

PAIMAS et PAIS

Producer-Archive Interface Methodology Abstract Standard Producer-Archive Interface Specification

Danièle Boucon (daniele.boucon@cnes.fr)

Présentation Groupe PIN

19 mai 2008

Le plan

- Des standards pour qui, quand et pourquoi ?
- Objectif
- Liens entre OAIS, PAIMAS, PAIS
- Principes de base du PAIMAS
- Principes de base du PAIS
- Le prototype
- Les applications
- Historique, état actuel
- Accéder aux documents
- Conclusion

Des standards pour qui, quand et pourquoi?

- **Pour qui ?**
 - ◆ **Producteur et Archive**

- **Quand ?**
 - ◆ **Très tôt au moment de la première prise de contact,**
 - ◆ **Tout au long de la période de production et de transfert des éléments d'information,**

- **Pourquoi ?**
 - ◆ **Pour prendre en compte tous les éléments possibles de la négociation,**
 - ◆ **Pour formaliser tout le processus,**
 - ◆ **Pour fournir des pistes afin d'anticiper sur les évolutions éventuelles.**

Objectif

- **Document de méthodologie PAIMAS**
 - ◆ **C'est un guide pratique visant à :**
 - dans un cadre générique :
 - identifier les différentes phases dans le processus de transfert d'information entre producteur et archive,
 - définir les objectifs de chaque phase, les actions associées et les sorties attendues,
 - appliquer ce cadre à la spécification pour une communauté identifiée (normes particulières, terminologie, objets d'info typiques du domaine, ...),
 - permettre l'identification et le développement d'outils facilitant le transfert d'information entre producteur et archive.
 - ◆ **Obtention d'un protocole de versement (« submission agreement ») entre producteur et archive.**

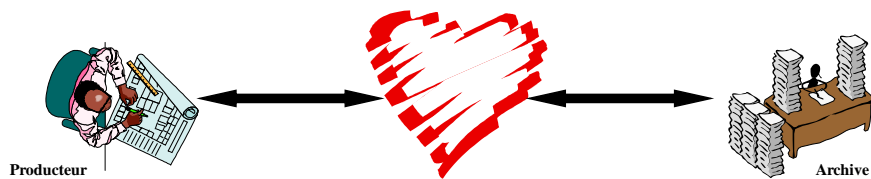
- **PAIS**
 - ◆ **Standard de mise en oeuvre de l'interface Producteur-Archive PAIMAS**

- **XFDU (XML Formatted Data Units)**
 - ◆ **Standard d'empaquetage des paquets d'information à verser (Submission Information Package SIP)**



Liens entre OAIS, PAIMAS, PAIS

- PAIMAS = **Interface** entre la fonction **Entrées « ingest »** du modèle OAIS d'une part, et le **producteur** d'autre part



Liens entre OAIS, PAIMAS, PAIS

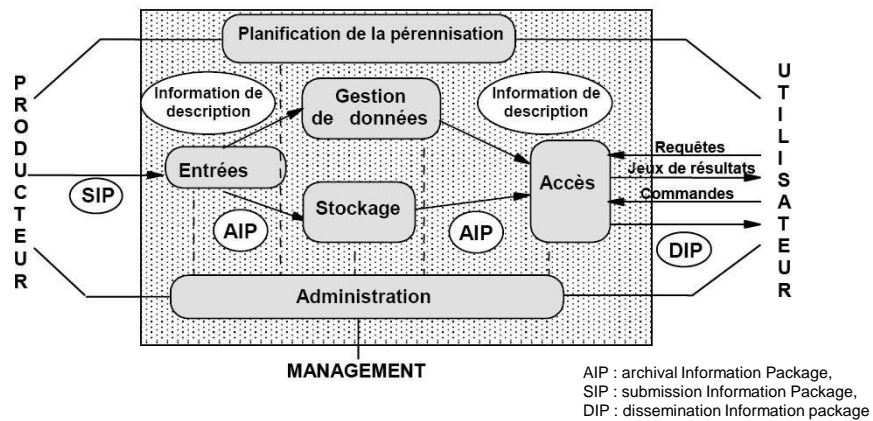
- Définition de la fonction Entrées « ingest » issue du modèle de référence OAIS

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.

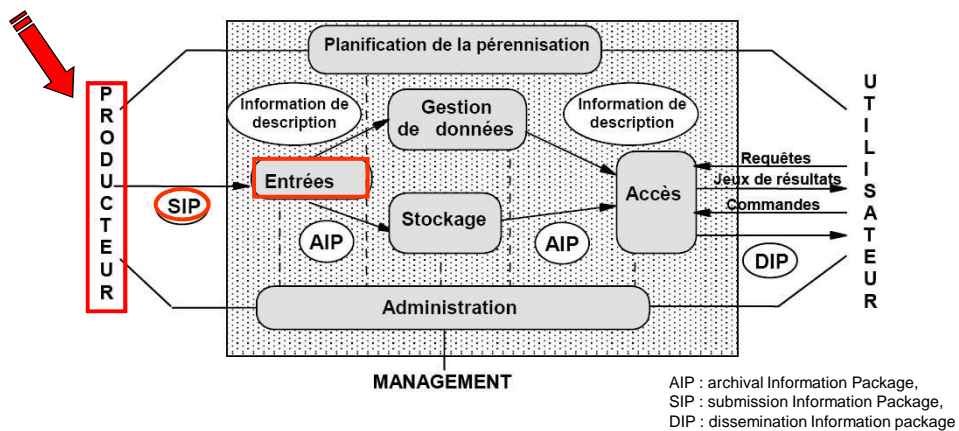
Liens entre OAIS, PAIMAS, PAIS

■ Les grandes fonctions du modèle OAIS



Liens entre OAIS, PAIMAS, PAIS

■ Les grandes fonctions du modèle OAIS



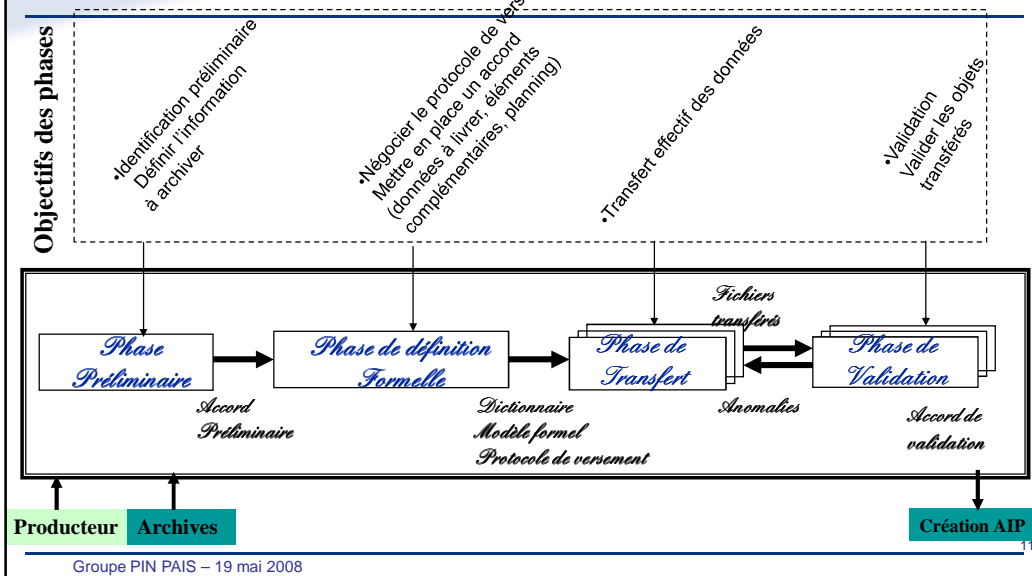
Liens entre OAIS, PAIMAS, PAIS

- **PAIS = standard de mise en oeuvre de l'interface Producteur-Archive PAIMAS**
 - ◆ **Conforme avec le standard PAIMAS**
 - ◆ **Conforme avec le standard XFDU**

Principes de base du PAIMAS

- **4 phases principales :**
 - ◆ **phase préliminaire,**
 - ◆ **phase de définition formelle,**
 - ◆ **phase de transfert,**
 - ◆ **phase de validation.**
- **Tableaux d'actions (exhaustivité, liste actions/sous actions, acteurs ...).**
- **Spécification pour une communauté.**

Principes de base : les phases du PAIMAS



Principes de base : les phases du PAIMAS

■ Phase préliminaire

Premier contact.	
Définition préliminaire, Faisabilité, évaluation	Information à archiver.
	Objets numériques et standards appliqués à ces objets.
	Quantification.
	Références des objets.
	Conditions de sécurité.
	Aspects légaux et contractuels.
	Opérations de transfert.
	Validation.
	Planning.
	Impact permanent sur l'Archive.
	Résumé des coûts et des risques.
	Points critiques.
Etablissement d'un accord préliminaire	

Id	Preliminary phase: quantification	Involves
P-19	Estimate the data volume to be transmitted to the Archive	Producer
P-20	Assess the permanent data volume to store	Archive
P-21	Assess the storage capability need for the ingest process	Archive
P-22	Assess the associated costs	Archive

Principes de base : les phases du PAIMAS

■ Phase de définition formelle

Mise en place de l'organisation		
Définition formelle	Information à préserver et modèle des objets de données à livrer.	Contexte général du projet et définition des objets d'information.
		Creation d'un dictionnaire.
		Construction d'un modèle formel.
	Formalisation des aspects légaux et contractuels.	
	Définition des conditions de transfert.	
	Définition de la validation.	
	Planning de livraison.	
	Gestion des changements.	
	Faisabilité et évaluation.	
Protocole de versement		

13

Principes de base : les phases du PAIMAS

■ Phase de transfert

Effectuer les tests
Gérer le transfert



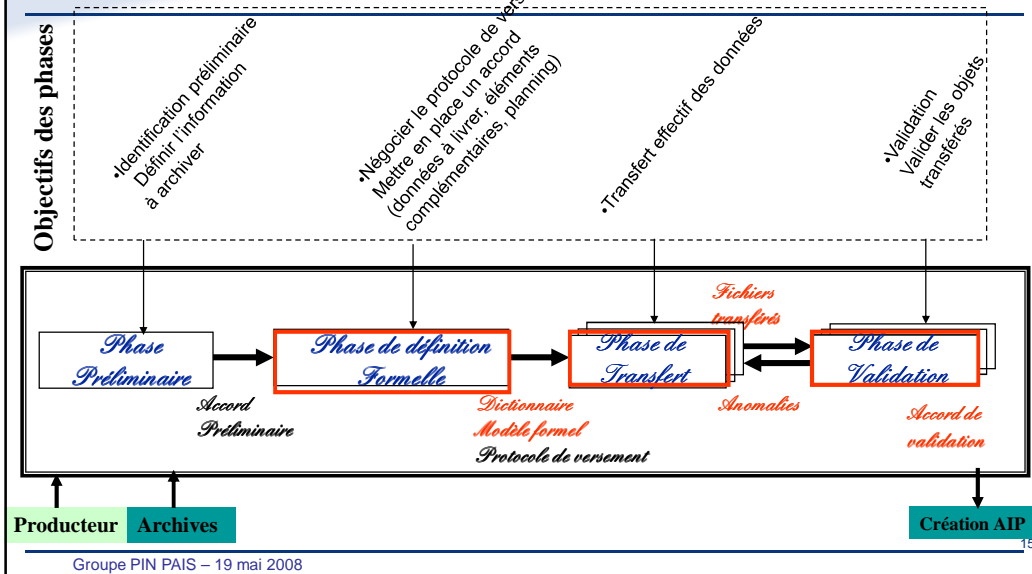
■ Phase de validation

Effectuer les tests de validation
Gérer la validation

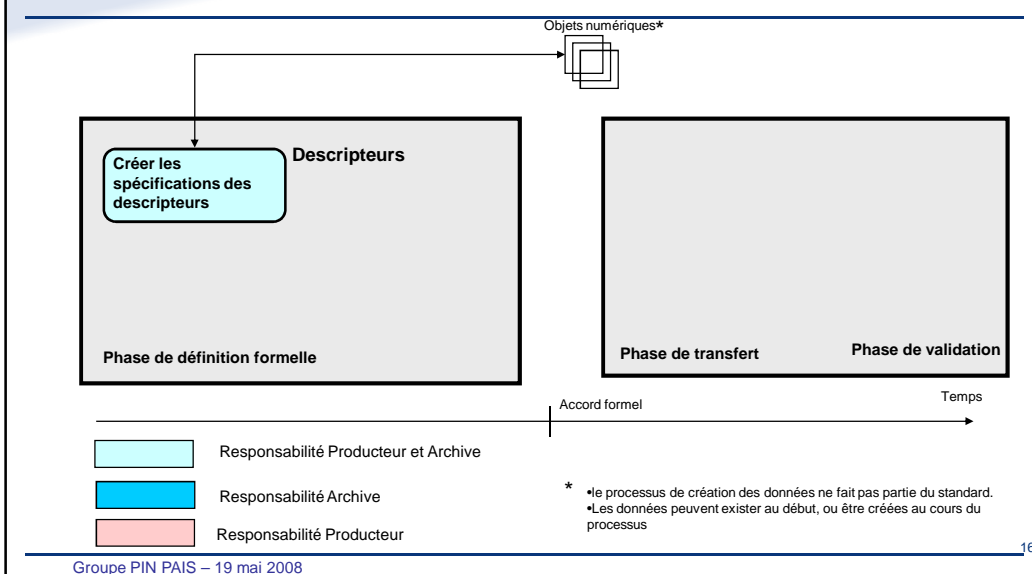


14

Principes de base du PAIS



Principes de base du PAIS



Principes de base du PAIS : description des objets à transférer

- Besoin constaté de formalisation de la description des données (défini aujourd'hui au cas par cas) :

=> Représentation exhaustive et formelle (interprétable par programme) des objets à transférer.

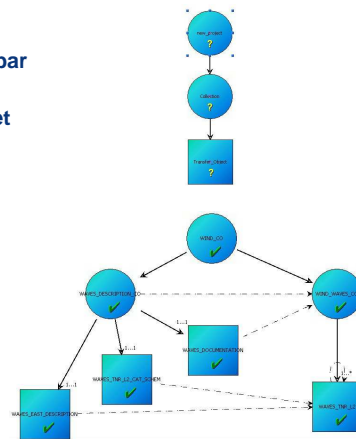
=> Chaque type d'objet à transférer ("Transfer Object") et chaque collection sont décrits par un "Descripteur"

- ♦ 2 XML schema génériques

- Ces descripteurs sont organisés sous forme d'arbre (relations parents/enfants) :

- ♦ MOT (Model of Objects for Transfer)

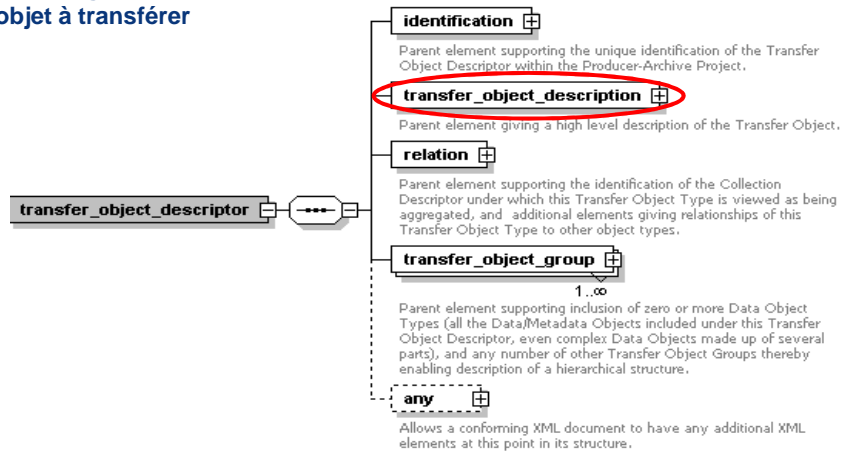
- Description possible d'associations entre différents objets du modèle.



17

Principes de base du PAIS : description des objets à transférer

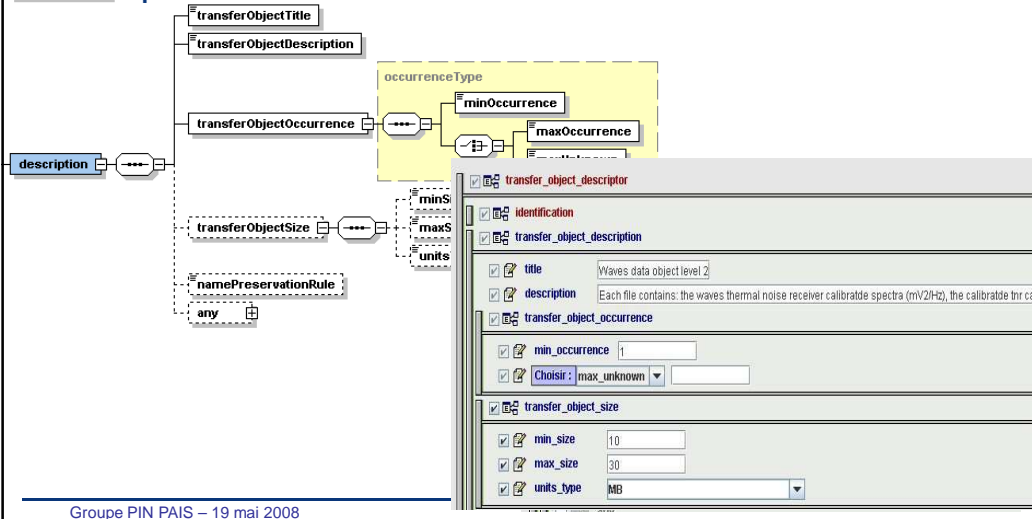
- Structure générale d'un descripteur d'objet à transférer



18

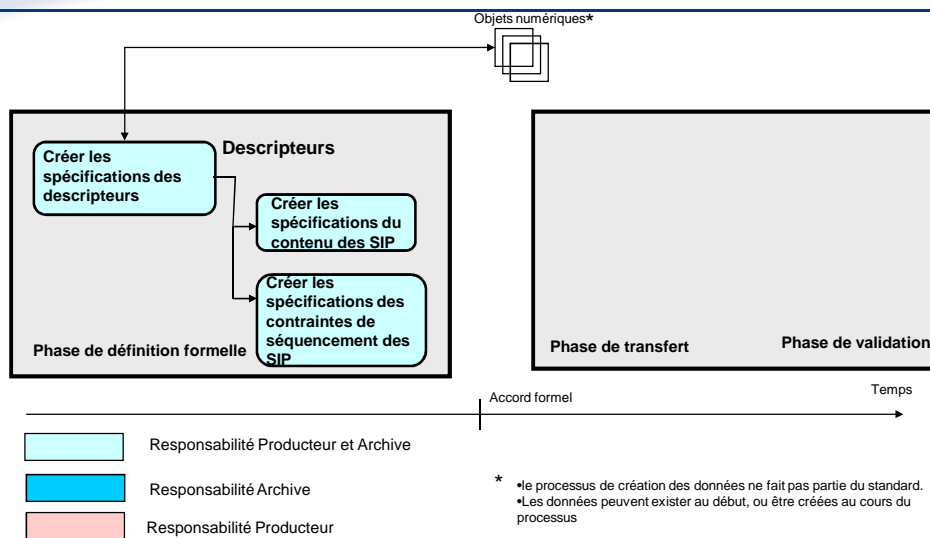
Principes de base du PAIS : description des objets à transférer

Description



Groupe PIN PAIS – 19 mai 2008

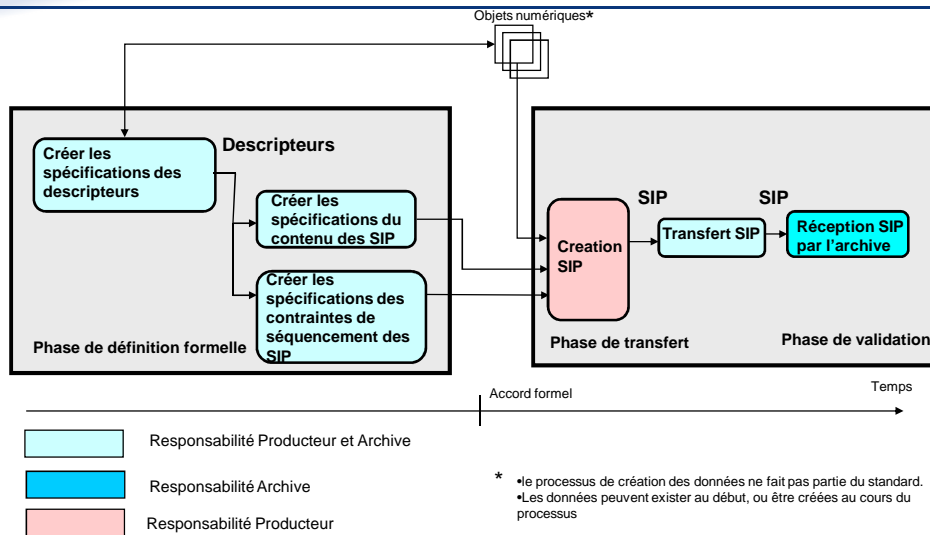
Principes de base du PAIS



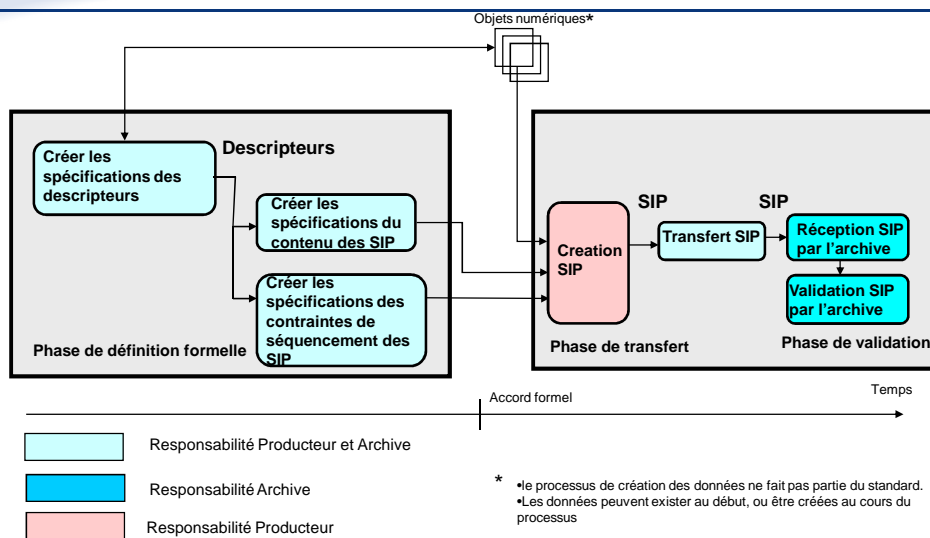
Groupe PIN PAIS – 19 mai 2008

20

Principes de base du PAIS



Principes de base du PAIS



Principes de base du PAIS : SIP et XFDU

- Paquet XFDU = XML manifest à l'intérieur d'un fichier ZIP => vue logique du contenu du paquet XFDU.
- Spécialisation du XFDU aux besoins du PAIS (identifiants du producteur, du projet, des descripteurs ...) => modèle de SIP
- Les validations sont effectuées sur le manifest XFDU (premier niveau de validation) :
 - ◆ Contenu des SIP,
 - ◆ Contraintes de séquençement des SIP,
 - ◆ Conformité avec les descripteurs :
 - Contenu des descripteurs : nombre d'objets attendus, etc,
 - Structure des objets de données inclus,
- Le niveau de validation requis est spécifié par le projet d'archive.
- Remarque : on pourrait utiliser un autre standard d'emballage.

23

PAIS : structure du document

- | | |
|---|--|
| <ul style="list-style-type: none"> ■ 1. Introduction ■ ... ■ 2 Overview <ul style="list-style-type: none"> ■ 2.1 General framework ■ 2.2 Formally describing Data Objects for Transfer ■ 2.2 SIP creation and validation ■ 3 Descriptions of Producer Data <ul style="list-style-type: none"> ■ 3.1 Transfer Object Descriptor Specification <ul style="list-style-type: none"> ■ <i>General and detailed description: abstract view + concrete implementation</i> ■ 3.2 Collection Descriptor Specification <ul style="list-style-type: none"> ■ <i>General and detailed description: abstract view + concrete implementation</i> ■ 3.3 Specialization of the Descriptor Models ■ 3.4 Design of the MOT | <ul style="list-style-type: none"> ■ 4 Description of constraints on types of SIPs <ul style="list-style-type: none"> ■ 4.1 SIP Content Types <ul style="list-style-type: none"> ■ 4.1.2 SIP Content Types Specification ■ 4.1.3 SIP Content Types Implementation ■ 4.2 SIP sequencing constraints <ul style="list-style-type: none"> ■ 4.2.1 SIP sequencing constraints Specification ■ 4.2.2 SIP sequencing constraints Implementation ■ 5 Abstract SIP Model ■ 6 Concrete SIP implementation using XFDU <ul style="list-style-type: none"> ■ 6.1 SIP XML schema for XFDU ■ 6.2 Mapping Abstract SIP to XFDU <ul style="list-style-type: none"> ■ 6.2.1 SIP Global Information ■ 6.2.2 Transfer Object identification and status information ■ 6.3 SIP specialization ■ ANNEX A: Transfer Object and Collection Descriptors XML schema ■ ANNEX B : legend for XML figures ■ ANNEX C : informative references |
|---|--|

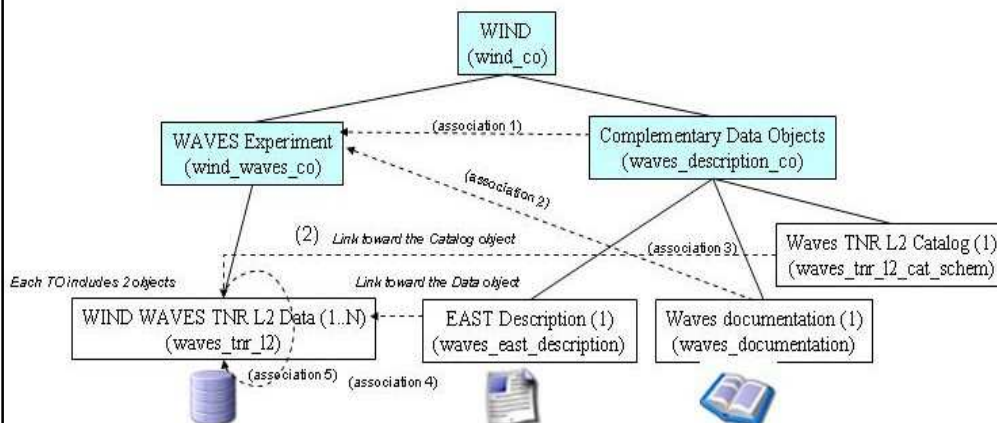
24

■ 2 parties :

- ♦ **Creation et visualisation du modèle** (au cours de la phase de définition formelle).
- ♦ **Validation des SIP** (contenu des SIP, contraintes de séquençement, structure des objets livrés),

Et : suivi et visualisation du transfert (au cours de la phase de transfert)

- Etat des objets livrés par rapport au modèle et visualisation du suivi.
- Avec la même représentation graphique.



Prototype pour la conception du MOT et le suivi du transfert

- XSD schema pour les modèles de descripteurs
- Fichiers XML pour les descripteurs
- XAmple, JGraph (Open source)

- Pré-requis
 - ◆ JAVA 1.5

- Compatible avec différents OS (PC, MAC, UNIX)

Les applications (1/2)

- PAIMAS :
 - ◆ Utilisation par la BnF dans le cadre du projet SPAR,
 - ◆ FEDORA (logiciel open source de gestion d'archive pour les bibliothèques) : « Ingest Guide for University Electronic Records », Fedora and the preservation of university records,
 - ◆ CNES :
 - Le PAIMAS fait partie du Référentiel Normatif du CNES.
 - Mise en application dans le cadre du CNES avec le document « Méthodologie pour l'archivage à long terme des données spatiales d'un projet » (DECLIC, ...).

- PAIS
 - ◆ Prototype sur un cas d'étude du CDPP,
 - ◆ NASA : étude pour la NARA.

Les applications (2/2)

■ OAIS, PAIMAIS et PAIS

- ◆ En cours au CNES : mise en application sur un prototype générique d'archivage au CNES dans un cadre plus large prenant en compte tout le processus d'ingestion (logique de modélisation et validation du PAIS, protocole de versement du PAIMAS, chaîne de récupération et d'archivage ...),

- SERAD Service de Référencement et Archivage des Données.

Historique

- OAIS : janvier 2002 – version livre bleue CCSDS, ISO 14721 en 2003
- PAIMAS : mai 2004 – version livre bleue CCSDS du PAIMAS, ISO 20652 en 2006
- PAIS

- ◆ Septembre 2004: concept paper

draft White Book versions



- ◆ Octobre 2008: proposition de norme CCSDS équivalente d'un Draft International Standard (DIS).

- ◆ Devrait être accompagnée d'un tutoriel

- Le standard PAIS inclut la définition des concepts et leur mise en oeuvre.
- Le tutoriel (Green Book) expliquera comment utiliser le standard avec des exemples.

Etat actuel

- **OAIS en cours de révision (« 5 year review »).**
- **Passage prévu pour l'automne 2008 du PAIS vers un « Red Book » norme CCSDS équivalente d'un Draft International Standard (DIS).**
- **Première version attendue d'un tutoriel.**
- **Prototype PAIS conforme à la version actuelle, développé en parallèle (CNES et NASA), nouvelle version pour octobre 2008.**

31

Accéder aux documents

- **OAIS** : via le serveur PIN pin.cnes.fr ou le serveur CCSDS <http://public.ccsds.org/publications/BlueBooks.aspx>
- **PAIMAS** : via le serveur PIN pin.cnes.fr ou le serveur CCSDS
 - ♦ *Producer-Archive Interface Methodology Abstract Standard*. Blue Book. CCSDS 651.0-B-1 Issue 1. May 2004, <http://public.ccsds.org/publications/BlueBooks.aspx>
- **XFDU**
 - ♦ Livre bleu, en cours de revue officielle pour la version définitive (sera fait sous peu) : via le serveur CCSDS <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206610R1/CCSDSAgency.aspx>
- **PAIS**
 - ♦ Me demander (daniele.boucon@cnes.fr) car pas encore de publication officielle

32

Conclusion

- **PAIMAS et PAIS : 2 standards pour gérer les processus d'entrée des données dans l'archive (méthodologie et mise en œuvre).**
- **Possibilité d'utiliser d'autres standards d'empaquetage (METS par exemple).**
- **Retours, questions ?**

33



Fin de la présentation

34

Rappel de définitions

1. 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).

2. 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.

3. 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.

4. Objet à Transférer (Transfer Object) : ensemble structuré et organisé d'un ou de plusieurs Objet-données qui doivent être transférés vers l'Archive.