

Pôle GLC

Présentation des activités

Resp. Michel RIVEILL

Adj. Philippe LAHIRE

- Effectifs
 - 26 permanents, environ 30 doctorants et 5 ingénieurs sous contrat
 - Augmentation notable du nombre de doctorants
- Un pôle jeune, fortement renouvelé
 - Très peu de départs à la retraite dans les 10 prochaines années
 - Nombreux recrutements récents au niveau MCF
 - 6 arrivées pour 1 départ
 - + nombreuses arrivées au cours de la période précédente
 - Reconversion, Intégration
 - Reprise d'une activité de recherche

- Très impliqué dans les activités d'enseignement et dans les structures de l'Université
 - Responsables de parcours du master IFI
 - 5 sur les 8 parcours
 - Responsable de la MIAGE (DU / L3 – M2)
 - 300 étudiants
 - Adjoint à la direction de l'UFR Sciences
 - Entreprise, recherche
 - Direction du département informatique de Polytech puis vice-président Sophia

- Fortement soutenu par le CNRS et l'INRIA

- CNRS :

- Même avec un seul chercheur
- + 2 ingénieurs supports (janv. 2007 et janv. 2011),
- Plusieurs délégations (2)
- Moyen mi-lourd pour l'Ubiquarium



- INRIA :

- Même en l'absence de projet commun
- Délégations (3) ou détachement (1)
- Vrai depuis 8 ans : a concerné presque toutes les thématiques

Présentation du pôle

Présentation
I3S 2011
pôle GLC

Présentation
Recherche
Bilan
Projets



- Fort sentiment d'appartenance
 - Réelle vie collective
 - Journées de pôle, séminaires
 - Le pôle est devenu une réalité aussi forte que les équipes
 - Équipes : lieu de vie au quotidien
 - Pôle : lieu de solidarité
 - Aide aux équipes les plus faibles
 - Scientifiquement
 - Financièrement
 - Nombreuses relations inter-équipes qui commencent à avoir des Résultats
 - Publications, contrats

Domaines scientifiques :

Logiciel + Connaissance

- RA 1 : Intergiciels

- Pour les systèmes distribués de grande taille, application à l'analyse d'images médicales ([Modalis](#))
- Pour l'informatique ambiante, application au bâtiment intelligent ([Rainbow](#))

- RA 2 : Composition logicielle et ingénierie des modèles

- Application à la fusion et migration des IHM ([Rainbow](#))
- Application à la construction et adaptation de workflows ([Modalis](#))
- Application à l'auto-adaptation des systèmes et des applications ([Modalis](#), [Rainbow](#))

- RA 3 : Ingénierie des connaissances et web sémantique

- Activité transverse
- Représentation des connaissances ([Kewi](#))
- Application à la modélisation et au support des communautés en ligne ([Kewi](#))

- RA 4 : Sécurité

- Activité transverse
- Protocoles sécurisés orientés propriétés ([Rainbow](#))

- RA 5 : Fouille de données

- Application aux données biologiques et à la relation client ([Keia](#))



RA3 - Web sémantique

Michel Buffa

Analyse de réseaux sociaux à l'aide de
Frameworks du web sémantique

- Intégration **S**émantique de l'**I**nformation par des **C**ommunautés d'**I**ntelligence en **L**igne
- **réseaux sociaux d'entreprise,**
- **veille technologique,**
- **communautés d'intérêt, de pratique, d'experts**
- Partenaires : ergonomes, sociologues, mathématiciens, ontologistes, informaticiens...



Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2
- RA4
- RA5
Bilan
Projets

ADEME



INRIA

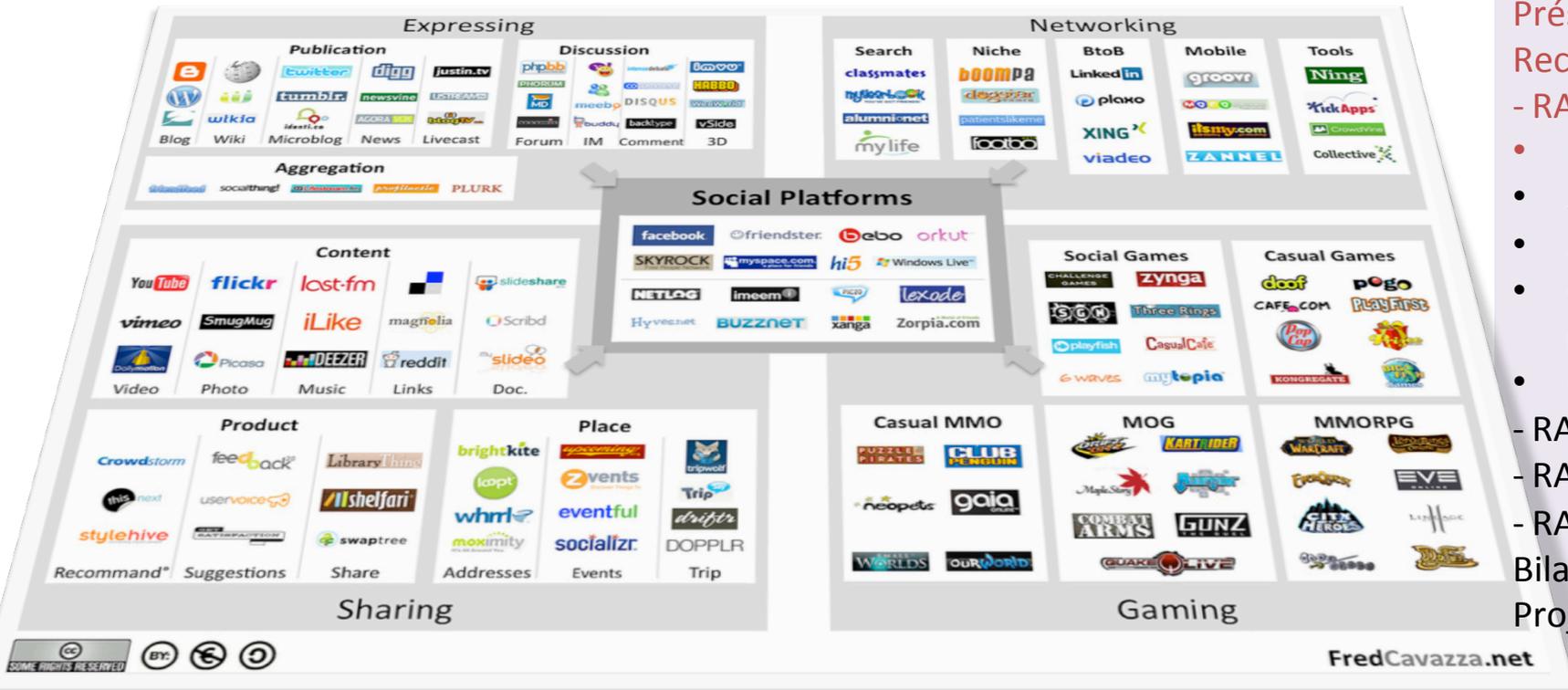


utt

université de technologie
Troyes



Social media landscape

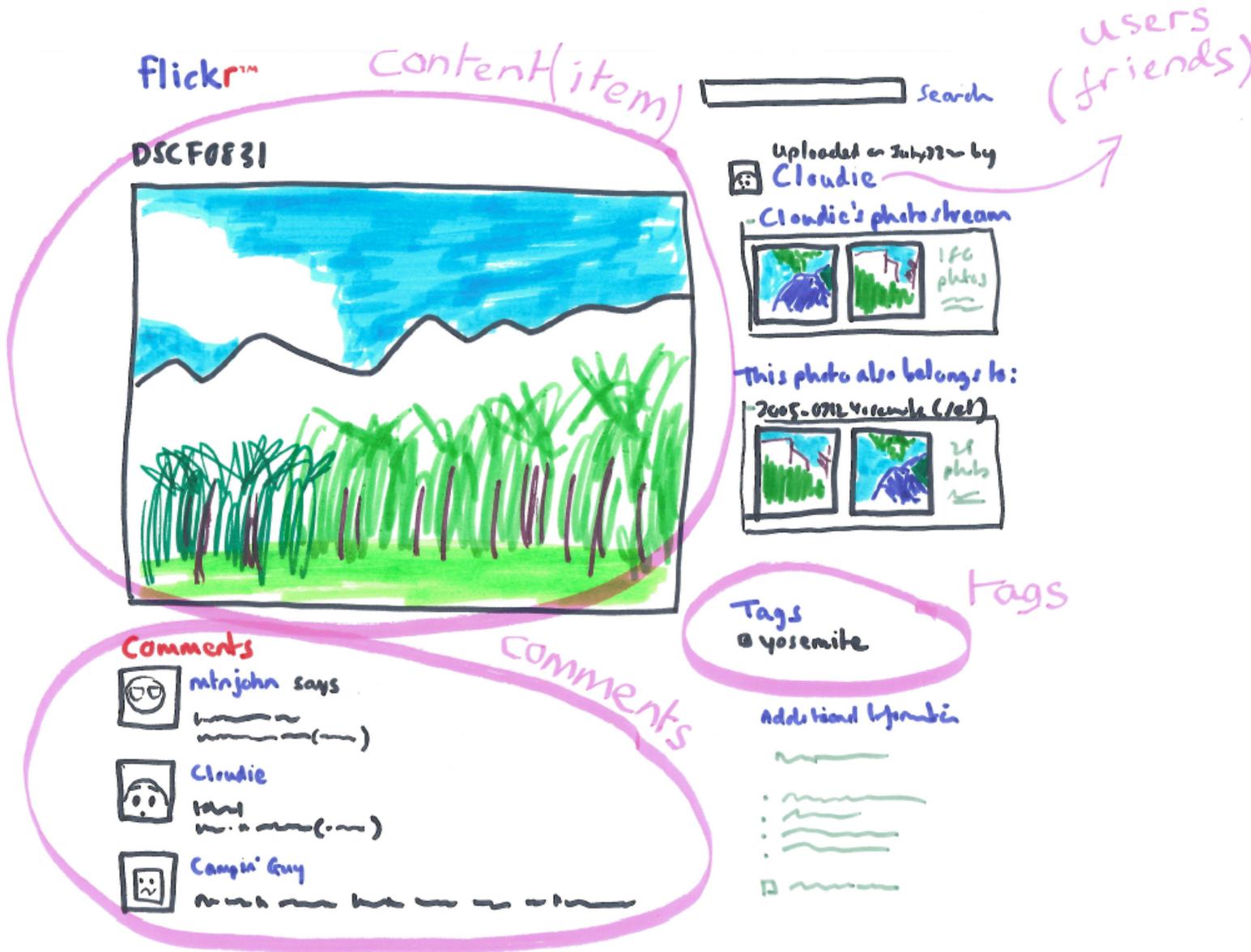


Présentation
Recherche
- RA3

- Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
- RA1 + RA2
- RA4
- RA5
Bilan
Projets

Réseaux sociaux explicites ou pas, mais les interactions révèlent des réseaux implicites. **L'entreprise 2.0 adapte ces outils et usages aux intranets**

Flickr/Youtube etc : classification par tags



- Présentation Recherche
- RA3
 - Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
 - RA1 + RA2
 - RA4
 - RA5
 - Bilan
 - Projets

- Rapprocher Web 2.0 et Web Sémantique
 - Rapprocher social tagging (folksonomies) et ontologies
 - Folksonomie = ensemble des tags utilisés, vocabulaire de la communauté
 - Ontologie = représentation formelle de la connaissance
 - Aller vers de l'analyse de réseaux sociaux "sémantique"
 - Modéliser les réseaux par des graphes RDF
 - Définir des modèles génériques pour représenter un réseau social, mais aussi les interactions entre les acteurs
 - Proposer des outils d'analyse et d'identification des communautés

Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2
- RA4
- RA5
Bilan
Projets

L'analyse des réseaux sociaux (SNA)

...propose des algorithmes basés sur des graphes pour caractériser la structure d'un réseau social, les positions et acteurs stratégiques, les activités remarquables, identifier les communautés, etc.



Présentation Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2

- RA4

- RA5

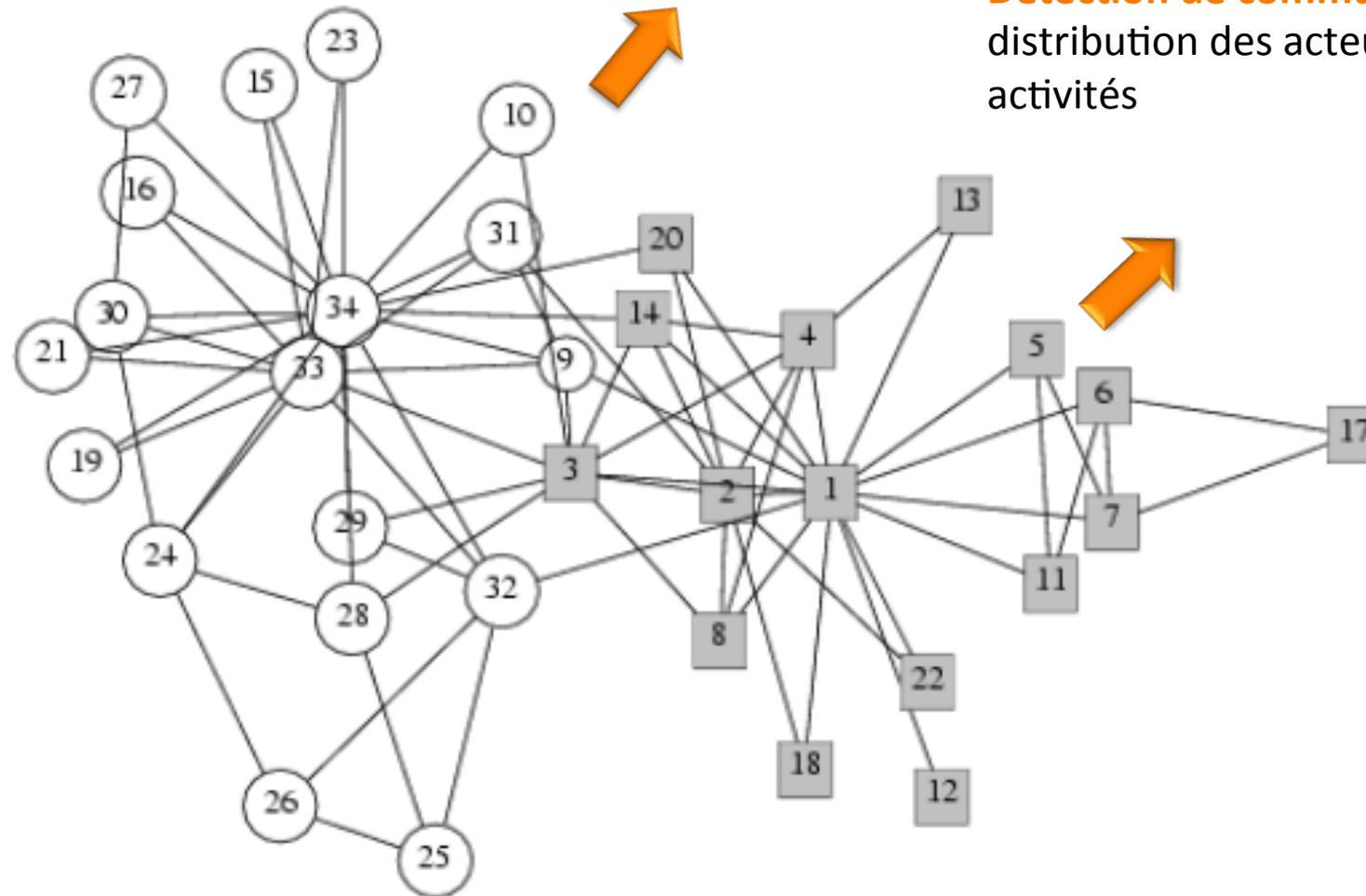
Bilan

Projets

L'analyse des réseaux sociaux (SNA)

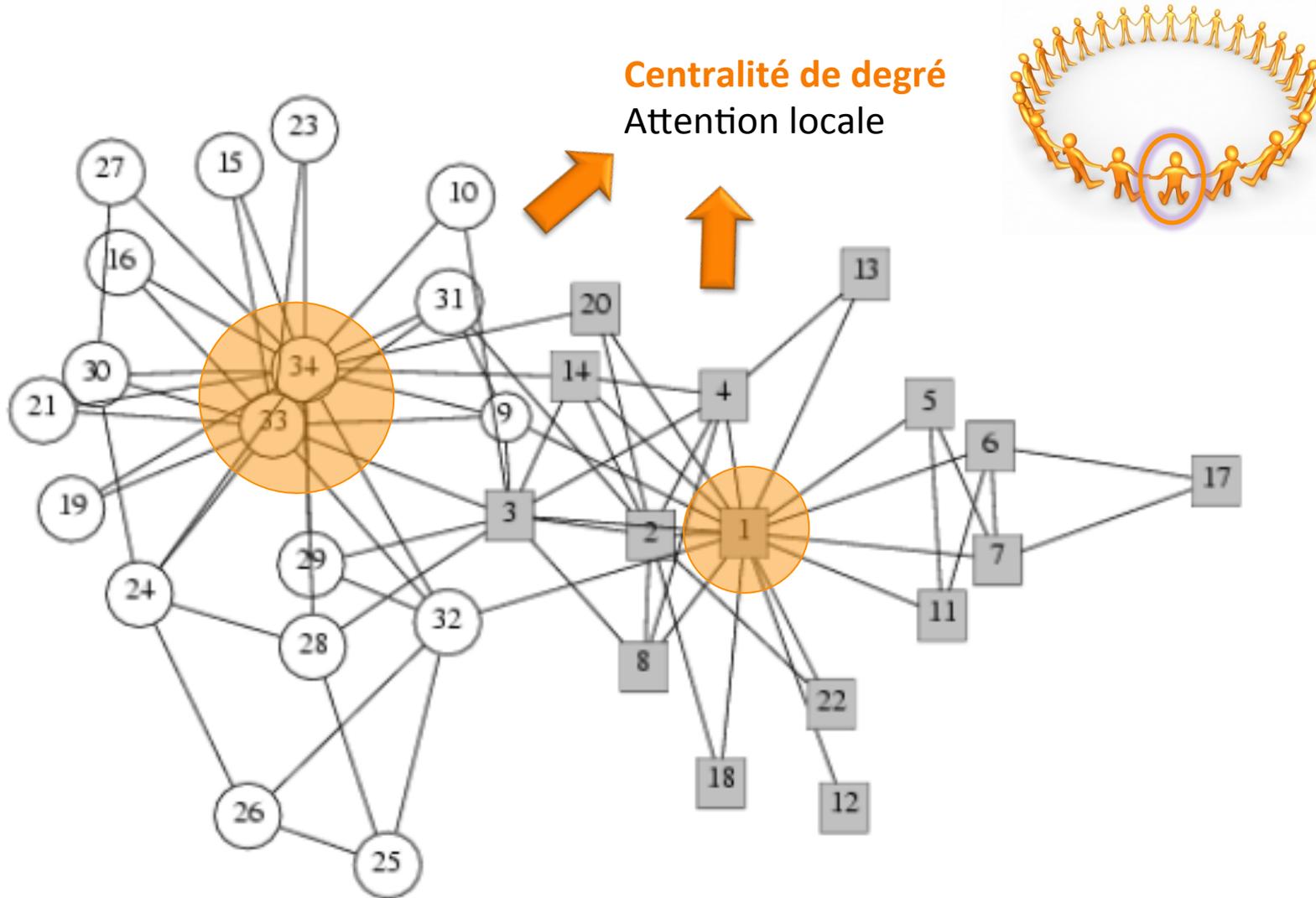
densité et diamètre
cohésion du réseau

Détection de communauté
distribution des acteurs et des activités



- Présentation Recherche - RA3
- Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
- RA1 + RA2
- RA4
- RA5
Bilan
Projets

L'analyse des réseaux sociaux (SNA)



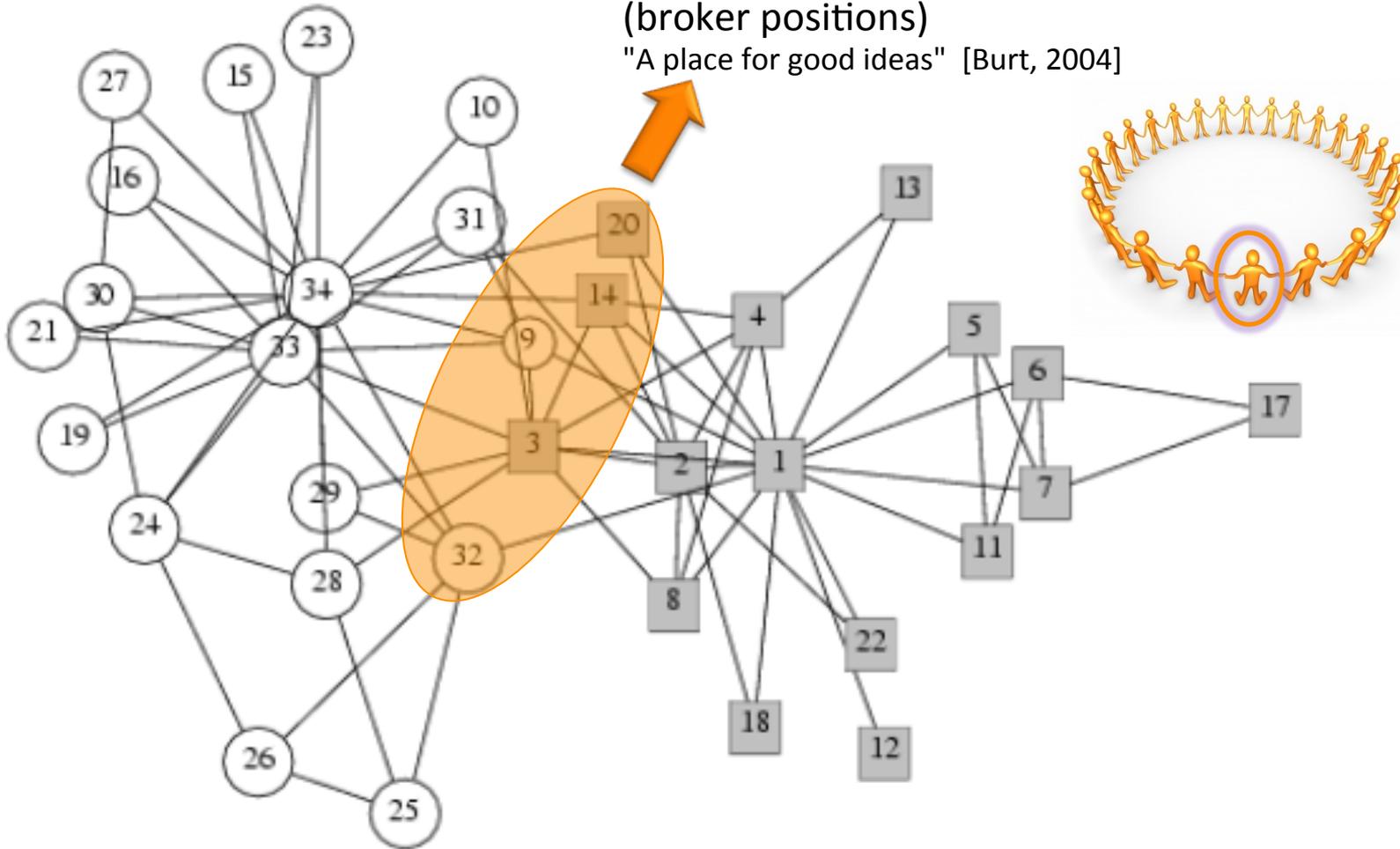
- Présentation
Recherche
- RA3
 - Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
- RA1 + RA2
- RA4
- RA5
Bilan
Projets

L'analyse des réseaux sociaux (SNA)

Centralité d'intermédiation

Révèle les positions intermédiaires
(broker positions)

"A place for good ideas" [Burt, 2004]



Présentation Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

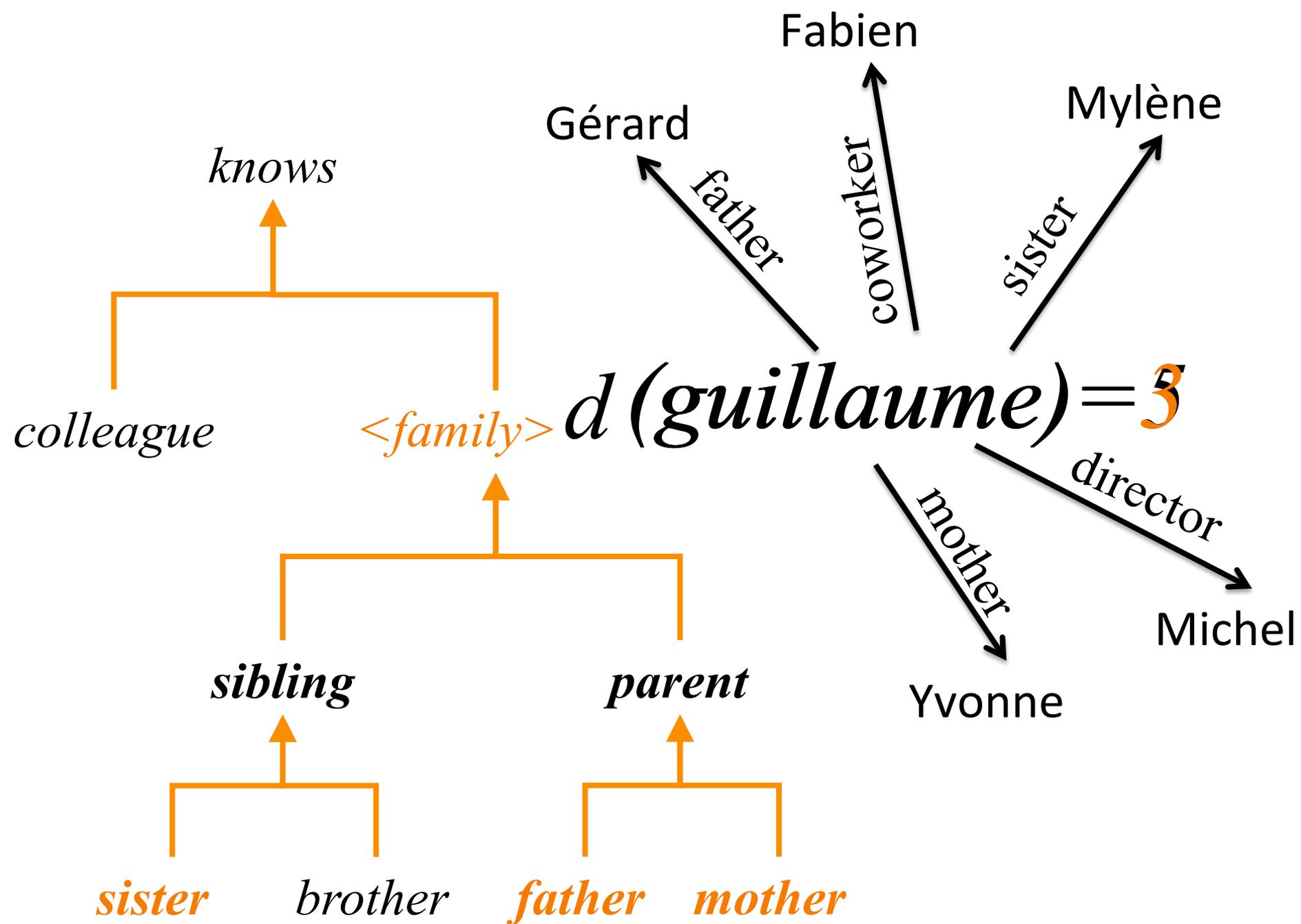
- RA1 + RA2

- RA4

- RA5

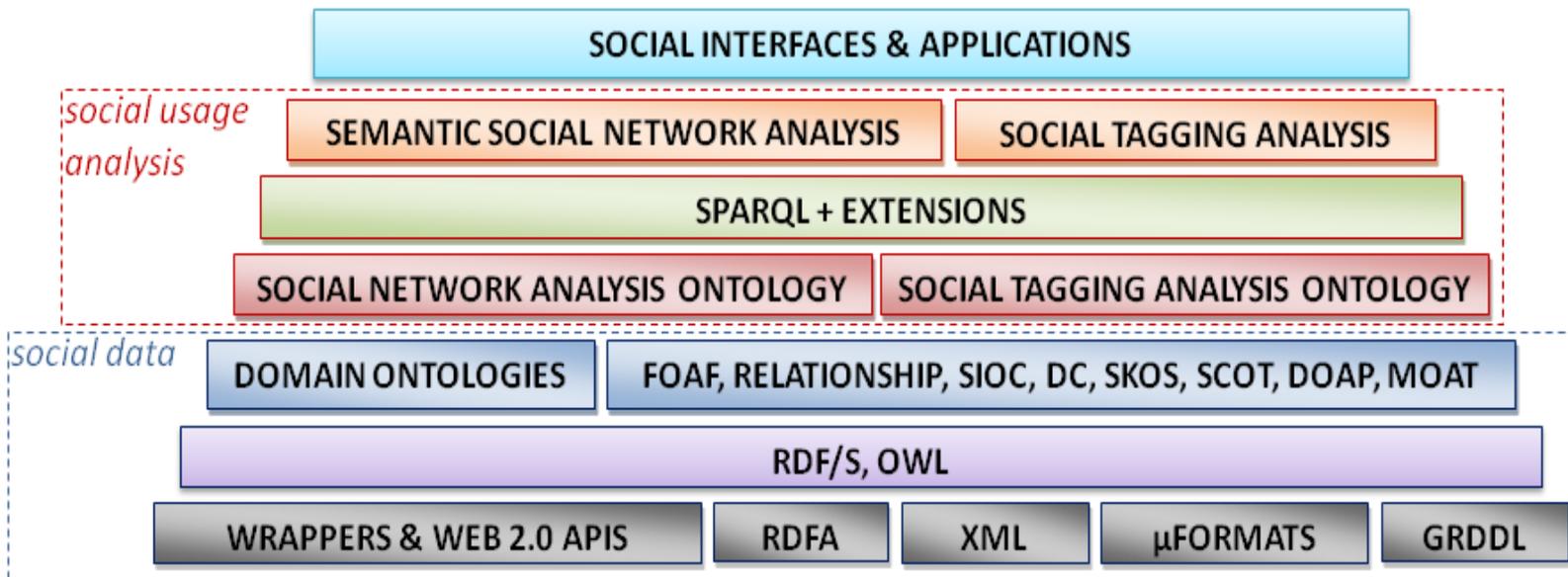
Bilan

Projets



Pile d'une SNA sémantique

- Exploiter la sémantique des réseaux sociaux
 - Contributions en rose
 - [ereteo et al, WebSci'09] [ereteo et al, ISWC2009]

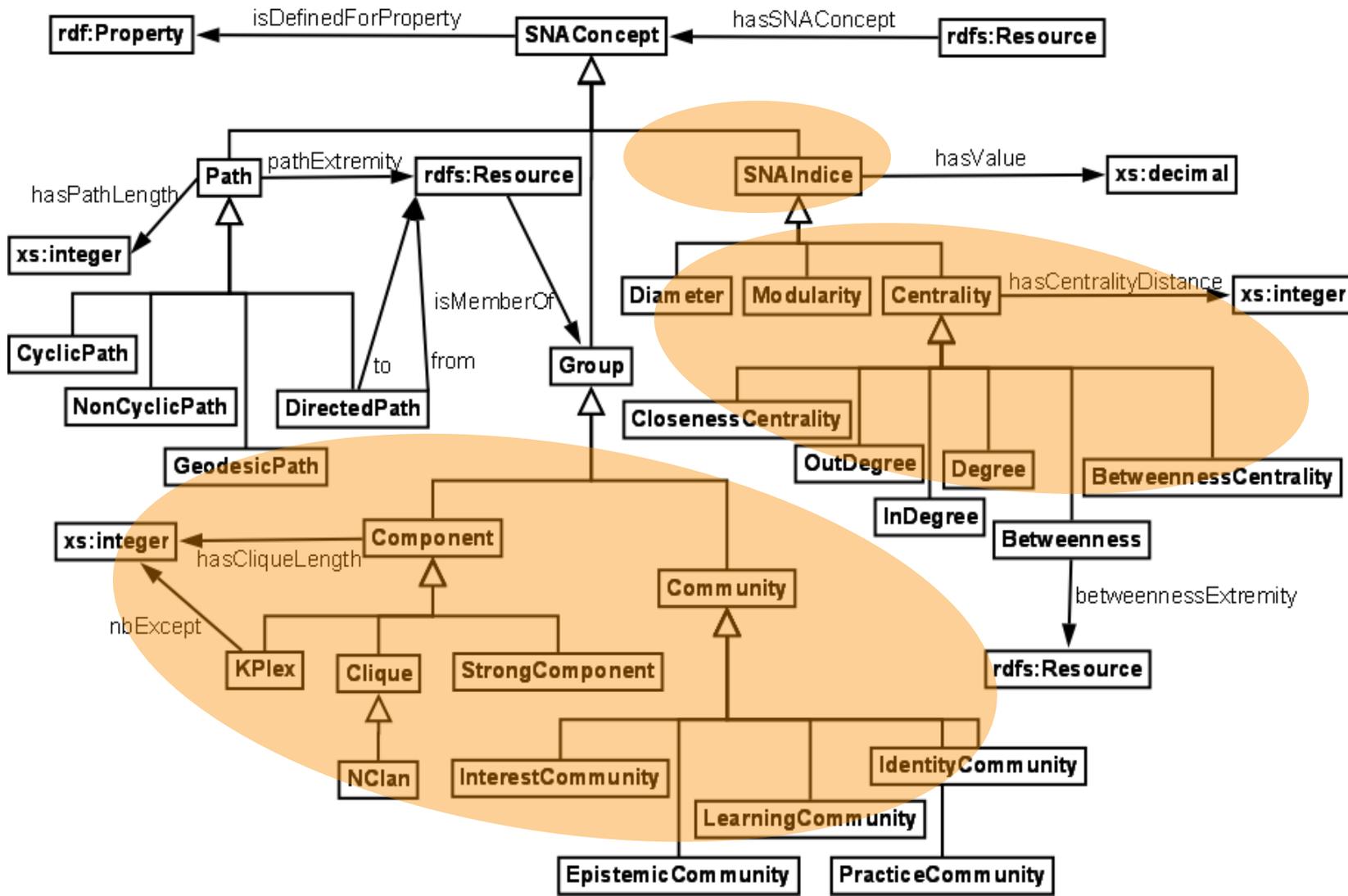


- Présentation
Recherche
- RA3
- Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
- RA1 + RA2
- RA4
- RA5
Bilan
Projets

SemSNA une ontologie de la SNA

- <http://ns.inria.fr/semsna/2009/06/21/voc>

Présentation
Recherche
- RA3
• Contexte
• Analyse
• SNA
• Communautés
• Contrib.
- RA1 + RA2
- RA4
- RA5
Bilan
Projets



Calcul des indices de la SNA en SPARQL

Nombre d'acteurs
impliqués dans une
relation donnée

```
 $nb_{<rel>}^{actor}(G)$ 
select merge count(?x) as ?nbactors from <G>
where{
  {?x param[rel] ?y}
  UNION{?y param[rel] ?x}
}
```

Nombre d'acteurs
impliqués dans un sujet

```
 $nb_{<rel>}^{subject}(G)$ 
select merge count(?x) as ?nbsubj from <G>
where{
  ?x param[rel] ?y
}
```

Composant

```
 $Comp_{<rel>}(G)$ 
select ?x ?y from <G> where {
  ?x param[rel] ?y
}group by any'
```

Degré entrant

```
 $D_{<rel,dist>}^{in}(y)$ 
select ?y count(?x) as ?indegree where{
  ?x (param[rel])*:: $path ?y
  filter(pathLength($path)g <= param[dist])
}group by ?y
```

Geodésique entre
from et to

```
 $g_{<rel>}(from,to)$ 
select ?from ?to $path pathLength($path) as
?length where{
  ?from sa (param[rel])*:: $path ?to
}group by ?from ?to
```

Nombre de géodesiques
entre *from et to*

```
 $nb_{<rel>}^g(from,to)$ 
select ?from ?to count($path) as ?count where{
  ?from sa (param[rel])*:: $path ?to
}group by ?from ?to
```

Diamètre

```
 $Diam_{rel}(G)$ 
select pathLength($path) as ?length from <G>
where {
  ?y s (param[rel])*:: $path ?to
}order by desc(?length) limit 1
```

Centralité de fermeture

```
 $C_{<rel>}^c(y)$ 
select distinct ?y ?to pathLength($path) as
?length (1/sum(?length)) as ?centrality
where{
```

Présentation
Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2

- RA4

- RA5

Bilan
Projets

Collaboration avec Ipernity.com

Présentation
I3S 2011
pôle GLC

beta
ipernity

Welcome to ipernity! [Sign up for free](#) | [Sign in](#)

share, meet explore

easy and free, have fun!

[Sign up!](#)

photos, videos, audios
all your contents at last together

family, friends,...
share with whoever you want

artist, author?
promote yourself to the world

Temptation "Shared by Christine Lebrasseur"

explore

[what's hot!](#) | [latest](#) | [world map](#) | [tags](#)

search

Enter one or more keywords [OK](#)

docs blogs groups members

© ipernity 2007-2009 [Català](#) | [Čeština](#) **NOVÉ** | [中文](#) | [Deutsch](#) | **English** | [Español](#) | [Esperanto](#) | [Ελληνικά](#) | [Français](#) | [Galego](#) | [Italiano](#) | [Nederlands](#) | [Português](#) | [More...](#)

[Login page](#) | [News](#) | [About ipernity](#) | [Jobs](#) | [Thanks!](#) | [Terms of service](#) | [Help](#)

Best photo sharing site [OPEN WEB AWARDS](#)

Sony World Photography Awards
Cannes 2009

Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2
- RA4
- RA5
Bilan
Projets

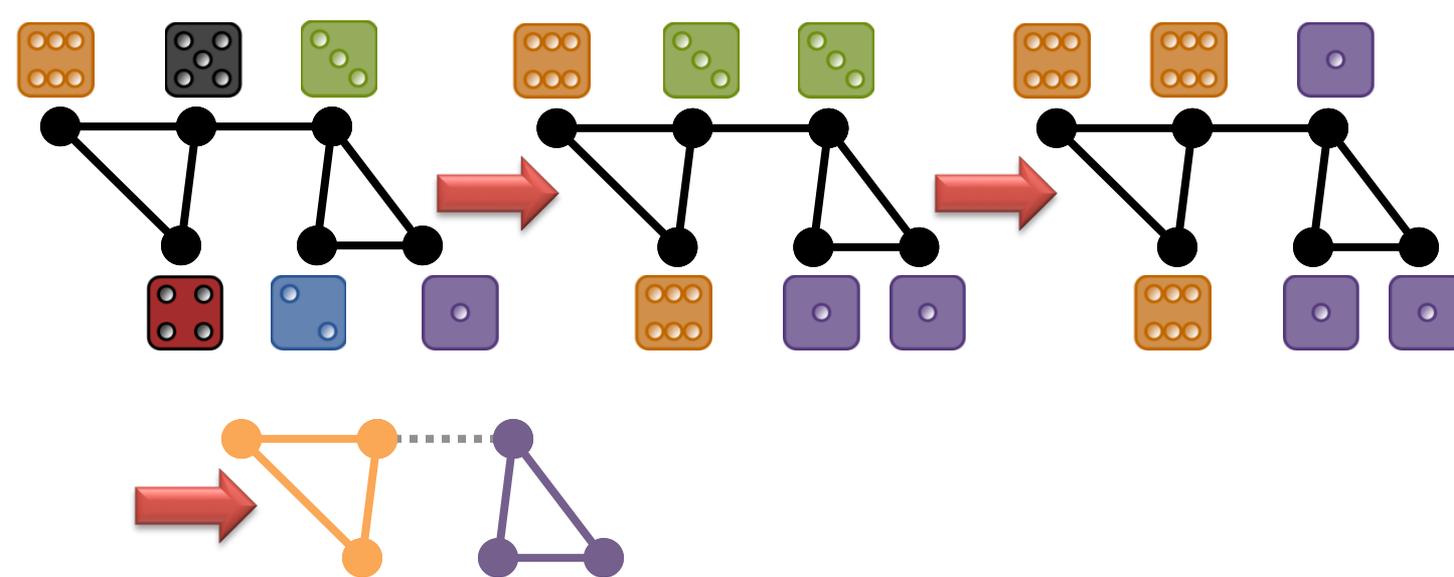
De vraies données !

- Jeu de données de ipernity.com en RDF: **61 937 acteurs & 494 510 relations**
 - 18 771 relations family entre 8 047 acteurs
 - 136 311 relations friend impliquant 17 441 acteurs
 - 339 428 relations favorite pour 61 425 acteurs
 - 2 874 170 commentaires par 7 627 acteurs
 - 795 949 messages échangés entre 22 500 acteurs

Résultats de l'analyse validés par Ipernity.com,

- **Publications : ISWC 2009, WebScience 2009**

- algorithmes basés sur la propagation de labels aléatoires



- On peut faire mieux : utiliser les tags !

- Contexte
- Analyse
- SNA
- Communités
- Contrib.

Des utilisateurs utilisant des tags sémantiquement proches forment des communautés d'intérêt

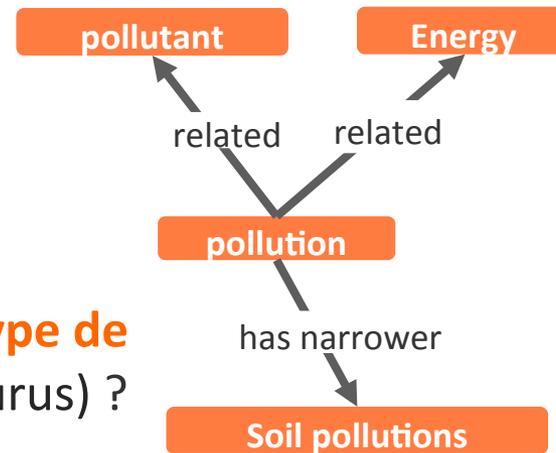
- Grand nombre de tags, comment les regrouper, les relier ?
- Difficile d'utiliser un dictionnaire (vocabulaire trop spécialisé, acronymes, typos, etc.)
- **Tirer parti des relations entre tags pour la détection de communautés**

communciation communication comparateur
 compensation comportement condition
 conference confiance **consommation**
 contact convention cosmetique coton
 covoiturage creation critique CSS cuisine
 dechet dell demarche developpement document
 donnee durable dynamique eau ecoconception
 ecolife ecologie ecomerce economie
 elevage emballage emission empreinte encre
 encyclopedie **energie** enquete enseigne
 entreprise **environnement** eolien epargne
 ericsson espece ethique etiquette etude
 europe evaluation evolution extensif faune
 federation fertilisation filiere film finance flore
 fondation forum france fruit GES
 goodguide grandedistribution graphique green
 greenpeace greenwashing grenelle guide
 habitat hitech HQE HTML hybride impact
 import imprimante inde indice industrie

Comment **transformer** les folksonomies ...



... En **ce type de structure** (thesaurus) ?



- Présentation Recherche
- RA3
 - Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
 - RA1 + RA2
 - RA4
 - RA5
 - Bilan
 - Projets

1 - Utiliser les distances syntaxiques

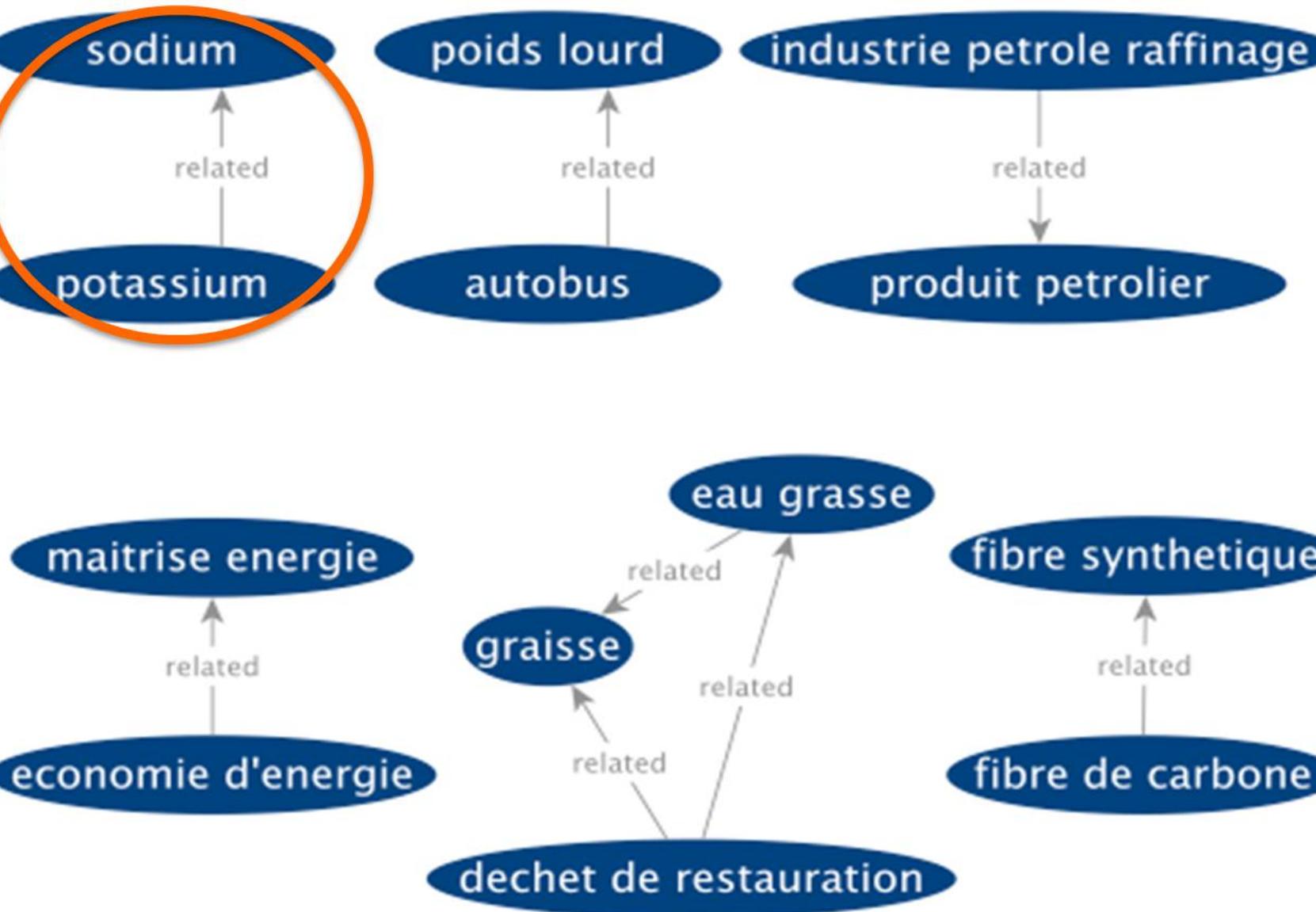


=> « pollution » **broader than** « soil pollutions »



=> « pollution » **related to** « pollutant »

2 – utiliser la co-occurrence entre tags

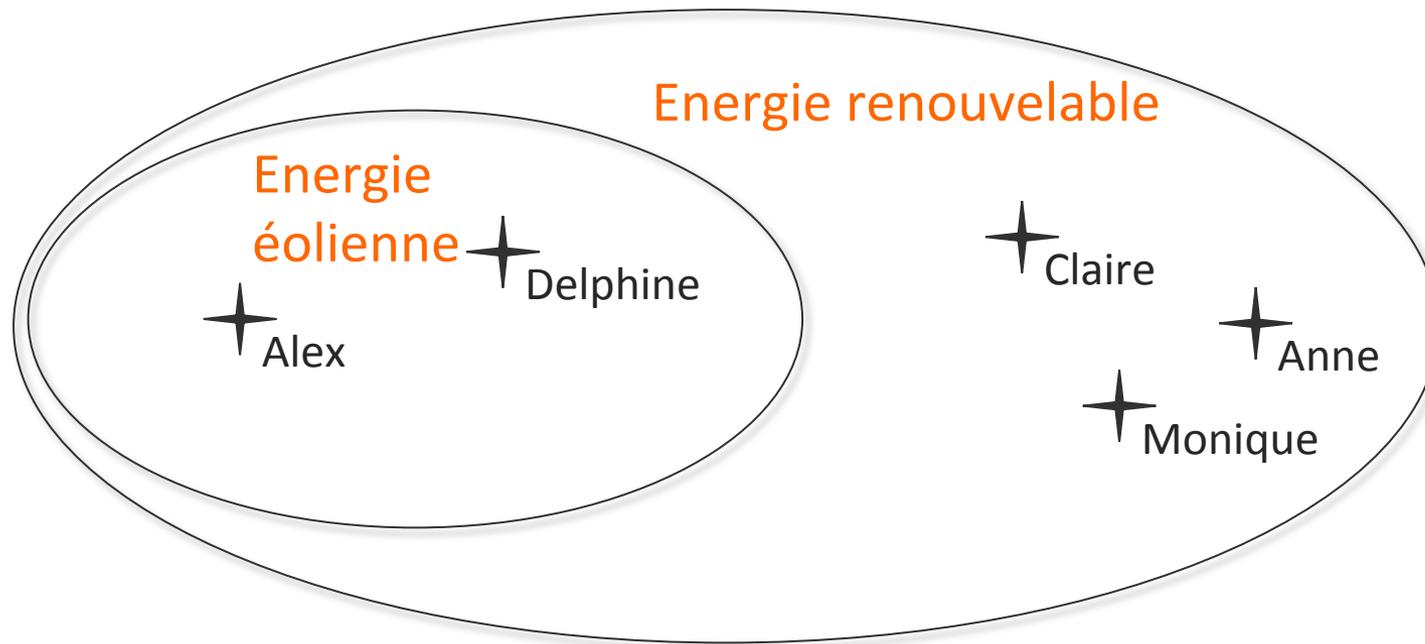


Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2
- RA4
- RA5
Bilan
Projets

3 – Utiliser les associations utilisateurs-tags

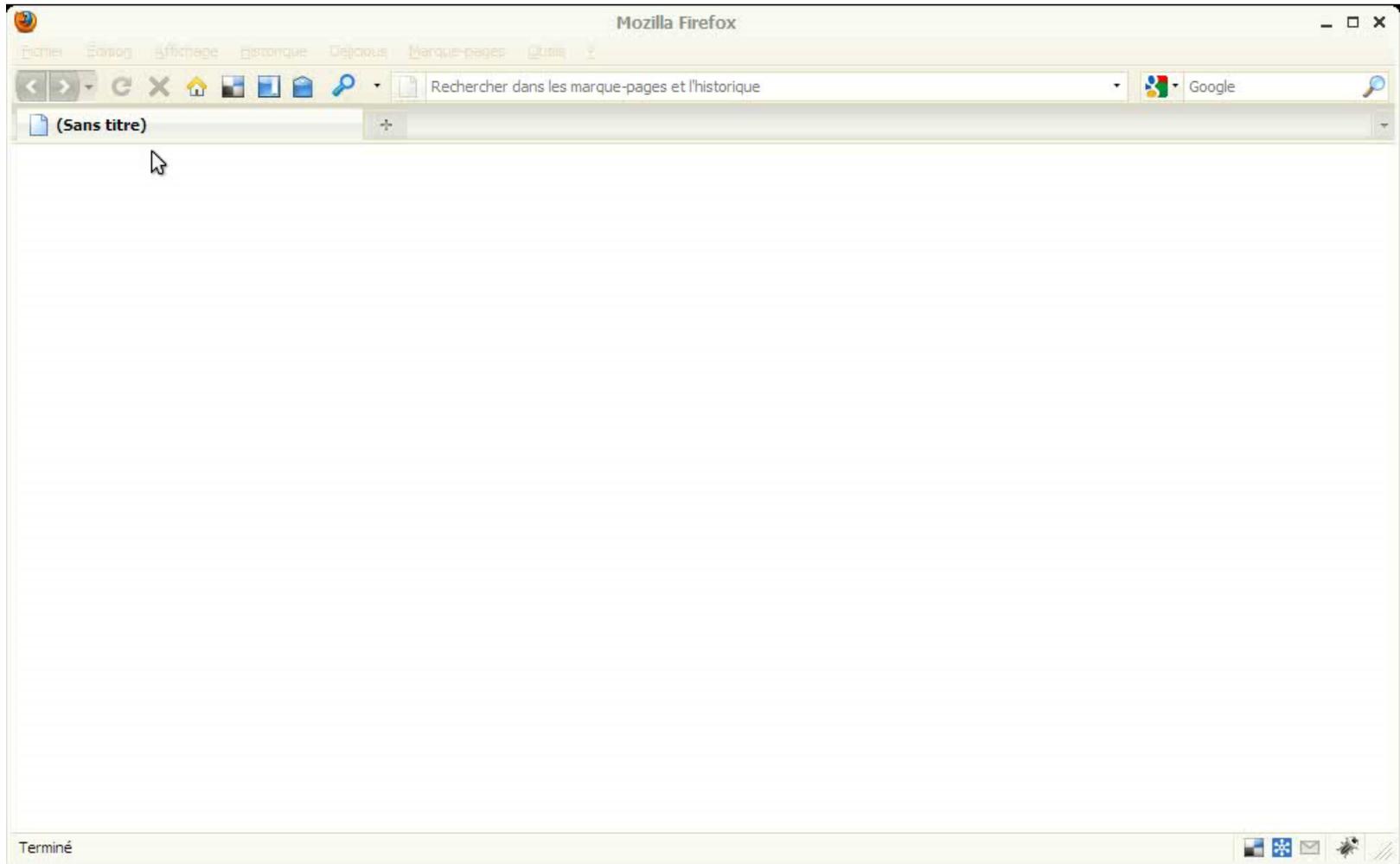


⇒ Relations hyponymes (**broader/narrower**) :

« énergie renouvelable » **broader than** « énergie éolienne »

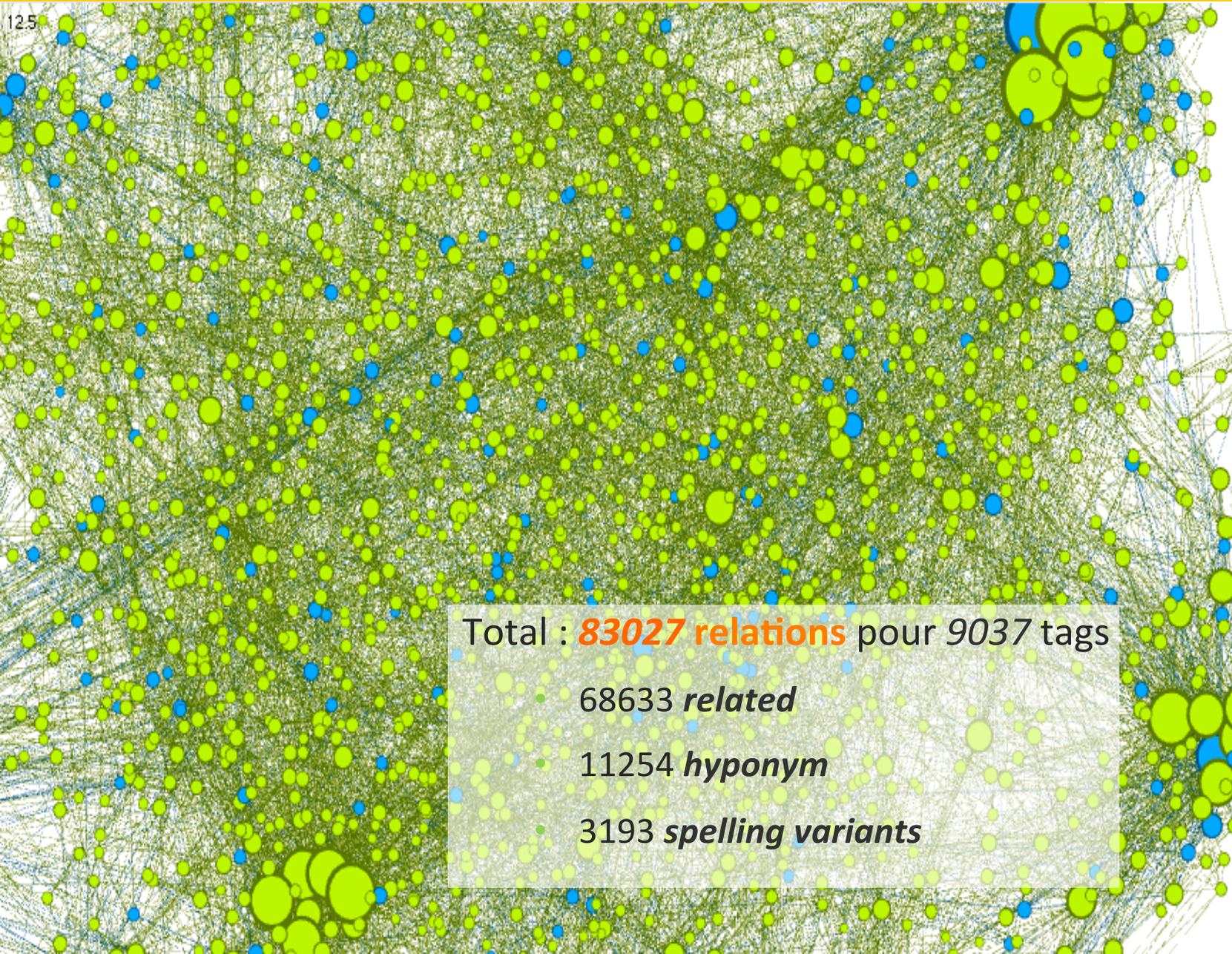
Corriger les erreurs : capturer les contributions des utilisateurs

*Inclure les tâches de **structuration** dans l'activité **quotidienne**
(par exemple : faire une recherche)*



Résultats sur le jeu de données complet

Présentation
I3S 2011
pôle GLC



Total : **83027 relations** pour 9037 tags

• 68633 *related*

• 11254 *hyponym*

• 3193 *spelling variants*

Présentation
Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2

- RA4

- RA5

Bilan

Projets

Exploiter la folksonomie de référence

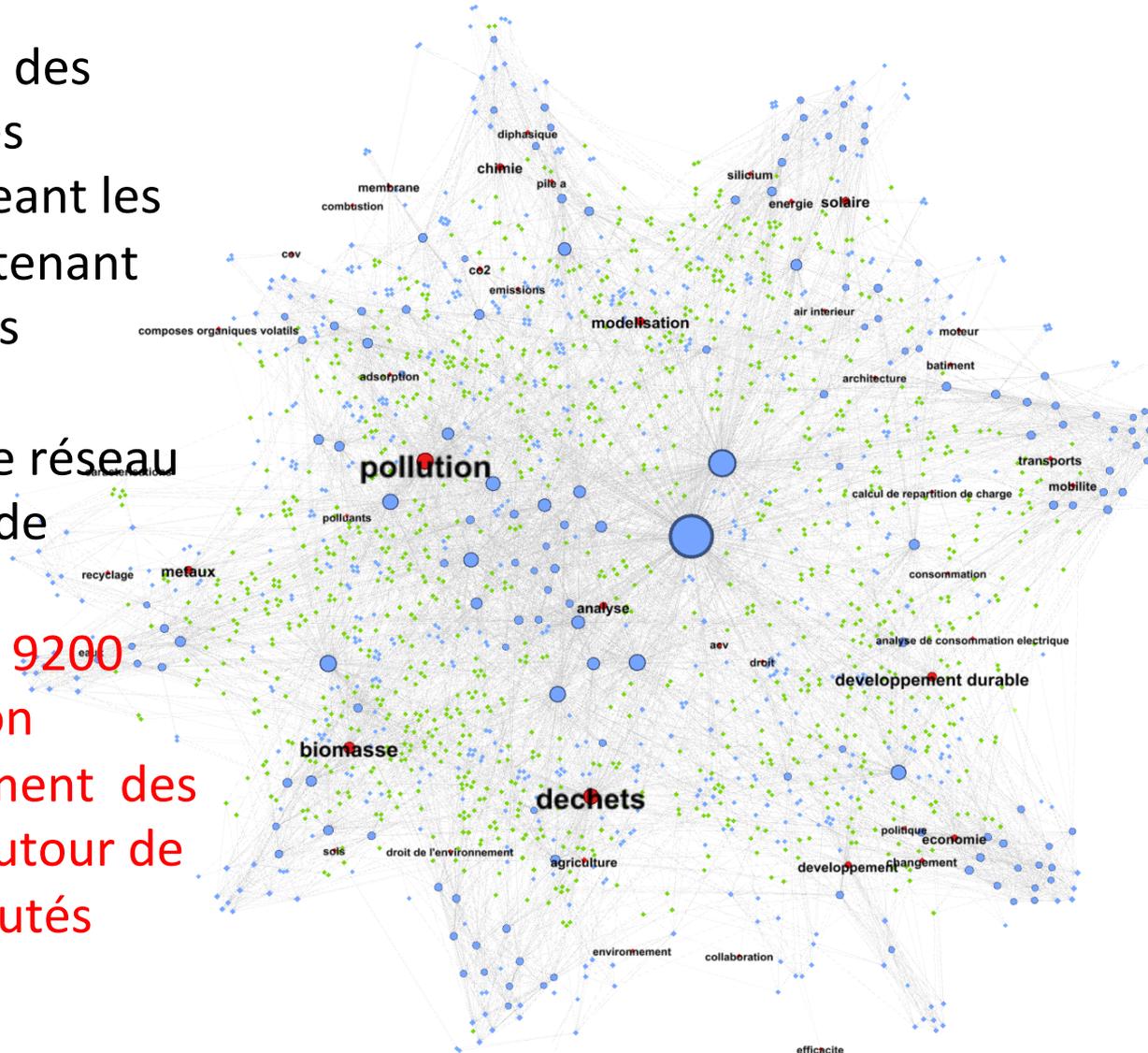
Pour le calcul des communautés

- En propageant les tags et en tenant compte des relations

- Testé sur le réseau TheseNET de l'ADEME

- On part de 9200 tags environ

→ regroupement des utilisateurs autour de 30 communautés thématiques



Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés

• Contrib.
- RA1 + RA2

- RA4

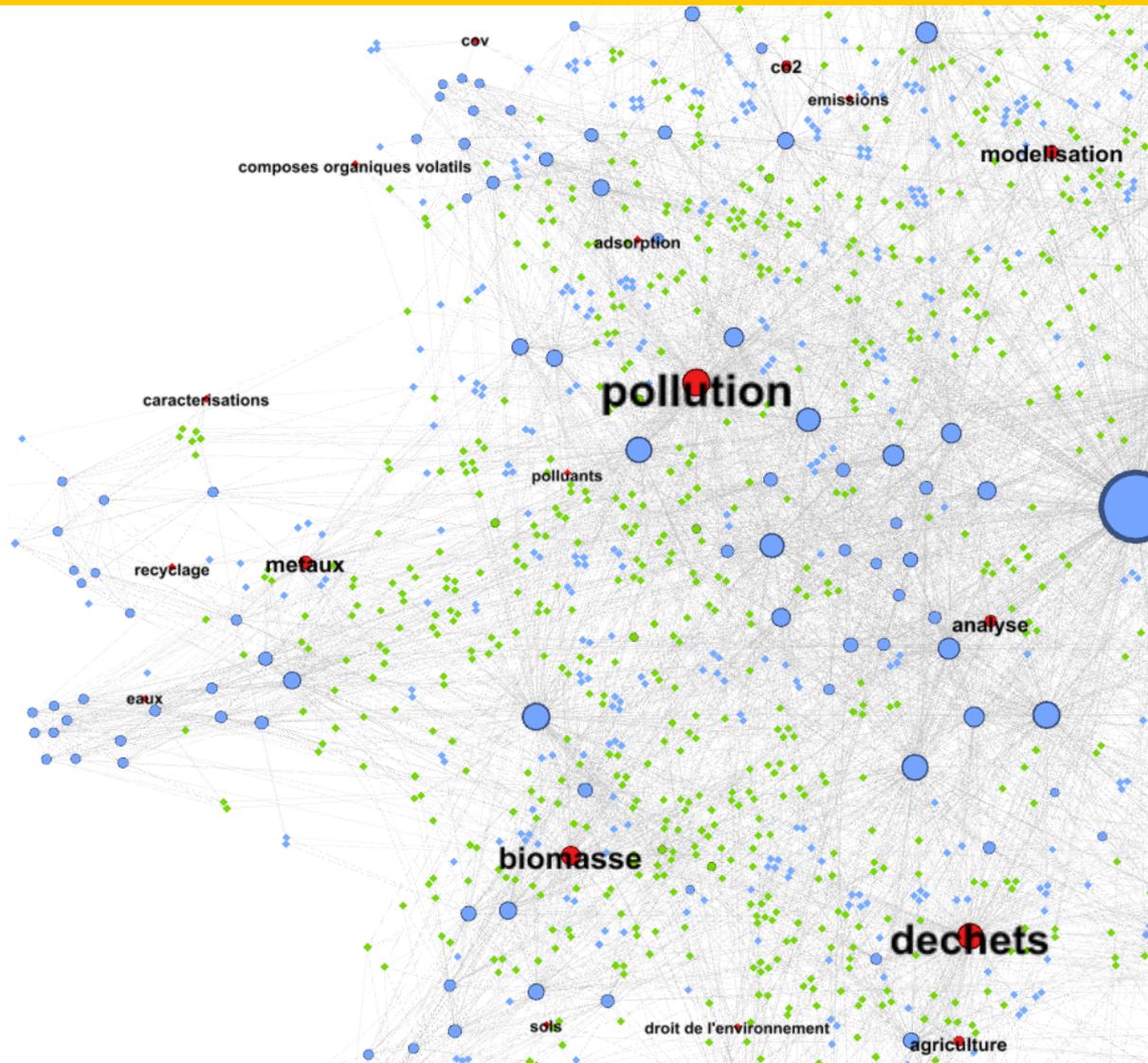
- RA5

Bilan

Projets

Résultats : zoom sur la communauté « pollution »

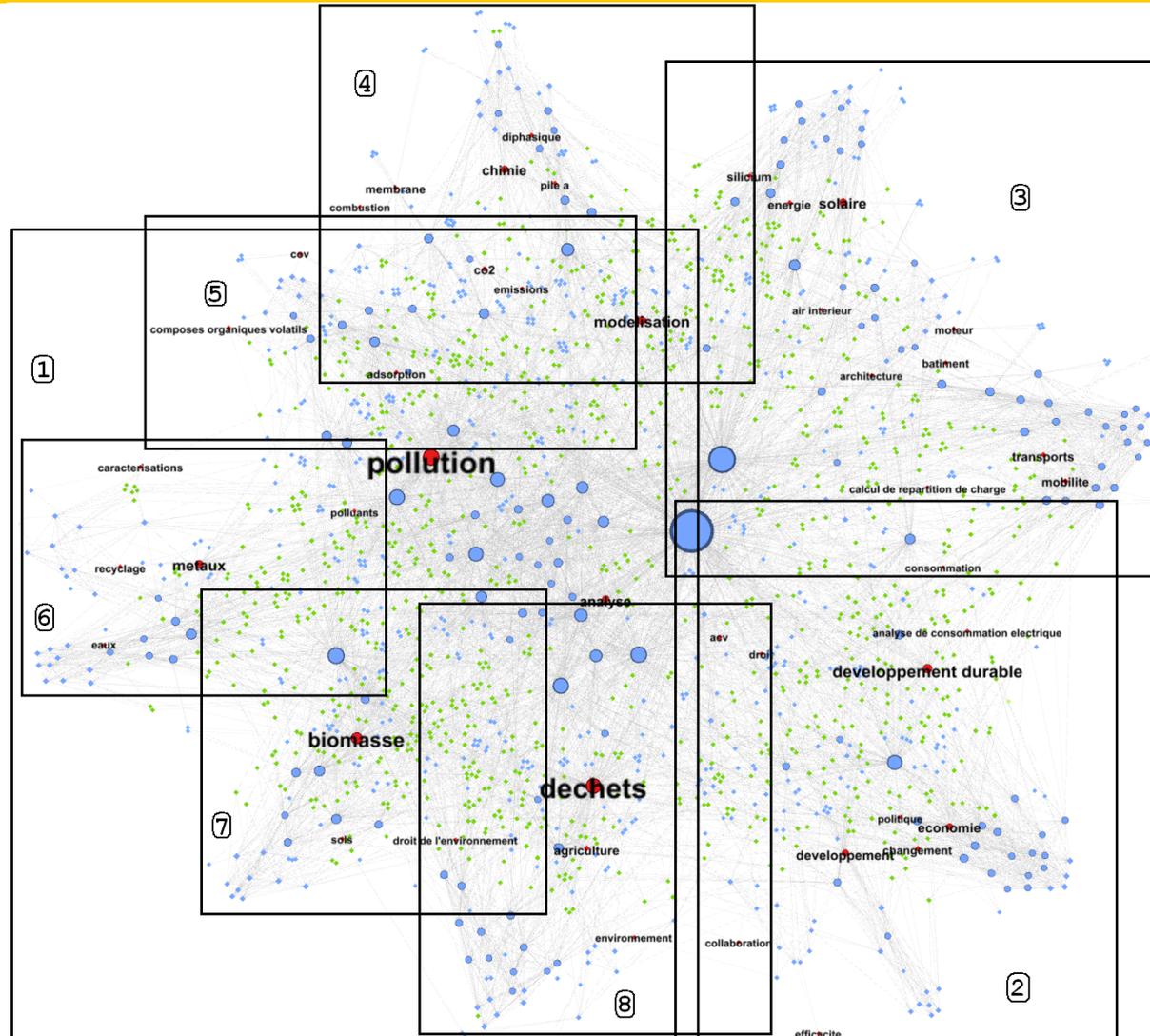
Présentation
I3S 2011
pôle GLC



- Présentation
Recherche
- RA3
 - Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
 - RA1 + RA2
 - RA4
 - RA5
- Bilan
Projets

Résultats (autour des 8 plus « gros » tags)

- Présentation Recherche
- RA3
 - Contexte
 - Analyse
 - SNA
 - Communautés
 - Contrib.
 - RA1 + RA2
 - RA4
 - RA5
- Bilan
Projets



1. pollution
2. développement durable
3. énergie
4. chimie
5. pollution de l'air
6. métaux
7. biomasse
8. déchets

Comparaison des algorithmes (critère : modularité)

Présentation Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés

• Contrib.

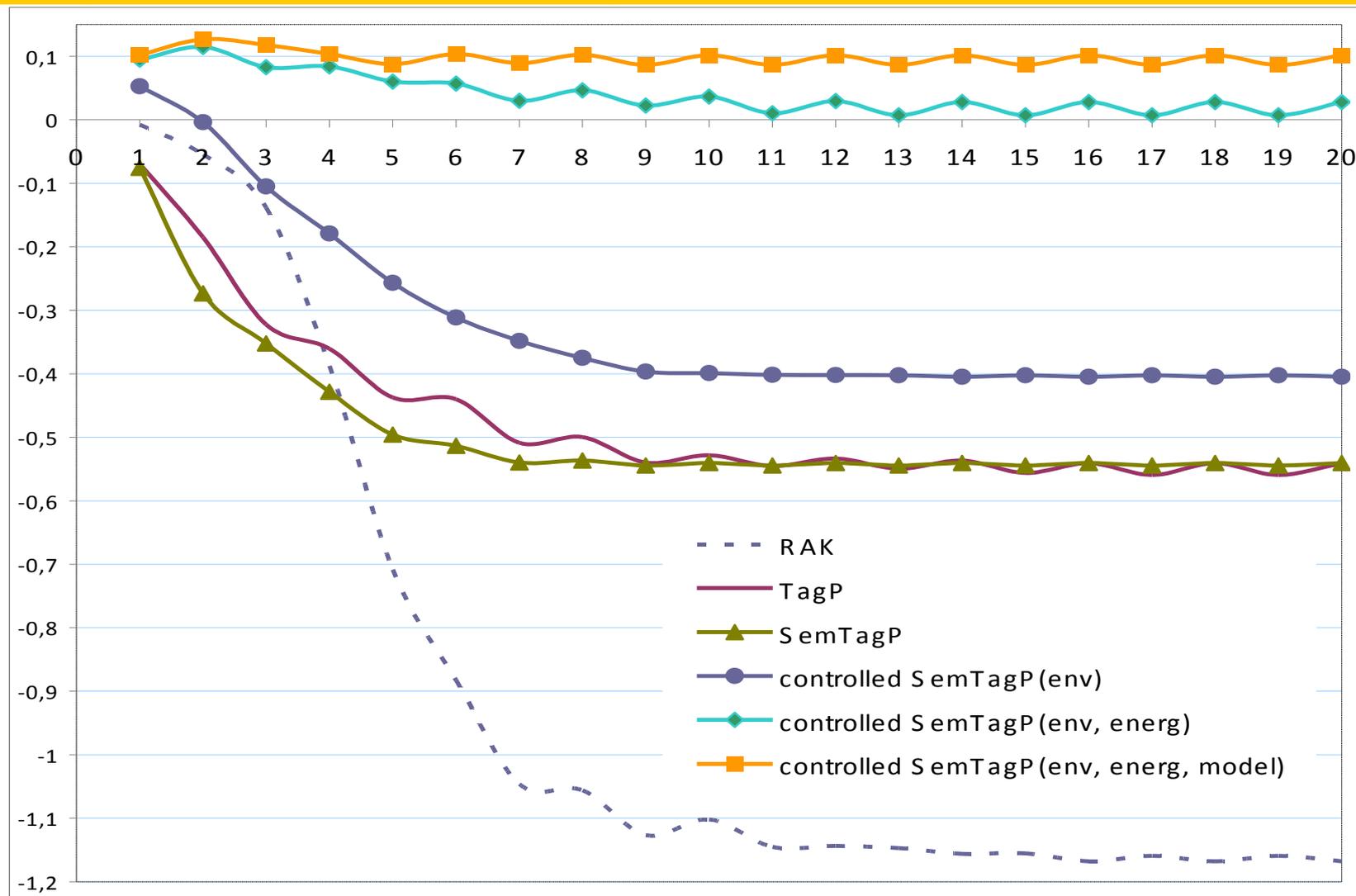
- RA1 + RA2

- RA4

- RA5

Bilan

Projets



- **Framework complet pour l'analyse des réseaux sociaux,**
 - Notamment : **ontologies SemSNA et SemSNI** pour organiser et structurer les données sociales
- Une approche pour **relier tagging et Web sémantique**
 - Un cycle de vie complet pour l'enrichissement de folksonomies
- **Un nouvel algorithme de détection de communautés :**
SemTagP
 - Testé sur le réseau TheseNet l'ADEME,
 - Qualité des communautés supérieure aux autres algorithmes basés sur LP/RAK (meilleure modularité)

Présentation
Recherche
- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2
- RA4
- RA5
Bilan
Projets

- Travaux complémentaires,
- **Implémentés et intégrés au framework ISICIL,**
- Collaboration forte avec :
 - INRIA, Ademe et Orange Labs pour l'analyse des usages et des besoins,
 - Ipernity.com pour la validation à grande échelle
- **Publications internationales** : Web Science 2009 & 2010, **ISWC 2009**, Web Intelligence 2009

Présentation
Recherche

- RA3

- Contexte
- Analyse
- SNA
- Communautés
- Contrib.

- RA1 + RA2

- RA4

- RA5

Bilan

Projets

RA 1 et RA 2

Intergiciels et composition logicielle

1. Informatique ambiante (Jean-Yves Tigli)
2. Modélisation d'infrastructures à grande échelle (Johan Montagnat)
3. Workflows scientifiques (Johan Montagnat)

1. Informatique ambiante

Intelligence Ambiante

↳ Informatique Ambiante

↳ Middlewares pour gérer l'adaptation des applications à l'évolution dynamique du **contexte**.



- Les **défis** autour de l'adaptation sont alors :
 - **Cohérence** de l'application
 - **Variabilité** et **Complexité** du contexte
 - **Dynamique** de l'adaptation pour des applications stables et utilisables (au moment opportun, avec un temps de réponse minimal)
 - ...
- Travailler à un niveau d'abstraction adapté

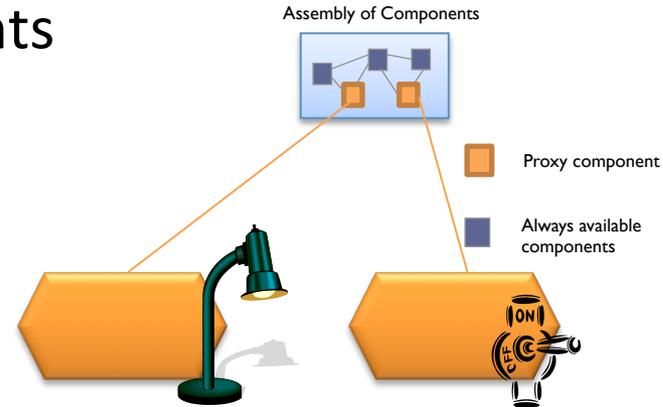
Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- Intergiciel et modèles de composition
 - Adaptation à l'exécution
 - Prise en compte des variations du contexte
- **Modèle de composition dynamique** par assemblage de composants et de services pour dispositifs
- **Modèle d'adaptations dynamiques et indépendantes** déclenchées par les variations du contexte
 - Garantir la cohérence de l'application
 - Maîtriser le temps de réaction aux variations du contexte

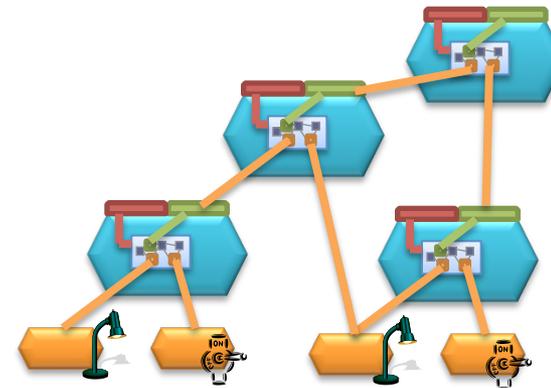
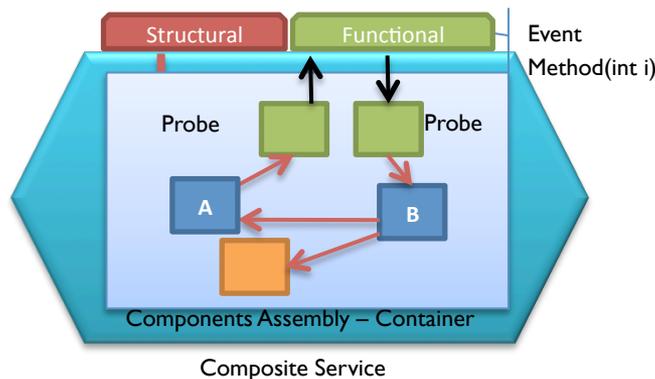
Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- Composition dynamique de Services pour Dispositifs par assemblages de composants (Modèle LCA)

- Flot d'événements
- Assemblages locaux de composants légers



- Composition de Services composites (modèle SLCA)

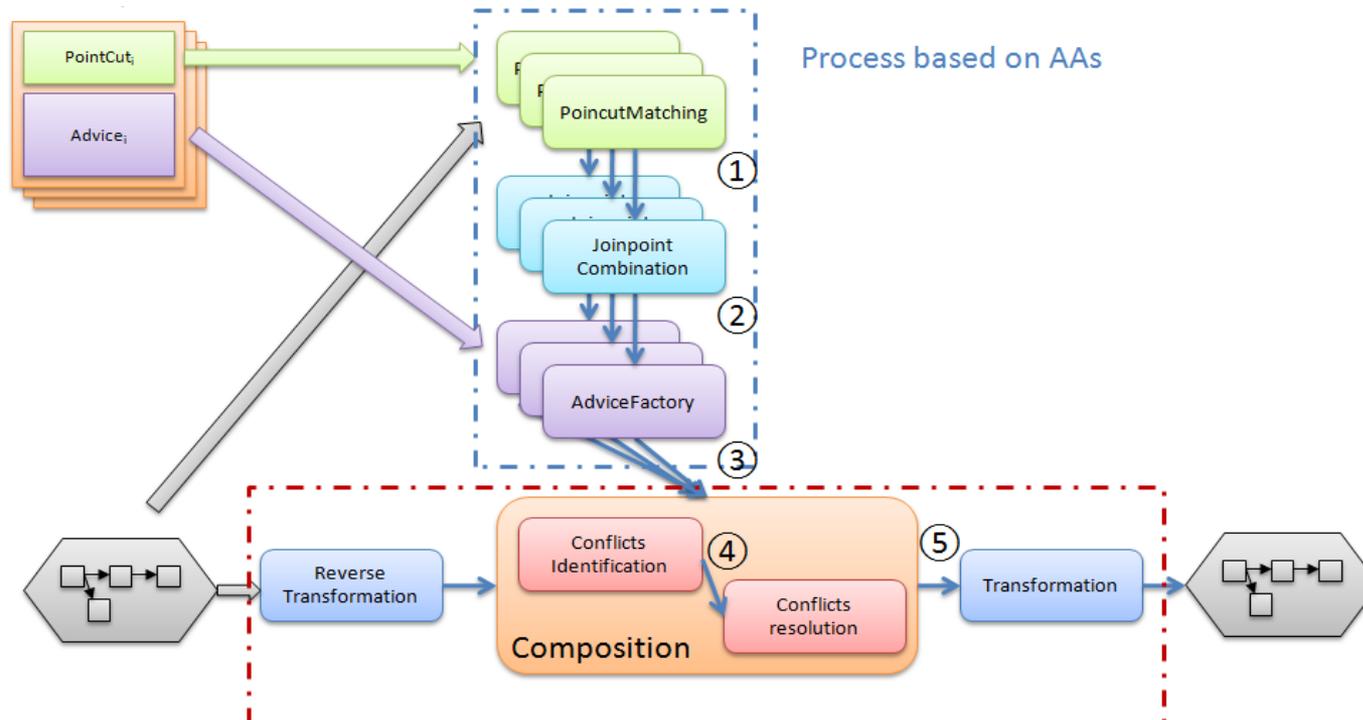


Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

Adaptations dynamiques et indépendantes : modèle des AA

- Indépendance : Aspect d'Assemblage (AA)
 - Aspects pour la modification d'assemblages de composants
- Dynamicité : Adaptation au contexte
 - Déclenchement du cycle de tissage d'AA sur modifications

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets



- Validation de propriétés logiques dans la gestion du tissage des AA :
 - Méthode : à partir de mécanismes de composition prouvés entre composants de sémantique connue :
 - Règles de fusion (Thèses D.Cheung et V. Hourdin)
 - Composition d'automates (Collab. A. Ressouche INRIA)
- Exemple :
 - Propriété de symétrie du tisseur

Commutativity : $AA0 \otimes AA1 = AA0 \otimes AA1$

Associativity : $(AA0 \otimes AA1) \otimes AA2 = AA0 \otimes (AA1 \otimes AA2)$

Idempotence : $AA0 \otimes AA0 = AA0$

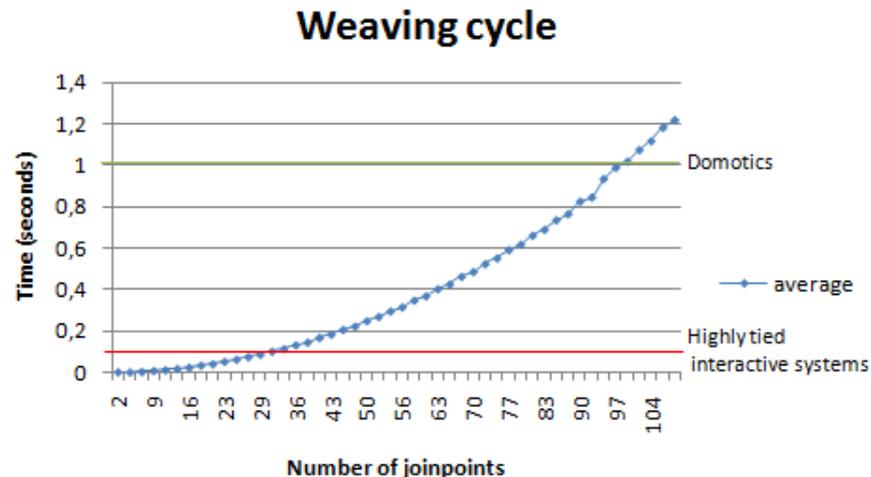
- Evaluation des performances dans l'adaptation dynamique sur modification

- de l'infrastructure
- de l'environnement

} Contexte

- Méthode :

- étude de la complexité des différentes étapes algorithmiques
- modèle et identification de paramètres dans l'analyse des performances



Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- Publications depuis 2009 (3 RI, 7 CI et WI, 2 RN)
- Nombreuses applications :
 - au bâtiment intelligent (Collab. CSTB)
 - à la ville intelligente (Collab. Asie)
 - aux travailleurs mobiles (Collab. Suez Environnement, Lyonnaise de eaux)
- Prototype Wcomp
 - Synergie entre formation d'ingénieurs et recherche dès 2003
 - URL : <http://www.wcomp.fr>
 - 200 000 lignes de code - 8 dépôts APP dont 3 depuis 2009
 - 1 ingénieur sous contrat, 4 doctorants, de très nombreux stagiaires et 3 MdC
 - **En cours de transfert** (VALORPACA)

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- WComp, support expérimental :
 - **Thèses** [Cheung 2009, Hourdin 2010, Ferry, Fathallah, Vergoni, Hassan (en cours)]
 - **Projets ANR** Faros / ANR Continuum / STIC-Asia UbiFlood
 - **Collaborations nationales** (IRIT, LIG ...) et **internationales** (ENIT/Kairouan-Tunisie, UL-Liban, U. Cluj-Napoca-Roumanie, TMSI-NUS-Singapour ...)
 - **Collaborations industrielles** (GFI Informatique, MobileGov, ...) (Bourse CIFRE, Région-Entreprise)
 - **Formations** niveau Master (Ubinet, IFI, ...)
- Une popularité de WComp en grande partie due aux efforts pour le maintien d'une plateforme d'expérimentation opérationnelle

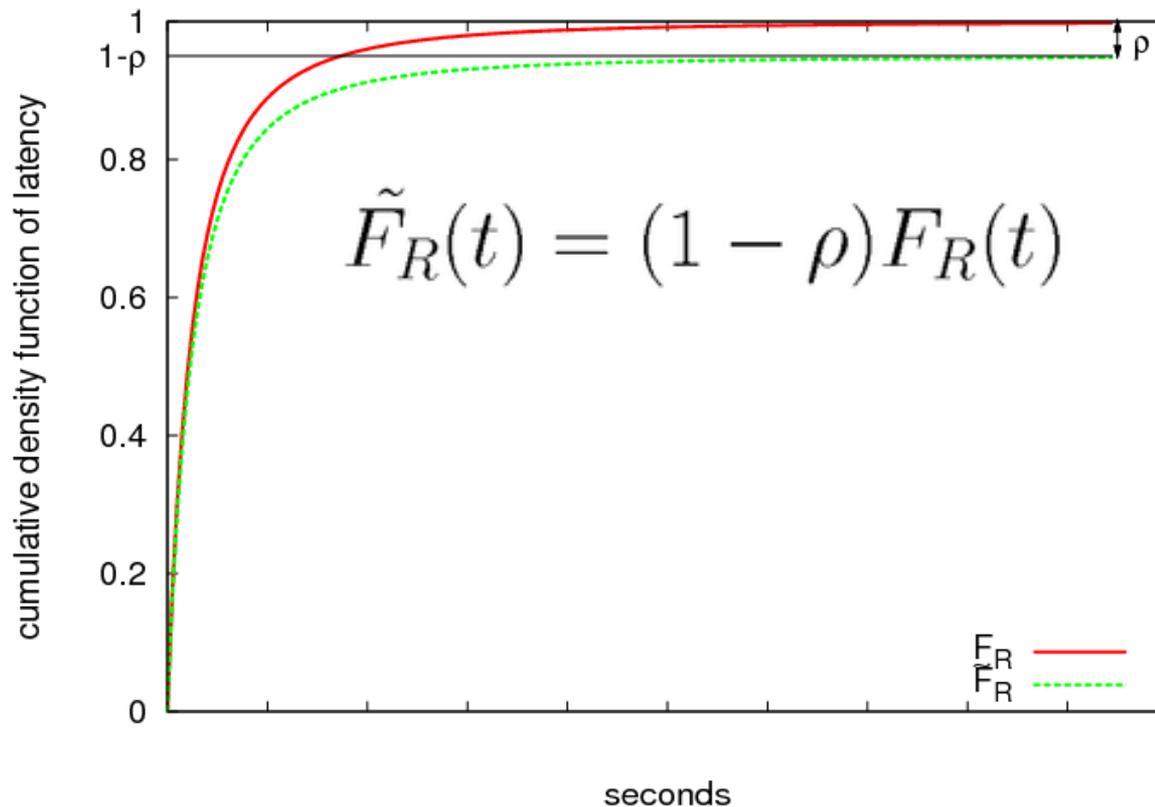
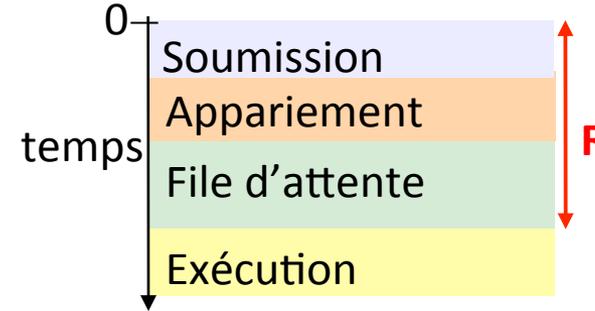
Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

2. Intergiciel et composition logicielle

- Modélisation des infrastructures sous charge: comprendre les systèmes distribués de grande taille
 - Infrastructures à grande échelle (non bornées) telles que les grilles : Modélisation probabiliste
 - Infrastructures de taille contrôlée telles que les Clouds (ou Grid'5000): Modélisation déterministe
- Flots de calculs scientifiques (workflows): modèle de programmation distribuée accessible
 - Langage orienté données, expression transparente du parallélisme
 - Composition formelle
 - Intégration de connaissances spécifiques au domaine

Modélisation probabiliste

- Tâches réussies et aberrations
 - Latence (R)
 - Loi de probabilité à queue lourde
 - Aberrations (ρ)



Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

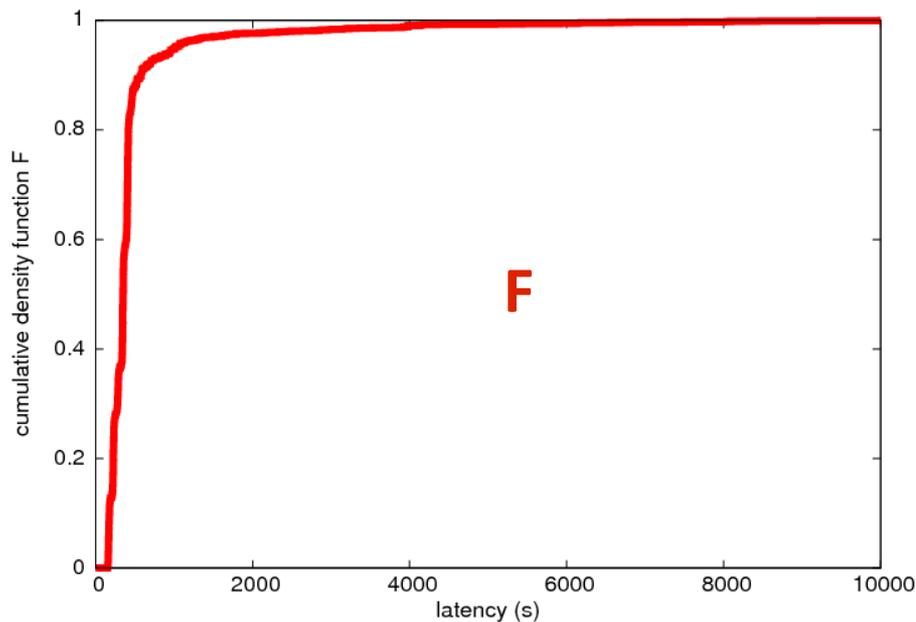
- Stratégie de re-soumission

- R = latence du système
- J = latence en prenant en compte les re-soumissions

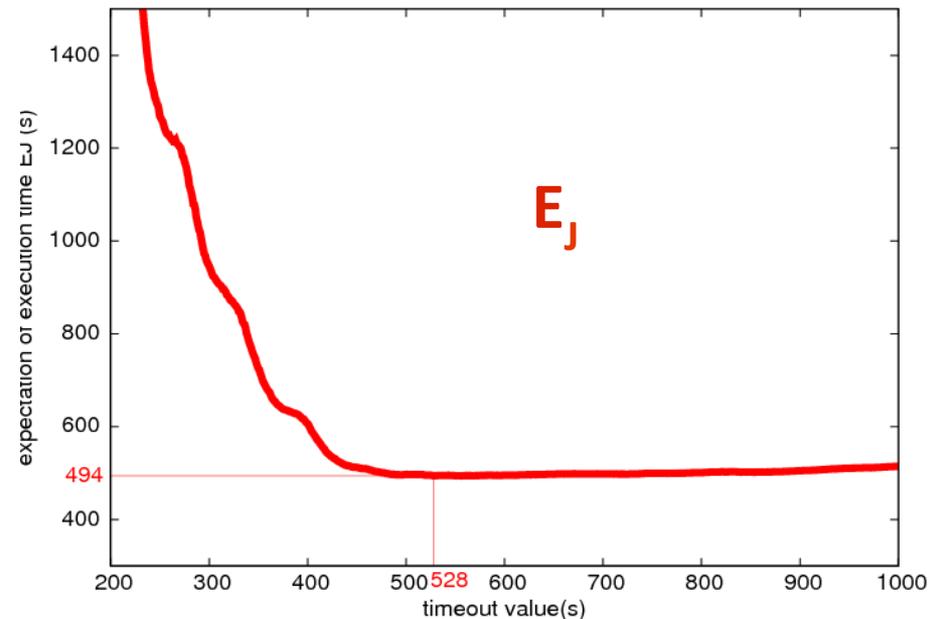
-> Espérance de la latence fonction du délai de re-soumission

$$E_J(t_\infty) = \frac{1}{F_R(t_\infty)} \int_0^{t_\infty} u f_R(u) du + \frac{t_\infty}{(1 - \rho) F_R(t_\infty)} - t_\infty$$

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows

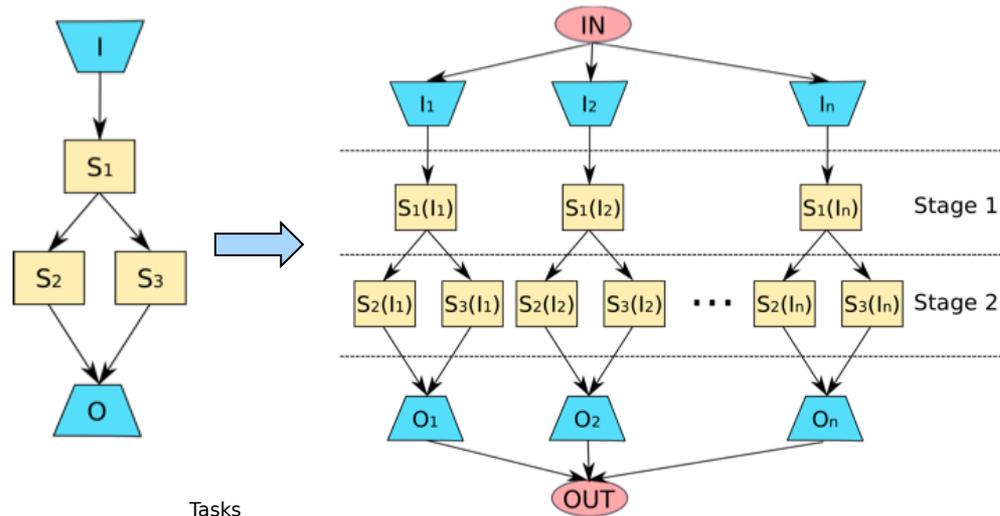


Valeurs de latence (s)

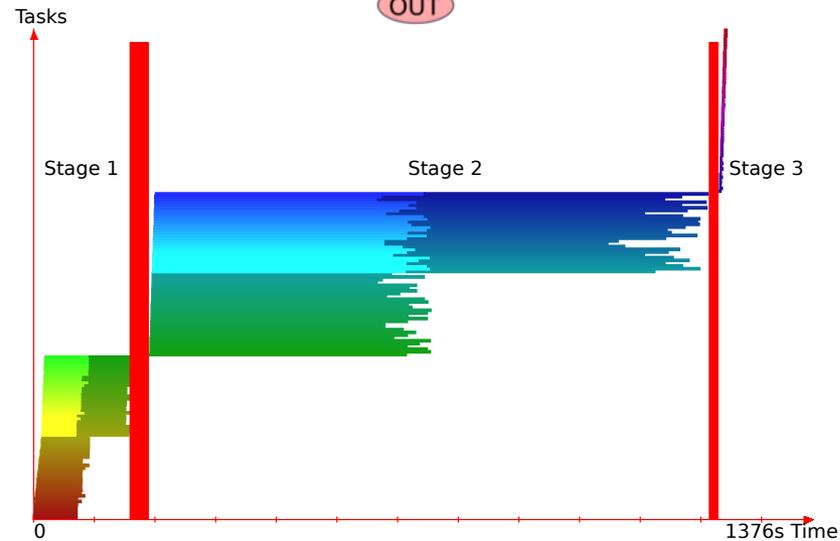


Délai de re-soumission (s)

- Plan d'exécution



- Allocation de ressources



- Bornes théoriques dans le cas de topologies simples

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- Journaux internationaux
 - JoGC(online), 2010. JoGC 8(2), 2010. ParCo 35(10-11), 2009. IJHPCA 20(1), 2006.
- Conférences internationales
 - Cloud'10, HPDC'09, CCGrid'09, CCGrid'08, CCGrid'07, HPDC'06
- Contrats, collaborations
 - EU EGEE, ANR HIPCAL, ANR NeuroLOG, PEPS GO, IdG-INRIASimGlite
 - Imperial College London (RTM), LRI (GridObservatory), CREATIS, LaBRI

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

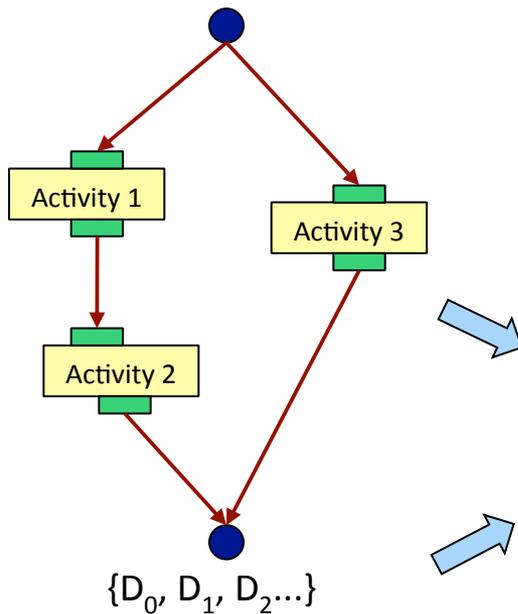
3. Flots de calcul scientifiques

- L'exploitation des systèmes distribués nécessite des modèles de programmation adaptés
 - Parallélisme de données (applications embarrassingly parallel, balayage de paramètres, SPMD, Map-Reduce...)
 - Performance élevée à un coût peu important
 - Faire face à la complexité du développement de logiciels distribués
 - Exploitation de sources de données distribuées et hétérogènes

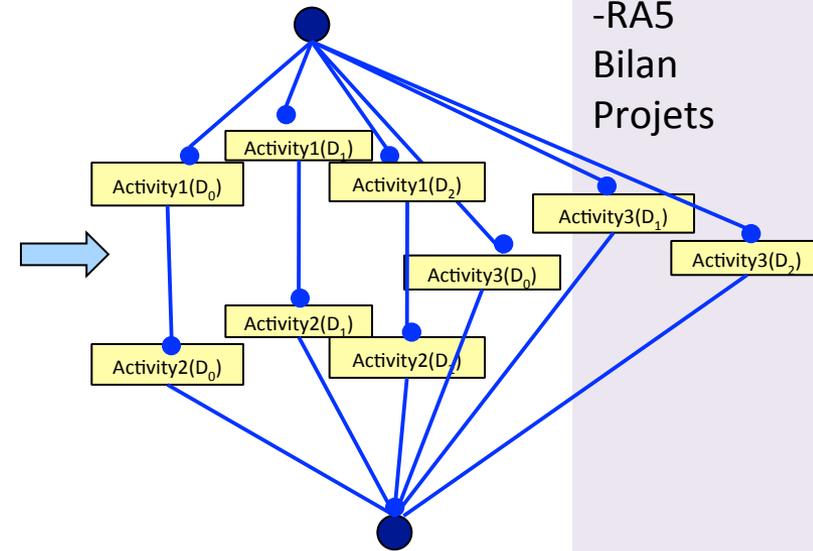
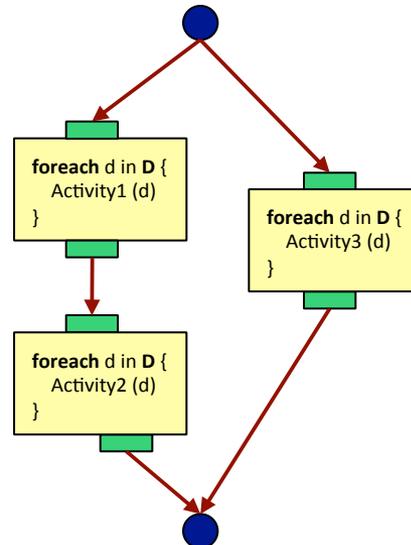
- Représentation abstraite compacte
 - Définition indépendante des flots de calcul et des jeux de données
- Génération dynamique des flots de données

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows

-RA4
-RA5
Bilan
Projets

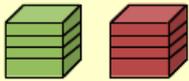


$D = \{D_0, D_1, D_2, \dots\}$ in:

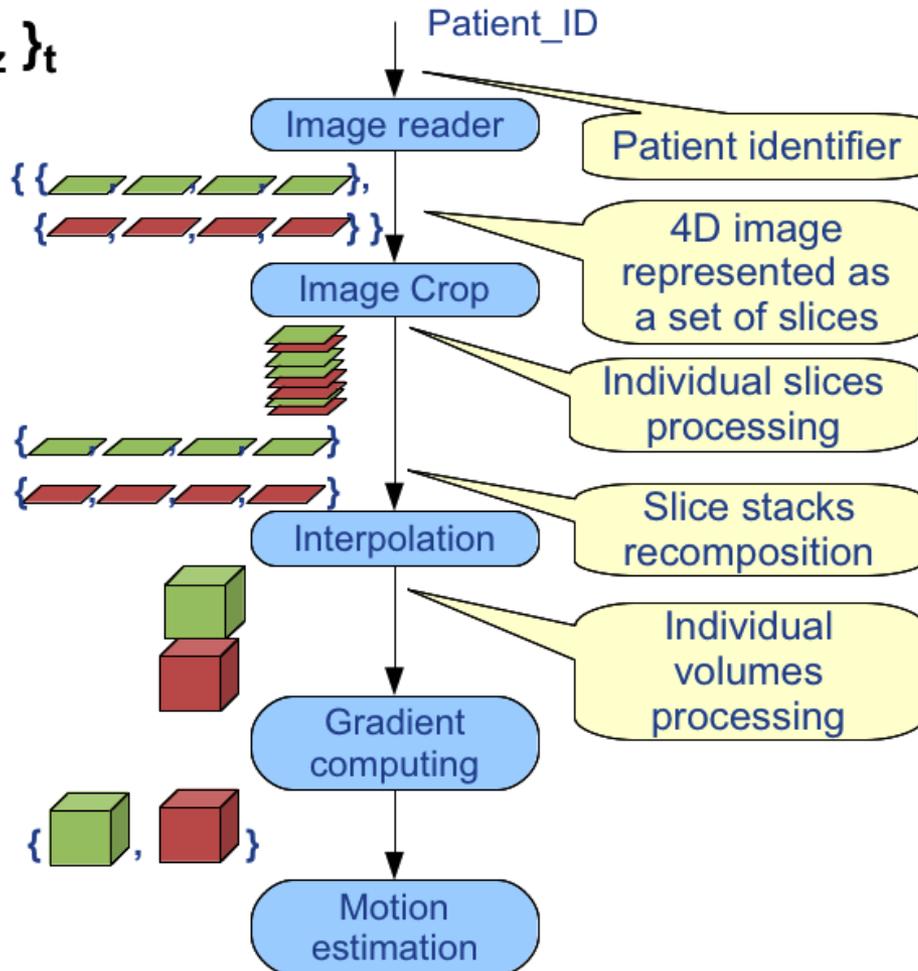


Structures de données centrées tableaux

- $4D = \{ 3D \}_t = \{ \{ 2D \}_z \}_t$



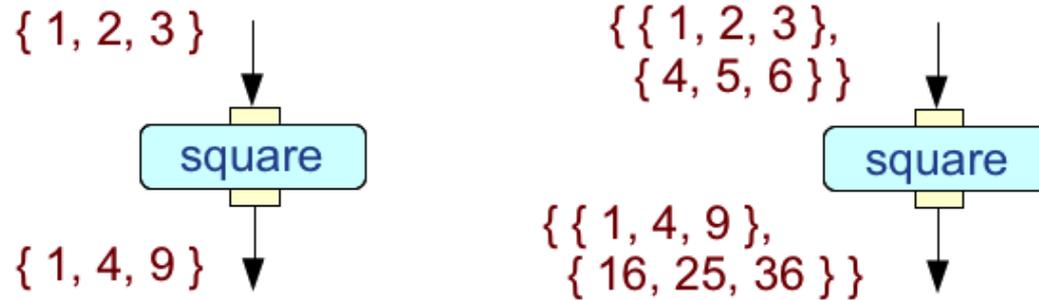
The input data set is a 4D image composed of 2 volumes (labelled green and red). Each volume is composed of 4 slices.



Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

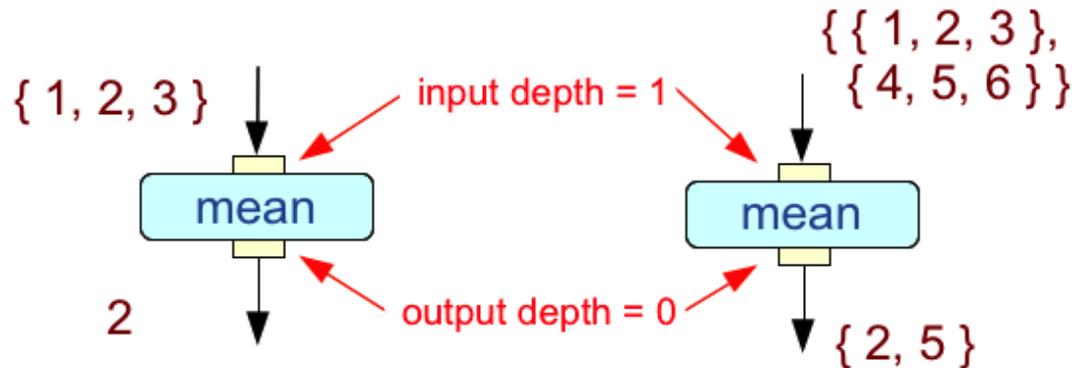
Tableaux imbriqués de profondeur unique

- Les structures de données sont des tableaux (imbriqués)

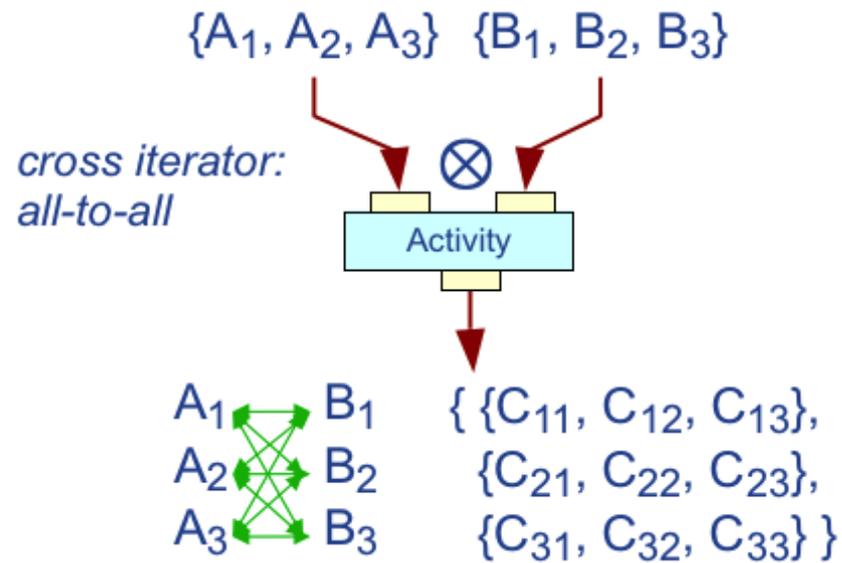


- A chaque port d'entrée/sortie est associée une profondeur

- Le niveau d'imbrication des tableaux est combiné à la profondeur des ports



- Exemple du cross-product

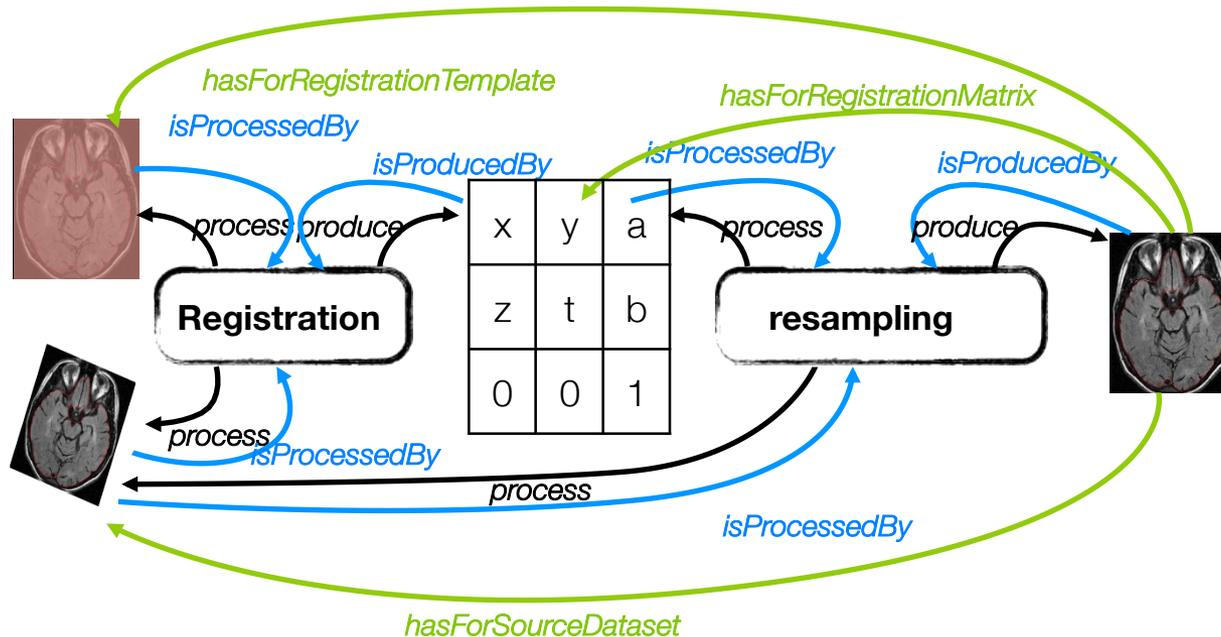


- Compatible avec les structures de contrôle
 - Contrôle orienté tableaux
 - Conditionnelles, boucles

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

Traitements enrichis par la sémantique

- Ontologie des jeux de données et des traitements
- Annotation des éléments de données (plateforme NeuroLOG)
- Paquetage d'outils annotés sémantiquement
- Production de connaissances au cours de l'exécution



Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

- Journaux internationaux:
 - SoSyM (online) 2010. JoGC 6(1), 2008. FGCS 24(7), 2008. IJHPCA 22 (3), 2008.
- Chapitres de livre
 - « Trans. on Aspect-Oriented Soft. Dev. », 200-233, LNCS 6210, 2010. « Workflows for e-Science », chap 18 (279-299), 2007.
- Conférences internationales:
 - **HTC**: WORKS'09, HPDC'06
 - **SOA**: ICWS'10 (WiP), MODELS'09, SLE'09, ECSA'08, MODELS'08, SCC'07, MODELS'07
 - **Applications**: HG'09, ICIP'08, MICCAI'06
- Contrats:
 - **Infrastructure**: EU SHIWA, EU EGEE
 - **Langage**: GWENDIA
 - **Applications**: ANR NeuroLOG
- Collaborations:
 - **Applications**: AMC(Amsterdam Medical Center), INRIA (IRISA, Sophia), INSERM (IFR49, GIN)...
 - **Systèmes de workflows**: U. Manchester (Taverna), USC ISI (Pegasus), U. Cardiff (Triana), U. Innsbruck (Askalon), MTA STZAKI (P-GRADE), U. Madison (CONDOR)

Présentation
Recherche
-RA3
-RA1 + RA2
• Info. Amb.
• Modélisation
• Workflows
-RA4
-RA5
Bilan
Projets

RA 2

Composition logicielle

C 1 : composition de services

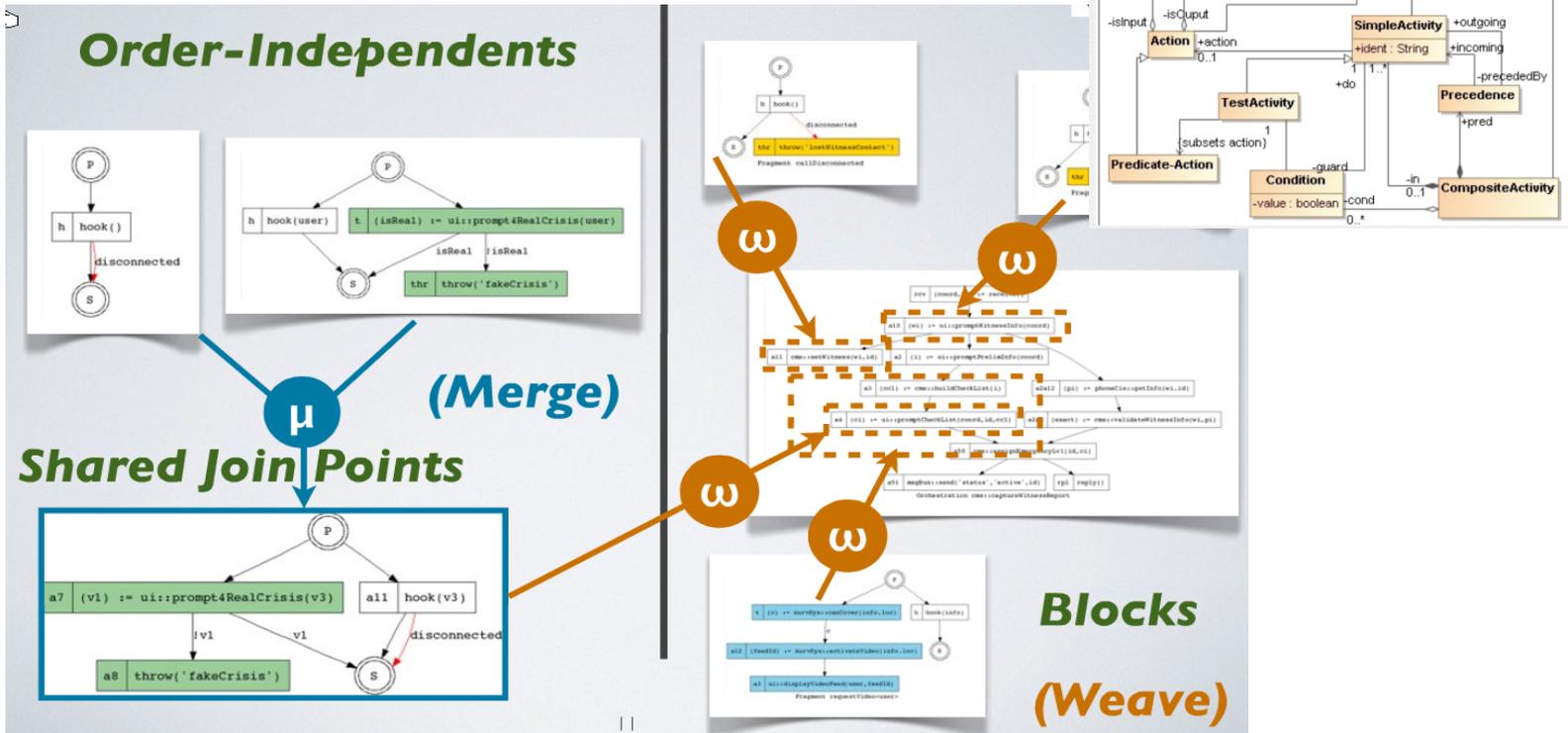
C 2 : composition d'IHM

C1. Software composition: Overall Approach

- Using model-driven techniques to provide sound models and supporting tools to tame its complexity
 - in a distributed context:
 - Service and Component Architectures
 - (Grid) Middlewares
 - on a large scale:
 - Software production through factories
 - Trade-offs between flexibility and reliability

Software Composition: On Orchestrations

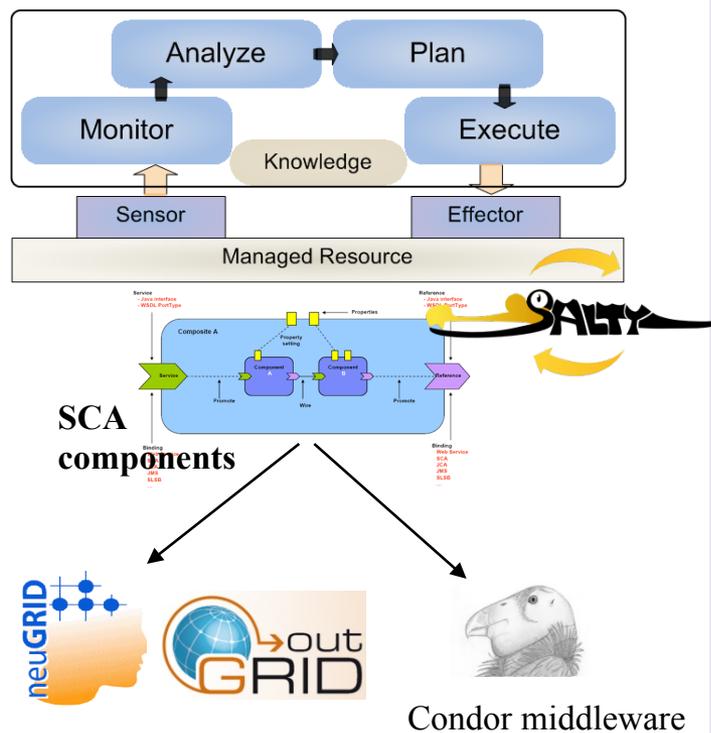
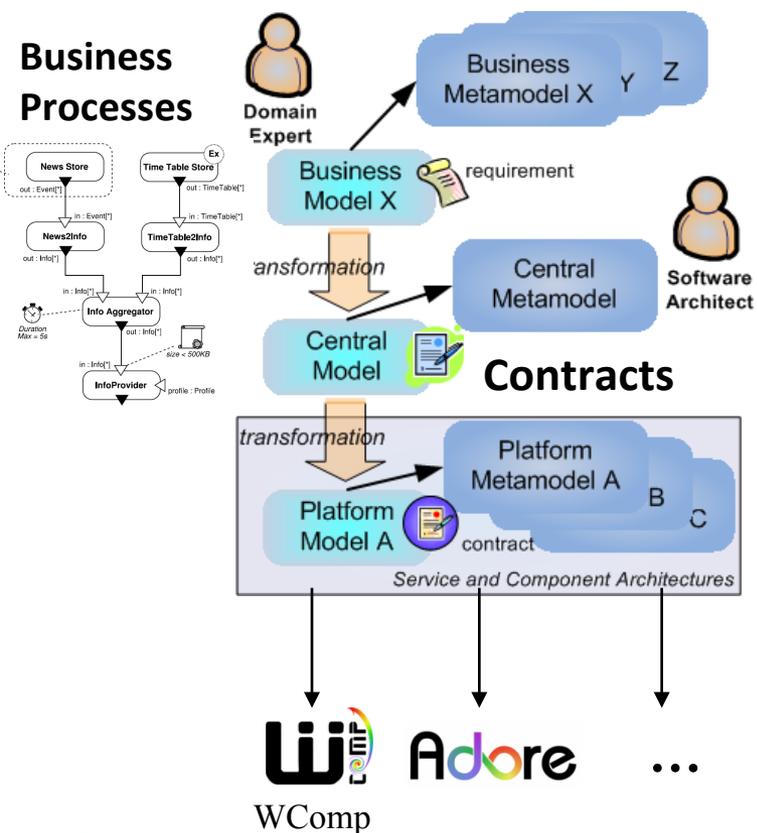
- Formal models and reasoning:



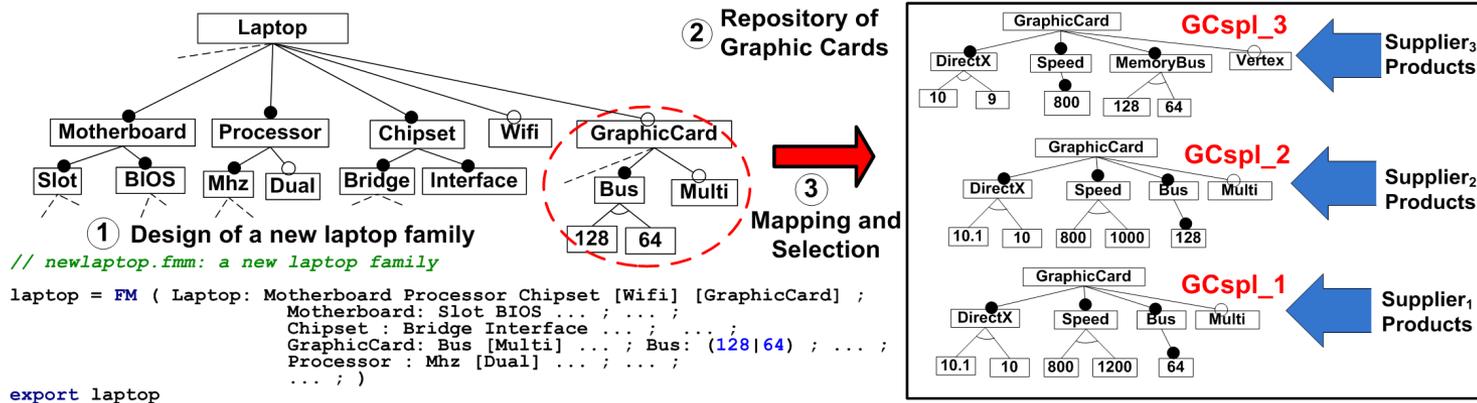
- Applications:
 - Business processes: Merging concerns while detecting conflicts
 - Data-driven (scientific) workflow : Merging overlapping workflow

Software Composition: On Contracts & Self-adaptation

- Model-driven Engineering Techniques for
- Contracts on Service and Component Architectures
- Autonomic Control Loops on large scale grid middlewares



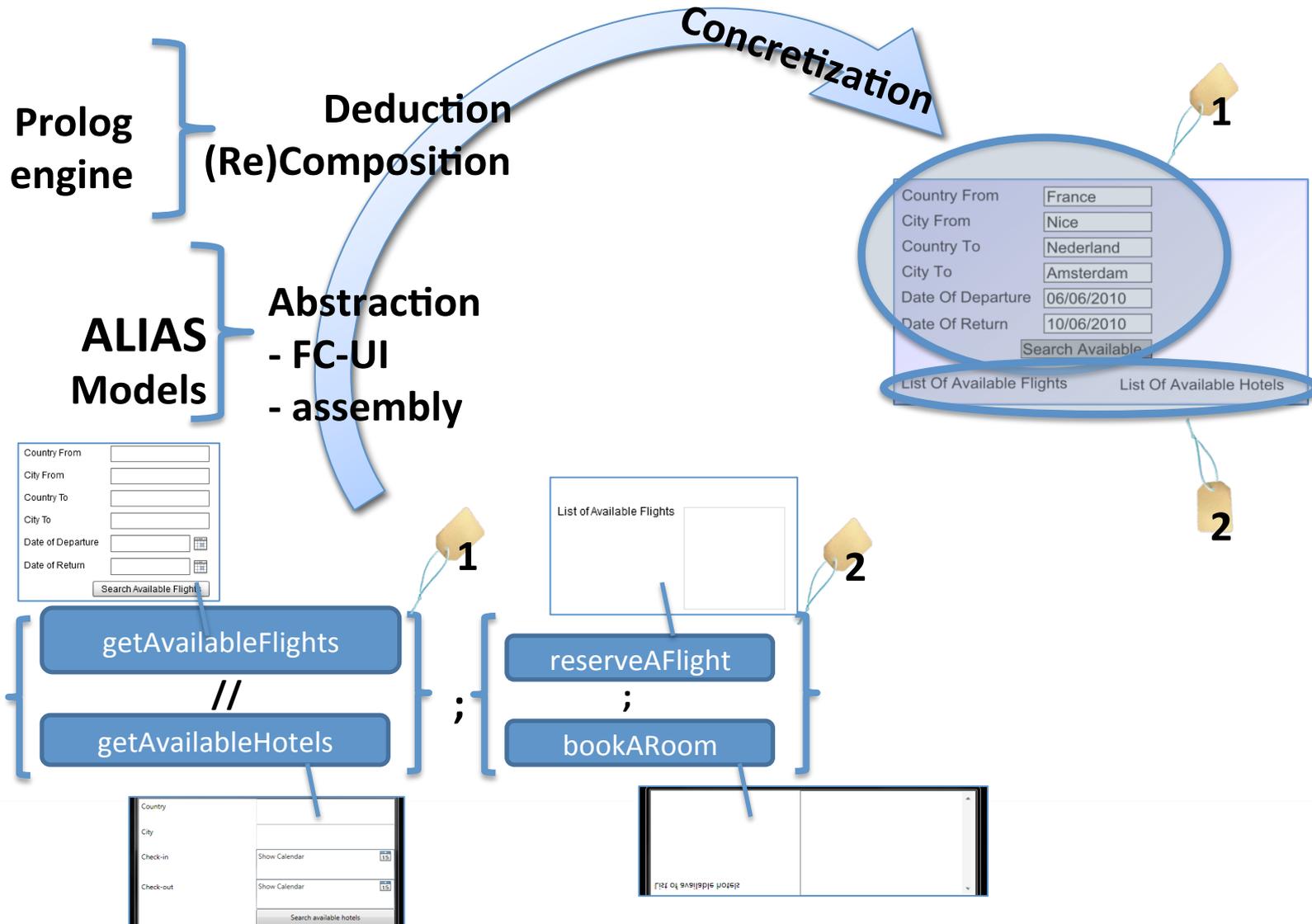
- Model-Driven Techniques and Languages for Large scale handling and merging of Feature Models
 - With sound formal basis



- Applications:
 - Scientific workflow incremental consistency checking
 - Management of very large multiple service product lines (Imaging services on the grid)
 - Versatile large scale automated manipulation of product lines

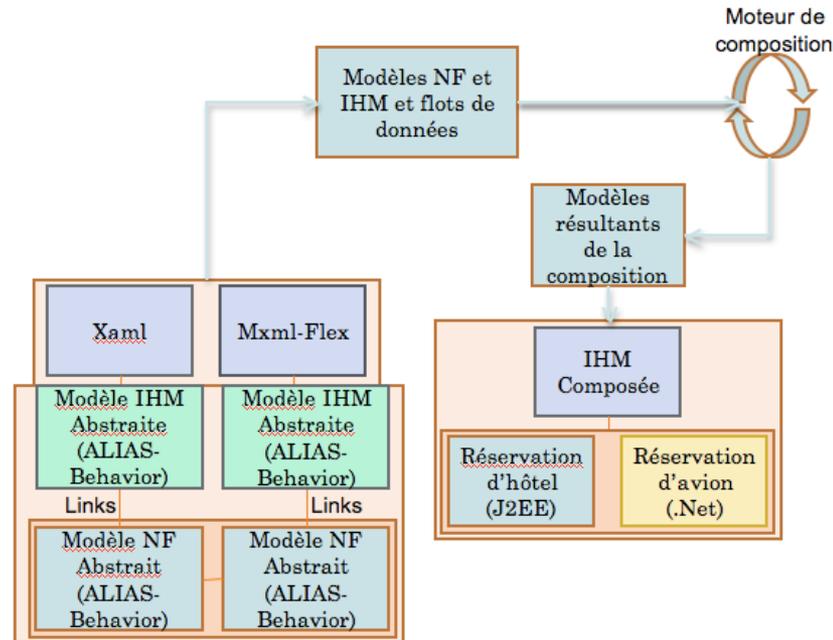
- Contracts & Self-Adaptation
 - Intl journals : JSW 2(2), 2007
 - Intl conferences: MODELS'07 & '08, IAWTIC'08, SOFSEM'07, ECBS'07, SC'06 & '07, ATC'06
 - Contracts and collaborations:
 - France Télécom, ANR FAROS, ANR SALTY
 - Internal partnership on platforms and “composition”

C2. Composition des IHMs



- Minimiser la réécriture des IHMs utilisant la composition de services
 - Attention aux pertes d'information
 - Lors de la composition au niveau du Noyau Fonctionnel
 - Lors de la réification des interfaces utilisateur (hétérogénéité des technologies utilisées)
 - Comment
 - Flot de données dans Noyau Fonctionnel
 - Lien entre IHM et le NF
 - Identifiant d'un méta-modèle facilitant la composition
 - Application de règles de transformations
 - Composition
 - Hétérogénéité

- Modélisation → raisonnement → Transformation



- Contrats
 - Mpub, STM3
- Collaborations
 - ST Microélectroniques, PME locales

- RA3
- RA1 + RA2
- RA4
- RA5

Bilan
Projets

RA 4 Sécurité

- Activité transversale, **nécessaire** dans le cadre de collaborations industrielles avec diffusion de logiciel
 - En particulier lorsque “l'on vend” la propriété d'auto-adaptation
- Mais aussi des travaux spécifiques :
 - Construction **à la demande** de protocoles de sécurité **orientés propriétés**
 - Sécurité de 'bout-en-bout'
 - Pour des développeurs non spécialistes
 - Propriétés étudiées :
 - Confidentialité, intégrité, authenticité, contrôle d'accès, non-répudiation
 - A l'étude : disponibilité, respect de la vie privée

1

• Utilisation d'une approche à base de composants

2

• Sécurisation des protocoles de communication

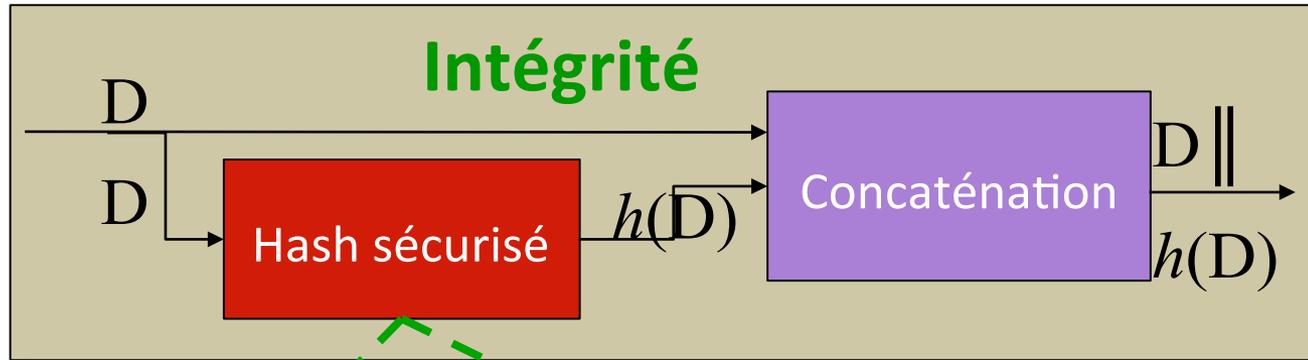
3

• Introduction de composants logiciels spécifiques

4

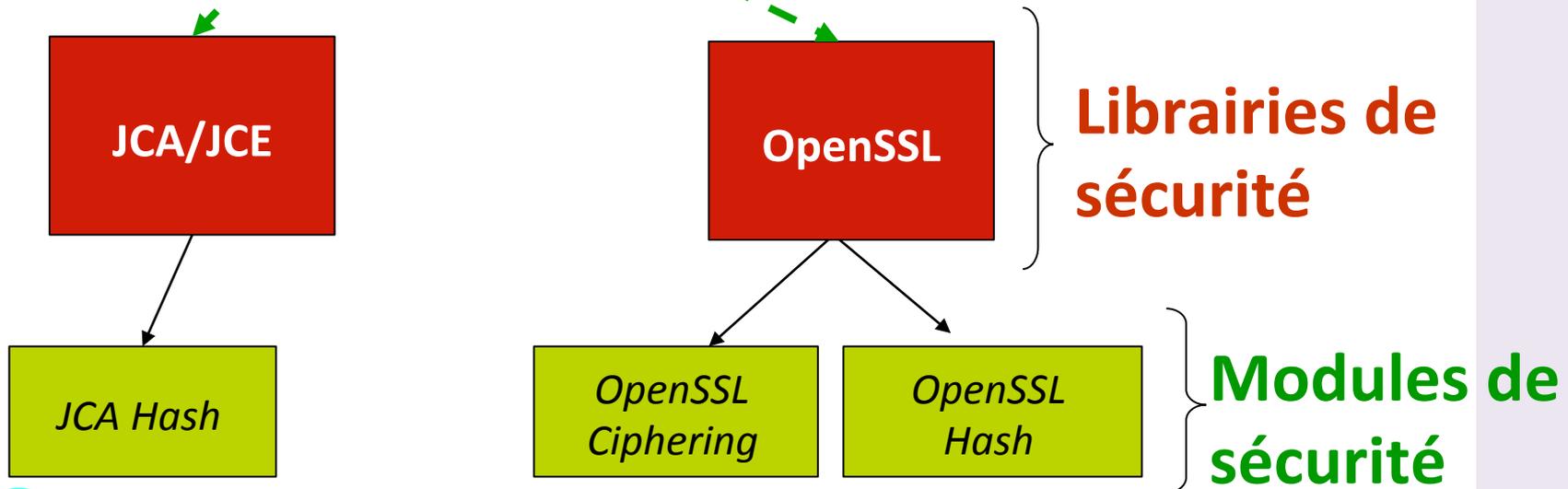
• Réutilisation de composants de sécurité

Présentation
Recherche
- RA3
- RA1 + RA2
- RA4
• Sécurité
- RA5
Bilan
Projets



Propriété
de sécurité

Sélection dynamique/statique d'une
bibliothèque cryptographique



- Travaux originaux
 - Il est possible de construire par composition de nouveaux protocoles ayant les propriétés désirées
 - Flexibilité, généricité, dynamicité
 - Réutilisation des composants / bibliothèques de sécurité disponibles
 - Expérimentation et validation basée sur différents modèles à composants : Wcomp, SCA, Fractal
 - 1 thèse soutenue (Région – entreprise)
- Projets de recherche
 - ACI Sécurité MetroSec, Projet Européen FastMatch ,
 - Aspects sécurité dans d'autres projets : ANR ISICIL, NeuroLog, Continuum et FUI M-PUB, STM3
- Publications
 - NOMS'2008, Mobility'2009, HealthGrid'09, EGC'2010, ASIGE'2010

Présentation
Recherche
- RA3
- RA1 + RA2
- RA4

- Sécurité

- RA5
Bilan
Projets

- RA3
- RA1 + RA2
- RA4
- RA5

RA5

Fouille de données

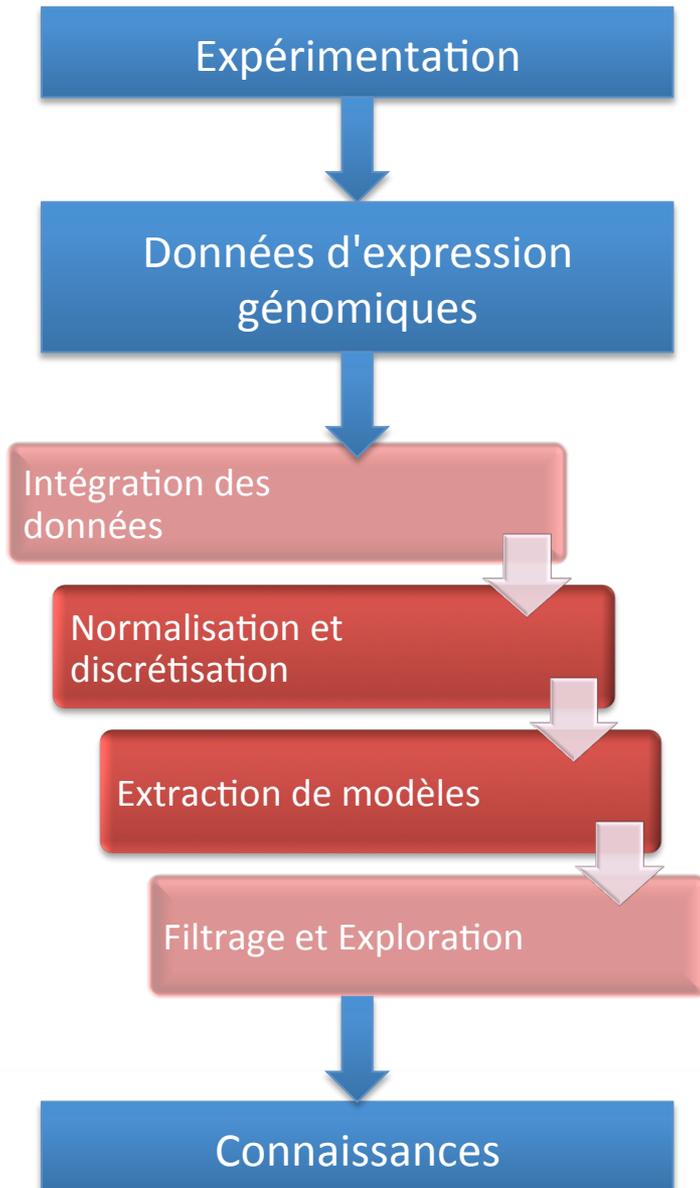
- **Activité qui se structure depuis janvier 2009**
 - **Essentiellement le travail de 2 personnes**
 - **- 1** promotion dans une autre université
 - Activité arrêtée : “gestion de la relation client”
 - **+ 2** en reconversion thématique
 - **+ 1** recrutement en septembre 2010
 - **Fouille sur des données biologiques issues :**
 - Des expérimentations de l'IBDC (UMR 6543 - Institute of Developmental Biology and Cancer)
 - Prédiction d'états cancéreux, construction de marqueurs
 - Des industriels du pôle PASS
 - Produire des solutions alternatives aux tests sur animaux

Présentation
Recherche

- RA3
- RA1 + RA2
- RA4
- RA5

• Fouille

Bilan
Projets



• Contributions

- Normalisation des données
 - En particulier dans la réponse des gènes (sur ou sous-exprimés)
- Extraction de modèles
 - Basés sur la théorie Gallois
 - Meilleure efficacité que d'autres approches
 - Le traitement assure aussi le filtrage
- Association des experts à chaque étape du processus de fouille

Présentation
Recherche
- RA3
- RA1 + RA2
- RA4
- RA5
• Fouille
Bilan
Projets

Associations de Type : *expressions* ⇔ *expressions*

Règle	Antécédent	Conséquent	Supp. (#)	Conf. (%)
1	heat4↓	heat3↓	83	79
2	cold4↓	heat3↓	66	68

règle 1

79 % des gènes sous-exprimés au 4^{ème} instant de mesure de l'expérience de choc thermique étaient déjà sous-exprimés au 3^{ème} instant de mesure

règle 2

Un ensemble de gènes sont sous-exprimés après à la fois un choc de chaleur et un choc au froid

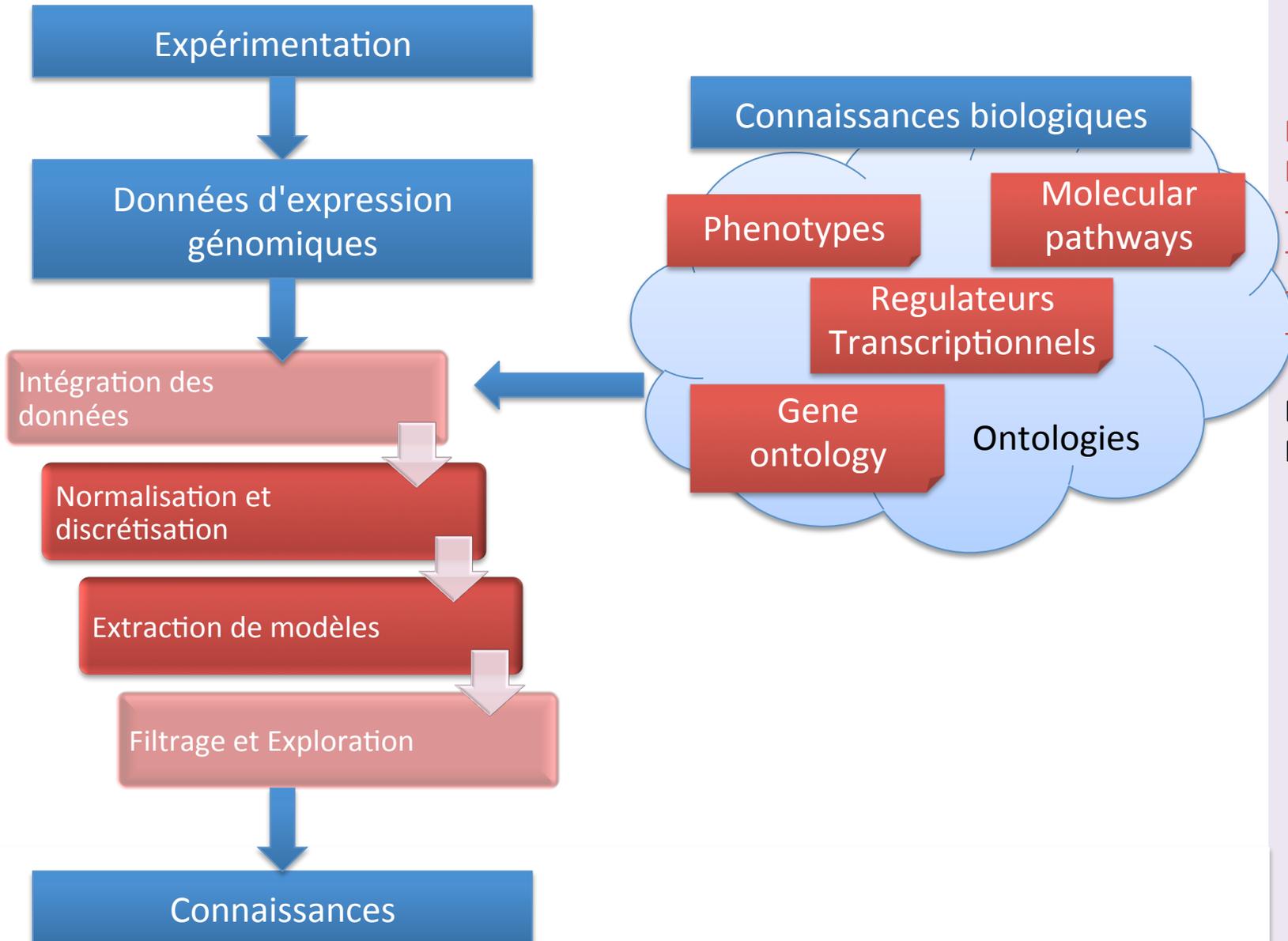
Présentation
Recherche

- RA3
- RA1 + RA2
- RA4
- RA5

• Fouille

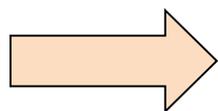
Bilan
Projets

Fouille de données biologiques



Associations de Type : *annotations* ⇔ *expressions*

Règle	Antécédent	Conséquent	Supp. (#)	Conf. (%)
1	Go:0006412 (translation) Go:0005840 (ribosome)	heat3↓	103	51
2	Go:0042254 (ribosome biogenesis and assembly) Go:0005840 (ribosome) Go:0005198 (structural molecule activity)	heat4↓	15	52
3	path:sce03010 (ribosome pathway)	heat4↓	69	53
4	tr:RAP1 tr:FHL1	heat3↓	71	62
5	pmid:5542014, pmid:9649613, pmid:3533916	heat3↓	12	100



Réduction de la synthèse de protéines, l'organisation ribosomique et la maintenance des cellules suivant un choc thermique

Présentation
Recherche
- RA3
- RA1 + RA2
- RA4
- RA5
• Fouille
Bilan
Projets

Associations de Type : *expressions* ⇔ *annotations*

Règle	Antécédent	Conséquent	Supp. (#)	Conf. (%)
1	heat3↓, heat4↓, heat5↓	Go:0006412 (translation)	35	88
2	heat2↓	Go:0006996 (organelle organization and biogenesis)	41	69
3	heat2↓	Go:0042254 (ribosome biogenesis and assembly)	39	66
4	heat2↑, heat3↑, heat4↑	Go:0006950 (response to stress)	15	52

Présentation
Recherche

- RA3
- RA1 + RA2
- RA4
- RA5

• Fouille

Bilan
Projets

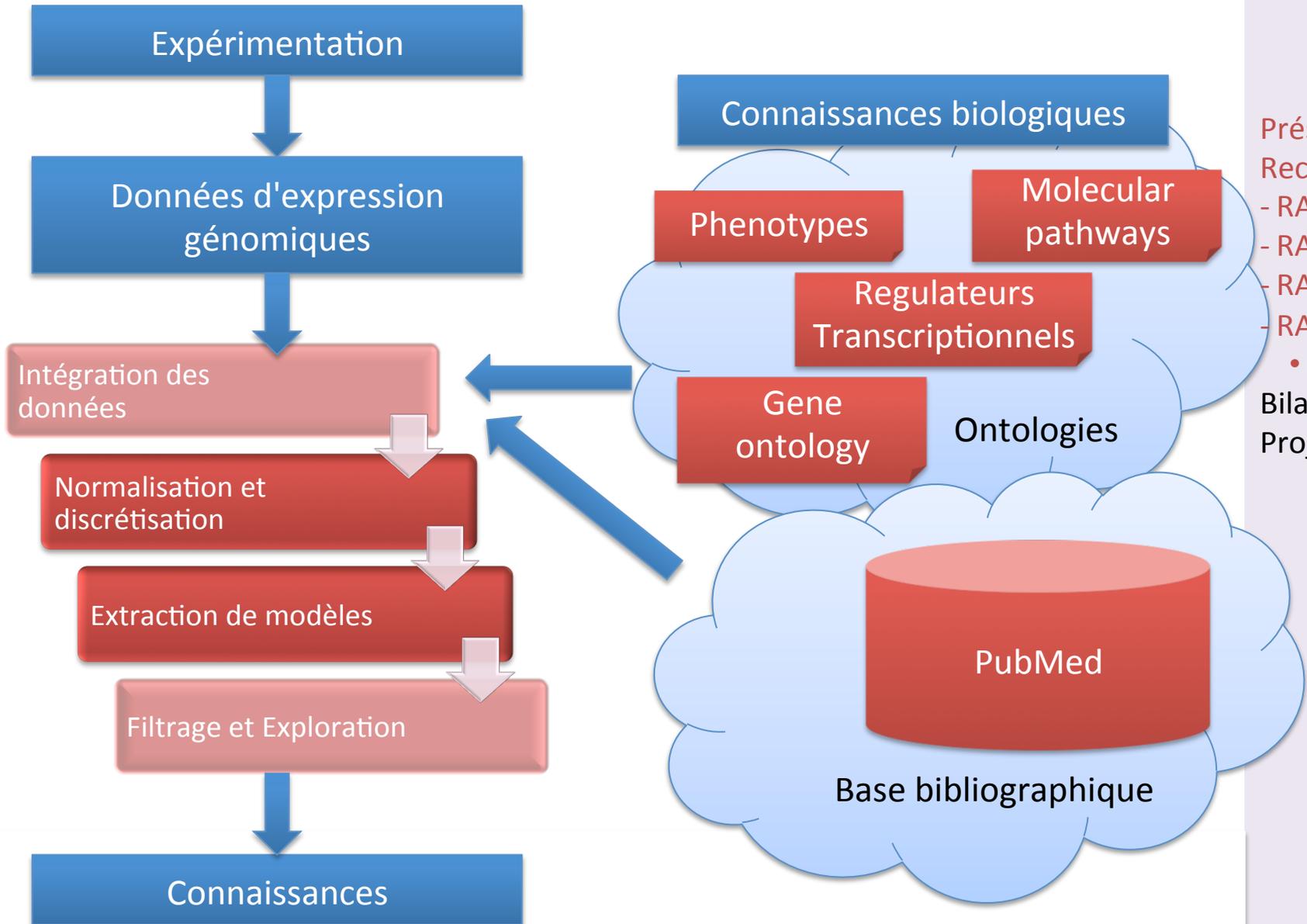
→ règles 1-3

Gènes sous-exprimés durant l'expérience de choc thermique sont impliqués dans la synthèse des protéines, l'organisation cellulaire et l'organisation ribosomique

→ règle 4

Gènes sur-exprimés durant l'expérience de choc thermique sont impliqués dans le processus de réponse au stress

Fouille de données biologiques



Associations de Type : *annotations* ⇔ *annotations*

Règle	Antécédent	Conséquent	Supp. (#)	Conf. (%)
1	Path:sce00190 (purine metabolism)	Go:0005737 (cytoplasm)	52	96
2	Pmid:16155567	Phenot:inviable	96	94
3	Tr:FHL1	Tr:RAP1	114	86
4	Tr:RAP1	Tr:FHL1	114	61

Présentation Recherche

- RA3
- RA1 + RA2
- RA4
- RA5

• Fouille

Bilan
Projets

→ règles 1-2

Révèlent des liens possibles entre annotations biologiques de différentes sources

→ règles 3-4

La liaison RAP1 est essentielle pour le recrutement de FHL1 (Identification de portions réseaux de régulation de gènes)

PMID:16155567: Davierwala et al. *The synthetic genetic interaction spectrum of essential genes*. Nat Genet. 2005 Oct. 37(10):1147-52. 2005 Sep 11.

- Mise en évidence de nouvelles relations
 - Issues des expérimentations
 - Mais aussi enfouies dans les bases de connaissances
 - A faire valider par des spécialistes
 - Collaboration avec l'IBMC, IPMC, industriels pôle Pass
- Logiciels
 - [NorDi](#) (*Normalized Discretization*)
 - [CGGA](#) (*Co-expressed Gene Groups Analysis*)
 - [GenMiner](#) (*Genomic data Miner* for extracting the equivalence classes)
- Publications
 - LNBI 2010 vol. 5488, CIBB'2010, CIBB'2008

Présentation
Recherche

- RA3
- RA1 + RA2
- RA4
- RA5

- Fouille

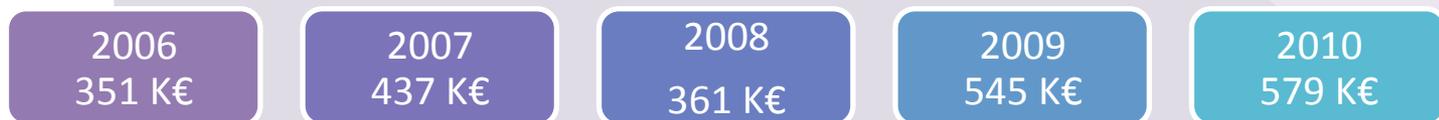
Bilan
Projets

Bilan général

- Logiciels : spécialité du pôle
 - Plus d'une dizaine de logiciels diffusés (500 Klines) dont certains sont largement utilisés en dehors du pôle
 - Wcomp, Moteur (RA1 + RA2)
 - jSeduite, Familiar (RA2)
 - SweetWiki (RA3)
- Publications : la qualité s'est notablement améliorée

	2006	2007	2008	2009	2010	Total	Moyenne par ETP sur 4 ans
RI	4	4	10	8	7	33	2.06
CI	30	27	34	32	37	160	10.00
Livres	10	8	5	10	6	39	2.44
Logiciels	7	8	5	4	1	25	1.56
ETP	11	12	13	14	14		

- Europe Fastmatch, Egee
- Industriels
 - Via pôle de compétitivité : Immunosearch, M-Pub, STM3
 - Directs : France Telecom + accompagnement bourses CIFRE / Région-entreprise
 - 2 créations de start-up : VuLog, SimplySim
 - Plusieurs accompagnements d'entreprises en incubation
- ACI Ergodyn, Metrosec, Agir
- ANR Mobivip, Faros, NeuroLOG, Gwendia, Hipcal, Isicil, Continuum, VIP, Salty
- Des ressources globalement en hausse, mieux réparties



Nécessaires

- pour le financement des doctorants
- pour la réalisation de plates-formes

« M-enaces O-pportunités F-orces F-aiiblesses »

	Pour	Contre
I N T E R N E	<p>Forces</p> <ul style="list-style-type: none"> ✓ Relations avec milieu industriel ✓ Fortement impliqué dans enseignement ✓ Proximité avec étudiants en Master (site polytech) ✓ Reprise d'une activité de recherche 	<p>Faiblesses</p> <ul style="list-style-type: none"> ✓ Coût dans le maintien des relations ✓ Coût des réalisations logicielles nécessaires à la validation des résultats ✓ Faiblesse en leader scientifique ✓ Faible nombre de chercheurs (1)
E X T E R N E	<p>Opportunités</p> <ul style="list-style-type: none"> ✓ Excellente relation avec INRIA (avec CNRS aussi) ✓ Insertion dans des communautés nationales (grille, informatique ambiante) ✓ Création campus STIC + nouveaux dépt ingénieur « bâtiment » ✓ Republication probable de nombreux postes de PR 	<p>Menaces</p> <ul style="list-style-type: none"> ✓ Surcharge de responsabilité administratives (3) ou pédagogiques ✓ Coût dans le montage/coordination des projets ✓ Nécessité d'être réactif pour maintenir relations industrielles

Projets

- **Génie logiciel (15 permanents)**

- Continuer ce qui marche pour mieux faire
- Maintien des 2 domaines d'application

1. Intergiciel

- Infrastructure à large échelle (ANR VIP, EU SHIVA)
- Informatique ambiante (ANR Continuum, CSTB, STIC Asia)

2. Composition logicielle (ANR Salty)

3. Interaction homme-machine (STM3)

4. Sécurité

- **Génie de la connaissance (9 permanents)**

- Renforcer la thématique : cadre scientifique

5. Web sémantique (ANR ISICIL, Emotica)

- Projet commun avec INRIA

6. Fouille de données (IBMC-IPMC, Gvt Monaco-Veolia)

1. Intergiciel
 1. Infrastructure à large échelle (grille et clouds)
 - Domaine principal : analyse d'image médicale
 - Activités :
 - Modélisation
 - Flots de calcul scientifiques
 2. Informatique ambiante
 - Domaine principal : bâtiment intelligent
2. Composition logicielle
 - Activité :
 - Contrôle des adaptations dans un système auto-adaptable
 - Composition de "feature modèles" pour architecture à large échelle
 - Réutilisation et composition de modèles sur étagères
3. Composition des Interactions
4. Sécurité
 - Sécuriser les échanges dans des environnements mobiles
 - Ajouter de la sécurité à une application non sécurisée via son IHM
5. Web sémantique
 - Représentation des connaissances
 - Modélisation des émotions
 - Modélisation, traitement et analyse des réseaux sociaux et des tags
6. Fouille de données (IBMC-IPMC, Gvt Monaco-Veolia)
 1. Processus de fouille
 2. Données biologiques
 3. Données médicales

1. Infrastructure à large échelle

- Modélisation
 - Modèle probabilistique pour architecture (non contrôlée) à très large échelle
 - Exploitation des traces collectées par l'observatoire de la grille
 - Evolution temporelle du comportement du système
 - Objectif : améliorer la précision des simulations
 - Infrastructure contrôlée (clouds)
 - Planification y compris si conditionnelles, boucles
 - Etendre le résultats sur la grille
- Flots de calcul scientifiques
 - Framework pour spécifier des flots extensibles (VIP)
 - Flots de calcul sémantiques (suite du projet NeuroLOG)
 - Exploiter la sémantique des données pour mieux les distribuer et les conserver
 - Interopérabilité des infrastructures au niveau des gestionnaires de flots (projet SHIWA)

2. Compositions de services et auto-adaptation

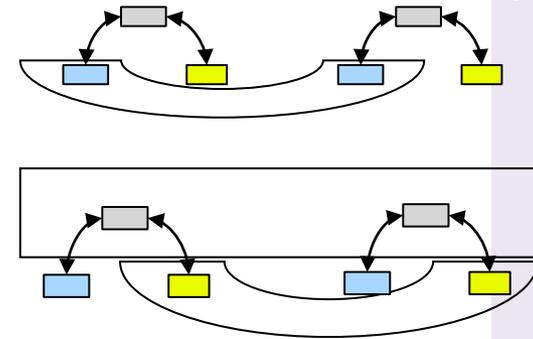
- Contrôle des adaptations dans un système auto-adaptable
 - Adaptation des boucles de contrôle, détection des adaptations potentiellement conflictuelles
 - Collaborations
 - gLite (ANR Salty) → NeuGRID / OutGRID
 - Condor (U. Wisconsin, clouds Amazone)
- Composition de “feature modèles” pour architecture à large échelle
 - Application aux workflows scientifiques et à la gestion de la composition dans les ligne de produits logicielles
 - Expérimentation sur services neurolog)
- Réutilisation et composition de modèles sur étagères
 - Collaboration avec R. France (U. Colorado State)

3. Informatique ambiante

- Adaptation logicielle, compromis entre autonomie et contrôle utilisateur :
 - Génération et sélection de schémas d'adaptations par l'utilisateur (U. Cluj Napoca, IIHM/LIG, U. Libanaise)
- Gestion de plusieurs niveaux de réactivité :
 - Mise en cascade des AAs pour en augmenter la modularité et augmenter la combinatoire des adaptations possibles (CSTB)
- Augmentation de l'expressivité des adaptations au contexte :
 - Prise en compte de l'environnement physique (CSTB, GFI)
 - Prise en compte d'informations sémantiques (HADAS-LIG)

4. Interaction Homme - Machine

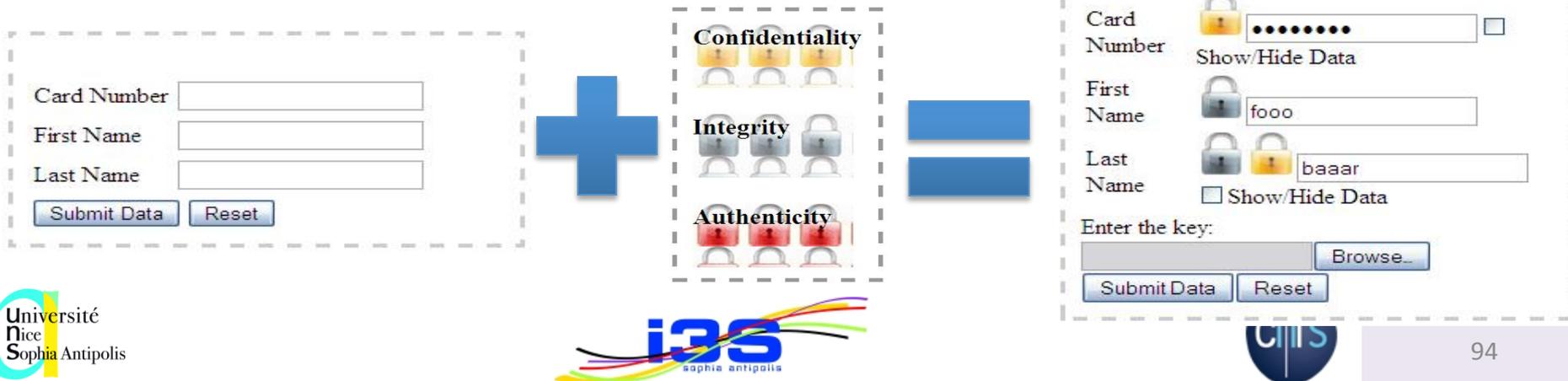
- Comment composer deux applications ou services en respectant l'architecture "Arche / MVC"
 - Composition services fonctionnels → composition IHMs
 - Composition interfaces → composition services fonctionnels
- Pourquoi ?
 - Adaptation :
 - nouveaux usages,
 - nouveaux besoins,
 - nouveaux dispositifs
 - Composition demandée par l'utilisateur
- Méthode
 1. Réification Interface / noyau fonctionnel
 2. Composition (composition de modèles)
 3. Transformation (cf. travaux en IHM/IDM)
- Collaboration



- ST-Microélectronique, Alcatel-Lucent + PMEs

5. Sécurité

- Sécuriser les échanges dans des environnements mobiles
 - Assurer juste le niveau désiré → préservation des ressources (mémoire, CPU, batterie)
 - Protocoles de sécurité standards (SSL, HTTPS, XMPPS) consomment beaucoup de ressources
 - SSL offre toutes les propriétés même si toutes ne sont pas nécessaire
- Ajouter de la sécurité à une application non sécurisée (IHM + Fonctionnelle)



6. Web sémantique

- En relation avec les autres équipes
 - Prise en compte des aspects sémantiques dans workflow, IHM, fouille
- Représentation des connaissances
 - Modélisation des émotions
- Modélisation, traitement et analyse des réseaux sociaux et des tags
 - Prise en compte du cycle de vie :
 - Nouveaux types de relations, relations obsolètes
 - Construction d'un Framework pour le web sémantique
 - Développement d'applications (serveur)
 - Intégration dans "pages web" (utilisateurs finaux)
- Collaborations
 - Orange Labs, Emotica

- Processus de fouille
 - Intégration de connaissances formalisées ou non à chaque étape du processus de fouille
 - Pose le pb de la représentation de la connaissance
 - Utilisation d'algorithmes évolutionnaires
- Domaine d'application
 - Données biologiques
 - Collaborations avec IDBC : prédiction d'états cancéreux, construction de marqueurs
 - Collaboration avec industriels du pôle PASS (immunosearch – produire des solutions alternatives aux tests sur animaux)
 - Données hydrologiques
 - Prédiction du courant marin en complément d'une modélisation mathématique du phénomène (Gvt de Monaco)
 - Prédiction du comportement d'une nappe phréatique (Véolia)