



## Un standard applicable au CNES

Archivage long terme des données spatiales

Règles et recommandations applicables aux projets

Claude Huc  
Présentation Groupe PIN  
11 octobre 2006



### Un projet spatial est organisé en phase

- Selon une logique de déroulement complètement normalisée
- Norme NF EN 13290-4
- Phase 0 - analyse de la mission, identification des besoins
- Phase A – faisabilité
- Phase B - définition préliminaire
- Phase C – définition détaillée
- Phase D – réalisation et qualification  
De la phase 0 à la phase D : 5 à 10 ans
- Phase E – utilisation
- Phase F – retrait de service
- la phase E varie de 2 à 15 ans suivant les missions



## Un projet spatial est organisé en phase

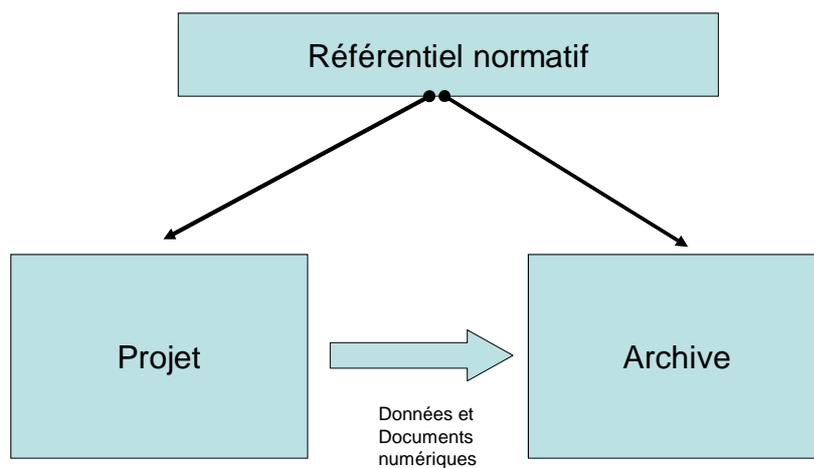
### ■ Au-delà de la phase D

- ◆ Les équipes en charge du développement sont pour l'essentiel réaffectées à d'autres missions

### ■ Au-delà de la phase F

- ◆ Les équipes en charge de l'exploitation du satellite) et des données qu'il produit sont réaffectées à d'autre missions

### ■ D'où la nécessité de considérer les projets comme des services versant par rapport à une archive pérenne





## Objectif de l'archivage des données spatiales

### ■ Permettre à des utilisateurs futurs

- ◆ de récupérer,
- ◆ utiliser,
- ◆ analyser,

dans un certain contexte, des données qui ont été créées dans un autre contexte.

**Il est donc essentiel pour cela, que les données puissent être parfaitement interprétées sans erreur, sans distorsion, par des chercheurs, des ingénieurs, des utilisateurs qui n'ont pas participé au projet initial.**



## 4 conditions minimales

1. Les données à préserver et les informations associées devront pouvoir être lues sur des plateformes informatiques futures, non connues à ce jour,
  1. conservation
  2. Neutralité par rapport au contexte de création
2. Les données doivent être décrites tant sur le plan de leur structure (la syntaxe) que de leur signification (la sémantique) afin qu'il soit toujours possible de passer du train de bits '000011000010100011.....' à une information intelligible,
3. Les données doivent pouvoir être retrouvées au sein de l'ensemble des données sous la responsabilité de l'Archive,



## 4 conditions minimales

4. L'usage et l'interprétation de ces données ne doivent généralement pas nécessiter une connaissance approfondie des instruments embarqués ; en effet la connaissance des technologies utilisées pour construire ces instruments peut s'avérer obsolète à terme et la conservation d'une expertise sur des technologies anciennes pose des problèmes encore plus complexes à résoudre que la préservation des données.

→ Ceci implique clairement le besoin d'archiver des données étalonnées.

- Ces règles générales minimales devront être complétées et enrichies en fonction des spécificités du Projet ou de l'Archive.
- Il appartiendra aux acteurs principaux, le Projet et l'Archive, de définir ces compléments éventuels ainsi qu'aux experts externes sollicités dans le cadre des revues, de formuler des recommandations dans ce domaine.



## Structure des règles et recommandations

<b>Référence</b> Règle ou Recom	<b>libellé</b>
------------------------------------	----------------

La *description* de la règle ou de la recommandation [si nécessaire],

- les éventuelles *exceptions* à la règle [si nécessaire],

- la **justification** de la règle ou de la recommandation [

- des illustrations par des **exemples** ou **contre-exemples** concrets de mise en oeuvre si nécessaire



### 3 règles de base de haut niveau

1. *Les informations qui constituent l'objectif principal de l'archivage doivent être définies avec précision, en étroite collaboration entre le Projet et l'Archive.*
2. *Les communautés utilisatrices de ces informations doivent être clairement définies.*
3. *L'archivage des informations issues d'expériences spatiales se traduit par :*
  - *L'archivage des données les contenant :*
    - ◆ *données principales,*
    - ◆ *données auxiliaires (orbite, attitude, etc.),*
  - *L'archivage des métadonnées associées aux données,*
  - *L'archivage de la documentation.*



### Règles relatives à la syntaxe

#### ■ Données neutres

- ◆ *Les fichiers livrés ne doivent pas contenir d'information spécifique du système d'exploitation créateur ou gestionnaire de ces données avant leur livraison.*

#### ■ Fichiers à structure d'enregistrement répétitifs

- ◆ Structures séquentielles

#### ■ Fichiers XML

- ◆ *Les fichiers structurés en XML sont autorisés, à condition que l'organisation des données soit conforme à un **schéma XML** qui doit se référer à une norme reconnue ou, s'il est spécifique, être livré avec les données.*



## Règles relatives à la syntaxe

### ■ Bases de données

- ♦ *Les bases de données qui peuvent être transmises à l'Archive doivent être des bases de données XML qui ne dépendent d'aucun SGBD particulier.*
- ♦ *Lorsqu'il s'agit de conserver les données organisées au sein d'une base **sans exigence de conserver la base en état de fonctionner**, le Projet devra procéder à l'extraction, sous une forme documentée, des données de la base et des schémas de définition de tables afin de les transférer à l'Archive sous une forme indépendante de tout logiciel.*



## Règles relatives à la syntaxe : codage

### ■ Codage des nombres dans les fichiers de données

*Les codages autorisés pour les nombres entiers et réels sont :*

- *le codage ISO 646,*
- *le codage binaire IEEE avec la convention MSB pour les nombres flottants,*
- *la représentation binaire en complément à 2 avec la convention MSB pour les nombres entiers.*

### • Codage des caractères

*Les codages autorisés pour les chaînes de caractères dans les fichiers de données sont les codages ISO10646, notamment le codage UTF-8.*



## Règles relatives à la syntaxe : Cas des images

### ■ Images raster

*On utilisera, selon le cas, le format PNG ou le format TIFF pour archiver des images « raster ».*

### ■ Formats vecteur

*On utilisera, selon le cas, le format CGM ou le format SVG pour archiver des images « vecteur ».*



## Règles relatives à la syntaxe : Données documentaires

### ■ Règle générale

*Les formats documentaires acceptés sont :  
les formats structurés XML et HTML,  
le format PDF/A.*

### ■ Règles applicables aux formats structurés

*Les documents XML doivent respecter les conditions suivantes :*

- le schéma ou la DTD les validant doit se référer à une norme reconnue ou, s'il est spécifique, être livré avec les documents,*
- absence de liens externes ou de liens pouvant dépendre de l'environnement de stockage.*

*Les directives de présentation des documents validés par un schéma XML ou une DTD spécifique seront livrées avec les documents sous la forme d'une feuille de style XSL.*



## Règles relatives à la syntaxe : Données documentaires

---

### ■ cas des documents issus des suites bureautique

- ◆ *Les fichiers issus de la bureautique doivent être mis en conformité avec la norme ISO 26300 Open Document Format for Office Applications (OpenDocument) v1.0*

### ■ Cas des documents HTML

- ◆ *Les documents HTML doivent respecter les conditions suivantes :*
- ◆ *les liens entre pages sont relatifs au document et ne dépendent pas de son environnement (fichiers ou serveurs externes),*
- ◆ *les caractères sont codés selon la norme ISO 10646.*

### ■ Format PDF/A

- ◆ *Les documents dont on sait qu'ils ne pourront être modifiés ou dont on voudra empêcher qu'ils le soient pourront être livrés au format PDF/A.*



## Règles relatives à la syntaxe : Nomenclature

---

- *Les noms des collections de données et les noms des fichiers doivent être uniques au sein du Projet.*
- *Le Projet définira une nomenclature cohérente et lisible, et si possible conforme à l'existant dans le domaine ou la discipline.*



## Règles relatives au contenu

- *Le Projet fournira un moyen de caractériser la validité et la qualité des données*
- *La représentation des temps et des dates sera conforme au niveau 1 de la recommandation CCSDS 301.0-B Time Code Formats*
- *Les latitudes, longitudes et altitudes seront représentées conformément à la norme ISO 6709 (cf. DR22). Dans le cas où cette norme ne serait pas retenue, il reste impératif que le Projet adopte un standard unique dans ce domaine et l'applique à toutes les données du Projet.*
- *Référencement spatio-temporel : Le référencement spatial d'un jeu de données suivra les recommandations de la norme ISO 19111 (cf. DR23), Le référencement temporel d'un jeu de données suivra les recommandations de la norme ISO 19108 (cf. DR24).*



## Métadonnées

- **Métadonnées pour la recherche des données dans l'archive**
  - ◆ **Description des collections**

*Les métadonnées caractérisant les collections de données devront être conformes à une norme de métadonnées, choisie, si possible, parmi l'une des normes Dublin Core, DIF, ISO 19115.*

*A défaut, on retiendra une norme reconnue dans la discipline concernée : astronomie, planétologie, physique des plasmas, etc.*
  - ◆ **Description des granules**

*Les métadonnées caractérisant les granules devront être identifiées*



## Métadonnées pour l'utilisation des données

■ *Pour chaque jeu de données à archiver, la description des données sera conforme aux règles pour l'élaboration de documents de description d'interfaces informatiques, définies dans le document RNC-CNES-E-40-507*

■ Ces descriptions doivent être :

- ◆ Exhaustives : Tous les champs de données doivent être décrits,
- ◆ Exactes : C'est à dire conformes aux données présentes dans les fichiers,
- ◆ Homogènes d'un jeu de données à l'autre, d'une expérience à l'autre, afin de faciliter la tâche de l'utilisateur de ces descriptifs.



## Autres sujets abordés

- Données auxiliaires
- Données d'aide à la recherche
- Documents et publications
- Recommandations spécifiques pour les logiciels



## Interface Projet-Archive

- *Le Projet et l'Archive suivront les recommandations de la norme « Producer-Archive Interface Methodology Abstract Standard » pour l'archivage des données*
- *Dans le processus d'archivage, l'Archive intervient au titre d'autorité habilitée pour vérifier le respect de la présente MP*



## Interface Projet-Archive

Processus de déroulement du Projet	Interactions Archive	Projet-	Documents livrés par le Projet	Commentaire
Travaux antérieurs				Identifier l'entité Archive.
Phase A - REP	<b>PHASE DE DÉFINITION PRÉLIMINAIRE</b>		Plan d'archivage, version préliminaire.	L'accord préliminaire entre le Projet et l'Archive doit être explicite au moment de la Revue des Exigences Préliminaires (REP).
Phase B - RDP	<b>Phase de définition formelle</b>		Plan d'archivage, version complète.	L'accord formel entre le Projet et l'Archive doit être explicite au moment de la Revue de Définition Préliminaire (RDP).
Phases C et D	<b>Phase de définition formelle</b>		Plan d'archivage en version définitive.	Définition, réalisation et qualification du segment sol.



## Interface Projet-Archive

Phase E - démarrage	<b>Phase de transfert et de validation</b> Recette préalable au lancement effectif du transfert : - Conformité des données par rapport à ce qui est attendu. - Bon fonctionnement du processus de transfert	Révision éventuelle du plan d'archivage en cours de phase E. Rapport de recette.	Une révision du plan d'archivage peut être envisagée en fonction des événements qui peuvent se produire pendant la phase E : défaillance totale ou partielle d'un instrument ou au contraire prolongation de la mission au-delà de la durée initialement prévue.
Phase E - routine	<b>Phase de transfert et de validation</b> Validation approfondie.		Les données et les documents à archiver ne sont pas tous livrés par le Projet au même moment. La cohérence et la complétude de l'ensemble doivent donc faire l'objet de validation en cours de phase E.
Phase F – retrait de service			L'Archive devient le site de référence unique pour les données du Projet qui sont archivées.