kolflow

Man-machine collaboration in continuous knowledge-construction flows

Pascal Molli
Université de Nantes

28/06/2012 – Revue ANR Kolflow (02/2011-2014)

http://kolflow.univ-nantes.fr

UNIVERSITÉ DE NANTES

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

INRIA

Nancy-Université
Université Henri Poincaré

Kolflow

* Kolflow focus on man-machine collaboration and aims to build a semantic web collaborative space to bring together human agents and software agents in order to foster knowledge-intensive collaboration
  * GDD Nantes, Orpailleur Nancy, Silex Lyon, Wimmics Sophia
* **What happened in Kolflow from 2 February 2011 ?**
* **What is the plan now ?**

Kolflow facts

* Kolflow is a 42 months project started in 1 February 2011
* PhDs and Postdocs hired in July-November 2011
  * 4 PhDs, 1 post-doc, 1 engineer
* We are nearly at 1/3 of the project

Feb 2011 June 2012 June

SWCS 2012@WWW2012

* We organized SWCS 2012 workshop in conjunction with WWW2012, 17 april 2012.
* Proceedings published in ACM DL http://www.swcs2012.org
* Validate Kolflow directions and gather a community around Kolflow topics
* SWCS2013…

Kolflow Expected results and Joined Publications

* Deliver Man-Machine collaboration scenarios and some reference corpus
  * *Champin, Pierre-Antoine, Cordier, Amélie, Lavoué, Elise, Lefevre, Marie, Skaf-Molli, Hala* - ***User assistance for collaborative knowledge construction*** *Proceedings of the 21st international conference companion on World Wide Web pp. 1065--1074, Lyon, France,2012*
  * Taaable Sparql endpoint : http://wikitaaablesparql.loria.fr/status/
* Make automated reasoning understandable by humans.
  * Hasan, Rakebul, Gandon, Fabien - **Linking justifications in the collaborative semantic web applications** Proceedings of the 21st international conference companion on World Wide Web pp. 1083--1090, Lyon, France,2012
* Manage inconsistencies generated by man-machine collaboration
  * *A. Cordier, J. Lieber, J. Sevenot* - ***Towards an operator for merging taxonomies (submitted)*** *Workshop on Belief change, Non-monotonic reasoning and Conflict resolution, with ECAI-2012 , Montpellier, France, August 2012*

Kolflow Expected results and Joined Publications

* Deliver Man-Machine collaboration and some reference corpus
  * *Skaf-Molli, Hala, Desmontils, Emmanuel, Nauer, Emmanuel, Canals, Gérôme, Cordier, Amélie, Lefevre, Marie, Molli, Pascal, Toussaint, Yannick - **Knowledge continuous integration process (K-CIP)** Proceedings of the 21st international conference companion on World Wide Web pp. 1075--1082, Lyon, France,2012*
  * *Cordier, Amélie, Gaillard, Emmanuelle, Nauer, Emmanuel - **Man-machine collaboration to acquire cooking adaptation knowledge for the TAAABLE case-based reasoning system** Proceedings of the 21st international conference companion on World Wide Web pp. 1113--1120, Lyon, France,2012*
* Build a social semantic space…
  * *Luis Daniel Ibanez, Hala Skaf-Molli, Pascal Molli, Olivier Corby **Synchronizing semantic stores with commutative replicated data types** Proceedings of the 21st international conference companion on World Wide Web pp. 1091--1096, Lyon, France,2012*

Knowledge Continuous Integration Process (K-CIP)

GDD, Orpailleur, Silex

Context

* In Social Semantic Web, Information, Ontology and Queries are mixed in the same space.
  * Semantic Wikis, Wikidata
* **How can I modify the ontology without breaking the queries ?** When I modify the ontology:
  * I want to know the impact of modifications on the queries
  * I want to ensure the non regression of the system

Approach

* Continuous integration in a social semantic space
  * Express the process: validation of changes
  * Express the quality: mainly through testing
  * Test the quality of changes for each incremental change
* Ensures that a modification of the ontology does not alter the system behavior
  * Which language to define tests ? How to write tests ?
  * How to collect test data ?
  * When and where to change the ontology ?
  * When and where to execute tests ?

Defining Tests

* A test T is defined as a set of assertions {Ai} on the results set RQT for a given query QT
  * (QT , {Ai})
* Assertions are defined as logical expressions using set operations on RQT, R+, R-, R?
  * R+: a set of relevant answers
  * R-: a set of irrelevant answers
  * R?: a set of unknown answers

* A modification does not reduce the positive answers of the query:

$$Assert(\ R^{+} \subseteq R_Q\ )$$

* A modification does not change the positive answers of the query:

$$Assert(\ R^{+} = R_Q\ )$$

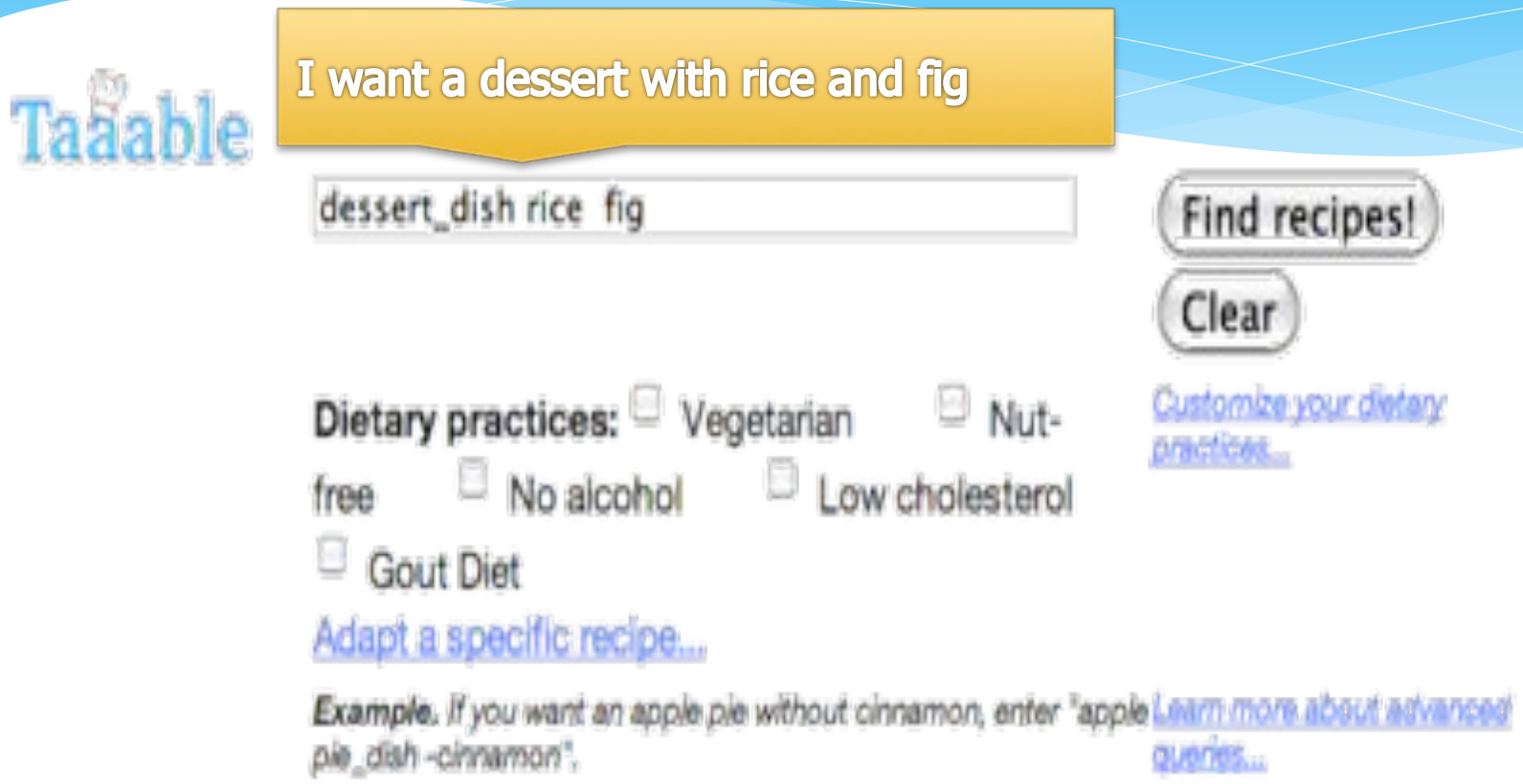
* A modification does not introduce unwanted results for a query

$$Assert(\ R_Q \cap R^{-} = \emptyset)$$

* More positive answers than unwanted ones negative

$$Assert(|\ R^{+} \cap R_Q\ | > |\ R^{-}\ |)$$

Collecting Tests



Taaable

dessert_dish rice fig

Find recipes! Clear

Dietary practices: Vegetarian Nut-free No alcohol Low cholesterol Gout Diet

Customize your dietary practices...

Adapt a specific recipe...

Example. If you want an apple pie without cinnamon, enter "apple pie_dish -cinnamon". Learn more about advanced queries...

Your request is: dessert_dish fig rice

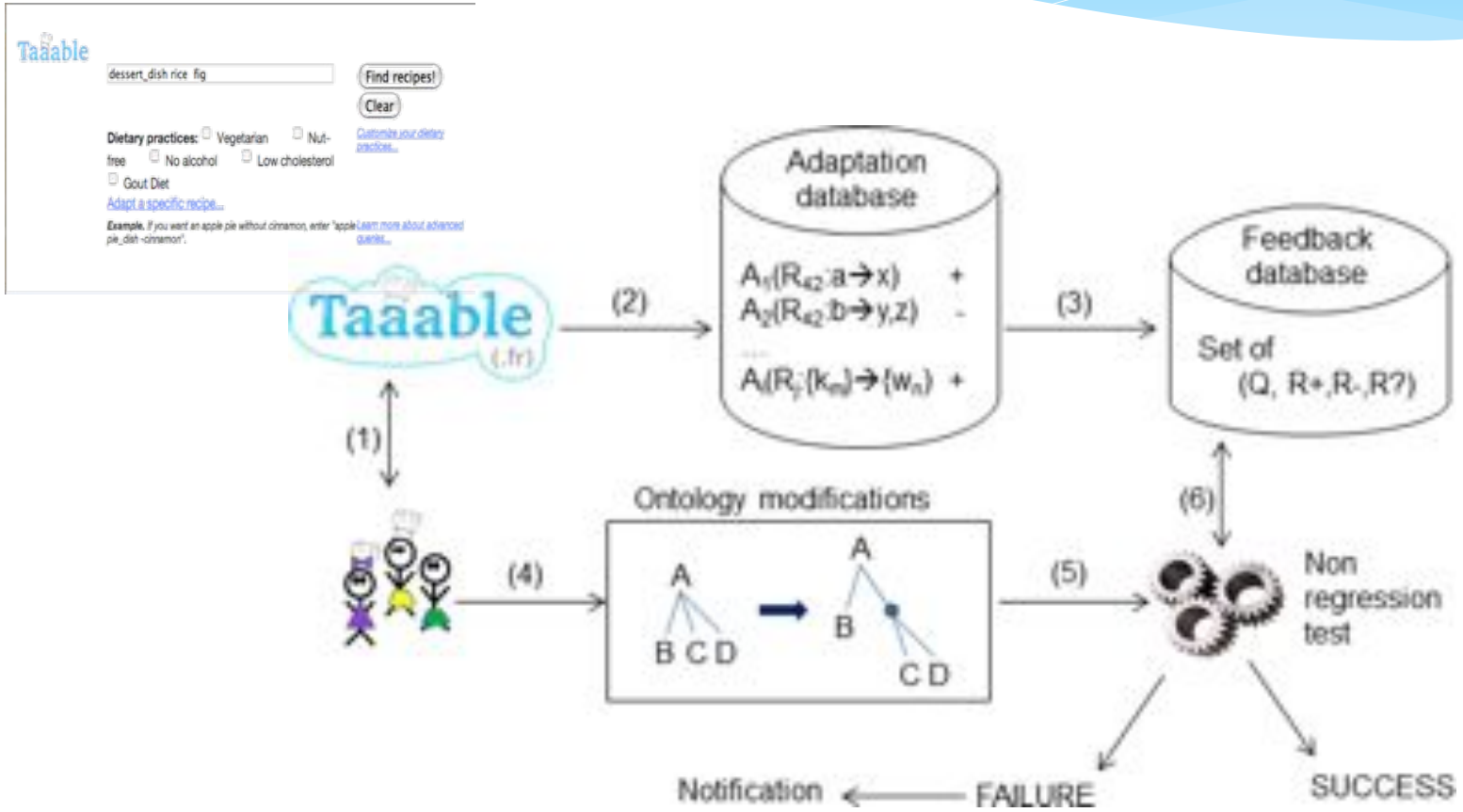
The request used for adaptation is: dessert_dish fig rice

#	Original recipe name (click to open recipe)	Adaptation overview (click to see the details)
1	Glutinous_rice_with_mangoes	Replace: Mango by Fig

Social Feedback for Collecting Test Data

* R+, R- and R? are collected from the social feedback
* When a user queries the system, according to the answer, she can decide to:
  * Agree : added to R+
  * Not agree: added to R-
  * Do not know : added to R?

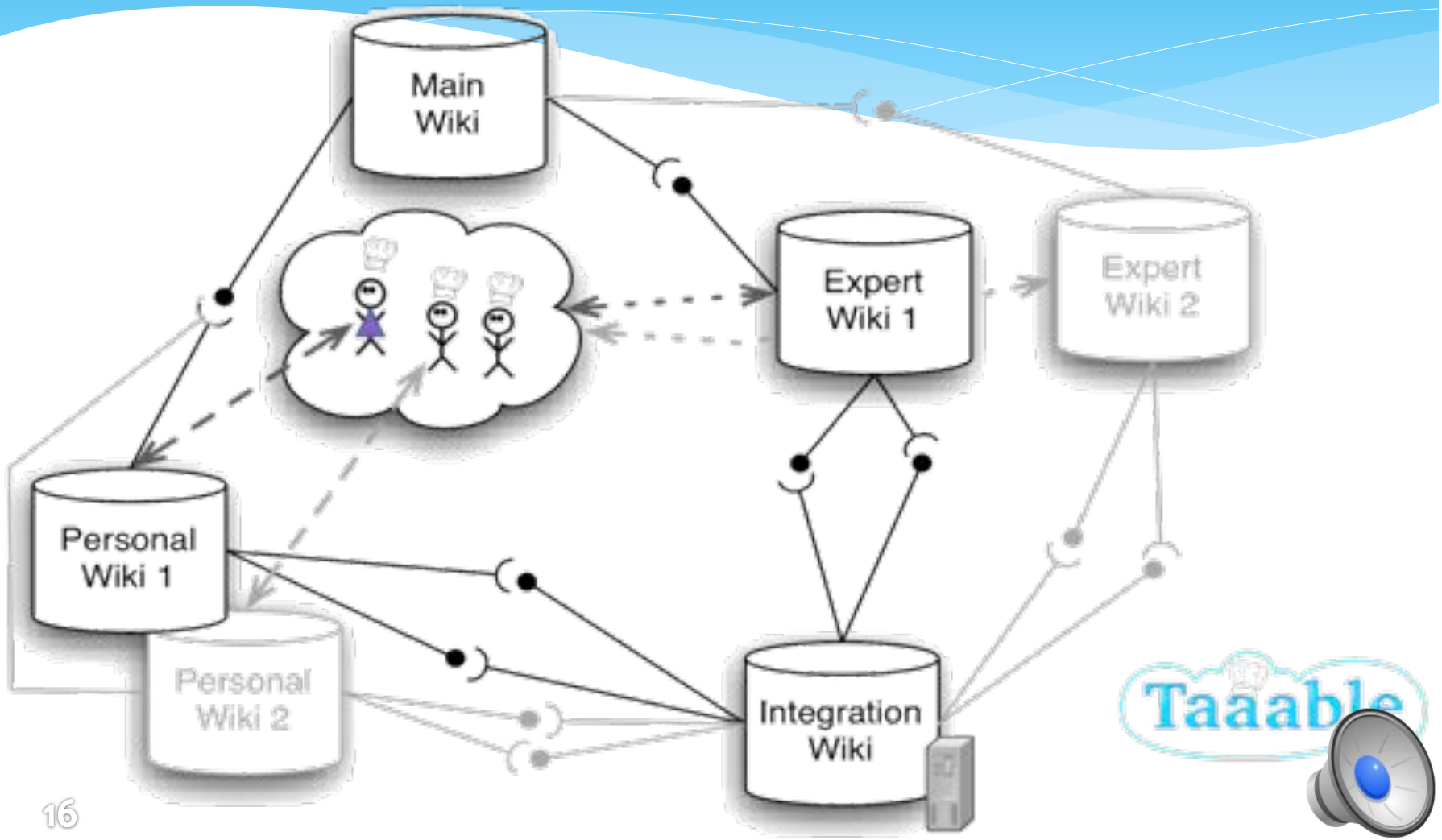
Test Data Collecting in WikiTaaable



Continuous integration in Man-Machine collaboration

* Ensure that a modification of the ontology does not alter the system behavior
  - ✓ Which language to define tests ? How to write tests ?
  - ✓ How to collect test data ?
  - * When and where to change the ontology ?
  - * When and where to execute tests ?

WikiTaaable in Distributed Semantic Wiki (DSMW)



16

Conclusion

* K-CIP is a continuous integration process for ontology evolution in the Social Semantic Web
* K-CIP prevent regression in a social semantic system..
* Time for enacting K-CIP:
  * We have tools : DSMW from GDD, DSMW+traces from Silex
  * Data, users and social feedback test collection : Orpailleur and Silex

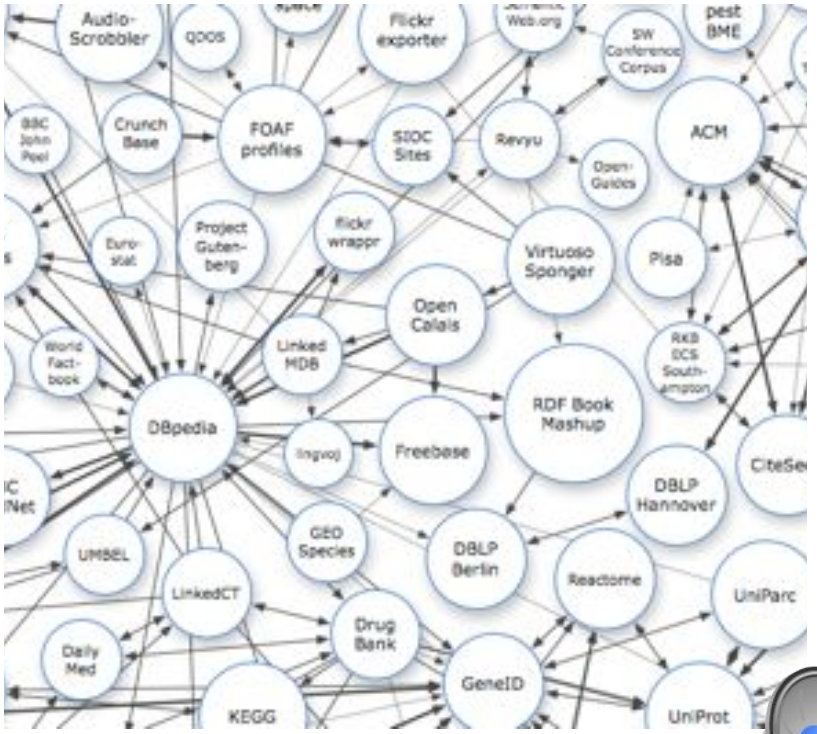
Live Linked Data

Synchronizing Semantic Stores with Commutative Replicated Data Types

GDD, Wimmics (Edelweiss)

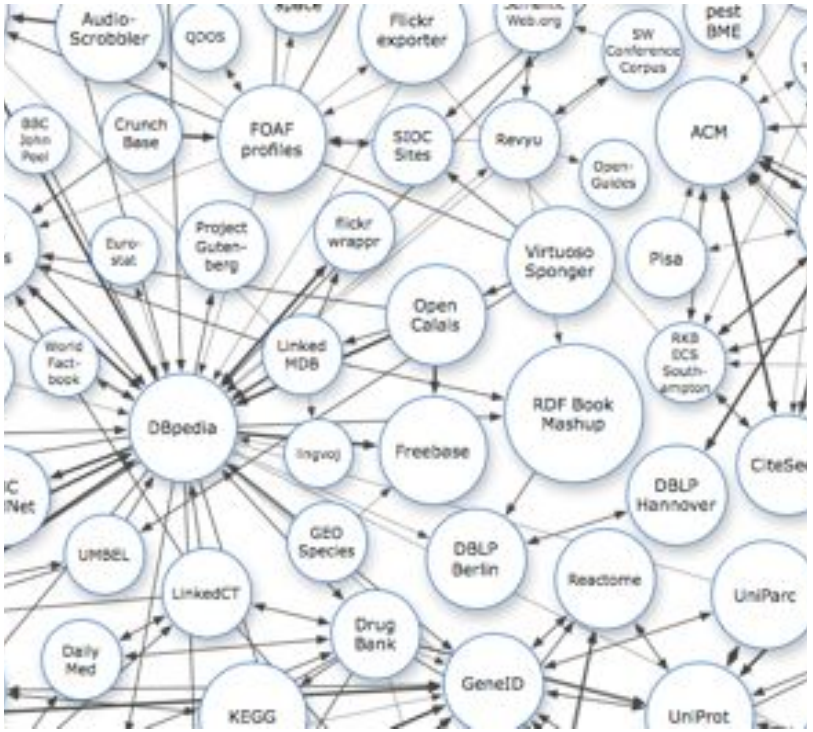
Context

* Social semantic web is composed by autonomous participants that are producing a continuous stream of evolving knowledge.
  * Ex: Wikipedia->Dbpedia
  * Linked data ?
* **How to make linked data writable ?**

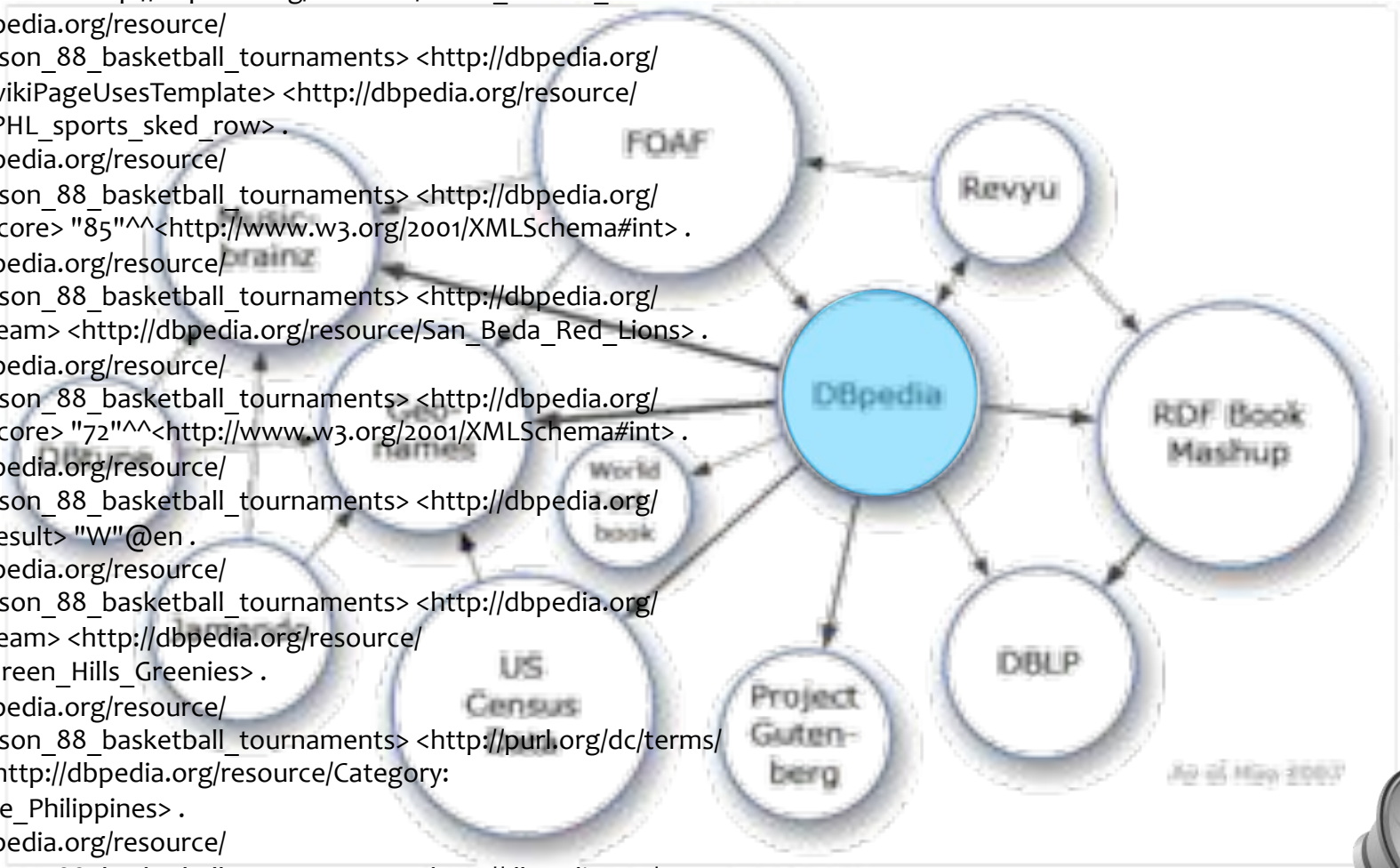


Context

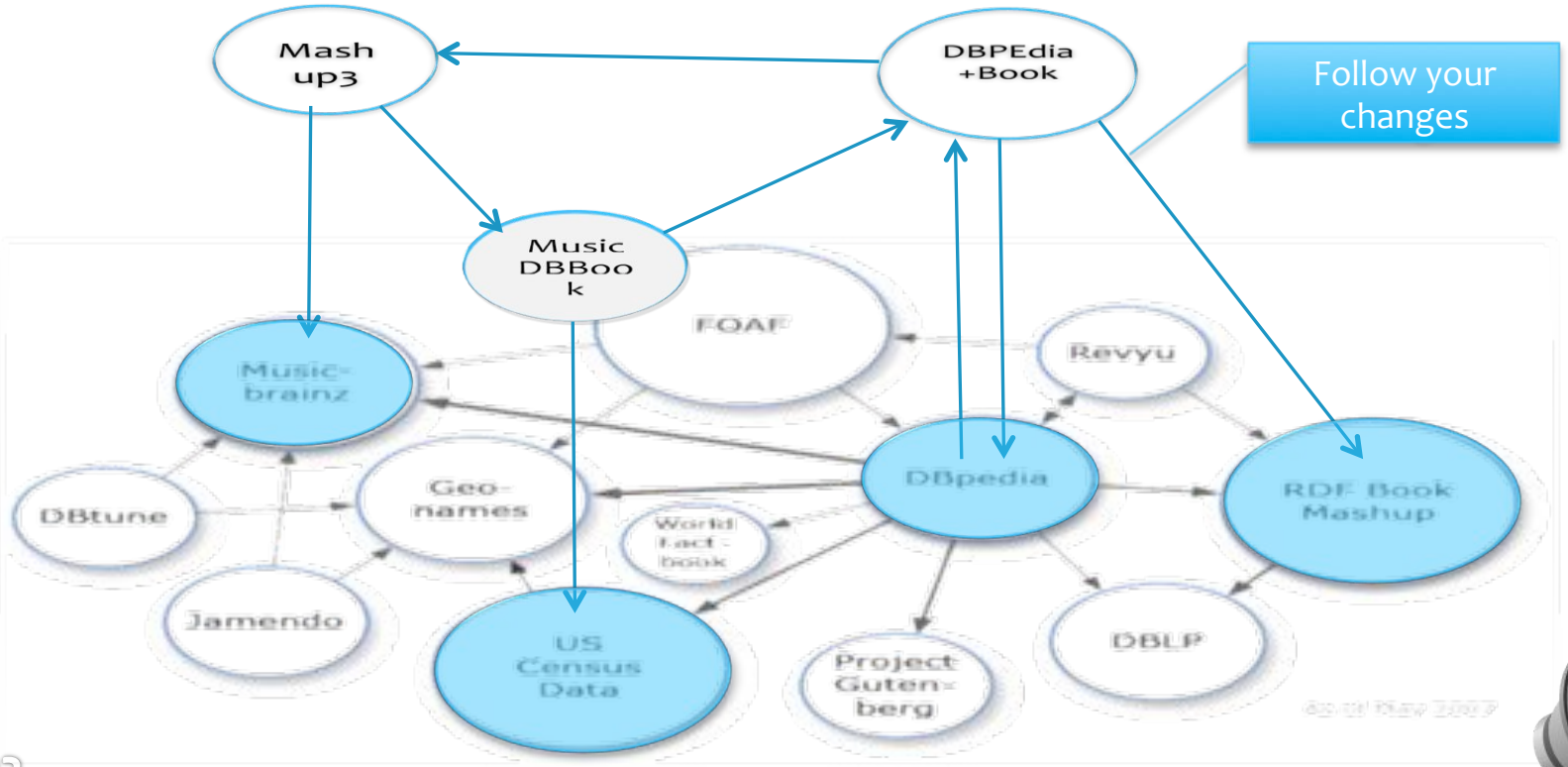
* If want to edit DBpedia
  * I have to copy DBpedia
* If I want to query 2 DBPedia and Freebase
  * Copy 2 datasets locally and query (freshness)
  * Distributed query (data availability)



<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/years> <http://dbpedia.org/resource/NCAA_Season_88> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/wikiPageUsesTemplate> <http://dbpedia.org/resource/
Template:PHL_sports_sked_row> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/score> "85"^^<http://www.w3.org/2001/XMLSchema#int> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/team> <http://dbpedia.org/resource/San_Beda_Red_Lions> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/score> "72"^^<http://www.w3.org/2001/XMLSchema#int> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/result> "W"@en .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
property/team> <http://dbpedia.org/resource/
La_Salle_Green_Hills_Greenies> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://purl.org/dc/terms/
subject> <http://dbpedia.org/resource/Category:
2012_in_the_Philippines> .
<http://dbpedia.org/resource/
NCAA_Season_88_basketball_tournaments> <http://dbpedia.org/
resource/Template:PHL_sports_sked_row> "result16"@en .
<http://dbpedia.org/resource/



Live Linked Data

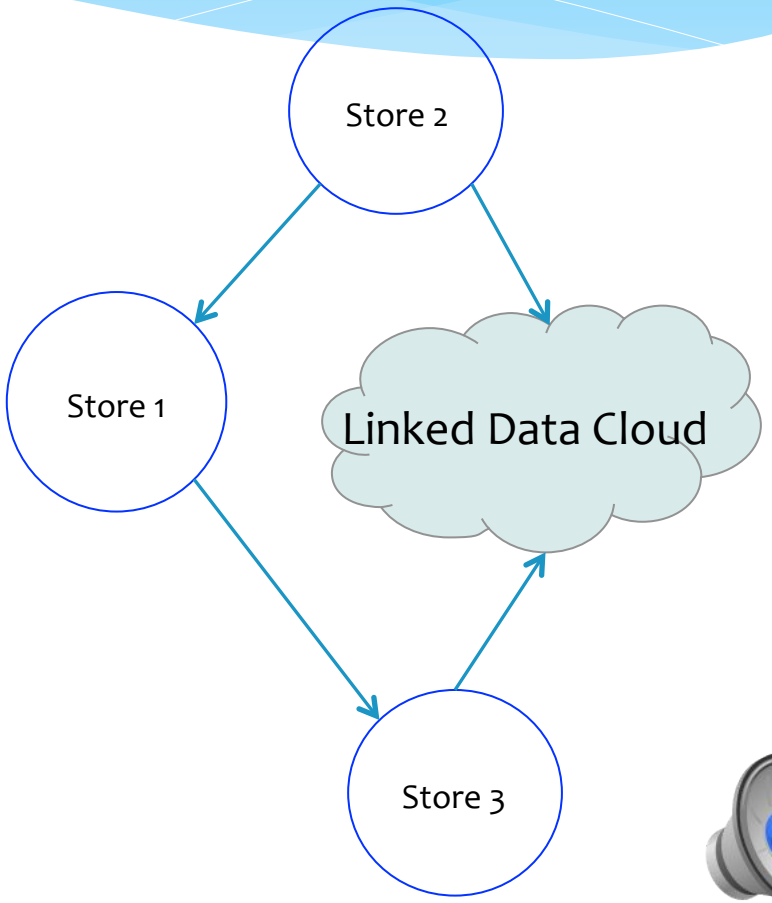


Live Linked Data

* A social network for linked data participant based on a « follow your change » relation
  * Makes Linked Data a « read/write » space : from linked data 1.0 to linked data 2.0
  * Creates assemblies of datasets and enable « synchronize and search » paradigm between warehousing approach and distributed queries approach.

Context

* What happens if participants start updating datasets?
  * We don't know who consumes from who...
  * Can get my own updates, multiple updates, conflicts...
* **What consistency criteria? And how to ensure it?**



Live Linked Data

* Let a social network of unknown number of linked data nodes linked by a *"follow your change"* relation
* Each linked data node:
  * Executes SPARQL Update queries locally.
  * Publishes these operations in "Live Streams"
    * **connection with ANR STREAM**
  * Other nodes consume and re-execute them.
* The system is correct if:
  * Convergence, Causality and Intention
  * **connection with ANR Concordant.**

SU-Set : A Sparql-Update Conflict-free Replicated Data Type

```
payload set S
    initial ∅
query lookup (triple t) : boolean b
    let b = (∃u : (t,u) ∈ S)
update insert (set<triple> T)
    atSource(T)
        let α = unique()
    downstream(T,α)
        let R = {(t,α) : t ∈ T}
        S := S ∪ R
update delete (set<triple> T)
    atSource(T)
        let R = ∅
        foreach t in T:
            let Q = {(t,u) | (∃u : | (t,u) ∈ S)}
            R := R ∪ Q
    downstream(R)
        // Causal Reception
        pre All add(t,u) delivered
        S := S \ R
```

Abstract operation is Sparql Update

Same id for all triples inserted together saves communication

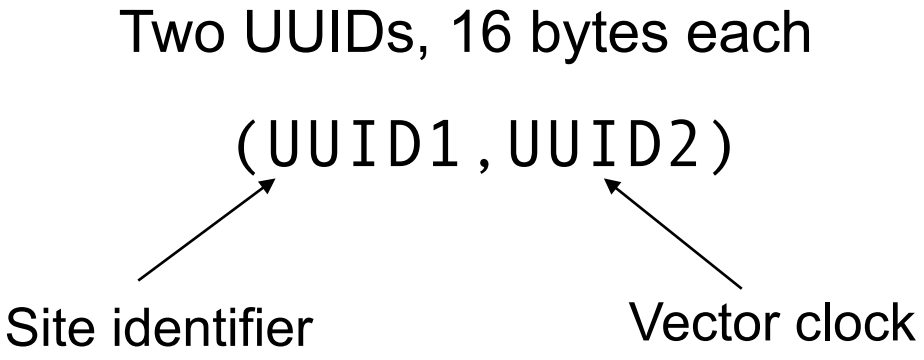
Delete all pairs associated to each triple. Can be expensive.

How much we need to pay to have eventual consistency ?

* Time Overhead :
  * Adding an id to each element is linear.
  * Selection and lookup is not affected by many pairs with the same triple.
* Round and # of messages Overhead :
  * Convergence after one round, one message per operation → Optimal

Validation – Space Overhead

* 32 bytes per 1 billion triples = 32 GB → 1 Ipod
* Semantic Stores already use an internal id → Reuse it
* Extra pairs produced by concurrent insertions could cause problems...



Communication Cost with DBPedia Live with SU-Set

* DBPedia Live generates one file with triples inserted and one with triples deleted approximately each 10 seconds
  * No pattern operations → No overhead here.
* Many more insertions than deletions
  * Insertions are cheap, they only need one id.
* Many triples per insertion
  * More triples inserted at a time is cheaper.

SU-Set Communication overhead on DBPedia Live

7 days of streaming
No concurrent insertions

Size (MB)

Operation	# of Triples	No ids	1 id per triple	1 id per operation
21957 Inserts	21762190	3403,4	4469,89	3404,6
21957 Deletes	1755888	238,46	324,5	324,5
Overhead			31,64%	2,39%

Conclusion

* Live Linked Data makes linked data “writable” and allow a new query paradigm
* SU-Set is a CRDT for RDF-Graphs updated with SPARQL-Update 1.1 that ensure eventual consistency on live linked data
* Time to embed in real system
  * Embed SU-Set in “Corese” engine of Wimmics

And now ?

* Kolflow at 1/3 of its life and delivered preliminary results
  * Many joint papers
  * Original assemblies of core skills of different partners
  * Relations with others running ANR Concordant and STREAM

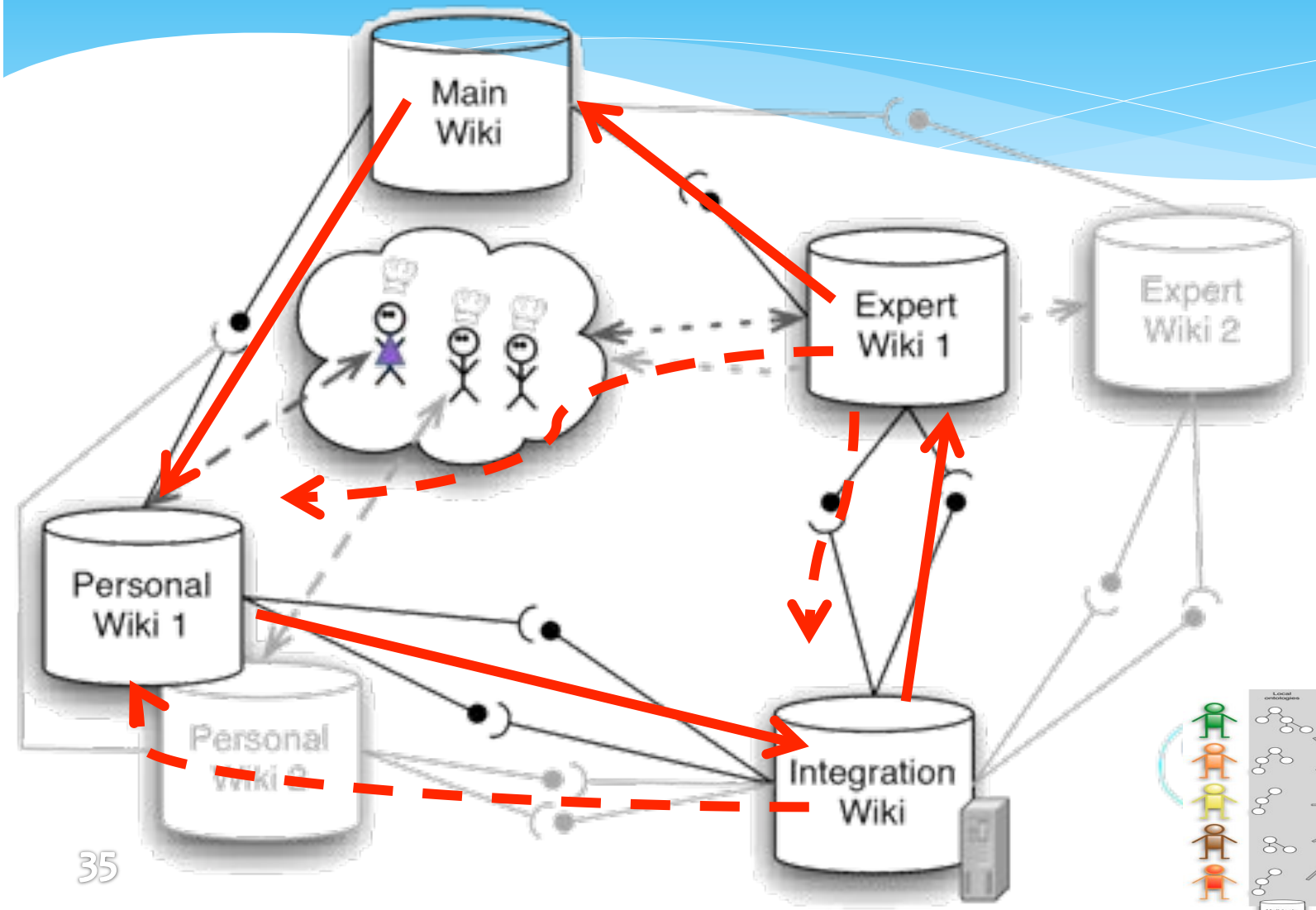
And now ?

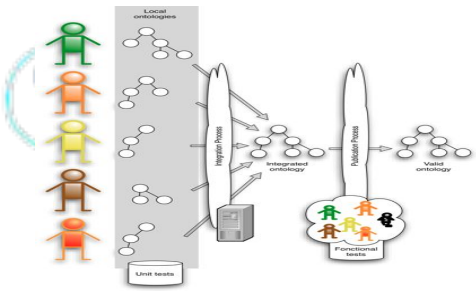
* Preliminary results established clear research directions
  * Skills, tools, data and users required for validations are already part of Kolflow project
* Our objective is now to transform workshop papers in major papers
  * Social semantic space
    * *L. Ibanez, H. Skaf-Molli, P. Molli, O. Corby - Live Linked Data: Synchronizing semantic stores with Commutative Replicated Data Types (Submitted), Journal of Metadata, Semantics and Ontologies (IJMSO) ,2012*
  * Deliver Man-Machine collaboration and some reference corpus
    * *Diego Torres, Pascal Molli, Hala Skaf-Molli, Alicia Diaz - From DBPedia to Wikipedia: Filling the gap by discovering Wikipedia conventions (submitted) The 2012 IEEE/WIC/ACM International Conference on Web Intelligence , Macau, China, December 2012*

And now ?

* Create a community around Kolflow topics with man-machine collaboration in social semantic spaces
  * Submit SWCS2013@WWW2013
  * Develop interactions with European and international partners

Changes Propagation in K-CIP in Distributed WikiTaaable





Tests Propagation in K-CIP