



Les critères d'évaluation des formats de représentation de l'information : synthèse et usage

Claude Huc
Présentation Groupe PIN
25 janvier 2006



Synthèse : règles de base

- **Un format de données doit :**
 - être entièrement et explicitement spécifié et sa spécification doit être connue de l'entité en charge de la pérennisation
 - être apte à représenter la sémantique et la complexité de l'information à préserver
- **L'usage de formats normalisés est recommandé. On évitera l'usage d'éléments propriétaires au sein d'un format normalisé**
- **Lorsque le besoin de pouvoir modifier un document est identifié, le choix du format doit prendre en compte cette contrainte**



Règles complémentaires

- ***critères relatifs à la création des données***
 - Le choix du format doit prendre en compte la disponibilité et le coût des outils et des facilités de création des données
- ***critères relatifs à la prise en charge des données par service d'archive***
 - Il doit être possible de vérifier automatiquement qu'un fichier de données respecte les spécifications du format et respecte également les règles d'utilisation et des restrictions qui auront été définies pour la pérennité
 - La possibilité d'extraire automatiquement tout ou partie des métadonnées à partir des données constitue un avantage certain



Règles complémentaires

- ***critères relatifs à la préservation des données***
 - L'usage de formats inutilement volumineux sera évité
 - Un format simple est préférable à un format complexe
- ***critères relatifs à l'utilisation des données***
 - Les formats largement reconnus et utilisés seront privilégiés
 - Le choix du format doit prendre en compte la disponibilité et le coût des outils de transformation des formats et de présentation des données
 - Le choix du format doit prendre en compte la disponibilité et les potentialités de développements de services à valeur ajoutée



Synthèse

- *Règles de base incontournables*
- *Un compromis à trouver sur les règles complémentaires*
- *Identifier très précisément les restrictions d'usage à imposer ou recommander*



Contexte normatif au CNES

- *Référentiel Normatif du CNES*
 - **Domaine « Ingénierie des données » :**
 - **3 niveaux de documents :**
 - Niveau 1 : Présentation du domaine, identification des besoins
 - Niveau 2 : Les standards d'exigence : quoi faire ?
 - Modèle OAIS
 - Règles applicables aux producteurs (avec des indications sur comment faire pour les formats les plus courants)
 - Règles d'évaluation des formats
 - Méthodologie d'interface Producteur-Archive
 -
 - Les standards de mise en œuvre : comment faire
 - Langage de description de données