéressent. Un Utilisateur ouvrira une Session de recherche avec l'Entité « Accès ». Au cours de cette Session de recherche, il se servira des Outils de recherche de l'OAIS pour identifier et examiner les collections potentiellement intéressantes. Ce processus de recherche sera certainement itératif, d'abord avec des critères très ouverts puis par affinement des critères sur la base des résultats des recherches précédentes. Quand les objets susceptibles d'intérêt sont identifiés, des Outils de recherche plus sophistiqués comme le feuilletage d'images ou des animations permettent d'affiner encore la définition du Jeu de résultats.

Une fois qu'il a identifié les objets de l'OAIS qu'il veut acquérir, l'Utilisateur met en œuvre un Outil de commande fourni par l'OAIS pour émettre une commande d'acquisition de données. L'Utilisateur définit une vue logique des AIP désirés et des Descriptions de paquets associées qui doivent être incluses dans le Paquet d'informations diffusé, et il spécifie les caractéristiques physiques de la session de diffusion de données telles que le type de support et le format d'objet. Ce processus n'implique pas forcément d'interaction visible entre l'Utilisateur et l'OAIS si des valeurs par défaut adéquates existent. Cette commande peut aussi spécifier les transformations qui doivent être appliquées aux AIP pour la création des DIP à la demande de l'Utilisateur.

L'Entité « Accès » enregistre alors le Protocole de commande pour l'Entité « Gestion de données ». Quand les conditions exigées par le Protocole de commande enregistré sont remplies (noter que pour de nombreux Protocoles de commande ces conditions sont remplies immédiatement ; dans le cas contraire l'Entité « Administration » avertit l'Entité « Accès » quand elles sont remplies) l'Entité « Accès » coordonne la réponse : elle sollicite les Entités « Stockage » et « Gestion de données » et leur demande les AIP et les Descriptions de paquets associées nécessaires à la constitution du DIP demandé par l'Utilisateur. Les Entités « Stockage » et « Gestion de données » créent les copies des objets demandés dans un espace de stockage temporaire.

L'Entité « Accès » transforme alors ce jeu d'AIP et de Descriptions de paquets associées en un jeu de DIP et stocke ces DIP sur des moyens de distribution (supports physiques ou moyens de communication) qui seront utilisés pour livrer ces DIP à l'Utilisateur via une session de diffusion de données. La complexité de ce processus de transformation peut varier énormément selon le niveau des services de traitement proposés par l'OAIS et demandés par l'Utilisateur dans le cadre d'une commande. Dans le cas le plus simple, le DIP contient des duplicata des AIP et des Descriptions de paquets associées provenant des Entités « Stockage » et « Gestion de données ». Dans des cas plus complexes, il est possible que le Contenu d'information souhaité soit extrait des Objets-information ou inséré dans des Objets-information auto-décrits et que l'encodage des Objets-information ou leur affectation à des fichiers physiques soient modifiés. Dans les cas les plus complexes, quand l'OAIS permet des services d'extraction de sous-ensembles, la granularité des Objets-information peut être modifiée et le processus de diffusion peut produire des DIP et des Descriptions de paquets associées reflétant cette nouvelle granularité. La correspondance entre les DIP et AIP est de 1 pour 1 si aucune transformation n'est demandée ; cependant, l'usage de services d'extraction de sous-ensembles et d'autres options de traitement de produit peut aboutir à la création de

plusieurs DIP issus d'un seul AIP ou d'un seul DIP à partir de la combinaison de plusieurs AIP.

## 5 PERSPECTIVES DE LA PERENNISATION

Cette section traite des diverses pratiques qui ont été - ou pourraient être - utilisées pour pérenniser l'information numérique et pour préserver les services d'accès à cette information numérique. Elle reprend les concepts de modélisation fonctionnelle et de modélisation de l'information décrits dans la sous-section 4.2 et les applique à ces pratiques, elle étend également la terminologie pour distinguer les aspects significatifs de ces pratiques. La sous-section 5.1 traite de la pérennisation de l'information numérique lorsqu'elle subit des migrations de support et de format. La sous-section 5.2 traite de la préservation des services d'accès à l'information numérique, lors d'évolutions technologiques et de portages vers de nouveaux systèmes, avec encapsulation pour maintenir la cohérence des interfaces ou émulation pour prendre en compte la conservation d'applications patrimoniales. Quelques problèmes critiques dans les différentes approches sont identifiés.

### 5.1 PERENNISATION DE L'INFORMATION

L'évolution rapide de l'industrie informatique et la nature éphémère des supports de stockage de données électroniques sont contradictoires avec l'objectif majeur d'un OAIS : pérenniser l'information. Quel que soit le niveau de qualité avec lequel un OAIS maintient ses collections courantes, il sera certainement nécessaire d'en migrer un grand nombre vers des supports différents et/ou vers d'autres plates-formes logicielles et matérielles pour en préserver l'accès. Les supports de données numériques actuels ne se conservent pas plus de quelques dizaines d'années avant que la probabilité d'une perte irréversible de données ne se concrétise. De plus, le rythme rapide de l'évolution technologique accroît le poids financier de nombreux systèmes après seulement quelques années d'existence.

La **Migration numérique** est définie comme le transfert d'information numérique au sein d'un OAIS dans une optique de conservation. Elle se distingue des transferts, dans leur acception générale, par trois caractéristiques :

- un objectif de conservation de l'intégralité des contenus d'information,
- la perspective dans laquelle la nouvelle implémentation du système d'archivage de l'information se substituera à l'ancienne,
- le pilotage et la responsabilité de tous les aspects du transfert comme parties intégrantes de l'OAIS.

Cette sous-section traite de la Migration numérique d'AIP au sein d'un OAIS.

### 5.1.1 LES FACTEURS DECLENCHEURS DE LA MIGRATION NUMERIQUE

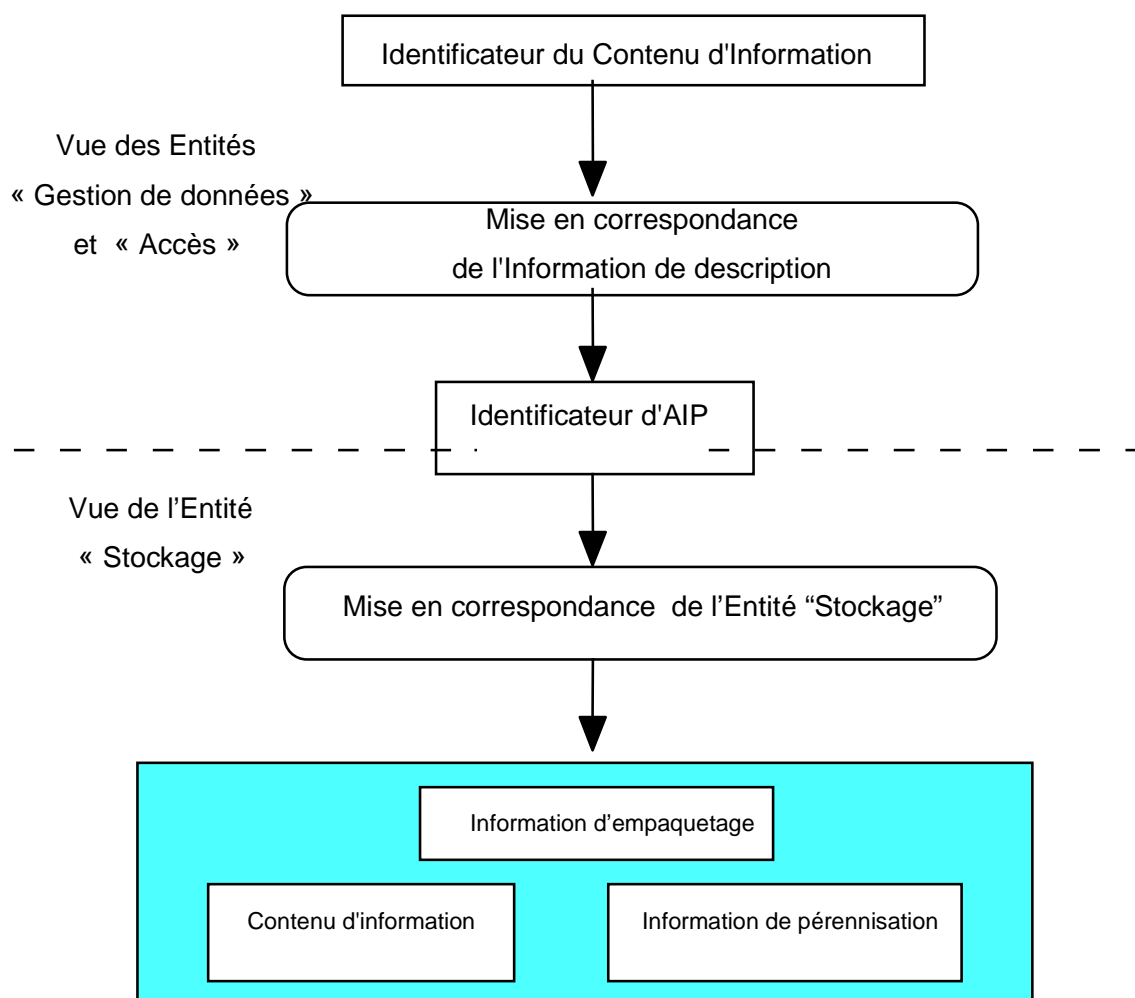
Trois facteurs majeurs peuvent conduire à une Migration numérique d'AIP au sein d'un OAIS. Ce sont :

- Une meilleure rentabilité : l'évolution rapide des matériels (lecteurs de disques et de bandes par exemple) et des logiciels permet un accroissement considérable des capacités de stockage et de transfert pour des coûts de plus en plus faibles. Cette évolution conduit aussi à une obsolescence de certains types de supports, bien avant qu'ils ne se dégradent. De plus, l'amélioration des empaquetages d'AIP peut induire une dépendance moindre de ces AIP vis-à-vis des supports et systèmes sous-jacents, ce qui simplifie les efforts de migration. Pour rester rentable, un OAIS doit s'appuyer sur ces technologies. Selon les technologies mises en œuvre, l'information d'AIP pourra être déplacée vers de nouveaux types de supports non pris en compte initialement et il peut s'avérer nécessaire de revoir l'implémentation de l'AIP pour en tirer parti.
- Nouvelles exigences de service à l'Utilisateur : les Utilisateurs d'un OAIS tirent aussi bénéfice des nouvelles technologies, de sorte que les types et niveaux de service qu'ils attendent d'un OAIS ne font que croître. Ces services étendus peuvent nécessiter de nouvelles formes de DIP pour servir certaines Communautés d'utilisateurs cibles, ce qui en retour peut conduire un OAIS à détenir de nouvelles formes d'AIP permettant de réduire les conversions en sortie. De plus, les AIP subissent généralement les effets de modes et l'OAIS peut devoir fournir des prestations d'accès de niveaux différents pour satisfaire à l'évolution des demandes d'Utilisateurs. Le transfert des AIP vers des supports différents offrant des accès plus ou moins performants est probablement un moyen d'y parvenir. En fin de compte, la Communauté d'utilisateurs cible pour un AIP donné peut s'élargir, avec pour conséquence la nécessité de réviser les formes de l'AIP de façon à ce qu'il soit compréhensible et utilisable par cette communauté élargie. Tout cela peut conduire à la migration d'AIP au sein d'un OAIS.
- Dégradation des supports : les supports numériques, au fil du temps, deviennent de moins en moins fiables en termes de sûreté de conservation des bits. Même ceux qui sont dotés de certains niveaux de corrections d'erreurs peuvent être à remplacer. Le résultat final de la dégradation des supports est que l'information d'AIP doit être transférée vers un nouveau support.

Les Migrations numériques prennent du temps, sont coûteuses et exposent l'OAIS à une augmentation des risques de perte d'information. C'est pourquoi un OAIS se doit de prendre en considération les problèmes de Migrations numériques et leurs différentes approches.

### 5.1.2 CONTEXTE DE MIGRATION

Les concepts clés du modèle fonctionnel et du modèle d'information de la section 4, sous l'angle de la migration, sont résumés dans le schéma 5-1.



**Schéma 5-1 : Vue conceptuelle des relations entre noms et composants d'AIP**

L'interface Utilisateur de l'OAIS de l'Entité « Accès » fournit un ou plusieurs identificateurs de Contenu d'information, avec les espaces de nommage associés, pour faciliter

l'identification de l'Objet-contenu d'information intéressant. Un ou plusieurs de ces identificateurs du Contenu d'information sont inclus dans l'Information d'identification du PDI associé à cet Objet-contenu d'information. Dans l'Entité « Gestion de Données », l'Information de description relie chacun de ces identificateurs à l'identificateur d'AIP correspondant. L'Entité « Accès » utilise cette information pour obtenir l'identificateur de l'AIP et le transmettre à l'Entité « Stockage » afin de récupérer l'AIP associé.

Dans l'Entité « Stockage », l'identificateur d'AIP est relié à la localisation de l'Information d'empaquetage de l'AIP par le système de mise en correspondance de l'Entité « Stockage ». L'Information d'empaquetage de l'AIP, à son tour, délimite et identifie logiquement le Contenu d'information et le PDI, et les associe en une entité unique pour la conservation. Par exemple, supposons que le Contenu d'information et le PDI sont définis comme le contenu de plusieurs fichiers, les pointeurs vers les documents décrivant les représentations de ces fichiers, et les documents eux-mêmes. Dans ce cas, l'Information d'empaquetage serait logiquement définie comme l'ensemble incluant : la référence au système de fichiers qui gère les trains de bits, la structure de données comprenant les pointeurs, l'information utilisée pour distinguer le Contenu d'information du PDI, et une structure de données englobante qui identifie les fichiers et autres structures de données telles que les composantes du Paquet d'informations archivé. Le système de mise en correspondance de l'Entité « Stockage » pourrait alors être implémenté en tant que base de données reliant l'identificateur de l'AIP à la localisation de la structure de données englobante.

Le transfert d'une partie quelconque du Contenu d'information, du PDI ou de l'Information d'empaquetage vers le même support ou un nouveau support, dans le but de remplacer cette partie de l'AIP précédent, est considéré comme une Migration numérique de l'AIP. A noter qu'un simple changement dans l'information de correspondance de l'Entité « Stockage », qui est extérieur au concept d'AIP, n'est pas considéré comme une migration de cet AIP, même si de tels changements doivent être soigneusement contrôlés pour s'assurer que l'accès à l'AIP est préservé.

Les modalités selon lesquelles les AIP sont implémentés auront une influence déterminante à la fois sur le niveau d'automatisation et sur la probabilité de perte d'information au cours des migrations. De bonnes conceptions d'AIP peuvent à la fois accroître l'automatisation de la migration et réduire les probabilités de perte d'information. Pour mieux apprécier les impacts de ces facteurs sur la migration d'AIP il est utile de catégoriser les migrations en différents types et d'étudier les problèmes liés à chacune des approches retenues.

### 5.1.3 TYPES DE MIGRATIONS

A partir des modèles et concepts ci-dessus, il est possible d'identifier quatre principaux types de Migration numérique classés ci-dessous par ordre croissant de risque de perte d'information :

- **Rafraîchissement de support** : Migration numérique dans laquelle un support, contenant un ou plusieurs AIP ou des parties d'AIP, est remplacé par un support du même type par copie bit à bit, support qui sera utilisé pour contenir les AIP. Au final, l'infrastructure de



mise en correspondance de l'Entité « Stockage » est toujours en mesure - sans modifications - de localiser et de donner accès à l'AIP.

- **Duplication** : Migration numérique dans laquelle il n'y a de changement ni dans l'Information d'empaquetage, ni dans le Contenu d'information, ni dans le PDI. Les bits représentant ces Objets-information sont conservés dans le transfert vers un support de même type ou vers un autre support. A noter que le Rafraîchissement est aussi une Duplication, mais une Duplication peut nécessiter des changements dans l'infrastructure de mise en correspondance de l'Entité « Stockage ».
- **Ré-empaquetage** : Migration numérique qui produit quelques changements de bits de l'Information d'empaquetage.
- **Transformation**: Migration numérique qui produit quelques changements dans les bits du Contenu d'information ou du PDI mais vise à conserver l'intégralité du Contenu d'information.

Le Rafraîchissement de support présente le plus faible risque de perte d'information puisque aucun des bits utilisés pour contenir l'information d'AIP ou pour permettre la recherche d'AIP et leur accès n'est modifié. Il y a également peu de risque de perte d'information dans le cas de la Duplication, puisque aucun des bits représentant l'information d'AIP n'a changé. Cependant, si on utilise un nouveau type de support, il sera nécessaire d'apporter quelques modifications à l'infrastructure de mise en correspondance de l'Entité « Stockage » (voir le schéma 5-1). Des modifications inattendues des bits risquent d'apparaître à la suite d'une anomalie au cours de l'opération. Le Ré-empaquetage admet quelques changements de bits, mais ils sont circonscrits à l'information utilisée pour délimiter le Contenu d'information et le PDI, de sorte qu'ils n'altèrent généralement pas l'information portée par le Contenu d'information ou le PDI. Il subsiste toujours un risque d'anomalie ou d'erreur, et il est parfois impossible d'éviter des interactions entre l'Information d'empaquetage, le Contenu d'information ou le PDI. Cela accroît les risques de perte d'information. Finalement, c'est la Transformation qui présente le plus de risques à cause des modifications apportées au Contenu d'information ou au PDI.

Pour comprendre plus clairement les enjeux de ces types de migrations, il est nécessaire d'étudier les différentes implémentations possibles. On verra que certaines migrations sont dans la pratique un mixte de Ré-empaquetage et de Transformation. Il est également important de rappeler que, pour tout AIP, l'OAIS doit en premier lieu identifier ce qui constitue le Contenu d'information ; ce n'est qu'ensuite que le PDI peut être identifié, et puis ensuite l'Information d'empaquetage. En fait, il n'existe pas de définition « correcte » unique de ce que doit être le Contenu d'information dans la mesure où c'est à l'OAIS de le définir pour chaque AIP élaboré et archivé. Tous ces points sont abordés en détail dans les sous-sections qui suivent, à travers une série de scénarios d'implémentation et de migration.

### 5.1.3.1 Rafraîchissement de support

Une migration est un Rafraîchissement de support quand elle a pour effet de remplacer un support par une copie suffisamment exacte pour que tout le matériel et les logiciels de

l'Entité « Stockage » continuent à fonctionner comme précédemment. Considérons le scénario suivant :

Le taux d'erreurs de bits à corriger sur un CD-ROM a atteint un seuil dangereux et on décide de lui substituer une copie exacte. Une fois la vérification de conformité effectuée, le nouveau CD-ROM remplace l'ancien et le Rafrâichissement de support est opéré. Les composants de l'AIP sur le CD-ROM restent inchangés.

### **5.1.3.2 Duplication**

Une migration est une Duplication s'il n'y a pas de modification des bits de l'Information d'empaquetage, ni du Contenu d'information, ni du PDI. Selon l'implémentation, vérifier qu'aucun de ces bits n'a été modifié peut représenter une tâche importante. Considérons le scénario suivant :

Le Contenu d'information et le PDI pour un AIP sont encapsulés dans une structure d'empaquetage standard, contenue dans un seul fichier. Il est facile d'effectuer une migration de type Duplication par simple copie de la séquence de bits du fichier vers un nouveau fichier, sur le même support ou un autre. Le besoin de localisation du fichier peut nécessiter des modifications de l'infrastructure de mise en correspondance de l'Entité « Stockage », mais aucune modification de l'Information d'empaquetage, ni du Contenu d'information, ni du PDI n'est intervenue. La Duplication, avec ce type d'Information d'empaquetage, facilite la migration vers de nouveaux types de support avec une automatisation maximum et un faible risque de perte d'information.

### **5.1.3.3 Ré-empaquetage**

Une migration est un Ré-empaquetage si le transfert s'accompagne de certaines modifications de l'Information d'empaquetage. L'Information d'empaquetage joue a minima un rôle critique pour délimiter et relier le Contenu d'information et le PDI. Si le Contenu d'information et le PDI sont eux-mêmes constitués de multiples composants, on peut demander à l'Information d'empaquetage de les délimiter et de les relier également. Ce sont des choix d'implémentation que l'OAIS doit explicitement reconnaître. Considérons le scénario suivant :

Tous les bits du Contenu d'information et du PDI pour un AIP sont contenus dans trois fichiers sur un CD-ROM. L'Information d'empaquetage est constituée par les bits utilisés pour implémenter la structure de fichiers et de répertoires qui donne accès à ces trois fichiers. Le contenu des trois fichiers est déplacé vers trois nouveaux fichiers sur un autre type de support, avec une nouvelle organisation de répertoires et de fichiers. Même si tous les noms de répertoires et de fichiers ont été préservés au cours du transfert, un Ré-empaquetage a bien eu lieu puisque les bits représentant l'Information d'empaquetage ont été modifiés.

#### 5.1.3.4 Transformation

Les migrations Numériques qui requièrent certaines modifications du Contenu d'information ou de la PDI sont appelées Transformations. Ces modifications concernent certains bits de l'Objet-contenu de données du Contenu d'information ou du PDI ainsi que les modifications correspondantes dans l'Information de représentation associée. Dans tous les cas, l'objectif est d'obtenir une conservation maximale de l'information. L'AIP résultant est censé remplacer totalement l'AIP qui est l'objet de la Transformation. Le nouvel AIP se définit comme une nouvelle **Versio**n du précédent AIP. La première Version de l'AIP est désignée comme AIP d'origine et peut être gardée en vue d'un contrôle de la conservation.

L'Information de représentation joue un rôle clé dans les Transformations, et les conséquences des modifications sur l'Information de représentation peuvent être utilisées pour définir les catégories de ces Transformations. Un Objet-information de représentation peut être modélisé comme étant constitué d'un ensemble élémentaire d'entités, un ensemble d'entités résultantes, et des règles de mise en correspondance, qui définissent les entités résultantes ainsi que leurs relations avec les entités élémentaires. Partant de ce modèle d'Objet-information de représentation, deux types de Transformations peuvent être définis : la **Transformation réversible** et la **Transformation irréversible**.

On parle d'une Transformation réversible quand la nouvelle représentation définit un ensemble (ou un sous-ensemble) d'entités résultantes équivalentes aux entités résultantes définies par la représentation d'origine. Ceci implique qu'il y ait une correspondance univoque vers la représentation d'origine et ses ensembles d'entités de base. On peut prendre pour exemple le remplacement d'une représentation qui utilise le jeu de caractères ASCII par une représentation utilisant le jeu de caractères UNICODE UTF-16. La Transformation va consister en un remplacement du codage sur 7-bits par un codage sur 16-bits dans l'objet AIP qui est l'objet du changement. La Transformation inverse peut ultérieurement être réalisée par le remplacement du jeu de caractères UNICODE UTF-16 par le jeu de caractères ASCII, et l'AIP d'origine est restauré.

On parle d'une Transformation irréversible quand on ne peut pas garantir que la transformation soit réversible. Par exemple, le remplacement d'un nombre représenté en virgule flottante IBM 7094 par un nombre représenté en virgule flottante IEEE constitue une Transformation irréversible, parce que les entités résultantes de ces deux représentations ne sont pas identiques. L'une est plus précise que l'autre. Cependant, elles peuvent être suffisamment équivalentes, selon l'usage qui est fait des valeurs qu'elles représentent, pour être effectivement interchangeables. Dans ce cas, une Transformation irréversible préserve effectivement le contenu de l'information. Pour des formats complexes, dans lesquels la définition des groupes et les relations entre groupes ont une signification, il peut être difficile d'affirmer qu'une Transformation irréversible a préservé convenablement le Contenu d'information. Des exemples de Transformations réversibles et irréversibles sont donnés dans les scénarios ci-après.

Le scénario suivant montre une Transformation réversible qui se produit lorsqu'on introduit une fonction de compression sans perte dans le Contenu d'information d'un AIP.

Tous les bits du Contenu d'information pour un AIP sont contenus dans trois fichiers sur un CD-ROM. L'Information d'empaquetage inclut les bits utilisés pour implémenter la structure de fichiers et de répertoires qui donne accès à ces trois fichiers. Les contenus des trois fichiers sont transférés vers un nouveau CD-ROM et sont compressés durant le processus avec un algorithme de compression sans perte. Ce transfert est une Transformation puisque la compression a modifié le Contenu d'information, et il s'agit d'une Transformation réversible parce qu'il existe un algorithme de décompression qui rétablira le contenu en bits du fichier d'origine. Les composantes utiles de l'Information de représentation du Contenu d'information d'origine doivent être mises à jour pour inclure l'algorithme de décompression, et l'information de la PDI doit être également mise à jour pour constituer la nouvelle Version de l'AIP.

Le scénario suivant montre une Transformation irréversible qui peut se produire quand le Contenu d'information a migré vers un nouveau format qui a la capacité d'exprimer un modèle de données plus diversifié que le format d'origine.

Tous les bits du Contenu d'information d'un AIP sont rassemblés dans trois fichiers sur un CD-ROM. L'Information d'empaquetage inclut les bits utilisés pour implémenter la structure de fichiers et de répertoires qui donne accès à ces trois fichiers. Les contenus des trois fichiers sont transférés vers un nouveau CD-ROM. Au cours du processus le troisième fichier est modifié parce qu'il n'existe plus d'outils disponibles permettant d'utiliser le contenu de ce fichier dans son format actuel. Le nouveau format, qui est communément utilisé, s'appuie sur un modèle de données différent de celui du format d'origine et il existe plusieurs façons de mettre l'information en correspondance dans le nouveau format. Cette mise en correspondance doit être réalisée soigneusement pour s'assurer qu'il n'y a pas de perte significative d'information pour la Communauté d'utilisateurs cible. La correspondance doit figurer dans le PDI et, bien entendu, l'Information de représentation décrivant le nouveau format va se substituer à celle du précédent format. Il en résulte une nouvelle Version de l'AIP. Il s'agit d'une migration de type Transformation qui constitue également une Transformation irréversible parce qu'il n'existe pas d'algorithme de restauration du fichier d'origine.

Le scénario suivant montre une Transformation réversible comprenant un Ré-empaquetage. C'est le cas lorsque le Contenu d'information contient un nom de fichier utilisé comme pointeur vers l'un de ses composants, et que ce Contenu d'information est transféré vers un nouveau type de support avec changement des noms des fichiers.

Le Contenu d'information d'un AIP est défini par le corps de trois fichiers sur un CD-ROM. Le premier de ces fichiers contient un nom interne qui le lie au troisième fichier et qui spécifie leur relation. L'Information d'empaquetage comprend la structure de répertoires et de fichiers permettant d'identifier ces trois fichiers. Au cours d'une migration vers un nouveau type de support, ces trois fichiers sont mis dans un nouveau répertoire et on leur donne de nouveaux noms. Ceci constitue une migration de type Ré-empaquetage parce qu'il s'agit d'une nouvelle implémentation

de la structure de répertoires et de fichiers, qui remplit la fonction d'empaquetage. Cependant, le nom interne doit être également être mis à jour pour maintenir le lien entre le premier et le troisième fichier. Cette mise à jour modifie le Contenu d'information et signifie que la migration est également une Transformation. Si le nom interne avait été un identificateur universel, il n'aurait pas été nécessaire de le changer. Cependant, le cadre standardisé qui définit l'identificateur universel aurait inclus l'information de mise en correspondance conduisant à la localisation du troisième fichier et aurait par conséquent nécessité une mise à jour. Cette approche est avantageuse pour un OAIS parce qu'elle permet des mises à jour centralisées et plus facilement gérées. Cependant, la technologie mise en jeu est plus complexe et il n'existe pas de consensus sur la technique d'identification à employer.

Ce dernier scénario montre une Transformation irréversible comprenant un Ré-empaquetage. C'est le cas lorsque le Contenu d'information inclut des noms de fichiers, une structure de répertoires, et des attributs de fichiers associés. Le Contenu d'information est alors migré vers un nouveau type de support admettant une implémentation différente des structures de répertoires et de fichiers qui acceptent moins d'attributs.

Les bits du Contenu d'information et du PDI d'un AIC sont définis comme une agrégation d'AIU dans laquelle chaque AIU est constitué de trois fichiers sur un CD-ROM, de leurs noms de fichiers, des attributs de fichiers et des noms de répertoires. L'Information d'empaquetage se compose des bits utilisés pour implémenter la structure de fichiers et de répertoires qui donne accès à chacune des instances des trois fichiers, mais elle n'inclut pas les noms réels de répertoires et de fichiers. Il est possible de trouver des milliers d'AIU sur un seul CD-ROM. Le transfert de cet AIC vers un nouveau type de support qui utilise une nouvelle représentation de structure de fichiers et de répertoires acceptant moins d'attributs de fichier peut conduire à une Transformation irréversible aussi bien qu'à une migration de type Ré-empaquetage. Il s'agit d'une Transformation parce que le Contenu d'information stocké à l'origine dans les structures de fichiers et de répertoires doit être redistribué dans les nouvelles structures de fichiers et de répertoires et probablement dans les fichiers eux-mêmes. Il s'agit d'une Transformation irréversible s'il n'existe pas de mise en correspondance algorithmique terme à terme entre, d'une part, les structures de fichiers et de répertoires et les contenus de fichiers résultants, et d'autre part, ceux d'origine. Il s'agit d'un Ré-empaquetage car on a affaire à une nouvelle implémentation de la structure de répertoires et de fichiers, laquelle constituait un sous-ensemble de l'Information d'empaquetage. La pratique qui consiste à encoder le Contenu d'information dans un nom de fichier ou de répertoire augmente le risque de perte d'information parce que l'évolution d'un environnement de gestion de données est facilitée par la possibilité de mise à jour des noms de fichiers et de répertoires selon les besoins.

#### **5.1.4 DISTINCTIONS ENTRE VERSIONS D'AIP, EDITIONS ET AIP DERIVES**

A moins qu'une Migration numérique n'implique une Transformation, il n'est ni envisagé de créer une nouvelle Version d'AIP, ni nécessaire de mettre à jour son PDI. En d'autres termes,

la Version d'AIP est considérée comme indépendante du Rafrâichissement, de la Duplication, et du Ré-empaquetage qui n'affectent pas le Contenu d'information ou le PDI. Ce qui ne signifie pas que l'OAIS ne garde pas trace de telles migrations, mais simplement qu'il n'est pas nécessaire de les tracer dans le PDI. Si ce processus est mené entièrement à l'intérieur de l'Entité « Stockage », l'identificateur d'AIP reste le même et il n'y a pas de conséquence sur les Notices descriptives ou les Outils d'accès.

Une Migration numérique qui entraîne une Transformation aboutit à une nouvelle Version de l'AIP comme cela est défini au 5.1.3.4. Dans ce cas, le PDI doit être mis à jour pour identifier l'AIP source et sa Version, et pour tracer ce qui a été fait et pourquoi. Le nouvel AIP est vu comme un remplaçant de l'AIP source dans lequel l'information a été conservée au mieux. L'identificateur d'AIP est également nouveau et la Notice descriptive doit être mise à jour. Ce qui n'implique pas que des modifications soient nécessaires dans les Outils d'accès sauf s'ils ont été implémentés avec des identificateurs d'AIP codés en dur.

Un AIP peut, dans certains environnements, être sujet à des mises à niveau ou améliorations dans le temps. Il ne s'agit pas d'une Migration numérique en ce sens que l'objectif n'est pas de conserver l'information, mais de l'enrichir ou de l'améliorer. Ce type de modification d'AIP peut être défini comme la création d'une nouvelle **Edition**. La nouvelle Edition est vue comme le remplacement de la précédente Edition, mais il peut être utile, à titre historique, de conserver la précédente Édition. Il en résulte également un nouvel identificateur d'AIP avec les mêmes conséquences sur les Notices descriptives et les Outils d'accès qu'une Migration numérique par Transformation.

Un OAIS peut aussi trouver avantage à fournir un AIP dérivé d'un AIP existant. Cela est possible par extraction de certaines informations ou par agrégation d'informations à partir de plusieurs AIP, afin de mieux servir les Utilisateurs. Ce type d'AIP résultant peut être désigné comme **AIP dérivé**. Il ne remplace aucun des AIP dont il est issu et ne constitue pas le résultat d'une Migration numérique. Il en découle également un nouvel identificateur d'AIP et de nouvelles Notices descriptives. Il se peut que cela requière de nouveaux Outils d'accès ou leur mise à jour, selon le type d'implémentation.

## **5.2 PRESERVATION DES SERVICES D'ACCES**

Il peut être souhaitable pour un OAIS de conserver les services d'accès d'un Utilisateur face aux changements technologiques. Cette sous-section présente un ensemble de scénarios afin de cerner les problèmes relatifs à la conservation des services d'accès et de définir une terminologie.

### **5.2.1 API DE DIFFUSION**

Le premier scénario suppose que la Communauté d'utilisateurs cible souhaite développer des applications qui permettent l'accès aux AIP en utilisant une Interface Programmatique (Application Programming Interface - API) maintenue par l'OAIS en tant que Logiciel d'accès. L'OAIS peut choisir de fournir cette API comme une implémentation alternative à la

production et à la livraison d'un DIP physique pour la diffusion. Ce type de service autorise l'Utilisateur, en tant que client, à développer des applications qui permettent d'accéder directement aux AIP. Ce type d'accès pourrait être très utile pour des applications telles que la fouille de données, peu adaptées à la création et l'envoi de DIP contenant de grands AIC. Cette API pourrait permettre à une application de naviguer virtuellement dans un AIC, de fournir à l'application les bits d'un Objet-contenu de données pour les AIU sélectionnés et d'identifier leurs localisations afin d'obtenir l'Information de représentation associée et les PDI. Cependant, du fait de l'évolution de la technologie, l'OAIS est porté sur de nouveaux matériels, supports et systèmes d'exploitation. Si l'OAIS entend maintenir la même API pour ses Utilisateurs, il sera nécessaire de fournir une surcouche autour d'une partie de sa nouvelle infrastructure pour mettre ses services en conformité avec l'API en place. Cette API devra être documentée de manière appropriée et testée pour s'assurer qu'elle délivre correctement le Contenu d'information de l'AIU utilisant ce nouveau Logiciel d'accès. Cette approche ne devrait pas entraîner de modifications pour les logiciels développés par la communauté des Utilisateurs. Quand l'API est applicable à une grande quantité d'AIU de l'OAIS ou qu'il y a un nombre significatif d'applications d'Utilisateurs basées sur l'API, cette approche en couche est clairement réalisable et peut engendrer un rapport coût/bénéfice favorable pour l'OAIS et sa Communauté d'utilisateurs cible. Le « Modèle d'Information en couche » présenté en Annexe C de ce document décrit plus en détail quelques API potentiellement standard.

### **5.2.2 CONSERVATION DU « LOOK AND FEEL » DU LOGICIEL D'ACCES**

Le deuxième scénario suppose que la Communauté d'utilisateurs cible souhaite conserver le « look and feel » d'origine du Contenu d'information d'un ensemble d'AIU tel qu'il se présente dans une application spécifique ou un ensemble d'applications. Conceptuellement, l'OAIS fournit un environnement qui permet à l'Utilisateur de visualiser le Contenu d'information d'AIU via les possibilités de transformation et de présentation de l'application. Par exemple, on peut vouloir utiliser une application spécifique qui extrait des données d'un CD-ROM ISO 9660 et les présente comme une image multi-spectrale. Cette application fonctionne sous un système d'exploitation particulier; elle nécessite un ensemble d'informations de contrôle, requiert l'utilisation d'un lecteur de CD-ROM et transfère l'information au pilote d'un périphérique d'affichage spécifique. Dans certains cas, cette application peut offrir à tous les membres de la Communauté d'utilisateurs cible un accès en profondeur à l'environnement et l'OAIS se contente alors de désigner l'Objet-contenu de données comme le train de bits utilisé par l'application. Dans d'autres cas, si ce type d'environnement est moins facile d'accès, l'OAIS peut le fournir via le Logiciel d'accès. Cependant, lorsque l'OAIS et/ou la Communauté d'utilisateurs cible migrent vers de nouveaux environnements informatiques, l'application finira par ne plus fonctionner ou ne fonctionnera pas correctement.

La réponse de l'OAIS pour conserver un Logiciel d'accès dépendra probablement, en grande partie, du fait qu'il possède, ou qu'il puisse obtenir ou non, le code source du logiciel. La sous-section 5.2.2.1 traite des méthodologies éprouvées pour conserver l'accès à ce logiciel au travers des changements technologiques. Le critère principal dans l'utilisation de ces

techniques doit être le rapport coût/bénéfice pour l'OAIS et sa Communauté d'utilisateurs cible. S'il n'y a pas la disponibilité du code source ou des passerelles du commerce et qu'il y a une exigence absolue pour l'OAIS de conserver le « look and feel », l'OAIS devra expérimenter la technique "d'émulation" qui fait actuellement l'objet de recherches dans le domaine des Bibliothèques numériques. Cette technologie est présentée en 5.2.2.2.

#### **5.2.2.1 Méthodologies impliquant la disponibilité du code source**

La réponse de l'OAIS pour conserver un service mettant en œuvre un Logiciel d'accès dépendra probablement, en partie, du fait qu'il dispose ou non du code source du logiciel. Si l'OAIS dispose du code source et de la documentation adéquate du logiciel, l'approche probable sera de porter ce logiciel vers le nouvel environnement et d'essayer de le valider suffisamment pour s'assurer qu'il fonctionne correctement. Comme décrit en 4.2, cette validation peut ne pas être évidente lorsque le logiciel tourne mais ne fonctionne pas correctement. Idéalement toutes les données de sortie doivent avoir été relevées au départ pour être utilisées comme référence afin de contrôler que le fonctionnement est correct après portage. Cependant, ce niveau de test risque d'entraîner un rapport coût/bénéfice inacceptable pour l'OAIS. Du fait que l'application a été compilée à partir du code source d'origine, les algorithmes sont supposés corrects. Il est alors probablement suffisant de réaliser une batterie de tests ou de réutiliser celle fournie dans la documentation de conception.

Si le Logiciel d'accès est un ensemble propriétaire, largement utilisé et disponible sur le marché, il est probable qu'il existe une passerelle logicielle commercialisée (c'est-à-dire la conversion) qui transforme les Objets-contenu de données courants en d'autres formes utilisées par le nouveau Logiciel d'accès ayant un « look and feel » semblable. Si on ne trouve aucune alternative commerciale, l'OAIS peut passer un accord avec le propriétaire du Logiciel d'accès d'origine pour développer et fournir le code source d'un outil simplifié pouvant lire - sans les modifier - les instances de données écrites avec ce format. Cette approche peut ne pas être viable pour des raisons de coûts ou de problèmes juridiques. Dans tous les cas, l'OAIS devra mettre en place les mécanismes pour vérifier qu'aucune information conservée n'a été perdue. Des critères doivent avoir été établis pour définir clairement ce qui constitue le Contenu d'information, comme traité en 5.1. De plus, l'OAIS doit s'assurer que le nouveau Logiciel d'accès est disponible pour la Communauté d'utilisateurs cible.

#### **5.2.2.2 Approches potentielles d'émulation**

Il se peut que la Communauté d'utilisateurs cible impose de maintenir le « look and feel » d'un Logiciel d'accès propriétaire, en raison du grand nombre d'AIU qui en dépendent. Dans ce cas, si l'OAIS est dans l'impossibilité d'obtenir le code source, il lui faudra envisager d'explorer une approche d'émulation.

L'OAIS peut étudier l'émulation du logiciel. Si le logiciel fournit un jeu bien connu de commandes et une API bien définie pour l'accès, l'API pourra être documentée et évaluée de façon appropriée pour tenter une émulation du logiciel. Cependant, si l'interface Utilisateur consiste surtout en un affichage ou d'autres dispositifs périphériques faisant appel aux sens humains (par exemple, le son), cette rétro-ingénierie est quasiment impossible. Comme décrit



en 4.2, cela peut ne pas être évident lorsque le logiciel tourne mais qu'il ne fonctionne pas correctement. Pour prévenir de telles situations, il est nécessaire de sauvegarder les sorties du Logiciel d'accès obtenues dans un fonctionnement correct, ainsi que l'Information de représentation adéquate et le PDI, afin que toute cette information soit conservée. L'ensemble doit être confronté aux résultats obtenus après la migration dans un nouvel environnement. Cela peut s'avérer délicat si le logiciel a plusieurs modes de fonctionnement. De plus, si les sorties du logiciel sont principalement dirigées vers un périphérique d'affichage, l'enregistrement de ce flux ne garantit nullement que l'affichage sera le même dans le nouvel environnement et, par conséquent, il se peut que la combinaison du logiciel et de l'environnement ne restitue plus une information totalement correcte à l'Utilisateur. Le maintien d'un « look and feel » cohérent peut requérir en premier lieu de l'enregistrer séparément pour l'utiliser en tant qu'information de validation.

Une autre approche est l'émulation de la plate-forme matérielle. L'avantage de cette émulation réside dans l'assurance que lorsqu'une plate-forme matérielle a été émulée avec succès, tous les systèmes d'exploitation et les applications qui ont tourné sur la plate-forme d'origine peuvent être basculés sans modification sur la nouvelle. Cependant, cela ne tient pas compte des contraintes liées aux périphériques d'entrée/sortie. L'émulation rencontre beaucoup de succès lorsqu'un système d'exploitation très répandu doit tourner sur un matériel pour lequel il n'a pas été conçu, comme une version de Windows <sup>TM</sup> sur un Apple <sup>TM</sup>. Cependant, même dans ce cas, lorsque les tendances fortes du marché favorisent cette approche, les applications ne tournent ou ne s'exécutent pas forcément correctement dans l'environnement émulé. Par exemple, il peut s'avérer impossible de simuler entièrement toutes les dépendances et contraintes de temps de matériels anciens, en raison des contraintes imposées par le nouvel environnement matériel. Par ailleurs, quand l'application présente l'information à une interface humaine, il est délicat de déterminer si le nouveau périphérique présente encore l'information correctement ; cela incite à prévoir un enregistrement séparé de la présentation de l'information pour la validation. Une fois le principe de l'émulation adopté, le système résultant est particulièrement sensible à des erreurs de logiciel inconnues jusqu'alors et qui peuvent sérieusement mettre en danger l'accès continu à l'information. Au vu de ces contraintes, les obstacles techniques et économiques à l'émulation de matériel ne sont pas négligeables.

Des recherches sur des approches d'émulation alternatives ont été faites, par exemple, le développement d'une architecture de machine virtuelle ou l'émulation au niveau du système d'exploitation. Ces approches résolvent certains des problèmes posés par l'émulation matérielle, mais soulèvent de nouvelles questions. Aucune des techniques d'émulation n'est à ce jour assez mûre pour en parler de manière significative. De plus, les travaux de recherche actuels sur l'émulation impliquent une architecture centralisée avec contrôle de tous les périphériques. Le niveau de complexité des interfaces et des interactions dans un contexte d'informatique entièrement distribuée (c'est-à-dire le WWW et JAVA) avec des clients hétérogènes peut impliquer des exigences qui vont au-delà du domaine des travaux actuels sur l'émulation.



## 6 INTEROPERABILITE DES ARCHIVES

Les utilisateurs d'Archives OAIS multiples peuvent avoir des raisons de souhaiter un certain degré d'uniformisation ou de coopération entre ces Archives. Par exemple, les utilisateurs de plusieurs Archives peuvent souhaiter disposer :

- d'Outils de recherche communs pour faciliter la localisation de l'information dans plusieurs Archives,
- d'un schéma de Description de paquet commun pour l'accès,
- d'un schéma de DIP commun pour la diffusion, ou
- d'un site unique d'accès global.

Les Producteurs peuvent vouloir disposer :

- d'un schéma de SIP commun pour les versements à différentes Archives, ou
- d'un dépôt unique pour toutes leurs productions.

Les gestionnaires peuvent vouloir disposer de solutions pour :

- réduire les coûts en partageant des matériels, des logiciels et des travaux de conservation onéreux, et
- accroître l'homogénéisation et la qualité des interactions entre plusieurs Archives.

Il peut donc se révéler avantageux pour des Archives de coopérer pour répondre à ces souhaits. L'initiative peut venir des Archives elles-mêmes ou d'une autorité ayant un certain poids pour pouvoir imposer cette coopération. Dans le premier cas, les motivations des Archives peuvent être diverses :

- réduire les coûts,
- satisfaire les Utilisateurs grâce à leurs produits,
- satisfaire les Utilisateurs grâce à leur qualité de service, ou enfin
- rivaliser avec d'autres Archives pour survivre ou se développer.

De telles situations peuvent avoir - et ont - motivé des accords, sans nécessiter pour autant la constitution formelle d'un Groupement d'archives instituant une autorité externe. Cependant,

dans les cas où un tel groupement est mis en place, l'autorité externe est représentée dans ce Modèle de référence par le Management.

Le but de cette section est d'examiner le degré d'interaction et de coopération entre Archives. La sous-section 6.1 se concentre sur les interactions techniques, tandis que la sous-section 6.2 traite des problèmes de management générés par l'opposition entre coopération et autonomie.

## **6.1 NIVEAUX TECHNIQUES D'INTERACTION ENTRE ARCHIVES OAIS**

Les associations d'OAIS peuvent être catégorisées, au plan technique, selon des facteurs externes et internes. Les facteurs externes prennent en compte les caractéristiques des communautés de Producteurs et d'Utilisateurs. Les facteurs internes peuvent prendre en compte les mises en œuvre communes des modèles d'information présentés en 4.2 ou le partage entre Archives d'un ou plusieurs domaines fonctionnels présentés en 4.1.

Cette sous-section définit quatre catégories d'associations d'Archives. Les trois premières catégories correspondent à des degrés croissants d'interaction :

- *Indépendantes* : Archives motivées par des considérants locaux, sans interaction technique ou de gestion entre elles.
- *Coopérantes* : Archives qui peuvent avoir des Producteurs communs, des standards communs de versement et de diffusion, mais pas d'Outil de recherche commun.
- *Groupements d'Archives* : Archives ayant à la fois une Communauté locale (c'est-à-dire la Communauté d'utilisateurs cible d'origine servie par l'Archive) et une Communauté élargie (c'est-à-dire une Communauté d'utilisateurs cible étendue). Cette Communauté « élargie » est intéressée par les fonds de plusieurs Archives OAIS et a convaincu ces Archives de lui fournir un accès au travers d'un ou plusieurs Outils de recherche communs. Les besoins d'accès de la Communauté locale ont généralement priorité sur ceux de la Communauté élargie. La diffusion et les versements globaux sont des fonctionnalités optionnelles.
- *Ressources Partagées* : Archives qui ont passé des accords avec d'autres Archives pour partager des ressources, et éventuellement réduire les coûts. Cela impose de partager différents standards propres aux Archives (comme les interfaces entre fonctions de versement et de stockage, ou encore entre fonctions de stockage et d'accès), mais ne change pas la vision de l'Archive par la Communauté d'utilisateurs.

Le reste de cette sous-section donne une vue plus détaillée de ces catégories d'association.

### **6.1.1 ARCHIVES INDEPENDANTES**

Une Archive indépendante est supposée servir une seule Communauté d'utilisateurs cible. L'Archive et la Communauté d'utilisateurs cible doivent s'accorder sur la conception des SIP, DIP et des Outils de recherche. Une Archive indépendante peut choisir de concevoir ces structures à partir de standards formels ou de facto, ce qui permettrait la coopération avec

d'autres Archives qui pratiquent les mêmes standards. Toutefois, les décisions d'utiliser ces standards ne sont pas dictées par la possibilité d'interopérabilité avec d'autres Archives, mais plutôt par des exigences locales et des soucis d'économie.

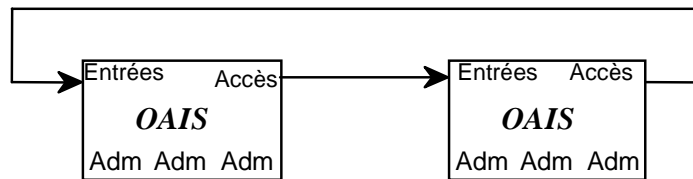
La qualification d'une Archive comme indépendante n'est liée ni à sa taille ni à l'existence d'une fonctionnalité distribuée. Une Archive indépendante peut occuper un site unique ou être physiquement répartie sur plusieurs sites. Elle peut utiliser plusieurs standards pour un élément interne donné. Cependant, s'il n'y a aucune interaction avec d'autres Archives, l'Archive est dite indépendante.

### **6.1.2 ARCHIVES COOPERANTES**

Les Archives coopérantes sont fondées sur des accords de standards entre deux ou plusieurs Archives. La forme la plus simple de coopération entre Archives apparaît lorsqu'une Archive agit comme Utilisateur de données pour une autre Archive. Dans ce cas l'Archive utilisatrice doit accepter le format de DIP de l'Archive productrice comme format de SIP. Les Archives coopérantes ont des communautés d'intérêt liées, elles commandent et versent ainsi des données provenant d'autres Archives coopérantes et peuvent éventuellement avoir des Producteurs de données communs. Cela ne suppose aucun standard d'accès, de versement ou de diffusion communs. La seule condition pour cette architecture est que les groupes de coopération acceptent au moins un format commun de SIP et de DIP pour des requêtes inter-Archives. Le mécanisme mis en œuvre pour cette sorte d'interopérabilité peut consister en des Demandes d'abonnement auprès de chaque Archive.

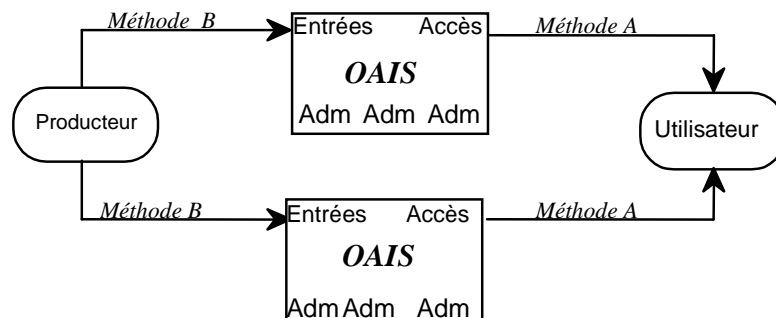
Les figures 6-1 et 6-2 illustrent le concept d'Archives coopérantes.

A un niveau élémentaire d'interaction entre Archives, le schéma 6-1 représente un simple accord d'échange mutuel d'information. (Note : dans ce schéma et les suivants, l'OAIS est représenté comme un dispositif à « cinq ports » selon la configuration du schéma 6-1. Dans chaque cas, et pour plus de clarté on montre un groupement de deux Archives, bien que le concept puisse être étendu indéfiniment). L'exigence essentielle pour ce groupement est l'existence d'un ensemble de Protocoles de versement, de Demandes d'abonnement et de standards d'interface utilisateur mutuels permettant aux DIP d'une Archive d'être réceptionnés en tant que SIP par une autre. Cela suppose donc que des compatibilités réciproques soient mises en place entre Archives. En revanche, des méthodes d'accès, de diffusion et de versement communes pour tous les participants ne sont pas obligatoires, bien que cela puisse permettre plus d'échanges. Ce niveau d'accord est aussi utile quand les fonds d'une Archive ont été consolidés/transférés dans une autre Archive en raison de problèmes de Management.



**Schéma 6-1: Archives coopérantes ayant un accord d'échange mutuel**

Le schéma 6-2 est un exemple d'Archives OAIS ayant standardisé leurs méthodes de versement et de diffusion au bénéfice des Utilisateurs. Aucun élément externe particulier n'est nécessaire pour cela. L'inconvénient réside dans le fait qu'il n'y a aucun mécanisme formel pour l'échange d'Information de description, de sorte que les Utilisateurs doivent avoir des Sessions de recherche séparées pour repérer les AIP qui les intéressent.



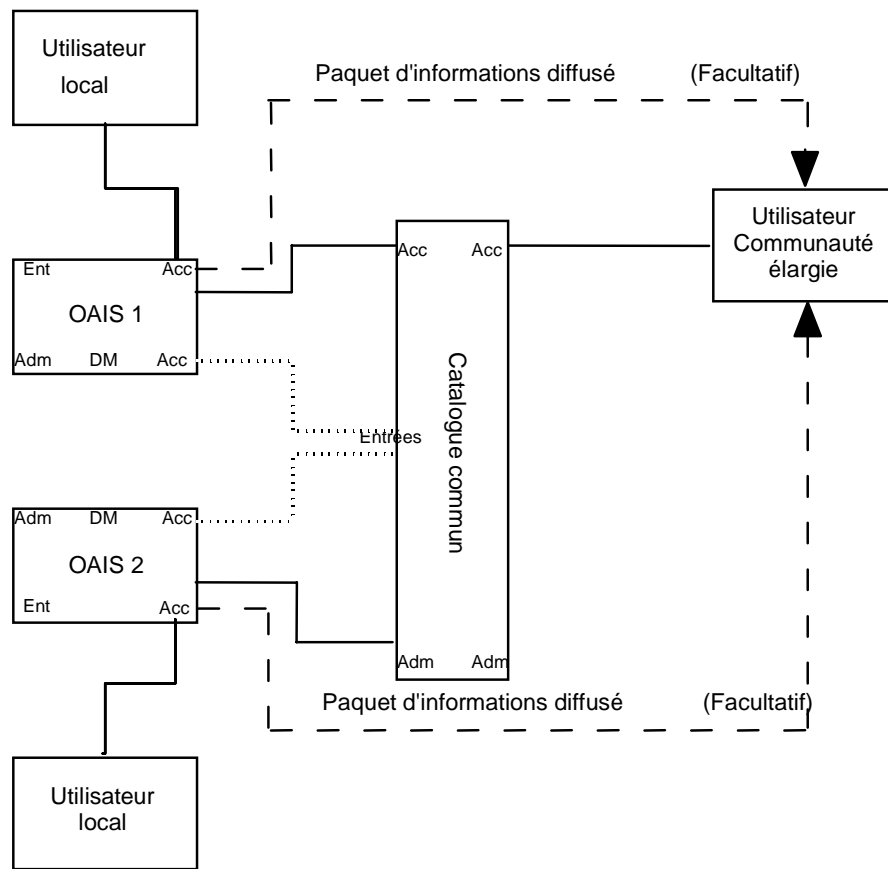
**Schéma 6-2 : Archives coopérantes avec des méthodes standard de versement et d'accès**

### 6.1.3 GROUPEMENTS D'ARCHIVES

Un Groupement d'archives est conceptuellement orienté Utilisateur. En plus de la Communauté locale (c'est-à-dire, la Communauté d'utilisateurs cible d'origine servie par l'Archive), il existe une Communauté élargie (c'est-à-dire, une Communauté d'utilisateurs cible étendue) qui est intéressée par les fonds de plusieurs Archives OAIS et les a convaincues de donner accès à leurs collections au travers d'un ou de plusieurs Outils de recherche communs. Cependant, les Utilisateurs locaux ont probablement la priorité d'accès sur ceux de la Communauté élargie.

Au niveau du Groupement, des éléments externes peuvent être introduits pour améliorer l'interopérabilité. Par exemple, le schéma 6-3 illustre une architecture fonctionnelle susceptible de résoudre le problème d'accès décrit en 6.1.2, en utilisant une entité externe aux groupements d'OAIS. Dans ce cas, deux Archives OAIS qui ont des Communautés cibles semblables ont décidé de se grouper pour permettre aux Utilisateurs de localiser les Paquets

d'informations qui les intéressent dans l'un ou l'autre OAIS lors d'une Session de recherche unique. Les catalogues et le gestionnaire de données communs représentent l'élément fédérateur global externe qui sert de point d'accès commun à l'information dans les deux Archives. Des DIP contenant les Outils de recherche de chaque OAIS sont versés dans le Catalogue commun comme indiqué par les pointillés du schéma 6-3. Le catalogue commun peut limiter son activité à servir d'Outil de recherche ; il peut aussi inclure la diffusion commune de produits de l'un ou l'autre des deux OAIS, comme indiqué par les tirets dans le schéma.



**Schéma 6-3 : Groupement d'OAIS utilisant un catalogue commun**

Les Groupements d'archives peuvent se subdiviser selon trois niveaux de fonctionnalité :

- *Site central* : l'accès global est réalisé par l'export d'une Notice descriptive selon un format standard vers un site global. Le site global gère indépendamment un ensemble de descripteurs de plusieurs Archives et possède des Outils de recherche pour localiser quelles Archives disposent d'une collection intéressante. On donne à l'Utilisateur une vue synthétique des fonds de plusieurs sites, maintenue de manière centralisée. Pour prendre

connaissance des détails relatifs aux documents archivés, l'Utilisateur doit accéder au site qui détient le document réel. Cela est plus facile quand les sites et les clients acceptent un jeu standard de protocoles.

- *Outil de recherche distribué* : l'accès global est réalisé par la mise en œuvre d'un nœud global qui peut distribuer une requête aux diverses Archives locales. Cela signifie que l'Entité « Gestion de données » locale doit stocker une Notice descriptive complémentaire dans le format global ou disposer d'un traducteur des requêtes globales vers les requêtes locales. Une option dans ce cas consiste à établir un format de DIP commun pour réduire la charge de travail des Utilisateurs susceptibles de commander des produits de plusieurs Archives.
- *Outil d'accès distribué* : à la fonctionnalité d'Outils de recherche distribués traitée ci-dessus, s'ajoute dans ce cas un mécanisme standard de commande et de diffusion disponible via des nœuds globaux. C'est un système fédéré entièrement fonctionnel. Ici, le système global peut influencer sur la conception du schéma de Notice descriptive dans chaque Archive locale. Il serait optimal de construire de nouvelles Archives locales fondées sur les schémas et les Outils de recherche globaux pour assurer de hauts niveaux d'interopérabilité.

Il y a plusieurs questions importantes de politique et/ou de technologie qui doivent être traitées quand un OAIS rejoint un Groupement d'archives ou quand plusieurs OAIS indépendants décident de créer un tel Groupement. Ces questions comprennent :

- **L'unicité des noms** pour chaque AIP dans le Groupement. Un OAIS a la responsabilité d'identifier de manière unique chacun de ses AIP. Quand un OAIS rejoint un Groupement, il n'y a aucune certitude qu'un de ses identificateurs courants d'AIP n'ait pas déjà été utilisé par d'autres membres du Groupement. On peut résoudre ce problème en définissant les identificateurs d'AIP au sein du Groupement par l'attribution d'un identificateur unique à chaque OAIS du Groupement et en le concaténant à chaque identificateur original d'AIP de cet OAIS. Ce nom d'OAIS pourrait être formaté selon une norme qui donne au Client ou à d'autres membres du Groupement l'information nécessaire pour établir une connexion avec l'OAIS qui détient l'AIP considéré. Ce peut être par exemple la norme ISO X.500, nommage des Services d'annuaire.
- **La duplication d'AIP** dans plusieurs OAIS différents ayant des identificateurs d'AIP différents. Ce problème vient du fait qu'avant d'adhérer au Groupement, un OAIS aura fait un double du Contenu d'information d'un AIP provenant d'autres OAIS pour permettre l'accès des Utilisateurs locaux. Dans ce cas un Utilisateur global verra tous les exemplaires comme des AIP distincts mais identifiés comme uniques. L'examen détaillé du PDI associé au Contenu d'information devrait permettre à l'Utilisateur de localiser l'exemplaire original qui fait foi, mais le processus de recherche risque d'être très laborieux. On peut aussi traiter ce problème à l'aide d'un champ spécifique dans la Notice descriptive de tout AIP qui identifie s'il s'agit de l'original ou d'une copie. Ce procédé n'est pas efficace si, avant le Groupement, deux ou plusieurs Archives ont reçu le Contenu d'information du Producteur pour archiver. Dans ce cas les Groupements d'archives verraient ces AIP doublonnés comme des AIP uniques et originaux.



- **Le maintien de l'accès du Groupement aux AIP** quand un OAIS cesse son activité. Il arrive souvent, malheureusement, que des Archives ferment alors que leurs fonds ont encore de la valeur pour la communauté du Groupement. Il faudrait alors que le Groupement souscrive un accord avec chaque OAIS membre afin de définir quel autre OAIS aurait la responsabilité de ses fonds en cas de fermeture.
- **L'authentification d'Utilisateur et la gestion d'accès** pour les utilisateurs de la Communauté élargie. Si un OAIS a une politique qui limite l'accès à certains de ses AIP ou facture la diffusion de certains Paquets d'informations, alors se pose le problème d'identifier et d'authentifier l'Utilisateur qui fait des demandes via le nœud central. Chaque OAIS aura mis en œuvre un système de gestion d'accès et d'authentification pour ses Utilisateurs locaux et l'infrastructure pour cette fonction devra être étendue aux Utilisateurs de la Communauté élargie. Voici quelques exemples de techniques utilisées dans les systèmes actuels :
  - un premier niveau par défaut où tous les membres du nœud global partagent un ensemble commun de contraintes d'accès, le nœud global prend en charge toute la chaîne d'authentification et vérifie que l'Utilisateur est bien un Utilisateur légitime du nœud global. L'authentification auprès de l'OAIS membre est réalisée en présupposant que toute demande parvenant du nœud global provient d'un Utilisateur légitime.
  - l'admission d'Utilisateur par identification où l'Utilisateur distant en question peut être authentifié par tout OAIS du Groupement, le nœud global agit simplement comme un intermédiaire dans le dialogue d'identification. Ceci est assez délicat du point de vue de la sécurité au regard des technologies existantes, mais la technique de certification X.509 du WWW est prometteuse.

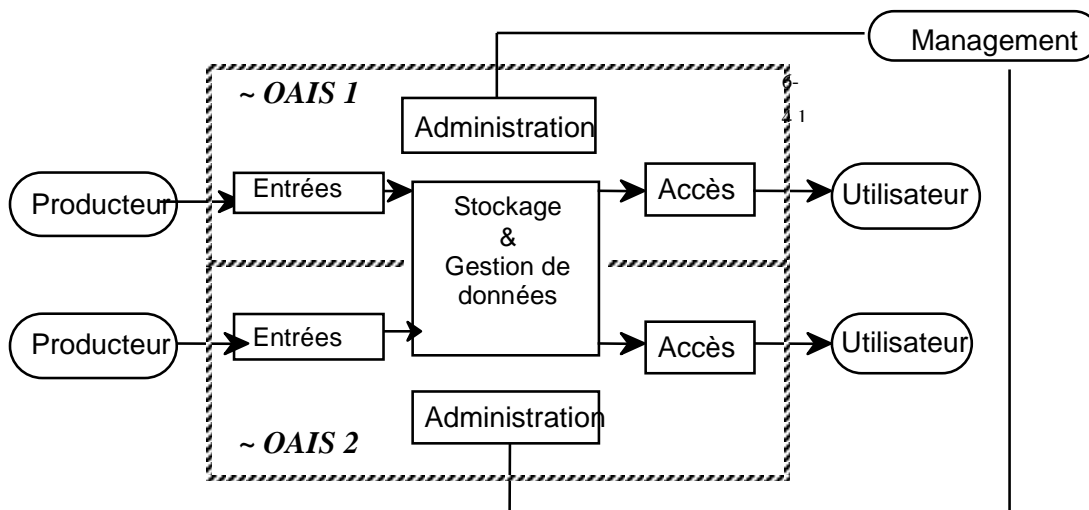
Il existe de nombreux critères pour décider laquelle de ces techniques devrait être utilisée par un Groupement d'Archives spécifique. Le critère majeur est la granularité des contraintes d'accès au sein du Groupement. S'il y a des données peu confidentielles et aucune facturation pour la diffusion de données, il est tout à fait raisonnable d'adopter une politique qui détermine les contraintes d'accès d'un Utilisateur à partir de la source de repérage et de commande des AIP. Cela implique peu de modifications du système OAIS d'identification : un simple ajout du nœud global comme Utilisateur. Le nœud global devra inclure des mécanismes pour identifier les Utilisateurs globaux au niveau de chacun des mécanismes d'identification des Groupements d'OAIS.

S'il y a facturation pour la diffusion de l'information archivée ou si des données privées suffisamment confidentielles nécessitent l'authentification individualisée des utilisateurs, les techniques de proxy seront insuffisantes. En ce cas, les techniques d'identification d'Utilisateur telles que mots de passe et certificats doivent être appliquées. Les technologies permettant la mise en œuvre de tels mécanismes continuent de se développer.

### 6.1.4 ARCHIVES AVEC DES DOMAINES FONCTIONNELS PARTAGES

Dans ce type d'association, le Management a passé des accords avec d'autres Archives dans le but de partager ou d'intégrer certains domaines fonctionnels. La motivation peut en être de devoir partager des ressources coûteuses comme un système de gestion hiérarchique des fichiers pour l'Entité « Stockage », un périphérique pour le versement ou la diffusion de Paquets d'informations ou enfin des super-calculateurs utilisés pour les transformations complexes entre SIP, AIP ou DIP. Cette association diffère fondamentalement des exemples précédents, en ce sens qu'on ne peut plus ignorer l'architecture interne des Archives.

Le schéma 6-4 illustre le partage entre deux Archives, OAIS 1 et OAIS 2 d'une fonction de stockage, constituée d'une Entité « Stockage » et d'une Entité « Gestion de données ». Les dispositifs pour l'accès, la diffusion et le versement peuvent se situer à tout niveau d'interopérabilité précédemment décrit. En fait, chaque Archive peut servir des communautés totalement indépendantes comme le montre le schéma suivant. Cependant, pour que l'Entité « Stockage » commune fonctionne de manière satisfaisante, l'application de standards est requise au niveau des interfaces internes entre l'Entité « Entrées » et l'Entité « stockage », ou entre stockage et accès. Le dispositif de l'Entité « Stockage » à Long terme du CNES est un exemple concret de cette architecture.



**Schéma 6-4 : Archives avec stockage partagé**

## 6.2 LE MANAGEMENT DES GROUPEMENTS D'ARCHIVES

Les exemples ci-dessus montrent que le modèle OAIS est compatible avec la mise en place de groupements dans le but de satisfaire tel ou tel objectif. Cependant, on pourrait aussi considérer que certains de ces objectifs peuvent être atteints à partir d'initiatives volontaristes. Il s'agit là d'un aspect important dans l'association de systèmes, et notamment pour les Archives, parce qu'il définit le degré d'autonomie de chaque système. Au cœur de la

question de l'autonomie réside la facilité avec laquelle une association peut être modifiée par l'un de ses membres. Citons quelques caractérisations possibles des niveaux d'autonomie :

- pas d'interactions donc pas d'association,
- les associations dont chaque membre garde sa propre autonomie. Le membre d'une telle association peut avoir à remplir certaines conditions mais il peut quitter l'association sans avertissement et sans répercussion. Un exemple est la participation à Internet, y compris comme opérateur de serveurs de noms de domaines. Tout membre devra satisfaire à certaines exigences, il devra notamment maintenir un site ayant des caractéristiques données, mais ce membre sera de fait rayé de l'association si le site en question n'est plus conforme à ces exigences. Il n'y a cependant aucune pénalité dans ce cas. Les membres restent toujours totalement autonomes, ils sont libres de leurs actions, sans pénalité particulière,
- les associations où l'adhésion est contractuelle. Pour changer la nature de cette association, un membre devra renégocier son contrat. Le degré d'autonomie est fonction de la difficulté de négocier ces changements. Cette difficulté est *a priori* proportionnelle au nombre de membres.

Ainsi, la dimension d'autonomie est un aspect clef pour les Archives qui interagissent. Elle détermine la facilité avec laquelle chacun peut apporter des modifications à la nature de l'association ainsi que les conséquences/pénalités pour chacun s'il veut recouvrer une pleine autonomie. Cette dimension diffère du degré d'homogénéité technique que l'association met en place ou accepte, mais elle n'en est pas totalement indépendante. Par exemple, un haut niveau d'homogénéité technique peut être réalisé dans une grande association, où chaque entité participante est libre de partir sans pénalité. Mais on ne peut garantir dans ce cas la permanence d'une telle association ; cette permanence peut être améliorée en rendant la renégociation du contrat d'adhésion plus difficile ou en prévoyant des pénalités en cas de volonté de retour à une pleine autonomie. De la même manière, un niveau donné d'homogénéité technique peut être atteint plus rapidement et à moindre coût si le contrat est plus contraignant.



## ANNEXE A : RELATIONS AVEC D'AUTRES NORMES OU EFFORTS

(La présente annexe **ne fait pas** partie de la Recommandation.)

Cette annexe décrit les relations entre le Modèle de référence OAIS et différents autres normes ou efforts. Elle fait un parallèle entre certains termes utilisés dans différents domaines et ceux utilisés dans le Modèle de référence OAIS.

- *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Conservation de l'information numérique : rapport du groupe de travail sur l'archivage de l'information numérique) (référence [2]):

Ce document a servi de base pour la définition de l'Information de pérennisation dans le modèle d'information OAIS détaillé à la section 4.2 du Modèle de référence OAIS. Le rapport sur la conservation de l'information numérique ne comprenait pas les classes distinctes d'Objet-information relatives à l'Information d'empaquetage et à l'Information de description qui ont été ajoutées au modèle d'information OAIS. Ainsi, les définitions suivantes des classes de PDI ne constituent qu'un sous-ensemble de celles abordées dans ce document, une partie de l'information étant affectée aux nouveaux Objets-information d'empaquetage et Objets-information de description. La principale différence entre le modèle d'information OAIS et le modèle d'information présenté dans le rapport sur la conservation de l'information numérique concerne :

- l'Information de contexte : cette information explique la relation entre le Contenu d'information et son environnement, notamment la justification de la création du Contenu d'information et son interaction avec les autres Objets-contenu d'information qui existent ailleurs. L'information de contexte du Modèle de référence OAIS est différente de la définition figurant dans le rapport sur la conservation de l'information numérique car elle n'inclut pas l'information utilisée pour associer l'information logique au support physique. Ce type d'information est affecté à l'Information d'empaquetage dans le Modèle de référence OAIS.
- *Z39.50 Profile for Access to Digital Collections* (Profil pour l'accès aux collections numériques) (référence [4]):

Ce document et les profils Z39.50 connexes ont servi de base aux concepts de descriptions associées et d'aides à la recherche traités dans les sections relatives aux données descriptives et à l'accès du Modèle de référence OAIS. Toutefois, le Modèle de référence OAIS a généralisé ces concepts de sorte que les définitions détaillées de protocole du document sur le profil des collections numériques ne sont plus applicables.

- *IEEE Reference Model for Open Storage Systems* (Modèle de référence IEEE pour les systèmes ouverts de stockage) (référence [6]):

Ce document établit un ensemble de fonctionnalités qui s'inscrivent dans l'Entité « Stockage » de l'Archive OAIS. Toutefois, la fonctionnalité de ce domaine

fonctionnel peut être supérieure, avec le stockage de supports physiques non numériques et l'accent sur les exigences de Pérennisation.

- *IEEE Guide to the POSIX® Open System Environment (OSE)* (référence [5]) (Guide pour l'environnement de systèmes ouverts POSIX®), et *Department of Defense Technical Architecture Framework for Information Management. Vol. 2, Technical Reference Model* (référence [7]): (Cadre technique d'architecture de la gestion des informations du Ministère de la défense, vol. 2, Modèle technique de référence)

Ces documents ont servi de base aux services communs définis dans la section 4.1 du Modèle de référence OAIS. Ils concernent le domaine des services de plate-forme communs de façon plus détaillée que cela n'a été nécessaire dans le Modèle de référence OAIS.

Les normes CCSDS du Panel 2 fournissent une mise en œuvre concrète de la plupart des concepts d'Objet-information et de Paquet d'informations traités dans la section 4.2 du Modèle de référence OAIS. Ces normes comprennent:

- *Standard Formatted Data Units—Structure and Construction Rules* (Unités de données à structuration normalisée (SFDU) - Règles de structure et de construction) (référence [8]).

Cette norme présente un mécanisme appliquant le concept de réseau d'Information de représentation et un Paquet d'informations indépendant de la plate-forme.

- *The Data Description Language EAST Specification (CCSD0010)* (Langage de description de données - Spécification EAST) (référence [10]):

Cette norme spécifie un langage approprié pour documenter les composants structurels de l'Information de représentation de la plupart des structures axées sur les enregistrements.

- *Data Entity Dictionary Specification Language (DEDSL)—Abstract Syntax (CCSD0011)* (Langage DEDSL - syntaxe abstraite).

Cette norme spécifie un ensemble d'attributs et une notation pour décrire certaines parties sémantiques des entités de données. Ce mécanisme peut être utilisé pour fournir des éléments sémantiques supplémentaires pour l'Information de représentation.

- *Data Entity Dictionary Specification Language (DEDSL)—XML/DTD Syntax (CCSD0013)* (Langage DEDSL - syntaxe XML).

Cette norme spécifie un ensemble d'attributs et une notation pour décrire certaines parties sémantiques des entités de données. Ce mécanisme peut être utilisé pour fournir des éléments sémantiques supplémentaires pour l'Information de représentation.

- *Producer-Archive Interface Methodology Abstract Standard* (référence [14]):

Cette norme prend en compte la terminologie et les concepts du Modèle de référence O AIS. Elle définit une méthodologie structurée en plusieurs étapes pour préparer et conduire dans les meilleures conditions les transferts d'objets numériques entre le Producteur et l'Archive.

Les termes suivants sont, dans certains contextes organisationnels, des correspondances proches des termes O AIS. Toutefois, ils ne sauraient être considérés comme des termes officiels de substitution O AIS.

Archives (archives traditionnelles) :	O AIS ou Archive O AIS
Acquisition (archives traditionnelles) :	Entrées
Article (archives traditionnelles) :	Contenu d'information
Public principal (journaux) :	Communauté d'utilisateurs cible





## ANNEXE B : INTRODUCTION RAPIDE AU LANGAGE DE MODELISATION UNIFIE (UML)

(Cette Annexe **ne fait pas** partie de la Recommandation.)

Le schéma B-1 présente un résumé des relations entre objets dans les diagrammes UML.

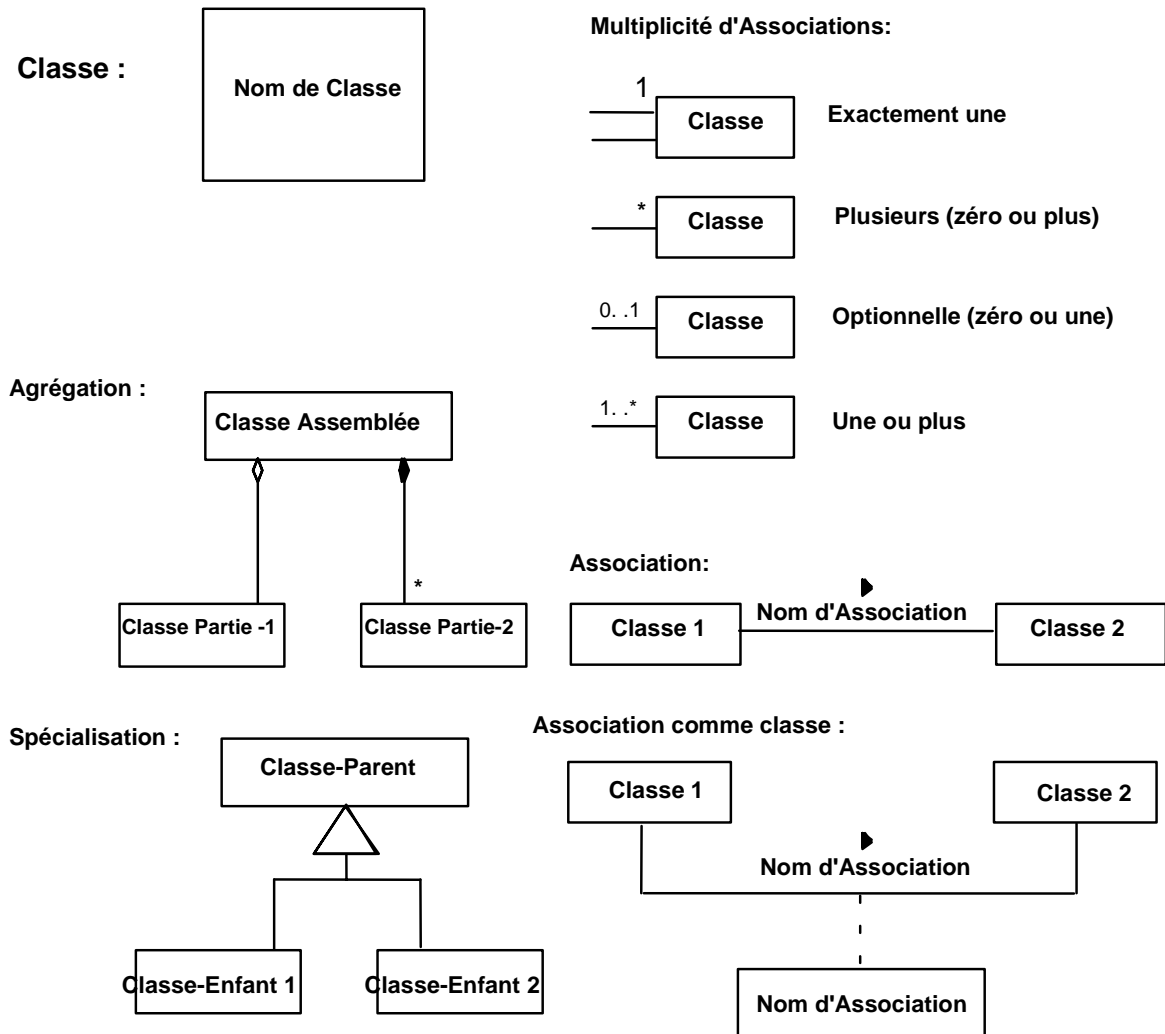


Schéma B-1 : les éléments clés dans les diagrammes UML

Une **classe** est indiquée par un rectangle contenant le nom de la classe. La représentation UML d'une classe est un rectangle à trois compartiments : le nom figure dans le compartiment du haut, les attributs dans le second compartiment et les méthodes dans le compartiment inférieur. Dans ce document les attributs et compartiments d'opérations sont toujours vides et UML établit que les compartiments vides peuvent être supprimés.

Les Classes et objets sont mis en relation les uns avec les autres au travers des **associations** et des **multiplicités** variées peuvent être attachées à ces associations. La multiplicité renvoie au nombre d'instances ou d'objets de cette classe impliqués dans la relation.

Une ligne continue connectant deux classes indique l'association générale entre deux classes. La ligne est renseignée par un nom d'association, indiquant la nature de l'association, et une pointe de flèche pleine indiquant le sens dans lequel la relation doit être comprise. La multiplicité de chaque Classe est indiquée à proximité de la classe et de la ligne d'association. Si l'association forme une classe qui peut posséder ses propres attributs ou méthodes, cette classe d'association est représentée comme un rectangle connecté à la ligne pleine par une ligne tiretée. La multiplicité peut être omise si l'association est de 1 pour 1.

Il existe deux associations particulières utilisées communément, agrégation et spécialisation, qui sont indiquées par des symboles spécifiques.

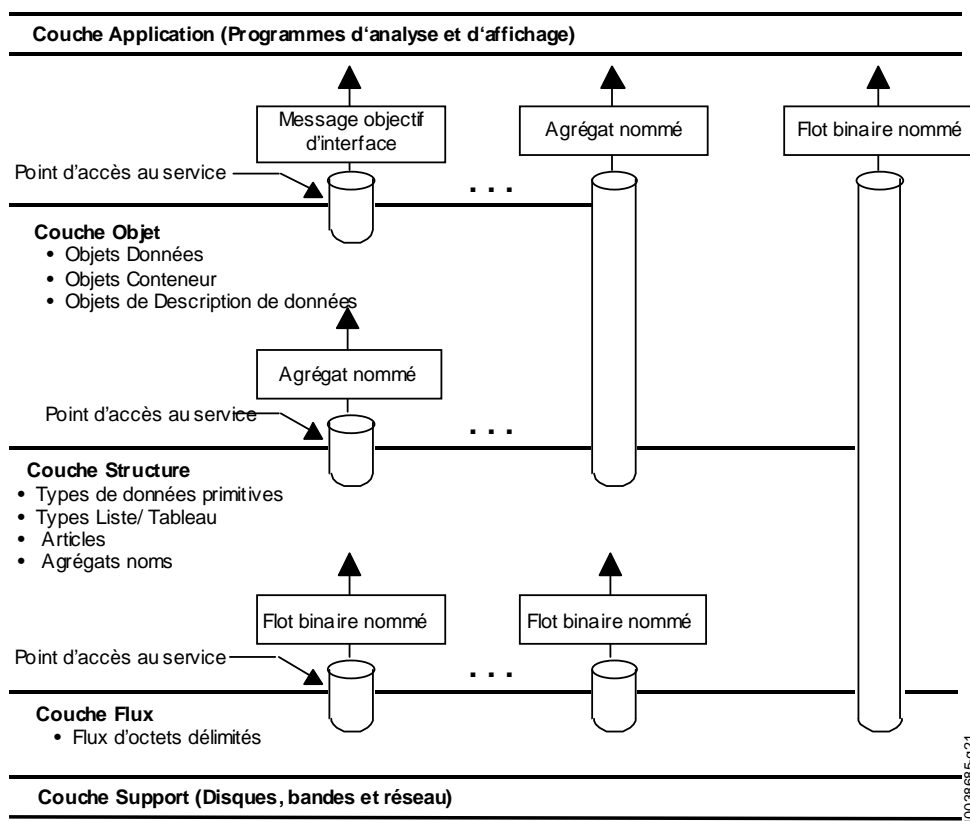
Une association de type **agrégation** est définie quand une classe est considérée comme partie d'une autre classe. En UML, un diamant connectant l'association agrégation à la classe agrégée montre une association. Deux types d'agrégations sont définis par UML. Une agrégation forte, où les classes Partie-de sont physiquement stockées comme parties de la classe agrégée, est indiquée par un diamant plein. Dans une agrégation forte, si la classe agrégée est détruite, les classes-enfant le sont également. Une agrégation faible, où les classes Partie-de sont référées par les classes agrégées, est indiquée par un diamant vide. Dans une agrégation faible, si la classe agrégée est détruite, les classes-enfant ne le sont pas et peuvent être agrégées à de nouvelles classes. Les agrégations fortes peuvent être vues comme des agrégations par valeur, tandis que les agrégations faibles peuvent être vues comme des agrégations par référence. Dans le schéma B-1, l'association agrégation exprime que la classe assemblée contient exactement une instance de classe Partie-de\_1 et zéro ou plus instances de classes Partie-de\_2. De même si une instance **Assemblée** est détruite l'instance Partie-de\_1 continuera à exister mais toutes les instances Partie-de\_2 seront détruites.

Une association **spécialisation** est une association où une classe-enfant hérite des attributs et méthodes de la classe-parent. En UML, un triangle connectant l'association agrégation à la classe-parent signale la spécialisation. Une instance d'une classe-enfant contient tous les attributs et méthodes de ses parents, donc une instance de la classe-enfant peut être utilisée dans toute opération où une instance de la classe parent serait valide. Cependant, la classe-enfant peut ajouter tout nombre de nouveaux attributs ou méthodes et ainsi une instance d'une classe-parent n'est pas nécessairement un substitut valide à la classe-enfant. Dans le schéma B-1, l'association spécialisation exprime que les attributs et méthodes de la Classe-Parent sont hérités par la classe-enfant\_1 et la classe-enfant\_2.

## ANNEXE C : MODELE POUR L'UTILISATION DE LOGICIELS DANS L'INFORMATION DE REPRESENTATION

(La présente annexe **ne fait pas** partie de la Recommandation.)

La sous-section 4.2 et la section 5 montraient que les logiciels sont souvent utilisés en fin de Réseau de représentation. Une façon de voir cette information est le Modèle d'information en couches illustré par le schéma C-1. Dans ce modèle, il y a cinq couches de logiciel. Chacune de ces couches a des interfaces bien définies avec les couches supérieures du modèle. Ces interfaces sont connues sous le nom d'API (interface programmatique) ou points d'accès au service dans d'autres modèles en couches. Voici un aperçu des fonctions de chaque couche et des données qui sont échangées à chaque interface. Cet aperçu illustre le processus d'acquisition de bits à partir du support et d'ajout de l'Information de représentation nécessaire pour que l'information soit utilisable par l'Utilisateur.



**Schéma C-1 : Modèle d'information en couches**

- La Couche Support modélise simplement le fait que les chaînes de bits sont stockées sur un support physique ou de communication, par exemple sous la forme d'éléments

magnétisés ou de tensions électriques. La fonction de cette couche est de convertir cette représentation binaire en une représentation binaire qui peut être utilisée à des niveaux supérieurs (c'est-à-dire, 1 et 0). Cette couche a une seule interface, qui permet aux couches supérieures de spécifier l'emplacement et la taille du flot de bits utile et de recevoir les bits comme une chaîne de bits 1 et 0. Dans les systèmes informatiques modernes, les pilotes de périphériques et les composants de l'interface de stockage physique fournissent en grande partie cette fonction.

- La Couche Flux masque les caractéristiques particulières du support de transport en enlevant tous les artefacts du processus de stockage ou de transmission (telles que les formats de paquet, les tailles de bloc, les espaces inter-enregistrements et les codes de correction d'erreur) et fournit aux niveaux supérieurs une vue cohérente des données indépendante de ce support. L'interface entre la Couche Flux et les couches supérieures permet aux couches supérieures de demander des Blocs de données par nom et de recevoir une chaîne bits/octetes représentant des Blocs de données. Le terme *nom* désigne ici toute clé unique permettant de localiser le flux de données utile. Par exemple, les noms de chemin d'accès pour des identificateurs de fichier ou de message pour les messages de télécommunication. Dans les systèmes informatiques modernes, les systèmes de gestion de fichiers du système d'exploitation fournissent souvent cette fonction.
- La Couche Structure convertit les flux bits/octetes de l'interface de la Couche Flux en structures adressables de types de données primitives qui peuvent être reconnus et traités par des processeurs informatiques et des systèmes d'exploitation. Pour toute implémentation, la couche structure définit les types de données primitives et les agrégations qui sont reconnus. Cela comprend généralement au moins les caractères, les nombres entiers et réels. Les types d'agrégation habituellement supportés, comprennent l'enregistrement (c'est-à-dire une structure qui peut contenir plusieurs types de données) et le tableau (où chaque élément correspond au même type de données). Les questions relatives à la représentation des types de données primitives sont résolues dans cette couche. L'interface entre la couche structure et les niveaux supérieurs permet aux niveaux supérieurs de demander des agrégations labellisées des types de données primitives et de les recevoir sous une forme structurée qui peut être adressable en interne. Dans les systèmes informatiques modernes, les compilateurs et interprètes de langage de programmation fournissent généralement cette fonction.
- La Couche Objet, qui convertit les agrégats labellisés de types de données primitives en information, représentée comme des objets qui sont reconnaissables et signifiants dans le domaine d'application. Dans le domaine scientifique, cela comprend les objets tels que les images, les spectres et les histogrammes. La couche objet ajoute un sens sémantique aux données traitées par les couches inférieures du modèle. Voici certaines fonctions spécifiques de cette couche :
  - Elle définit les types de données à partir du contenu de l'information et non de la représentation de ces données dans la couche structure. Par exemple, différents types d'objets – images, cartes et tableaux – peuvent être implémentés au niveau structure en utilisant les tableaux ou les articles. Dans la couche objet, les images,

cartes et tableaux sont reconnus et traités comme des types d'information distincts.

- Elle présente des applications avec une interface cohérente vers des types similaires d'Objets-information, sans souci des représentations sous-jacentes. L'interface définit les opérations qui peuvent être réalisées sur l'objet, les entrées requises pour chaque opération et les types de données de sortie pour chacune.
- Elle fournit un mécanisme pour identifier les caractéristiques des objets qui sont visibles pour les utilisateurs, les opérations qui peuvent être appliquées à un objet, et les relations entre objets.

L'interface entre la couche Objet et la couche application permet aux niveaux supérieurs de spécifier l'opération qui doit être appliquée à un objet, les paramètres nécessaires pour cette opération et la forme sous laquelle les résultats de l'opération lui seront retournés. Une interface spéciale permet à l'Utilisateur de prendre connaissance des informations sémantiques relatives aux objets telles que les opérations disponibles et les relations avec d'autres objets. Dans les systèmes informatiques modernes, les bibliothèques de sous-programmes et d'objets caractérisés par les méthodes applicables et par leurs interfaces assurent cette fonction.

- La couche application contient des programmes spécifiques pour analyser les Objets-données et présenter le résultat de l'analyse ou l'Objet-données lui-même sous une forme compréhensible par un Utilisateur des données. Dans les systèmes informatiques modernes, les programmes d'application fournissent cette fonction.

Le problème qui se pose en utilisant un logiciel dans les Réseaux de représentation finaux est que les programmes qui sont sauvegardés ne comprennent pas l'information nécessaire pour permettre aux niveaux inférieurs du modèle en couches d'extraire l'information des bits sur le support. Ces services sont généralement fournis par les systèmes d'exploitation, les pilotes de périphériques et les systèmes de fichier fournis par le distributeur. Lorsque les données sont déplacées vers d'autres supports ou des plates-formes logicielles différentes, les interfaces entre ces niveaux peuvent changer. Le processus de migration est traité de façon plus détaillée dans la section 5 du présent document.



## ANNEXE D : REFERENCES INFORMATIVES

(La présente annexe **ne fait pas** partie de la Recommandation.)

- [1] *Procedures Manual for the Consultative Committee for Space Data Systems*. CCSDS A00.0-Y-7.4. Livre jaune. Edition 7.4 (interim update). Washington, D.C.: CCSDS, Janvier 2001.
- [2] *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Washington, D.C.: Commission on Preservation and Access, Mai 1996. <<http://www.rlg.org/ArchTF/>>
- [3] *Unified Modeling Language*. Version 1.1. Cupertino, CA: Rational Software Corporation, Septembre 1997. <<http://www.rational.com/uml/resources>>
- [4] *Z39.50 Profile for Access to Digital Collections*. Draft 7. Washington, D.C.: Library of Congress, Mai 1996. <<http://lcweb.loc.gov/z3950/agency/profiles/collections.html>>
- [5] *IEEE Guide to the POSIX® Open System Environment (OSE)*. IEEE 1003.0-1995. Piscataway, NJ: IEEE, Février 1995.
- [6] IEEE Storage System Standards Working Group. *Reference Model for Open Storage Systems Interconnection—Mass Storage System Reference Model Version 5*. New York: IEEE, Septembre 1994. <[http://www.ssswg.org/public\\_documents.html](http://www.ssswg.org/public_documents.html)>
- [7] *Department of Defense Technical Architecture Framework for Information Management*. Vol. 2, *Technical Reference Model*. Version 2. Arlington, VA: DISA, 1994.
- [8] *Standard Formatted Data Units—Structure and Construction Rules*. Recommandation pour les Standards sur les Systèmes de Données Spatiales, CCSDS 620.0-B-2. Livre Bleu. Edition 2. Washington, D.C.: CCSDS, Mai 1992.
- [10] *The Data Description Language EAST Specification (CCSD0010)*. Recommandation pour les Standards sur les Systèmes de Données Spatiales, CCSDS 644.0-B-2. Livre Bleu. Edition 2. Washington, D.C.: CCSDS, Novembre 2000.
- [11] *Data Entity Dictionary Specification Language (DEDSL)—Abstract Syntax (CCSD0011)*. Recommandation pour les Standards sur les Systèmes de Données Spatiales, CCSDS 647.1-B-1. Livre Bleu. Edition 1. Washington, D.C.: CCSDS, Juin 2001.
- [12] *Data Entity Dictionary Specification Language (DEDSL)—XML/DTD Syntax (CCSD0013)*. Recommandation pour les Standards sur les Systèmes de Données Spatiales, CCSDS 647.2-B-1. Livre Bleu. Edition 1. Washington, D.C.: CCSDS, Juin 2001.

- [13] *Information Processing—Volume and File Structure of CD-ROM for Information Interchange*. ISO 9660:1988.
- [14] *Producer-Archive Interface Methodology Abstract Standard*. . Recommandation pour les Standards sur les Systèmes de Données Spatiales, CCSDS 651.0-B-1. Livre Bleu. Edition 1. Washington, D.C.: CCSDS, Mai 2004.



## ANNEXE E : INDEX

(La présente annexe **ne fait pas** partie de la Recommandation.)

Cette annexe contient un index croisé des principaux mots clés sélectionnés (acronymes et termes du glossaire) et de leur utilisation dans le document.

---

### A

AIC. 1-7, 1-9, 1-10, 4-39, 4-40, 4-43, 4-44, 4-45, 4-46, 4-47, 4-55, 5-9, 5-11  
 AIP1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-13, 1-14, 1-15, 1-16, 2-7, 2-8, 2-11, 2-12, 3-2, 3-3, 3-5, 3-6, 4-1, 4-2, 4-6, 4-7, 4-8, 4-9, 4-10, 4-11, 4-13, 4-15, 4-16, 4-17, 4-33, 4-34, 4-35, 4-36, 4-37, 4-38, 4-39, 4-40, 4-41, 4-43, 4-44, 4-45, 4-46, 4-47, 4-48, 4-52, 4-53, 4-54, 4-55, 5-1, 5-2, 5-3, 5-4, 5-5, 5-6, 5-7, 5-8, 5-9, 5-10, 6-4, 6-6, 6-7, 6-8  
 AIU.... 1-7, 1-10, 1-15, 4-39, 4-40, 4-41, 4-42, 4-43, 4-44, 4-45, 5-9, 5-11, 5-12  
 Archives coopérantes ..... 1-9, 6-3, 6-4

---

### C

Collection virtuelle..... 1-9, 4-47  
 Copie de substitution..... 1-9

---

### D

DIP. 1-7, 1-9, 1-14, 2-7, 2-8, 2-11, 2-12, 3-6, 4-12, 4-17, 4-33, 4-34, 4-35, 4-43, 4-55, 5-2, 5-11, 6-1, 6-2, 6-3, 6-5, 6-6, 6-8

---

### E

Edition..... 1-10, 5-10

---

### G

Groupement d'Archives ..... 1-11, 6-7

---

### I

Information d'empaquetage ... 1-9, 1-11, 1-14, 1-15, 2-5, 2-7, 2-8, 2-10, 3-2, 4-8, 4-31, 4-32, 4-34, 4-41, 4-43, 5-4, 5-5, 5-6, 5-8, 5-9  
 Information d'identification ..... 1-12, 4-29  
 Information d'intégrité..... 1-12, 3-3, 4-30  
 Information de contexte ..... 1-11, 4-30  
 Information de description .. 1-11, 2-5, 2-7, 2-10, 2-11, 3-2, 4-1, 4-2, 4-6, 4-7, 4-10, 4-17, 4-32, 4-34, 4-36, 5-4, 6-4  
 Information de provenance .. 1-12, 4-30, 4-36, 4-45  
 Information de représentation 1-9, 1-11, 1-12, 1-13, 1-15, 2-3, 2-4, 2-5, 2-6, 2-7, 2-10, 3-2, 3-3, 3-4, 3-5, 4-20, 4-21, 4-22, 4-23, 4-24, 4-25, 4-26, 4-27, 4-28, 4-34, 4-41, 5-7, 5-8, 5-11, 5-13, 2, 3, 1  
 Information de structure ..... 1-12, 4-22

---

### M

Migration numérique . 1-5, 1-9, 1-13, 1-15, 3-3, 5-1, 5-2, 5-4, 5-5, 5-9, 5-10

---

### N

Notice descriptive .. 1-10, 1-13, 4-36, 4-37, 4-43, 4-45, 4-46, 4-47, 4-53, 5-10, 6-5, 6-6

---

### O

Opération de diffusion de données. 1-13, 2-11, 2-12

Opération de versement .. 1-14, 1-15, 2-10,  
2-11, 4-52, 4-53  
Outil d'accès .1-13, 1-14, 4-36, 4-37, 4-43,  
6-6  
Outil de commande 1-14, 4-37, 4-38, 4-41,  
4-43, 4-47, 4-55  
Outil de recherche .. 1-14, 4-37, 4-38, 4-43,  
4-47, 4-54, 6-2, 6-5, 6-6  
Outil de récupération... 1-14, 4-37, 4-43, 4-  
47

---

**P**

PDI. 1-7, 1-9, 1-11, 1-12, 1-14, 1-15, 1-16,  
2-5, 2-7, 2-8, 2-10, 3-2, 3-3, 3-4, 3-5, 4-  
8, 4-9, 4-13, 4-29, 4-30, 4-31, 4-32, 4-  
34, 4-35, 4-36, 4-37, 4-38, 4-40, 4-41,  
4-43, 4-45, 4-46, 4-47, 4-52, 4-53, 5-4,  
5-5, 5-6, 5-7, 5-8, 5-9, 5-10, 5-11, 5-13,  
6-6  
Protocole de commande.. 1-13, 1-15, 2-11,  
2-12, 4-55  
Protocole de versement 1-14, 1-15, 2-10, 2-  
11, 4-11, 4-13, 4-51

---

**R**

Rafraîchissement de support 1-15, 4-8, 5-4,  
5-5, 5-6

---

**S**

SIP 1-8, 1-9, 1-11, 1-14, 2-7, 2-10, 2-11, 4-  
1, 4-2, 4-6, 4-7, 4-11, 4-12, 4-13, 4-15,  
4-33, 4-34, 4-35, 4-51, 4-52, 4-53, 6-1,  
6-2, 6-3, 6-8

---

**T**

Transformation..1-15, 4-8, 5-5, 5-7, 5-8, 5-  
9, 5-10

---

**V**

Version..... 1-16, 5-7, 5-8, 5-9, 5-10