WWW.IARIA.ORG

PANEL IMMM/ICIW

Bridging the Web of Data with the Web of Documents

→ Information

Information: discovery, interpretation (semantic, ontology, formal/non-formal)

Nature: text, photos, voice

Representation: text, tables, graphics, functions

Context
BIG/SMALL Data
Linked Data
Open Data

Panel

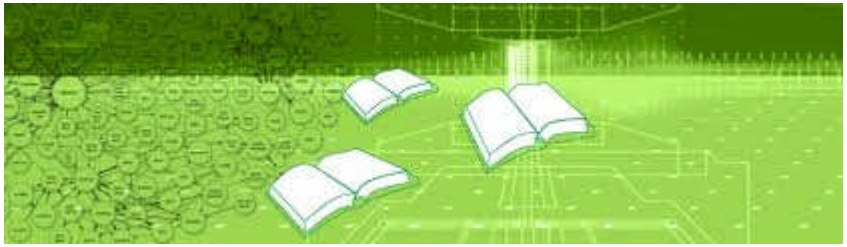
- **Moderator**
  > Petre Dini, Concordia University, Canada/ China Space Agency Center, China

- **Panelists**
  > Awatef Hicheur Cairns, Altran Research, France
  Is it more appropriate to use process discovery techniques or sequence mining techniques to extract important patterns in students' behaviors and to make recommendations for the best learning paths for students to follow

- **> António Teixeira, University of Aveiro, Portugal**
  Applications of extraction of semantic information from publicly available documents regarding Health and local government

- **> Athanassios Jimoyiannis, University of Peloponnese, Greece**
  Participative-learning affordances of Social Media and Web 2.0, e.g. Participatory Web, Open Web, Collaborative Web, and Social Web.
  A new pedagogical and learning context, which combines formal, non-formal and informal features; e.g., autonomous and self-directed learning, collaborative learning and ubiquitous learning, networking and community learning.

Bridging the Web of Data with the Web of Documents



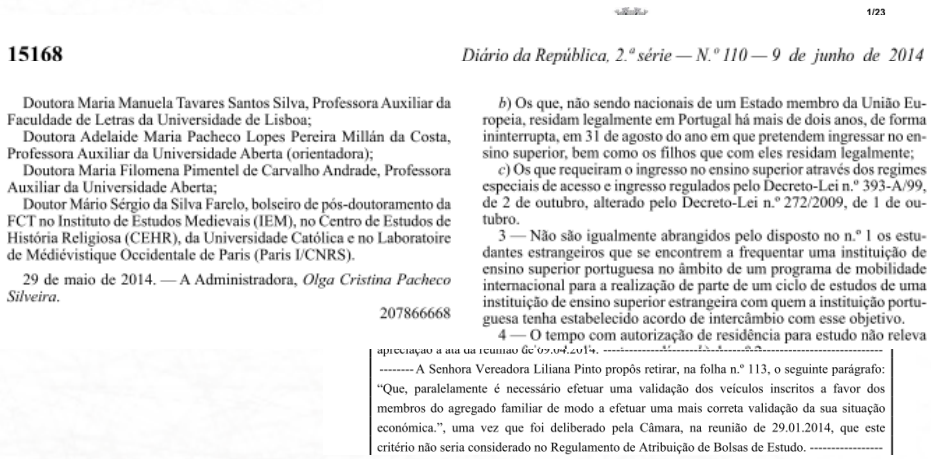
The need and applications of extraction of semantic information from publicly available documents

(with concrete examples for Health and eGov)

António Teixeira
Dep Electronics Telecom. & Informatics/IEETA
University of Aveiro

Government documents

- Governments produce large amounts of documents
  - At the several levels (local, central …)
  - Of many types:
    - Laws
    - Regulations
    - Minutes of meetings
    - Deliberations
    - …



15168 Diário da República, 2.ª série — N.º 110 — 9 de junho de 2014

Doutora Maria Manuela Tavares Santos Silva, Professora Auxiliar da Faculdade de Letras da Universidade de Lisboa;
Doutora Adelaide Maria Pacheco Lopes Pereira Millán da Costa, Professora Auxiliar da Universidade Aberta (orientadora);
Doutora Maria Filomena Pimentel de Carvalho Andrade, Professora Auxiliar da Universidade Aberta;
Doutor Mário Sérgio da Silva Farelo, bolseiro de pós-doutoramento da FCT no Instituto de Estudos Medievais (IEM), no Centro de Estudos de História Religiosa (CEHR), da Universidade Católica e no Laboratoire de Médiévistique Occidentale de Paris (Paris I/CNRS).

29 de maio de 2014. — A Administradora, *Olga Cristina Pacheco Silveira.*

207866668

b) Os que, não sendo nacionais de um Estado membro da União Europeia, residam legalmente em Portugal há mais de dois anos, de forma ininterrupta, em 31 de agosto do ano em que pretendem ingressar no ensino superior, bem como os filhos que com eles residam legalmente;
c) Os que requeiram o ingresso no ensino superior através dos regimes especiais de acesso e ingresso regulados pelo Decreto-Lei n.º 393-A/99, de 2 de outubro, alterado pelo Decreto-Lei n.º 272/2009, de 1 de outubro.

3 — Não são igualmente abrangidos pelo disposto no n.º 1 os estudantes estrangeiros que se encontrem a frequentar uma instituição de ensino superior portuguesa no âmbito de um programa de mobilidade internacional para a realização de parte de um ciclo de estudos de uma instituição de ensino superior estrangeira com quem a instituição portuguesa tenha estabelecido acordo de intercâmbio com esse objetivo.

4 — O tempo com autorização de residência para estudo não releva

apreciação a ata da reunião de 09.04.2014. ----------

-------- A Senhora Vereadora Liliana Pinto propôs retirar, na folha n.º 113, o seguinte parágrafo: "Que, paralelamente é necessário efetuar uma validação dos veículos inscritos a favor dos membros do agregado familiar de modo a efetuar uma mais correta validação da sua situação económica.", uma vez que foi deliberado pela Câmara, na reunião de 29.01.2014, que este critério não seria considerado no Regulamento de Atribuição de Bolsas de Estudo. ----------

Problem(s)

- Relevant information produced in (written) natural language
- How to make available as really usefull information for citizens?
- How to make it usable by machines?

eGov

Face to face interaction:

Working hours

In loco

Relevant Information

Keyword information search:

Anytime

Anywhere

Difficult to find relevant data in large result set

Search that understands language:

Anytime

Anywhere

Relevant Information

Health

- More and more healthcare institutions store vast amounts of information
  - about users, procedures, and examinations, as well as the findings, test results, and diagnoses respectively.
- Other institutions, such as the Government, increasingly disclose health information on varied topics of concern to the public writ large.
- Health research is one of the most active areas, resulting in a steady flow of publications reporting on new findings and results.
- In recent years, the Internet has become one of the most important tools to obtain medical and health information.

Examples of data - Health

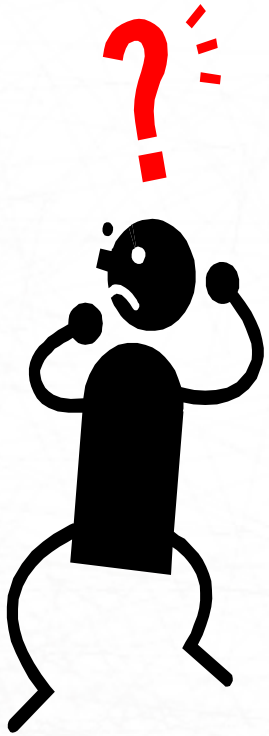
- Public
  - Medical journals
  - Newspapers
  - Masters and PhDs
  - Books
  - …
- Not public
  - Hospital Recordings
  - Doctors documents
  - …

Problem(s)

- General search engines do not allow the end-user to obtain a clear and organized presentation of the available health information.
  - Instead, it is more or less of a hit or miss, random return of information on any given search.
  - In fact, medicine-related information search is different from other information searches, since users often use medical terminology, disease knowledge, and treatment options in their search (Wang et al. 2012).
- Much of the information that would be of interest to private citizens, researchers, and health professionals is found in unstructured documents.

The Gap between Data/Docs and Humans (and Machines)

DOCUMENTS



The need for semantic search

- Efficient access to this information implies the development of search systems capable of handling the technical lexicon of the domain area, entities such as drugs and exams, and the domain structure.

- Such search systems are said to perform semantic search as they base the search on the *concepts*
- Semantic search maintains several advantages over search based on surface methods
  - such as those that directly index text words themselves rather than underlying concepts.

- Three main advantages of concept-based search are:
  - they usually produce smaller sets of results, as they are able to identify and remove semantically duplicated results and/or semantically irrelevant results;
  - they can integrate related information scattered across documents; frequently answers are obtained by compounding information from two or more sources;
  - they can retrieve relevant results even when the question and answer do not have common words, since these systems can be aware of similar concepts, synonyms, meronyms, antonyms, etc.

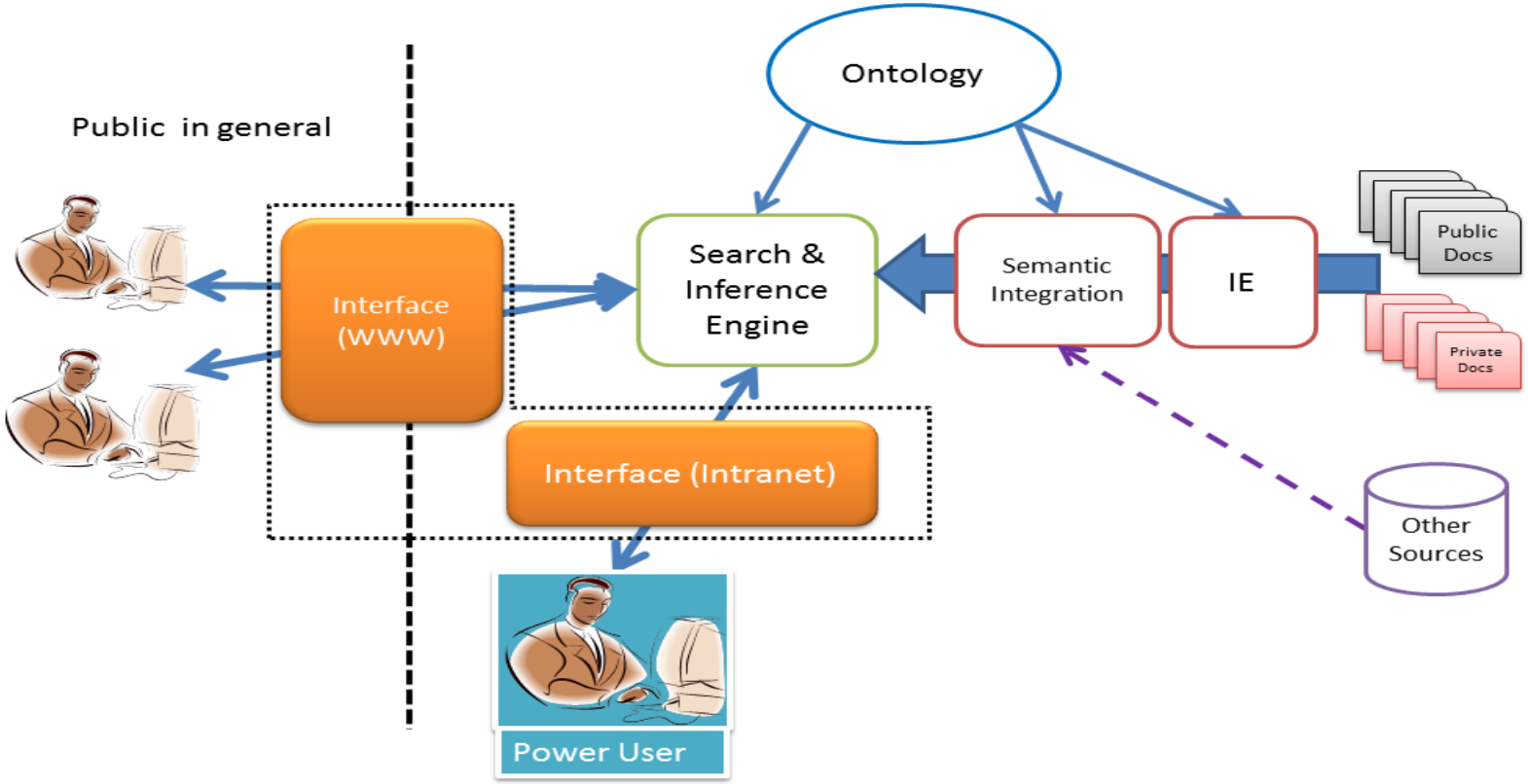
Requirements …

- But the semantic information must be derived from the documents…
  - Using techniques from Information Extraction, NLP …
- Obtained information must be easy to explore
- Complex and more natural queries must be possible

Examples

From my
Recent Work

Example - HealthInX



Chapter: Online Health Information Semantic Search and Exploration:
Reporting on Two Prototypes for Performing Information Extraction on both a Hospital Intranet and on the World Wide Web

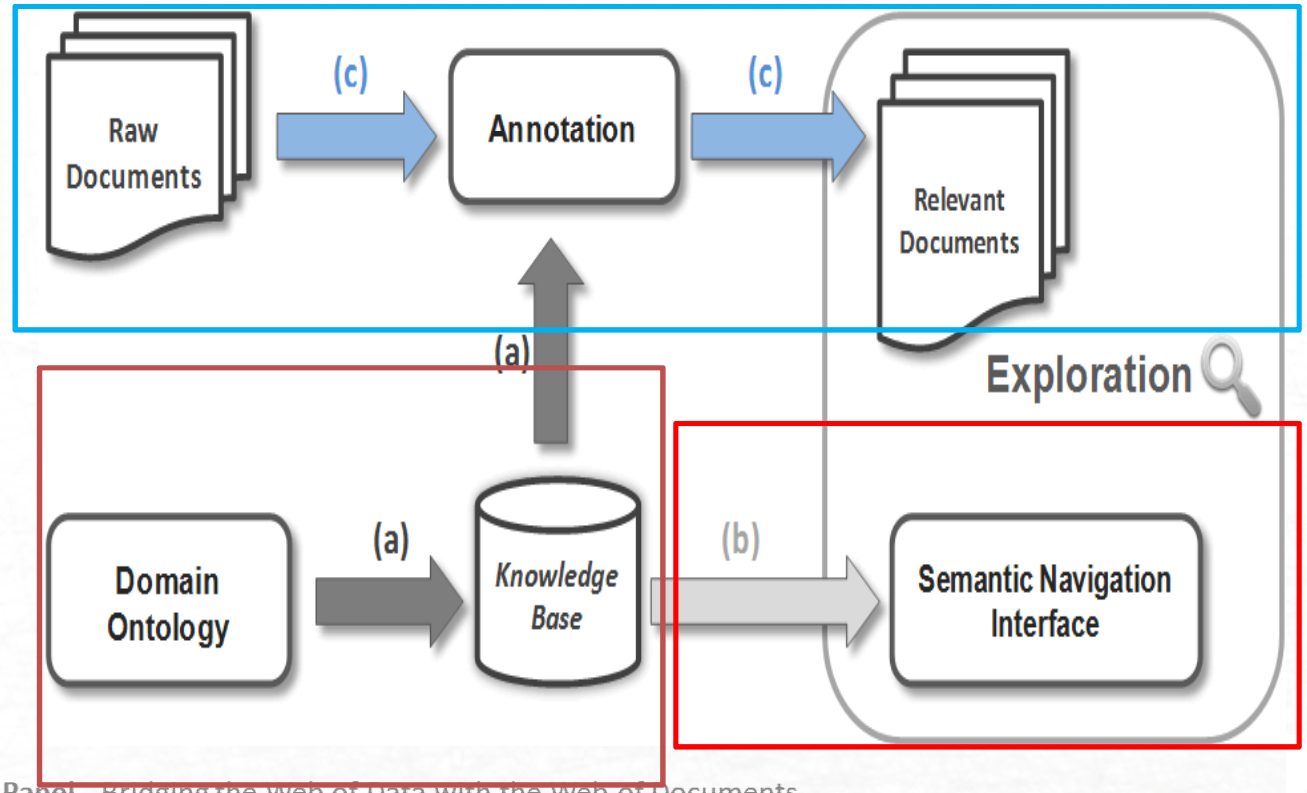
Ontology-based Health information Search (in project World Search)

System consists of three principal modules:

a) semantic annotation;

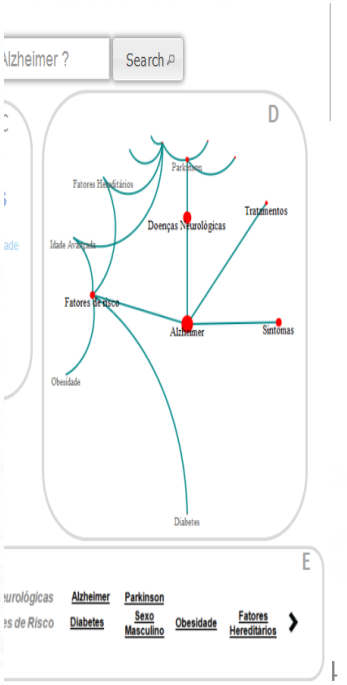
b) semantic navigation interface; and

c) search platform

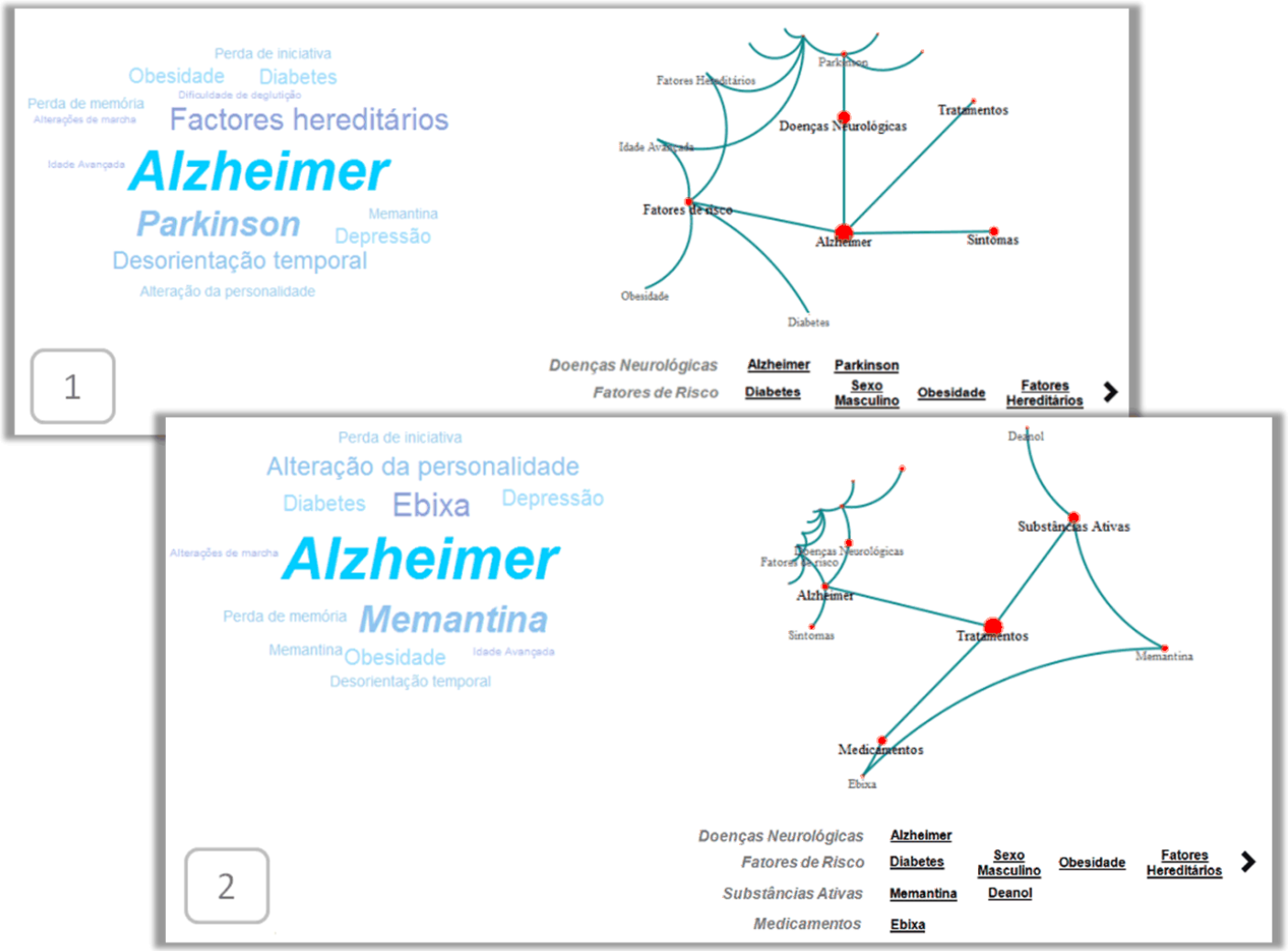


Semantic navigation/exploration

- This tool allows navigation of the semantic information
- When users search for something, the platform analyzes the query and verifies in the knowledge base the semantic of this text, constructing a knowledge graph



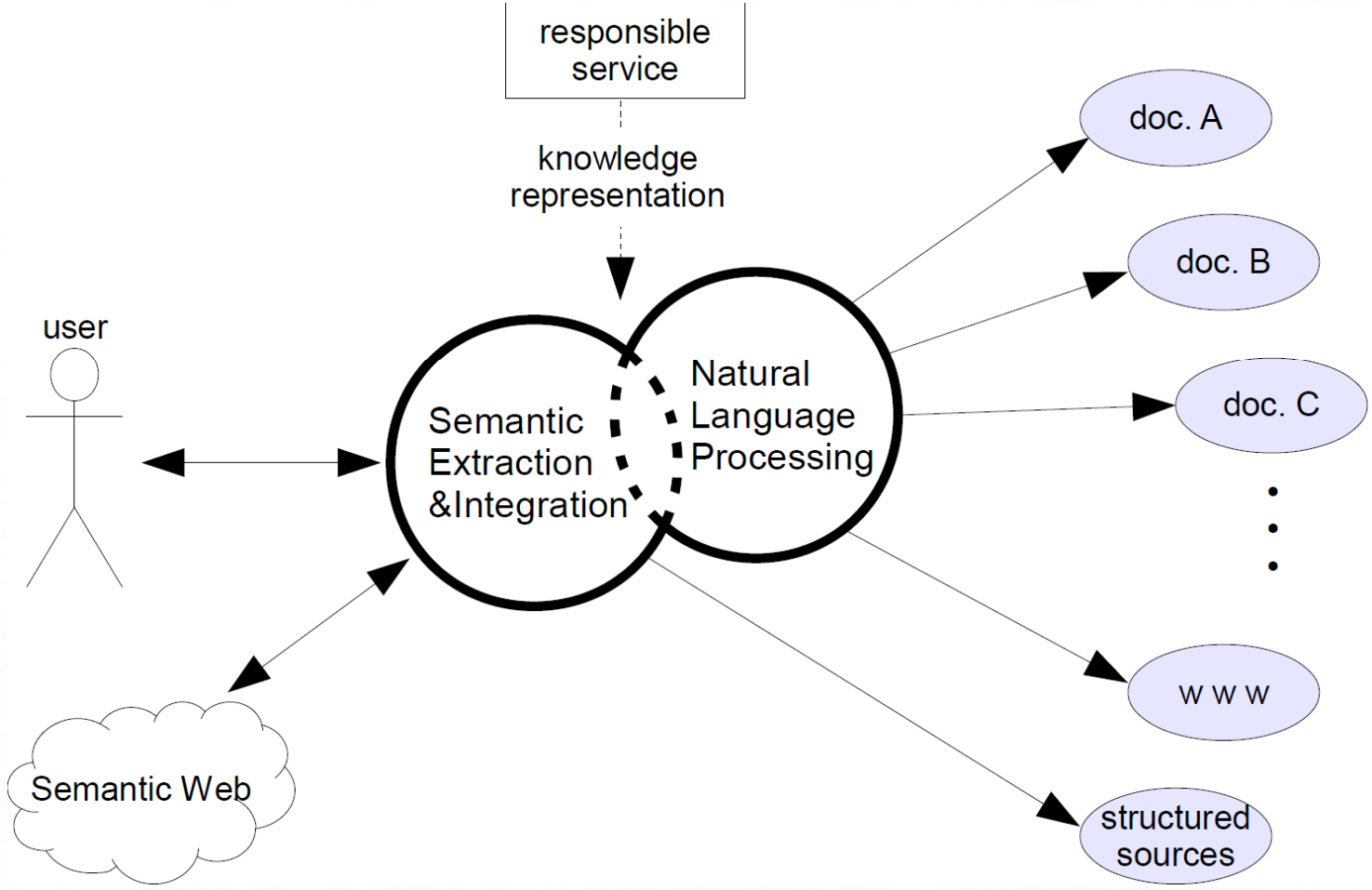
Example of Use



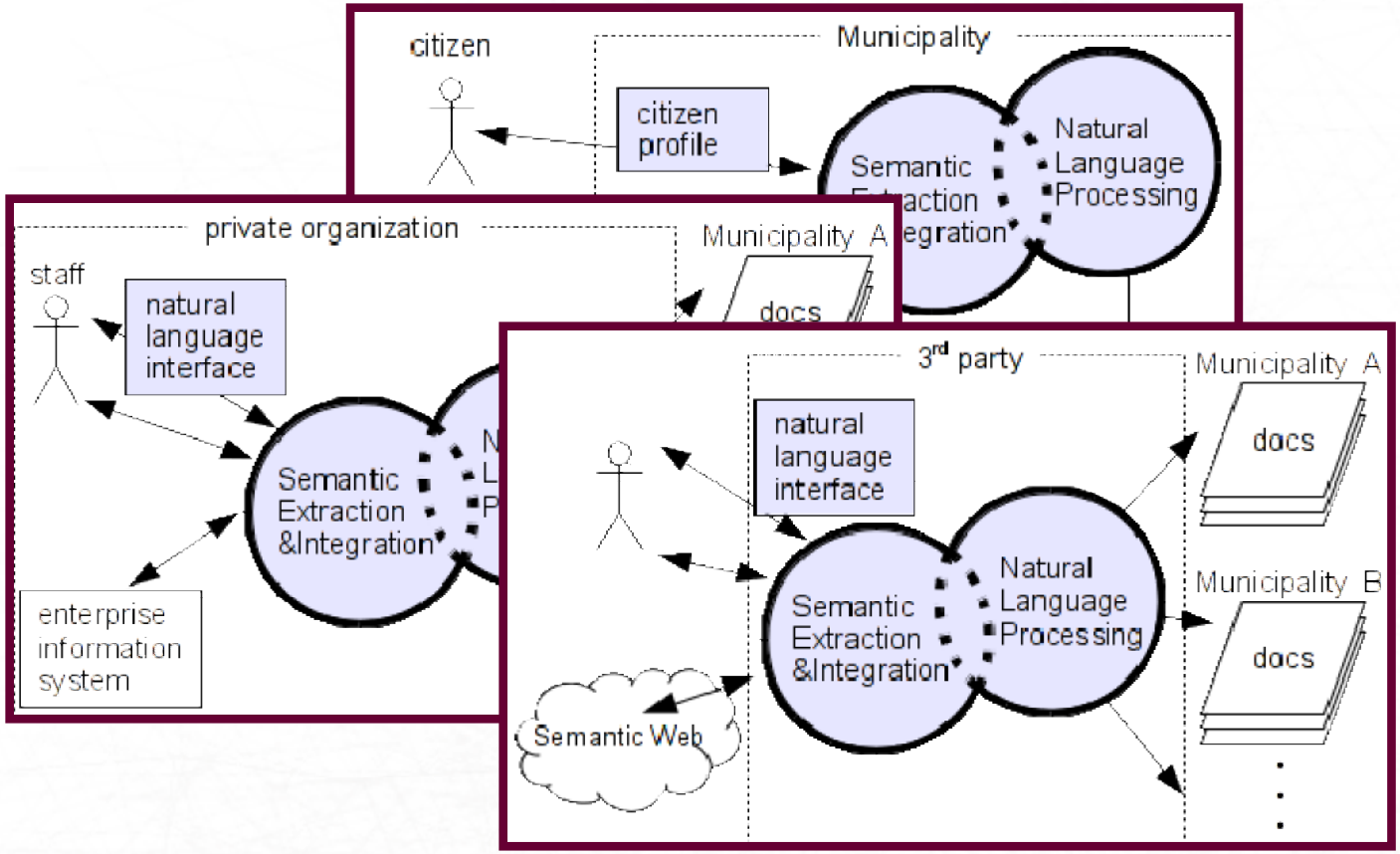
Memantina is one of the terms identified in the tag cloud selected.

If the user wants to know what memantina is or what its relationship is with these diseases and risk factors, selecting memantina the system will add this term to the query

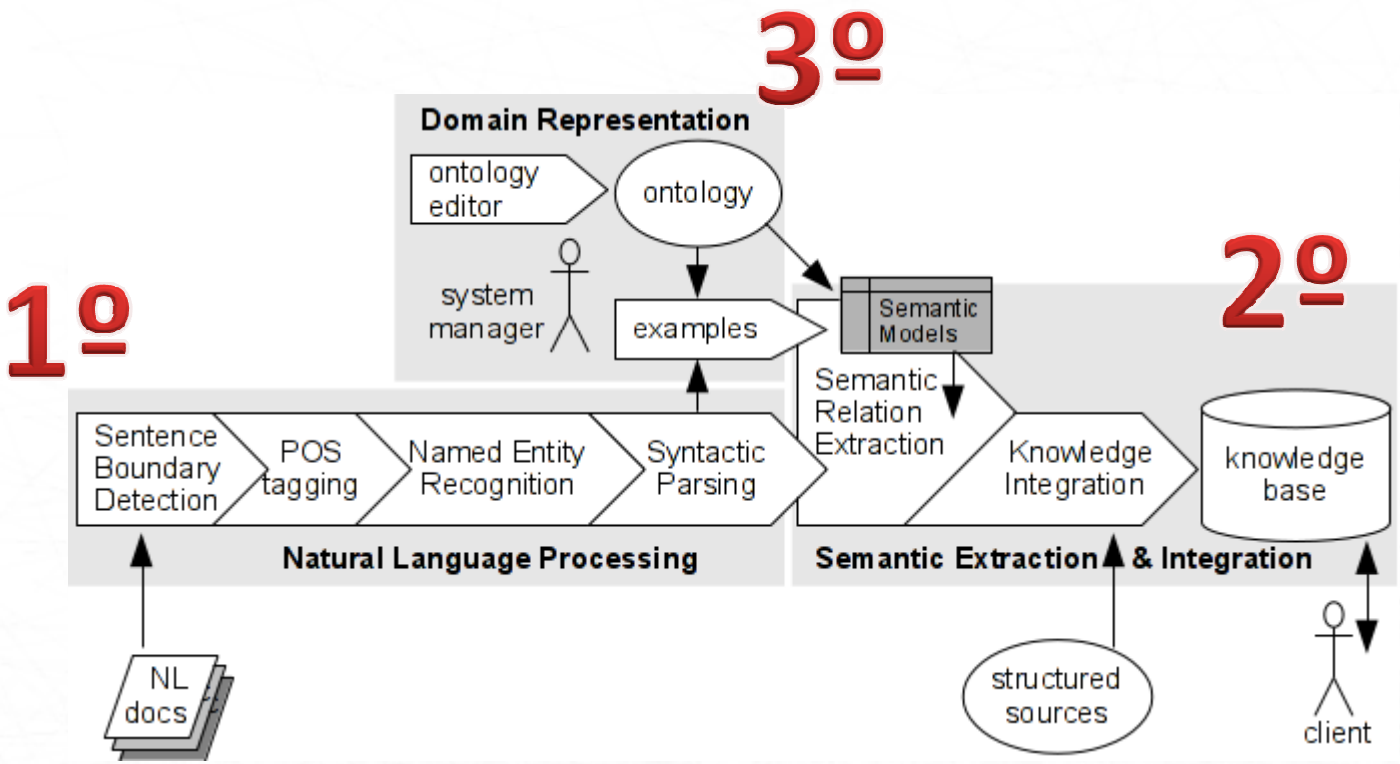
eGov Scenarios – Conceptual Model



Use scenarios



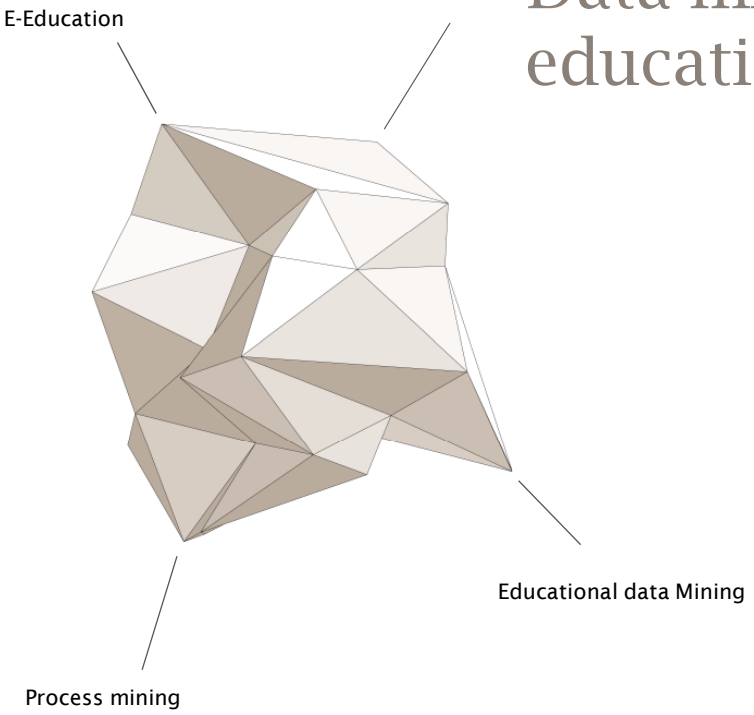
Architecture



Thank you for your attention.

- Acknowledgments
  - Mário Rodrigues and Liliana Ferreira
  - Part of the mentioned work was part of World Search project, a QREN project (QREN 11495), co-funded by CC and FEDER.
  - Also acknowledged the support by IEETA Research Unit, FCOMP-01-0124-FEDER-022682 (FCT-Pest C/EEI/UI0127/2011).

The Fourth International Conference on Advances in Information Mining and Management



Data mining and Process Mining in the educational field

Awatef Hicheur Cairns
Altran Reseach
Contact : awatef.hicheurcairns@altran.com

ALTRAN

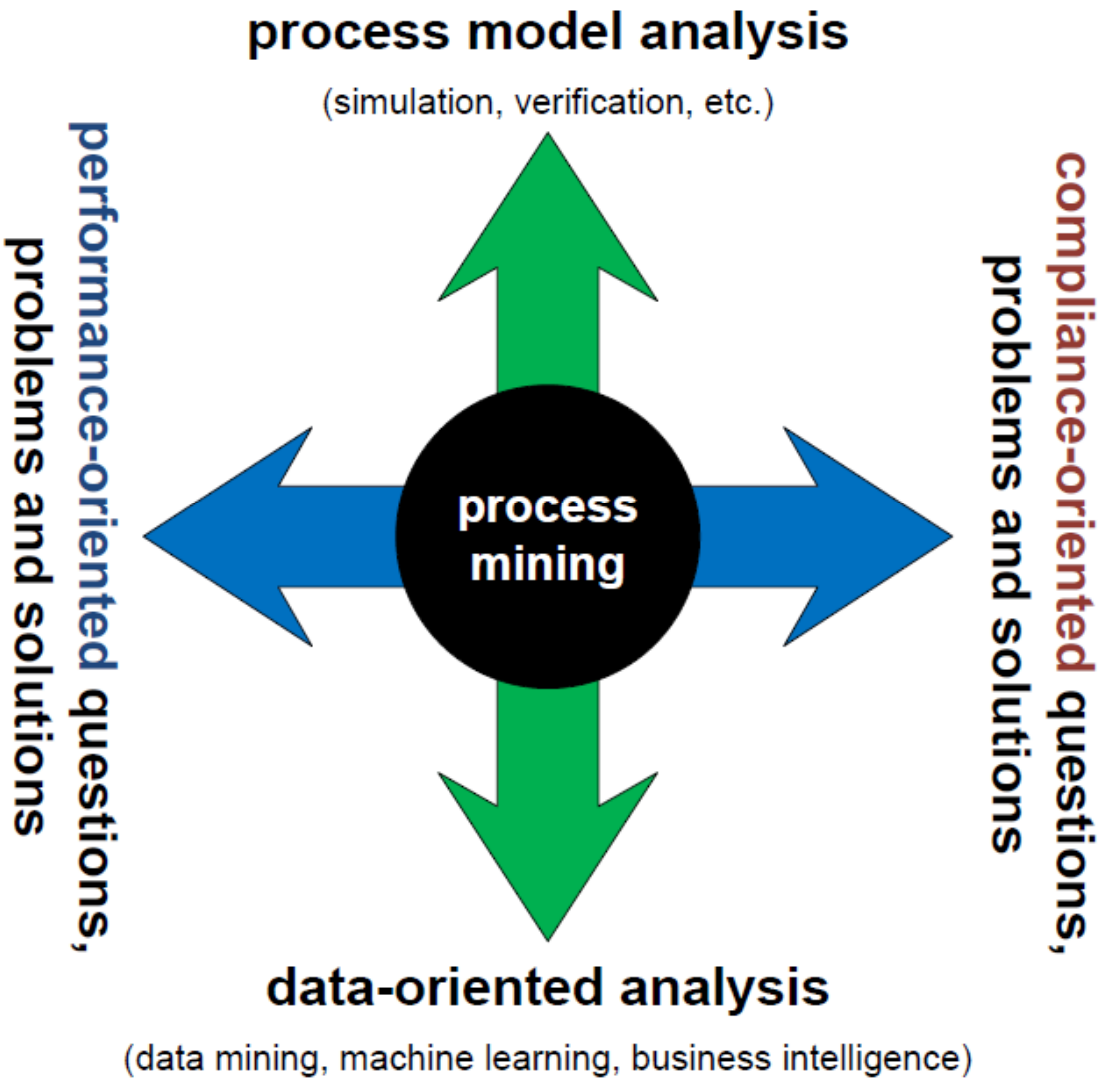
Educational data Mining

- Specific data mining methods to
  - explore the unique types of educational datasets
  - get a better understanding of students and the settings on which they learn

- Educational data mining methods
  - Prediction
  - Clustering
  - Relationship Mining
  - Discovery with Models
  - Distillation of Data For Human Judgment

- Limitation of the data mining techniques
  - Data centric, not process centric
  - Focus on data or sequential structures rather than whole process models with concurrency patterns

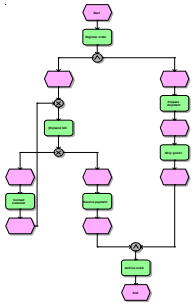
Positioning Process Mining



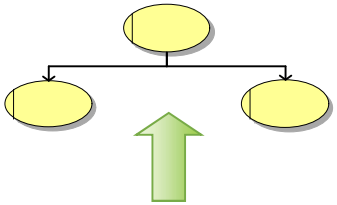
ALTRAN

Process Mining spectrum

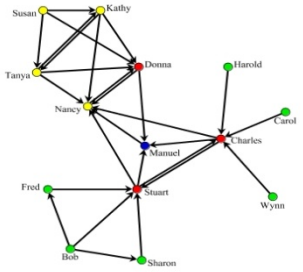
1) process model



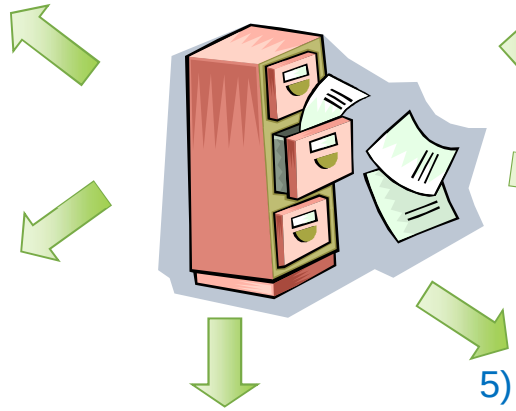
2) organizational model



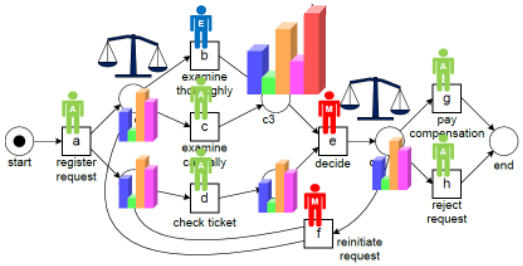
3) social network



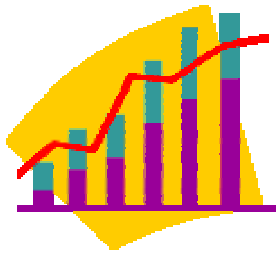
Model extraction



7) model extension



Extending models with frequencies, temporal information, decision points



4) basic performance metrics

6) auditing/security

Conformance checking

5) operational support

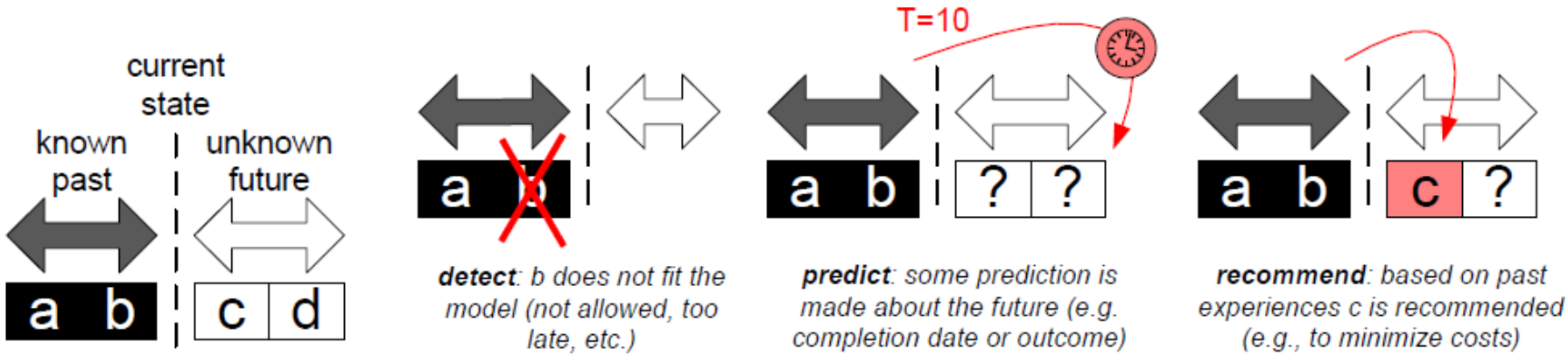
If …then …

Constructing predictive models
Operational support (prediction, recommendation, etc.)

Replay model against event log

ALTRAN

Process mining : Operational support



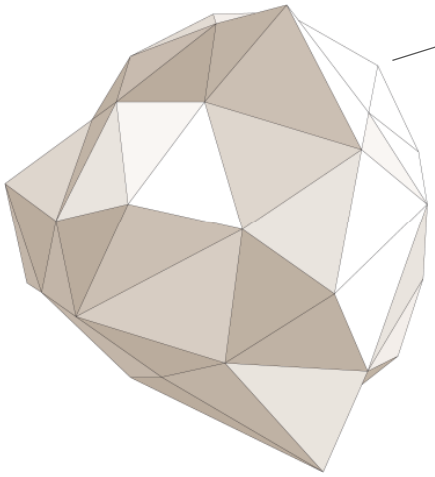
- Possible recommendations:
  - next activity;
  - suitable resource; or
  - routing decision.

- A recommendation is always given with respect to a specific goal. Examples of goals are:
  - minimize the remaining flow time;
  - minimize the total costs;
  - maximize the fraction of cases handled within 4 weeks;
  - maximize the fraction of cases that is accepted; and
  - minimize resource usage.

ALTRAN

Process Mining challenges in the educational field

- ✓ Voluminous Data - Large number of cases or events in event logs
  - ➢ Possible solution: clustering techniques to partition logs and distributed algorithms to parallelize computation
- ✓ Heterogeneity and Complexity - Large number of distinct traces and activities in event logs
  - ➢ Possible solution: filtering, abstraction or clustering techniques may help reducing the complexity
- ✓ Concept drift - Educational processes may change while being analyzed
- ➢ Possible solution: splitting the event log into smaller logs
- ✓ Usability and understability for end users
- ➢ Proposed solutions: visualization techniques and notation simplification



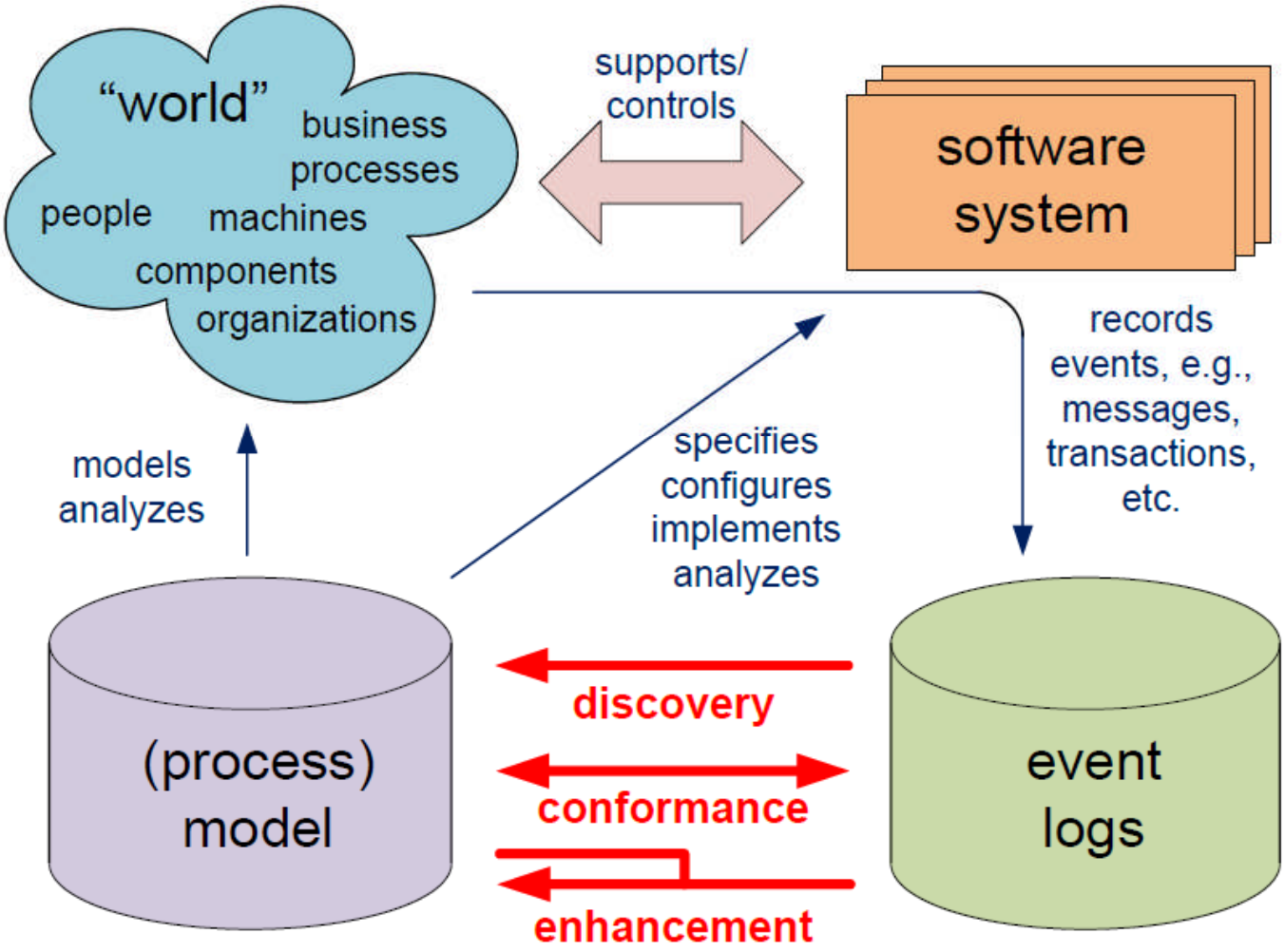
INNOVATION MAKERS

Thank you for your attention !

Questions ?

ALTRAN

The three main types of process mining: discovery, conformance, and enhancement



ALTRAN

Exemple of process discovery algorithms

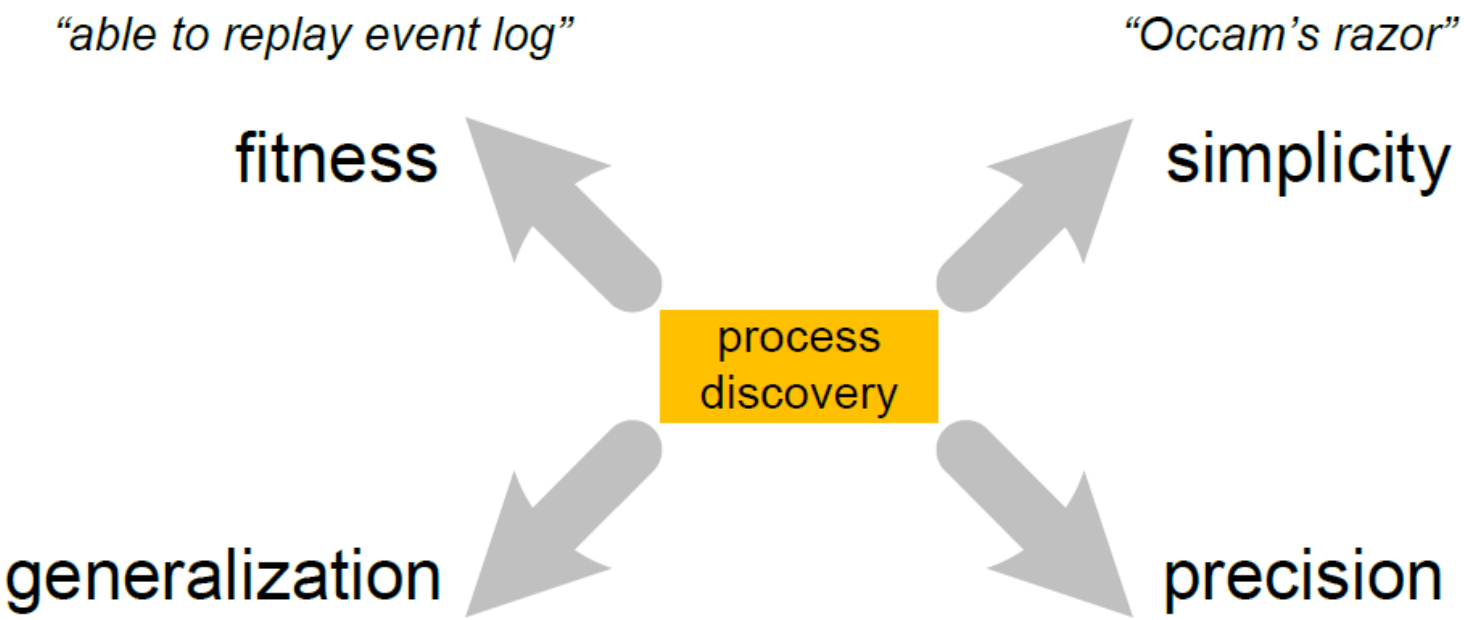
- **Algorithmic techniques**
  - Alpha miner
  - Alpha+, Alpha++, Alpha#
  - FSM miner
  - Fuzzy miner
  - Heuristic miner
  - Multi phase miner

- **Genetic process mining**
  - Single/duplicate tasks
  - Distributed GM

- **Region-based process mining**
  - State-based regions
  - Language based regions

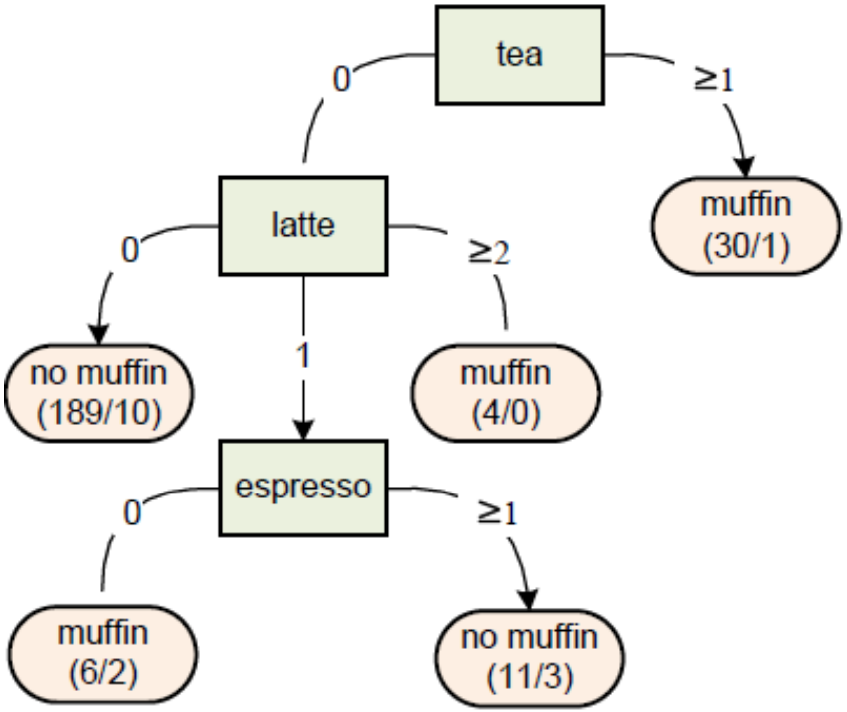
- **Classical approaches not dealing with concurrency**
  - Inductive inference (Mark Gold, Dana Angluin et al.)
  - Sequence mining

Process discovery Challenge: Four Competing Quality Criteria



ALTRAN

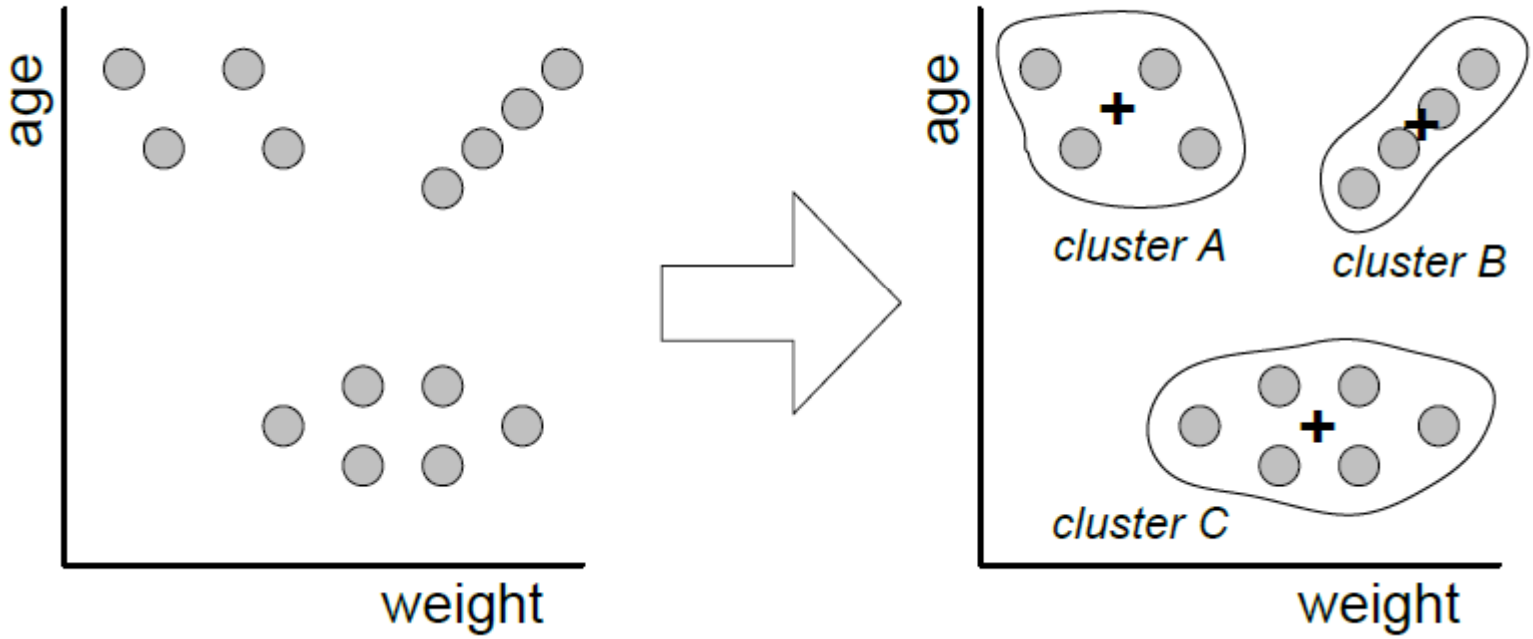
Decision tree learning



cappuccino	latte	espresso	americano	ristretto	tea	muffin	bagel
1	0	0	0	0	0	1	0
0	2	0	0	0	0	1	1
0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	1	2	0
0	0	0	1	1	0	0	0
...	...	...	...	...	...	...	...

ALTRAN

Clustering



Association rule learning

Règle de la forme “IF X THEN Y” $X \Rightarrow Y$

cappuccino	latte	espresso	americano	ristretto	tea	muffin	bagel
1	0	0	0	0	0	1	0
0	2	0	0	0	0	1	1
0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	1	2	0
0	0	0	1	1	0	0	0
…	…	…	…	…	…	…	…

$$tea \wedge latte \Rightarrow muffin$$

$$tea \Rightarrow muffin \wedge bagel$$

ALTRAN

Sequence mining

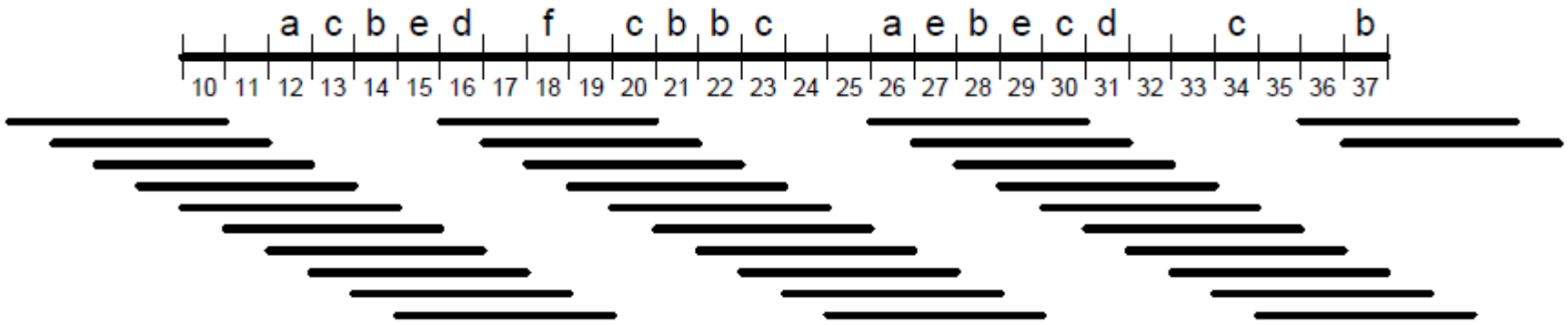
$$X \Rightarrow Y$$

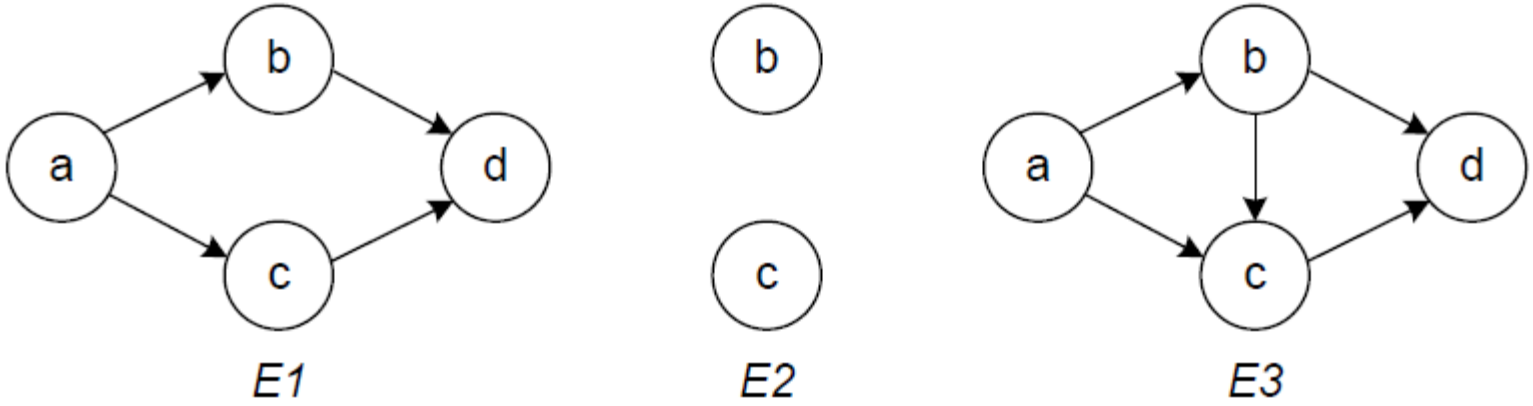
customer	seq. number	timestamp	items
Wil	1	02-01-2011:09.02	{*cappuccino*}
	2	03-01-2011:10.06	{*espresso,muffin*}
	3	05-01-2011:15.12	{*americano,cappuccino*}
	4	06-01-2011:11.18	{*espresso,muffin*}
	5	07-01-2011:14.24	{*cappuccino*}
	6	07-01-2011:14.24	{*americano,cappuccino*}
Mary	1	30-12-2010:11.32	{*tea*}
	2	30-12-2010:12.12	{*cappuccino*}
	3	30-12-2010:14.16	{*espresso,muffin*}
	4	05-01-2011:11.22	{*bagel,tea*}
Bill	1	30-12-2010:14.32	{*cappuccino*}
	2	30-12-2010:15.06	{*cappuccino*}
	3	30-12-2010:16.34	{*bagel,espresso,muffin*}
	4	06-01-2011:09.18	{*ristretto*}
	5	06-01-2011:12.18	{*cappuccino*}
...	...	...	...

$$X = \langle \{cappuccino\}, \{espresso\} \rangle$$

$$Y = \langle \{cappuccino\}, \{espresso\}, \{latte, muffin\} \rangle$$

Episode mining





32 time windows of length 5

ALTRAN

Towards the Web of Learning

Panel "Bridging the Web of Data with the Web of Documents"

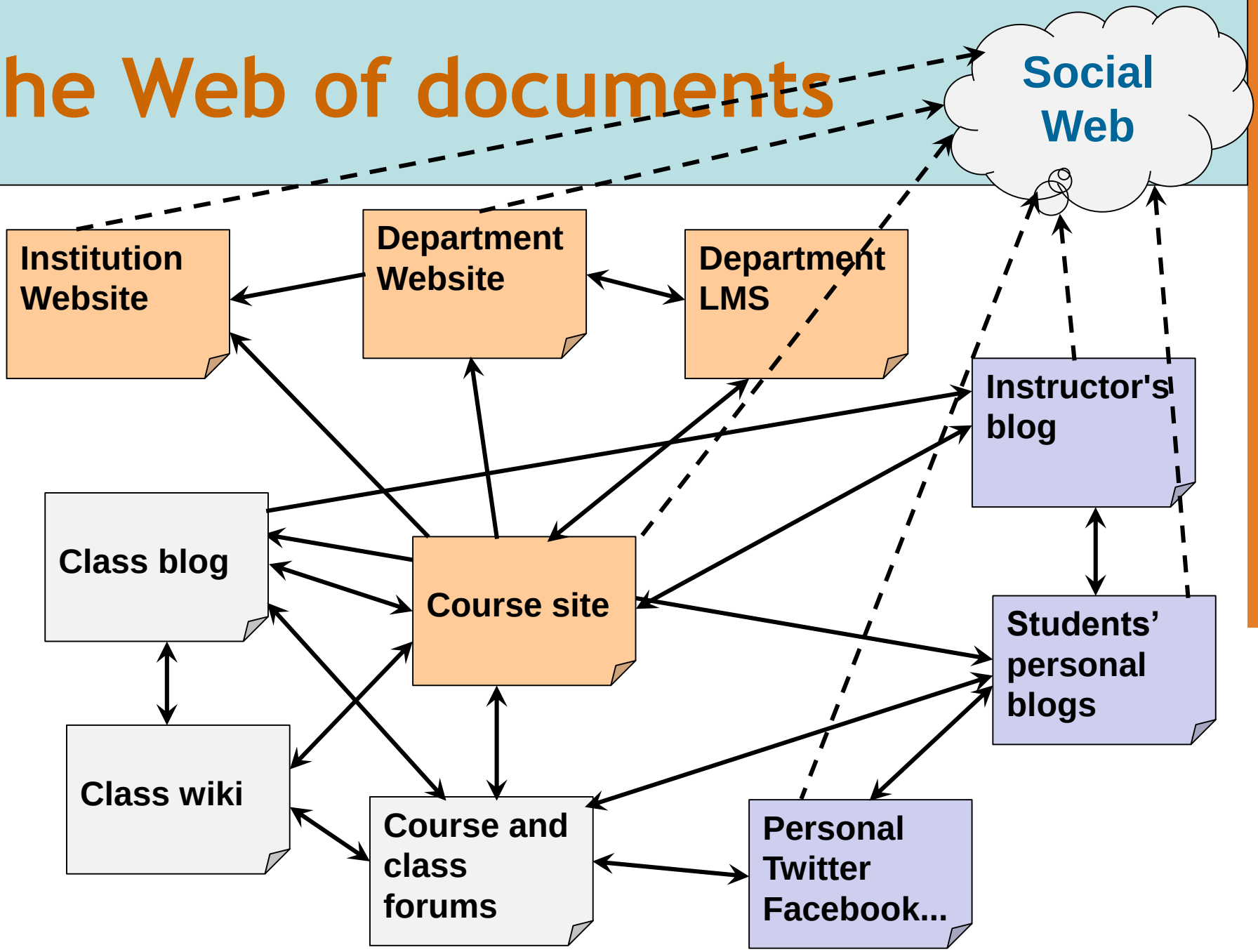
Prof. Athanassios Jimoyiannis

Department of Social and Educational Policy
University of Peloponnese, Greece

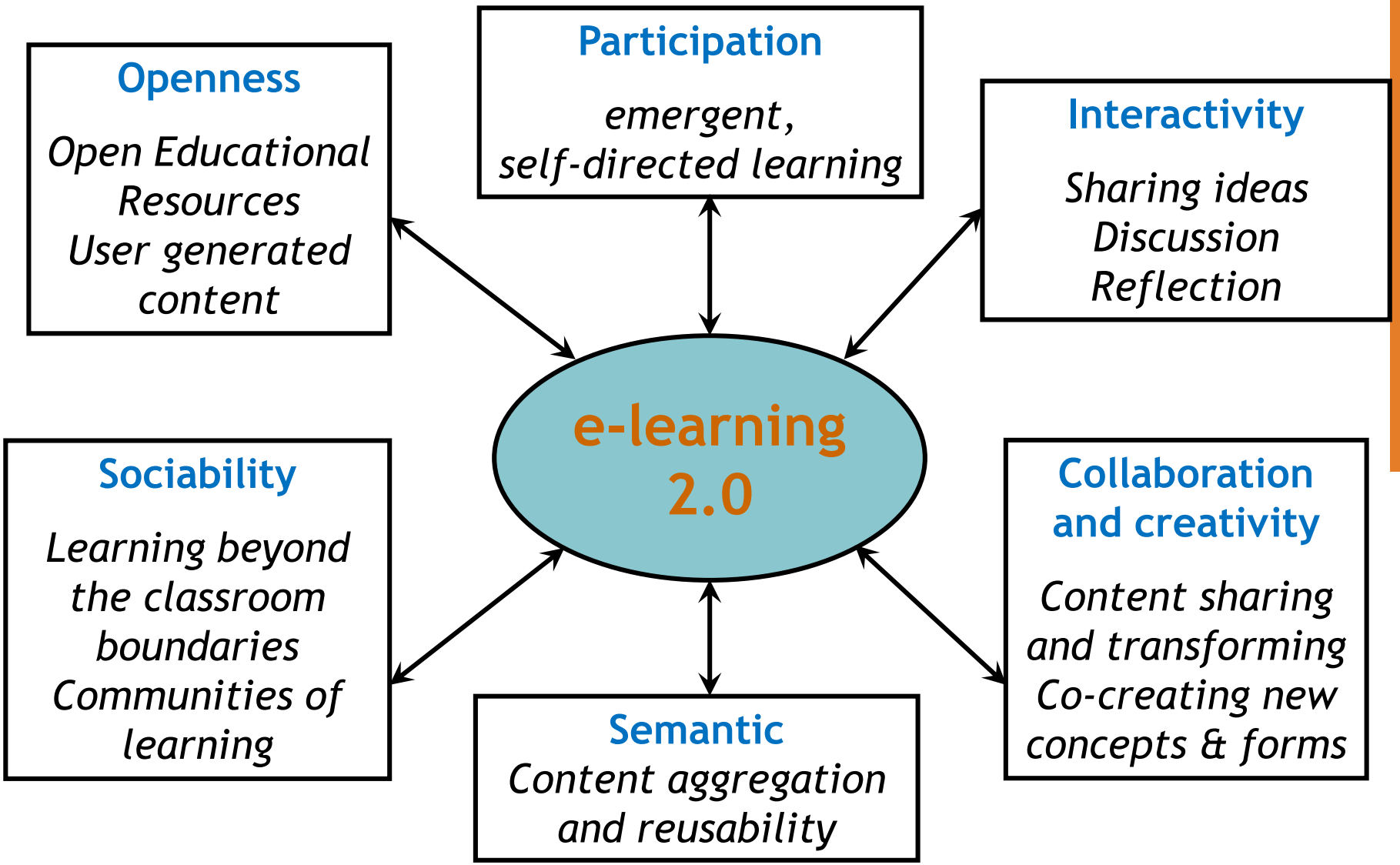
ICIW2014 Conference, 20-24 July 2014

Paris, France

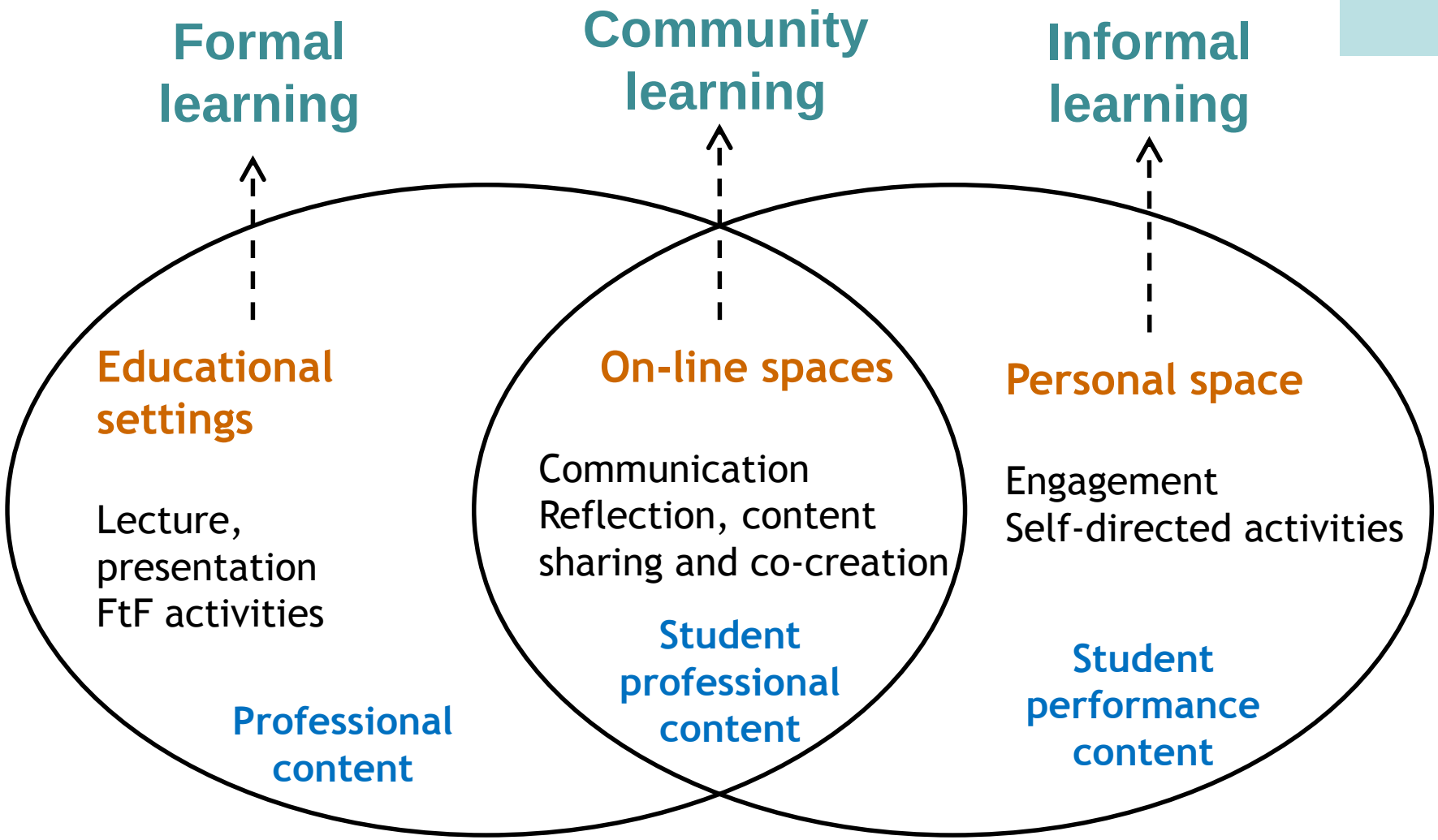
The Web of documents



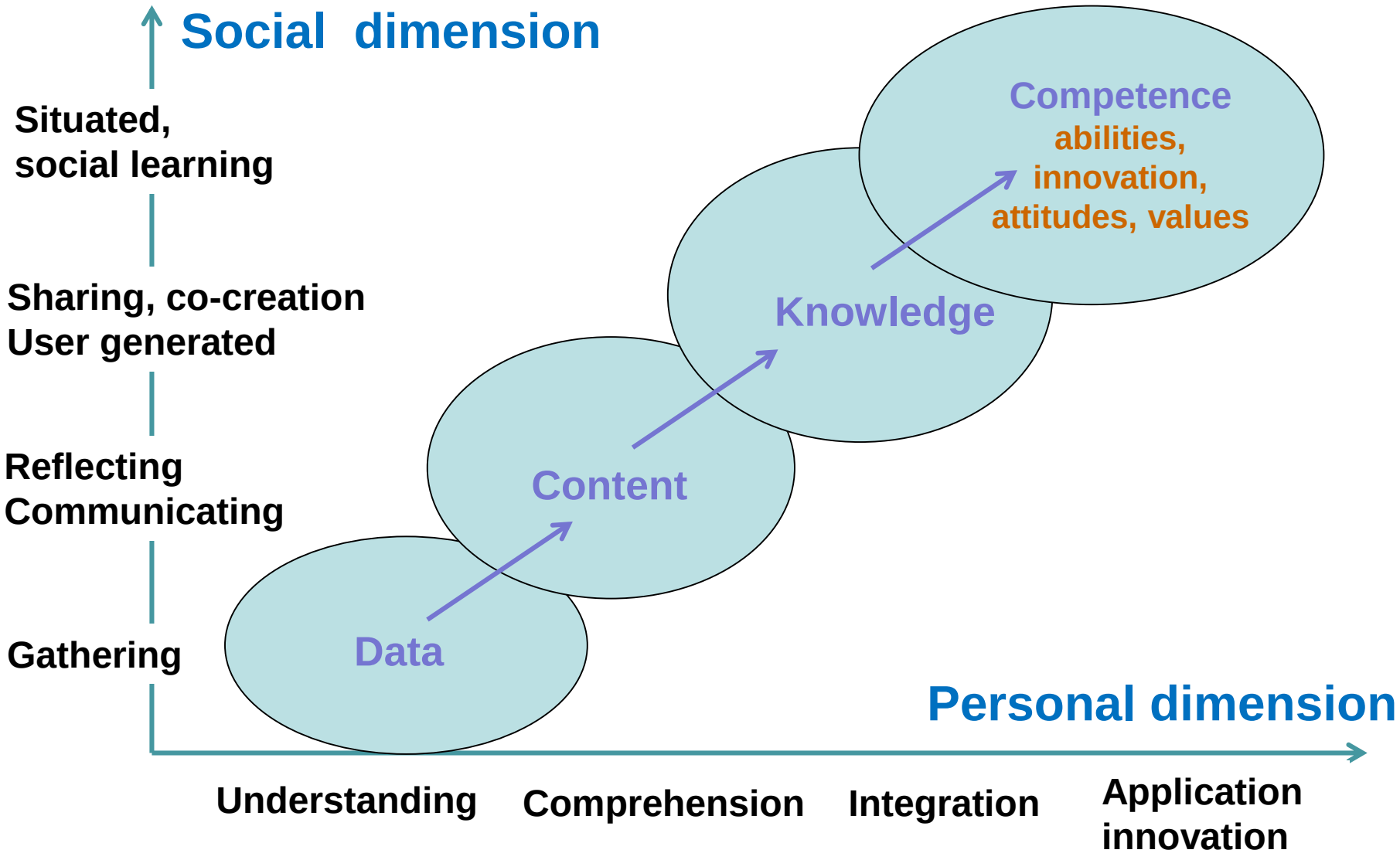
Web as a learning platform: e-Learning 2.0



The blended structure of Web-based learning



From data to competence



Towards Pedagogy 2.0

- **Beyond didactic paradigm (knowledge transfer)**
- **Beyond constructivism**
- Adding **Community** to the **Content**
- Shifting the focus from the **teaching-content** of a subject to the **learning activities**
  - Shared content and resources
  - Collaborative learning
  - Learners as content creators
  - Networked and collective intelligence
  - Self-directed and ubiquitous learning

Learning 2.0: Open questions

- How do we design **educational programs** and **learning tasks** through Web-based, social learning approaches?
- What are the best practices to **support** and **scaffold students** in a system where **learning** is expected **outside of school settings**?
- Are instructors ready - properly prepared to effectively respond to their **changing role**?

Conclusion message

"No teacher left behind!"

Thanks to the participants who contributed to this topic/debate with their **criticism**, **comments**, **ideas** and **experiences**