

# Alimentation de l'annuaire - problématique des sources multiples

Benoit Branciard

Université Paris 1 Panthéon - Sorbonne

tutoJRES 11

Annuaire : mise en oeuvre de SupAnn 2008

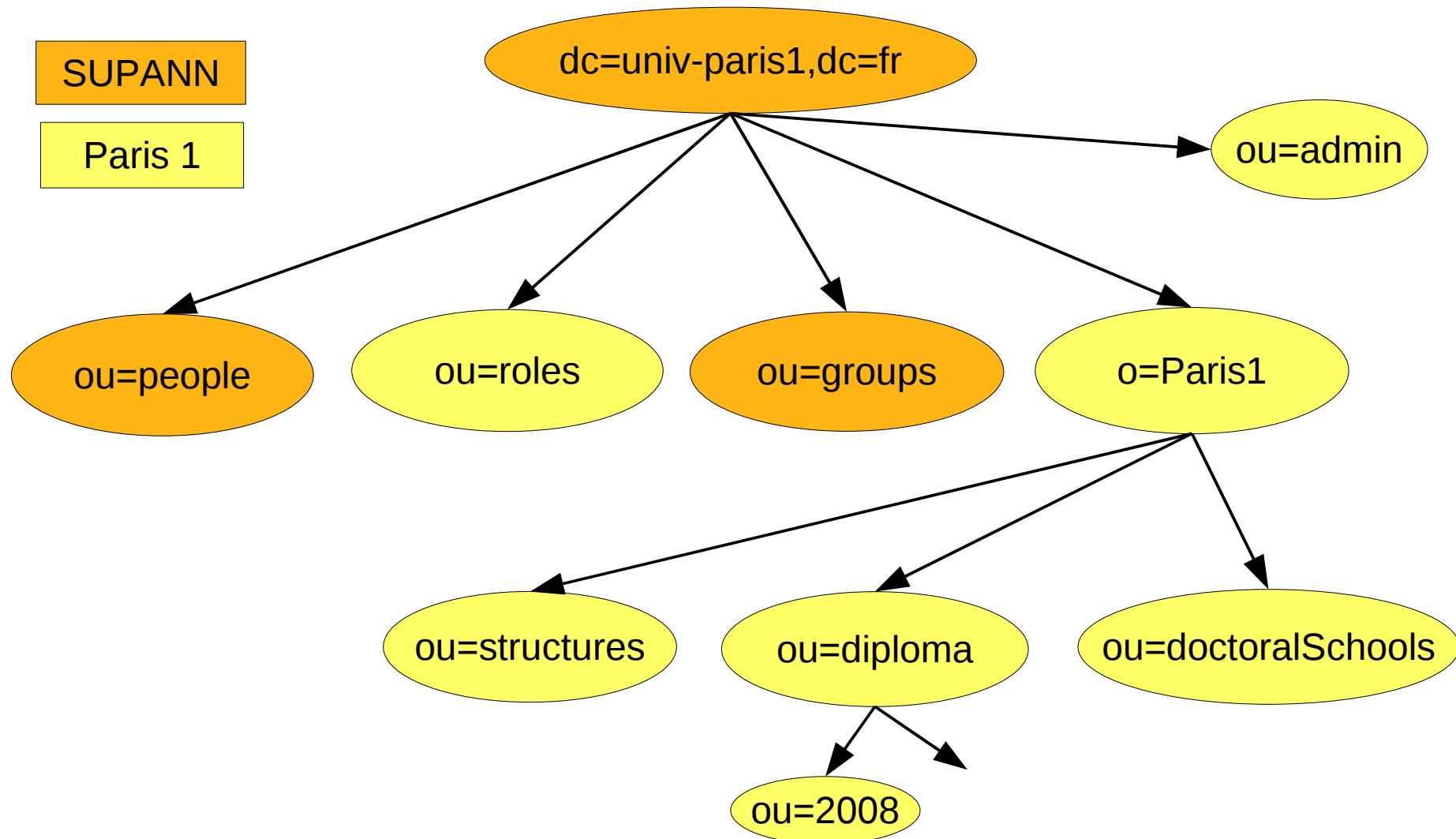
1 avril 2009

# Contexte et Historique

- Paris 1: 45 000 étudiants, 3 000 personnels
- Déploiement de SUPANN en 2004
  - OpenLDAP, un maître et deux réplicats.
  - SUPANN v1 avec très peu d'extensions « maison »: branche « structures », « diploma », « doctoralSchools »
  - Enseignement: groupes LDAP pour les matières
- Evolution du schéma selon les besoins:
  - Ajout d'une branche « ou=roles » pour les comptes email fonctionnels.
  - Attributs « maison »: nom patronymique, date de naissance, adresses mail alternatives
  - intégration progressive de Supann 2008

# Schéma LDAP Paris 1

(avant Supann 2008)



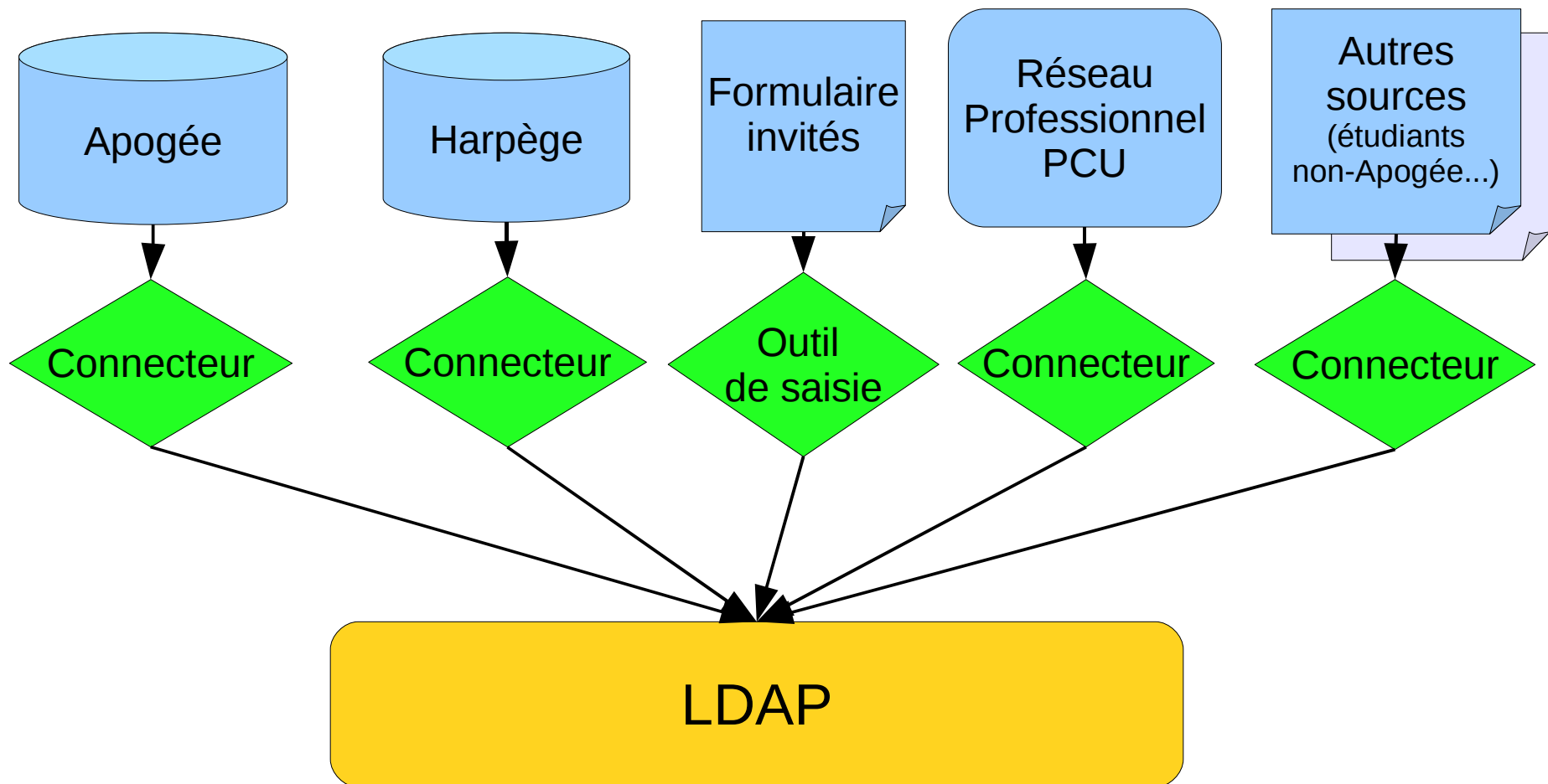
# Usages

- **Identification, authentification:**
  - pam\_ldap: messagerie, stockage
  - CAS: applications Web et ENT
  - LDAP: webmail, diverses applications non CASsifiées
  - RADIUS: Eduroam, 802.1x
- **Routage de messagerie**
  - Aliases et forwards LDAP
- **Gestion de groupes, populations et privilèges**
  - Sympa: listes de diffusion sur filtres LDAP
  - Ciblage de population au niveau applicatif
  - Shibboleth: exportation d'attributs
- **Annuaire public, pages web personnelles**

# Alimentation

- Ensemble de « connecteurs » maison opérant de façon indépendante:
  - Apogée: scripts Perl, modules temps réel et asynchrone
  - Harpège: appli « cuisine » en Java, asynchrone
  - Saisie directe pour les « invités »: scripts Perl
  - Réseau professionnel PCU: script Perl
  - Sources « clandestines »: étudiants de diplômes cohabilités ou de matières « sous-traitées »
  - ...

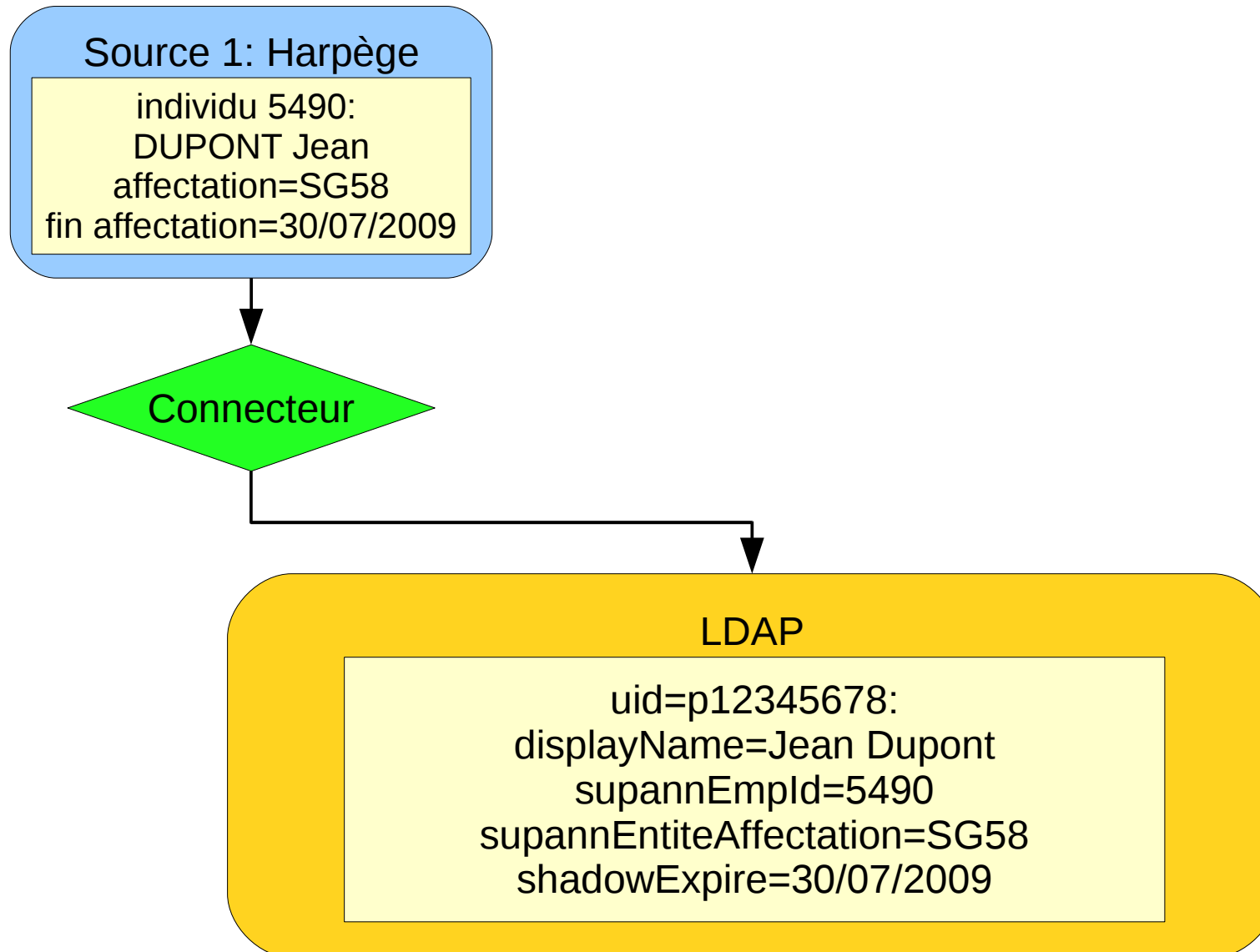
# Schéma d'alimentation



# Difficultés rencontrées

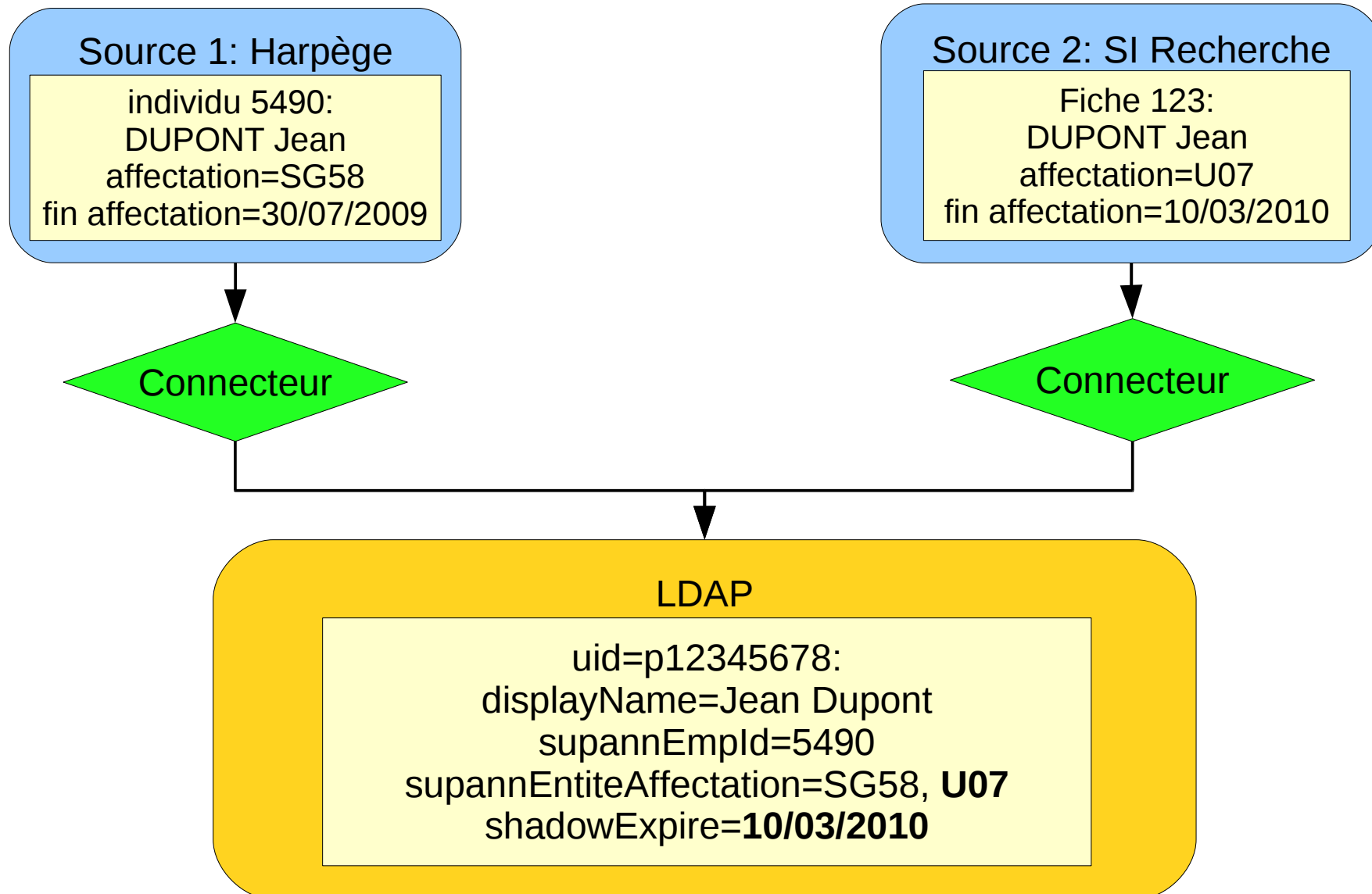
- **Dédoublonnage entre les sources:**
  - Pas d'identifiant commun: les entrées à sources multiple ne peuvent pas être rapprochées automatiquement. Selon les cas le rapprochement est manuel (nom, prénom, date de naissance...., ex. invité → Harpège) ou inexistant (thésard + ATER = deux compte pour la même personne)
  - Pas de remontée des homonymes à la saisie en dehors de ceux enregistrés dans l'application même où est effectuée la saisie
- **Contraintes d'intégrité faible:**
  - LDAP n'est pas une base relationnelle, toutes les vérifications d'intégrité doivent se faire à l'alimentation par les connecteurs.
- **Gestion de la durée de vie des attributs:**
  - Chaque entrée LDAP a une date d'expiration globale, mais en cas de relai d'une source à l'autre, il est difficile de déterminer quels attributs sont encore valides.

# Durée de vie des attributs 1/3

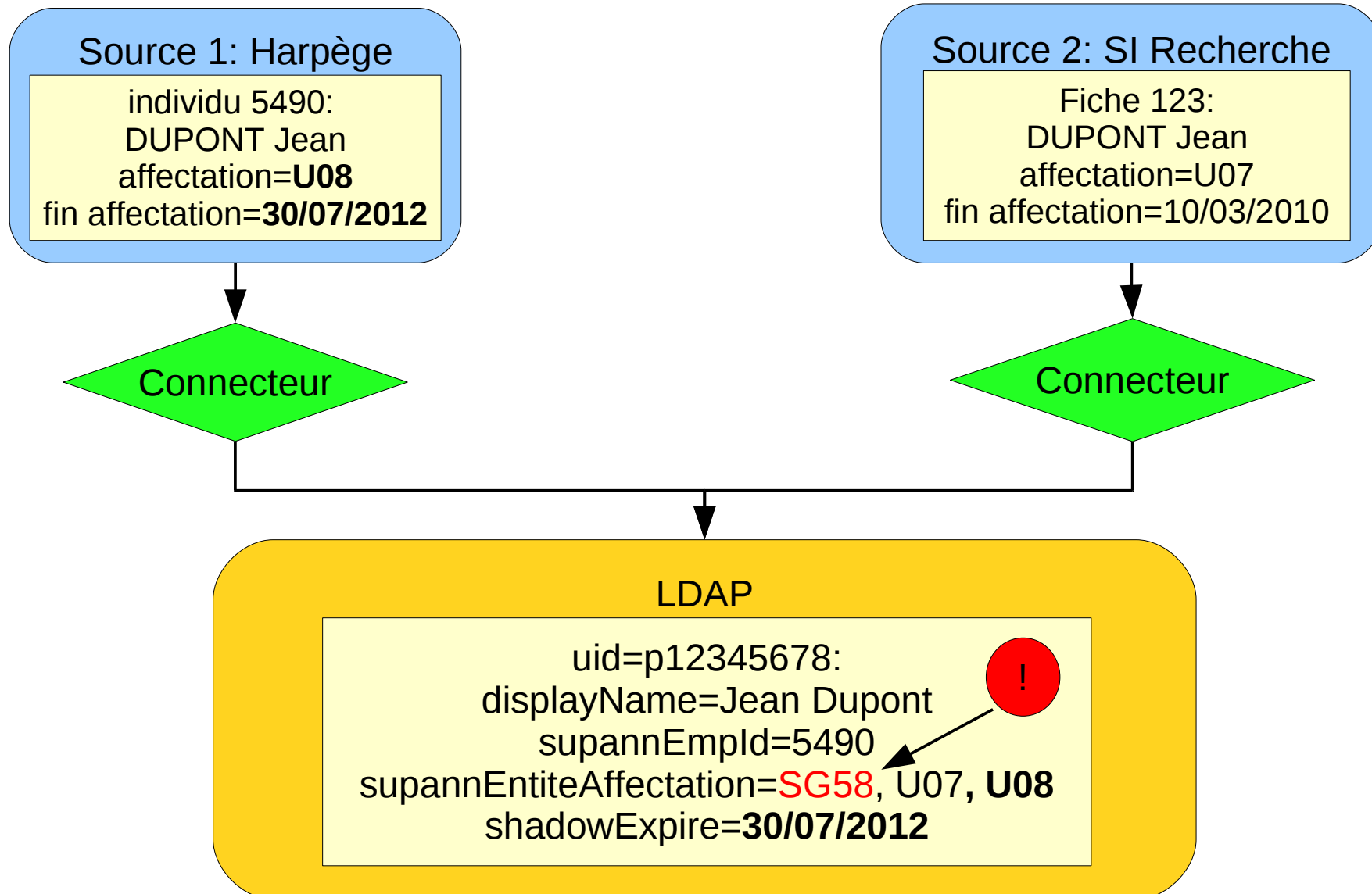




# Durée de vie des attributs 2/3



# Durée de vie des attributs 3/3



# Solution 1: connecteur multiprise ?

## Source 1: Harpège

individu 5490:  
DUPONT Jean  
affectation=U08  
fin affectation=30/07/2012

## Source 2: SI Recherche

Fiche 123:  
DUPONT Jean  
affectation=U07  
fin affectation=10/03/2010

Connecteur

LDAP

uid=p12345678:  
displayName=Jean Dupont  
supannEmpId=5490  
supannEntiteAffectation=U07, U08  
shadowExpire=30/07/2012

# Solution 2: référentiel

## Source 1: Harpège

individu 5490:  
DUPONT Jean  
affectation=U08  
fin affectation=30/07/2012

## Source 2: SI Recherche

Fiche 123:  
DUPONT Jean  
affectation=U07  
fin affectation=10/03/2010

## Référentiel

uid=p12345678  
affectation1=U08, source=source1, fin=30/07/2012  
affectation2=U07, source=source2, fin=10/03/2010

## LDAP

uid=p12345678:  
displayName=Jean Dupont  
supannEmpId=5490  
supannEntiteAffectation=U07, U08  
shadowExpire=30/07/2012

# Une solution: le référentiel unique

- Base de données relationnelle
  - Contrôle d'intégrité fort, pas de redondance
- Stockage de méta-informations d'arbitrage
  - source de chaque attribut avec durée de validité
  - identifiants utilisés dans chaque source
- Outils de dédoublonnage puissants
  - Système de "scoring", rapprochement phonétique, interface web...
  - Connecteurs de remontée des homonymes à la saisie
- Kit de déploiement de connecteurs uniformisés
  - Couche ETL configurable, protocoles standardisés: Java Queues, ESB, web services...
- **Le LDAP devient un « instantané » du référentiel**

# Futur

- Projet de référentiel de l'AMUE ?
- Evolutions de GRHUM (consortium Cocktail) ?
- Et si le LDAP n'était qu'un protocole d'interrogation, greffé directement sur le référentiel ?
  - Traduction dynamique filtre LDAP → requête SQL, avec mécanisme de cache...