Helping online communities to semantically enrich folksonomies

Freddy Limpens, Fabien Gandon

Edelweiss, INRIA Sophia Antipolis
{freddy.limpens, fabien.gandon}@inria.fr

Michel Buffa
Kewi, I3S, Université Nice-Sophia Antipolis

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

INRIA

i3s
sophia antipolis

Edelweiss

WebScience 2010, Raleigh, NC, USA, 26-27 april 2010

Social tagging for bookmarking

http://delicious.com

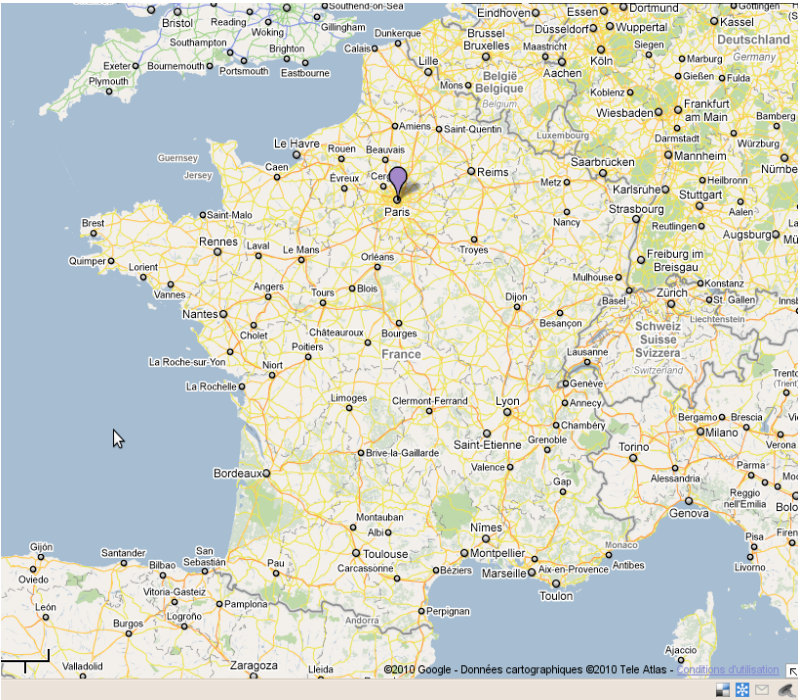# … and the resulting

FOLKS - ONOMIES

Spelling variations of tags
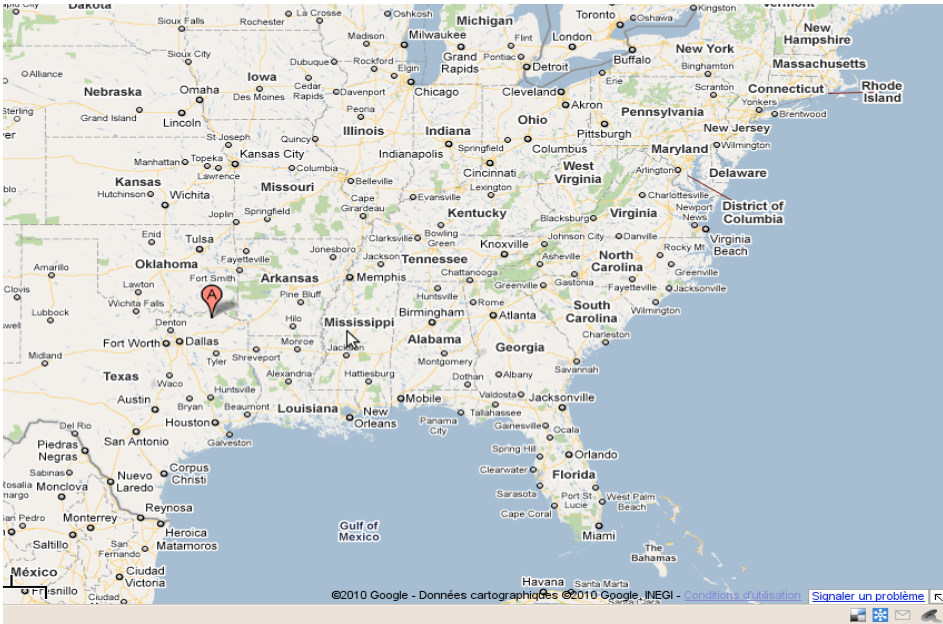
newyork = new_york = nyc

# **Amibiguity** of tags

paris

.. in France ?

… or in Texas ?

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   ecommerce   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
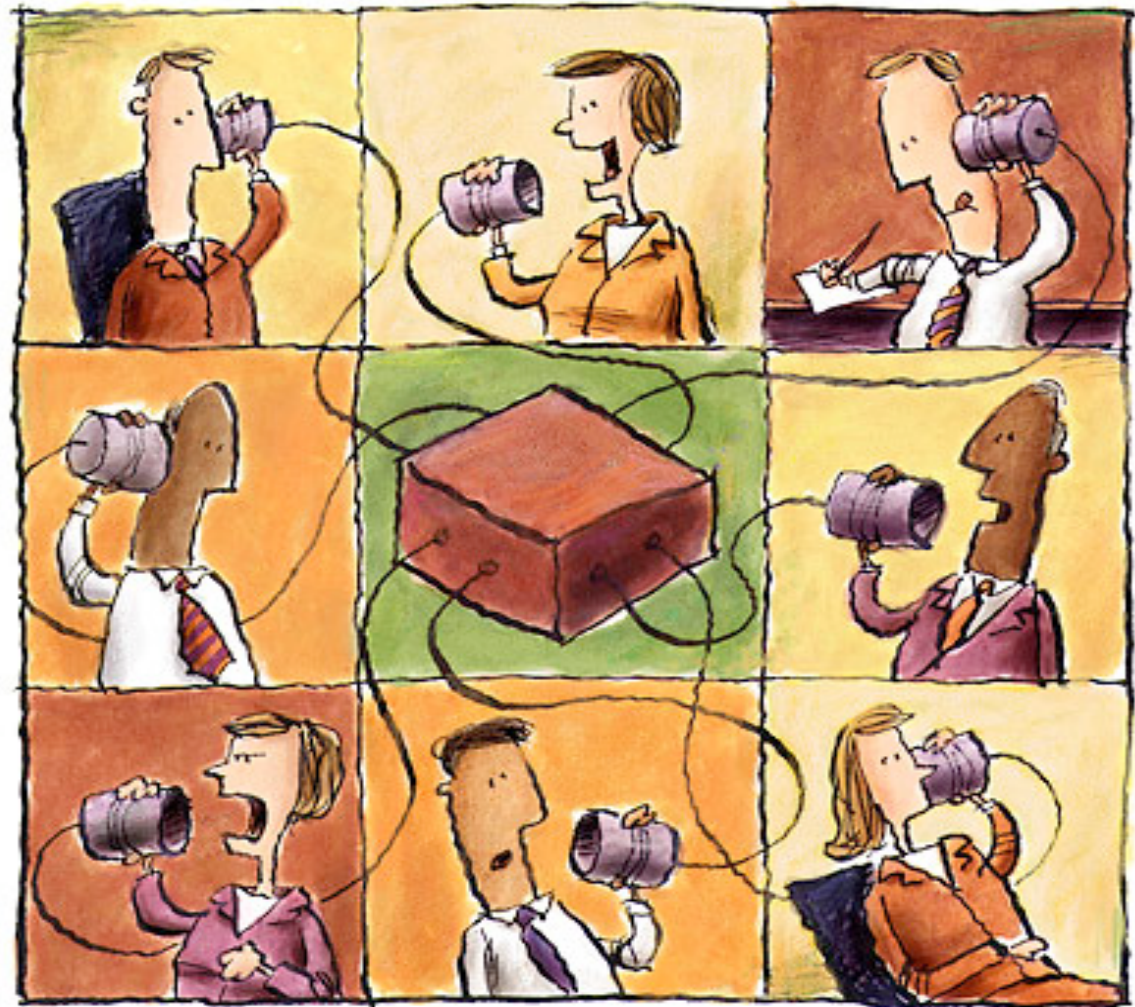import   imprimante   inde   indice   industrie

How to turn folksonomies ...

?
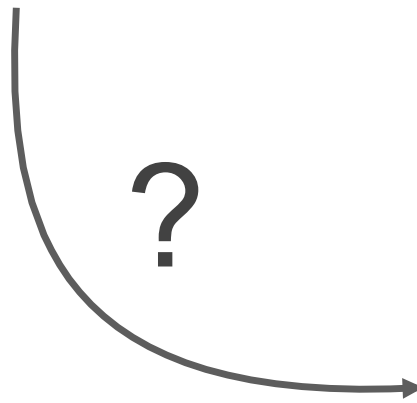
...into comprehensible topic structures ?

pollutant   **Energy**

related   related

pollution
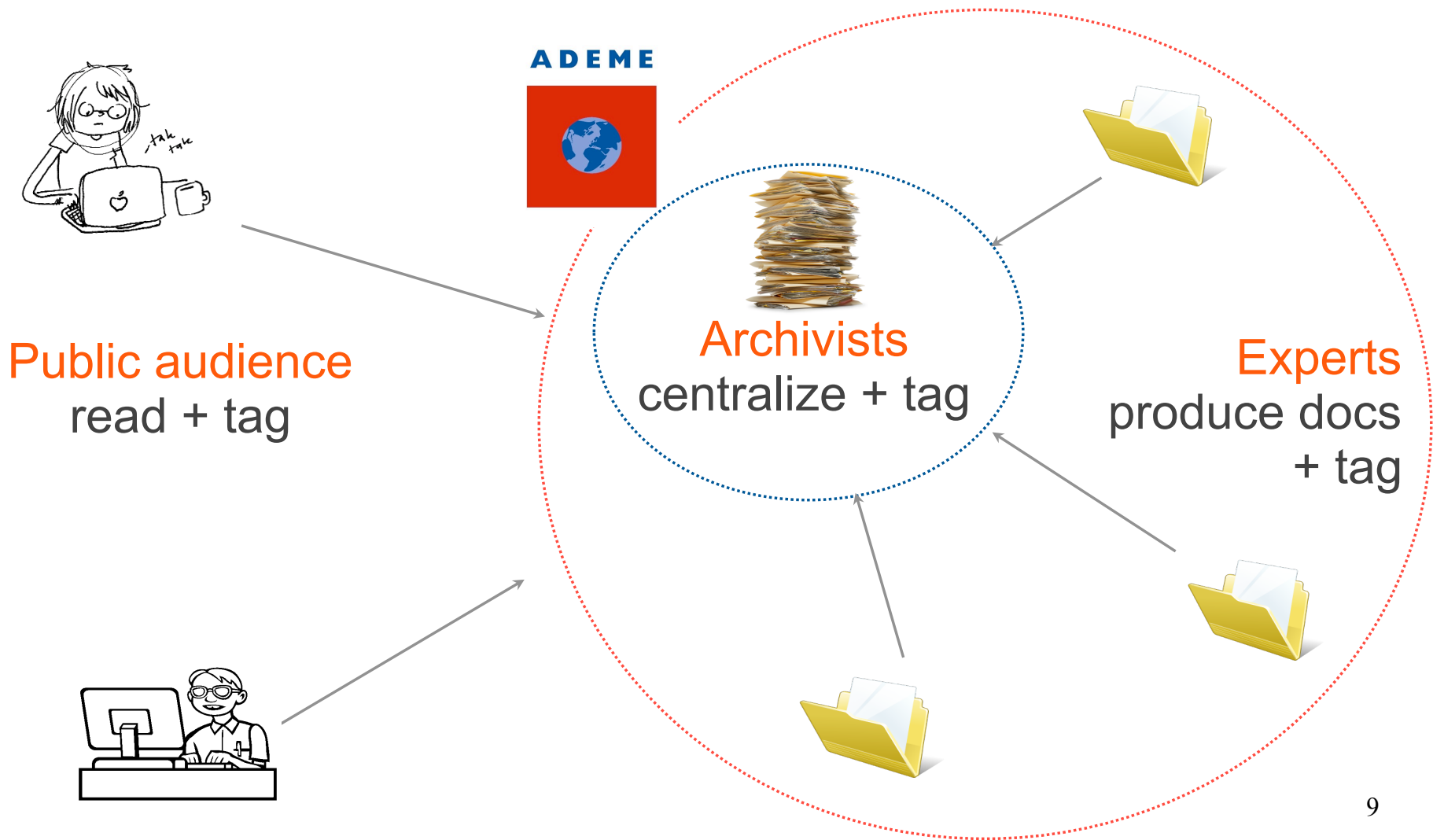
has narrower

Soil pollutions

6
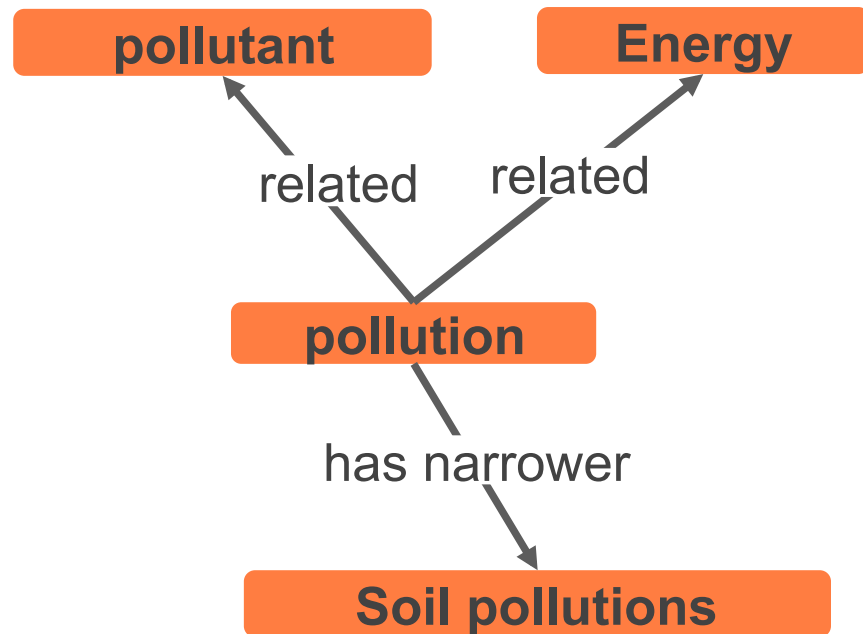
… by collecting all user's expertise into the process

Our approach

- Integrate usage-analysis for a tailored solution

- Supporting diverging points of view

- Automatic processings

    +

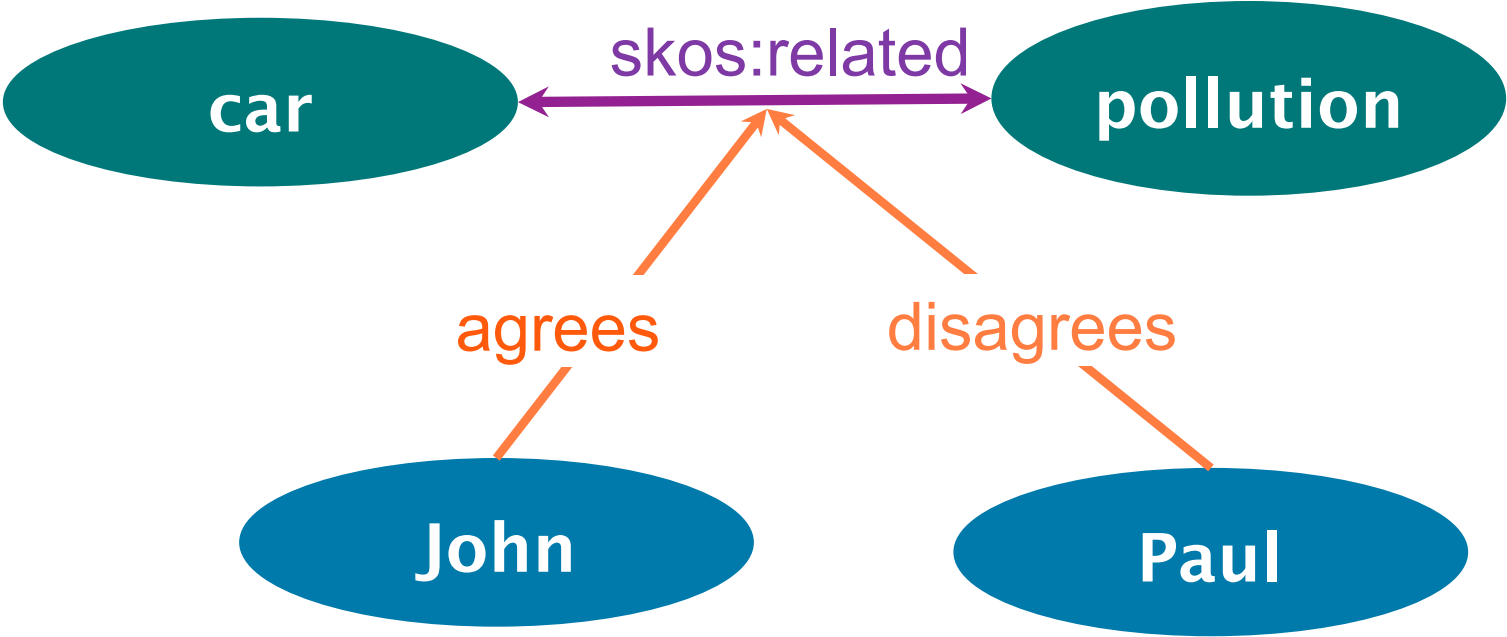- Human expertise through user-friendly interfaces

Concrete scenario

ADEME

Public audience
read + tag

Archivists
centralize + tag

Experts
produce docs
+ tag

Supporting diverging points of view
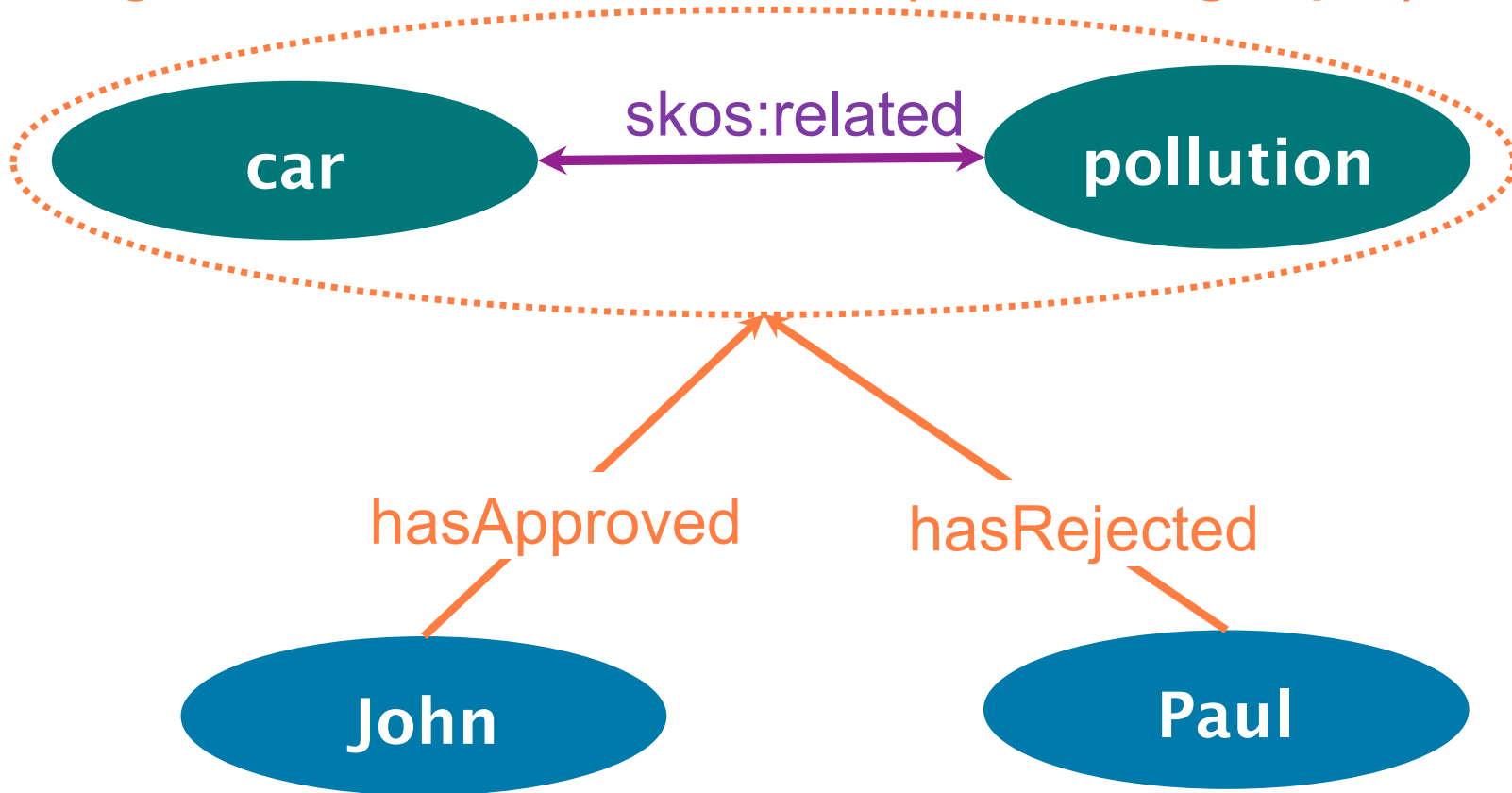

car — skos:related — pollution

Supporting diverging points of view

Supporting diverging points of view

tagSemanticStatement (named graph)

Folksonomy enrichment life-cycle

Flat folksonomy

Automatic processing

ADDING TAGS

User-centric structuring

Detect conflicts

Global structuring
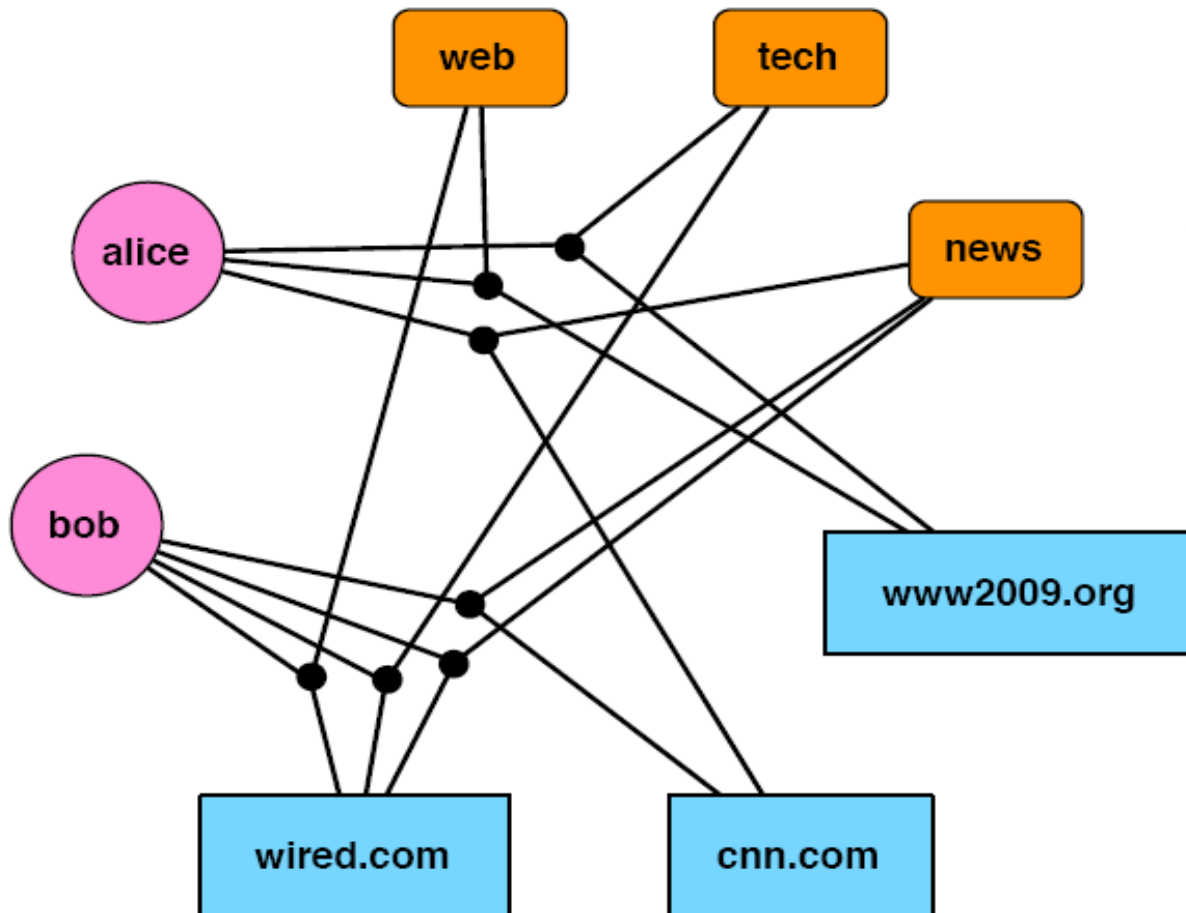
Structured folksonomy
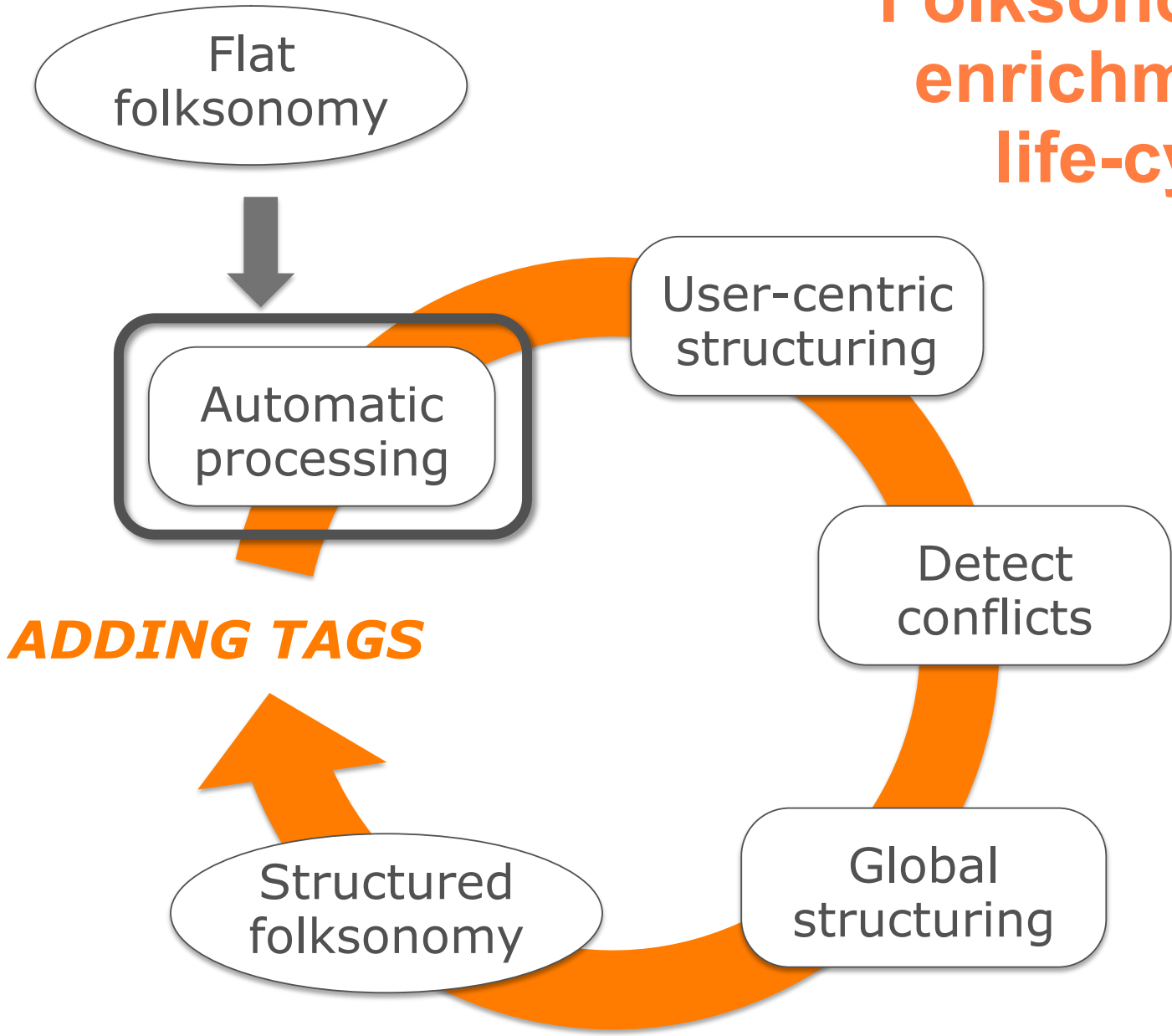
1. Comparing Tag labels
with string edit distances
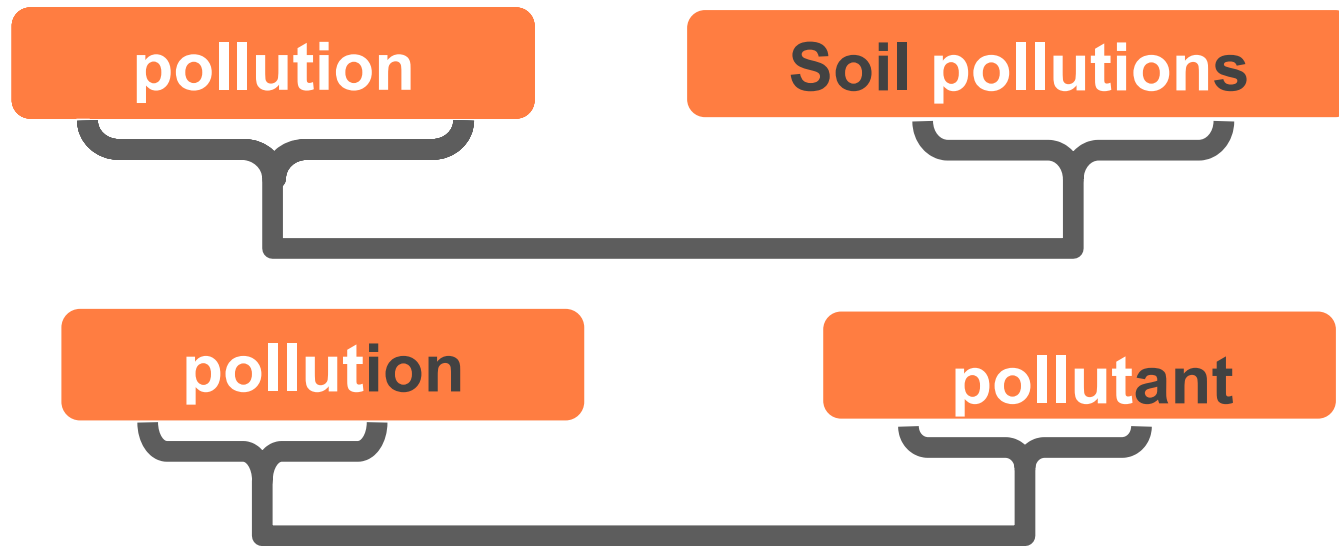
1. Comparing Tag labels
with string edit distances

- Evaluation of 30 edit distances

- Combining the best metrics

- Needs complement !
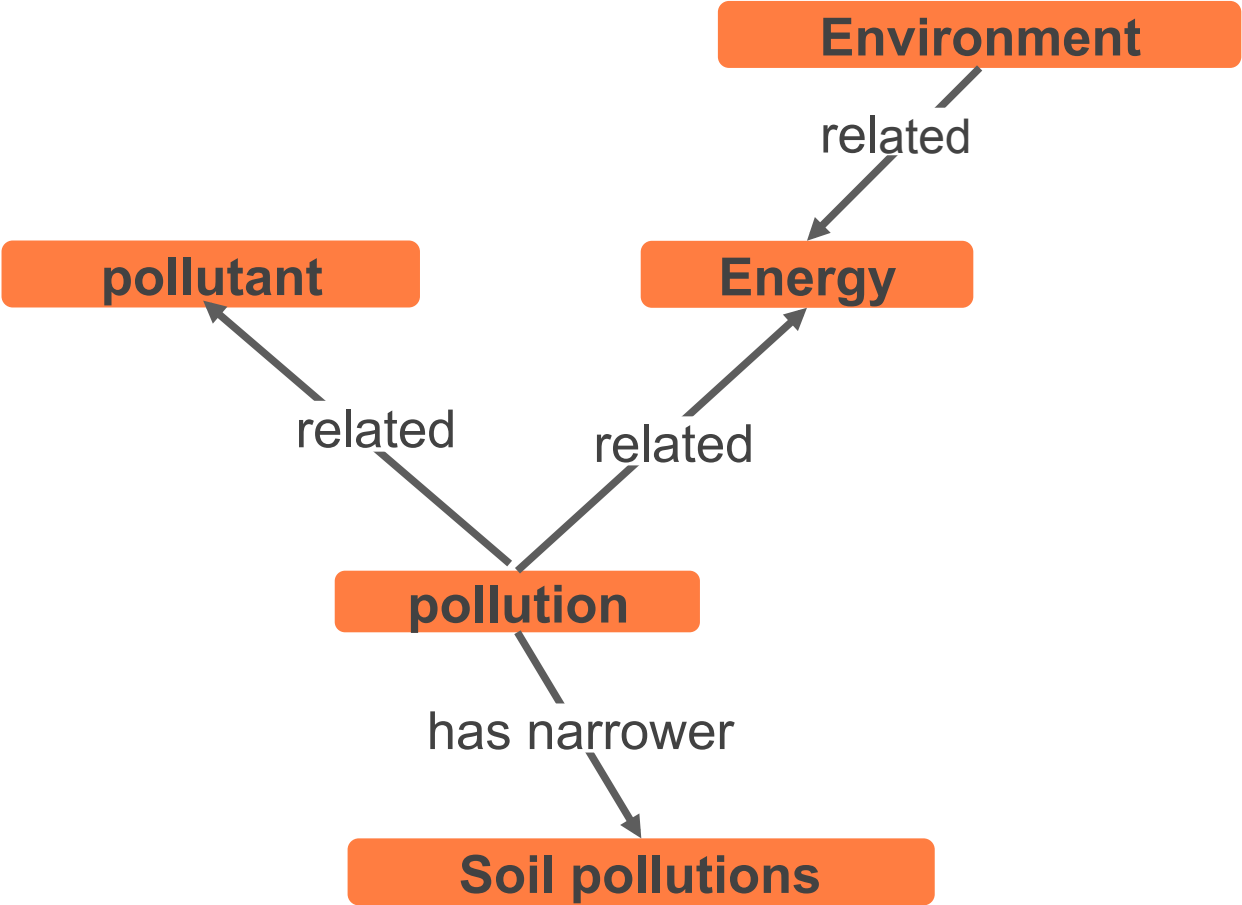
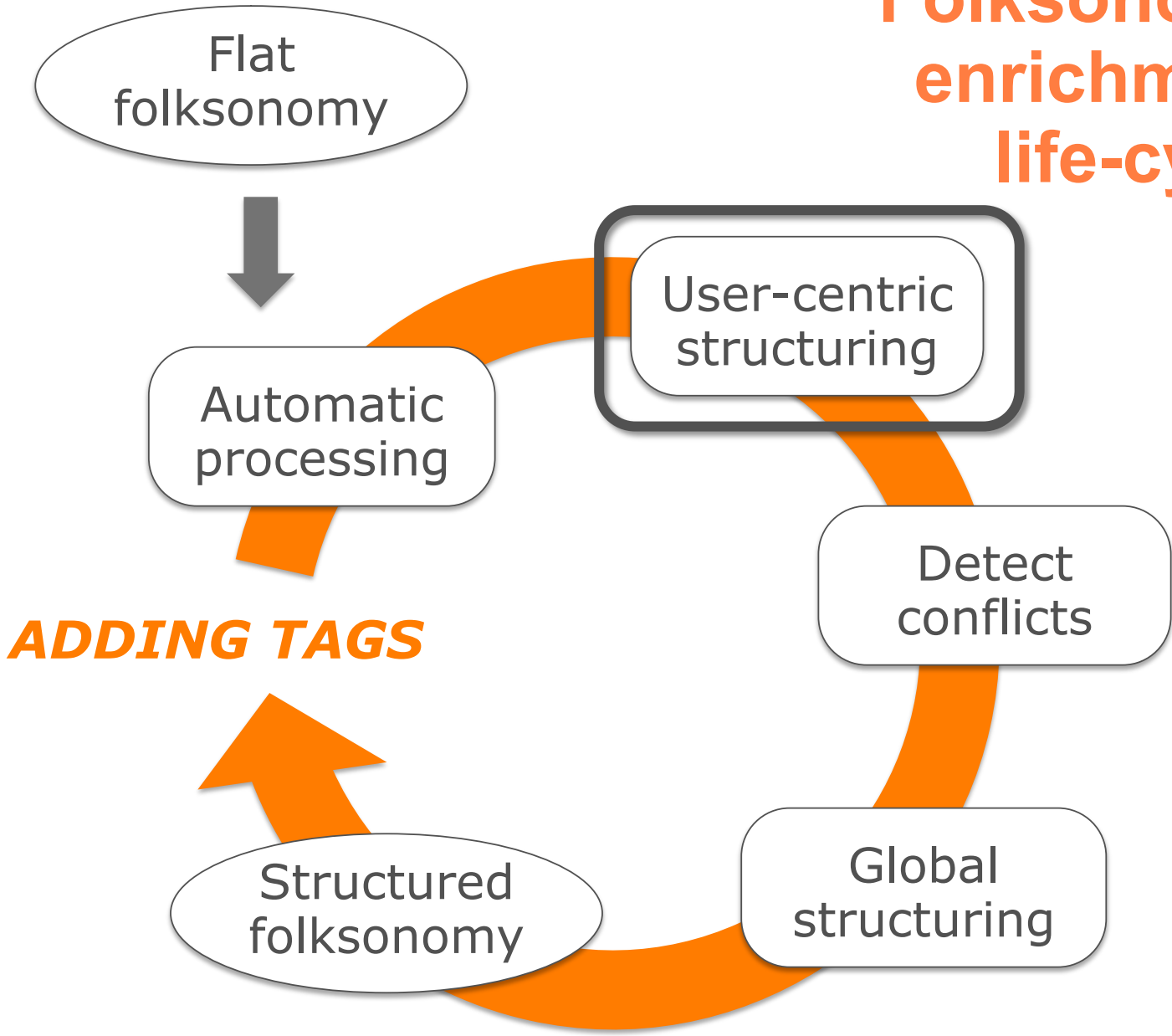# 2. Analyzing the tri-partite structure of folksonomies

Fig. Markines *et al.* (2009)

Association via:

- Users

- tags

Result: automatically suggested semantics

Folksonomy enrichment life-cycle

Flat folksonomy

Automatic processing

User-centric structuring

Detect conflicts

Global structuring

Structured folksonomy

ADDING TAGS

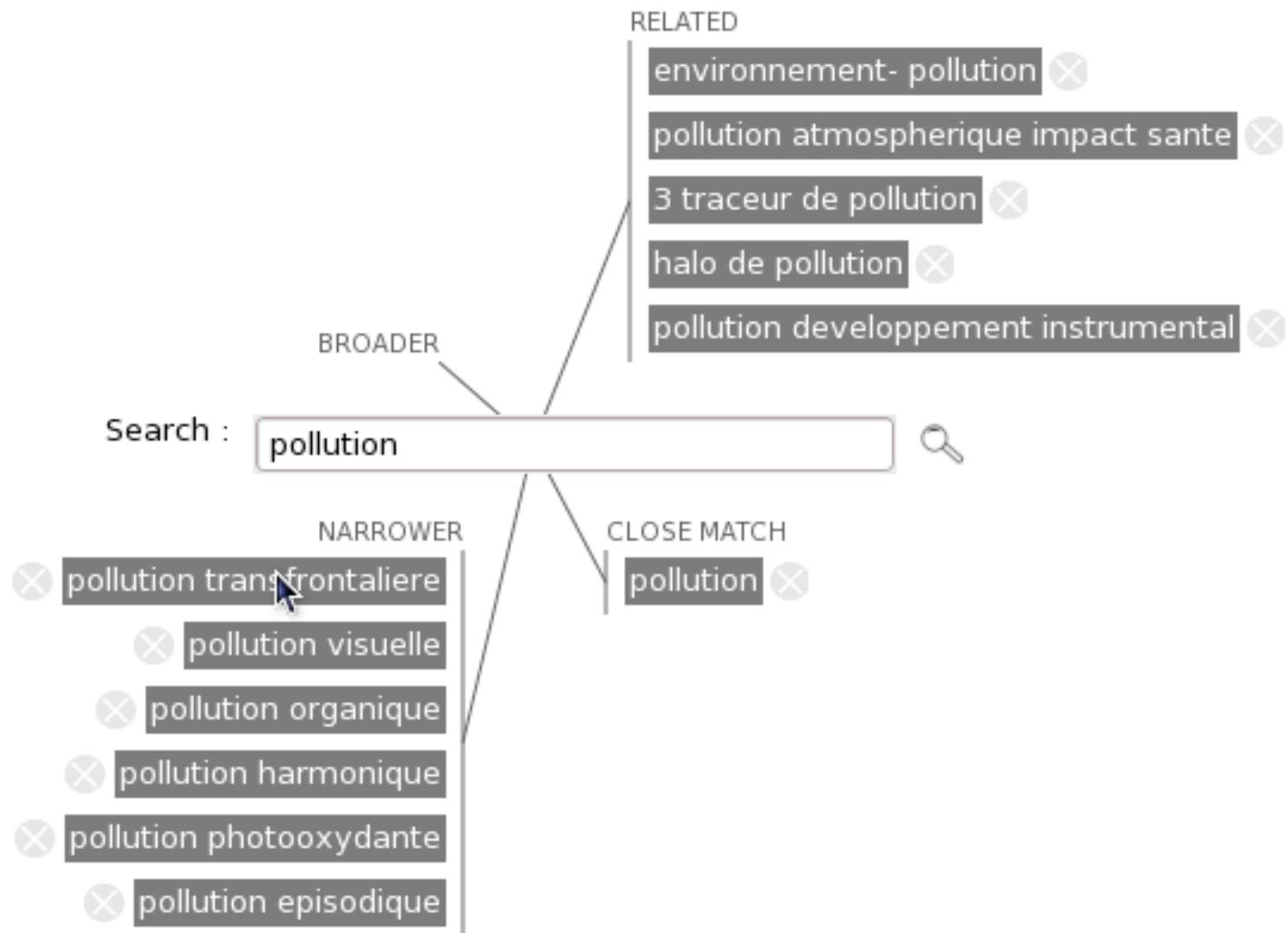# Embedding structuring tasks within everyday activity (searching *e.g*)

Embedding structuring tasks within everyday activity (searching *e.g*)
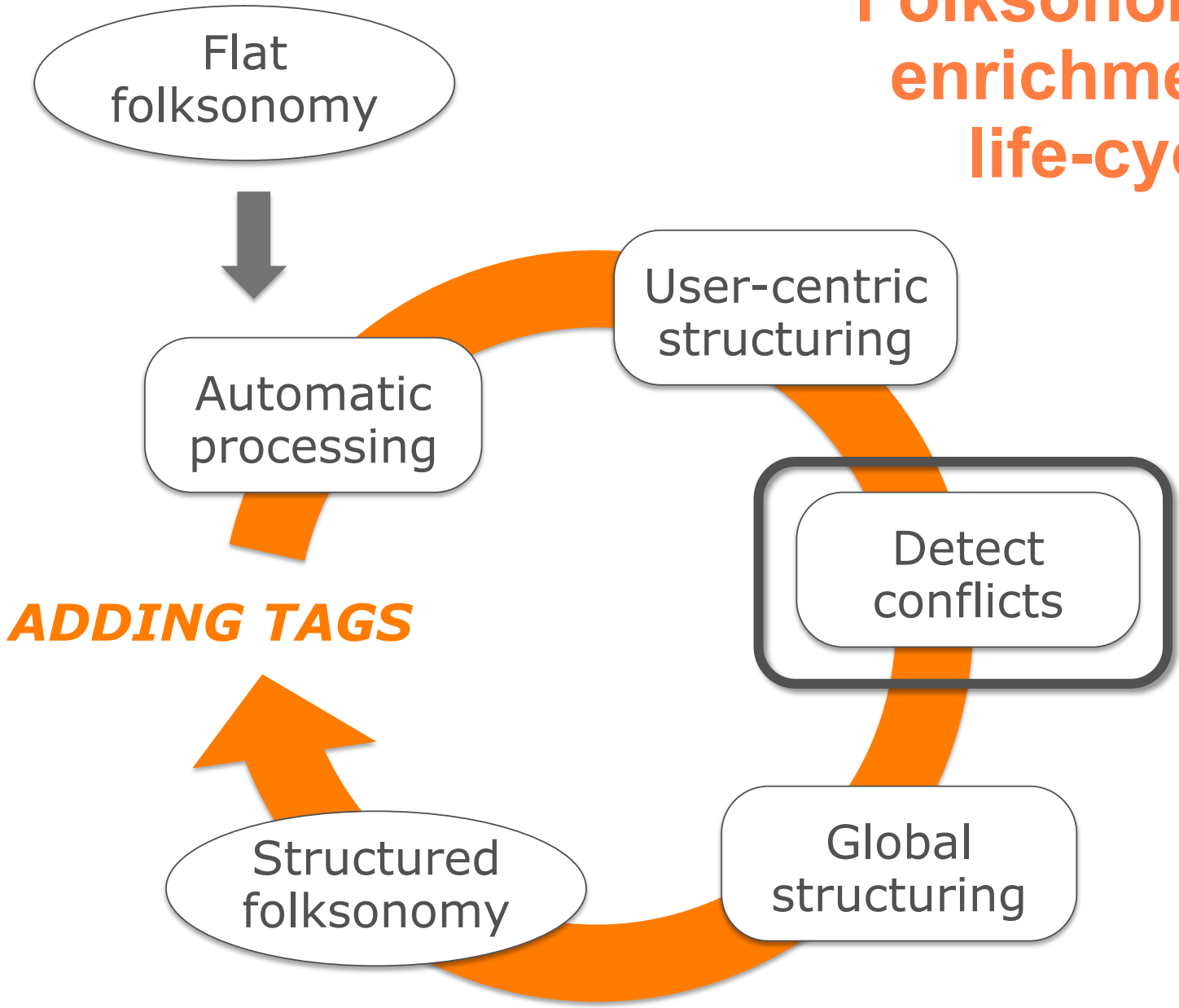
List of resources tagged with "pollution"

NB: Includes results for spelling variant tags (pollution)

- **Comptes (Les) �conomiques de l'environnement : Donn�es �conomiques de l'environnement, rapport de la commission des comptes et de l'�conomie de l'environnement. Editions 2001, 2002, 2003, 2005**
  Tagged by **svic_referent** with : administration publique cout pollution politique de l environnement
  ...

- **Strat�gie nationale de d�veloppement durable. Enraciner l'avenir dans l'action.**
  Tagged by **svic_referent** with : territoire peche climat sante ...

- **air des beffrois. ATMO Nord/Pas-de-Calais prend son envol (L')**
  Tagged by **svic_referent** with : polluant reseau mesure surveillance no2 ...

- **pollutions olfactives des installations class�es : de l'�valuation de la g�ne aux techniques de r�duction. salon de l'aveyron le 8 et 9 f�vrier 2005**
  Tagged by **svic_referent** with : emission odeur air agriculture ...

- **eurodeur biorodeur 2005, les solutions de d�pollution ext�rieures et int�rieures les impacts sanitaires de la pollution atmosph�rique. Actes des conf�rences 22 et 23 juin 2005**
  Tagged by **svic_referent** with : qualite air epidemiologie traitement odeur ...

Capturing user's point of view

Folksonomy enrichment life-cycle

Flat folksonomy

Automatic processing

User-centric structuring

ADDING TAGS

Detect conflicts

Structured folksonomy

Global structuring

Conflict detection

narrower

environment → pollution

broader

Conflict detection

narrower

environment → **pollution**
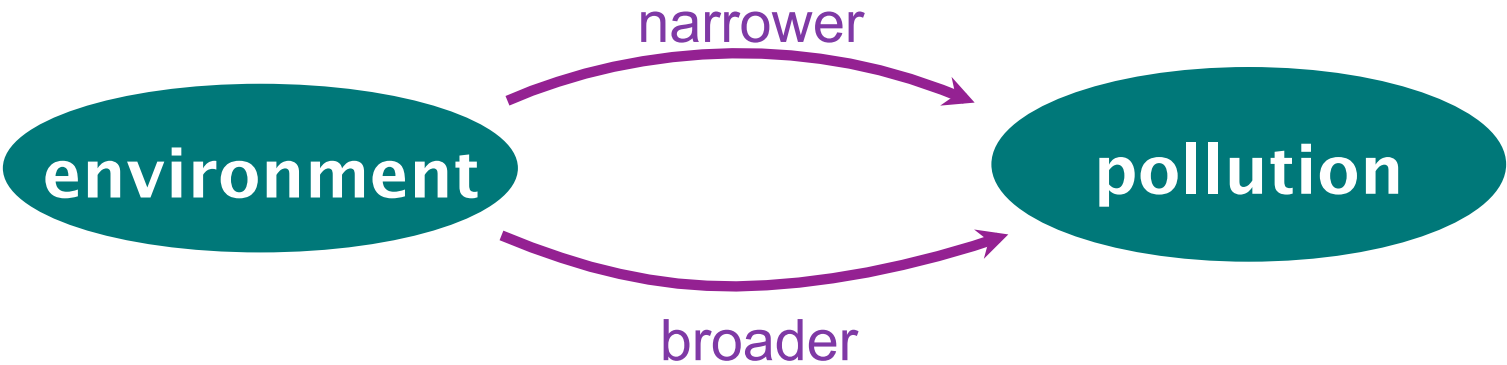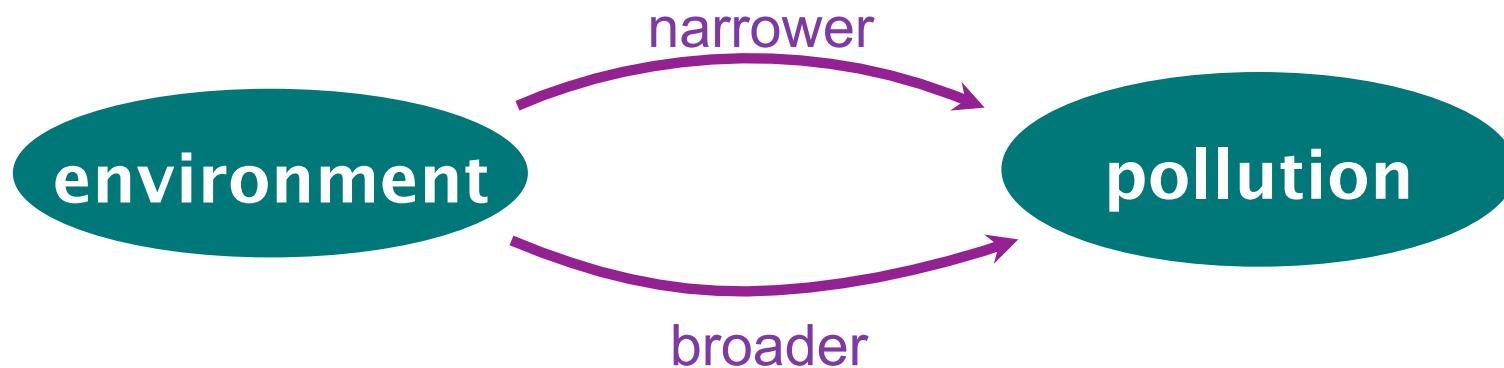
broader

Using rules *e.g*:

```
IF num(narrower)/num(broader) ≥ c
THEN narrower wins
ELSE 'more generic' wins
```

Conflict detection

narrower

environment ← related → pollution

broader

related
more generic   more generic
broader   narrower

Folksonomy enrichment life-cycle

Flat folksonomy

Automatic processing

ADDING TAGS

User-centric structuring
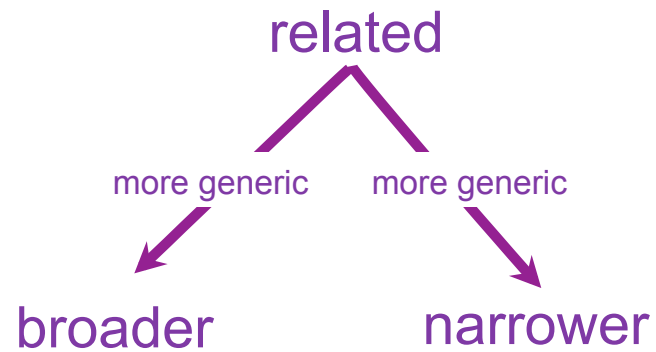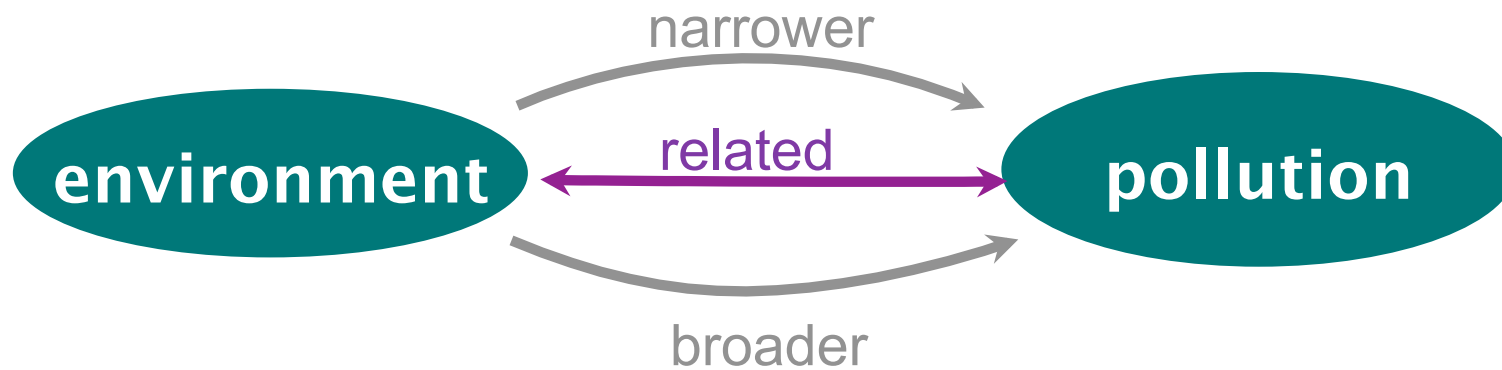
Detect conflicts

Global structuring

Structured folksonomy
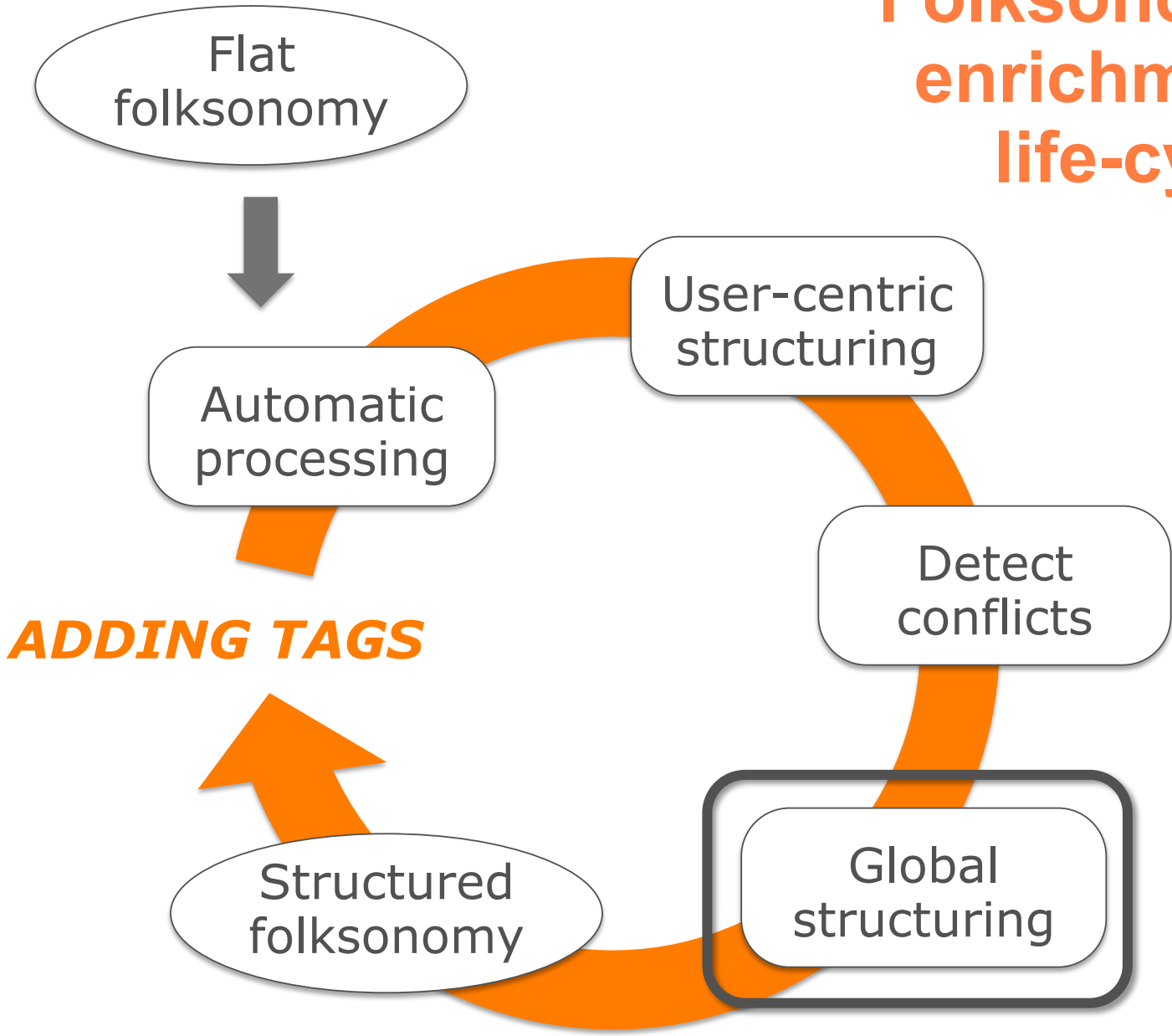
Global structuring by Referent

Folksonomy enrichment life-cycle

Flat folksonomy

User-centric structuring

Automatic processing

Detect conflicts

ADDING TAGS

Global structuring
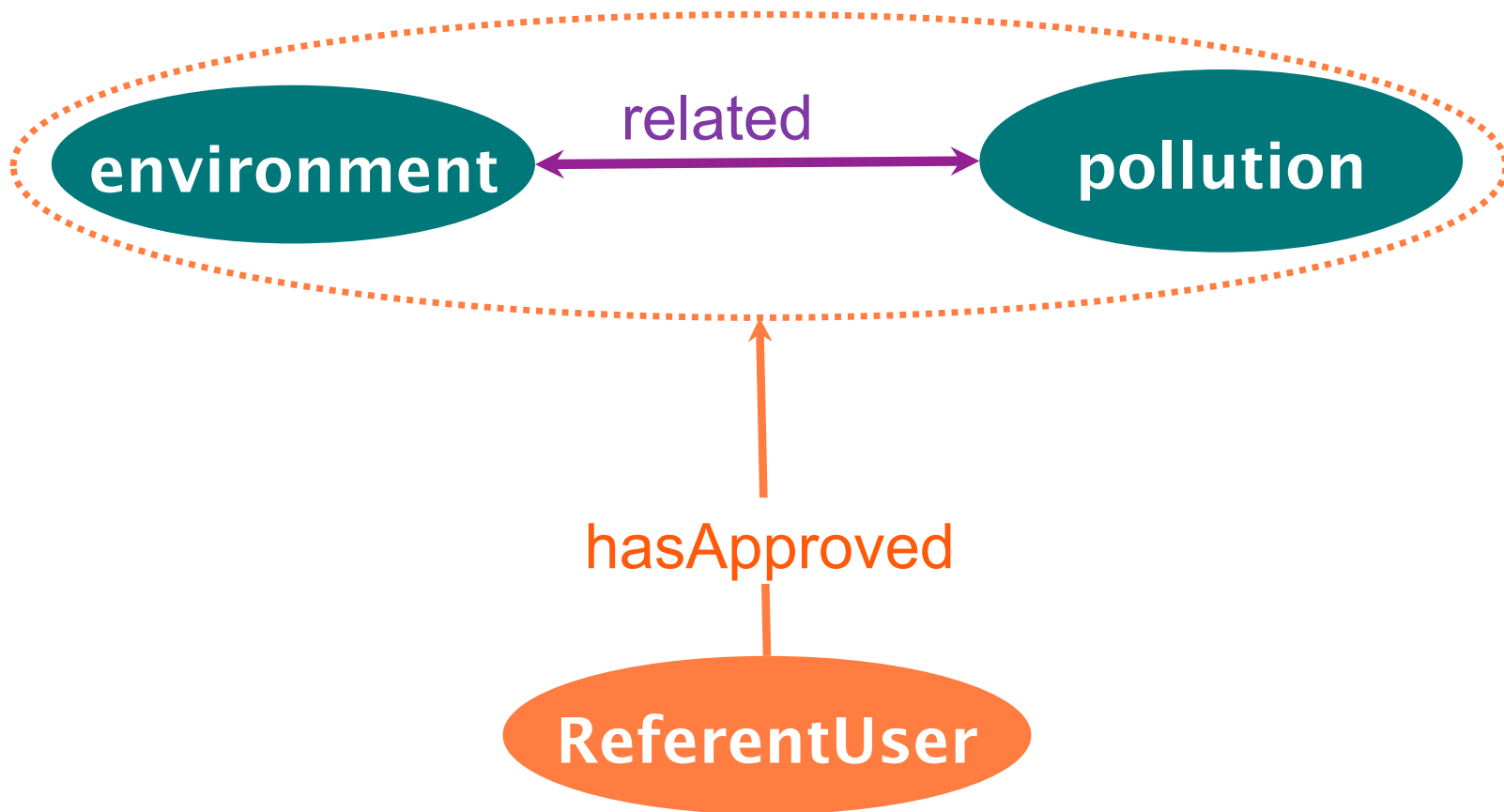
Structured folksonomy
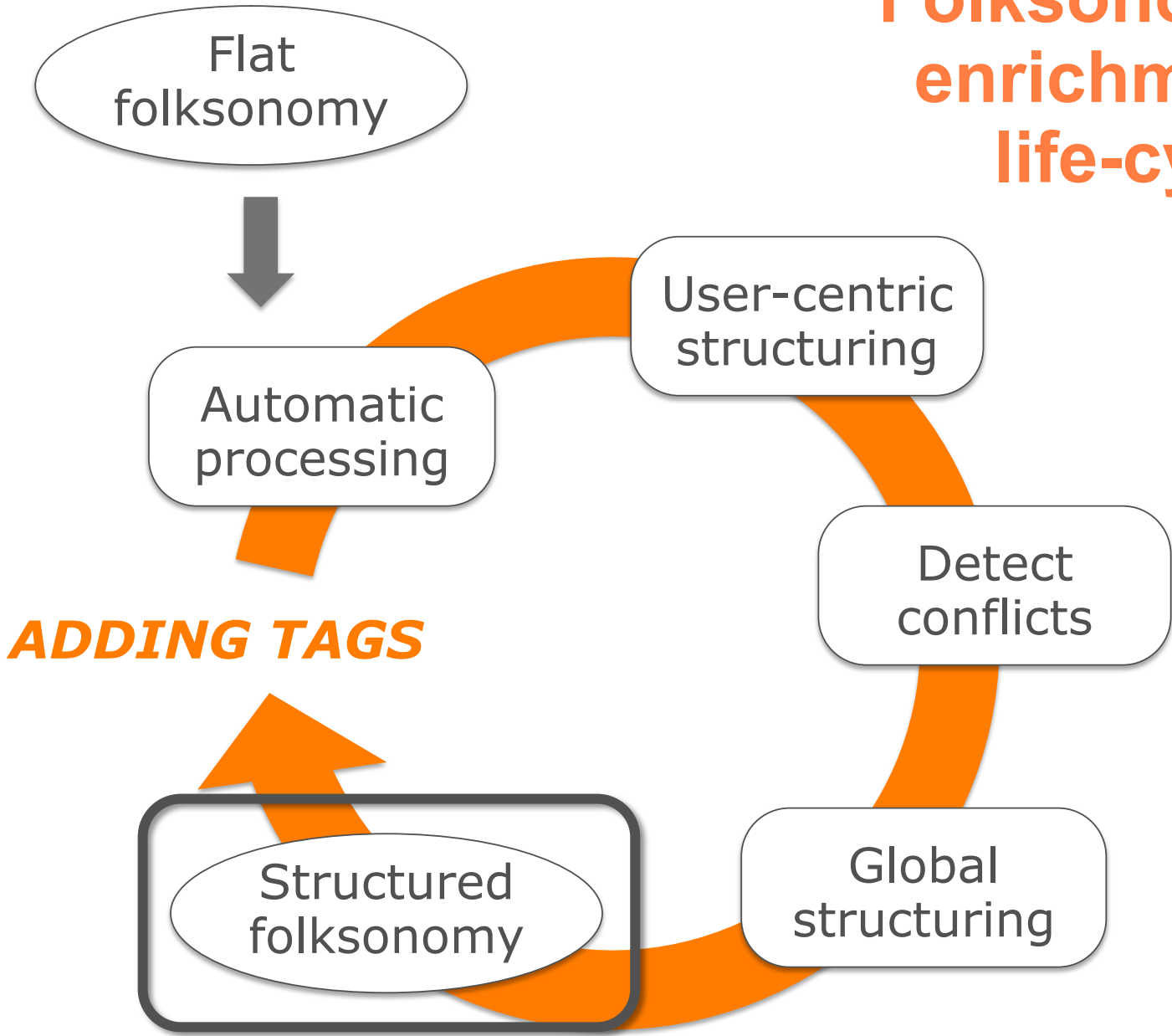
Take away message
(conclusion)

What we do :

Help communities

structure their tags

Our contributions:
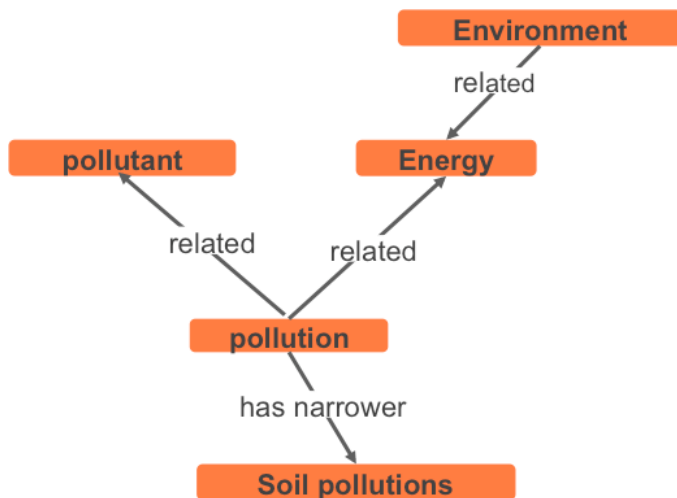
- Usages analysis

- Automatic processing of tags

- Tag structuring embedded in every-day tools

- Supporting multi-points of view

Implementation & tests

- ADEME dataset (~10000 tags)

- Tag server

- Tag searching interface

Perspectives

- More automatic methods

- More ontological resources

- Other interfaces (tagging, global structuring)

- Test + Evaluation @ Ademe & Orange Labs

Thank you !

freddy.limpens@inria.fr

http://www-sop.inria.fr/members/Freddy.Limpens/

http://isicil.inria.fr