



CENTRE NATIONAL D'ÉTUDES SPATIALES

Intégration du traitement automatique des courriels dans la chaîne de traitement des archives



Direction du Système d'Information
Sous-Direction « **S**upport et **D**éveloppement »
Service Gestion de l'Information et de la connaissance

SOMMAIRE

- **Contexte**
- ✿ **Intégration du traitement des courriels**
- ✿ **Evolutions : Catégorisation sémantique (personnel/privé)**

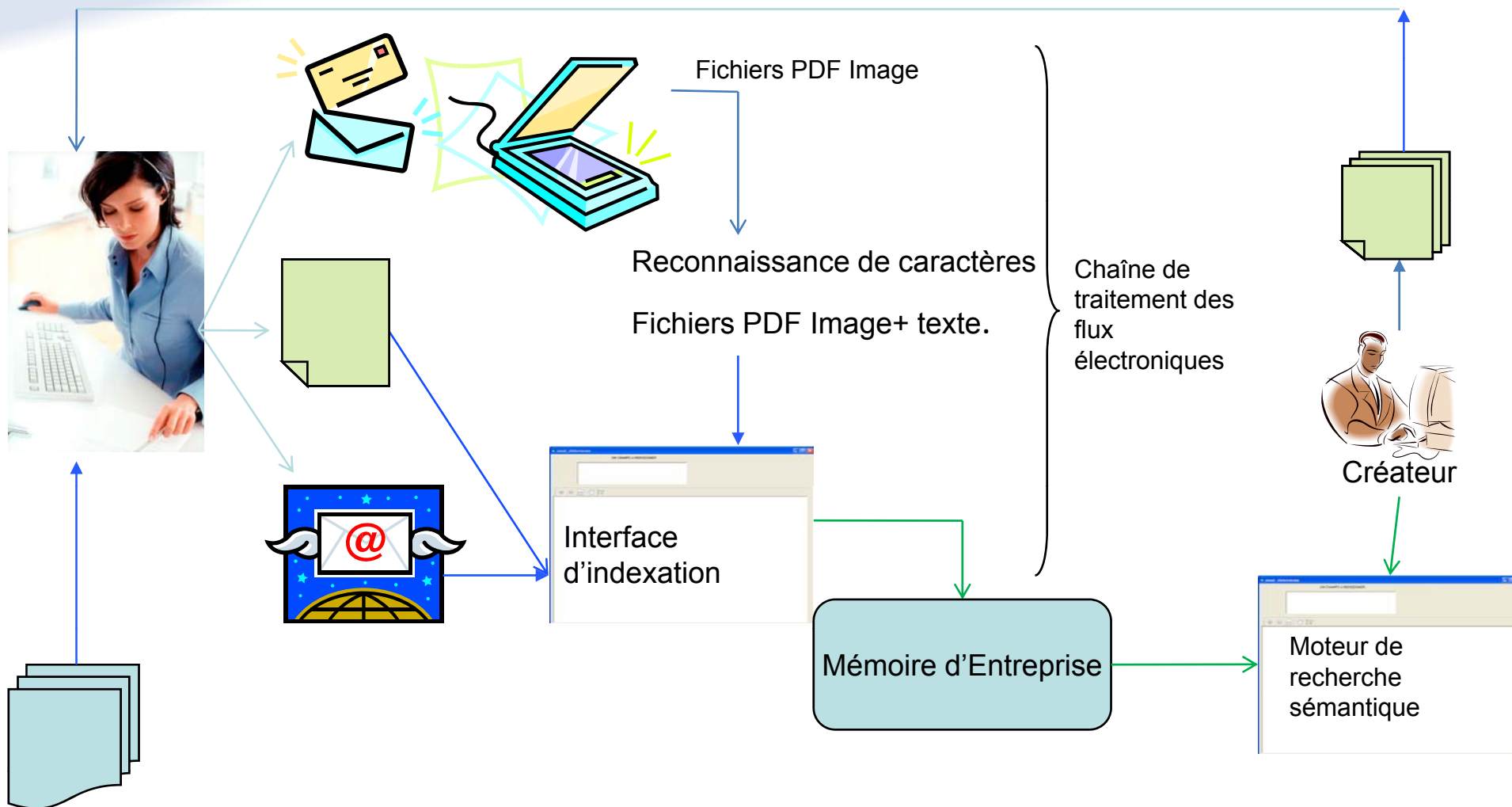
- CNES-EPIC- Archives donc les courriels, ADHG OK
- Dématérialisation des archives
 - **Numérisation océrisation papier, tri, indexation, formatage, inclusion des métadonnées, versements.**
 - **Importance du tri des courriels : professionnels, privés, redondants, sans intérêts...**
- Valorisation via la ME
 - **Mise à disposition de l'ensemble des agents du CNES via un moteur de recherche sémantique qui s'appuie sur une Base de Connaissances CNES**

- **Pas de politique générale de collecte au sein du CNES activité trop « jeune »**

- **Actions ponctuelles :**
 - ◆ Tests sur des lots de mèls personnels par exemple suite à un départ en retraite (tri) et avec l'autorisation de la personne
 - ◆ Collecte de mèls spécifiques (Bulletins d'Ambassades...)

- **Collecte à la source**
 - ◆ Documents : intervention opérateur (numérisation, OCR pour le papier, puis indexation...)
 - ◆ Mèls entrants/sortants : une appli sur le PC fait la collecte puis le transfert vers un serveur de traitement

Collecte (préconisation)



■ TRI

- ◆ Personnel/privé ? Aujourd'hui nous ne traitons que des courriels spécifiques : bulletins d'ambassades...
- ◆ Redondances : traitement des doublons sur date/taille/contenu ...
- ◆ Intérêt du courriel ? Fonction de la taille ?

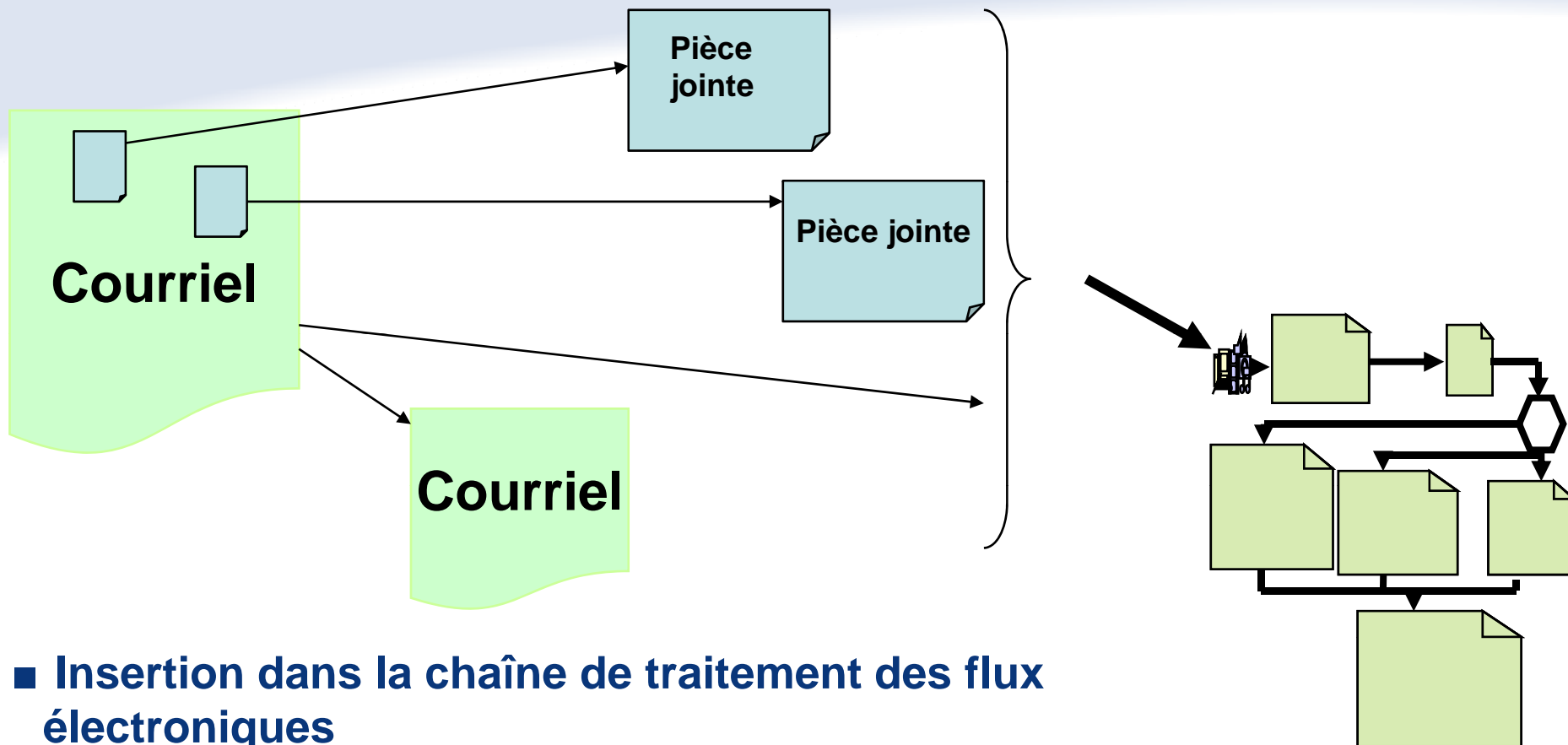
■ Formatage

- ◆ Extraction des pièces jointes, décompression automatique, formatage PDF (texte sous l'image)

■ Indexation automatique

- ◆ Objet (titre)
- ◆ Date
- ◆ Auteur (dans l'adresse pour le CNES)
- ◆ Conservation du source complet (avec fichiers joints...)

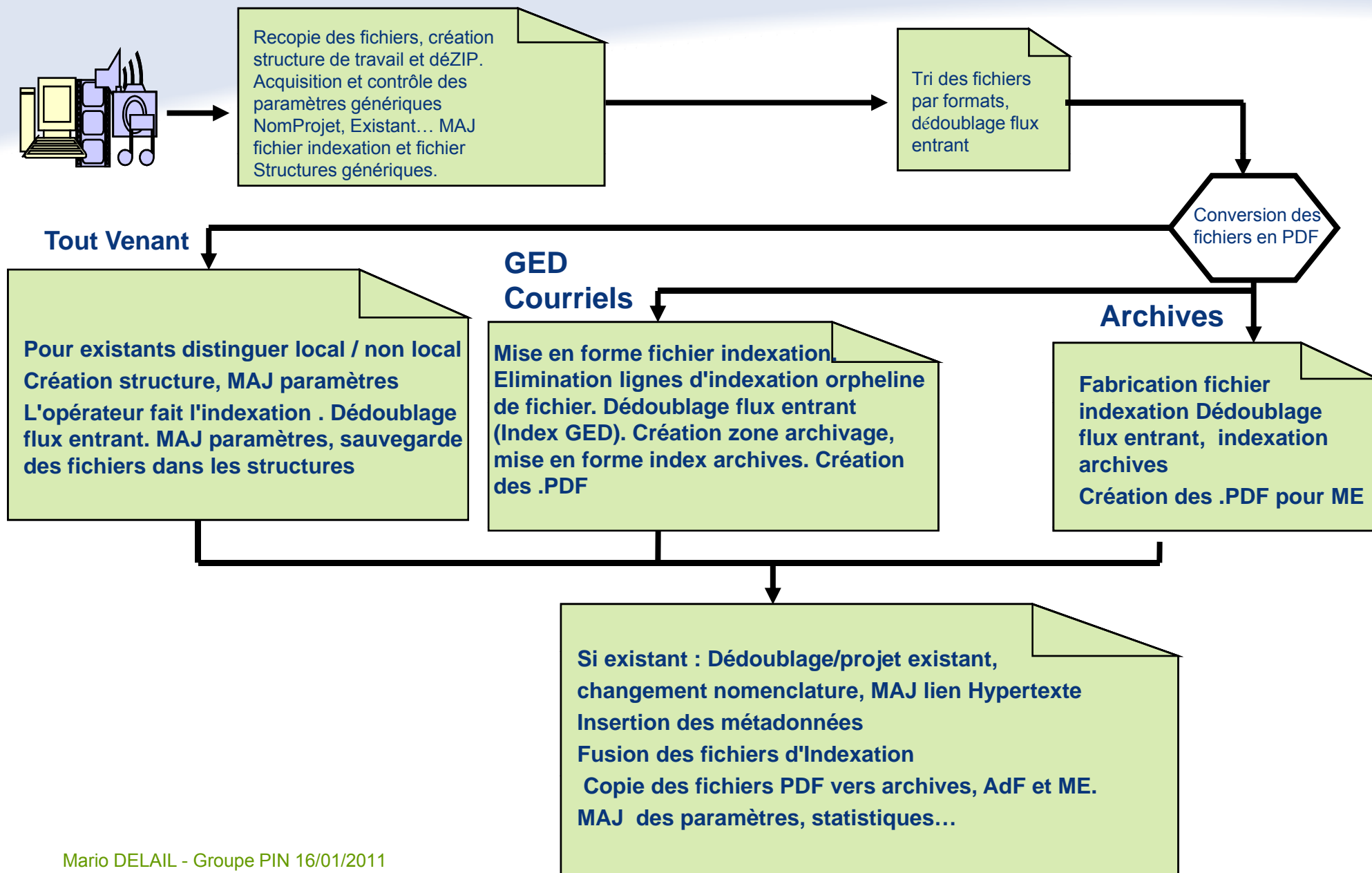
- **Sortir les pièces jointes et les traiter comme des documents unitaires**
- **Conserver le message source avec les fichiers joints dans la structure de travail (Source) avec un lien dans le fichier d'indexation**
- **Traiter la partie message comme un document si taille minimum**



■ Insertion dans la chaîne de traitement des flux électroniques

- ◆ Mise en forme (séparation fichiers joints, attributs, conversions, formatage, dédoubleage flux entrant...), indexation (automatique/aidée) et versements (archives CNES, AdF, ME)

Le traitement des flux électroniques



Collecte, dédoubleage flux entrant, création structure logique de travail, préservation du contexte des sources (minimum chemin)

Flux à traiter dans structure de travail

Séparation fichiers joints, normalisation des attributs des fichiers, décompression des .Zip, conversion en PDF texte caché

Fichiers PDF-Texte caché

**Indexation des fichiers : automatique ou aidée.
Métadonnées explicite minimum : titre, auteur, date**

Base de données

Dédoublage par rapport à l'existant, Insertion des métadonnées dans le PDF, versements en archives CNES, aux Archives de France, en Mémoire d'Entreprise

Partie message

- **Distinction messages personnels/messages professionnels**
 - ◆ **Élimination si objet contient « personnel », « perso », « privé »**
 - ◆ **Tri et classement automatique en utilisant la BC CNES (en test)**

- **Mèl (taille très réduite) + fichiers joints : élimination du message pour le traitement, conservation du mèl complet en source**

- **Mèls RE (RE-RE...), ne conserver que la date la plus récente (possibilité de perte d'information)**

Pièces jointes

- **Dédoublage flux entrant, dédoublage par rapport à l'existant**

- **Contextualisation, garder le lien avec le message**
 - **Problèmes intérêt du texte msg ? → non classé ?**
 - **Document CNES ? Auteur CNES ? Référence CNES ? (éliminer les autres ?)**
 - **Fichiers particuliers : compressés, cryptés, liste de diffusion restreinte...**

Utilisation des outils de l'Atelier d'Ingénierie de la Connaissance (ATIC) et de la Base de Connaissances CNES pour :

- **Trier les mès : professionnels/privés**
 - ◆ Analyse du contenu en fonction des concepts et des réseaux sémantiques de la BC

- **Classer les mès : par activités, par projets [par mission]**
 - ◆ Intérêt des non classés ?



CENTRE NATIONAL D'ÉTUDES SPATIALES

MERCI POUR VOTRE ATTENTION