

***Aristote : Groupe PIN, 15/01/08***

Gestion des formats de fichiers :  
une question clé pour l'archivage  
Un exemple de stratégie dans ce  
domaine : le pilote Pil@e

Jean-Louis Pascon

### ***Gestion des formats : Nécessité***

- Liée au choix de la stratégie d'archivage (ni émulation, ni conservation de l'environnement)
- Une obligation due au grand nombre de formats (+ de 200 an matière d'images fixes) et à leur grande variabilité (presque une version nouvelle tous les ans pour DWG)
- Maîtrise de la volumétrie (mon appareil photographique peut produire jusqu'à 21 Mo toutes les 10 secondes)

## ***Définition d'une politique***

- Choix des formats en entrée
  - Utilisé largement au sein des services
  - Normes ou standard
- Choix des formats des formats cibles
  - TRES utilisé largement aux seins des services mais aussi dans le domaine considéré (exemple JPEG 2000 pour l'image fixe)
  - Normes ou standard stables
  - Plusieurs fournisseurs de visualisateur

## ***Tableau des formats***

- Formats entrée et version du format
  - Référence (norme, standard, etc.)
- Formats en cible et version du format
  - Référence (norme, standard, etc.)
- Outils de test
  - Fournisseurs, nom, version, prix
- Outils de conversion
  - Fournisseurs, nom, version, prix
- Test des systèmes

## **Les formats en entrée**

- ◆ **Images fixes**
  - ◆ PNG \*
  - ◆ GIF
  - ◆ JPEG 2000 (jp2) \*
  - ◆ JPEG (jfif, jpg) \*
  - ◆ TIFF \*
  - ◆ BMP
  - ◆ PCX
- ◆ **Images animées et sons**
  - ◆ MPEG-1 Layer III (MP3) \*
  - ◆ WAV \*
  - ◆ OGG VORBIS \*
  - ◆ MPEG-2 \*
  - ◆ MPEG-4 \*
- ◆ **Formats structurés/Non structurés**
  - ◆ HTML
  - ◆ XHTML \*
  - ◆ OpenDocument \*
  - ◆ MSoffice (DOC/XLS/PPT)
  - ◆ PDF/A \*
  - ◆ PDF \*
  - ◆ RTF
  - ◆ CSV \*
  - ◆ TXT \*

## **Les formats en entrée**

- ◆ **Formats composites**
  - ◆ ZIP \*
  - ◆ WinRAR \*
  - ◆ SIT \*
  - ◆ E-mail \*
- ◆ **Documents structurés**
  - ◆ XML \*
  - ◆ XSD \*
  - ◆ XML validé par XSD \*
  - ◆ SGBD
- ◆ **Données graphiques**
  - ◆ SVG \*
  - ◆ PDF/E \*
  - ◆ OpenDWG \*\*
  - ◆ DWG
  - ◆ DXF
  - ◆ CGM \*\*
  - ◆ STEP \*\*

## ***Progiciels de conversion***

- ◆ Disponibles sous LINUX
- ◆ Doit exister en mode commande
- ◆ Si possible disponible sous Windows (pour effectuer des tests préalables)
- ◆ Prix raisonnable (quelques dizaines d'€ en vue de généralisation)
- ◆ Support technique disponible
- ◆ Temps de tests ou de conversion compatible avec la vitesse globale de la chaîne d'archivage

## ***Les difficultés***

- ◆ Plate-forme homogène (LINUX) pas toujours de logiciels disponibles pour cet OS
  - ◆ Conséquences des coûts qu'il faut maîtriser (souvent les progiciels licence à l'utilisation)
- ◆ La gestion des tests : qu'est ce qu'un fichier en anomalie (exemple MPEG-4) ?
- ◆ L'identification des formats sur la signature : peu utile (exemple DROID) : on peut les tromper facilement

## ***Les difficultés***

- ◆ L'intégration dans les chaînes de traitement (que faire des codes d'anomalie tous différents suivant les progiciels ?)
- ◆ La taille de certains fichiers (exemple PDF/A sur PDF 1.4)
- ◆ La gestion de la qualité des conversions : comment s'assurer que tous les éléments d'un fichier sont bien transformés ?
- ◆ Les formats composites ZIP/Messages : régressivité combien de niveaux ?

## ***Des problèmes non résolus***

- ◆ Les coûts (filtrage, conversion, etc.)
- ◆ Conversion au fil de l'eau ou opérations commandos ?
- ◆ Implication des services informatiques lors de la conception des applications
- ◆ Pourquoi autant de formats ?
- ◆ La complexification des formats (BMP = 10 pages / JPEG 2000 = des milliers de pages)

**Conclusions**

**Questions ?**