



CENTRE NATIONAL D'ÉTUDES SPATIALES

PAIMAS et PAIS

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

Danièle Boucon (daniele.boucon@cnes.fr)

Présentation Groupe PIN

7 octobre 2010

Le plan

- **Des standards pour qui, quand et pourquoi ?**
- **Objectifs**
- **Liens entre OAIS et PAIMAS**
- **Description du PAIMAS**
- **Principes de base du PAIS, prototype**
- **Les applications**
- **Etat actuel, accès aux documents**
- **Conclusion, questions**

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 du projet d'archivage,**
- ◆ **Pour fournir des pistes afin d'anticiper sur les évolutions éventuelles.**

Objectifs

■ Document de méthodologie **PAIMAS**, guide pratique, visant à :

- Cadre générique** {
 - ◆ Identifier les différentes **phases** du projet d'archive pour le transfert d'informations entre producteur et archive,
 - ◆ Définir les **objectifs** de chaque phase, les **actions** associées et les **sorties** attendues,
- Projet particulier** {
 - ◆ **Spécifier** ce cadre pour une communauté identifiée (normes particulières, terminologie, objets typiques du domaine ...),
- ◆ Obtenir un **protocole de versement** (« submission agreement ») entre producteur et archive.

=> Eviter que chaque service d'archives élabore sa propre méthodologie (travail dupliqué, manque de généricité, diversité des outils)

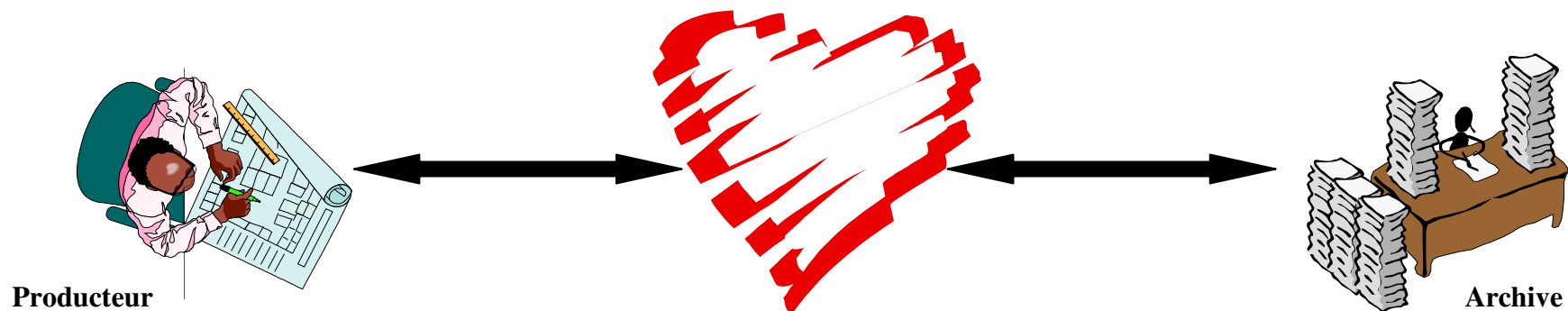
■ PAIS

- ◆ Standard de **mise en oeuvre** de l'interface Producteur-Archive PAIMAS
- ◆ S'appuie sur XFDU (XML Formatted Data Units), standard d'empaquetage des paquets d'information à verser (Submission Information Package SIP)



Liens entre OAIS, PAIMAS (1/2)

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



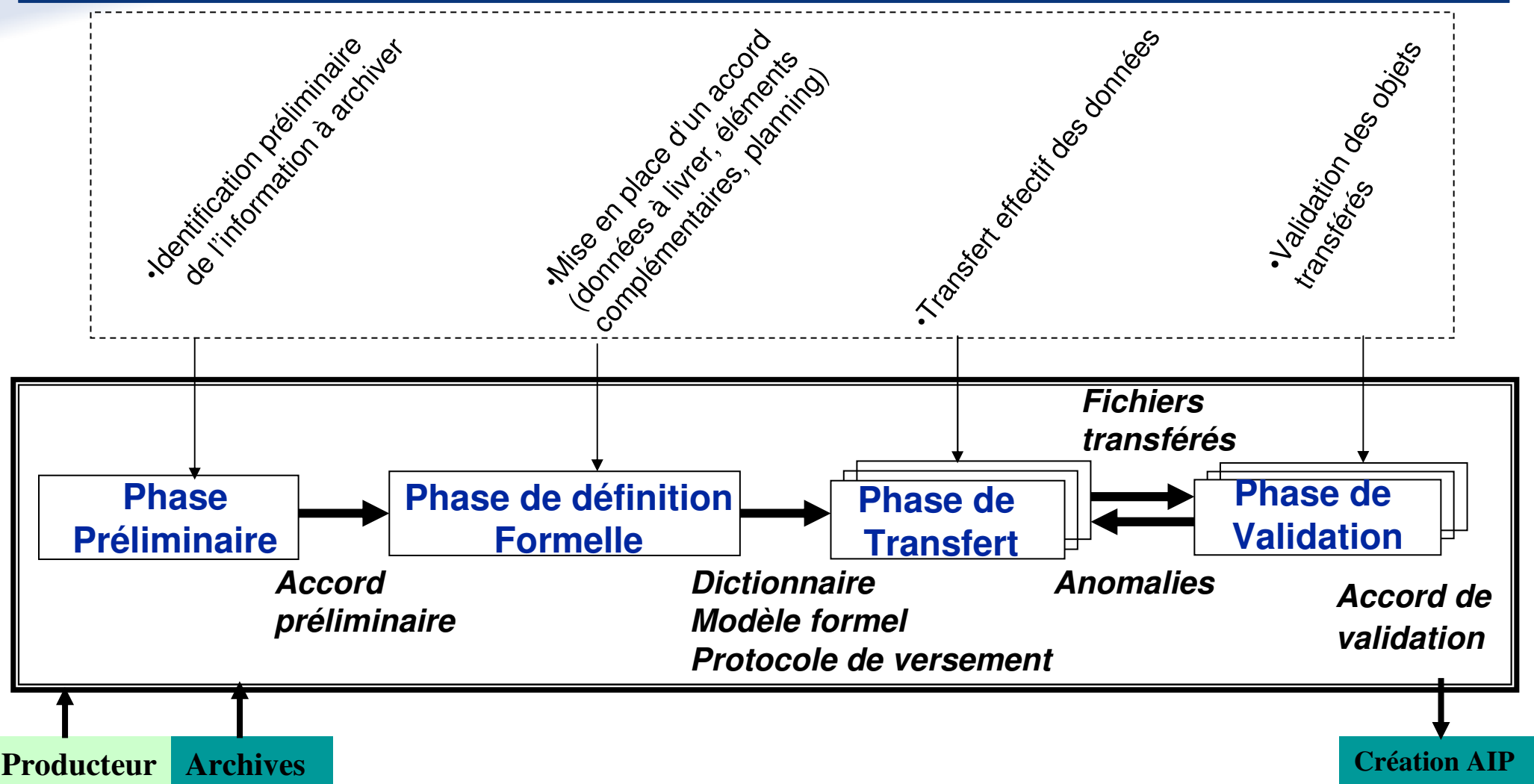
Principes de base du PAIMAS

- **4 phases principales, découpées elles mêmes en sous-phases (ordre chronologique) :**
 - ◆ phase préliminaire,
 - ◆ phase de définition formelle,
 - ◆ phase de transfert,
 - ◆ phase de validation.

- **Chaque sous-phase est constituée d'un ensemble d'actions regroupées dans des Tableaux d'actions (exhaustivité, liste actions/sous actions, acteurs ...)**
(pas d'ordre chronologique).

- **Spécification pour une communauté.**

Principes de base : les phases du PAIMAS (objectifs, résultats)



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	

P-19 Estimate the data volume: The Producer must estimate the volumes to be transmitted in the short, medium and long term (global volume, minimum, average, and maximum planned size of files, number of files), as well as the frequency of the transfer sessions. These elements have an influence on the technique used for the transfer.

Id	Phase préliminaire : quantification	Acteurs
P-19	Evaluer le volume des données à transmettre à l'Archive	Producteur
P-20	Evaluer le volume de données permanents à stocker	Archive
P-21	Evaluer les capacités de stockage nécessaire dans le processus d'ingestion	Archive
P-22	Evaluer les coûts associés	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.	
	Definition des conditions de transfert.	
	Définition de la validation.	
	Planning de livraison.	
	Gestion des changements.	
	Faisabilité et évaluation.	
Protocole de versement		

Principes de base : les phases du PAIMAS

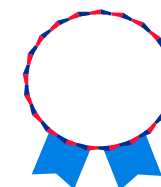
■ Phase de transfert

Effectuer les tests
Gérer le transfert

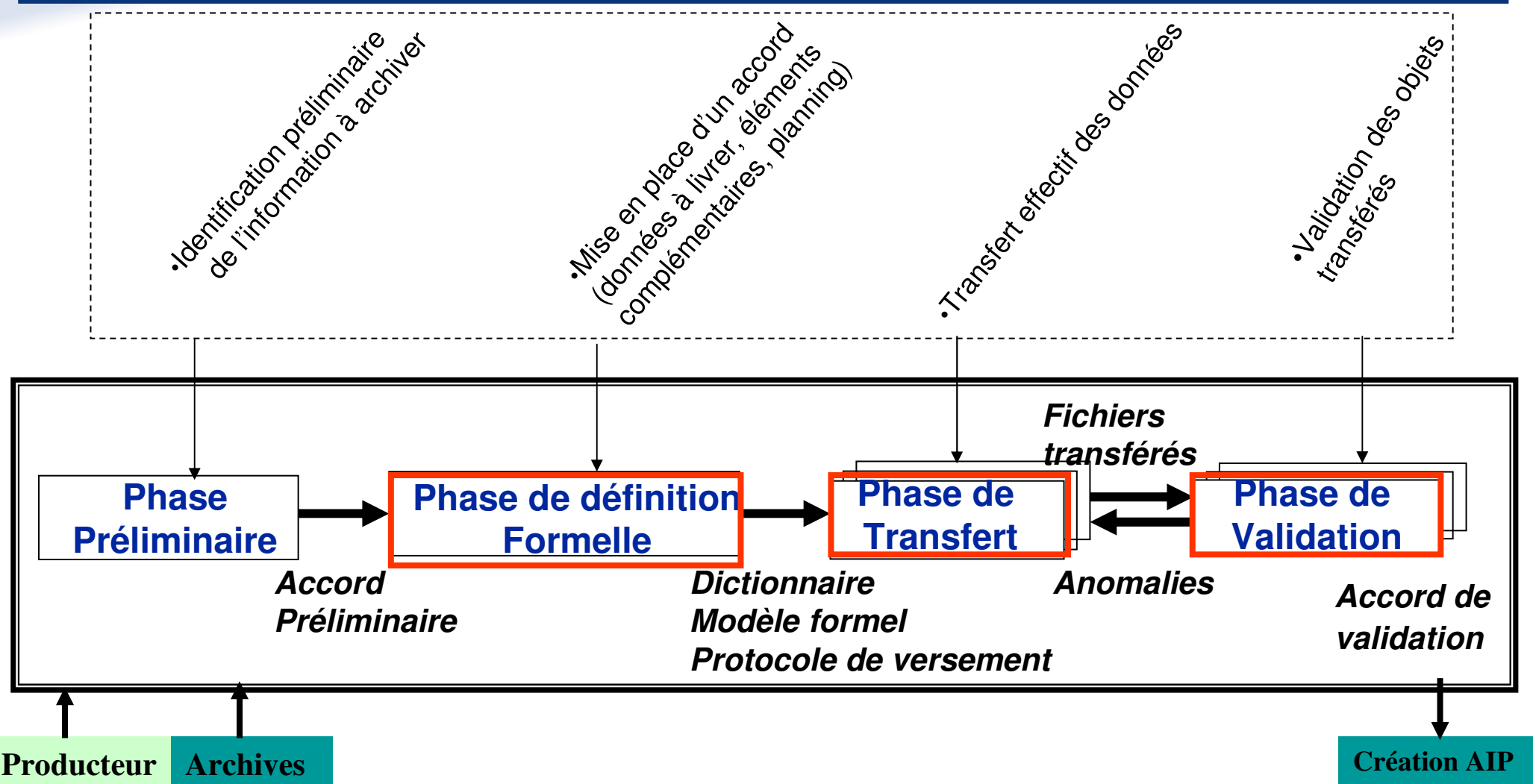


■ Phase de validation

Effectuer les tests de validation
Gérer la validation



Principes de base du PAIS

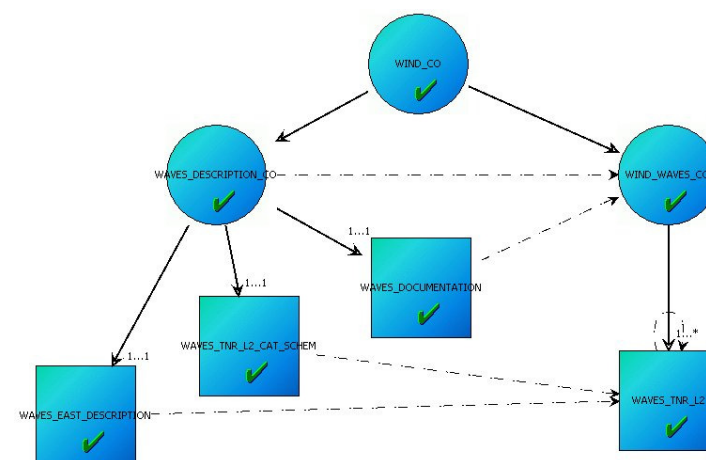


Principes de base du PAIS (phase de définition formelle)

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		

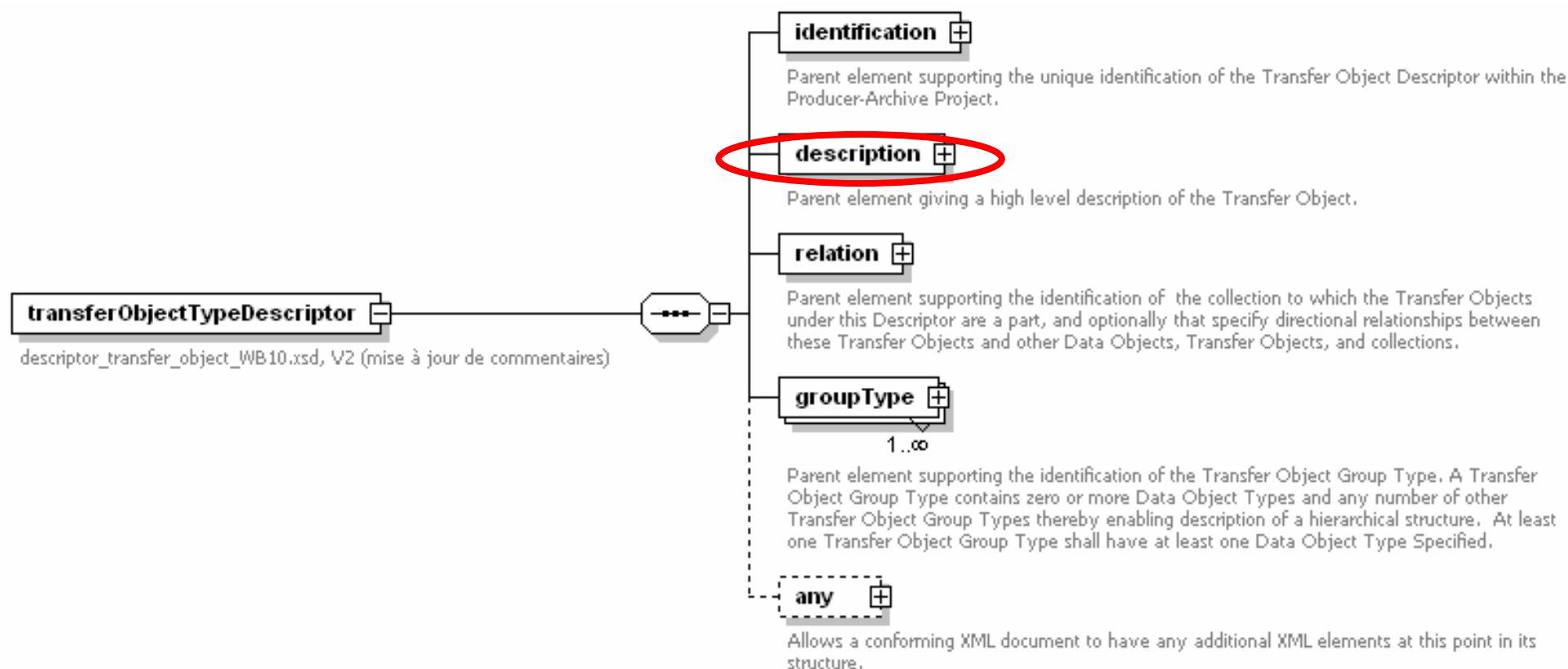
- Dictionnaire générique pour les objets à transférer (« Transfer Object ») et collections : descripteurs, vue abstraite et implémentation (XML schema)
- Ces descripteurs instanciés sont organisés sous forme d'arbre (relations parents/enfants, possibilité de décrire les relations entre objets) pour créer le “Model of Objects for Transfer” (MOT)

- Modèle générique de SIP : vue abstraite et implémentation (XML schema)
- Modèle de contraintes s'appliquant sur les SIP : vue abstraite et implémentation (XML schema)
- Empaquetage XFDU intégrant le modèle de SIP



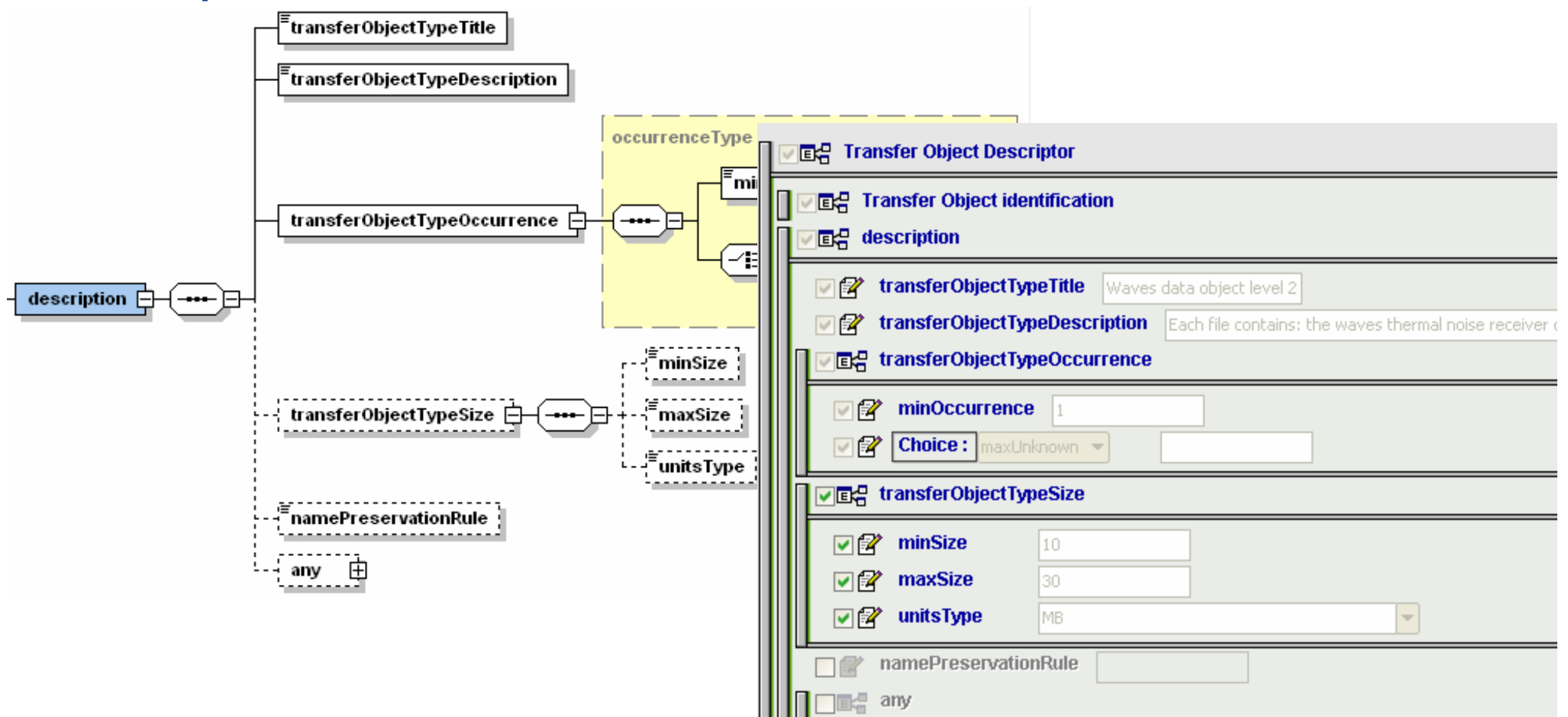
Principes de base du PAIS : description des objets à transférer

■ Structure générale d'un descripteur d'objet



Principes de base du PAIS : description des objets à transférer

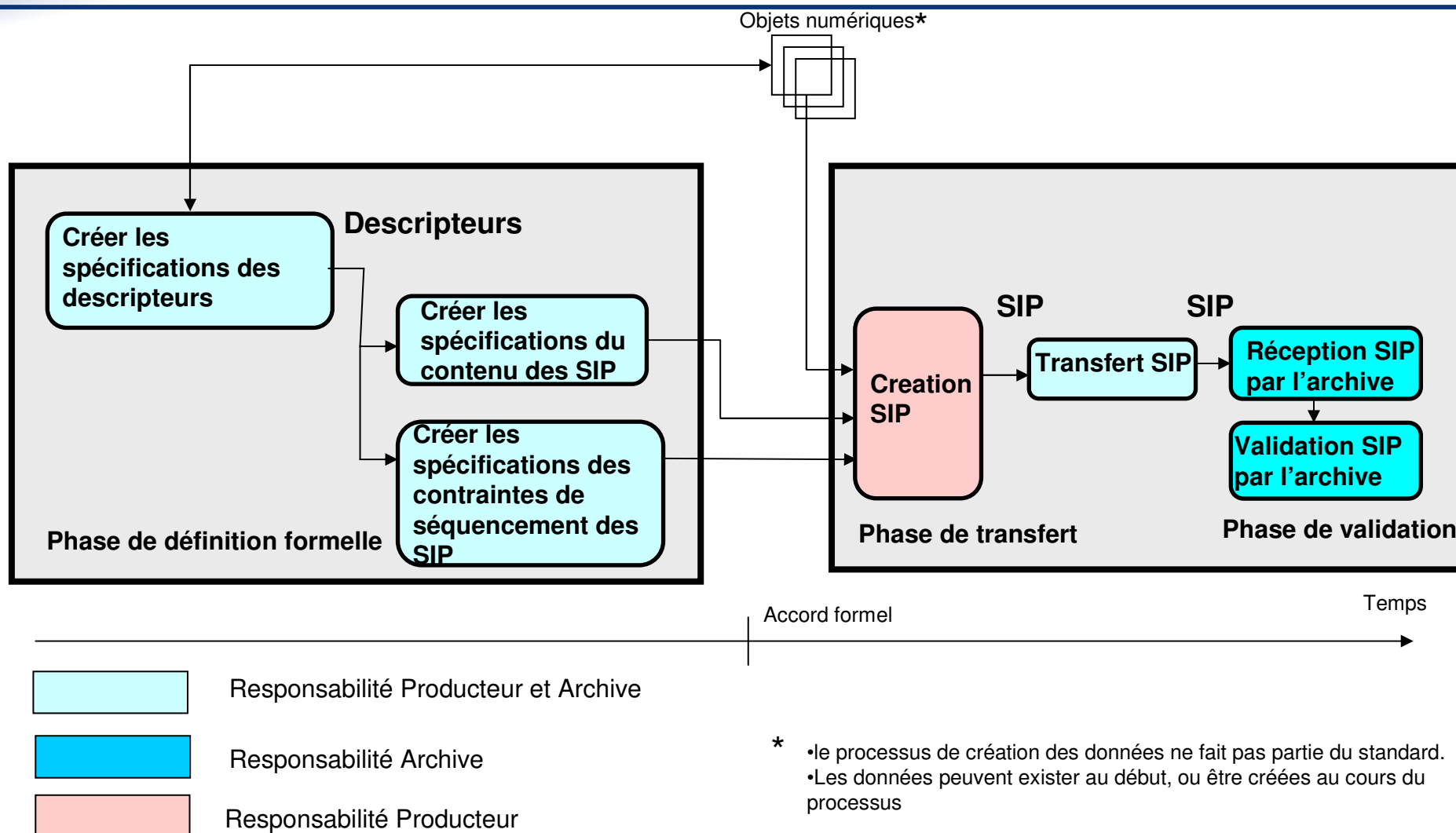
■ Description



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.**

Principes de base du PAIS



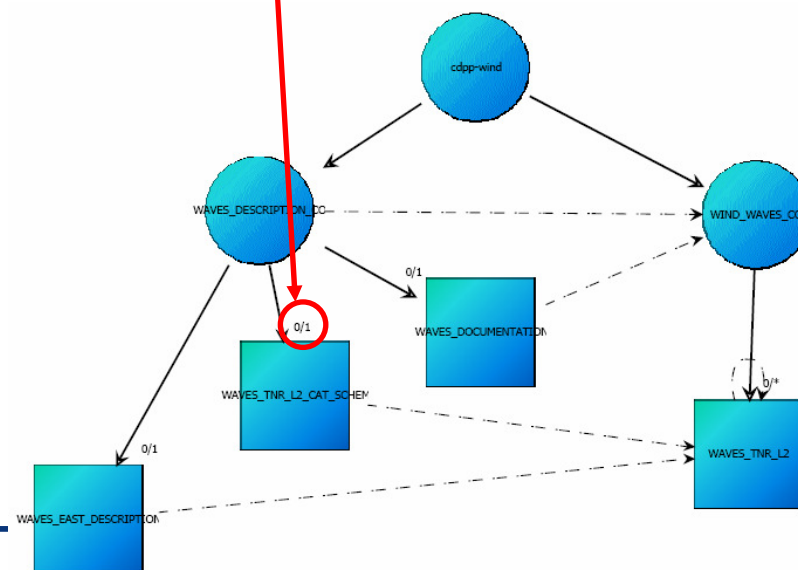
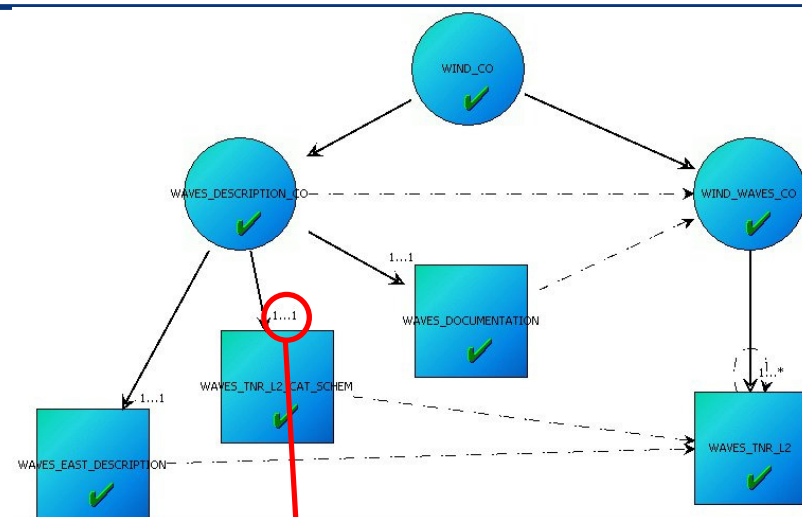
Prototype pour la conception du MOT et le suivi du transfert

■ 2 parties :

- ◆ **Création 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.



Etat actuel du prototype

- **Prototype PAIS développé par le CNES, suit l'évolution du PAIS.**
- **La NASA développe de son côté un prototype pour générer des SIP conformes XFDU et intégrant le modèle de SIP.**

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) pour les formulaires et les graphes
- Base PostgreSQL

- Pré-requis
 - ◆ **JAVA 1.6**

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

Les applications (1/2)

■ PAIMAS :

- ◆ **Utilisation par la BnF dans le cadre du projet SPAR** (phase initiale pour construire les fonctionnalités),
- ◆ **FEDORA** (logiciel open source de gestion d'archive pour les bibliothèques) : « **Ingest Guide for University Electronic Records** », **Fedora and the preservation of university records**, <http://dca.lib.tufts.edu/features/nhprc/reports/ingest/index.html>
- ◆ **CINES** (cf présentation d'O. Rouchon)
- ◆ **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 » et plus particulièrement du CDPP* -spécialisation du PAIMAS au contexte particulier des données spatiales : interface avec des scientifiques, diversité des formats, gros volumes de données, ...
 - Sert de base à l'élaboration des plans d'archivage (des jeux de données des missions spatiales).

=> Application dans le cadre du SERAD (Service de Référencement et d'Archivage des Données), cadre plus formel pour un élargissement à l'ensemble des données, projet en cours de développement.

* CDPP Centre des Données de la Physique des Plasmas

Etat actuel des documents, accès

- **OAIS** Blue Book CCSDS et norme ISO (CCSDS 650.0-B-1 January 2002, ISO 14721:2003) :
 - ◆ via le serveur PIN http://www.aristote.asso.fr/PIN/documents/OAIS_650_0b1.pdf ou le serveur CCSDS <http://public.ccsds.org/publications/BlueBooks.aspx>, version française http://www.aristote.asso.fr/PIN/documents/norme_oais_version_francaise.pdf
 - ◆ magenta book CCSDS, en cours de révision ISO, versions officielles prévues courant 2011 (ainsi que la mise à jour de la version française).

- **PAIMAS** Magenta Book CCSDS et norme ISO (CCSDS 651.0-M-1 May 2004 , ISO 20652 : 2006)
 - ◆ via le serveur CCSDS <http://public.ccsds.org/publications/MagentaBooks.aspx>

- **PAIS** toujours en cours de standardisation CCSDS.
 - ◆ Me demander (daniele.boucon@cnes.fr) car pas encore de publication officielle

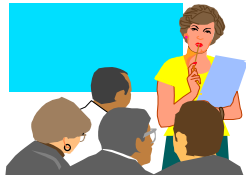
- **XFDU** Blue Book CCSDS et norme ISO (CCSDS 661.0-B-1 September 2008, ISO 13527:2010)
 - ◆ via le serveur CCSDS <http://public.ccsds.org/publications/BlueBooks.aspx>

Conclusion

- **PAIMAS (méthodologie) : standard pour gérer les processus d'entrée des données dans l'archive et les relations entre Producteur et Archive.**

- **PAIS (mise en œuvre) :**
 - ◆ **contient les modèles génériques (modèles abstraits et XML schema),**
 - ◆ **l'empaquetage est basé sur XFDU (possibilité d'utiliser d'autres standards d'empaquetage comme METS).**

- **Retours, questions ?**



Fin de la présentation

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.