kolflow

Man-machine collaboration in continuous knowledge-construction flows

Pascal Molli

Université de Nantes

21/02/2011 – Kickoff Kolflow

UNIVERSITÉ DE NANTES

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

Nancy-Université

Université Henri Poincaré

Kolflow Context

* Social web is producing and updating a huge amount of information… **continuously**.
  * Ex : Wikipedia, Blogs, Twitter,
* Information need to be transformed **continuously** into knowledge in order to provide better search and better navigability on the web : semantic web vision
  * This transformation is costly (KDD Process)
  * Ex: Wikipedia -> DBPedia,

Kolflow Context

Calgary

Downtown Calgary.

Government	
- Mayor	Dave Bronconnier (Past mayors)
- Governing body	Calgary City Council
- Manager	Owen A. Tobert
Area [1]	
- City	726.50 km² (280.5 sq mi)
- Metro	5,107.43 km² (1,972 sq mi)
Elevation	1,048 m (3,438.3 ft)
Population (2006)[1]	
- City	988,193
- Density	1,360.2/km² (3,522.9/sq mi)
- Metro	1,079,310
- Population rank	3rd
- Metro rank	5th

```
http://en.wikipedia.org/wiki/Calgary
```

```
<http://dbpedia.org/resource/Calgary>
  dbpedia:native_name "Calgary" ;
  dbpedia:elevation "1048" ;
  dbpedia:population_city  "988193" ;
  dbpedia:population_metro "1079310" ;
  mayor_name
       dbpedia:Dave_Bronconnier ;
  governing_body
       dbpedia:Calgary_City_Council ;
 ...
```

DBPedia 2.0 in July 2007
DBPEdia 3.5 in March 2010
DBPedia 3.6 in October 2010
Release every 3 to 5 months… Time needed by a trained team to update knowledge from contents…

Kolflow context

* The transformation of Wikipedia to Dbpedia is not executed within Wikipedia as social tool or in a social semantic tool as Semantic Wikipedia.
  * The DBPedia team takes dumps from Wikipedia, filter noisy data, scan changes on templates, evaluate if DBPedia ontology needs to be updated, modify their mappings from template to ontology…
* Transformations are done **outside the social semantic web.**

Kolflow context

* Knowledge needs to be **continuously** revised with other knowledge sources
  * Ex : DBPedia interlinked with others Linked Data sources, interlinked with foreign language wikipedia
* Contents needs to be continuously updated to reflect knowledge updates
  * Ex: New property coming from interlinked Wikipedia -> new field in Wikipedia template, suggestions while editing templates …

Kolflow motivations

* Contents and Knowledge **co-evolve…**
* How to assist and improve the **continuous** transformation of contents into knowledge ?
* How to assist and improve the **continuous** revision and enrichment of knowledge ?

Kolflow approach

* **Build a social semantic space where humans collaborate with smart agents in order to produce knowledge understandable by humans and machines**
  * Humans are able to understand the actions of smart agents
  * Smart agent are able to understand and take into account actions of humans.
* **Man-Machine collaboration** should be the key to ensure **co-evolution** of contents and knowledge.

A social semantic space…

* Building a social semantic space means that the KDD process and knowledge revision are really executed within a single space (that can be distributed)…
* When done, contents, at all steps of transformations, knowledge at any step of elaboration and all the history of transformation or knowledge revision are available within this semantic space…
* Humans and machines can access and edit at any time any of these elements…(except history…)

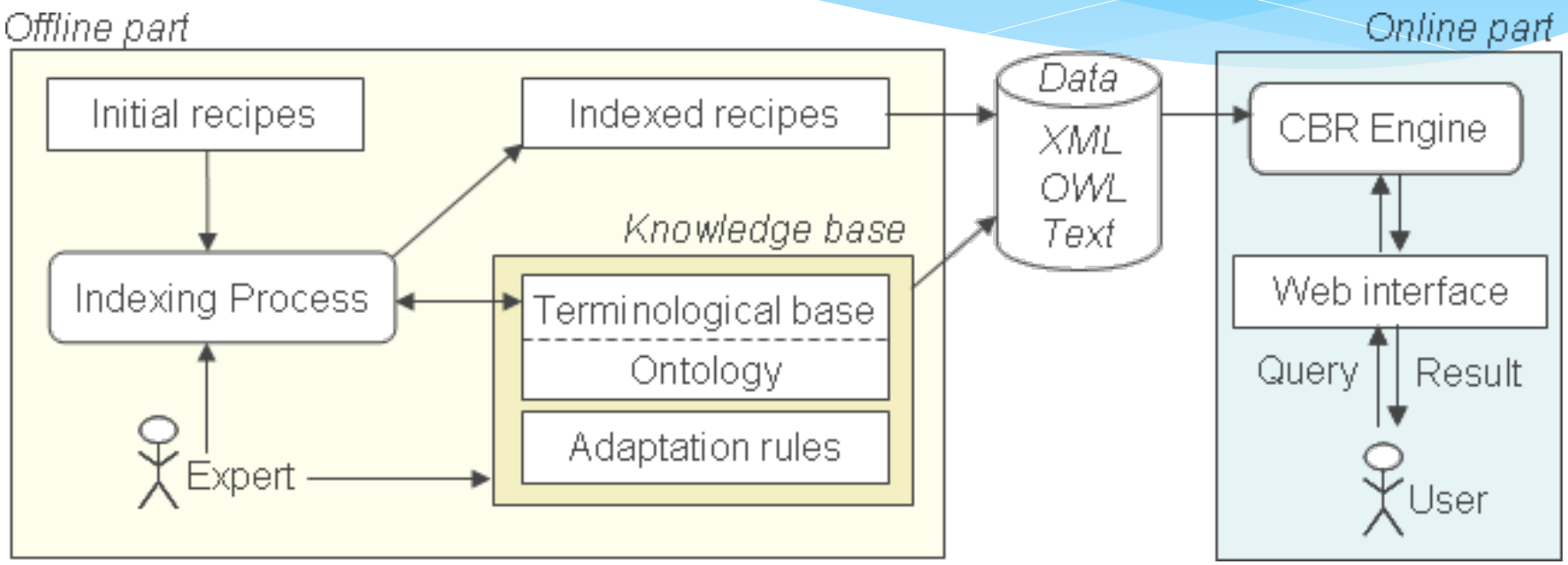
Kolflow & Taaable : Man-Machine Collaboration scenario

* The Taaable project aims to take recipes data, transform it into knowledge and use case based reasoning to adapt existing recipes to recipe requests
* Ex : I want a rhubarb pie, but there is no recipe for that in the knowledge base.
  * The CBR select an apple pie, replace apple by rhubarb and add some sugar to overcome rhubarb acidity

http://taaable.fr

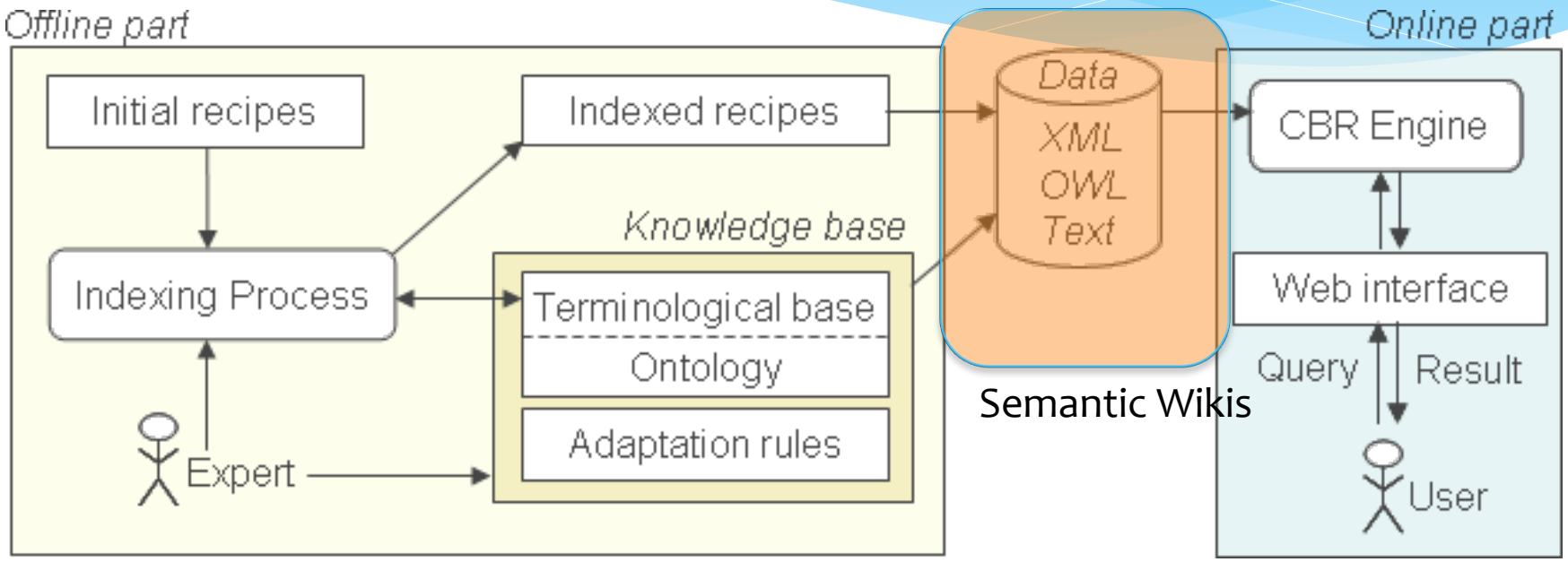
french_dish apple –sugar ?
Results on http://taaable.fr/

Kolflow/Taaable 1.0 (2008)



* The KDD Process is done offline (as DBPedia)
* The CBR execution is done on-line and sometimes give bad results…
* evolution of adaptation rule and ontology is not possible online
* So it it not possible to prevent Chocolate Lasagna online, ontology revision not supported.

Kolflow/Taaable 2.0 (2009)



* The KDD Process still done offline.
* The CBR access SPARQL Endpoint of Semantic wiki to access knowledge…
* Annotated recipes and ingredients ontology can be edited online

Log in

page | discussion | view source | history

Taaable

navigation

- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search

Go Search

toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Browse properties

Easy french rice

Ingredients

- 1 c Rice category:rice (1, c, ?, ?, ?)
- 2 tb Margarine category:margarine (2, tblsp, ?, ?, ?)
- 1 cn (8 oz) sliced water chetnuts, drained category:water (1, can, sliced,drained, (8 oz), chetnuts)
- 1 cn (8 oz) mushrooms (liquid reserved) category:mushroom (1, can, ?, (8 oz)(liquid reserved), ?)
- Enough water to equal 1 1/3 category:water (?, ?, ?, ?, enough to equal 1 1/3)
- C with mushroom liquid category:liquid (?, c, ?, ?, with)
- 1 cn (10 3/4 oz) French onion soup, defated category:onion (1, can, ?, (10 3/4 oz), french)

Preparation

- Combine all ingredients in a 2 quart casserole and mix well. Cook, covered, in a 350 degree oven for one hour. remove from oven and let rest for 10 minutes.

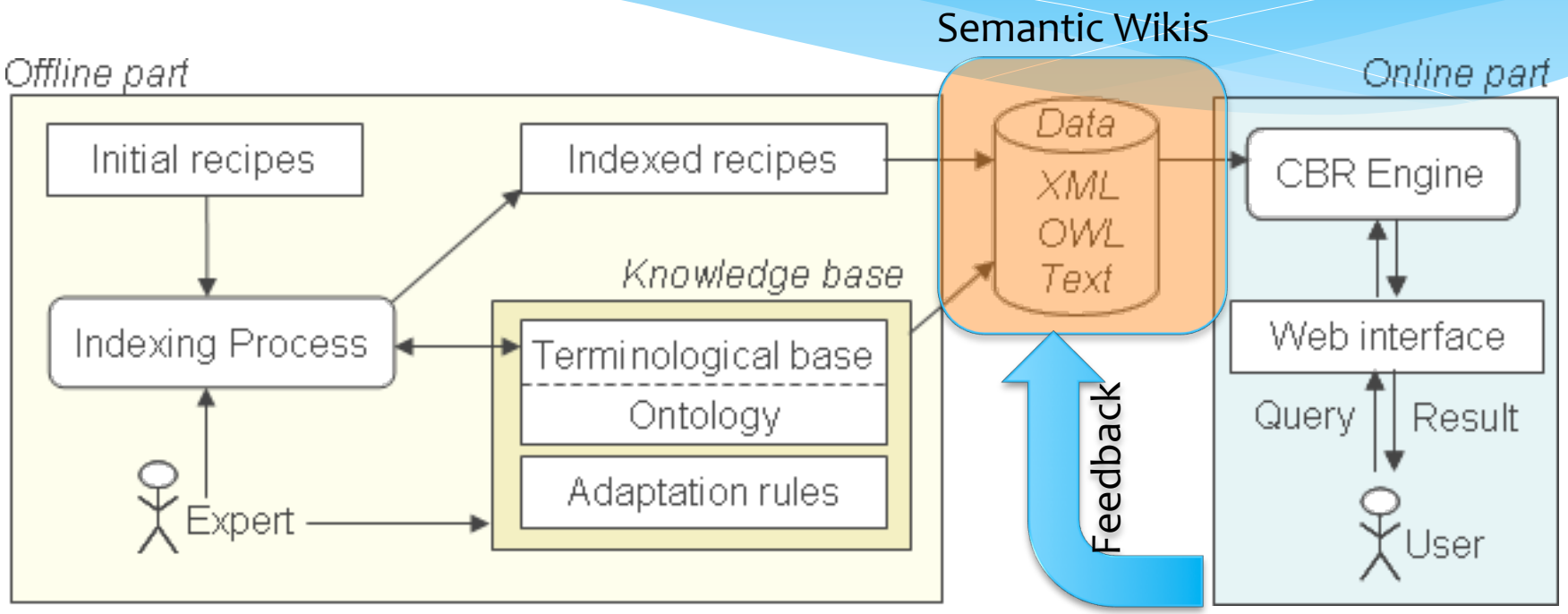
Formal Preparation

- slice: chetnut_0(chetnut) => chetnut_1(chetnut)
- drain: chetnut_1(chetnut) => chetnut_2(chetnut)
- reserve: mushroom_0(mushroom) => mushroom_1(mushroom)
- reserve: mushroom_0(mushroom) => mushroom_1(mushroom)

Easy french rice preparation graph

Categories: Unachievable formal preparation | Recipe | RecipeCompulsary | Rice dish | French dish | Side dish

Kolflow/Taaable 3.0 (2009)

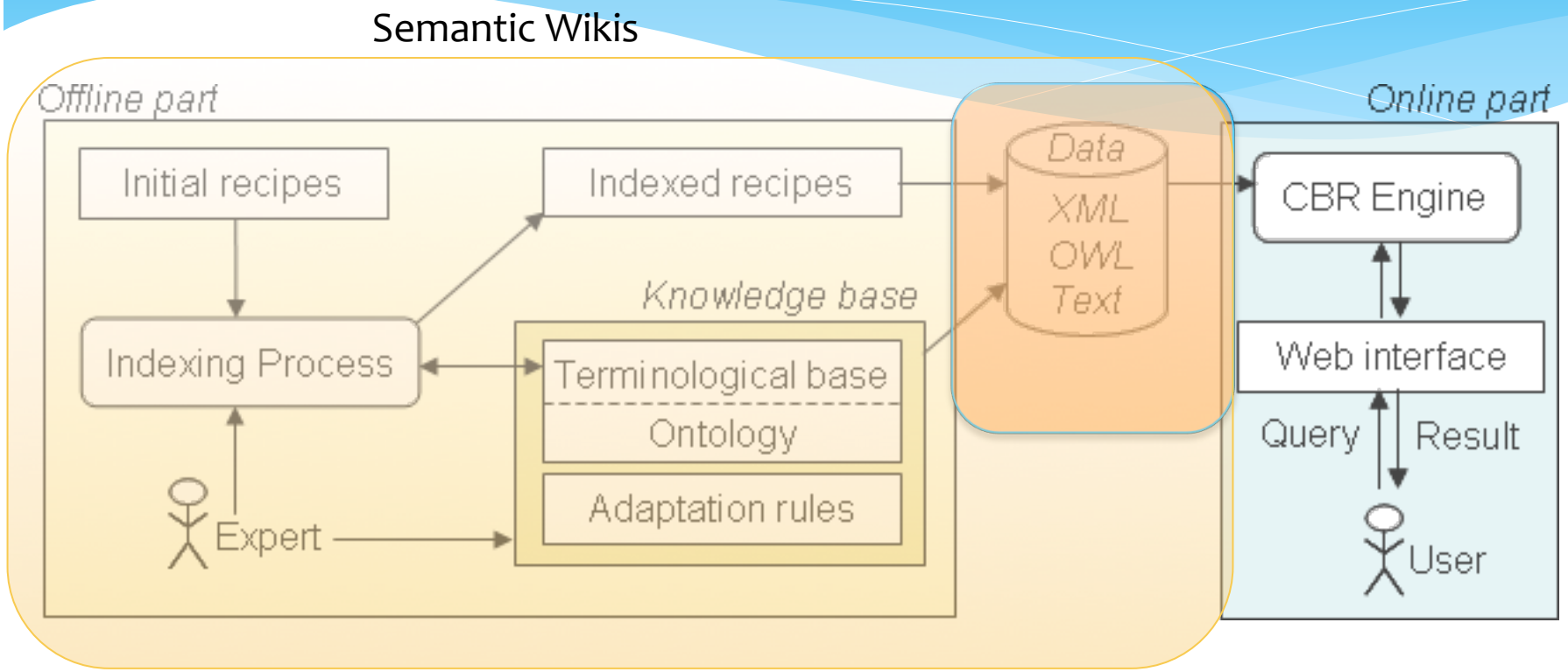


* The KDD Process still done offline.
* Feedback of users can modify the knowledge base, interaction with CBR results

Taaable 3.0 (2010)

french_dish apple –sugar ?
Results on http://taaable.fr/

Taaable 4…

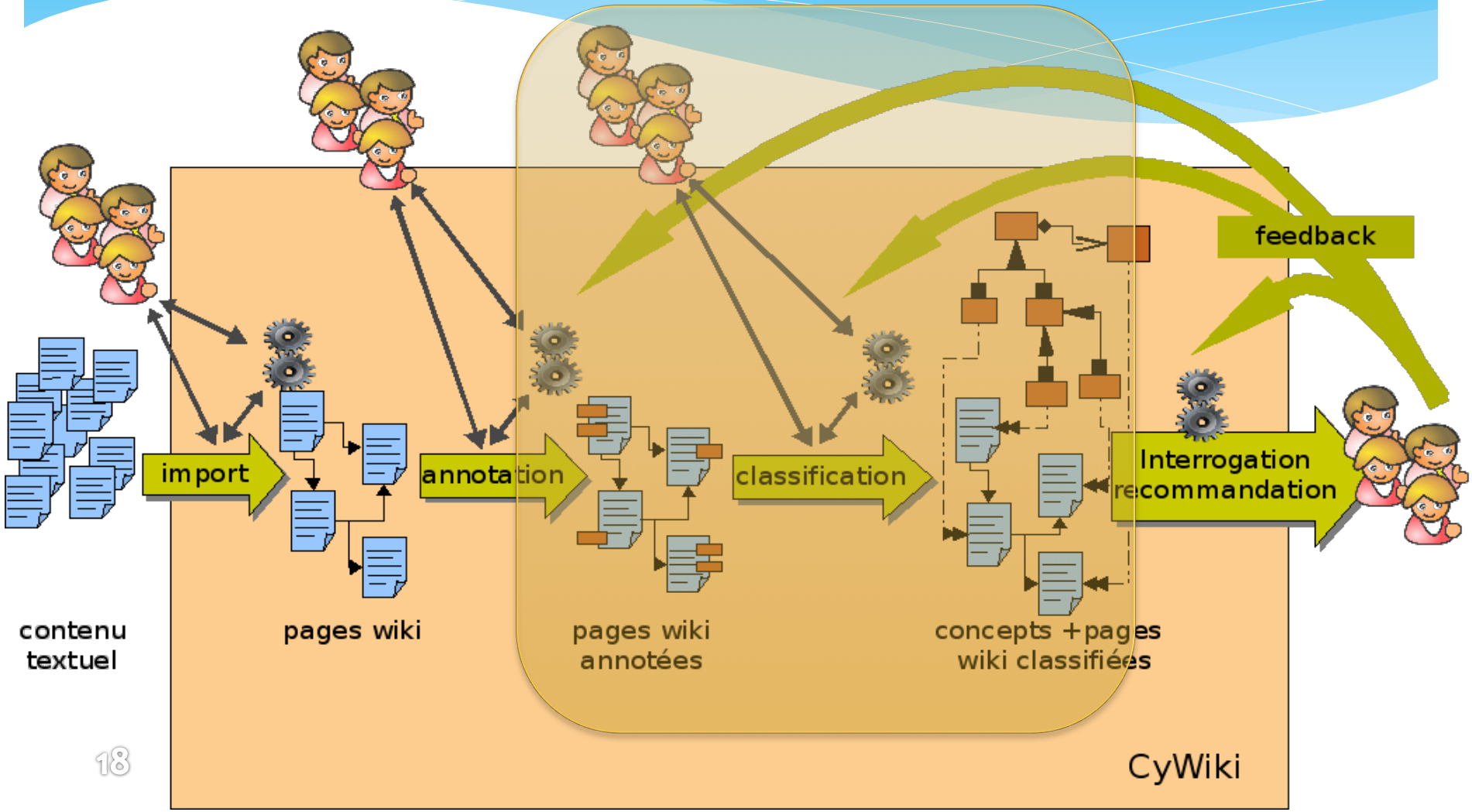


* Try to make the KDD process within semantic wiki…

Kolflow Taaable scenarii : raised issues

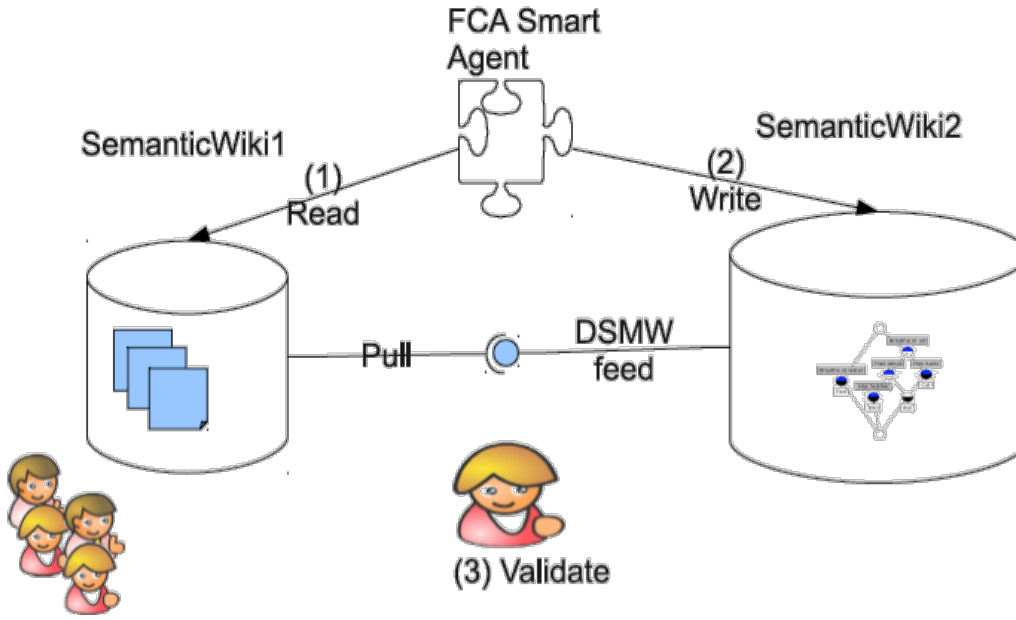
* Taaable illustrates a Man-Machine collaboration scenario:
  * Contents and Knowledge are available online for man and machine. They can edit anything at anytime
  * Feedback of users updates knowledge
  * Humans can annotate recipes, add new recipes, update ingredients…
* Taaable also illustrates system regression issues:
  * Ex : Changing an adaptation rule online can break the CBR, changing the ingredient ontology can force CBR to give bad results

Kolflow 2nd scenario : Cywiki Score/orpailleur



Human-Machine Collaboration Process

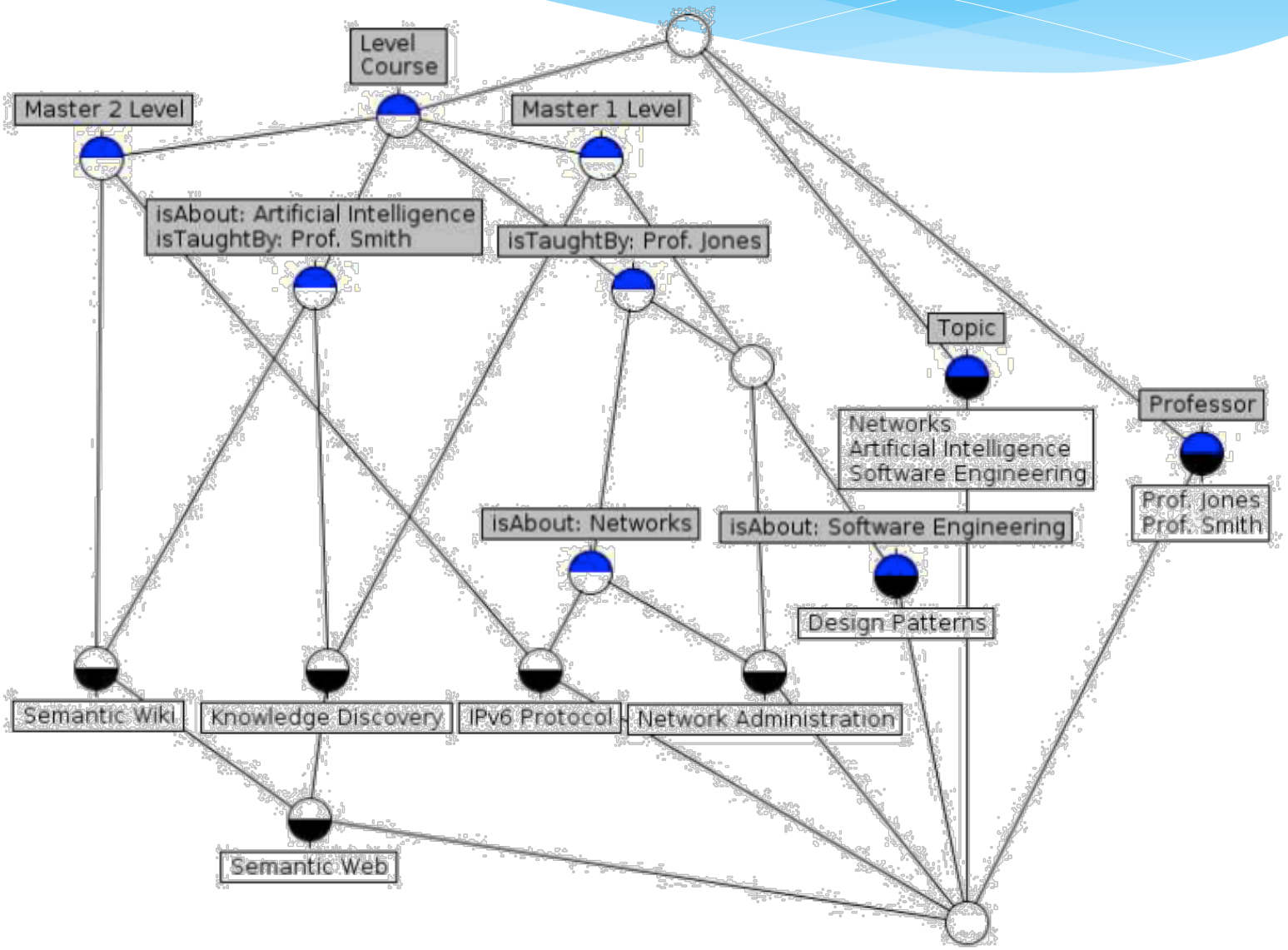
* Semantic Wiki1 initial Wiki (usual wiki), from Semantic Wiki1, FCA agent creates the lattice in Semantic Wiki2
* Humans correct, refine the content of Semantic Wiki2, push the content of Semantic Wiki2 of a push feed
* Administrator of Semantic Wiki1 can pull validated modifications from Semantic Wiki2 into Semantic Wiki1



Human-machine Collaboration for Enriching Semantic Wikis using Formal Concept Analysis – Alexandre Blansché, Hala Skaf-Molli, Pascal Molli and Amedeo Napoli - SemWiki2010

Scenario: Initial Semantic Wiki

* Category:
  * Professor, Topic, Course, Level
  * Level has two subcategories: Master1 Level and Master 2 Level
* Property:
  * isTaughtBy(Course, Proessor) , isAbout(Course, Topic)


Level
Course
Master 2 Level
Master 1 Level
isAbout: Artificial Intelligence
isTaughtBy: Prof. Smith
isTaughtBy: Prof. Jones
Topic
Professor
Networks
Artificial Intelligence
Software Engineering
Prof. Jones
Prof. Smith
isAbout: Networks
isAbout: Software Engineering
Design Patterns
Semantic Wiki
Knowledge Discovery
IPv6 Protocol
Network Administration
Semantic Web

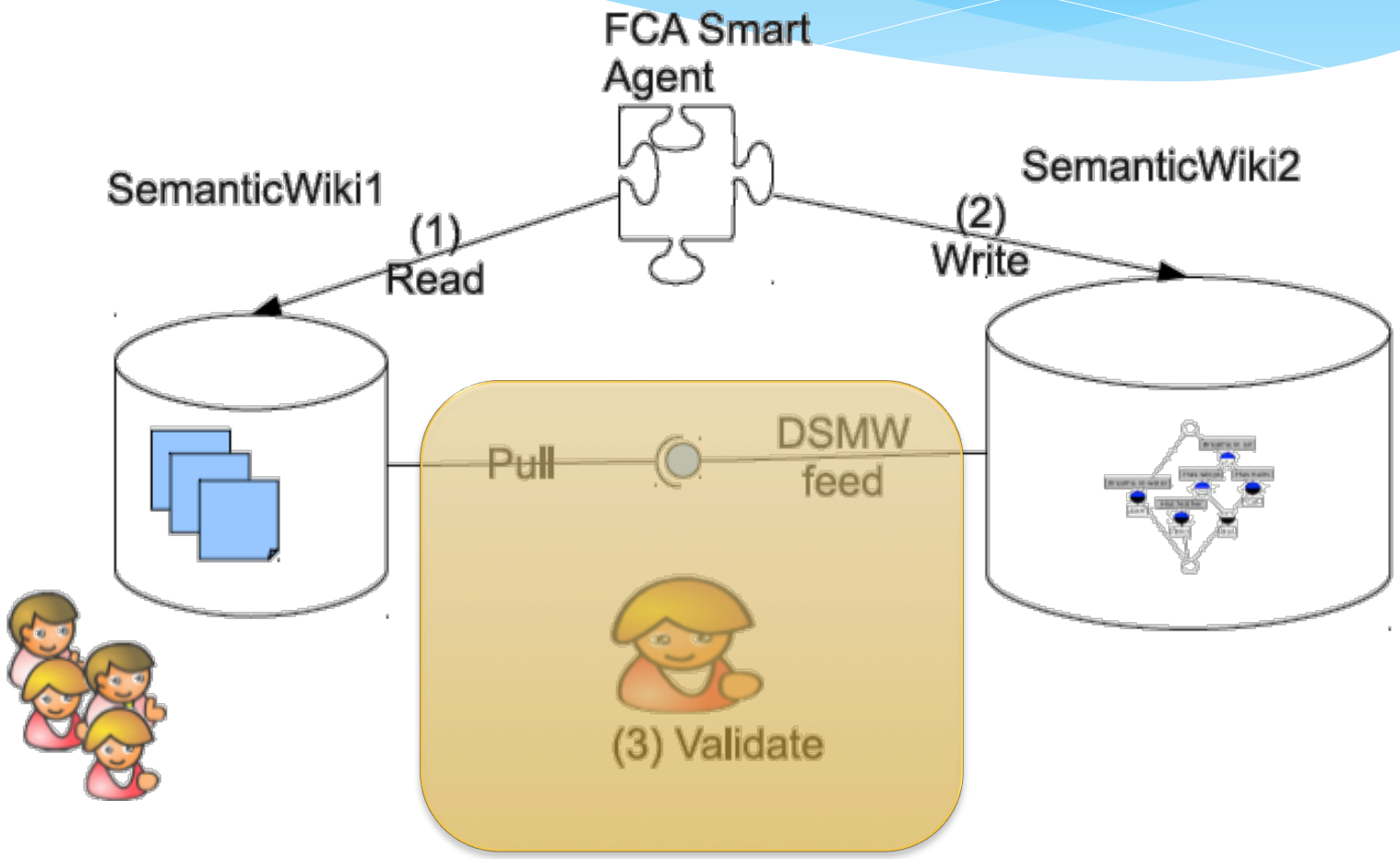
Mapping Categories and Lattice concepts

* If a concept maps a single category then the category will be preserved
* If a concept maps two categories or more then the categories are merged
* If a concept does not map any category then a new category will be created ..

Enriched wiki

* Category:
  * Professor, Topic, Course, ~~Level~~
  * ~~Level~~ Course has two subcategories: Master1 ~~Level~~ Course and Master 2 ~~Level~~ Course
  * Artificial Intelligence Course SubCat of Course
  * Master 1 Artificial Intelligence Course subCat of Master 1 Course and subCat of Artificial Intelligence
  * Etc ..
* We have 14 categories instead of 4 categories in the initial wiki

Enrichment and validation



Kolflow and cywiki: raised issues

* Cywiki example illustrates Man-Machine Collaboration
  * Users annotates pages and propose first classification through a wiki
  * A smart agent (FCA) propose an revision of the classification
  * Users validate (or not not)
* It also show severe issue:
  * How Users can understand result of FCA ?
  * How to process a second FCA Loop i.e. users refuse some class and FCA regenerates them ? Ex of Divergence in MMC…

Kolflow Issues

* How to ensure that Man-Machine Collaboration ensures non-regression of the system ? How to understand how Man-Machine collaboration works ?
  * Man-Machine Collaboration for humans : understand actions from machines
  * Man-Machine Collaboration for machines : understand actions from humans

Man-Machine Collaboration for humans issues

* How to make formal knowledge and its evolution accessible, usable, editable and understandable by humans ?
* So they can observe, control, evaluate the output of Smart Agents ?

Man-Machine Collaboration for machines issues

* How to take into account the unpredictable behavior of humans?
* How to make reasoning with the risk of uncertainty and inconsistency ?
* How automated reasoning's can adapt their behaviors and results according to feedback of users ?

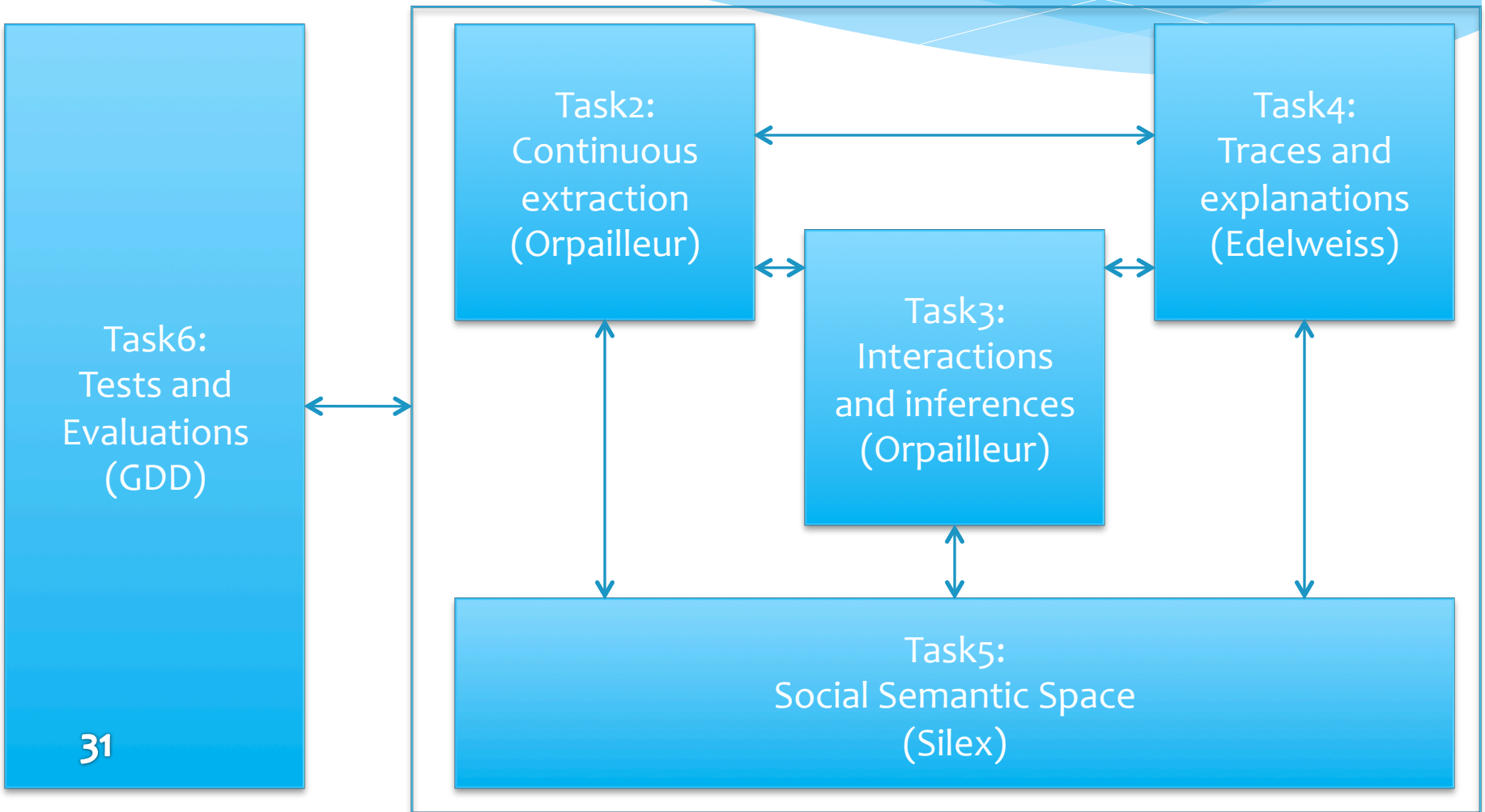
Kolflow expected results

* Deliver man-machine collaboration scenarii and some reference corpus. These scenarii drive the project and evaluate the overall progress
* Build a social semantic space based on distributed semantic wikis. This space behaves as blackboard for Man-Machine collaboration
* Make histories of knowledge building understandable by man and machine. Make MMC histories a source of knowledge.

Kolflow expected results

* Make automated reasoning understandable by humans. Smart agents must explain what they did and why they did that
* Manage inconsistencies generated by man-machine collaboration by allowing interactive reasoning

Task coordination (GDD)



Task 1 : Coordination
GDD (Nantes)

* The web site is started and is announced…
  * http://kolflow.univ-nantes.fr
* Encourage collective publications
  * Ph.D. Thesis co-advisement.
  * Ph.D. Students will visit other partners…

Task 1 : Coordination

* Develop relationships with other ANRs Kolflow has links with others ANRs
  * Concordant (2010-2013) is an ANR Project dedicated to the development of Commutative Replicated Data Types. CRDTs are used to build a Social Semantic Space in Kolflow
  * STREAMS (2010-2013) is an ANR Project dedicated to Real-time web. CRDTs are also used, Kolflow can give experimental results to STREAMs.
  * DataLift (2010-2013) is an ANR Project dedicated to the transformation of data to linked data.

Task 6 : Test and Evaluation GDD (Nantes)

* Deliver man-machine collaboration scenarii and some reference corpus. These scenarii drive the project and evaluate the overall progress
* Taaable project as corpus provider. We have data from:
  * Computer Cooking Contest 2008 : Taaable1 vice-champion
  * CCC 2009 : Taaable2 vice-champion
  * CCC 2010 : Taaable3 champion.
* Taaable 4 an opportunity to collect first corpus for Kolflow!

Task6 deliverables

* T0+6 Man-machine collaboration scenarios
* T0+12 Man-machine collaboration scenarios progress report
* T0+24 Man-machine collaboration scenarios progress report
* T0+36 Man-machine collaboration scenarios progress report

Task 2 : Continuous Extraction Orpailleur (Nancy)

* Extracting knowledge units from texts using data mining
  * Automatic knowledge extractor takes data from social semantic space and take care of history (user feedback)
  * Users are able to edit the result of automatic knowledge extraction through a wiki interface.
* Ex : Cywiki scenarii, how FCA tool is aware of user change ? How to make FCA processing with users constraints ?

Task 2 : Continuous extraction subtasks

* Collecting data :
  * T0+6 Building a corpus for experimenting continuous knowledge extraction. (should be the result og CCC2011)
* Formal methods for knowledge extraction :
  * T0+12 Integrating a knowledge extraction system from texts in a semantic wiki. methodology, modules, and experiments.
* Continuous extraction of knowledge :
  * T0+18 Specification of a continuous knowledge extraction system
  * T0+36 Final report.
* Semantic annotation
  * T0+36 Dynamic semantic annotation in a semantic wiki: definitions and specifications

Task 3 : Interaction and Inferences Orpailleur (Nancy)

* Design and experiment strategies for dealing with semantic inconsistencies
* Merging 2 consistent knowledge bases modified from the same source can be inconsistent.
* How to repair interactively the resulting base ?
* Subtasks:
  * Revision and merging of ontologies
  * Human-Computer Interaction for managing consistency
  * Interactive reasoning with a globally inconsistent family of ontologies.
* Ex : In Taaable scenario, a user introduced a inconsistency in ingredient ontology. How to detect and interactively repair ?

Task3 : Interaction and Inferences deliverables

* T0+12 : Algorithms and architecture of the interactive semantic inconsistency managing system
* T0+24 : prototype and tests of the interactive semantic inconsistency managing system
* T0+36 : prototype of the extension integrating management of ”human-machine” and ”logical” inconsistencies

Task 4 : Traces and explanations Edelweiss (Sophia)

* Design and experiment tracing and explanation approaches
* Subtasks:
  * Opening query solving problems :
    * Explain the execution of a distributed query
    * Express search strategy and inferences, performances, errors
  * Alter-ego assistant :
    * Explain knowledge available in the system by using interaction traces :
    * T0+18 : Test and evaluation of the alter ego assistant with regard to the scenarios

Task4 : Traces and explanation deliverables

* T0+6 : State of the art on the different topics addressed in this task
* T0+10 : Algorithm to explain basic query mechanisms
* T0+15 : Algorithm to provide performance and errors indicators
* T0+20 : Algorithm to ontology-based processing
* T0+26 : Algorithm to suggest queries and changes to queries

Task 5 : Social Semantic Space Silex (Lyon)

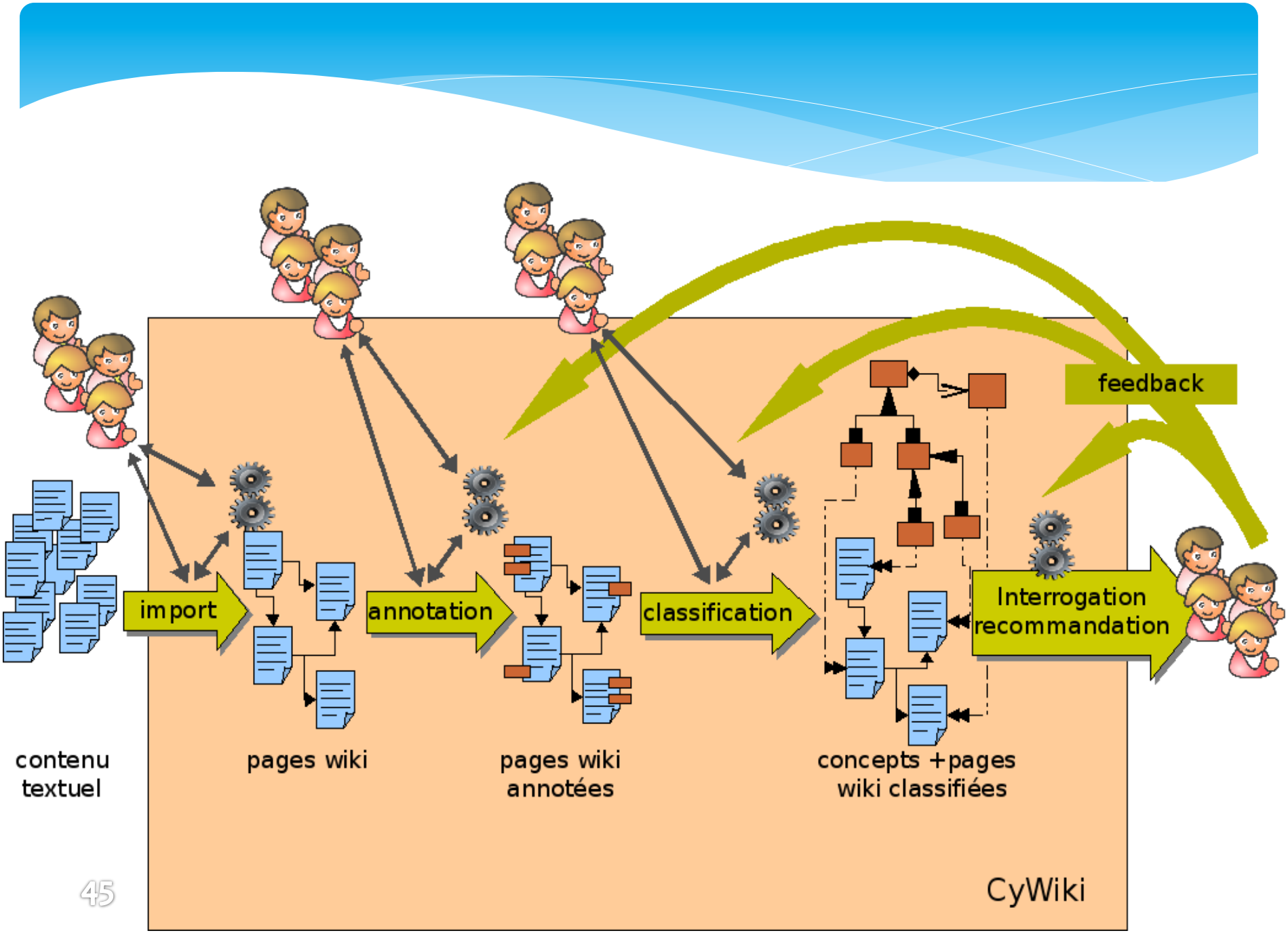
* Design and experiment a of social semantic space
  * **Distributed Semantic Wikis :** extending semantic wikis to a distributed context. Using Commutative Replicated Data Types for semantic web.
  * **Distributed semantic Queries** : Indexing and publishing content of bases, decompose and route sub queries, handle conflict detection between sources
  * **Distributed alter ego assistant :** allow alter-ego to communicate and to exchange information…

Task5 : Social semantic space

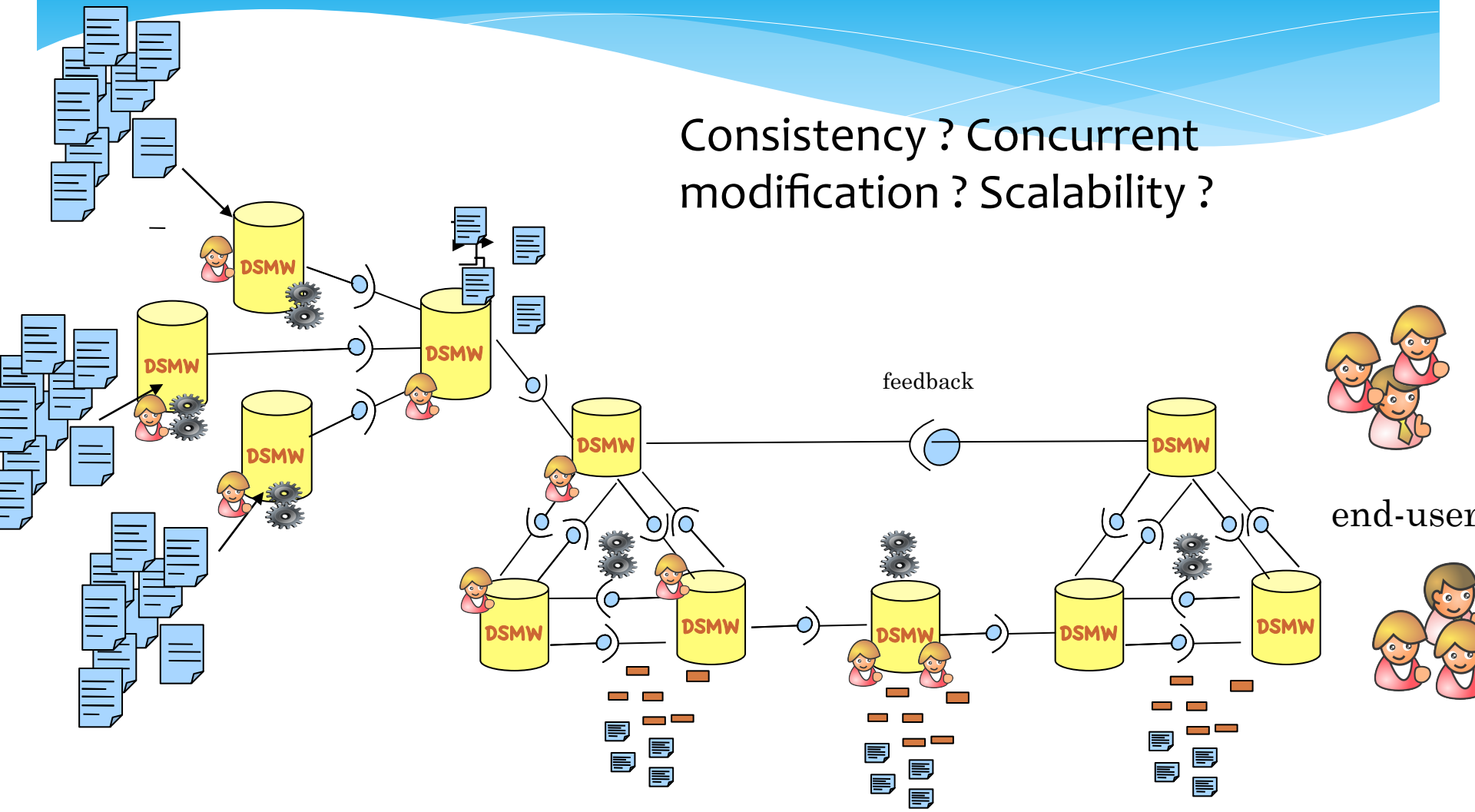
* a state of the art on semantic wikis and distributed semantic wikis (T0+6);
* Algorithms and ontology for multi-synchronous edition in a distributed semantic wiki (T0+18);
* Process language and process machine for man-machine coordination in a distributed semantic wiki, (T0+34);
* validation by using the scenarios of tasks 6 (T0+36)

Task5 : Social Semantic Spaces deliverables

* T0+6 State of the art on the three topics studied (Distributed Semantic Wikis, distributed querying distributed alter-ego assistant),
* T0+12 Algorithm to summarize the content of a base
* T0+18 Algorithms and architecture for a distributed semantic wiki and a distributed alter-ego assistant
* T0+28 Algorithm to distribute query-solving
* T0+34 Algorithm to explain and document the distributed solving, Process language and machine for man-machine coordination
* T0+36Algorithm adapted to conflict detection
* T0+36Test and evaluation according to the project scenarios


feedback
import
annotation
classification
Interrogation
recommandation
contenu
textuel
pages wiki
pages wiki
annotées
concepts +pages
wiki classifiées
CyWiki

Consistency ? Concurrent modification ? Scalability ?

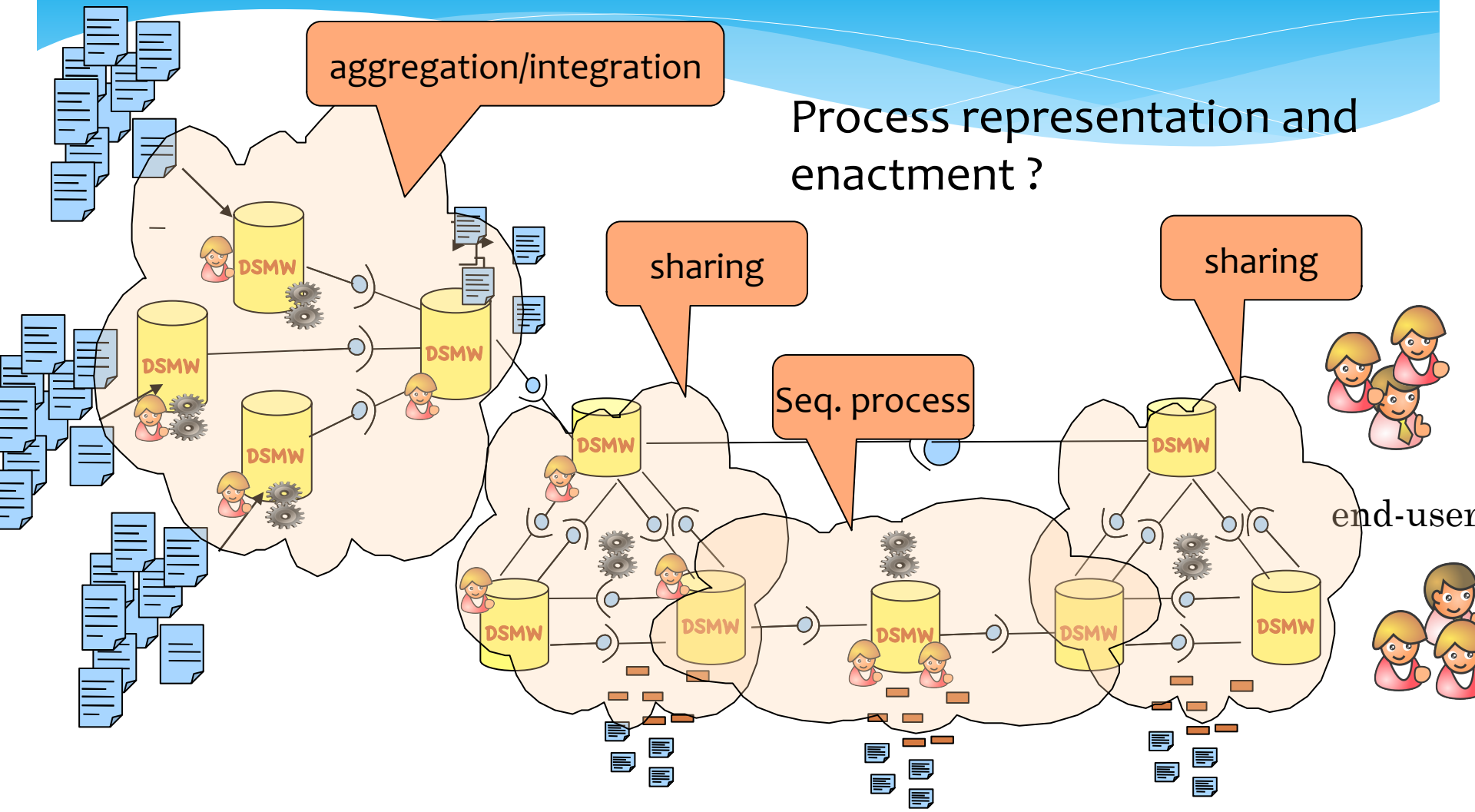


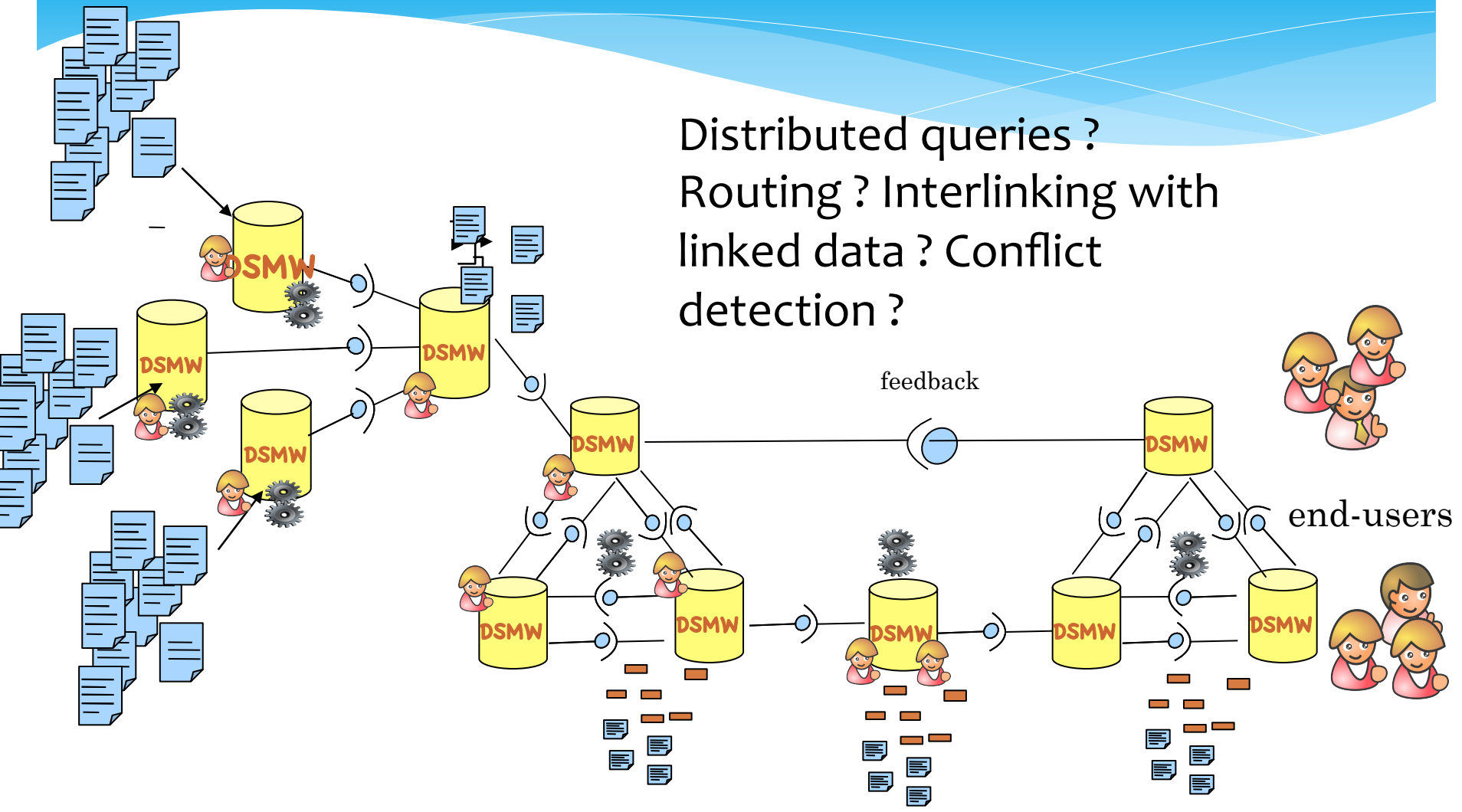
Importation

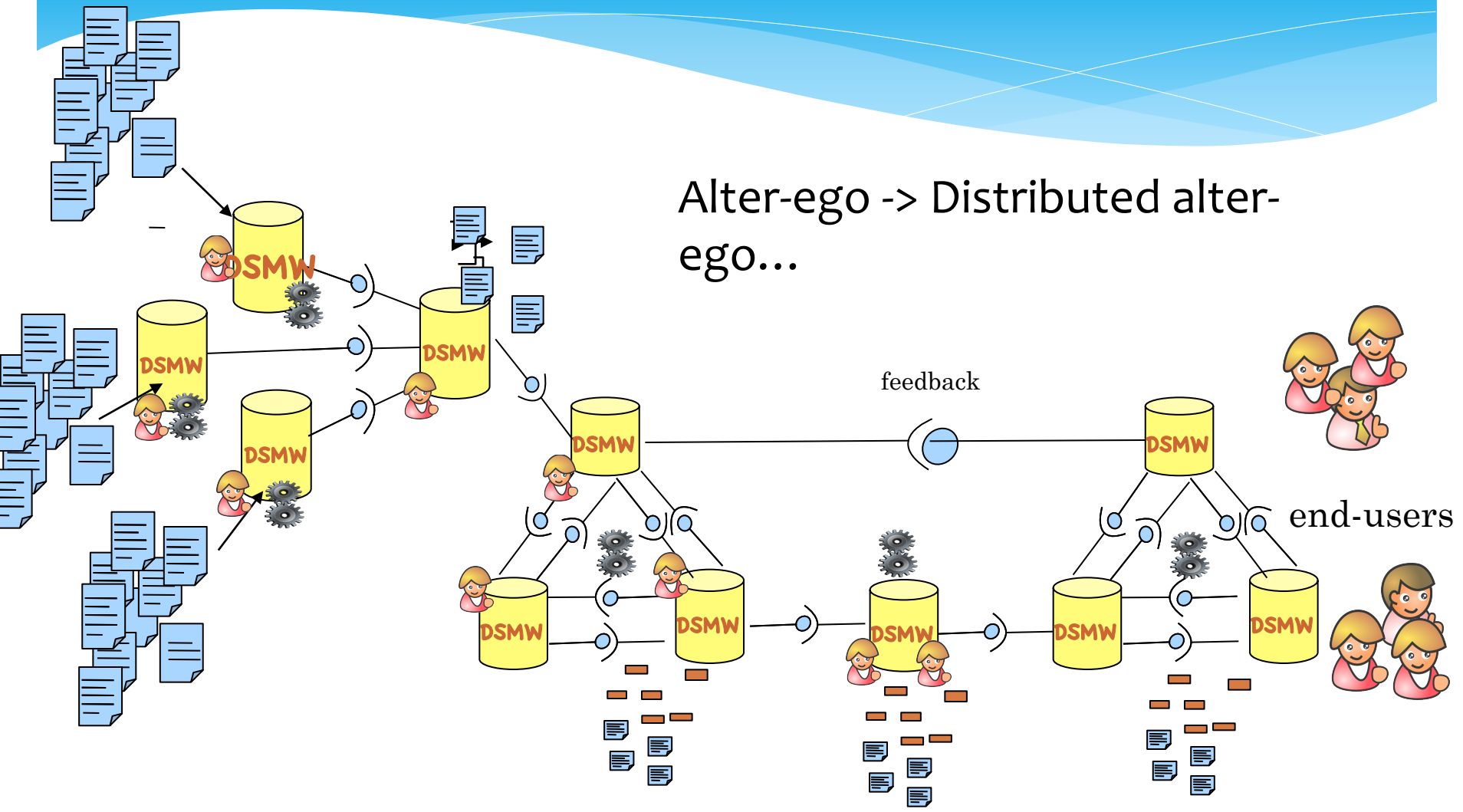
Annotation/ Classification

Validation

Query/Recommendation


aggregation/integration
Process representation and enactment ?
sharing
sharing
Seq. process
DSMW
end-user
Importation
Annotation/ Classification
Validation
Query/Recommendation


Distributed queries ? Routing ? Interlinking with linked data ? Conflict detection ?
feedback
DSMW
end-users
Importation
Annotation/ Classification
Validation
Query/Recommendation


Alter-ego -> Distributed alter-ego…
DSMW
feedback
end-users
Importation
Annotation/
Classification
Validation
Query/Recommendation

Ressources

* Ph.D. Co-advised by 2 partners, Ph.D. should spend 2 months with other partners during the Ph.D.
* 3 Phds:
  * Man-machine collaboration processes for continuous knowledge extraction using distributed semantic wikis (GDD)
  * Solving problems upstream and downstream of a distributed query on the semantic web (Edelweiss)
  * Inferences and interactions for dealing with semantic inconsistencies (Orpailleur/Silex)
* 2 postdoc :
  * Alter-ego assistant (Silex)
  * Evaluation (Orpailleur)

Summary

* Kolflow : managing co-evolution of contents and knowledge
* Social semantic tools open the way to Man-machine collaboration for continuous content to knowledge transformation and continuous knowledge enrichment.
* Kolflow aims to develop approaches, models and algorithms to understand and to enact safe man machine collaboration.