

*ingestion de métadonnées dans un système
multi missions :
des concepts à la pratique*

*Michel Nonon-Latapie
(michel.nonon@cnes.fr)*

CNES -Service "Valorisation des Données et Outils d'ingénierie"

Plan de la présentation

- **historique et objectifs recherchés**
- **solutions retenues :**
 - modèle de données et dictionnaires
 - exemples
- **cas d'utilisation opérationnelle**
- **études en cours et perspectives**

Historique et objectifs recherchés



- depuis plusieurs années, le CNES a développé des systèmes logiciels génériques permettant la mise en place de **Centres de données scientifiques** ou de systèmes d'accès à des données, notamment :
 - » le SIPAD (« Système d'Information, de Préservation et d'Accès aux Données »)
 - » et son successeur le SIPAD-NG (« SIPAD Nouvelle Génération »)

SERVER DE DONNÉES MAPS
Ce serveur rend accessibles les produits générés par le groupe MAPS de la mission Cassini. Ces produits étant à accès restreint, il faut s'inscrire sur le serveur pour les récupérer :
[Formulaire d'inscription](#)

Flats résumés (quelliers)
[Accès aux Données](#)

[Récupérer les fichiers commentés](#)
[Charger un contexte de travail existant](#)

[Cassini MAPS à l'Observatoire Midi Pyrénées \(OMP\)](#)

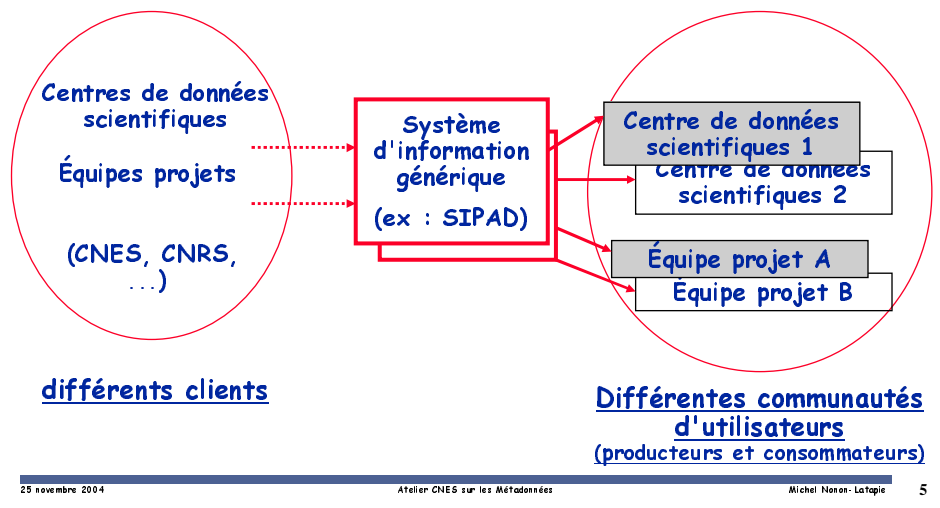
MAPS Data Server
This server provides access to the data processed by the MAPS group of the Cassini mission. These products having restricted access rights, you have to register before retrieving them :
[Fill the registration form](#)

Summary Flats (quelliers)
[Data Access](#)

[Retrieve ordered files](#)
[Load an existing working context](#)

[Cassini MAPS at the "Observatoire Midi Pyrénées" \(OMP\)](#)

Historique et objectifs recherchés



Historique et objectifs recherchés



- rôle capital des métadonnées dans ces systèmes : ces systèmes doivent permettre de sélectionner des informations, de les exploiter (transformations, analyses), et d'accéder à ... des "informations sur ces informations" (que sont les métadonnées !)
- or chaque projet client a en général des besoins de définition, de gestion et d'utilisation des métadonnées différents ... surtout d'une thématique scientifique à une autre :
 - définition et fourniture de métadonnées spécifiques
 - utilisation de certaines métadonnées pour rechercher des données
 - affichage structuré de métadonnées
 - ...etc...

trouver une solution alliant rigueur et souplesse

- ...rigueur pour :
 - définir précisément le modèle de données de chaque projet client
 - récupérer de façon exhaustive et fiable les métadonnées définies
- ...souplesse pour :
 - s'adapter aux besoins des projets clients (et de leur thématique), afin de ne pas être contraint par les choix de généricité maximale et de réutilisation logicielle retenus pour les systèmes d'accès
 - accepter les évolutions inhérentes aux projets scientifiques (évolutions des produits, mais aussi évolutions des besoins d'accès aux informations)

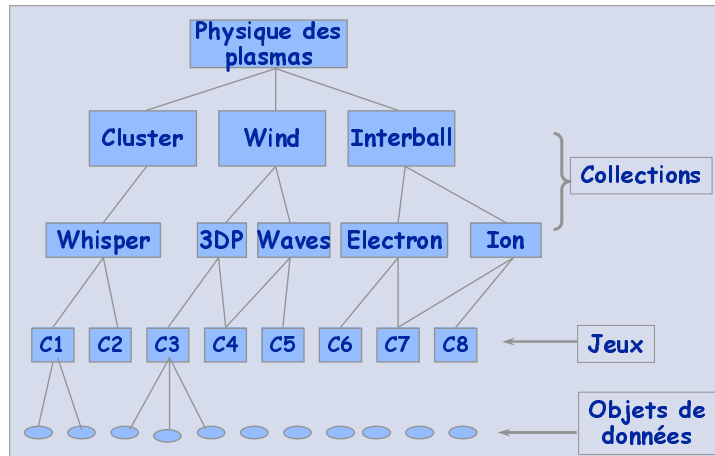
Définir avec le projet client son modèle de données

- structuration des informations qu'il souhaite archiver et mettre à disposition
 - » différentes catégories d'informations :
données expérimentales ou auxiliaires, produits graphiques, éléments documentaires, ...
 - » notions de : objets de stockage, objets (données, produits graphiques), jeux, collections, graphes, ...
 - » identification d'entités et de liens entre ces entités
- définition des attributs caractérisant chacune de ces entités (prise en compte des besoins d'accès et de gestion de l'archivage)
 - le dictionnaire des entités et des attributs de ces entités
 - ↔ la spécification des métadonnées pour ce projet

Solutions retenues



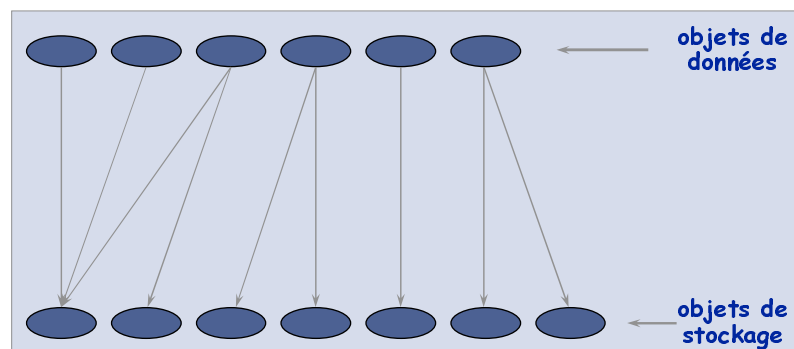
- Exemple de modèle de données :



Solutions retenues



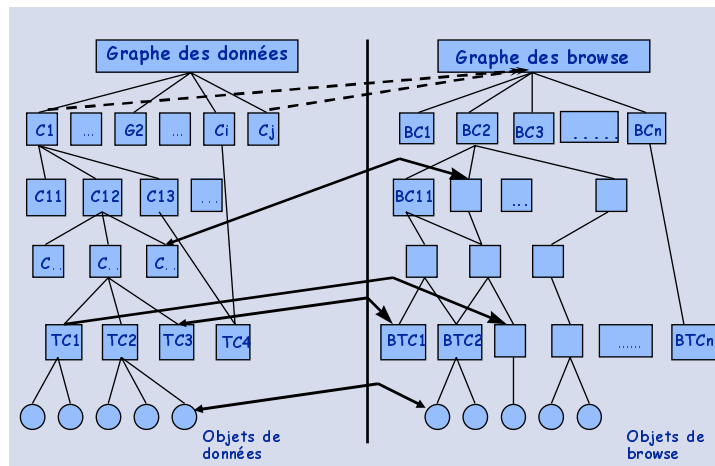
- Exemple de modèle de données (suite) :



Solutions retenues



- Exemple de modèle de données (suite) :



browse = produit graphique

Solutions retenues



• Les dictionnaires de métadonnées

- le principe : définir dans le dictionnaire tout ce que le projet client pourra livrer concernant son modèle de données : entités et attributs de ces entités, liens entre entités, entités composites, ...
- "définir" signifie :
 - identifier l'entité et identifier les éventuelles entités la composant (fil conducteur : la sémantique),
 - identifier les attributs la caractérisant, en préciser le type, la syntaxe autorisée (ex. : énuméré des valeurs possibles, nombres min et max de caractères pour les chaînes, squelette de saisie, ...)
 - définir la cardinalité d'occurrence de l'entité ou de l'attribut d'une entité (ex. : facultatif, plusieurs valeurs autorisées, ...)
 - définir les liens éventuels avec les objets numériques décrits (fichier de données, fichier graphique, document électronique, ...)

Solutions retenues



- **Ensuite toute livraison de métadonnées doit être conforme au dictionnaire qui lui est applicable ("contrat")**
 - préparation (souvent par différents fournisseurs) indépendante du système qui va les gérer
 - contrôles possibles dès la phase d'ingestion
 - rebouclages aisés avec un fournisseur si besoin
 - pas d'influence d'une erreur ou d'une omission sur le fonctionnement du système ou sur l'utilisation des informations récupérées par son intermédiaire
- origine des dictionnaires de (méta)données :
CCSDS (Comité Consultatif pour les Systèmes de Données Spatiales), travaux sur les DEDSL (Data Entity Dictionary Specification Language, CCSDS 647.1-B-1, <http://www.ccsds.org/CCSDS/documents/647x1b1.pdf>)

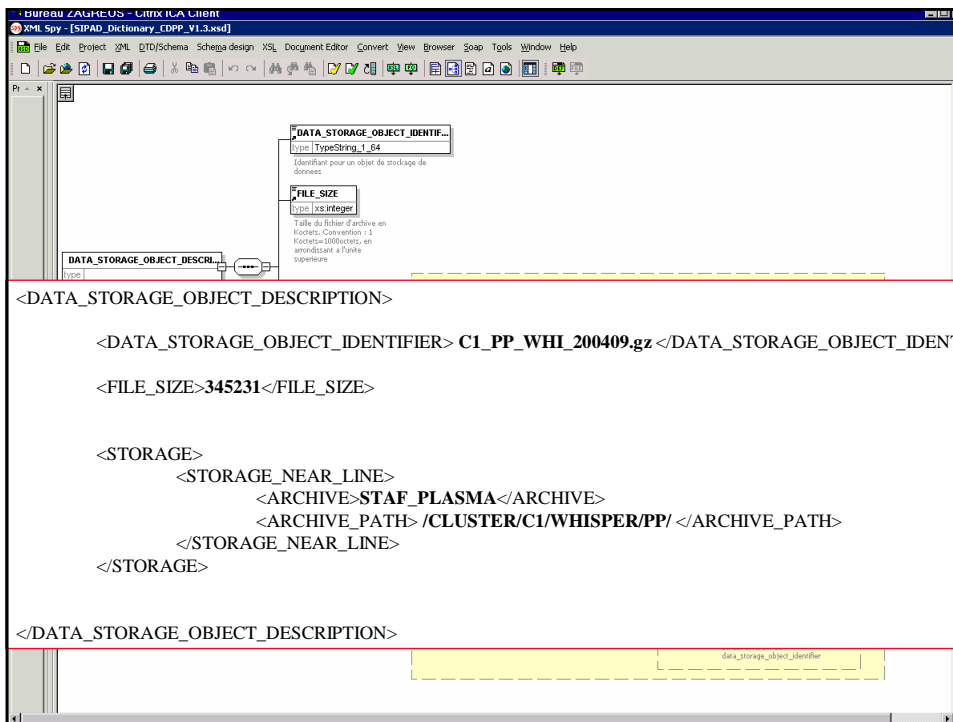
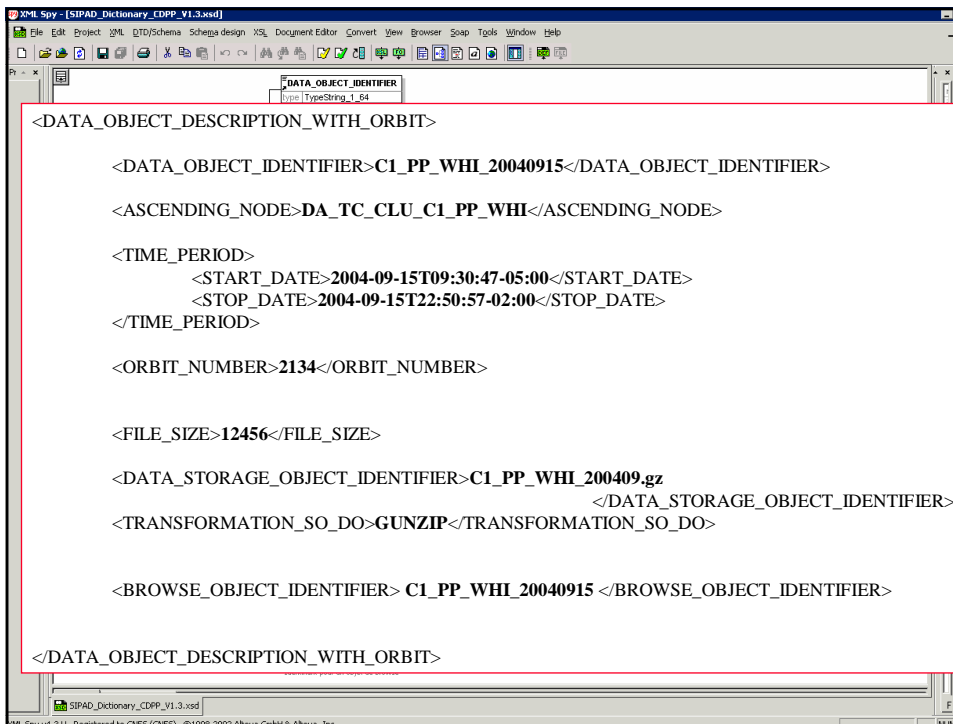
Solutions retenues

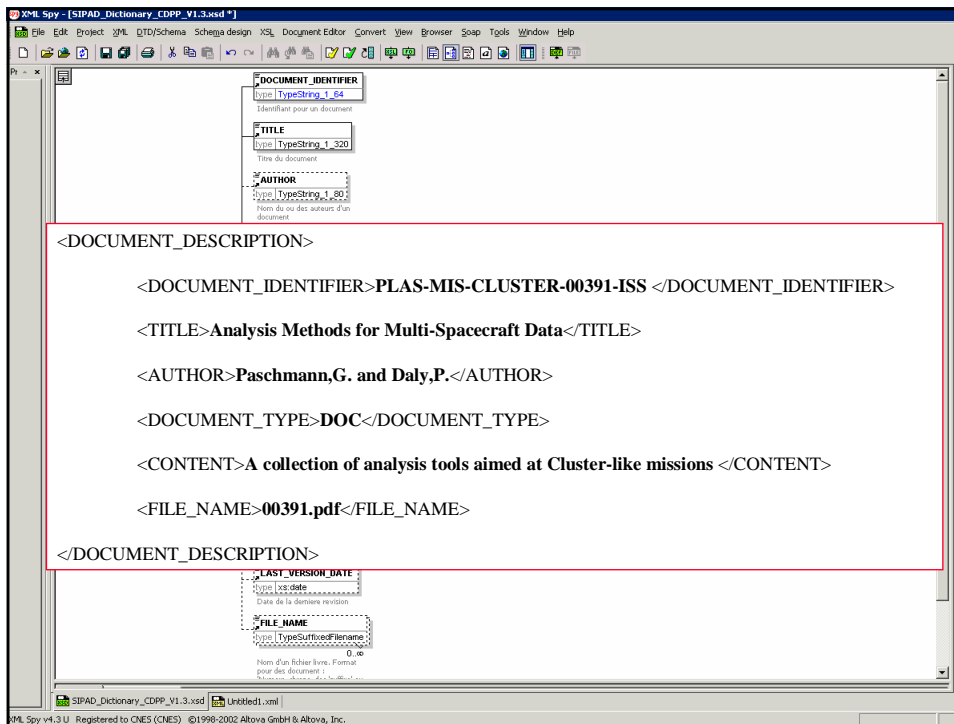


Les technologies ... :

- première application opérationnelle au CNES sur le SIPAD (3 instances de ce système) :
de 1999 à 2005
implémentation en langage de description PVL (Parameter Value Language)
exemple : FILE_SIZE=1234;
- deuxième application opérationnelle sur le SIPAD NG : de 2005 à ...
implémentation en langage de description XML (eXtensible Markup Language)
au delà de l'effet de mode (!), XML est bien adapté :
 - » "naturellement" structurant
 - » 1 dictionnaire = 1 schéma XML
 - » logiciels de "parsing" existants (et gratuits !)

exemple : <FILE_SIZE>1234</FILE_SIZE>





Cas d'utilisation opérationnelle



- **CDPP (Centre de Données de la Physique des Plasmas, <http://cdpp.cesr.fr>), un seul projet client d'une instance SIPAD, utilisation d'un dictionnaire, en exploitation depuis 1999**
- **SAM (Service d'Accès Multi missions), plusieurs projets clients d'une même instance SIPAD, un dictionnaire commun à tous, en exploitation depuis 2001**
- **SIPAD NG, plusieurs projets clients d'une même instance, chaque projet a "son" dictionnaire, gestion centralisée et maîtrisée au niveau de l'instance, mise en exploitation en 2005**
 - avantage de l'adaptation aux besoins et au vocabulaire de chaque client
- **en parallèle de ces applications au CNES, travaux de normalisation en cours au niveau CCSDS et ISO pour "diffuser" ces approches (modèles de données et dictionnaires) vers d'autres agences et organismes (spatiaux et non spatiaux)**

Préparation des fichiers XML à livrer :

- **pour les entités définies une seule fois (ex : les collections ou les jeux) :**
 - saisie via un éditeur XML ou un formulaire
 - recherche de solutions permettant la génération automatisée du formulaire à partir du schéma XML (dictionnaire) d'un projet
 - » solutions "logiciels libres" , attente de produits conformes à la norme récente Xform ...
 - » solutions commerciales (Microsoft, Adobe, ...)
 - » développement spécifique "maison"

- **pour les entités fréquemment livrées (ex : les objets) :**
 - production automatisée par logiciels, support aux projets pour leur masquer la syntaxe XML (bibliothèque d'utilitaires fournie)

- **actuellement "mapping" des informations livrées en XML vers une base de données relationnelle**

- **prise en compte de l'adaptabilité et des évolutions des dictionnaires : possible grâce au choix d'un "méta schéma" au niveau de la base de données relationnelle**

- **dans le futur, intérêt des bases de données XML ?**

Merci de votre attention ... avez-vous des questions ?!