

## Interface Producteur-Archive, méthodologie

Présentation Groupe PIN  
15 janvier 2003

**Danièle Boucon**  
**SILOGIC**  
daniele.boucon@cnes.fr

**Claude Huc**  
**CNES**  
claude.huc@cnes.fr



## Le plan

- ◆ Constat
- ◆ Pour qui, objectif du standard
- ◆ Contexte
- ◆ Historique
- ◆ Description du document
- ◆ Les phases
- ◆ Application à une communauté particulière
- ◆ Les études de cas
- ◆ Conclusion

### Le constat (1/3)

- ◆ Les relations entre archives et producteurs de données sont rarement simples et faciles, comme le prouvent nombre de difficultés (constatées dans tous les domaines) :
  - non conformité des objets reçus par les archives,
  - ce que le producteur livre n'a pas été clairement défini,
  - planning de livraison non respecté,
  - erreurs dans les données archivées détectées tardivement,
  - ...
- ◆ Ces difficultés peuvent nuire à la qualité de l'information archivée et au coût de l'opération.



### Le constat (2/3)

- ◆ La diversité de plus en plus grande des producteurs (et les contraintes accrues de planning).
- ◆ Existence du modèle de référence OAIS.
- ◆ L'expérience acquise dans le domaine des Centres de Données Scientifiques (formalisation : modèles d'information, dictionnaire de données), confrontée à de nombreuses autres archives (état des lieux).



- ➔ Perception du besoin d'une méthodologie d'interface Producteur-Archives (application du RM OAIS).
- ➔ Maturité dans l'archivage, expérience dans le domaine du dictionnaire --> niveau de formalisation.

### Le constat (3/3)

- ◆ Besoin confirmé par une étude préliminaire de cas à partir d'éléments fournis par :
  - CAC - Centre des Archives Contemporaines (MN Leblanc)
  - Royal Library of Netherland, projet NEDLIB
  - CNES / CDPP - Centre des Données de la Physique des Plasmas (C. Huc, D. Boucon)
  - NASA / NSSDC (National Science Space Data Center), Projet Image
  - NARA - the U. S. National Archives and Records Administration
  
- ◆ **Chaque service d'archives a élaboré sa propre méthodologie : travail dupliqué, pas de généralité, diversité des outils ...**

### Un standard pour qui et quand ?

- ◆ Producteur et Archives,
  - Très tôt au moment de la première prise de contact,
  - Pour prendre en compte tous les éléments possibles de la négociation,
  - Pour formaliser tout le processus,
  - Tout au long de la période de production et de transfert des éléments d'information,
  - Pistes pour anticiper sur les évolutions éventuelles.

### Objectif du document de méthodologie

- ◆ 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 archives,
    - 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 archives.
- ◆ Obtention d'un accord de soumission entre producteur et archives.



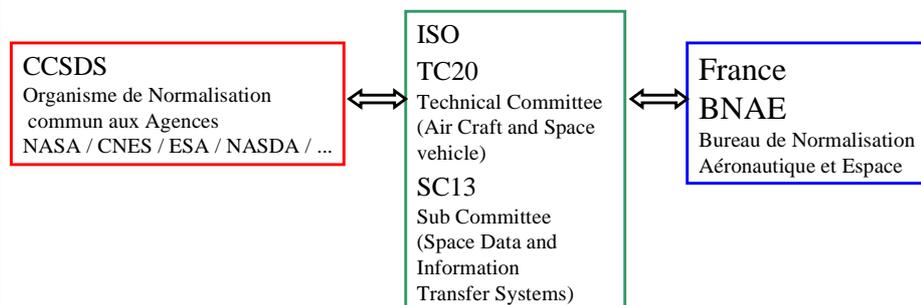
15 janvier 2003

Interface Producteur-Archive

7

### Le contexte (1/6)

- ◆ Travail mené dans le cadre du CCSDS (Consultative Committee for Space Data Center)



- ◆ D'autres communautés ont été invitées à participer à ce travail : archivistes (NARA, ...) et bibliothécaites (BNF, RLG-Research Library Group).

15 janvier 2003

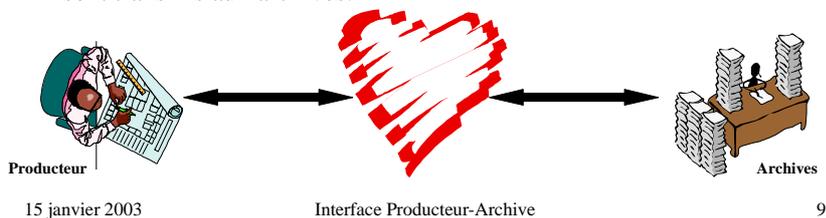
Interface Producteur-Archive

8

**Le contexte (2/6)**

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

- Vue des archives : rappel des grandes fonctions du modèle OAIS et de la partie « ingest »,
- Vue du producteur : fonctions qui visent à produire des objets qui sont transmis aux archives.



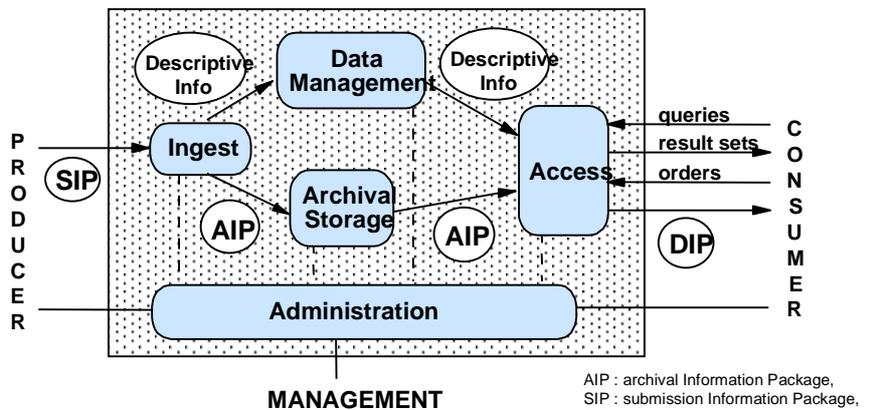
15 janvier 2003

Interface Producteur-Archive

9

**Le contexte (3/6)**

◆ Les grandes fonctions du modèle OAIS



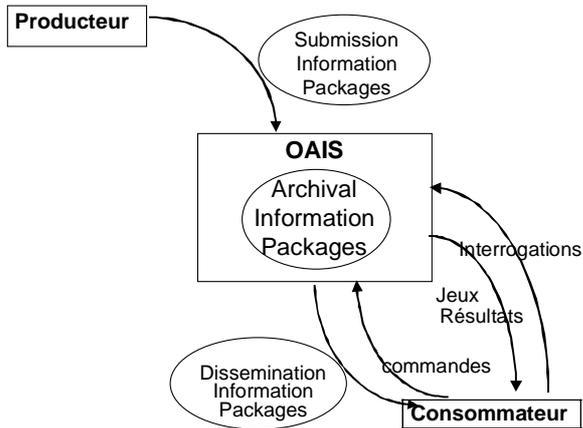
15 janvier 2003

Interface Producteur-Archive

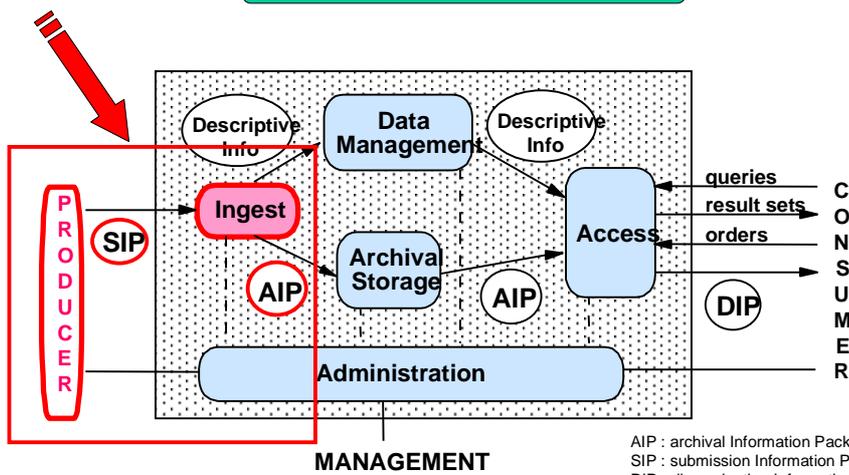
10

**Le contexte (4/6)**

◆ Le flot des données



**Le contexte (5/6)**



AIP : archival Information Package,  
 SIP : submission Information Package,  
 DIP : dissemination Information package

### Le contexte (6/6)

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

*The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepare Archival Information Packages for storage,*

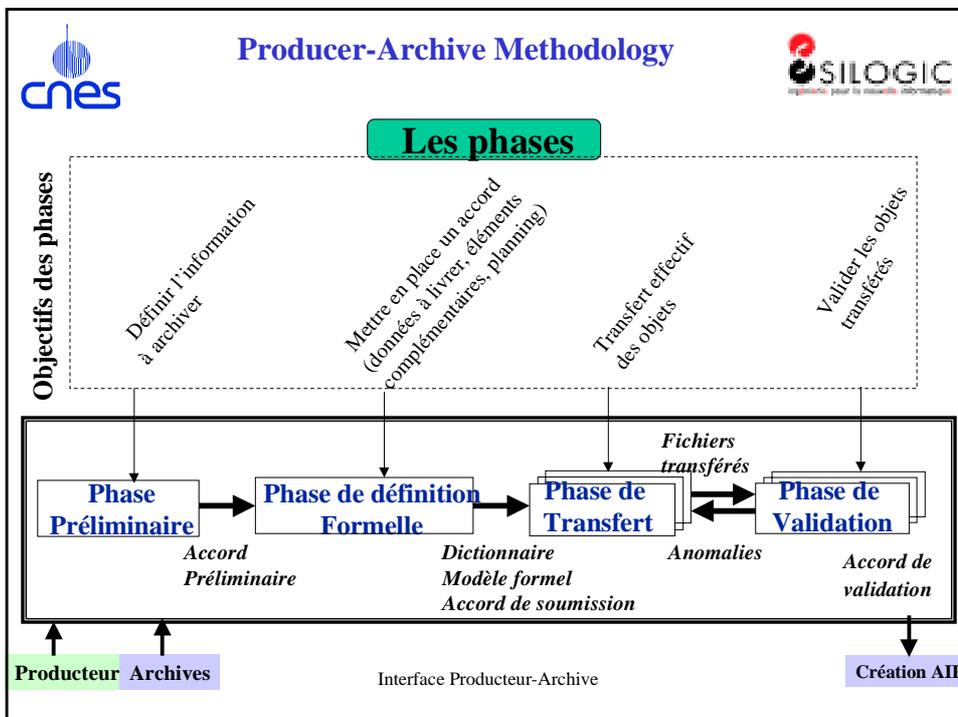
*and ensure that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.*

### Historique

- ◆ **Novembre 2000** - 3 cas d'étude (édition, archives contemporaines, données scientifiques) et 1ère analyse des principales étapes du processus d'ingest.
- ◆ Mai 2001- *ingest process concept paper issue 2*. Identification des phases.
- ◆ Juillet 2001 - *ingest process methodology* - version préliminaire
- ◆ Octobre 2001 - *Producer - Archive interface methodology guide*. 1ère version
- ◆ Décembre 2001 - *Producer-Archive interface methodology Abstract Standard* (White Book).
- ◆ **Juin 2002** - *Producer-Archive Interface Methodology, Abstract Standard, CCSDS-651.0-W-2 White Book*.
- ◆ **Novembre 2002** : **Draft Red Book**.
  
- ◆ **Décembre 2002** : **Red Book proposal, équivalent d'un DIS (Draft International Standard)**

**Description du document**

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



### Les phases : Phase préliminaire

<b>Premier contact.</b>	
<b>Définition préliminaire, Faisabilité, évaluation</b>	Information à archiver.
	Objets numériques et standards appliqués à ces objets.
	<b>Quantification.</b>
	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.
<b>Etablissement d'un accord préliminaire</b>	

15 janvier 2003

Interface Producteur-Archive

17

### Les phases : Phase 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

15 janvier 2003

Interface Producteur-Archive

18

**Les phases : Phase de définition formelle**

<b>Mise en place de l'organisation</b>		
<b>Définition formelle</b>	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.		
<b>Accord de soumission</b>		



**Les phases : Phase de définition formelle**

Dictionnaire : ex de quelques types d'objets définis dans le dictionnaire du CDPP

- ◆ Description de collections (de données, de documents, d'images)
- ◆ nœuds et liens du graphe
- ◆ Description d'objets de données, de stockage
- ◆ Description de documents textuels (forme électronique)
- ◆ Références bibliographiques
- ◆ ...

**Les phases : Phase de définition formelle**

Dictionnaire : ex de structure d'un objet de données, et ingestion d'un objet de données -issu du CDP (langage PVL, 1995)

Ce qui est dans le dictionnaire

```
BEGIN_GROUP = ENTITY_DEFINITION;
NAME = DATA_OBJECT_DESCRIPTION;
MEANING = "Description d'un objet de donnees";
COMPONENT = (DATA_OBJECT_IDENTIFIERS,1..1);
COMPONENT = (ASCENDING_NODE,1..N);
COMPONENT = (START_DATE,0..1);
COMPONENT = (STOP_DATE,0..1);
COMPONENT = (ORBIT_NUMBER,0..N);
COMPONENT = (TRANSFORMATION_SO_DO,0..1);
COMPONENT = (DATA_STORAGE_OBJECT_IDENTIFIERS,1..N);
END_GROUP = ENTITY_DEFINITION;
```

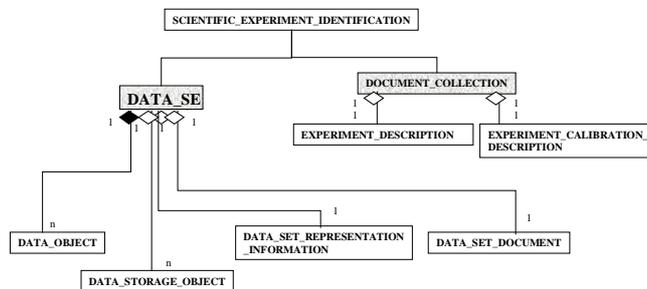
Ce qui est livré avec le fichier de données

```
BEGIN_GROUP=DATA_OBJECT_DESCRIPTION;
DATA_OBJECT_IDENTIFIERS=ROBE_1981_1125_1819;
ASCENDING_NODE=DA_TC_SPECTRO_ROBE;
START_DATE=1981-11-25T18:19:00;
STOP_DATE=1981-11-25T18:40:36;
DATA_STORAGE_OBJECT_IDENTIFIERS=ROBE_1981_1125_1819;
END_GROUP=DATA_OBJECT_DESCRIPTION;
```



**Les phases : Phase de définition formelle**

Modèle formel : exemple dans le cas du CDP



**Les phases : Phase de transfert**



Effectuer les tests
Gérer le transfert

**Les phases : Phase de validation**

Effectuer les tests de validation
Gérer la validation



### Application à une communauté particulière (1/2)

- ◆ Adapter le standard générique à une communauté particulière (allant d'un organisme international, à un simple service d'archives)
- ◆ les étapes pour définir un standard de communauté :
  - remplacer la terminologie générique par la terminologie du domaine,
  - définir un dictionnaire des données et un modèle d'information qui seront utilisés par les projets d'archivage,
  - identifier d'autres standards utiles au projet,
  - identifier des outils communs

### Application à une communauté particulière (2/2)

- ◆ Analyser chaque action du standard générique afin de statuer si :
  - elle est applicable au contexte,
  - elle ne s'applique pas,
  - elle doit être modifiée,
- ◆ Procéder à l'addition de nouvelles actions si nécessaire.
- ◆ Conseils pour la création du standard de communauté :
  - définir une communauté aussi large que possible,
  - définir un comité de rédaction du standard représentatif (du domaine et de sa diversité),
  - publier régulièrement le travail en cours (pour le soumettre à critique, nouvelles propositions, validation, ...),
  - soumettre le standard de communauté à un organisme de normalisation si nécessaire.

### Les études de cas (1/2)

L'objectif était de valider l'application du standard à différents contextes

- ◆ CNES - Centre des Données de la Physique des Plasmas (CDPP).
- ◆ NASA - (life science research in a government agency, Space Life Sciences Data Archive - 2 cas).
- ◆ National Archives and Records Administration (US Federal Government records : National Health Interview Survey).
- ◆ BnF - Site WEB.

### Les études de cas (2/2)

- ◆ Cohérence entre la méthodologie présentée et les cas d'étude --> faisabilité.
- ◆ Ont permis de proposer quelques modifications (ajouts/précisions/suppressions).
- ◆ Pas de difficulté apparente pour appliquer le standard.
- ◆ Phase de réflexion sur les processus actuels et comment les améliorer.

## Conclusion



- ◆ Evolution du standard vers une forme définitive.
- ◆ Application dans le cadre du CDPP : pour l'étude préliminaire, les tests et validation, le dictionnaire des objets ...
- ◆ Outil d'aide à l'application : réflexion en cours pour un prototype.
- ◆ Travaux visant à préciser tout l'aspect formalisation.
- ◆ Interrogation : l'intérêt in fine que les différentes communautés vont pouvoir tirer de ce standard



**Fin de la présentation**