BIBLIOTHÈQUE SCIENTIFIQUE NUMÉRIQUE
DIGITAL SCIENTIFIC LIBRARY

Conditor, vers un référentiel national des publications scientifiques françaises

Une recommandation de BSN3 (signalement)

Construire un référentiel de la production scientifique française en s'appuyant sur des référentiels communs aux universités et organismes de recherche (unités, thématiques, auteurs,…)

Journée d'étude sur les affiliations Renatis/Urfist

19 juin 2015

Annie Coret (CNRS-DIST)
Alain Zasadzinski (CNRS-INIST)

Sommaire

- **1 Présentation du projet**
  - Périmètre
  - Positionnement et grands principes de construction
- **2 L'expérimentation : une preuve de concept**
  - Partenaires, rôles et données fournies
  - Méthode de construction du corpus
  - Résultats et conclusions
- **3 Zoom sur le traitement des affiliations**
  - Principes
  - Apport du RNSR
- **4 Suite des travaux**
  - Macroplanning
  - Organisation du projet

1 Présentation du projet : périmètre

Un **réservoir de métadonnées de qualité :**

- **références bibliographiques**
  - **Titre, auteurs, titre de revue, ISSN,…**
- **structures ESR et tutelles**
- identifiants numériques d'articles et auteurs (DOI…)
- lien vers le texte intégral
- + financements, données liées, classifications / indexations…

Un **référentiel national** :

- détenant **tous les « objets numériques » pertinents**
- **utilisable par tous les dispositifs** de l'ESR :
  - Archives
  - Bases bibliométriques
  - Systèmes de gestion de la recherche (Crac, Ribac, Caplab…)
  - Systèmes décisionnels
- géré par une équipe de **professionnels de l'IST**

1 Présentation du projet : périmètre

~~Archive~~ ~~Outil décisionnel~~ ~~Portail~~ ~~Système d'Information Recherche~~

des métadonnées + un applicatif + une équipe d'administration

- **Captation de flux de notices**
  - réservoirs ESR
  - bases internationales…
- **Homogénéisation**
- **Dédoublonnage**
- **Enrichissements**

- **Fonctionnalités d'administration**
  - recherche, consultation,
  - création, modification, suppression, validation

- **Mise à disposition** des métadonnées normalisées et enrichies aux dispositifs ESR, leur permettant d'**éviter** la saisie et la curation de ces métadonnées

Utilisateurs « directs »

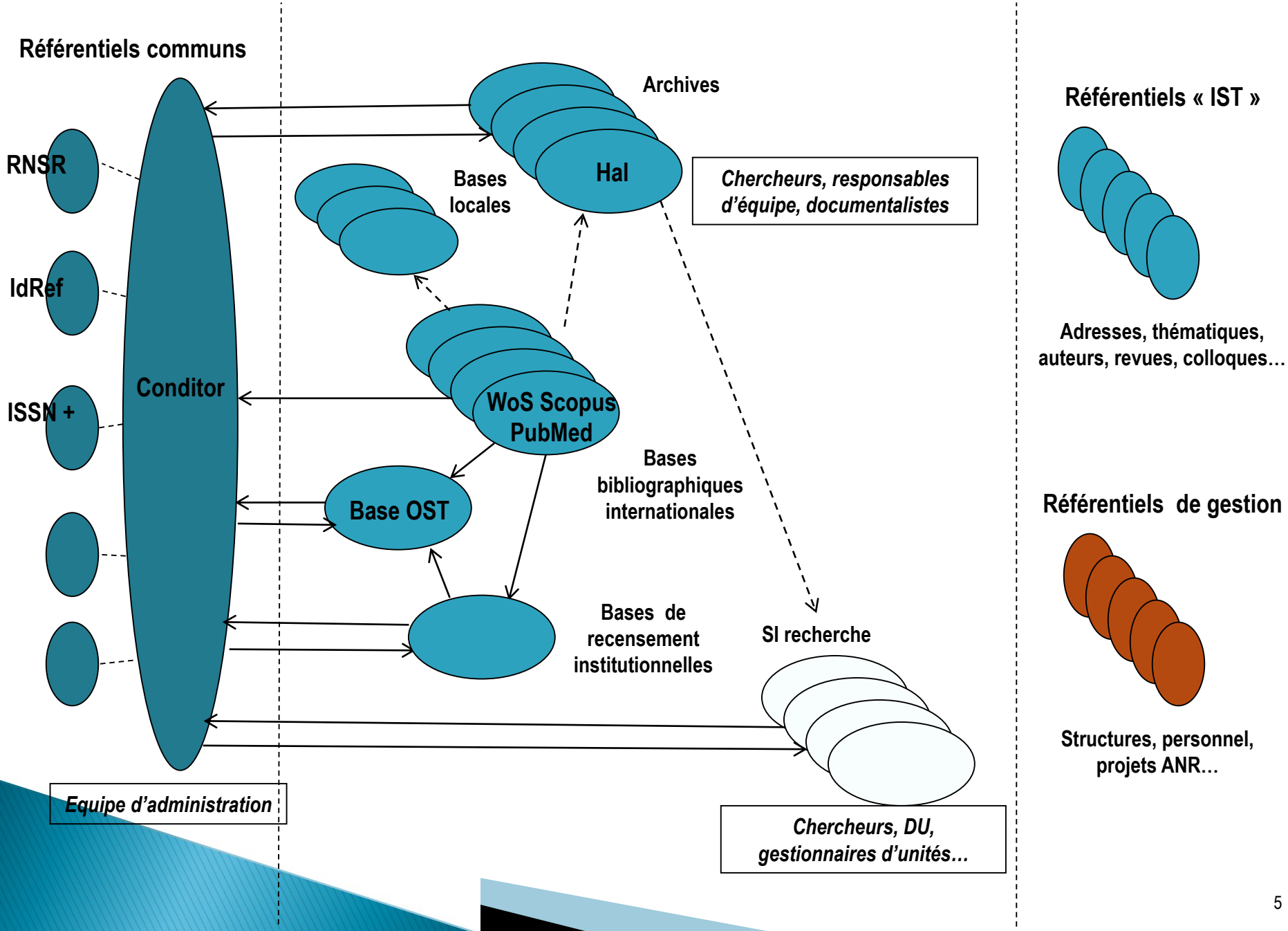
Une équipe de professionnels IST
Les Systèmes d'Information de l'ESR

Bénéficiaires « indirects »

Chercheurs, responsables d'équipes, producteurs d'indicateurs, décideurs…

Au travers d'applicatifs dédiés

1 Présentation du projet : positionnement



1 Présentation du projet : grands principes de construction

Privilégier le **pragmatisme**

Miser sur la **complémentarité** et la **mutualisation**

- Utiliser les données dont on dispose
- S'organiser progressivement pour minimiser les efforts de chacun
- Avancer en faisant => Une expérimentation pour :
  - Apprécier concrètement
    - les difficultés **de tous ordres** à surmonter
    - les bénéfices **potentiels** pour chaque partenaire
    - l'implication nécessaire de chacun
  - Constituer un groupe de partenaires volontaires
    - pour agir ensemble et de façon coordonnée

Equipe

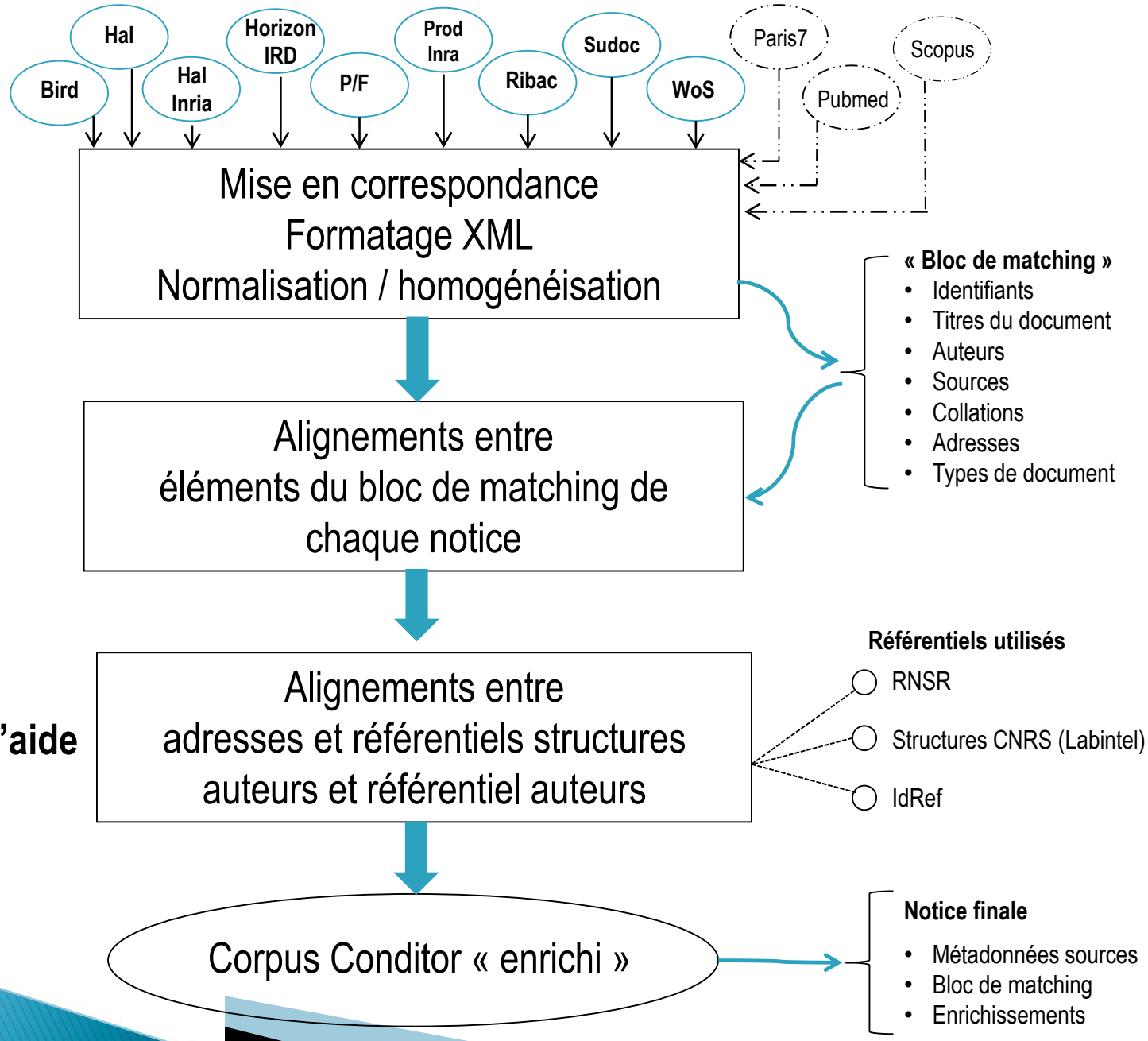
Multi-partenaires et multi-métiers

2 L'expérimentation : partenaires, rôles, données fournies

Partenaires	Fournisseurs de corpus de notices	Fournisseurs de référentiels	Autres
Abes	**Sudoc** : catalogue collectif des documents accessibles dans les bibliothèques et centres de documentation de l'ESR	**IdRef** : référentiel auteurs du catalogue Sudoc ISSN : données présentes dans le Sudoc	
CNRS-CCSD	**Hal** : archive nationale		
CNRS-Dastr-SAP2S		Dictionnaire des **différentes formes d'adresses** de chaque unité CNRS Extrait du **référentiel des structures CNRS** Labintel	
CNRS-Dist			Organisateur
CNRS-Inist	**Pascal-Francis** : bases de données bibliographiques multidisciplinaires		Opérateur technique
CNRS-INSHS	**Ribac** : base de données décrivant l'activité des chercheurs en SHS		
Inra	**ProdInra** : archive institutionnelle		
Inria	**Hal Inria** : archive institutionnelle		
IRD	**Horizon** : archive institutionnelle **Base bibliométrique**		
MESR		**RNSR** : Répertoire national des structures de recherche	
Université de Bordeaux			Expert
Université Paris Dauphine	**Bird** : archive ouverte institutionnelle		

Engagement réciproque
de non-réutilisation des données

2 L'expérimentation : méthode de constitution du corpus



Etape 1
Préparation des MétaDonnées (MD)

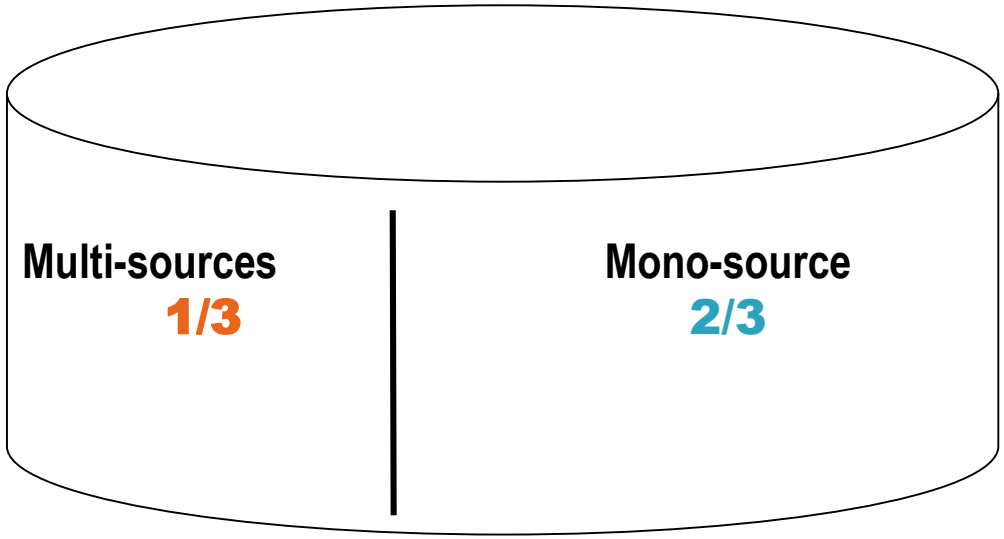
Etape 2
Détection des doublons

Etape 3
Enrichissements à l'aide des référentiels

2 L'expérimentation : résultats

Une expérimentation grandeur nature

une année de production (2011)



Notice présente dans 3 réservoirs ESR : Bird, Hal, Hal Inria

Optimal control problems of BV trajectories with pointwise state constraints More like this

Auteurs Forcadel, Nicolas ; Rao, Zhiping ; Zidani, Hasnaa ;

Source 18th IFAC World Congress. Milan. 2011-08-28.

Type Conference

id-conditor 148376-conditor

Origine E

Bird Hal Hal Inria

Curation et enrichissement

Liens externes

HAL inria-00639021
BIRD 7436
DOI 10.3182/20110828-6-IT-1002.01694
pdf [FoRaZi_IFAC.pdf,]

IDREF
FORCADEL, Nicolas131176897
LABO UMR7534; UMR7641;

Financement
SADCO: Sensitivity Analysis for Deterministic Controller Design.

Données proposées à Bird

Adresse affiliation HAL
ENSTA ParisTech, UMA, Unité de Mathématiques Appliquées, FRANCE
INRIA, UMR7641 CNRS, Polytechnique - X, ENSTA ParisTech, INRIA Saclay - Ile de France, COMMANDS, Centre de Mathématiques Appliquées Ecole Polytechnique Route de Saclay 91128 Palaiseau, FRANCE
UMR7534 CNRS, Université Paris IX - Paris Dauphine, CEREMADE, CEntre de REcherches en MAthématiques de la DEcision, Place du Maréchal de Lattre de Tassigny 75775 - Paris Cedex 16, FRANCE

Adresse affiliation INRIA
ENSTA ParisTech, UMA, Unité de Mathématiques Appliquées, France
UMR7534 CNRS, Université Paris IX - Paris Dauphine, CEREMADE, CEntre de REcherches en MAthématiques de la DEcision, France
UMR7641 INRIA, CNRS, Polytechnique - X, ENSTA ParisTech, INRIA Saclay - Ile de France, COMMANDS, France

Structure d'affiliation (equipe, labo, ...)
CENTRE DE MATHEMATIQUES APPLIQUEES - CNRS: UMR7641 # RNSR: 199719340P [1997-] UMR7641 (CNRS - EC POLYTECHNIQUE)
CENTRE DE RECHERCHES EN MATHEMATIQUES DE LA DECISION - CNRS: UMR7534 # RNSR: 199712592E [1997-] UMR7534 (CNRS - PARIS DAUPHINE)
CONTROLE OPTIMISATION MODELES METHODES ET APPLICATIONS POUR LES SYSTEMES DYNAMIQUES NON LINEAIRES - RNSR: 200718995W [2009-] EPI (INRIA)
SACLAY ILE DE FRANCE - RNSR: 200818248E [2008-] CRI (INRIA)

Tutelles
CNRS; ECOLE NATIONALE SUPERIEURE DE TECHNIQUES AVANCEES (ENSTA) PARISTECH; ECOLE POLYTECHNIQUE; INRIA; UNIV PARIS 9 - PARIS DAUPHINE;

Notice présente dans 4 réservoirs ESR : Hal, Hal Inria, Inist, WoS

Nonlinear methods for inverse statistical problems More like this

Auteurs Barbillon, Pierre ; Celeux, Gilles ; De Rocquigny, Etienne ; Grimaud, Agnes ; Lefebvre, Yannick ;

Source Comput. Stat. Data Anal. 55(1):132-142, JAN 1 2011.

Journal Computational statistics & data analysis ;

ISSN 0167-9473 - Print ; 1872-7352 - Online ;

ISSN liaison 0167-9473 ;

Type Article

id-conditor 000077-conditor

Origine

Hal, Hal Inria, base Inist (P/F), WoS

Editeur

ELSEVIER SCIENCE BV, AMSTERDAM.

Adresse affiliation HAL

Ecole Centrale Paris, MAS, Mathématiques Appliquées aux Systèmes - EA 4037, Grande Voie des Vignes 92295 Châtenay-Malabry, FRANCE
Schlumberger, GIER Schlumberger, GIER Schlumberger, 12 place des Etats-Unis 92120 Montrouge, FRANCE
UMR6206 CNRS, Université de la Méditerranée - Aix-Marseille II, IML, Institut de mathématiques de Luminy, case 907 163 Av de Luminy 13288 MARSEILLE CEDEX 09, FRANCE
UMR8628 CNRS, Université Paris XI - Paris Sud, LM-Orsay, Laboratoire de Mathématiques d'Orsay, FRANCE

Adresse affiliation INRIA

Ecole Centrale Paris, MAS, Mathématiques Appliquées aux Systèmes - EA 4037, France
Schlumberger, GIER Schlumberger, GIER Schlumberger, France
UMR6206 CNRS, Université de la Méditerranée - Aix-Marseille II, IML, Institut de mathématiques de Luminy, France
UMR8628 CNRS, Université Paris XI - Paris Sud, LM-Orsay, Laboratoire de Mathématiques d'Orsay, France

Pré cisions de Hal et Hal Inria

Adresse affiliation Inist

Ecole Centrale Paris, FRANCE.
Institut Mathematiques de Luminy, FRANCE.
Schlumberger, FRANCE.
Université Paris-sud 11, INRIA Saclay, FRANCE.

Adresse affiliation WOS

Ecole Cent Paris, Paris, France.
UFR Sci, Math Lab, Batiment 425, F-91405 Orsay, France.
Univ Paris 11, INRIA Saclay, Orsay, France.

Enrichissement via le RNSR

Structure d'affiliation (equipe, labo, ...)

INSTITUT DE MATHEMATIQUES DE LUMINY - CNRS: UMR6206 # RNSR: 200412255H [2004-2011] UMR6206 (AIX MARSEILLE 2 - CNRS) - RNSR: 201220249M [2012-] FRE3529 (AIX MARSEILLE - CNRS)
LABORATOIRE DE MATHEMATIQUES D ORSAY - CNRS: UMR8628 # RNSR: 199812953T [1998-] UMR8628 (CNRS - INRIA - PARIS 11)
MATHEMATIQUES APPLIQUEES AUX SYSTEMES - RNSR: 200615286U [2006-] EA4037 (CIRAD - EC CENTRALE PARIS - INRIA)

Tutelles

CIRAD; CNRS; ECOLE CENTRALE PARIS; INRIA; UNIV AIX MARSEILLE; UNIV PARIS 11 - PARIS SUD;

Liens externes

HAL inria-00441967
WOS WOS:000283017900012
INIST 23321081
DOI 10.1016/j.csda.2010.05.030
pdf [RR-7156.pdf,]

IDREF
BARBILLON, Pierre156566834
CELEUX, Gilles02951598X
GRIMAUD, Agnes110881036
See more

2 L'expérimentation : conclusions

- **Des points à améliorer dans la méthode de construction**
  - Détection des doublons
  - Ajout de données à partir de **référentiels nationaux** de structures et d'auteurs
- **Des enseignements**
  - Le tout automatique ne suffit pas
  - Conditor est « faisable » et nécessaire
- **Les apports potentiels**
  - Eviter d'avoir à saisir ce qui l'est déjà ailleurs grâce à un « **entrepôt commun** de notices descriptives »
  - Contribuer à l'amélioration des référentiels nationaux, à l'interopérabilité entre référentiels voire à la mise en place de **référentiels communs** (structures, tutelles, auteurs, colloques…)

Un zoom sur le traitement des adresses : le principe
Enrichissement via les référentiels de structures

Notice bibliographique Conditor :

Adresses « France »

Référentiel de structures :

RNSR

Codes de l'unité (fonction des organismes)
Acronymes de l'unité
Noms de l'unité

+

Ville
Tutelle

Un zoom sur le traitement des adresses : le principe
Enrichissement via les référentiels de structures

Notice bibliographique Conditor :

Adresses « France »

Métadonnées Hal

U1016 INSERM, UMR8104 CNRS, Université Paris V - Paris Descartes, Institut Cochin, 22 rue Méchain, 75014 Paris, FRANCE

Métadonnées P/F Inist

Inserm, U1016, Institut Cochin, Paris, France and Cnrs, UMR8104, Paris, FRANCE

Référentiel de structures :

RNSR

Codes de l'unité (fonction des organismes)
u1016 umr8104
Acronymes de l'unité
um3
Noms de l'unité
centre de recherche institut cochin

+

Ville
paris
Tutelle
inserm
chr chu
inra
paris 5
cnrs

Un zoom sur le traitement des adresses : le principe
Enrichissement via les référentiels de structures

Notice bibliographique Conditor :

Adresses « France »

Métadonnées Hal

U1016 INSERM, **UMR8104** CNRS, Université Paris V - Paris Descartes, Institut Cochin, 22 rue Méchain, 75014 **Paris**, FRANCE

Métadonnées P/F Inist

Inserm, **U1016**, Institut Cochin, Paris, France and Cnrs, **UMR8104**, **Paris**, FRANCE

Référentiel de structures :

RNSR

Codes de l'unité (fonction des organismes)
u1016 **umr8104**
Acronymes de l'unité
um3
Noms de l'unité
centre de recherche institut cochin

+

Ville
paris
Tutelle
inserm
chr chu
inra
paris 5
cnrs

Alignements exacts de chaînes de caractères
entre
adresses de chaque notice
et éléments du référentiel structure

Un zoom sur le traitement des adresses : le principe
Enrichissement via les référentiels de structures

Notice bibliographique Conditor :

Adresses « France »

Métadonnées Hal

U1016 INSERM, **UMR8104** CNRS, Université Paris V - Paris Descartes, Institut Cochin, 22 rue Méchain, 75014 **Paris**, FRANCE

Métadonnées P/F Inist

Inserm, **U1016**, Institut Cochin, Paris, France and Cnrs, **UMR8104**, **Paris**, FRANCE

Référentiel de structures :

RNSR

Codes de l'unité (fonction des organismes)
u1016 umr8104
Acronymes de l'unité
um3
Noms de l'unité
centre de recherche institut cochin

+

Ville
paris
Tutelle
inserm
chr chu
inra
paris 5
cnrs

Alignements exacts de chaînes de caractères entre
adresses de chaque notice
et éléments du référentiel structure

Ajout des identifiants nationaux et enrichissements avec des éléments :
Tutelles
Types de structures et Codes UAI

Centre de recherche institut cochin
200217519N

Inserm ; CHR CHU ; INRA ; Paris 5 ; CNRS

Un zoom sur le traitement des adresses : apport potentiel du RNSR

Notice Conditor

Métadonnées Hal

U1016 INSERM, UMR8104 CNRS, Université Paris V - Paris Descartes, Institut Cochin, 22 rue Méchain, 75014 Paris, FRANCE

Métadonnées P/F Inist

Inserm, U1016, Institut Cochin, Paris, France and Cnrs, UMR8104, Paris, FRANCE

Eléments « matchant »avec le RNSR

Enrichissements

Item RNSR

<code>200217519N
- <LIBELLE>CENTRE DE RECHERCHE INSTITUT COCHIN
- <SIGLE>UM3
- <VILLE>PARIS
- <LABO>
  - <CODE_UNITE>U1016
  - <TUTELLE_SIGLE>INSERM
- <LABO>
  - <CODE_UNITE>UM3
  - <TUTELLE_SIGLE>CHR CHU
- <LABO>
  - <CODE_UNITE>UM3
  - <TUTELLE_SIGLE>INRA
- <LABO>
  - <CODE_UNITE>UM3
  - <TUTELLE_SIGLE>PARIS 5
- <LABO>
  - <CODE_UNITE>UMR8104
  - <TUTELLE_SIGLE>CNRS

Mise à disposition de l'Inra d'une notice Conditor absente du corpus ProdInra potentiellement pertinente

Notice présente dans 2 réservoirs ESR : Hal, Inist

AMP-activated protein kinase and metabolic control. More like this

Auteurs Andreelli, Fabrizio ; Viollet, Benoit ;

Source Handb. exp. pharmacol. 203:303-330, 2011.

Journal Handbook of experimental pharmacology ;

ISSN 0171-2004 - Print ; 1865-0325 - Online ;

ISSN liaison 0171-2004 ;

Type Article

id-conditor 091691-conditor

Origine HAL IN

Pays GERMANY

Editeur
Springer, Heidelberg.

Financement
ANR-06-PHYSIO-026: AMP-activated et S6 kinases: roles opposés dans l'adaptation du muscle squelettique à l'état nutritionnel et à l'exercice.
EXGENESIS: Health benefits of exercise: identification of genes and signalling pathways involved in effects of exercise on insulin resistance, obesity and the metabolic syndrome.
Association Française contre les Myopathies (AFM) Association pour l'Etude des Diabètes et des Maladies Métaboliques (ALFEDIAM) Institut Benjamin Delessert.

Adresse affiliation HAL
Assistance publique - Hôpitaux de Paris (AP-HP), Hôpital Pitié-Salpêtrière, Université Pierre et Marie Curie (UPMC) - Paris VI, Service de diabétologie Pitié-Salpêtrière, 47-83, boulevard de l'Hôpital 75651 Paris Cedex 13, FRANCE
U1016 INSERM, UMR8104 CNRS, Université Paris V - Paris Descartes, Institut Cochin, 22 rue Méchain, 75014 Paris, FRANCE

Adresse affiliation Inist
Department of Diabetology, Pitie-Salpetriere Hospital AP-HP, Univ Pierre et Marie Curie-Paris 6, Paris, FRANCE.
Fakultat Lebenswissenschaften, Molekulare Pharmakologie und Toxikologie, Technische Universitat Braunschweig, Beethovenstrasse 55, 38106 Braunschweig, GERMANY.
Inserm, U1016, Institut Cochin, Paris, France and Cnrs, UMR8104, Paris, FRANCE.
Univ Paris Descartes, Paris, FRANCE.

Structure d'affiliation (equipe, labo, ...)
CENTRE DE RECHERCHE INSTITUT COCHIN - RNSR: 200217519N [2002-] U1016 (INSERM) ; UM3 (CHR CHU - INRA - PARIS 5) ; UMR8104 (CNRS)
INSTITUT COCHIN - CNRS: UMR8104

Tutelles
AP-HP; CNRS; INRA; INSERM; UNIV PARIS 5 - PARIS DESCARTES; UNIV PARIS 6 - PIERRE ET MARIE CURIE;

Liens externes

HAL inserm-00597010
INIST 24091751
DOI 10.1007/978-3-642-17214-4_13
pdf [AMPK_metabolic_control_HAL_.pdf, inserm-00597010_edited.pdf, index.xhtml,]

IDREF
ANDREELLI, Fabrizio035798025
LABO UMR8104;

HAL Base Inist (P/F)

Information présente dans Hal mais pas dans la base Inist

Curation et enrichissement

Adresse présente dans la base Inist mais pas dans Hal

Donnée ajoutée grâce au RNSR

Un zoom sur le traitement des adresses : apport potentiel du RNSR

Extrait d'une notice RNSR

<code>199712570F

<LIBELLE>**INSTITUT ELIE CARTAN DE LORRAINE**

<SIGLE>**UMR7502**

<VILLE>**VANDOEUVRE LES NANCY**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**CNRS**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**INP NANCY**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**INRIA**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**LORRAINE**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**NANCY 1**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**NANCY 2**

<LABO>

<CODE_UNITE>**UMR7502**

<TUTELLE_SIGLE>**PRES NANCY**

Fournir au PRES de Nancy ses publications et désormais aux COMUEs

4 La suite des travaux : phase projet

Macroplanning (après organisation du projet)

Année 1	Année 2

Négociation des ressources et des conventions

Constitution d'un corpus de notices sur plusieurs années

Définition des métadonnées à gérer et de la stratégie de construction puis mise en œuvre dont formatages, …

Corpus 0	Corpus 1

Conception et développement itératif de l'applicatif

Mise en place progressive des fonctionnalités

Recherche, modification, suppression, validation, mise à disposition	Collecte, formatage, dédoublonnage Enrichissements via les référentiels Industrialisation des échanges

Conception et mise en place du service Conditor

Organisation de l'équipe d'administration et de la structure pérenne de régulation, gestion de la montée de puissance...

4 Suite des travaux : organisation du projet

Equipes multi-partenaires et multi-métiers à tous les « niveaux »

- Grands ensembles de tâches du projet (= work packages) :
  - Négociation des ressources et des conventions
  - Constitution d'un corpus de notices sur plusieurs années
  - Conception et développement itératif de l'applicatif
  - Conception et mise en place du service Conditor
- Comités techniques
- Pilotage stratégique et opérationnel

Implication des partenaires

Fourniture d'un corpus (notices ou référentiel)
Contribution à une ou plusieurs tâches, à un comité technique …
Responsabilité d'une ou plusieurs tâches
Pilotage ou copilotage d'un workpackage ou d'un comité technique
…