

## Proposition

### Projet d'étude technique sur les formats de représentation de l'information

Présentation du groupe PIN  
24 septembre 2002

*Claude HUC*

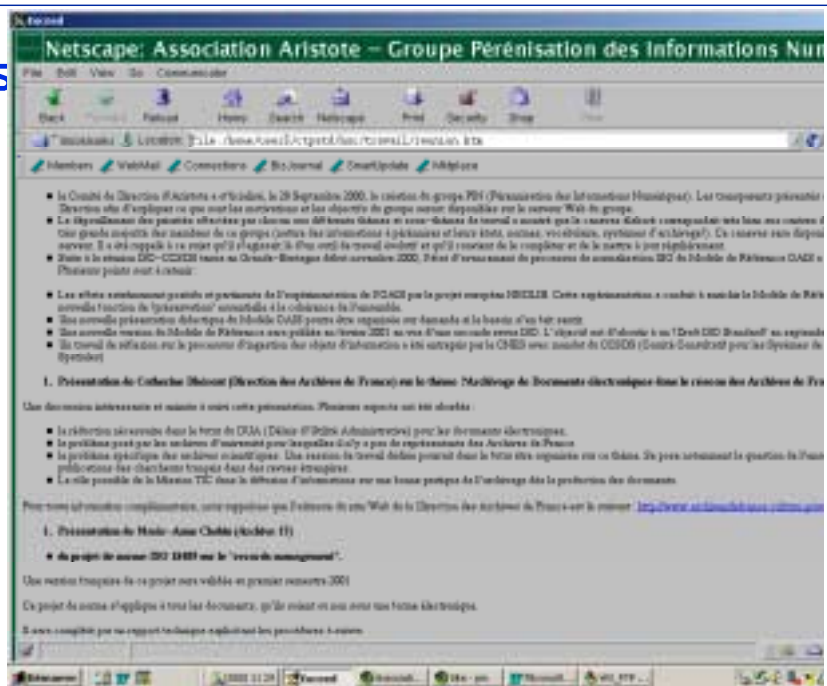
## Un exemple

➤ Le format HTML

➤ Dans le guide de la MTIC, le format HTML

- est considéré comme un format possible pour la conservation sans que ce soit le plus recommandé
- sans restriction précisément identifiée

- Pourtant



La pérennisation de l'information numérique - 24-09-2002

Claude HUC - CNES

- le Comité de Direction d'Aristote a officialisé, le 28 Septembre 2000, la création du groupe de travail de la Direction afin d'expliquer ce que sont les motivations et les objectifs du groupe serveur
- Le dépouillement des priorités affectées par chacun aux différents thèmes et sous-thèmes a été réalisé à la très grande majorité des membres de ce groupe (nature des informations à pérenniser, modèle de référence, serveur. Il a été rappelé à ce sujet qu'il s'agissait là d'un outil de travail évolutif et collaboratif
- Suite à la réunion ISO-CCSDS tenue en Grande-Bretagne début novembre 2000, l'établissement d'un modèle de référence OAIS. Plusieurs points sont à retenir :
  - Les effets extrêmement positifs et pertinents de l'expérimentation de l'OAIS par le groupe de travail de la Direction
  - Une nouvelle fonction de préservation essentielle à la cohérence de l'ensemble.
  - Une nouvelle présentation didactique du Modèle OAIS pourra être organisée sur demande
  - Une nouvelle version du Modèle de Référence sera publiée en février 2001 en vue d'être utilisée par le groupe de travail de la Direction
  - Un travail de réflexion sur le processus d'ingestion des objets d'information a été effectué

La pérennisation de l'information numérique - 24-09-2002

Claude HUC - CNES

## Origine de ce fichier ?

- Fichier Word 97
- Enregistré au format HTML
- remis en forme sous l'éditeur HTML Dreamweaver

➤ Début du fichier :

<HTML>

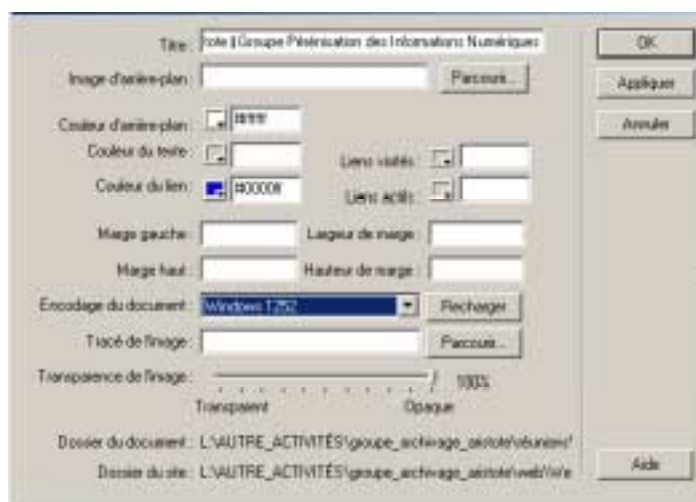
<HEAD>

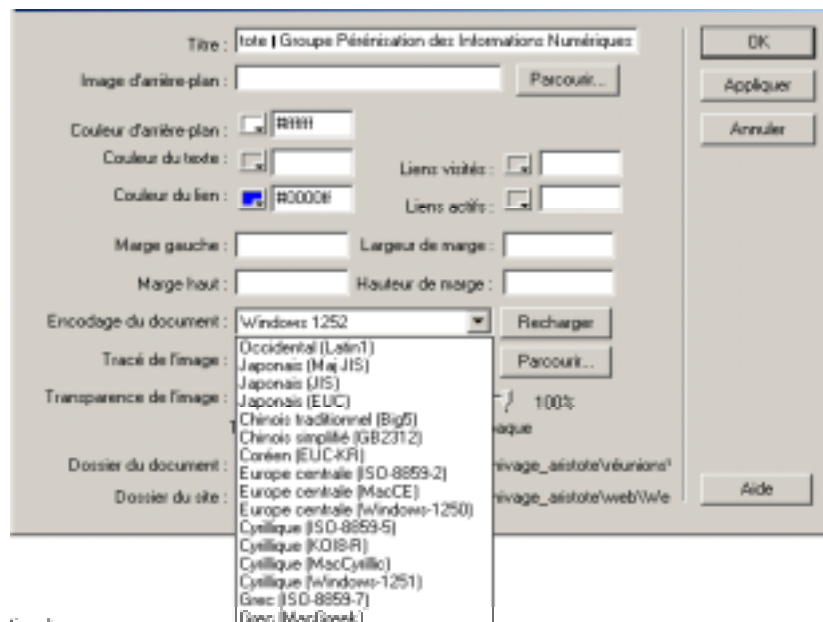
<META HTTP-EQUIV="Content-Type" CONTENT="text/html;  
charset=windows-1252">

<TITLE>Association Aristote – Groupe Pérennisation des Informations  
Numériques</TITLE>

## Examen sous Dreamweaver

- Menu Modifier/ propriété de la page





- le Comité de Direction d'Aristote a officialisé, le 28 Septembre 2000, la création du groupe Direction afin d'expliquer ce que sont les motivations et les objectifs du groupe seront di
- Le dépouillement des priorités affectées par chacun aux différents thèmes et sous-thèmes a été fait par une très grande majorité des membres de ce groupe (nature des informations à pérenniser et serveur. Il a été rappelé à ce sujet qu'il s'agissait là d'un outil de travail évolutif et qu'il c
- Suite à la réunion ISO-CCSDS tenue en Grande-Bretagne début novembre 2000, l'état d'avancement de l'outil est le suivant :  
Plusieurs points sont à retenir :
- Les effets extrêmement positifs et pertinents de l'expérimentation de l'OAIS par le projet ont été notés. La nouvelle fonction de 'préservation' essentielle à la cohérence de l'ensemble.
- Une nouvelle présentation didactique du Modèle OAIS pourra être organisée sur demande.
- Une nouvelle version du Modèle de Référence sera publiée en février 2001 en vue d'une mise à jour.
- Un travail de réflexion sur le processus d'ingestion des objets d'information a été entrepris.

## Examen du fichier ' source '

➤ Nouvelle entête

```
<HTML>
```

```
<HEAD>
```

```
<META HTTP-EQUIV="Content-Type" CONTENT="text/html;  
  charset=iso-8859-1">
```

```
<META NAME="Generator" CONTENT="Microsoft Word 97">
```

```
<TITLE>Association Aristote &#150; Groupe Pérennisation des  
  Informations Numériques</TITLE>
```

## Conclusion

- Les recommandations sur les formats utilisables pour la conservation des documents
  - sont souvent insuffisantes
  - les précautions à prendre, les restrictions, les pièges ne sont pas décrits
  - les moyens de valider la bonne application des règles ne sont pas identifiés
- D'ou l'idée de mettre en chantier la constitution d'une base de connaissance pratique sur les formats disponibles au filtre des principes de la pérennisation

## Démarche

- Réutilisation des travaux synthétiques existants : guides, recommandations
- réutilisation des travaux réalisés par tel ou telle institution sur un format particulier
- recherche de l'existant sur Internet
- recours aux experts du domaine

**En vue d'une mise à disposition de l'information sur le serveur PIN**

## Domaines d'étude

- Terminologie :
  - qu'est ce qu'on entend par format standard, fermé, ouvert, propriétaire, publié,
- l'encodage de l'information de base : de l'ISO 646 (ASCII) à UNICODE et UTF, mais aussi le codage des nombres, des couleurs, du son, de la vidéo...
- les formats généralistes pour la représentation de l'information

## Constituer des fiches de référence (1)

➤ Pour chaque format, on élaborera une fiche de référence dont le contenu se conformera au plan type suivant :

- Acronyme :  
Nom complet du format :  
Numéro de version :
- Domaine d'utilisation : image/document textuel/....
- Auteur de l'analyse :  
Date :
- Analyse mise à jour par :  
date :
- Archivage : format conseillé, possible, déconseillé
- Diffusion : : format conseillé, possible, déconseillé

## Constituer des fiches de référence (2)

➤ Références documentaires

- du document descriptif du format
- des documents d'analyse existants
- des sites Internet utiles sur le format lui-même et sur son utilisation (exemples d'application)

➤ Résumé descriptif : description succincte, caractéristiques générales, compressions avec ou sans perte, performances, volume, particularité à connaître, limitations...

## Constituer des fiches de référence (3)

- Logiciels (mais aussi système d'exploitation, compilateur...) utilisables pour la création : logiciels libres, logiciels commerciaux.
- Logiciels (système d'exploitation, compilateur...) utilisables pour la lecture
  
- Moyens de validation
  
- Exemple de fichiers
  
- Autres commentaires

## Autres questions

- Comment, à partir des logiciels couramment utilisés, produire des fichiers respectant les règles de conservation
  - Que peut-on faire en pratique avec un fichier Excel ?
  
  - Quid de Word ou de Powerpoint ?
  
  - Etc.



## Questions à PIN

- Est-ce une tâche utile au groupe ?
- Est-ce une tâche faisable ?
- Quelle répartition du travail ?
- Comment procéder ?