

Centre de Documentation et d'Exploitation de l'Information

# XediX : une plateforme XML pour la pérennisation des données

**Didier Courtaud & Pierre Brochard**

**CEA / DAM Ile-de-France / CDEI**

**Didier.Courtaud@cea.fr**  
**Pierre.Brochard@cea.fr**

1

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Ile-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Plan

- **Brève présentation du CEA**
- **Problématique**
- **Projet ATOS**
  - Objectifs
  - Choix stratégiques
  - Solution retenue
- **Conclusion**

2

Réunion du Groupe de travail PIN d'Aristote

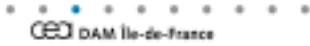
CEI DAM Ile-de-France  
18 Septembre 2001

## Le CEA en quelques chiffres

Centre de Documentation et d'Exploitation de l'Information

- Etablissement public de recherche ( EPIC )
- 16139 personnes (fin 2000) dont  
~ 46% d 'ingénieurs-chercheurs
- Budget : ~19 Milliards de francs  
soit 2,9 Milliards d 'euros
- 7500 brevets en portefeuille

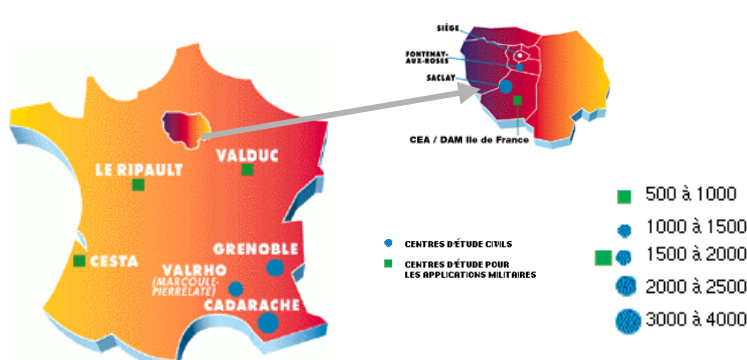
3



Réunion du Groupe de travail PIN d 'Aristote 18 Septembre 2001

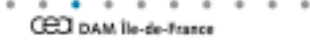
## Le CEA : implantations géographiques

Centre de Documentation et d'Exploitation de l'Information



Symbol	Personnel Count	Center Type
Green Square	500 à 1000	Centres d'étude pour les applications militaires
Blue Dot	1000 à 1500	Centres d'étude civils
Green Square	1500 à 2000	Centres d'étude pour les applications militaires
Blue Dot	2000 à 2500	Centres d'étude civils
Blue Dot	3000 à 4000	Centres d'étude civils

4



Réunion du Groupe de travail PIN d 'Aristote 18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Problématique

### Problème à résoudre

- **Engranger rapidement les connaissances des experts**
  - Arrêt des activités
  - Départ en retraite des experts
- **Assurer la pérennité des systèmes**
  - Normes internationales
  - Standards
- **Utiliser les nouvelles technologies multimédia**
  - Convivialité
  - Habitudes de travail des futurs utilisateurs
  - Indispensables à la sauvegarde de certaines connaissances

5

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Problématique ( II )

- **Connaissance écrite ( imprimable )**
  - Textuelle : rapports, notes techniques, fax, courrier électronique
  - Visuelle : images de synthèse, photos argentiques, photos numériques
- **Connaissance orale**
  - Auditive : comptes-rendus, discussions, *brainstormings*
- **Connaissance non exprimée**
  - Savoir-faire : technique, méthode, conditions expérimentales

**Nous devons en tenir compte pour bâtir des applications informatiques chargées de maintenir ces connaissances**

6

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Recueil de connaissances

- **Recueil des connaissances formalisées et non formalisées**
  - textes, documentations techniques
  - savoir-faire, processus de fabrication
- **Méthode MKSM ( CEA / DIST )**
- **Interviews**
  - de responsables hiérarchiques
  - de techniciens

7

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Recueil de connaissances ( II )

- **Production des interviews**
  - Formalisation textuelle
    - décomposition fonctionnelle MKSM
  - Photos
  - Croquis
  - Planches CAO ( format A0 )
  - Interviews originels sous forme audio
  - Films vidéo
- **Documentation scientifique et technique associée**
  - Rapports
  - Notices techniques

8

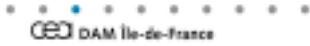
Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Objectifs du projet

- **Construire un système ouvert, évolutif et pérenne dans le temps**
  - Système modulaire : les briques de bases sont remplaçables et évoluent indépendamment
  - Extensions possibles :
    - soit par la société de service (Euroclid -> Valoris) sur cahier des charges CEA
    - soit par le CEA car nous avons les sources et la maîtrise technique des développements
  - Choix de la norme internationale SGML -> XML
- **Montrer l'intérêt de la documentation structurée, pour améliorer la qualité de la recherche**
  - Séparation des structures logique et physique du document
  - Lors d'une recherche on peut tenir compte de la structure
  - Exemple : je cherche le mot énergie uniquement dans les titres de documents ou dans les mots clés ou dans un titre de chapitre, ...




Réunion du Groupe de travail PIN d'Aristote 18 Septembre 2001

9

Centre de Documentation et d'Exploitation de l'Information

## Choix stratégiques

- **Format des documents : norme SGML ( puis XML )**
  - Une DTD DOCCEA adaptée mais générique a été définie pour ce projet
- **Editeur SGML**
  - Saisie et édition des documents SGML / XML
  - Configuré pour la DTD DOCCEA
- **Base de données**
  - Permet de stocker et gérer des documents au format XML après création du filtre adéquat
  - En particulier, gestion possible d'éléments multimédia
  - Nécessité d'une conception objet
    - Décomposition arborescente du document
    - Héritage
  - Tenue aux grands volumes



Réunion du Groupe de travail PIN d'Aristote 18 Septembre 2001

10

Centre de Documentation et d'Exploitation de l'Information

## Choix stratégiques (II)

- **Moteur de recherche**
  - Recherche statistique
  - Recherche par mots clés
  - Recherche contextuelle utilisant la connaissance de la structure du document XML
  - Recherche par concept *arborescent*
- **Outils Web**
  - Standards de la famille XML : SMIL , SVG, MathML
  - Serveur web
  - Butineurs WWW pour l'utilisateur final
  - Utilisation de Java et Javascript
    - Possibilité de portage sur des *clients léger*

11

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Projet ATOS

### Accès Thématique aux Ouvrages Scientifiques

**A terme :**

- **75000 documents**
- **Documents**
  - anciens : Rétroconversion XML
  - actuels : XMLisation et/ou import direct
- **3 millions pages**
- **Une ou plusieurs bases de l'ordre de 200 Goctets**

12

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Projet ATOS ( II )

### Choix retenus

- **Base de données**
  - Première implémentation sous O2, puis sous Clio
  - Tenue aux grands volumes
  - Performances générales
- **Moteur de recherche**
  - Search '97 de Verity
  - Possibilité d'indexer selon la structure
  - Arborescence de concepts

13

Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Projet ATOS ( III )

- **Interface Homme machine**
  - Navigateur Web
  - Affichage HTML créé dynamiquement
  - Affichage XML natif dès que possible au niveau des navigateurs
- **Intégration**
  - Réalisé par Euroclid sur cahier des charges CEA
  - Générique
  - Paramétrable
  - Multi DTDs

14

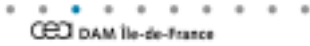
Réunion du Groupe de travail PIN d'Aristote

CEI DAM Île-de-France  
18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Projet ATOS ( IV )

- **En amont, mise en place d'une chaîne documentaire pour :**
  - Faciliter l'intégration des documents futurs
  - Autoriser la capitalisation de connaissances *en ligne*
- **S'appuie sur :**
  - l'utilisation d'un éditeur SGML/XML
  - l'adoption d'une charte documentaire d'entreprise
    - consensuelle
    - pragmatique
- **Possibilité d'intégrations d'abstraction supérieure**
  - Gestion de bibliothèques
  - Workflow



15


Réunion du Groupe de travail PIN d'Aristote

18 Septembre 2001

Centre de Documentation et d'Exploitation de l'Information

## Conclusion

- **Produit résultant performant**
  - Pertinence de la réponse ( grâce aux requêtes structurées )
  - Bonne tenue aux volumes
  - Grande modularité
  - Polymorphe
- **Evolution**
  - Facile grâce à la modularité
    - Base de données : Clio, Oracle, MySQL, ..
    - Moteur de recherche : linguistique, thématique ( images , vidéo )
  - Plate-forme d'intégration
    - Applications *métier*



16

Réunion du Groupe de travail PIN d'Aristote

18 Septembre 2001