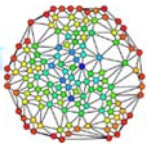


# Estrazione e Deduzione della Conoscenza via Modelli Semantici: From Social Network to Smart City

***seminario per il** Corso di Dottorato 2014*


***Prof. Paolo Nesi***

Dipartimento di Ingegneria dell'Informazione

University of Florence

Via S. Marta 3, 50139, Firenze, Italy

tel: +39-055-4796523, fax: +39-055-4796363

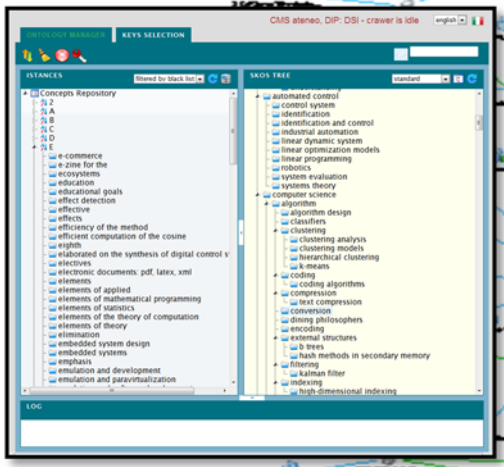
**DISIT Lab**

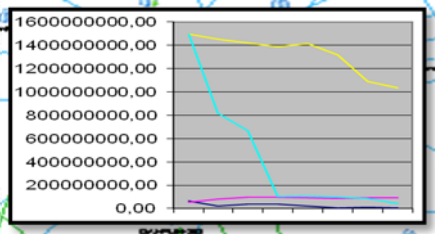
Paolo.nesi@unifi.it

**http://www.disit.dinfo.unifi.it/**

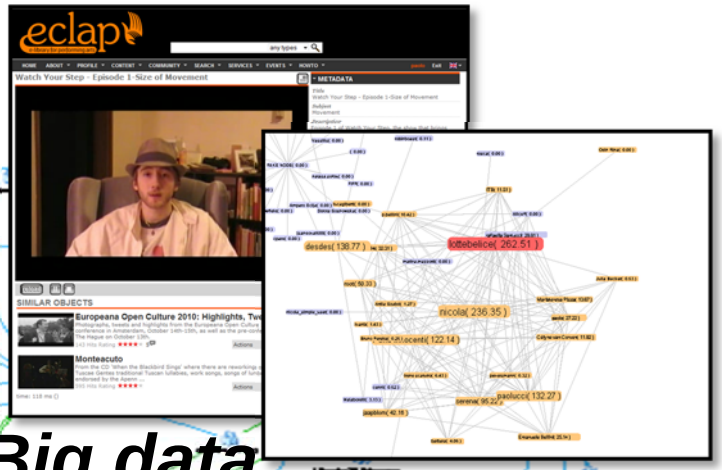
http://www.disit.dinfo.unifi.it

# Knowledge Acceleration




## Data Analytics, Big data

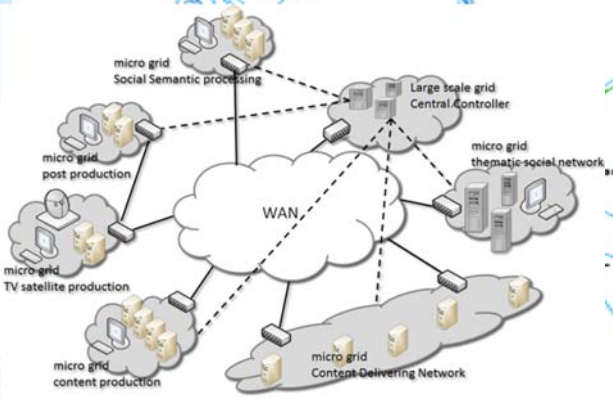



## Social Media


## Mobile Computing



## Smart Cities




## Cloud and Grid Computing

## Staff http://www.disit.dinfo.unifi.it

- Prof. Paolo Nesi, paolo.nesi@unifi.it
- More than 20 among postdocs, PhD students and fellows

## Current research topics

- Social media, collaborative work, Mobile computing, OpenData, LOD
- SmartCity, BigData, data analytics
- Railway signaling, autonomous driving systems, formal methods
- Cloud Computing, grid computing, smart cloud
- Data Mining, Knowledge Acceleration, natural language processing

## Main research results

- Knowledge Management and Natural Language Processing: OSIM, CoSkoSAM
- Content and Protection Management, grid computing: AXMEDIS AXCP
- Social Media, recommendations and tool: ECLAP.eu, MyStoryPlayer, Social Graph, IPR Wizard…
- Mobile Computing: Mobile Medicine, Mobile Emergency, etc.….
- Music Transcode, winner of MIREX for piano
- Awards: IEEE ICECCS, DMS, Italia degli Innovatori, etc.

## Main sources of funding

- **European Commission**: ECLAP (social media, CH), AXMEDIS (DRM, protection, automation e grid computing), WEDELMUSIC, IMAESTRO, VARIAZIONI, IMUTUS, MUSICNETWORK, MOODS, MUPAAC, OFCOMP, etc. ……
- **Italian Ministry**: Smart Cities COLL@BORA (collaborative work, social media), FIRB e PRIN
- **Regional**: SACVAR (knowledge mining and reasoning), TRACE-IT (Railway signalling), RAISSS (Railway signalling), ICARO (cloud)
- **Fondations**: MatchMaking (NLP), OSIM (Knowledge Acceleration, NLP)

# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities

# Profilo degli utenti

- *Informazioni statiche:*
  - *Informazioni generali:*
    - *nome, cognome, sesso,*
    - *foto, data di nascita,*
    - *descrizione personale,*
    - *località di provenienza (ISO 3166),*
      - *Nazione*
      - *Suddivisione*
      - *Provincia*
    - *lingue parlate (ISO 369)*
  - *Informazioni di contatto:*
    - *lista di contatti di instant messaging*
  - *Scuola e Lavoro:*
    - *scelta del livello scolastico,*
    - *nome della scuola,*
    - *tipo di lavoro,*
    - *nome del posto di lavoro*
  - *Interessi:*
    - *Vettore contenente la lista di valori del campo Type degli oggetti scelti dall'utente*
- *Informazioni dinamiche:*
  - *Lista di oggetti preferiti*
  - *Lista di amici*
  - *Lista gruppi*
  - *Voti positivi ad oggetti*
  - *Commenti ad oggetti*
  - *…*
  - *…*
  - *Informazioni sulle preferenze sulla base delle visualizzazioni degli oggetti*
    - *Format*
    - *Type*
    - *Taxonomy*

# Content Searching in Social Network

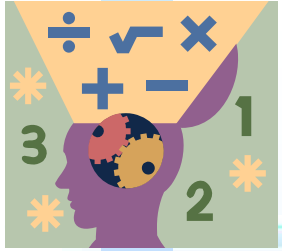
- Traditional Classification based on Metadata
  - Faceted search
- Taxonomies
- Free Tags, as Folksonomy
- Geotagging, GPS data
- Annotations
- Votes

# User Profile Problems

- **Different data types**:
  - Numbers: age, votes, #kids, ..
  - Enumerates/symbolic: language, nationality, etc.
- **Multiple Values / Selections**:
  - languages, nationalities, preferences, etc…
- **Non-Symmetrical Distances**, for instance:
  - Preferences: *Dim ({Pref(A)}) ≠ Dim ({Pref(B)})*
- **Dynamic information**
  - related computational complexity
- **Different Languages** of comments, descriptions,
  - Language processing and understanding
  - Dictionaries, Semantics, Taxonomy, etc.

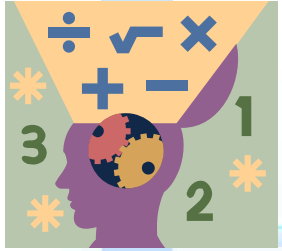
# friendship propagation

- User links and friendship propagation….
- **Mechanisms for invitation**
  - User A invites N Users
  - Among these N Users, M Accept the invitation
- **Viral Indicator**
  - If M > N a mechanism of viral grow is started
  - It can exponentially grow up or to simply produce a small pike
- **Users have:**
  - Direct Friends---------- for example: 90
  - Indirect Friend of different levels ---------: level 1: 900
  - Friends via groups (see LinkedIn) ---------: 14000



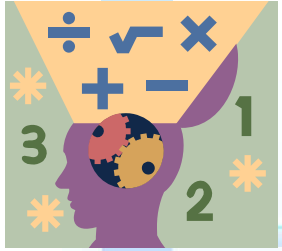
# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
  - Semantic Modeling
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities



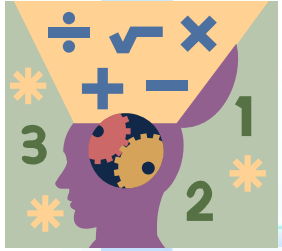
# Semantic Descriptors and info 1/3

- **user profile descriptions** collected via user registration and dynamically on the basis of user actions, migrated also on the mobile:
  - selected content, performed queries,
  - preferred content, suggested content, etc.;
- **relationships among users/colleagues** (similarly to friendships, group joining) that impact on the user profile and are created via registration, by inviting colleagues, etc.;
- **user groups descriptors** and their related discussion forums and web pages (with taxonomic descriptors and text);



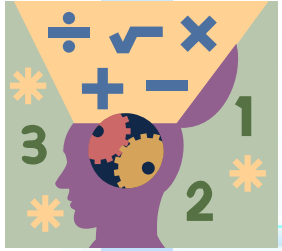
# Semantic Descriptors and info 2/3

- **content descriptors** for simple and complex content, web pages, forums, comments, etc.;
- **device capabilities** for formal description of any acceptable content format and parameters, CPU capabilities, memory space, SSD space;
- **votes and comments on contents, forums, web pages**, etc., which are dynamic information related to users;



# Semantic Descriptors and info 3/3

- **lists of elements marked as preferred by users,** which are dynamic information related to users;
- **downloads and play/executions** of simple and/or complex content on PC and mobiles, to keep trace of user actions as references to played content, which are dynamic information related to users preferences;
- **uploads and publishing** of user provided content on the portal (only for registered users, and supervised by the administrator of the group). Each Content element has its own static metadata, descriptors and taxonomy; while the related action of upload is a dynamic information associated with the User who performed it. In addition, Content elements can be associated with Groups.

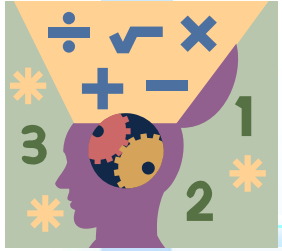


# Group Descriptors

- **Groups** of users they may have specific descriptors and those inherited by the users:

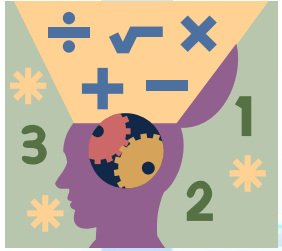
- **static** aspects of the groups such as:
  - objectives, topics, web pages, keywords, taxonomy, etc.;

- **dynamic** aspects related to:
  - users belonging to the group; users may: join and leave the group, be more or less active over time;
  - content associated with the group: files, comments, etc., with their taxonomical classification, metadata and descriptors.



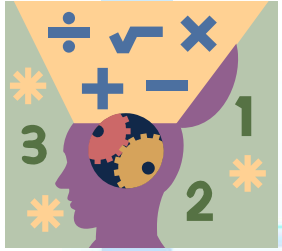
# Content Descriptors

- **Static aspects** : more relevant since the content description is typically not changing over time. They are:
  - metadata, keywords extracted from description, comments, etc.;
  - technical description (as the Format in the following): audio, video, document, cross media, image,..;
  - content semantic descriptors such as: rhythm, color, etc.; genre, called Type in the following;
  - groups to which the content has been associated with;
  - taxonomies classification to which the content has been associated, taking into account also the general taxonomy;
- **dynamic aspects** are marginally changed and may be related to:
  - user's votes, user's comments;
  - number of votes, comments, download, direct recommendations, etc.;



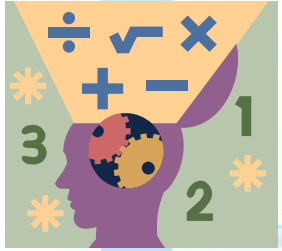
# Semantic Descriptors

- **Modeling descriptors with formalisms:**
  - XML
  - MPEG-7, metamodel for descriptors and descriptors
  - MPEG-21: item descriptor and/or package
- **Audio, Video, images:**
  - **Low level** fingerprint/descriptors
    - Hash, MD5, etc.
  - **High level** fingerprint/descriptors
    - Genre, rhythms, color, scenes/movements, etc.
    - Evolution of them along  the time, along the file
- **Documents**:
  - Keywords extractions, multilingual agnostic, …
  - Summarization
  - Paragraphs modeling and descriptions



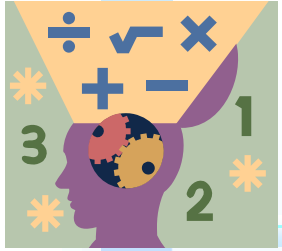
# Usage/Prod of Semantic Information

- **content ingestion.** semantic tagging while technical descriptors about digital resources are added during the automated adaptation and icon production;
- **repurposing and publication for several kinds of end-user devices**
- **extraction of semantic technical descriptors** from simple and complex essences,
- **content indexing** to prepare and accelerate the process of search.
- **packaging content and semantics into MPEG-21/AXMEDIS** binary format: integrating digital essences with metadata and descriptors
- **exporting content** to other databases, or posting them on other social networks or portals, publishing on P2P networks
- **estimating similarities among users, objects/content,** to pose the basis of generating suggestions and reasoning;
- **producing suggestions** about potential colleagues, interesting content, and groups;



# Content Enrichment

- The content, UGC, reaches the Social Network with partial information
- **Content Enrichment is needed** to get enough semantic information for
  - indexing/querying and producing suggestions
- **Content enrichment** may be performed by:
  - Addition/Extraction of semantic descriptors
  - Multilingual translation for metadata
  - Addition of annotations, textual and audiovisual
  - Association of SKOS/taxonomical terms
  - Association of Tags → folksonomy
  - ***Comments, rating, citations, etc.***
  - ***Creation of Aggregations: collection, courses, play lists***



# Extraction of semantic descriptors

- **Technical Information**
  - duration, resolution, size, dimension, video rate, sample rate and size, file format, MIME type, number of included files, file extension, etc.
  - libraries or tools can be used to extract information: FFMPEG for video and audio, ImageMagik for images, etc.

- **Context information:**
  - summary and extract keywords;
  - Video processing to: segment in major scenes, understand them, identify objects, colors, etc.
  - audio processing to extract tonality, rhythm, etc.
  - Images processing to extract contained objects, etc.

# High Level Reasoning Semantic Computing, 1/2

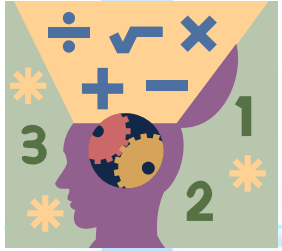
- **Linguistic processing:** assessment of intentions, understanding
  - Extraction of positive/negative impressions
  - Technical instruments:
    - Ontology production, integration, augmentation
    - Ontology merging, engines
    - Processing OWL
    - Triple database, Semantic SQL

- **Semantic meaning** of high level information
  - Dictionaries: to compare/infer multilingual keywords
  - Folksonomies: production of free keywords
  - Taxonomies: specialization relationships
  - Ontology: a range of relationships

# High Level Reasoning Semantic Computing, 2/2

- **Taking decision** on the basis of Descriptors and their relationships
  - Technical instruments:
    - Taking decision engines
    - inferential engines such as Jena,
    - rules based systems,
    - script-based rules,
    - constraint programming,
    - First logic, temporal logic engine, etc.

- **Recommendations/suggestions**, production of
  - Technical instruments:
    - Clustering among elements: content, users, groups, ..
      - on the basis of distances/similarities among descriptors
    - Clustering models: K-means, k-medoid, hierarchical clustering



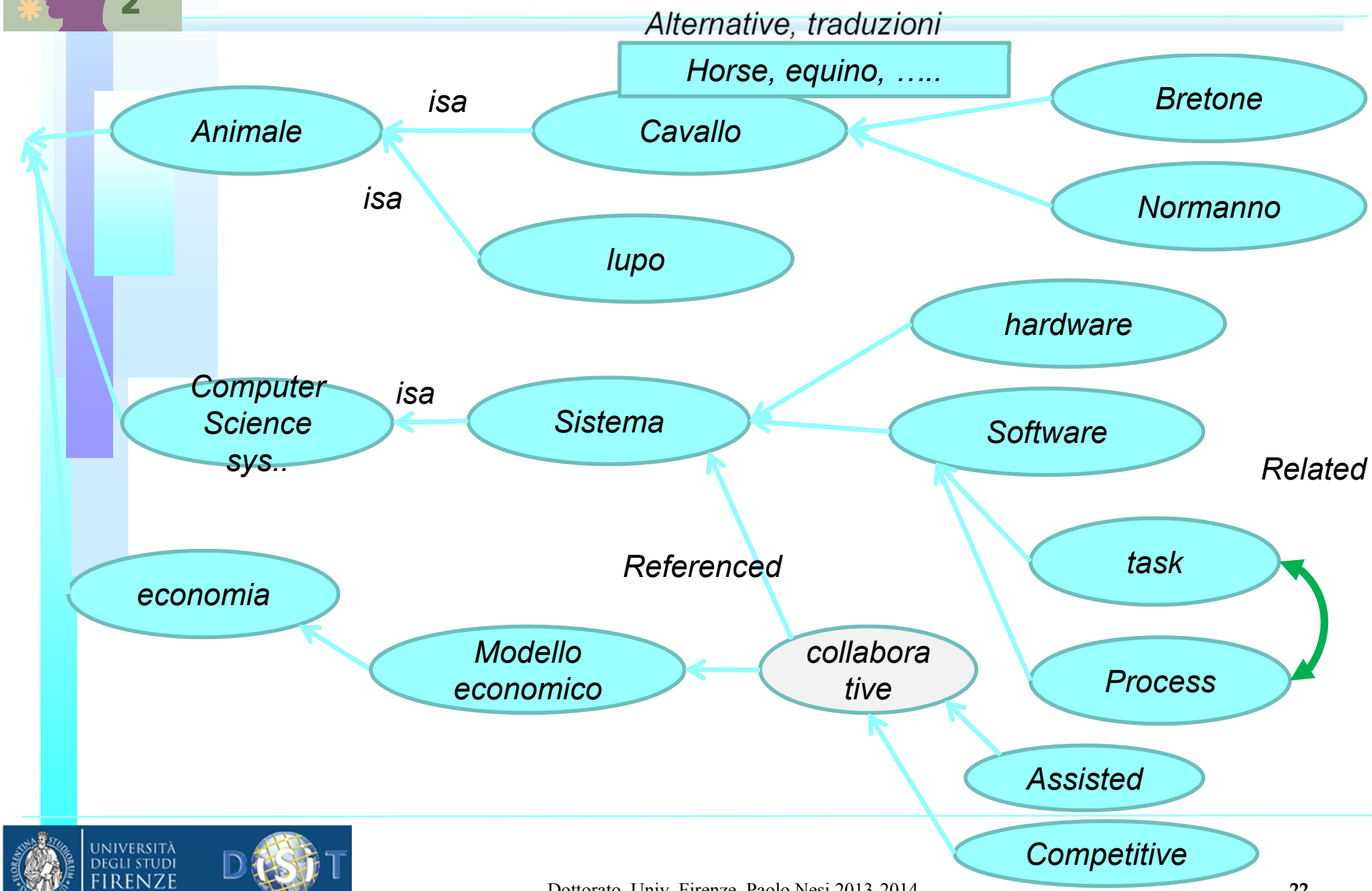
# Esempio di Tassonomia

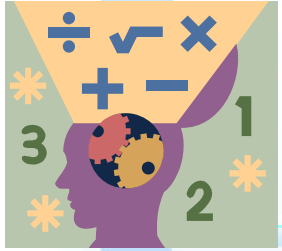
- Classificazione secondo vari assi
- Dominio Specifico
- Multilingua
- Istanze connesse a più nodi
- Preso da: http://mobmed.axmedis.org



UNIVERSITÀ DEGLI STUDI FIRENZE

# SKOS simple knowledge organization system

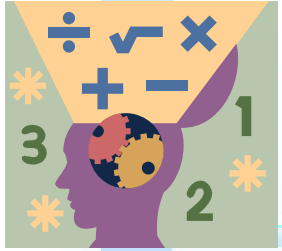




# FOAF: Friend of a Friend

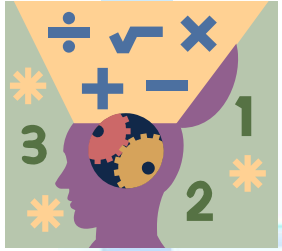
- **Friend of a Friend (FOAF):**
  - A format for supporting description of people and their relationships
  - a vocabulary in OWL for sharing personal and social network information on the Semantic Web
  - **Based on AAA principle:**
    - ➔ Anyone can say anything about any topic

- **Modeling Information:**
  - Organization at which people belong
  - Documents that people have created/co-authored
  - Images that depict people
  - Interests/skill of people,…
  - …..



# FOAF: Friend of Friend

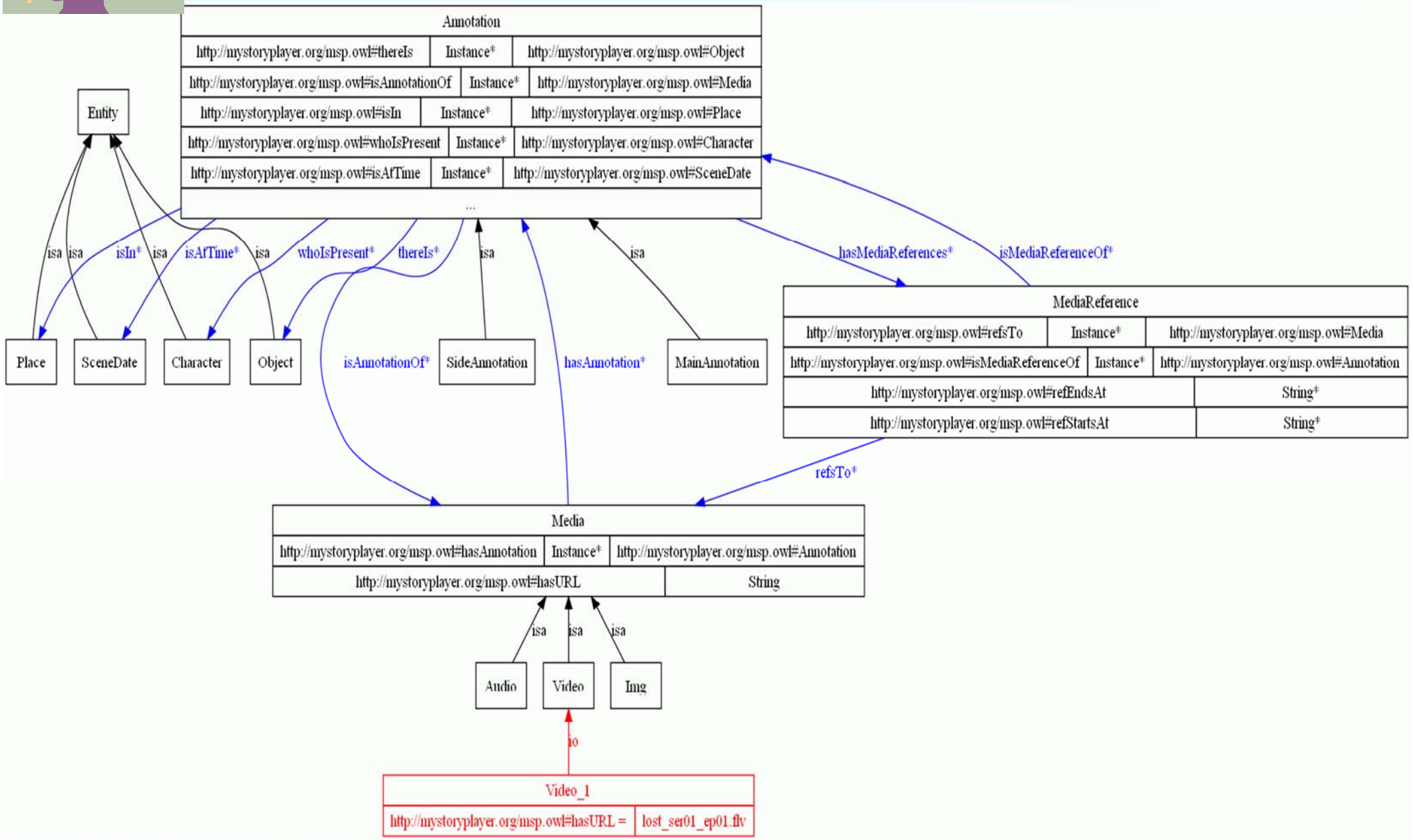
```
<foaf:Person>
    <foaf:name>Peter Parker</foaf:name>
    <foaf:gender>Male</foaf:gender>
    <foaf:title>Mr</foaf:title>
    <foaf:givenname>Peter</foaf:givenname>
    <foaf:family_name>Parker</foaf:family_name>
    <foaf:mbox_sha1sum>cf2f4bd069302febd8d7c26d803f63fa7f
      20bd82</foaf:mbox_sha1sum>
    <foaf:homepage rdf:resource="http://www.peterparker.com"/>
    <foaf:weblog
      rdf:resource="http://www.peterparker.com/blog/"/>
    <foaf:knows> <foaf:Person>
      <foaf:name>Aunt May</foaf:name></foaf:Person>
    </foaf:knows>
</foaf:Person>
```

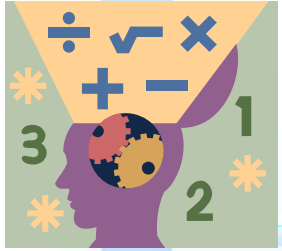


# Ontologie

- L'ontologia è una specificazione *formale esplicita* di una *concettualizzazione* di un *dominio*
- Rappresentano:
  - Concetti e oggetti: modelli, categorie, proprietà,..
  - Relazioni fra concetti e fra relazioni
- Idealmente mirano a modellare in modo "*esaustivo*" un dominio

# Un Esempio

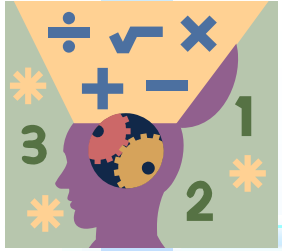

Annotation
http://mystoryplayer.org/msp.owl#thereIs
Instance*
http://mystoryplayer.org/msp.owl#Object
http://mystoryplayer.org/msp.owl#isAnnotationOf
Instance*
http://mystoryplayer.org/msp.owl#Media
http://mystoryplayer.org/msp.owl#isIn
Instance*
http://mystoryplayer.org/msp.owl#Place
http://mystoryplayer.org/msp.owl#whoIsPresent
Instance*
http://mystoryplayer.org/msp.owl#Character
http://mystoryplayer.org/msp.owl#isAtTime
Instance*
http://mystoryplayer.org/msp.owl#SceneDate
...
Entity
isa
isa
isIn*
isa
isAtTime*
isa
whoIsPresent*
thereIs*
isa
isa
hasMediaReferences*
isMediaReferenceOf*
Place
SceneDate
Character
Object
isAnnotationOf*
SideAnnotation
hasAnnotation*
MainAnnotation
MediaReference
http://mystoryplayer.org/msp.owl#refsTo
Instance*
http://mystoryplayer.org/msp.owl#Media
http://mystoryplayer.org/msp.owl#isMediaReferenceOf
Instance*
http://mystoryplayer.org/msp.owl#Annotation
http://mystoryplayer.org/msp.owl#refEndsAt
String*
http://mystoryplayer.org/msp.owl#refStartsAt
String*
refsTo*
Media
http://mystoryplayer.org/msp.owl#hasAnnotation
Instance*
http://mystoryplayer.org/msp.owl#Annotation
http://mystoryplayer.org/msp.owl#hasURL
String
isa
isa
isa
Audio
Video
Img
io
Video_1
http://mystoryplayer.org/msp.owl#hasURL =
lost_ser01_ep01.flv



# Base Ontologica

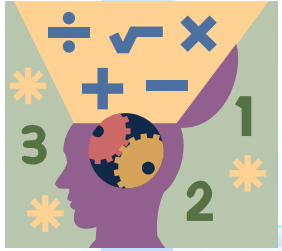
- Si può formalizzare in OWL (ontology web language), XML (Extensible Markup Language)

- Unifica/Generalizza modelli come:
  - Tassonomie
  - Tesauri
  - Vocabolari
  - SKOS: simple knowledge organization system
  - FOAF: friend of a friend



# Base Ontologica

- Le ontologie sono specifiche di un dominio
- Spesso prodotte in team e formalizzate in OWL, vi sono strumenti di:
  - Editing, e.g., Protégé
  - Database semantici, e.g., Sesame in RDF (Resource Description Language)
  - Inferenza su database
  - Query semantiche, per esempio formalizzati in SPARQL (Simple Protocol and RDF Query Language)
  - ….



# Base di Conoscenza

- Si può vedere come: Ontologia + istanze
- Le istanze dei concetti, delle relazioni, popolano la base di conoscenza connettendosi all'ontologia con la relazione di *instance-of (io)*, e.g.:
  - il documento afkagf.pdf connesso al nodo Analgesia della tassonomia MobMed
  - Carlo Rossi e' un Paziente
  - Carlo Rossi e' figlio di Giovanni Rossi

# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
  - Raccomandazioni / suggerimenti
  - Metrics Similarity Distances
  - Clustering algorithms comparison
  - Performances, Incremental Clustering
  - Suggerimenti U→U an improvement
  - Validazione del modello di suggerimento
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities

# Recommendations

- **Different Recommendations/Suggestions**
  - U → U: a user to another user on the basis of user profile
  - O → U: an object to a user on the basis of user profile
  - O → O: an object on the basis of a played object of a user
  - G → U: a group to a user on the basis of user profile
  - Etc…
- **Objects can be:**
  - Advertising, Content, Events, etc.
  - Some of  them may have specific descriptors…

# Different Recommendations

- **FOR YOU: Suggesting Users to another Users since they**
  - have similar preferences
  - like/prefer what you like/prefer
  - are friends of your friends
  - are in one or more of the your groups
  - are new of the Social Network
  - are the most linked, grouped, active
  - etc.

- **FOR THE SN: Suggesting Users to another Users since they**
  - *are important for the SN and do not have to left alone, the new entry*
  - *are the only contact path for Connecting a remote group, if the path is left a peripheral group will be completely disjoined with respect to the rest of the SN*
  - …

# Different Recommendations

- **FOR YOU: Suggested objects/contents/events/groups since they**
  - are the less, most viewed, most played, most played in your group, ..
  - are similar to your highest voted/ranked objects
  - are similar to what you usually play, pay, print, upload, etc.
    - The most played/../voted in absolute
    - The most played/../voted in the last Month/Day, week, etc…
    - The most played/../voted in your area, country, group, etc..
  - are new for the SN
  - belongs to the preferred of your friends, …
  - have been posted/commented by your friends, in your group, …
  - have been recommended by a your friend

- **FOR BUSINESS: Suggested objects/…./groups since they**
  - are new for the SN, and thus are new for the market/business of the SN
  - are commercially proposed and have to be commercially promoted for the business of the SN
  - belong on the log tail of the content distribution/usage

# Recommendations

| | | Recipient of the suggestions | | |
|---|---|---|---|---|
| | | **User** | **Content (played by a user)** | **Group (leader or members)** |
| **Suggested elements** | **Users** | Proposing to a user possible colleagues / friends | --no sense-- | Proposing at a group responsible possible interested colleagues to be invited |
| | **Contents** | Proposing to a user possible interesting contents | Proposing at a play of a content similar content items | Proposing at a group members possible interesting content (not much different with respect to C-C combination) |
| | **Groups** | Proposing to a user possible interesting groups | Proposing at a play of a content possible interesting groups in which similar contents are discussed | --no sense-- |
| | **Ads** | Proposing to a user possible interesting ads | Proposing at a play of a content the possible interesting ads | Proposing at a/all group member/s possible interesting ads |

# Similarity Distance

- The simplest solution for the recommendations/suggestion is to estimate the closest Users or Objects with respect to the reference User/Object
- The estimation of the closest entity between two entities described with multiple symbolic description is an instance of multidomain symbolic similarity distance among their descriptors.
- We can suppose for a while to have the possibility of estimating the similarity distance among descriptors.
- Some indexing tools, such as  Lucene/Solr, may help in doing this with a query based on information of the reference user/object.

# Complexity of Recommendation 1/2

- **Each day:**
  - N new users reach the SN
- The SN has 1 Million of users: U=10^6

- The SN has to suggest the possible friends to the new N users immediately:
  - Complexity is an O(NU)
  - N*U distances should be estimated in real time/per day
  - If N=10^6 such as on YouTube
  - Thus: 10^12 estimations of 10ms,
    - → 10^10s → which are 317 years !!!

# Complexity of Recommendation 2/2

- **Each day:**
  - M new UGC items are uploaded on the SN,
- The SN has
  - 1 Million of content: C=10^6
  - 1 Million of users: U=10^6
- The SN has to estimate the distance of that content with respect to all the other items/objects and users:
  - Complexity is an O(MC+MU)
  - M*C distances to be estimated in real time/per day
  - M*U distances to be estimated in real time/per day
  - If M= 1 Million
  - Thus: 10^12 estimations of 10ms, thus 10^10s, 2*317 years !!!

# Technologies for Recommendations

- **Objective:**
  - To provide targeted elements on the basis of the elements descriptors

- **Technical solutions**
  - **create distance matrices** and matching via direct distance or similarities estimations, very unfeasible for millions of elements would be too expensive
  - **making queries on the basis of element profile** to get the most similar. For millions or elements with several aspects or dimensions in descriptors would be very complex
  - **use some clustering to create group of elements**, also based on distances or similarities. If the groups are too many, the precisions can be low while the costs are contained.

# Similarity Distances

| Suggested element | Recipient of the suggestions: User | Recipient of the suggestions: Content (played by a user) | Recipient of the suggestions: Group (leader or members) |
|---|---|---|---|
| **Users** | D(U(s,d);U(s,d)) | **--no sense--** | D(U(s,d);G(s,d)) |
| **Contents** | D(C(s);U(s,d)) | D(C(s);C(s)) | D(C(s);G(s,d)) |
| **Groups** | D(G(s,d);U(s,d)) | D(G(s,d);C(s)) | **--no sense--** |

# General Distances Models

- Weighted Models:

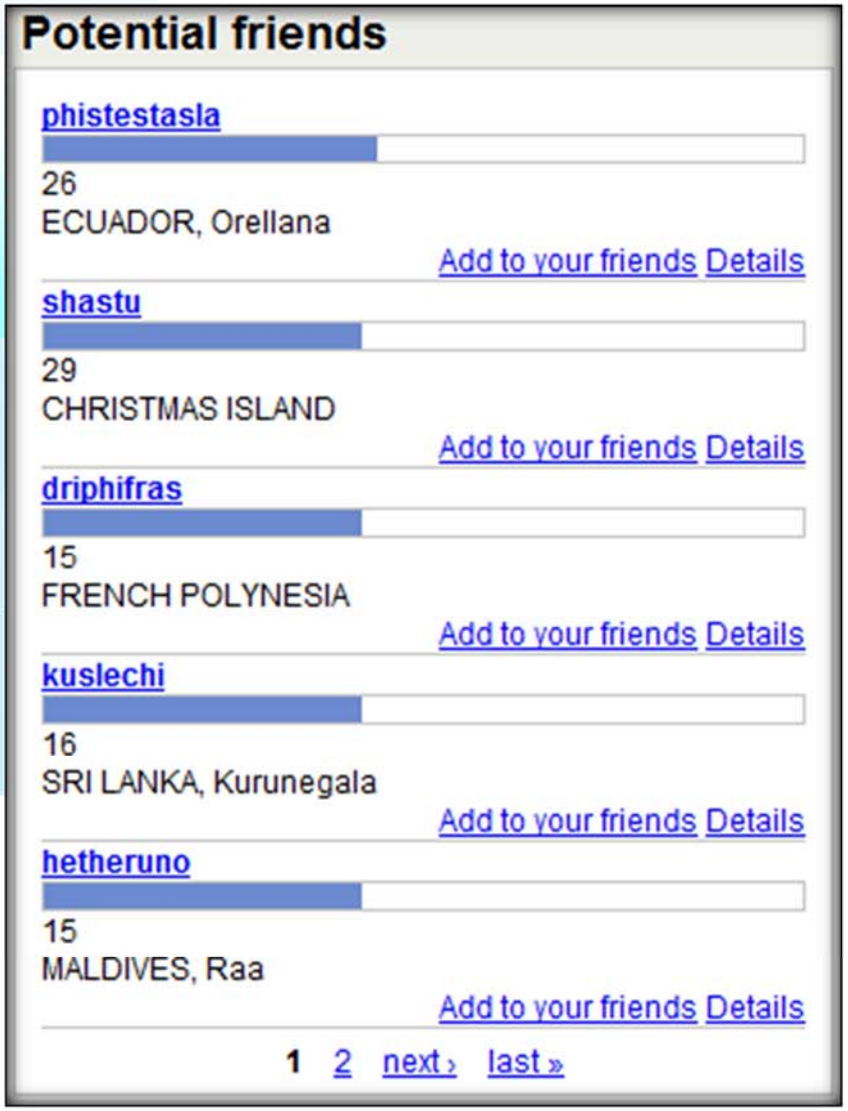
$$D(U1; U1) = k_s \sum_{i=1}^{T_1} x_i Sd_i(U1, U2) + k_d \sum_{i=1}^{T_2} y_i Dd_i(U1, U2),$$

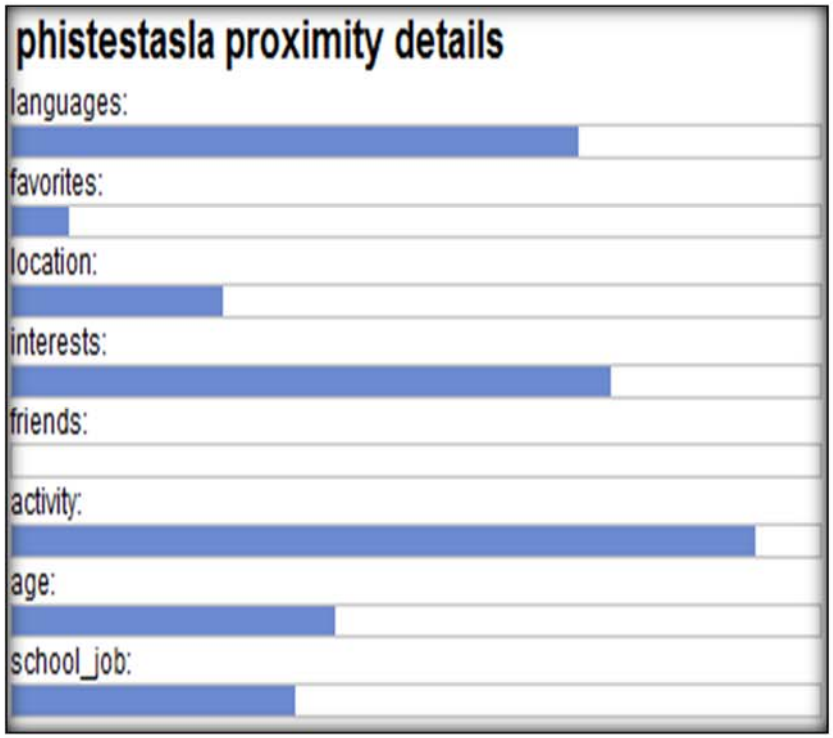
- Vector weighted models:

$$D(U1; U2) = \begin{Bmatrix} K_s(x_1 Sd_1(U1,U2), x_2 Sd_2(U1,U2), \ldots, x_n Sd_{Ts}(U1,U2)), \\ K_d(y_1 Dd_1(U1,U2), y_2 Dd_2(U1,U2), \ldots, y_n D_{Td}(U1,U2)) \end{Bmatrix}$$

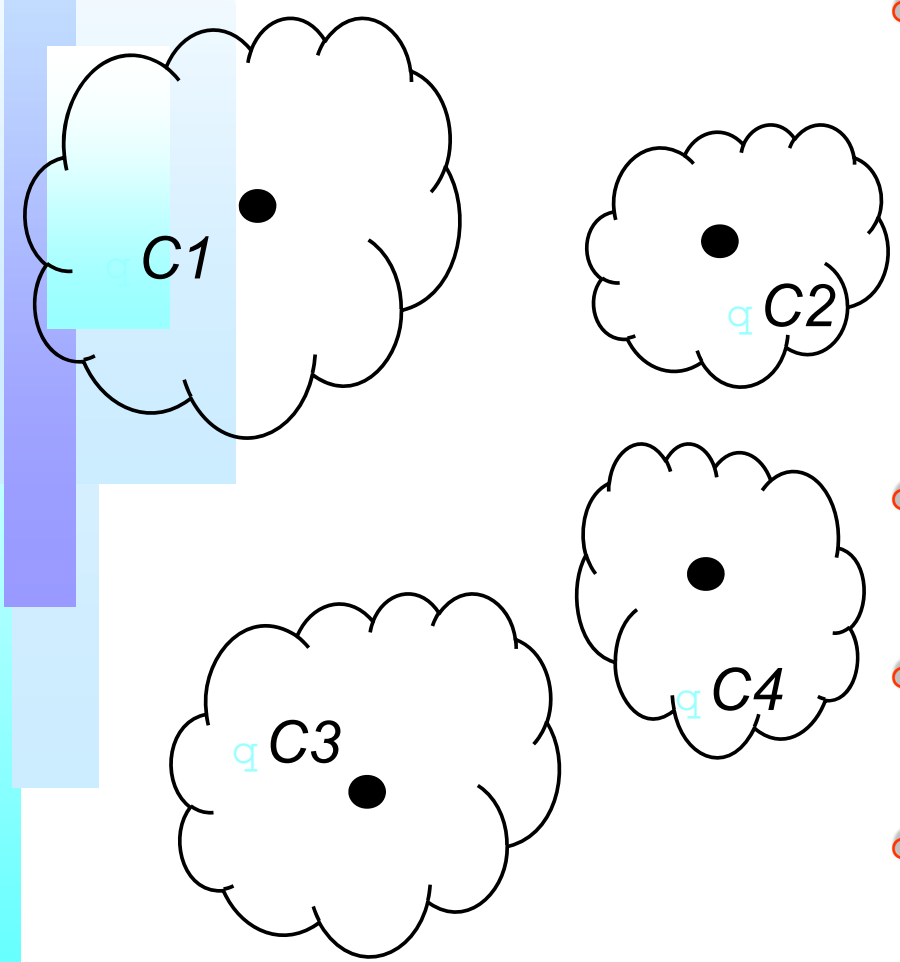
- The weights can be defined according to the SN goals.
- They can be determined by using multi-linear regressions techniques.

# Visualizzazione di Suggerimenti e dist



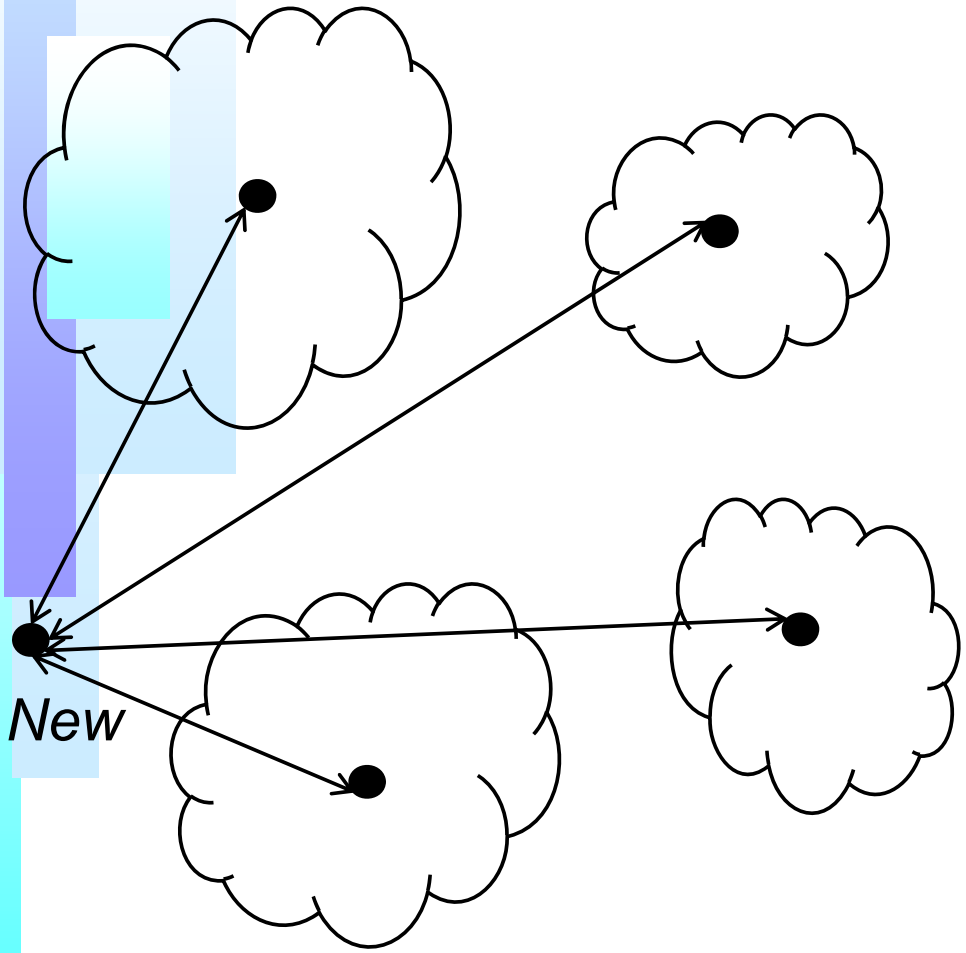



# Clustering among descriptors




- *K-Means* clustering
  - Based on a multidimensional distance model among each other
  - Define the number of clusters
  - Estimation process to maximize the cohesion among clusters
- Some items can be spare
  - They are classified in any case
- Millions of content items, thousands of clusters, …
- Periodic re-clustering taking into account all the content/objects/users

# Clustering among descriptors




- Millions of content items,
- ONLY thousands of clusters

- At each New Object
  - Distance of the new object with respect to cluster Centers
  - Reduction of complexity

- Usable on recommendations:
  - UU, UO, OO, etc.

# Clustering k-means

- Good performance in terms of scalability;
  - discovery of clusters with arbitrary shape;
  - ability to deal with noise and outliers;
  - insensitivity to order of input records;
  - support for high dimensionality.

- Complexity of an *O(NKI)*, where *N* is the number of elements, *K* the number of clusters and I the number of iterations.

- k-means has demonstrated the best performances when *N* is largely bigger than *K* and *I* (Everitt, Landau, Leese, 2001).

# K-means problems

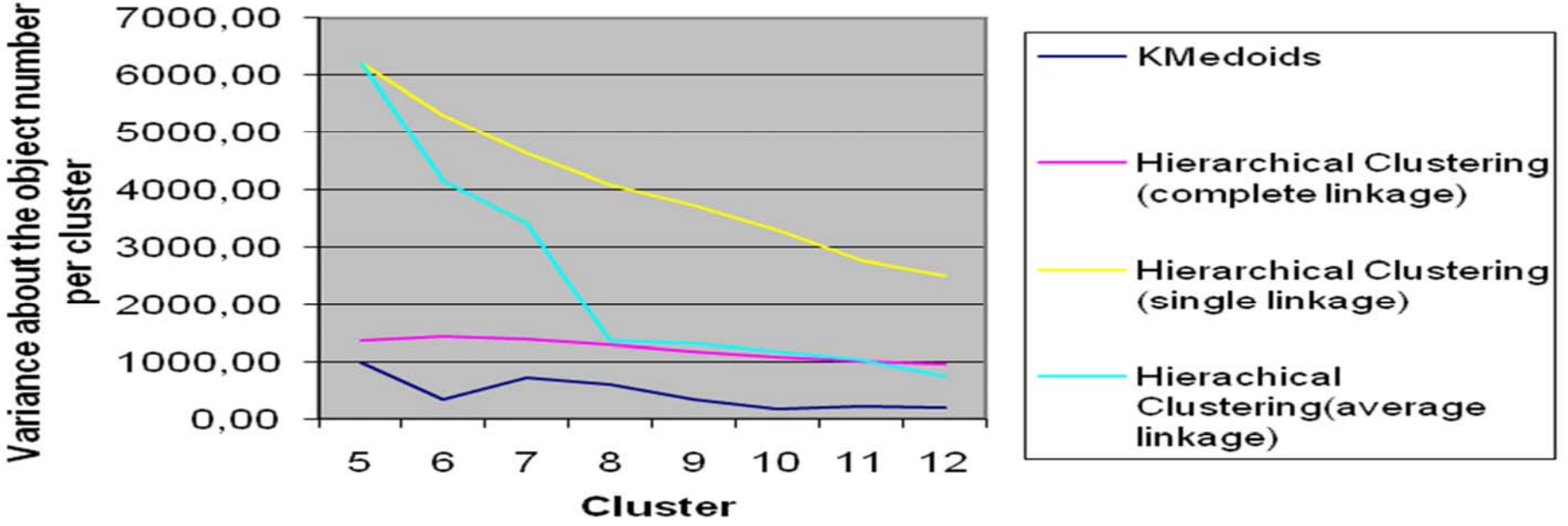
- **dependency on** the availability of numerical absolute distance estimations between two numerical values

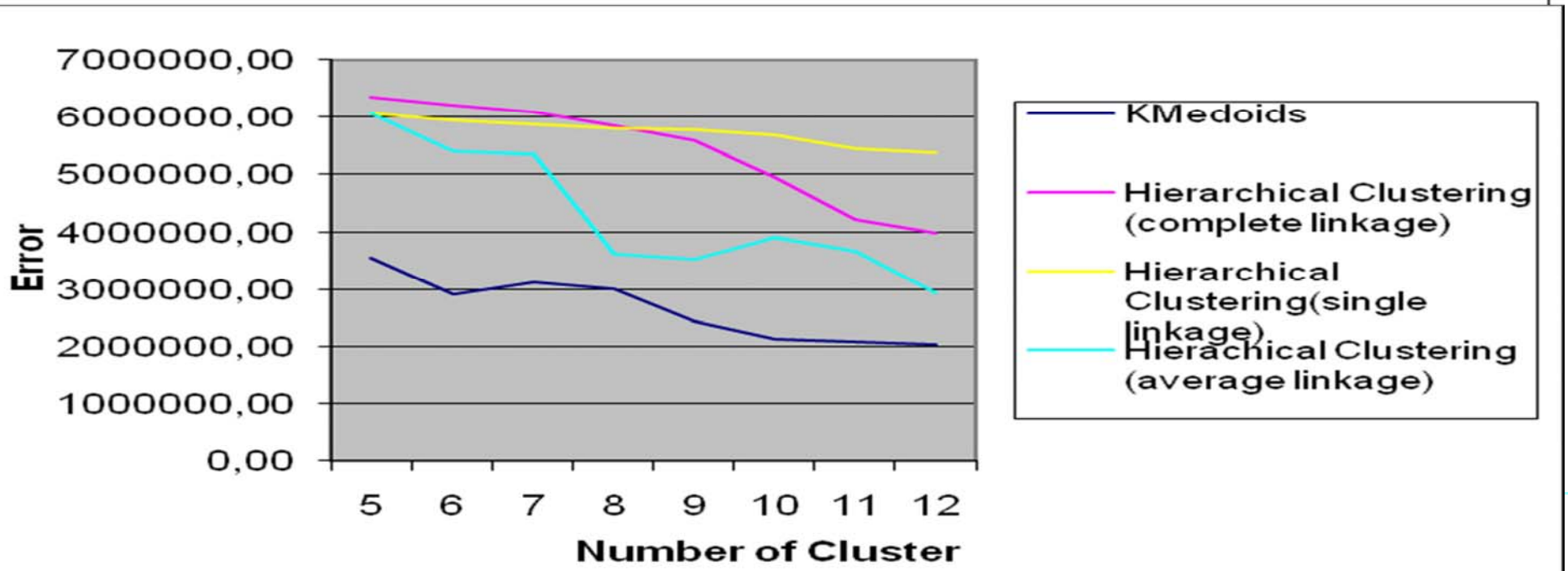
- **Unfortunatly elements descriptors** are
  - mainly symbolic and in some cases with multiple values,
  - coming from both the semantics and concepts they describe.

# K-medoids clustering

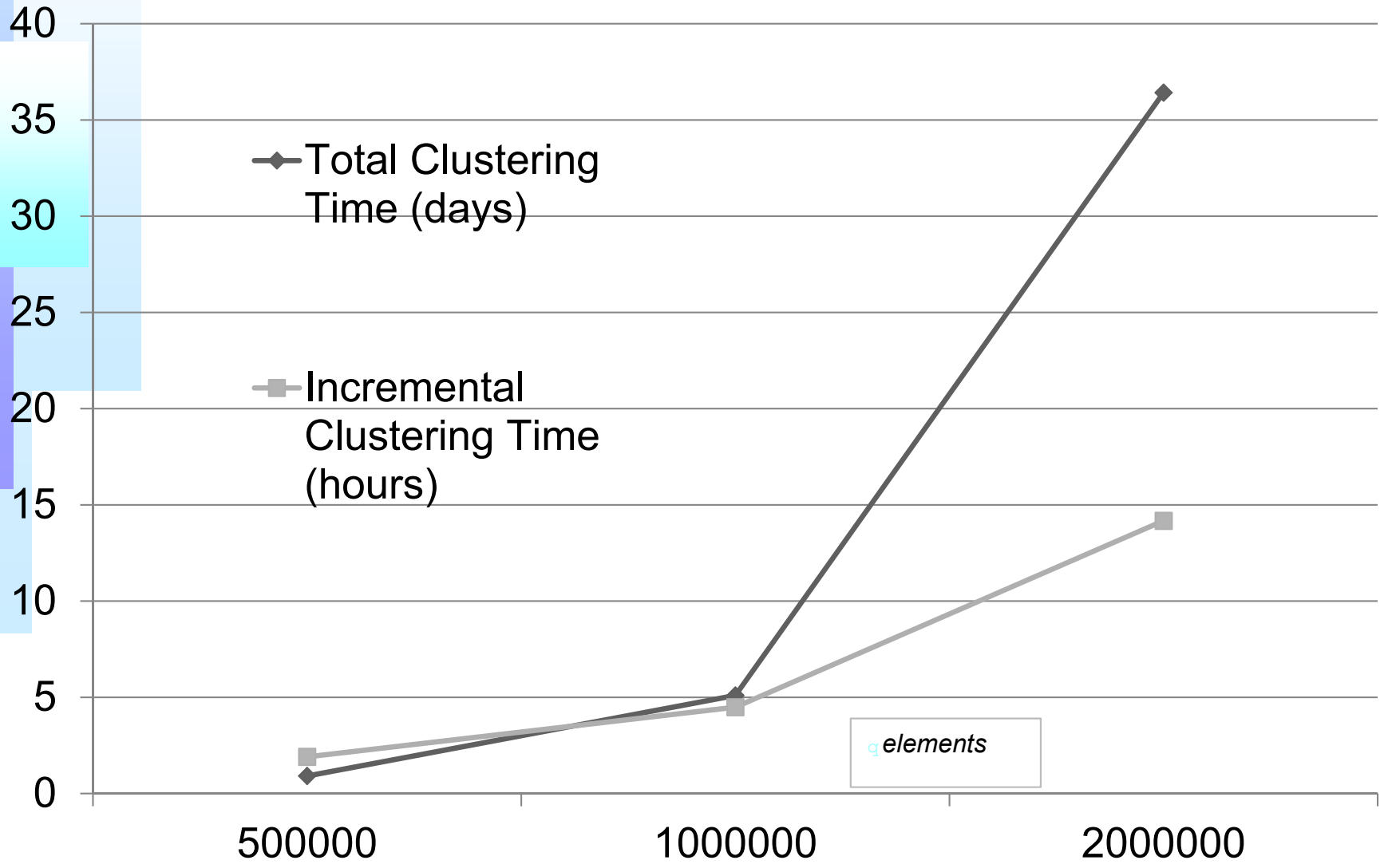
- **K-medoids** adopts as a center of the cluster the element which has the minimal average (or the median) distance among the others involved in the cluster.
- This means that the complexity is grounded on $O(K(N-K)^2)$, that for $N>>K$ is an $O(N^2)$.
  - N are the elements
  - K are the medoids/clusters
- initially the clusters centers are some selected elements (Xui & Wunsch, 2009).

# Comparison of Clustering algs

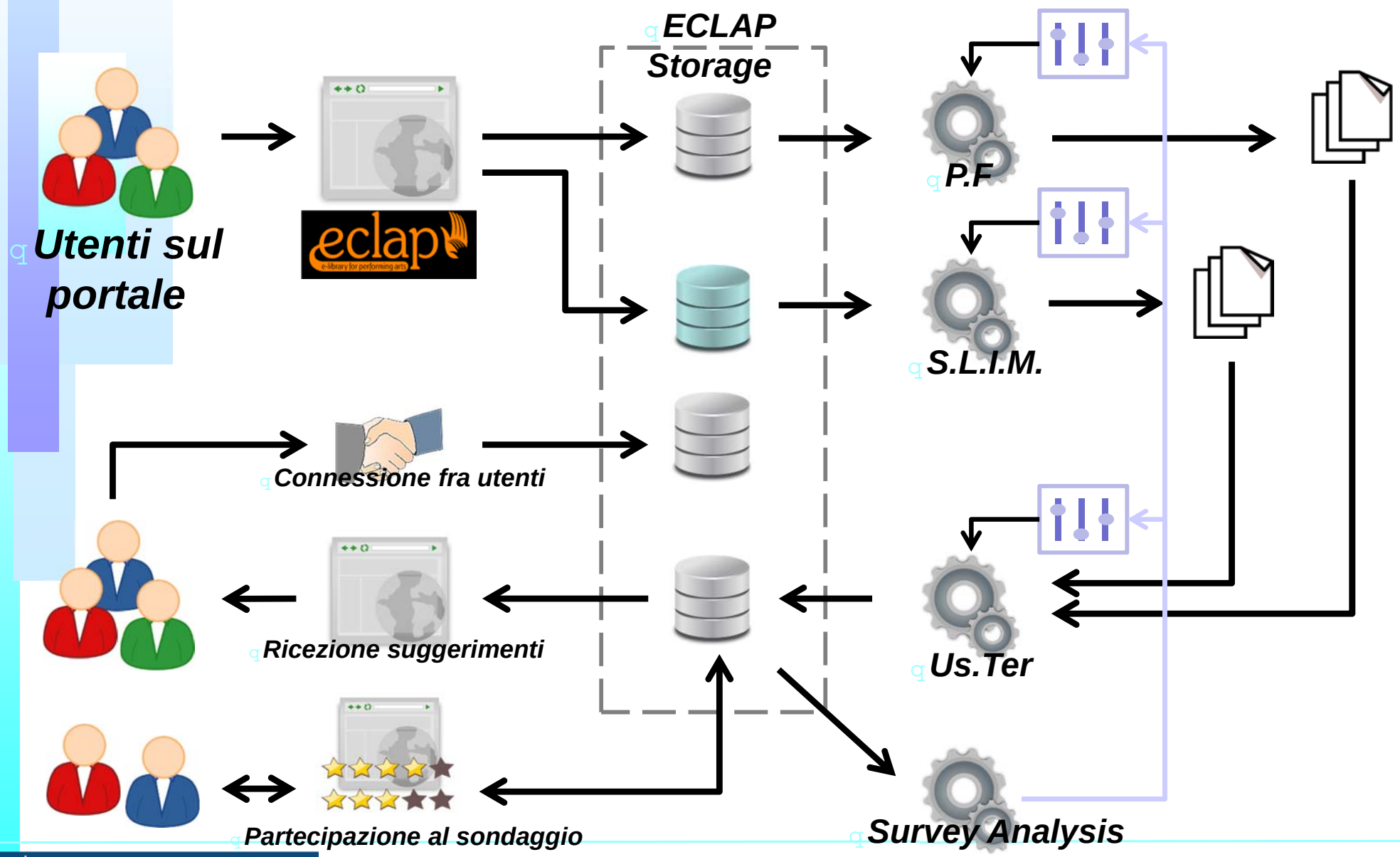


# Incremental Clustering



# Architettura del sistema



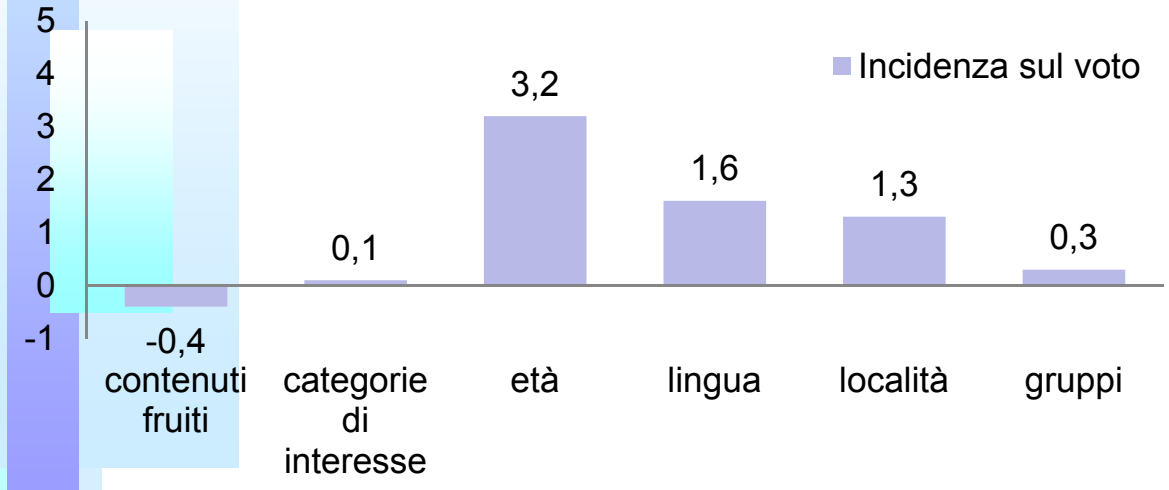
# La validazione

- I suggerimenti proposti nel sondaggio sono un sottoinsieme di quelli elaborati dal sistema e presentano una serie di informazioni relative agli utenti.
  - ✓ Viene chiesto di votare quanto un suggerimento è ritenuto interessante.
  - ✓ In questo modo non si valida la qualità delle metriche ma l'efficacia che hanno i suggerimenti sulla base delle informazioni che vengono fornite.

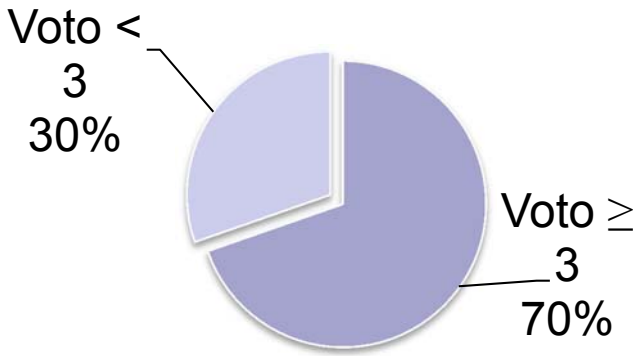

- *I parametri relativi alla generazione di una raccomandazione sono tutti tracciati e tramite i valori delle metriche e il voto lasciato dagli utenti si stima l'efficacia che ha mostrare un dettaglio relativo ad un utente o meno.*
  - ✓ *Questo è possibile solo perché i dati sulla similarità sono stati calcolati tramite le procedure precedenti.*
- *È possibile indagare quali sono le tipologie di raccomandazione che vengono gradite maggiormente.*
- *L'analisi dei dati viene effettuata tramite regressione multilineare per ottenere un modello nella forma*

$$y(A,B) = v_{language}(A,B) \cdot P_{language} \cdot \gamma_{language} + v_{location}(A,B) \cdot P_{location} \cdot \gamma_{location} + v_{friends}(A,B) \cdot P_{friends} \cdot \gamma_{friends} + v_{groups}(A,B) \cdot P_{groups} \cdot \gamma_{groups} + v_{age}(A,B) \cdot P_{age} \cdot \gamma_{age} + v_{taxonomy}(A,B) \cdot P_{taxonomy} \cdot \gamma_{taxonomy} + proximity_{dynamic}(A,B) \cdot P_{proximity_{dynamic}} \cdot \gamma_{proximity_{dynamic}}$$

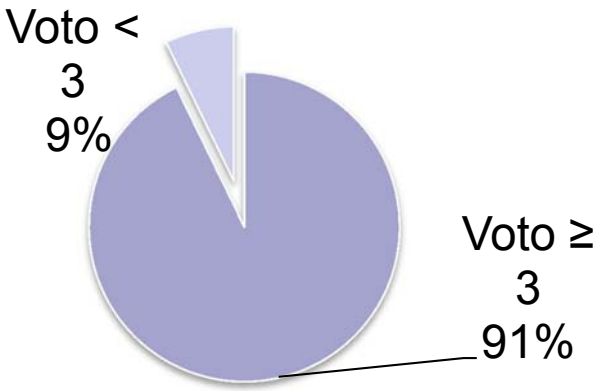
# La validazione



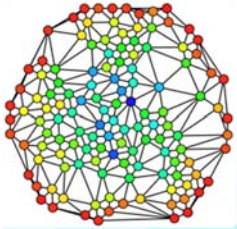

| Statistica della regressione | |
|---|---|
| R multiplo | 0, 9624 |
| F - Value | 131,7795 |
| Significatività di F | 2,3389E-33 |




- *Tipologia Serendipity*
  - ✓ *Competenze*
  - ✓ *Gruppi di appartenza*

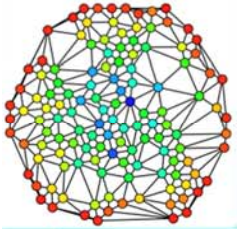



- *Tipologia Strategici*
  - ✓ *Popolarità*



# Struttura del Seminario

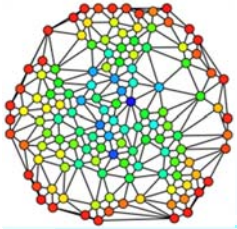
- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities



# Natural Language Processing NLP (1)

- **Scenario / Requisiti**
  - Dotare l'IA delle abilità linguistica proprie dell'essere umano
  - Comprensione e generazione del testo
  - Contesto multi-language: differenti regole e strutture a seconda della lingua

- **Applicazioni**
  - Generalizzazione delle query nei motori di ricerca
    - *"Chi si occupa di sistemi distribuiti nell'Università di Firenze ?"*
  - Supporto automatizzato per Help-Desk
  - Tutoring assistito (e-tutoring, e-teaching…)
  - Summarization: creare compendi da una collezione eterogenea di documenti
  - Machine translation: tradurre testi in lingue diverse

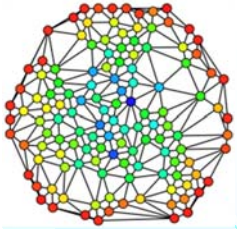


# Ambiguità dei linguaggi naturali (1)

- **Scenario / Requisiti**
  - I linguaggi naturali sono ambigui.
  - Le ambiguità si possono avere a 4 livelli:
    - ✓ *Ambiguità lessicale*: *«attacco» (verbo, sostantivo)*
    - ✓ *Ambiguità strutturale*: *«Ieri ho visto l'uomo col telescopio»*
      *«Una vecchia legge la regola»*
    - ✓ *Ambiguità semantica*: *«acuto» (persona intelligente, tipo di suono)*
    - ✓ *Ambiguità pragmatica:* *«se Buffon non gioca contro la Spagna, l'Italia perderà»*

      *L'intensione comunicativa viene recepita diversamente dagli interlocutori:*
      - *interpretazione emotiva: l'assenza di Buffon è psicologicamente fondamentale per i tifosi*
      - *Interpretazione referenziale: l'Italia senza Buffon è più debole*

**Ciò rende il processo di elaborazione del linguaggio naturale molto complicato**



# Fasi dell'elaborazione in Linguaggio Naturale (1)

**Morphological Analysis**: le parole vengono analizzate (distinzione dei morfemi che le compongono) ed i simboli (punteggiature) vengono separati dalle parole .

**Syntactic Analysis:** Le sequenze di parole sono trasformate in strutture che mostrano come le parole sono in relazione l'una con l'altra.

**Semantic Analysis**: Viene assegnato un significato alle strutture sintattiche trovate.

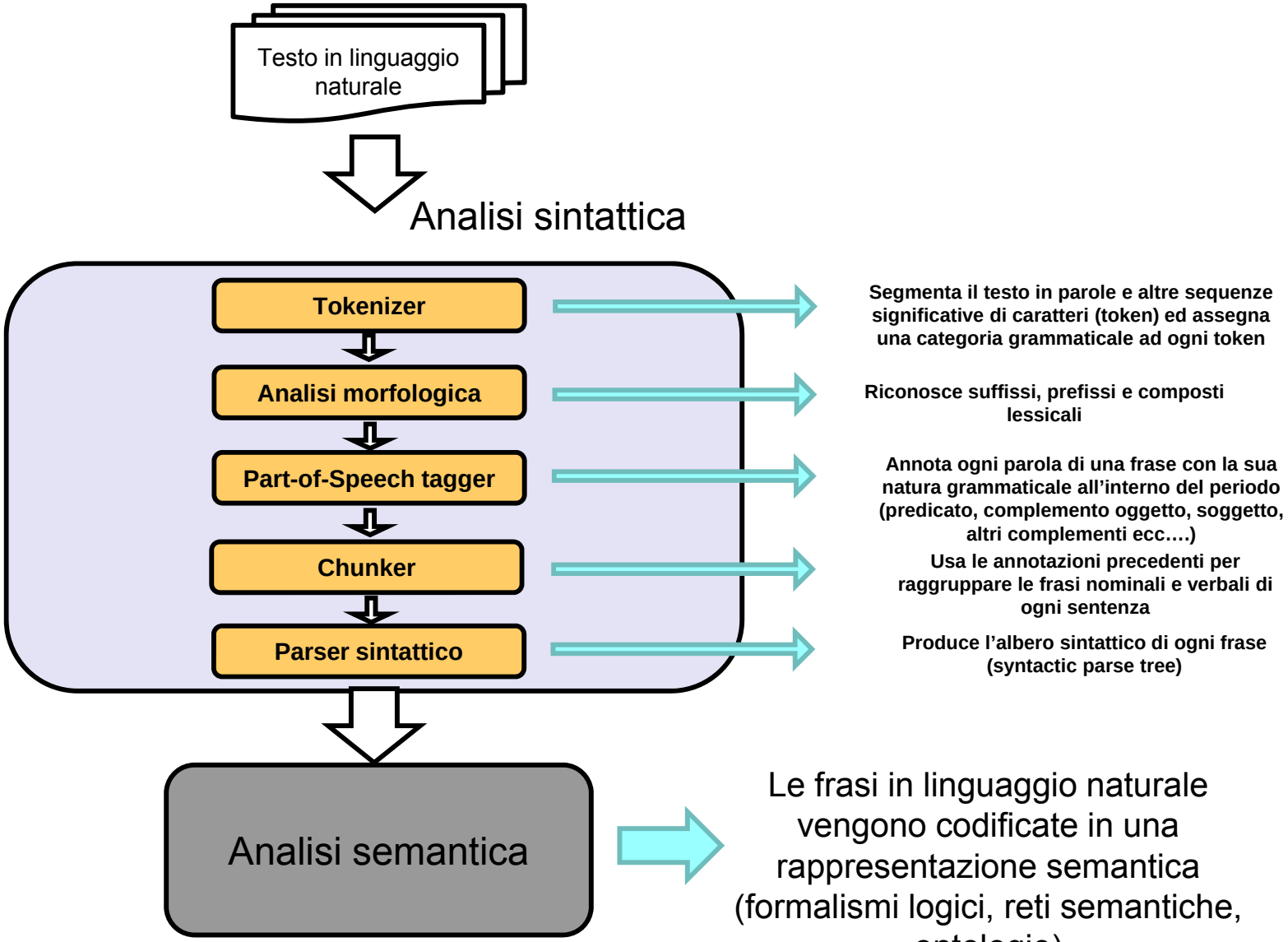
**Discourse integration**: il significato di una frase **spesso dipende dalla frase che la precede** e può influenzare quello della frase che la segue.

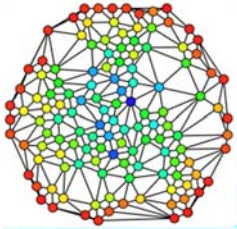
**Pragmatic Analysis:** la frase è reinterpretata per determinare il significato specifico della frase stessa.

"la porta è aperta" necessita di conoscere quale è stata l'intenzione dell'interlocutore:

- Si è creata una corrente d'aria…
- Invito ad entrare liberamente…
- Richiesta affinché qualcuno chiuda la porta…

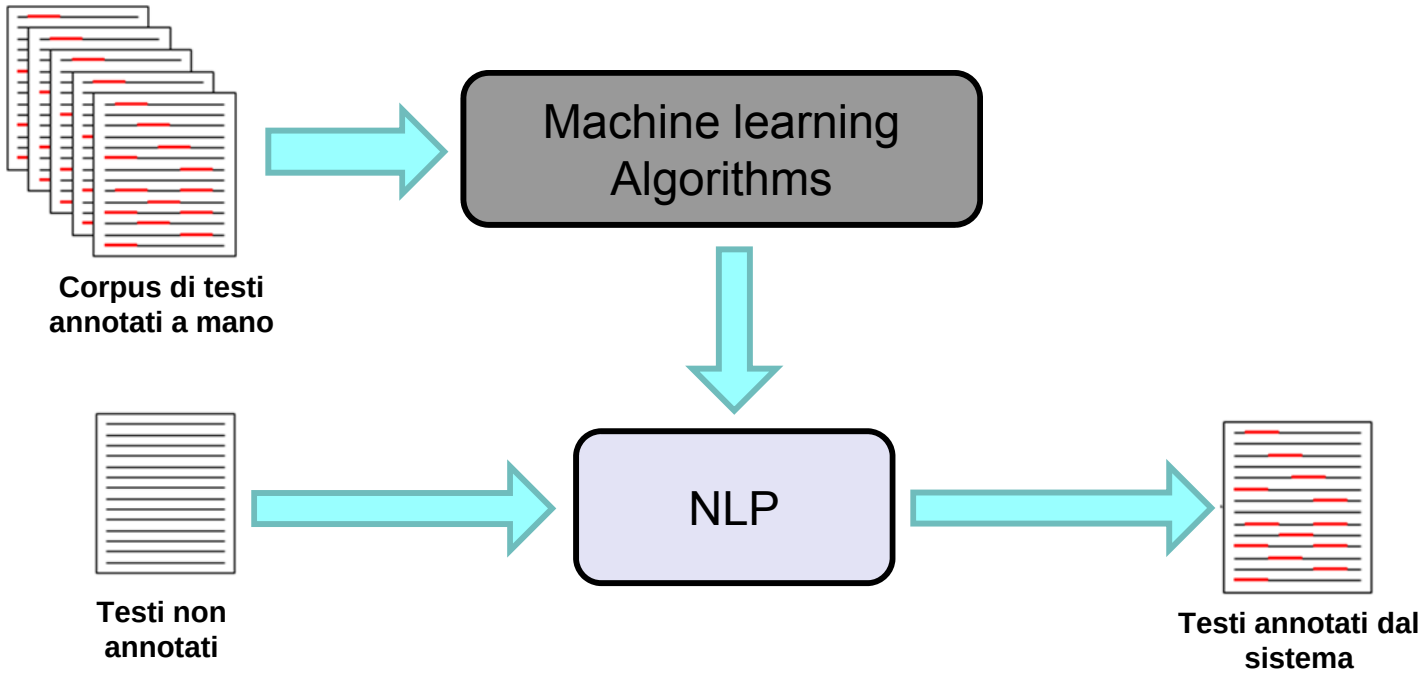
# Fasi dell'elaborazione in Linguaggio Naturale (2)

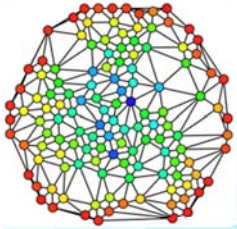




# Machine Learning

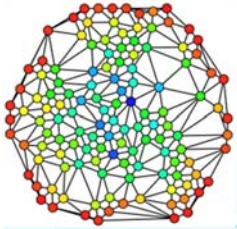
I sistemi di NLP usano principalmente algoritmi di machine learning addestrati su corpus di testi annotati a mano





# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities



# Forme di conoscenza

**Implicita**

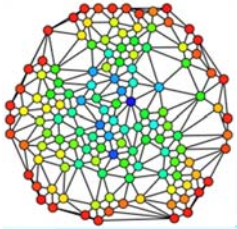
- Posseduta dalle persone
- Comunicabile in forma verbale o scritta

**Tacita**

- Presente nelle menti degli individui
- Difficile da comunicare verbalmente *(importante è l'esperienza sensoriale)*

**Esplicita**

- Strutturata (data base, XML+DTD, XML+Shema, ecc.)
- Semi-strutturata (XML, ecc.)
- Debolmente strutturata (HTML, testi tabulati, ecc.)
- Non strutturata (documenti in linguaggio naturale)



# Acquisizione e Conservazione

## Fonti di Conoscenza

### Esperienza diretta

Interazione del soggetto con il suo ambiente

### Ragionamento

Deduttivo/inferenza (conclusioni premesse)

Abduttivo (possibili cause effetti osservati)

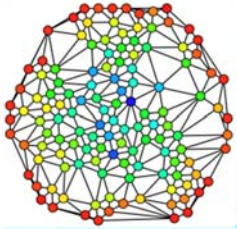
Induttivo (regole generali fatti specifici)

### Comunicazione

Uso di sistemi di segni *(in particolare il linguaggio naturale)* per trasferire informazioni da un soggetto a un altro.

## Funzione della memoria

**Capacità di Conservare** nel tempo elementi di conoscenza e soprattutto di reperirli con efficienza quando occorre farne uso.



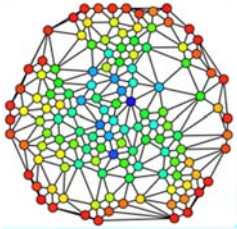
# La Logica simbolica

## Problema

**Rappresentare** la conoscenza in formato **machine-readable** *(un computer può "leggere" tale conoscenza rappresentata e utilizzarla per eseguire compiti d'interesse applicativo)*

## Soluzione

**Rappresentazione dichiarativa** tramite **logica simbolica** *(formale)*, ed in particolare la **logica dei predicati del primo ordine** *(first order logic, FOL)*



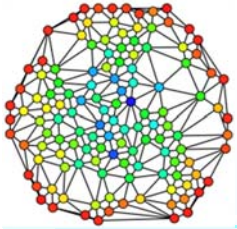
# First Order Logic

In **FOL** tutte le rappresentazioni riguardino un **insieme** non vuoto di **individui** detto **universo** *(o dominio)*.

Di questi **individui** possiamo rappresentare **proprietà** oppure **relazioni** che li leghino fra loro. Un **fatto** è dato dal sussistere:

di una **proprietà** di un determinato individuo *(es. "Barbara è bionda", "Luigi ha 21 anni")*

oppure di una **relazione** fra più individui *(es. "Alberto è più alto di Barbara", "Alberto ha dato il suo cellulare a Barbara", ecc.)*



# Linguaggi di rappresentazione

Un linguaggio per la rappresentazione di conoscenze è un **linguaggio formale**, con sintassi testuale o grafica, le cui espressioni sono utilizzate per rappresentare elementi di conoscenza.

**Esempio**: rappresentare il significato del **termine** "madre" come "donna con almeno un figlio".

Linguaggio naturale:

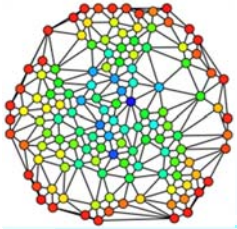
**(*x* è una madre) se e solo se (*x* è una donna ed esiste almeno un *y* tale che *x* è genitore di *y*)**

First Order Logic (FOL):

$\forall x$ **(MADRE(*x*)** $\leftrightarrow$ **DONNA(*x*)** $\wedge \exists y$ **GenDi(*x*,*y*))**

Logic Programming (LP):

```
madre(X) :- donna(X), genDi(X,Y).
```



# La deduzione

Nel contesto in cui ci stiamo muovendo, per "**ragionamento**" s'intende il ragionamento **deduttivo** *(o deduzione)*

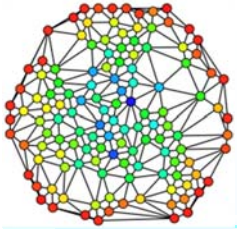
Una **deduzione** è un processo che fa passare da alcune espressioni (dette premesse o ipotesi) a un'espressione (detta conclusione o tesi), in modo tale da conservare l'eventuale verità delle premesse: in altre parole, se le premesse sono vere, lo sarà anche la conclusione.

Ad esempio, dati come premesse

1. la **definizione** di "madre"
2. il fatto che ***laura*** è una **`DONNA`**
3. il fatto che ***laura*** è **`GenitoreDi`** di ***franco***

si può dedurre come conclusione che

***laura*** è una **`MADRE`**



# Le logiche descrittive

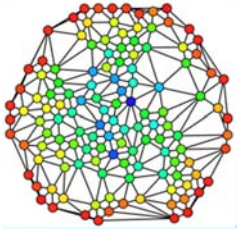
I sistemi di questo tipo hanno preso il nome di **logiche descrittive** *(description logic, DL)*. Le DL utilizzano una sintassi semplificata rispetto a FOL.

Ad esempio, le tre premesse

1. $\forall x\,(\mathtt{MADRE}(x) \leftrightarrow \mathtt{DONNA}(x) \wedge \exists y\,\mathtt{GenDi}(x,y))$
2. $\mathtt{DONNA}(laura)$
3. $\mathtt{GenDi}(laura, franco)$

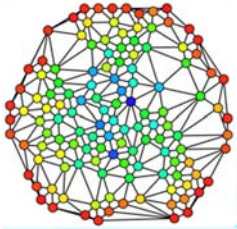
in logica descrittiva verrebbero rappresentate come

1. $\mathtt{MADRE} \equiv \mathtt{DONNA} \sqcap \exists \mathtt{GenDi}$
2. $\mathtt{DONNA}(laura)$
3. $\mathtt{GenDi}(laura, franco)$



# Risorse

- *Tutte le cose descritte con espressioni RDF vengono dette **Risorse**.*
- *Una risorsa può essere:*
  - *un'intera pagina Web (http://www.pippo.it/pluto.html)*
  - *una parte di una pagina Web*
  - *un'intera collezione di pagine (un sito Web)*
  - *un oggetto non direttamente accessibile via Web (un libro stampato)*
- *Le risorse sono sempre definite da URI*
- *Qualsiasi cosa può avere associato un URI*

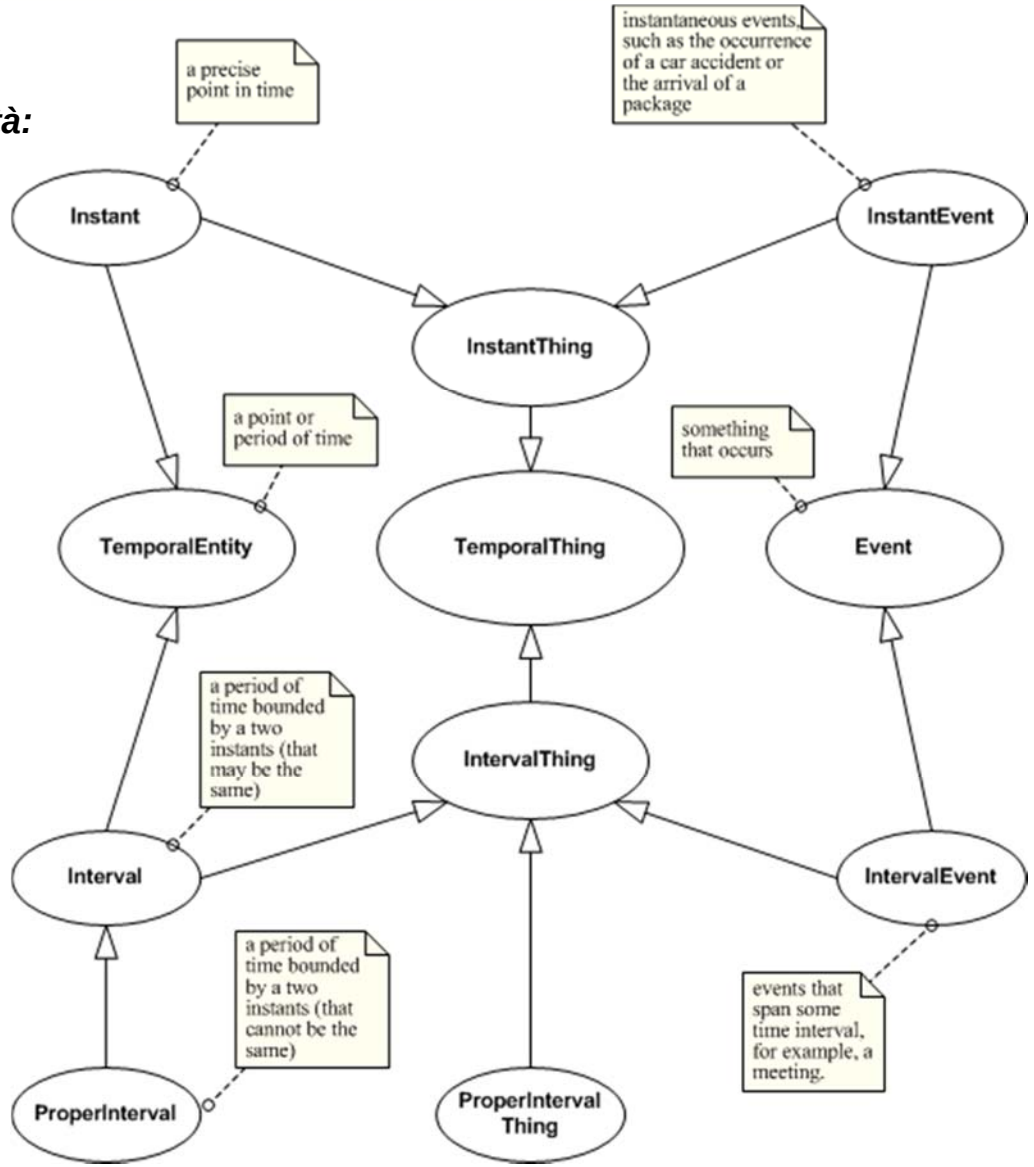


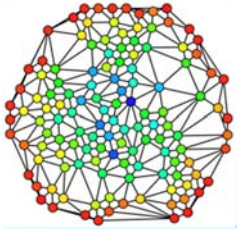
# OSIM Ontology

- Http://openmind.disit.org
- **L'ontologia di dominio di OSIM è composta da 4 ontologie diverse**
  - *Academy Life Ontology (Unifi) modella l'ateneo fiorentino in termini di docenti, corsi, strutture di affiliazione, facoltà,* gruppi di ricerca, laboratori, ecc…
  - *Friend of a Friend* (FOAF) modella le persone in termini di professori, ricercatori, phd e relazioni tipo nome, indirizzo, e-mail, settore scientifico, relazioni di conoscenza, di co-autore di pubblicazioni, ecc …..
  - Simple Knowledge Organization System (SKOS) che modella ed organizza semanticamente le competenze delle persone e dei corsi.
  - Time Ontology (TIME) che modella I concetti di intervalli ed istanti temporali per quantificare temporalmente i fatti asseriti nell'ontologia.

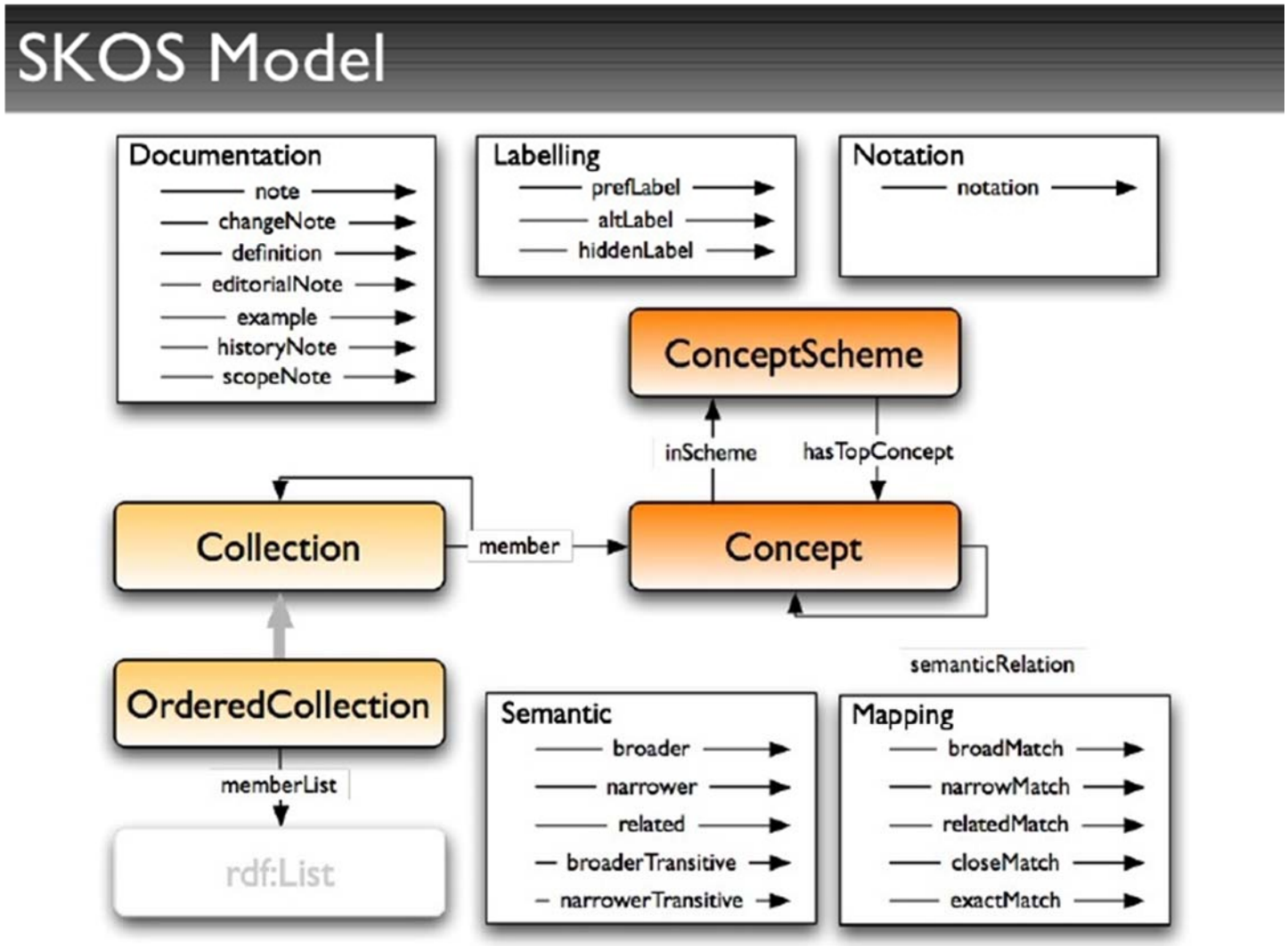
# Time Ontology

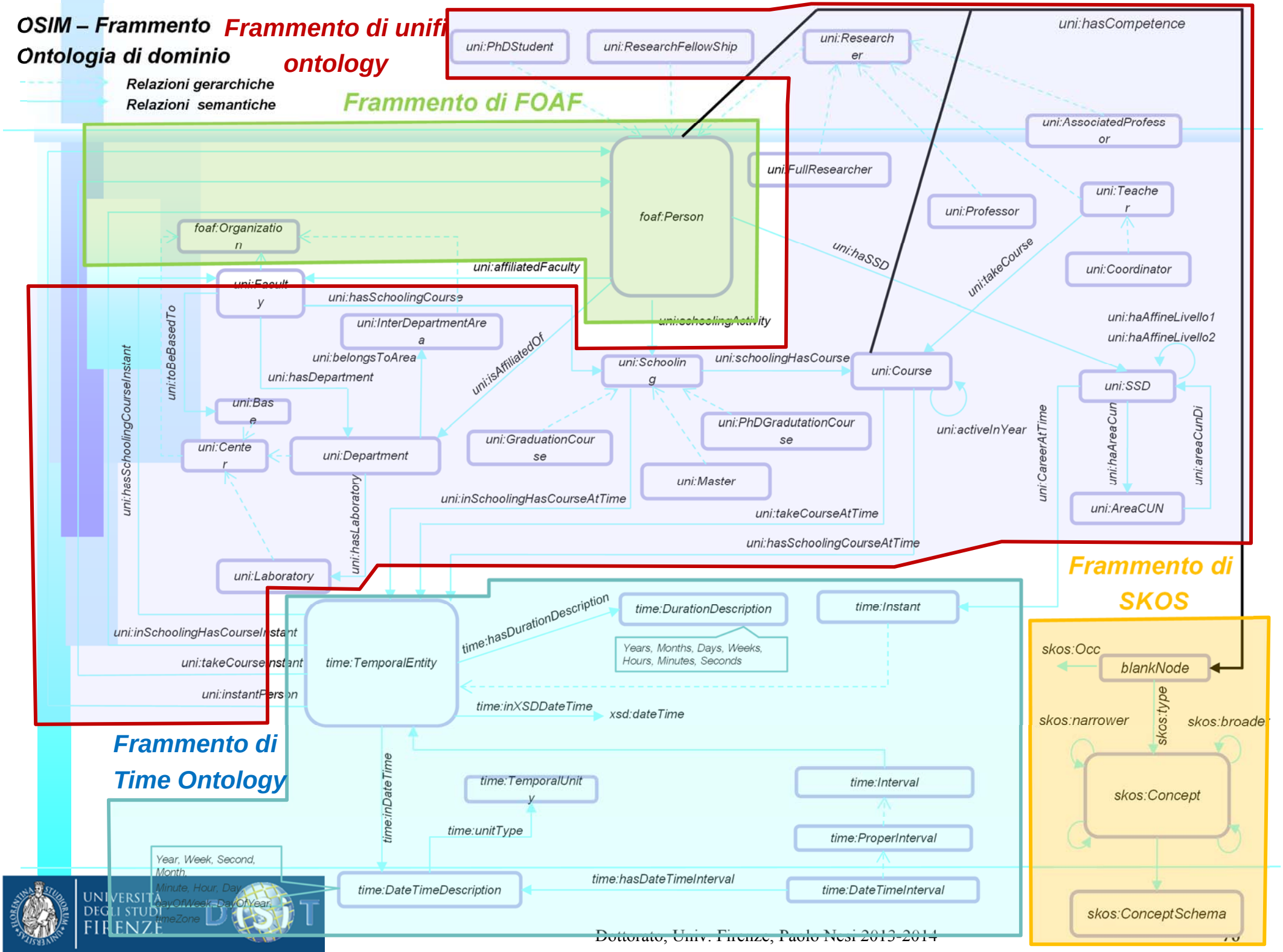
- ***Si specifica il valore di un istante ad una certa granularità:***
  - ***- Ora***
  - ***- Giorno***
  - ***- Mese***
  - ***- Anno***
  - ***- …***

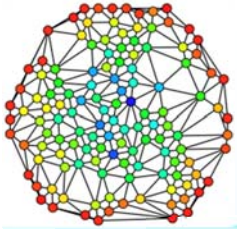




# Simple Knowledge Organization System (SKOS)

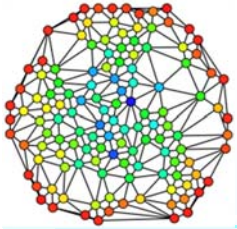



OSIM – Frammento
Ontologia di dominio
Relazioni gerarchiche
Relazioni semantiche
Frammento di unifi
ontology
Frammento di FOAF
uni:PhDStudent
uni:ResearchFellowShip
uni:Research
er
uni:hasCompetence
uni:AssociatedProfess
or
uni:FullResearcher
uni:Professor
uni:Teache
r
foaf:Person
foaf:Organizatio
n
uni:haSSD
uni:takeCourse
uni:Coordinator
uni:affiliatedFaculty
uni:Facult
y
uni:hasSchoolingCourse
uni:schoolingActivity
uni:haAffineLivello1
uni:haAffineLivello2
uni:InterDepartmentAre
a
uni:toBeBasedTo
uni:hasSchoolingCourseInstant
uni:belongsToArea
uni:hasDepartment
uni:isAffiliatedOf
uni:Schoolin
g
uni:schoolingHasCourse
uni:Course
uni:SSD
uni:Bas
e
uni:Cente
r
uni:Department
uni:GraduationCour
se
uni:PhDGradutationCour
se
uni:activeInYear
uni:CareerAtTime
uni:haAreaCun
uni:areaCunDi
uni:Master
uni:inSchoolingHasCourseAtTime
uni:takeCourseAtTime
uni:hasSchoolingCourseAtTime
uni:hasLaboratory
uni:AreaCUN
uni:Laboratory
Frammento di
SKOS
uni:inSchoolingHasCourseInstant
uni:takeCourseInstant
uni:instantPerson
time:TemporalEntity
time:hasDurationDescription
time:DurationDescription
time:Instant
Years, Months, Days, Weeks,
Hours, Minutes, Seconds
skos:Occ
blankNode
time:inXSDDateTime
xsd:dateTime
skos:narrower
skos:type
skos:broade
Frammento di
Time Ontology
time:inDateTime
time:TemporalUnit
y
time:Interval
skos:Concept
time:unitType
time:ProperInterval
Year, Week, Second,
Month,
Minute, Hour, Day,
DayOfWeek, DayOfYear,
TimeZone
time:DateTimeDescription
time:hasDateTimeInterval
time:DateTimeInterval
skos:ConceptSchema



# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities



# Tipi di ragionamento

- ***Compito*** *di ragionamento (reasoning task)*

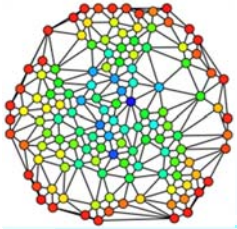
  *è caratterizzato dal tipo di enunciati che si desidera dedurre da una base di conoscenze*

- ***Procedura*** *di ragionamento*

  *l'algoritmo che consente la deduzione degli enunciati*

- ***Servizio*** *di ragionamento*

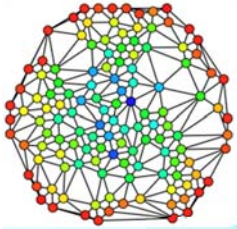
  *un servizio effettivamente implementato da uno strumento e messo a disposizione delle applicazioni che accedono alla base di conoscenze.*



# Interrogare la conoscenza (3)

- ***Built-in SPARQL***
  - *Logical:* `!`, `&&`, `||`
  - *Math:* `+`, `-`, `*`, `/`
  - *Comparison:* `=`, `!=`, `>`, `<`, ...
  - *SPARQL tests:* `isURI`, `isBlank`, `isLiteral`, `bound`
  - *SPARQL accessors:* `str`, `lang`, `datatype`
  - *Other:* `sameTerm`, `langMatches`, `regex`

- ***E' possibile effettuare l'unione di più graph paths tramite la clausola UNION***

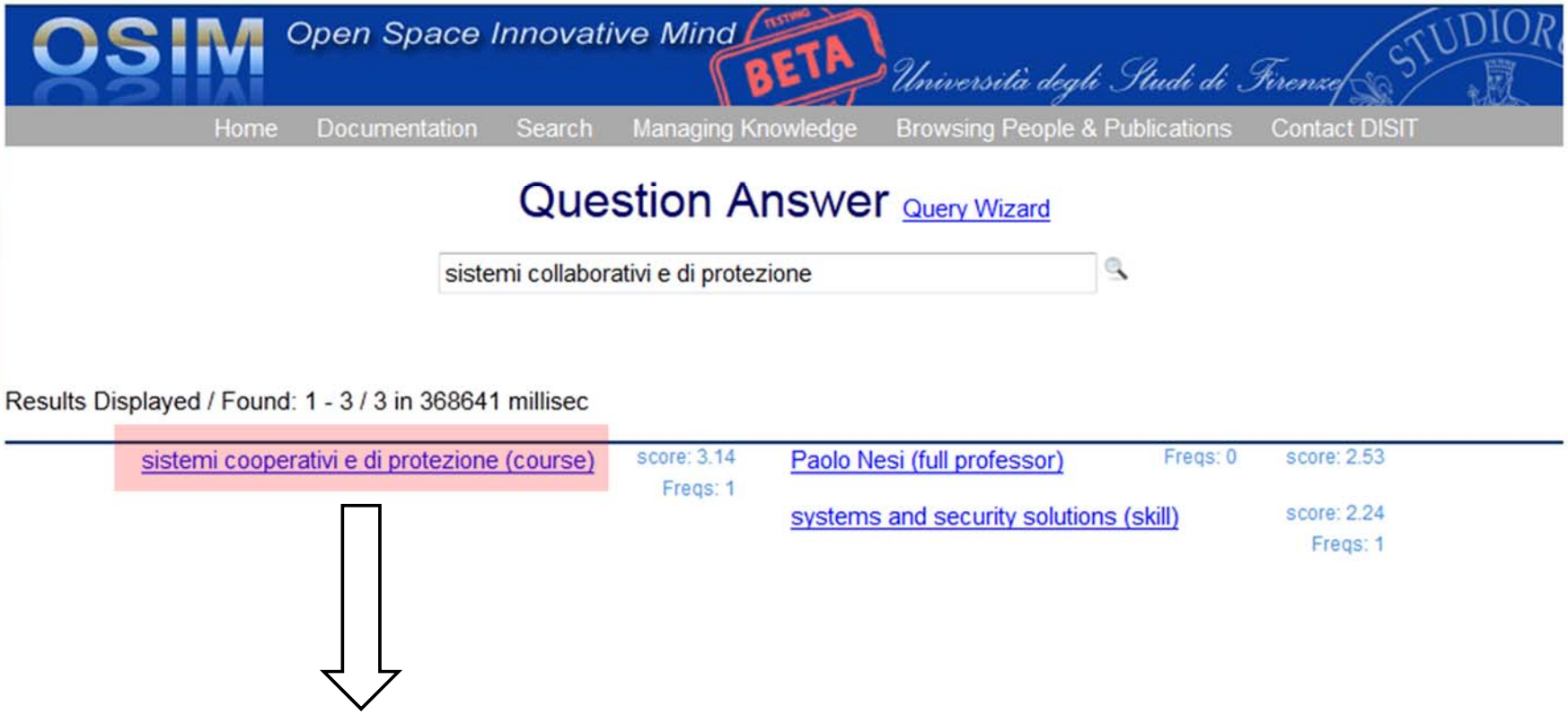


# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities

# OSIM: Conoscenza

http://openmind.disit.org




- *"navighiamo" la conoscenza relativa al corso*

# OSIM: Conoscenza relativa alla competenza «programming»

**programmazione (skill )**

Competenze collegate

Concetti broader

informatica

Concetti narrower

esperienza nella programmazione
espressioni regolari
fondamenti di programmazione
funzioni
lingua
linguaggio di programmazione
linguaggio macchina
livello di programmazione
modello di programmazione

Concetti related

Nessun concetto presente

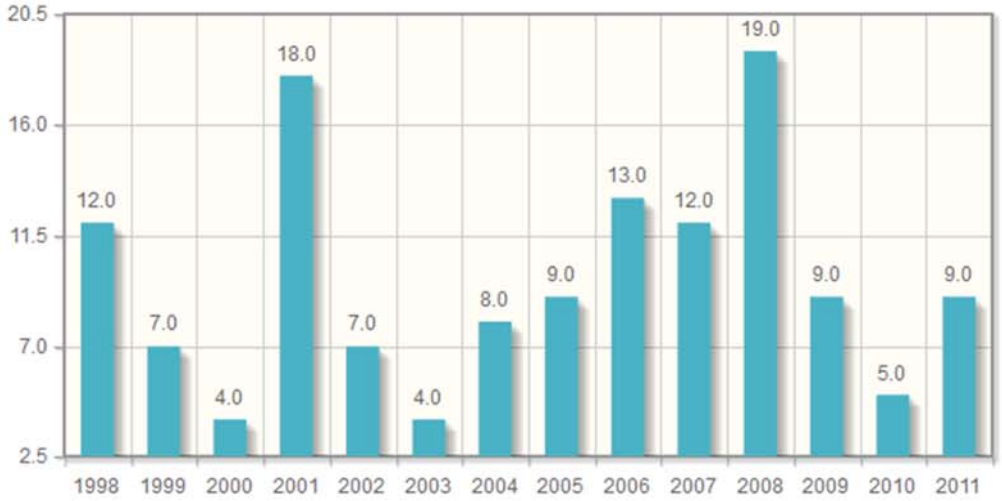
Persone

Persone con questa competenza

Michele Boreale (4/170)
Michele Loreti (1/170)
Paola Cappanera (1/170)
Paolo Frasconi (9/170)
Paolo Nesi (12/170)
Pierluigi Crescenzi (7/170)
Pietro Pala (1/170)
Rosario Pugliese (9/170)
Simone Marinai (2/170)
Stefano Berretti (13/170)

# OSIM: Conoscenza relativa alle pubblicazioni e i coautori di «Paolo Nesi»

**Totale pubblicazioni: 167**




**Anno:**

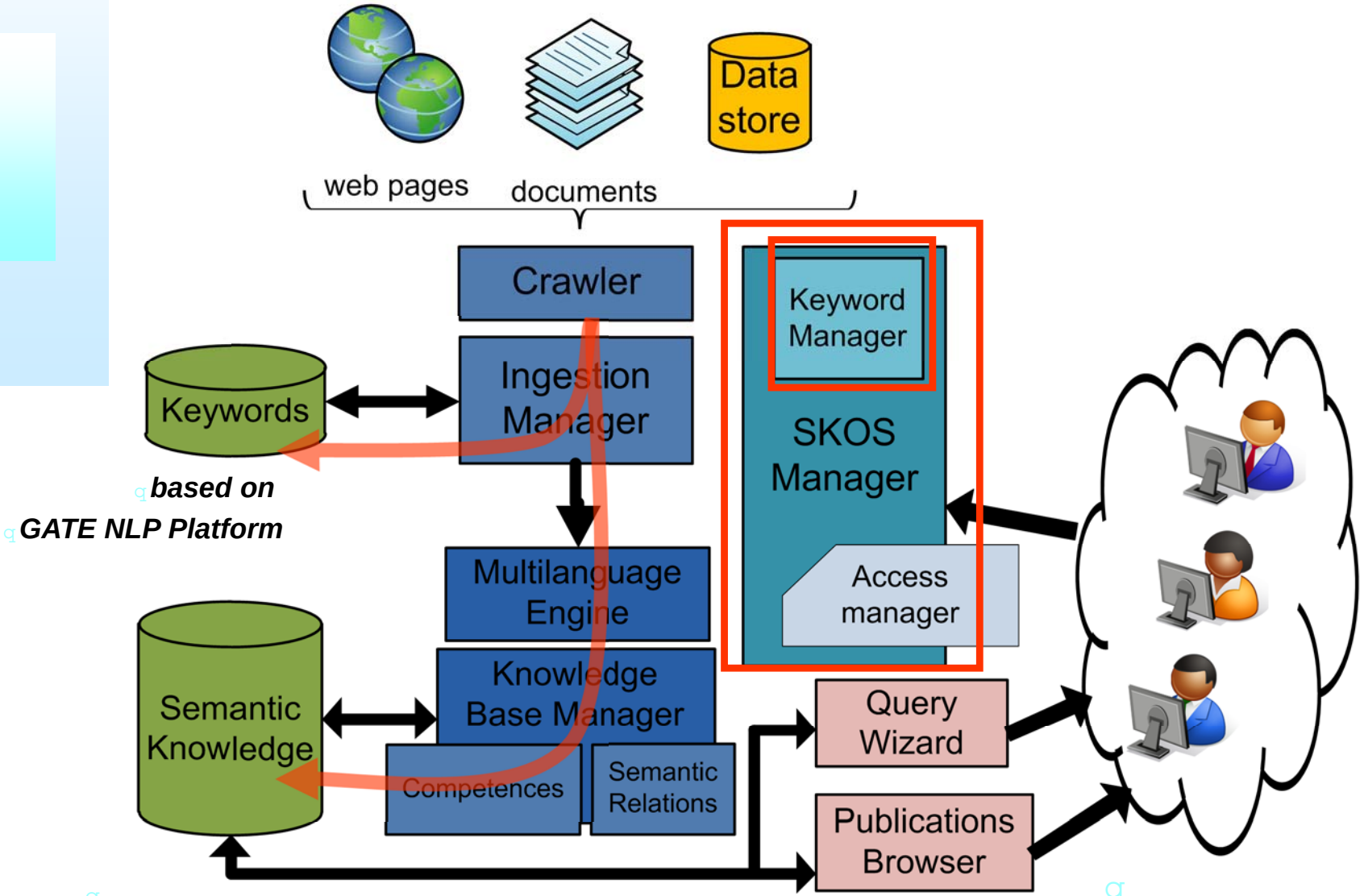
1998 (12) 1999 (7) 2000 (4) 2001 (18) 2002 (7) 2003 (4) 2004 (8) 2005 (9) 2006 (13) 2007 (12) 2008 (19) 2009 (9) 2010 (5) 2011 (9)

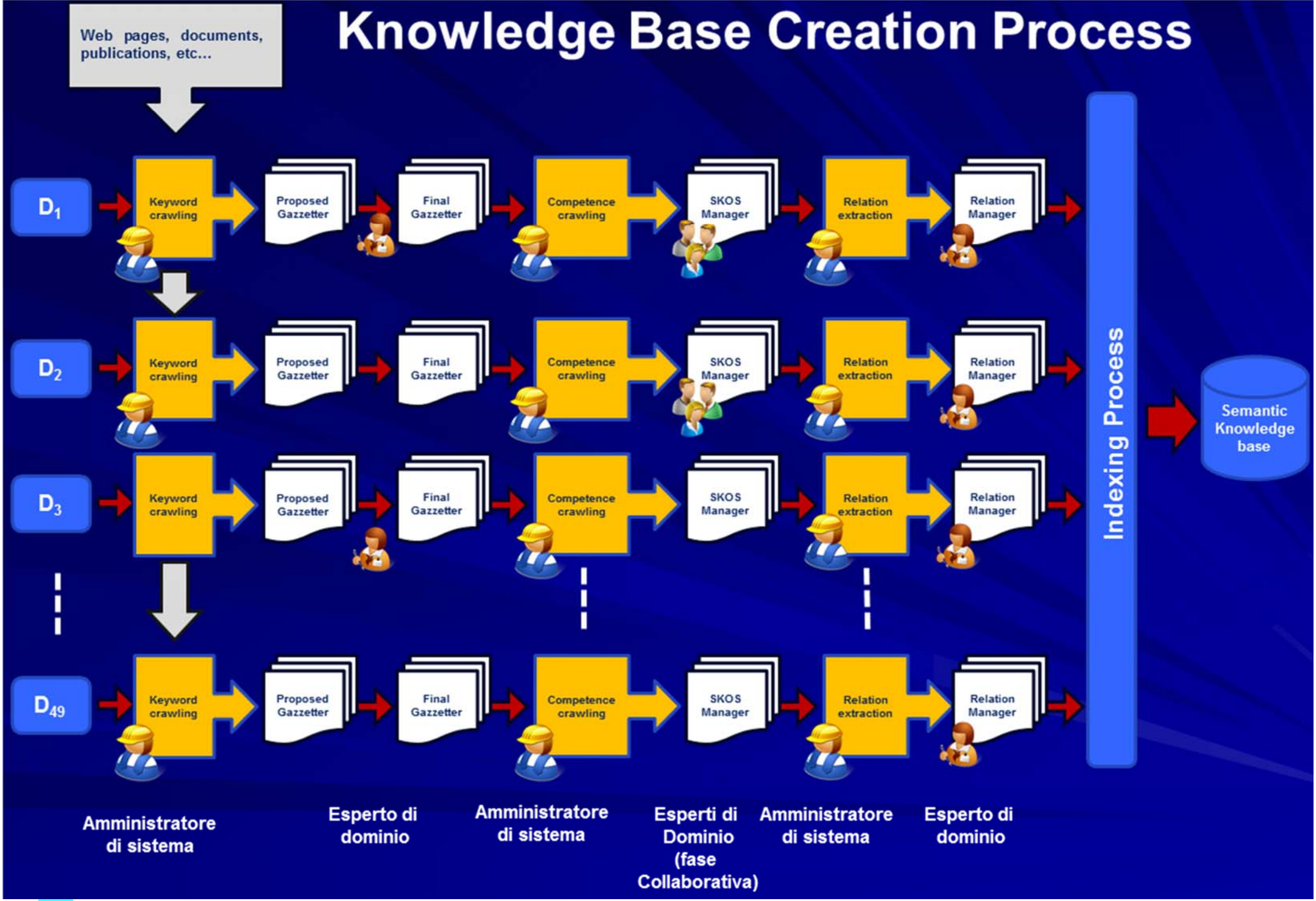
Elenco di tutte le pubblicazioni (167)

**Autori che hanno lavorato con questa persona:**

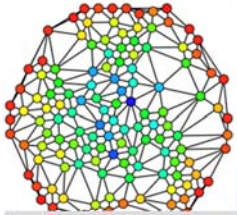
| | |
|---|---|
| ARGENTI FABRIZIO (Registrato CINECA) | Visualizza le pubblicazioni in comune (2) |
| BALDASSARRE ANTONIO (Registrato CINECA) | Visualizza le pubblicazioni in comune (1) |
| BELLINI PIERFRANCESCO (Registrato CINECA) | Visualizza le pubblicazioni in comune (41) |
| BRUNO IVAN (Registrato CINECA) | Visualizza le pubblicazioni in comune (22) |
| BUCCI GIACOMO (Registrato CINECA) | Visualizza le pubblicazioni in comune (5) |
| CENNI D. (Non Registrato CINECA) | Visualizza le pubblicazioni in comune (3) |
| CENNI DANIELE (Registrato CINECA) | Visualizza le pubblicazioni in comune (5) |
| DE LUCIA M. (Non Registrato CINECA) | Visualizza le pubblicazioni in comune (1) |
| DE LUCIA MAURIZIO (Registrato CINECA) | Visualizza le pubblicazioni in comune (1) |
| EVANGELISTI ATTILIO (Registrato CINECA) | Visualizza le pubblicazioni in comune (1) |

# OSIM: Architettura




Knowledge Base Creation Process
Web pages, documents, publications, etc...
D1
D2
D3
D49
Keyword crawling
Proposed Gazzetter
Final Gazzetter
Competence crawling
SKOS Manager
Relation extraction
Relation Manager
Indexing Process
Semantic Knowledge base
Amministratore di sistema
Esperto di dominio
Amministratore di sistema
Esperti di Dominio (fase Collaborativa)
Amministratore di sistema
Esperto di dominio

# Keyword management



Welcome root Logout OSIM Managing Knowledge HOME

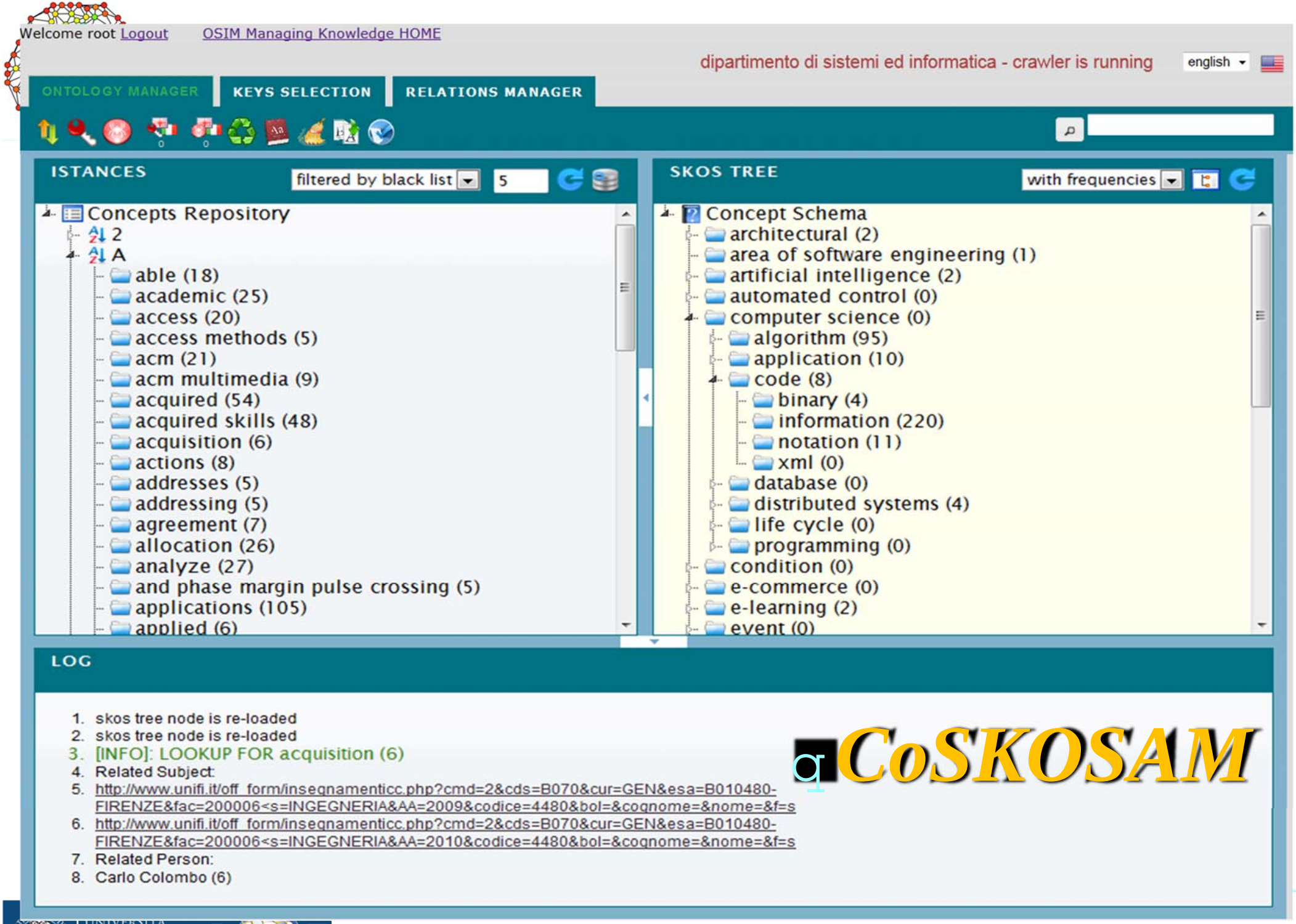
dipartimento di matematica per le decisioni - crawler is running english

ONTOLOGY MANAGER | KEYS SELECTION | RELATIONS MANAGER

all 20/113 (#2258)

| id | value | translated values | occurrencies | gazetteer | black list | no action | lang | Proposed |
|---|---|---|---|---|---|---|---|---|
| 9390 | algebra | algebra | 9 | | | | en | 0 |
| 1201 | complementary | complementare | 9 | | | | en | 0 |
| 9139 | changes | variazioni | 9 | | | | en | 0 |
| 9143 | horizon | orizzonte | 9 | | | | en | 0 |
| 7611 | decomposition | decomposizione | 9 | | | | en | 0 |
| 9148 | infinite | infiniti | 9 | | | | en | 0 |
| 7365 | laboratory | laboratorio | 9 | | | | en | 0 |
| 8646 | angles | angoli | 9 | | | | en | 0 |
| 8397 | embrechts | embrechts | 9 | | | | en | 0 |
| 9427 | bond | obbligazione | 9 | | | | en | 0 |
| 9431 | fields | campi | 9 | | | | en | 0 |
| 9689 | edition | edizione | 9 | | | | en | 0 |
| 8673 | min | min | 9 | | | | en | 0 |
| 482 | decomposition | scomposizione | 9 | | | | en | 0 |
| 8426 | year-old | anni | 9 | | | | en | 0 |

Trova: mate Successivo Precedente Evidenzia Maiuscole/minuscole

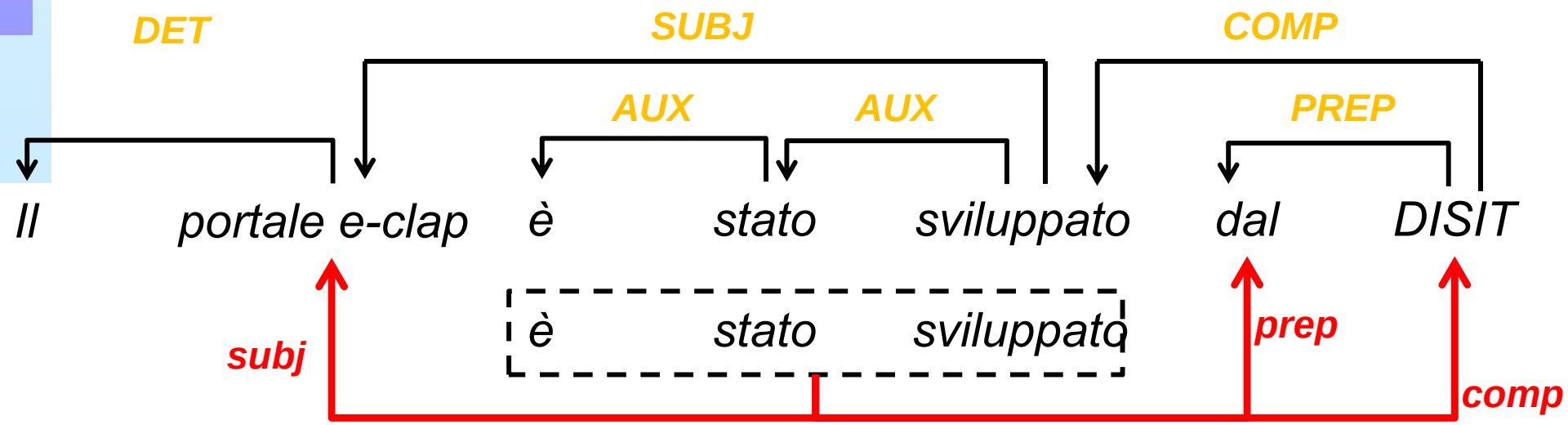

Welcome root Logout
OSIM Managing Knowledge HOME
dipartimento di sistemi ed informatica - crawler is running
english
ONTOLOGY MANAGER
KEYS SELECTION
RELATIONS MANAGER
ISTANCES
filtered by black list
5
Concepts Repository
2
A
able (18)
academic (25)
access (20)
access methods (5)
acm (21)
acm multimedia (9)
acquired (54)
acquired skills (48)
acquisition (6)
actions (8)
addresses (5)
addressing (5)
agreement (7)
allocation (26)
analyze (27)
and phase margin pulse crossing (5)
applications (105)
applied (6)
SKOS TREE
with frequencies
Concept Schema
architectural (2)
area of software engineering (1)
artificial intelligence (2)
automated control (0)
computer science (0)
algorithm (95)
application (10)
code (8)
binary (4)
information (220)
notation (11)
xml (0)
database (0)
distributed systems (4)
life cycle (0)
programming (0)
condition (0)
e-commerce (0)
e-learning (2)
event (0)
LOG
1. skos tree node is re-loaded
2. skos tree node is re-loaded
3. [INFO]: LOOKUP FOR acquisition (6)
4. Related Subject:
5. http://www.unifi.it/off_form/insegnamenticc.php?cmd=2&cds=B070&cur=GEN&esa=B010480-FIRENZE&fac=200006<s=INGEGNERIA&AA=2009&codice=4480&bol=&cognome=&nome=&f=s
6. http://www.unifi.it/off_form/insegnamenticc.php?cmd=2&cds=B070&cur=GEN&esa=B010480-FIRENZE&fac=200006<s=INGEGNERIA&AA=2010&codice=4480&bol=&cognome=&nome=&f=s
7. Related Person:
8. Carlo Colombo (6)
CoSKOSAM

# Relazioni verbali (1)

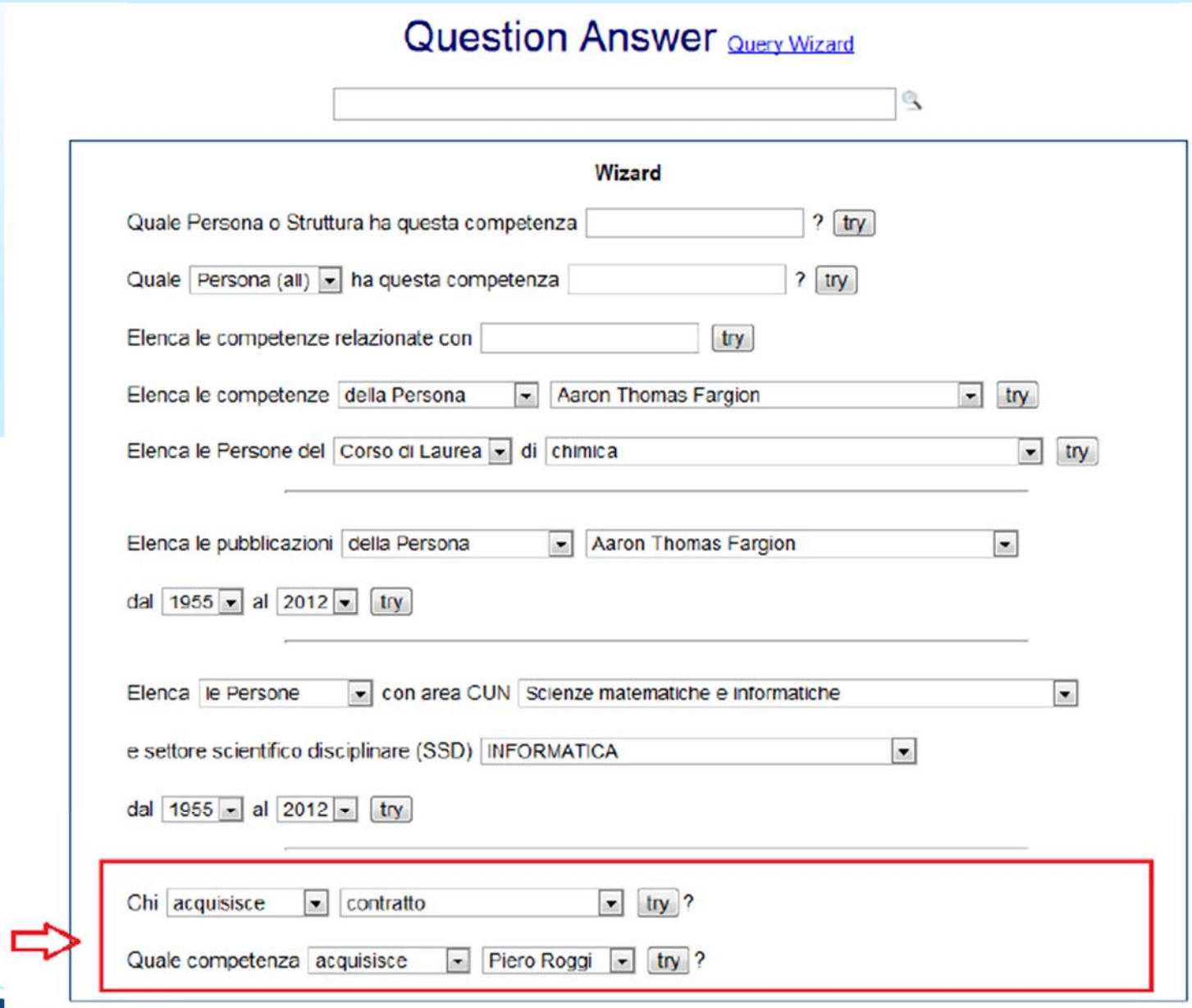
- Il sistema permette di estrarre relazioni semantiche non tassonomiche dalle pagine dell'Universita' di Firenze che riguardano i corsi e le persone

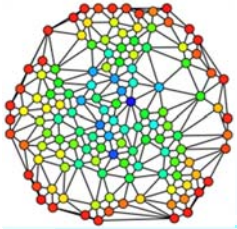
- Per estrarre le relazioni semantiche dai testi, il sistema analizza le frasi, genera un grafo a partire dall'albero delle dipendenze di ogni frase e individua eventuali **PATTERN PREDEFINITI** presenti sul grafo

  - ***Esempio**: Il portale e-clap è stato sviluppato dal DISIT*



# Relazioni verbali - Query





# Validazione (1)

- OSIM è stato validato e confrontato con Marsilius
  - La performance dei due sistemi, su un sottoinsieme di 8 dipartimenti, è stata misurata e confrontata
- Per sistemi di IR viene tipicamente utilizzato lo standard di validazione TREC (Text REtrieval Conference)

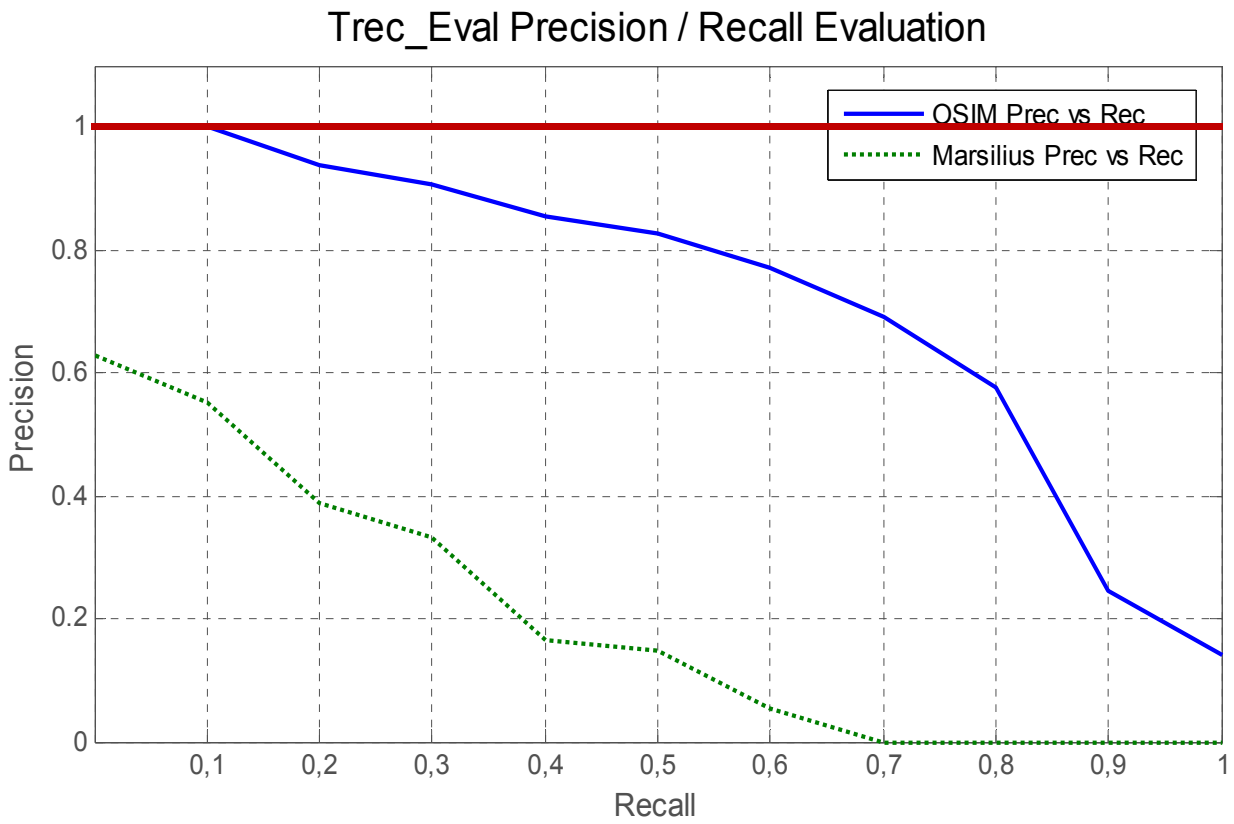
$$Precision = \frac{\#(\text{relevant items retrieved})}{\#(\textit{retrieved}\ \text{items})}$$

$$Recall = \frac{\#(\text{relevant items retrieved})}{\#(\textit{relevant}\ \text{items})}$$

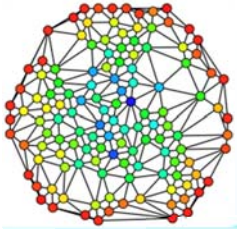
- I documenti / risultati sono considerati rilevanti se soddisfano il tipo di informazione richiesta, non solo perché contengono tutte le keywords immesse nella ricerca.
- Query di validazione eseguite da esperti dei vari domini di conoscenza analizzati

# Validazione (2)

- Set di query su un sottoinsieme di 4 dipartimenti
- Profondità dei risultati fissata N = 20
- Curva *Precision – Recall* ottenuta con il software standard *Trec_Eval*




- ***Ottimo ideale***



# Struttura del Seminario

- Social Networking and knowledge
- Semantic and Social Networks
- Recommendations and Suggestions
- Natural Language Processing System
- Knowledge Representation System
- Reasoning System
- Sistema OSIM
- Smart Cities

# Main Smart City and OD Projects

- Linked Open Graph: http://log.disit.org
- Sii-Mobility http://www.sii-mobility.org
- SmartCityOntology Coll@bora
- SACVAR and OSIM

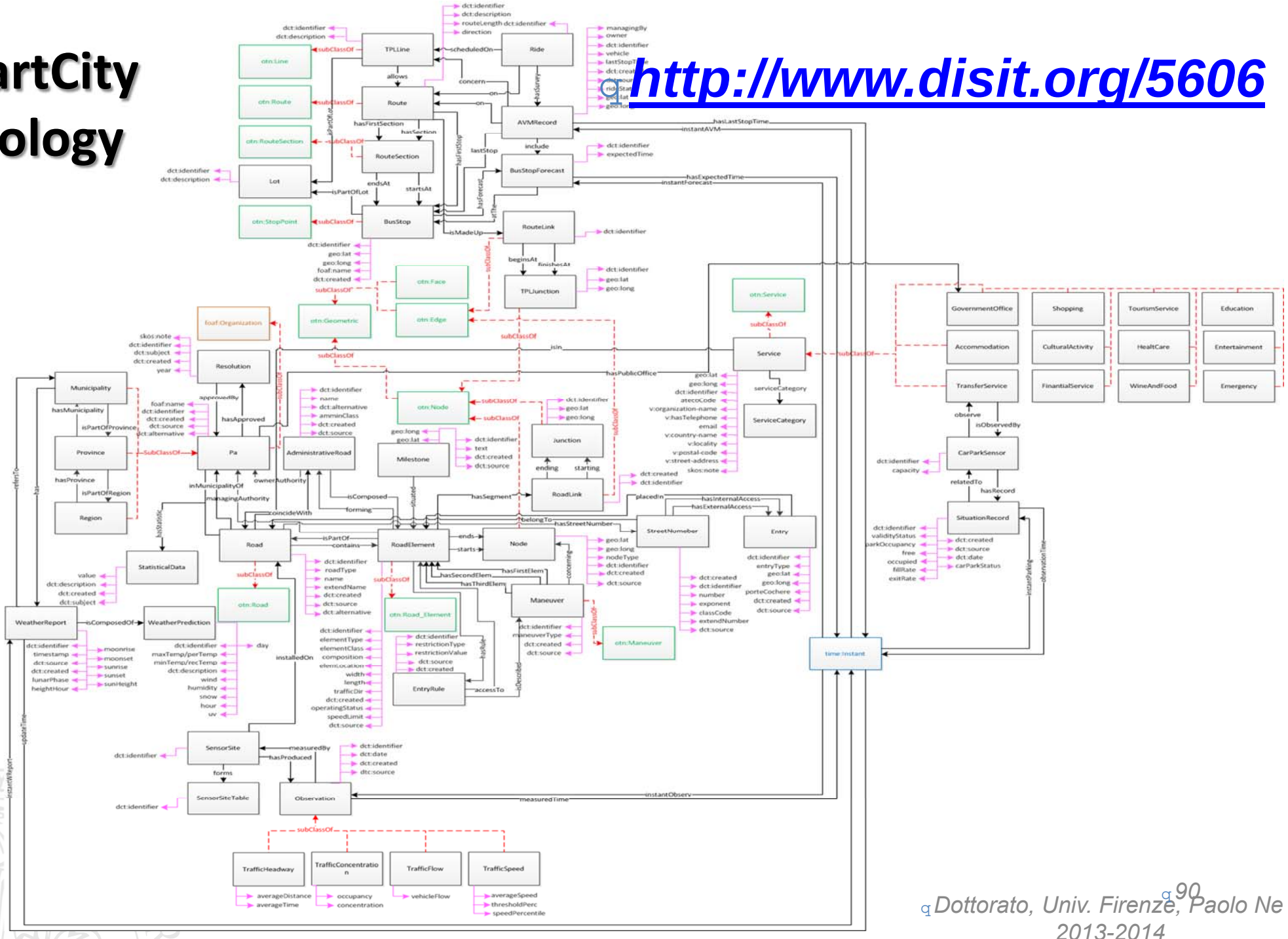
- see them it via http://log.disit.org
- see http://www.disit.org

# SmartCityOntology http://www.disit.org/5606

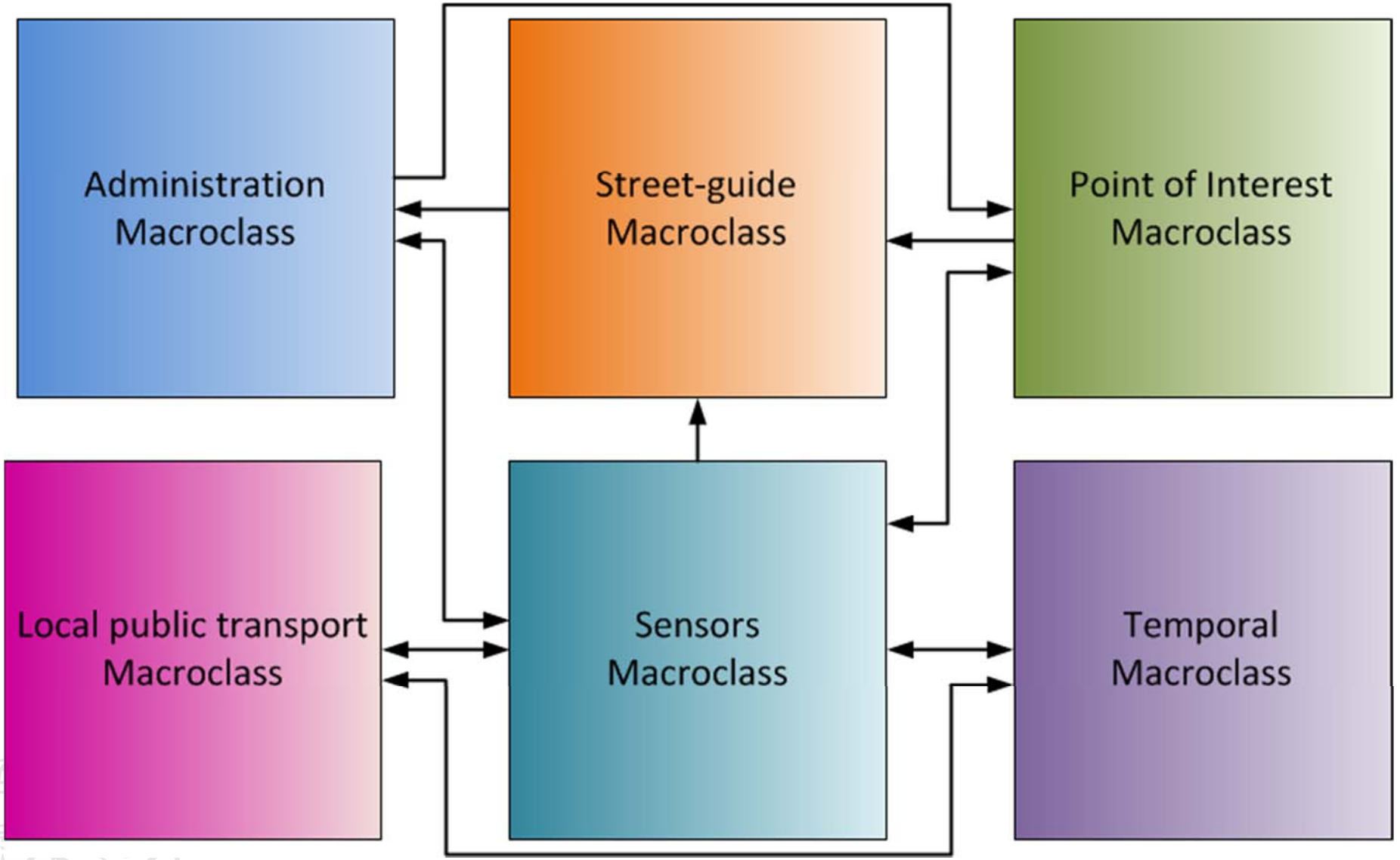
- **Title:** *SmartCity Ontology for Service Inference*
- **Duration**: *12 months*
- **Objectives**: *create an ontology that allows to combine all the data provided by the city of Florence and the Tuscan region:*
  - *509 OpenData (Municipality of Florence)*
  - *119 OpenData (Tuscany Region)*
  - *Timetable TPL*
  - *Street Graph*
  - *Punti di interesse*
  - *Real Time Data from traffic sensors*
  - *Real Time Data from parking sensors*
  - *Real Time Data from AVM systems*
  - *Weather Forrecast (consortium Lamma)*
- **Link**: http://log.disit.org, http://www.disit.org/5606

# SmartCity Ontology

http://www.disit.org/5606

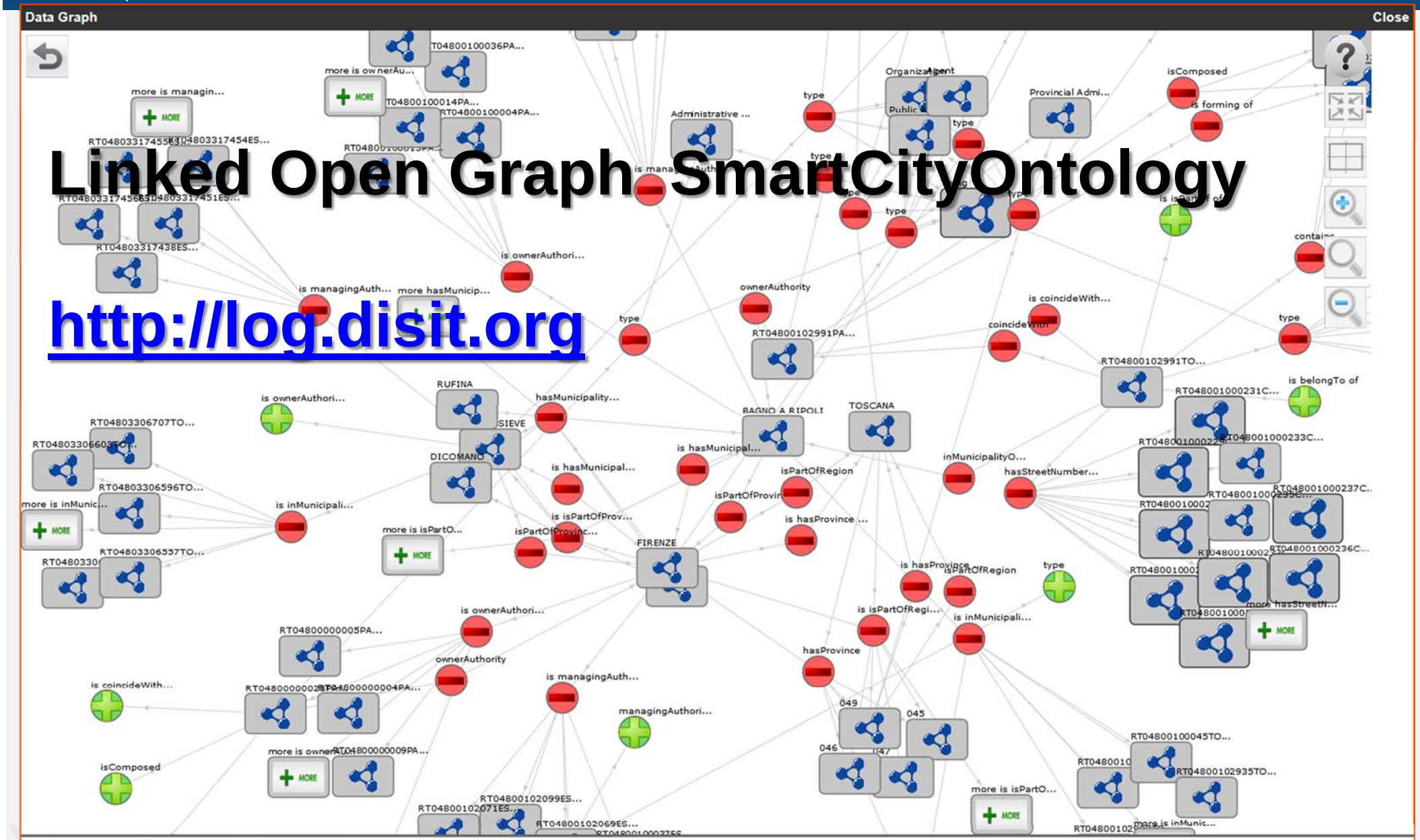


# SmartCityOntology … major macroclasses



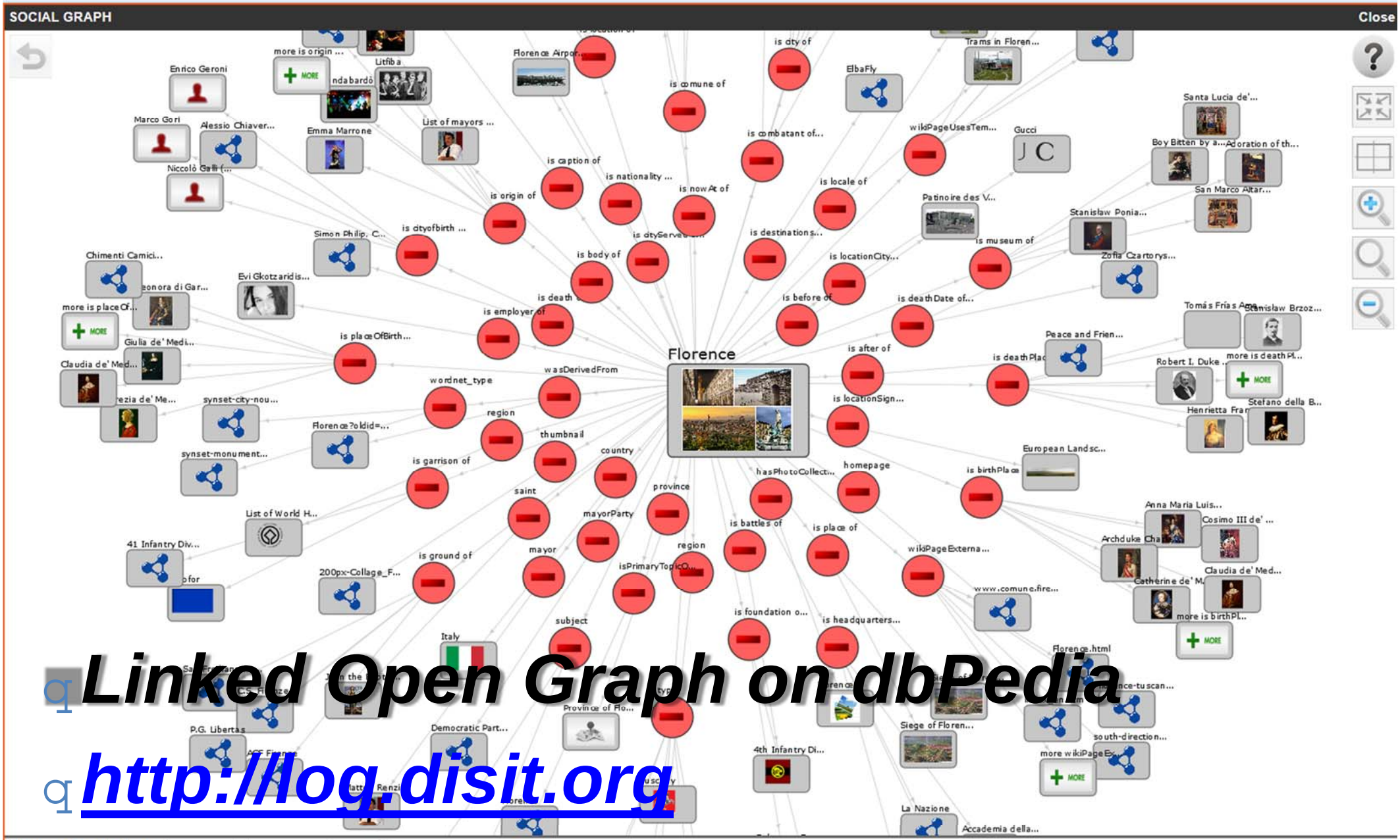
# Linked Open Graph

- ***Title**: Linked Open Graph: Social and Linked Open Data Navigation*
- ***Duration**: 24 months, derived from EC project ECLAP*
- ***Objectives**:*
  - *Design and develop tools for graphical navigation on Open Data and Linked Open Data*
- ***Link**:*
  - *LOG LOD [http://log.disit.org](http://log.disit.org)*
  - *Also used in www.ECLAP.eu Social Graph: [http://www.eclap.eu/116088](http://www.eclap.eu/116088)*
- ***Examples for: dbPedia, Europeana, British Museum, LinkedGeo Data, Cultura Italia, Sii-Mobility, ICARO Cloud, MyStoryPlayer, OSIM Knowledge Modeling and reasoning***


Data Graph
Close
Linked Open Graph SmartCityOntology
http://log.disit.org
Type of relations
Select all
Deselect all
Invert
sameAs
depiction
seeAlso
type
contains
coincideWith
inMunicipalityOf
hasStreetNumber
is isPartOf of
is belongTo of
ownerAuthority
isComposed
is forming of
isPartOfProvince
is managingAuthority of
is hasMunicipality of
isPartOfRegion
is hasProvince of
ends
starts
is placedIn of

# Piattaforma di sviluppo (beta)


http://servicemap.disit.org




# Linked Open Graph on dbPedia

## http://log.disit.org

# Linked Open Graph for LOD

A browsing tool to explore LOD sparql services via their entry point. To explore RDF elements and view which contents or users are linked with that.
With just one click (or tap) over a node you can see appear the navigation panel that allows you to:

- Explore/Reduce a node of the graph.
- Focus the visualization over a node.
- Open a specified content and view it's info.
- Direct accessing to the info associated with an entity, attributes and their values.
- Filtering relationships, inverting the filtering.
- Save your linked open graph with your preferences and navigations and get their access via email, that you can share with your colleagues for reading and further browsing and change.

A list of check buttons, one for each relation kind, to turn on/off the visualization of relations from the LOG.

LOG tool is free of use for no profit organizations. You can embed the LOG tool in your web pages.

# Sii-Mobility (Smart City nazionale)

- *Titolo:* **Supporto di Interoperabilità Integrato per i Servizi al Cittadino e alla Pubblica Amministrazione**
- **Ambito: Trasporti e Mobilità Terrestre**
- **Obiettivi**:
  1. ridurre i costi sociali della mobilità
  2. semplificare l'uso dei sistemi di mobilità
  3. Sviluppo di soluzioni e applicazioni funzionanti e sperimentazione
  4. Contribuire al miglioramento degli standard nazionali ed internazionali


- **Coordinatore Scientifico:** ***Paolo Nesi, DISIT DINFO UNIFI***
- **Partner:** ECM; Swarco Mizar; University of Florence (svariati gruppi+CNR); Inventi In20; Geoin; QuestIT; Softec; T.I.M.E.; LiberoLogico; MIDRA; ATAF; Tiemme; CTT Nord; BUSITALIA; A.T.A.M.; Sistemi Software Integrati; CHP; Effective Knowledge; eWings; Argos Engineering; Elfi; Calamai & Agresti; KKT; Project; Negentis.
- **Durata**: 36 months; **Costo**: circa 14 Meuro
- **Link**: http://www.disit.dinfo.unifi.it/siimobility.html

DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it



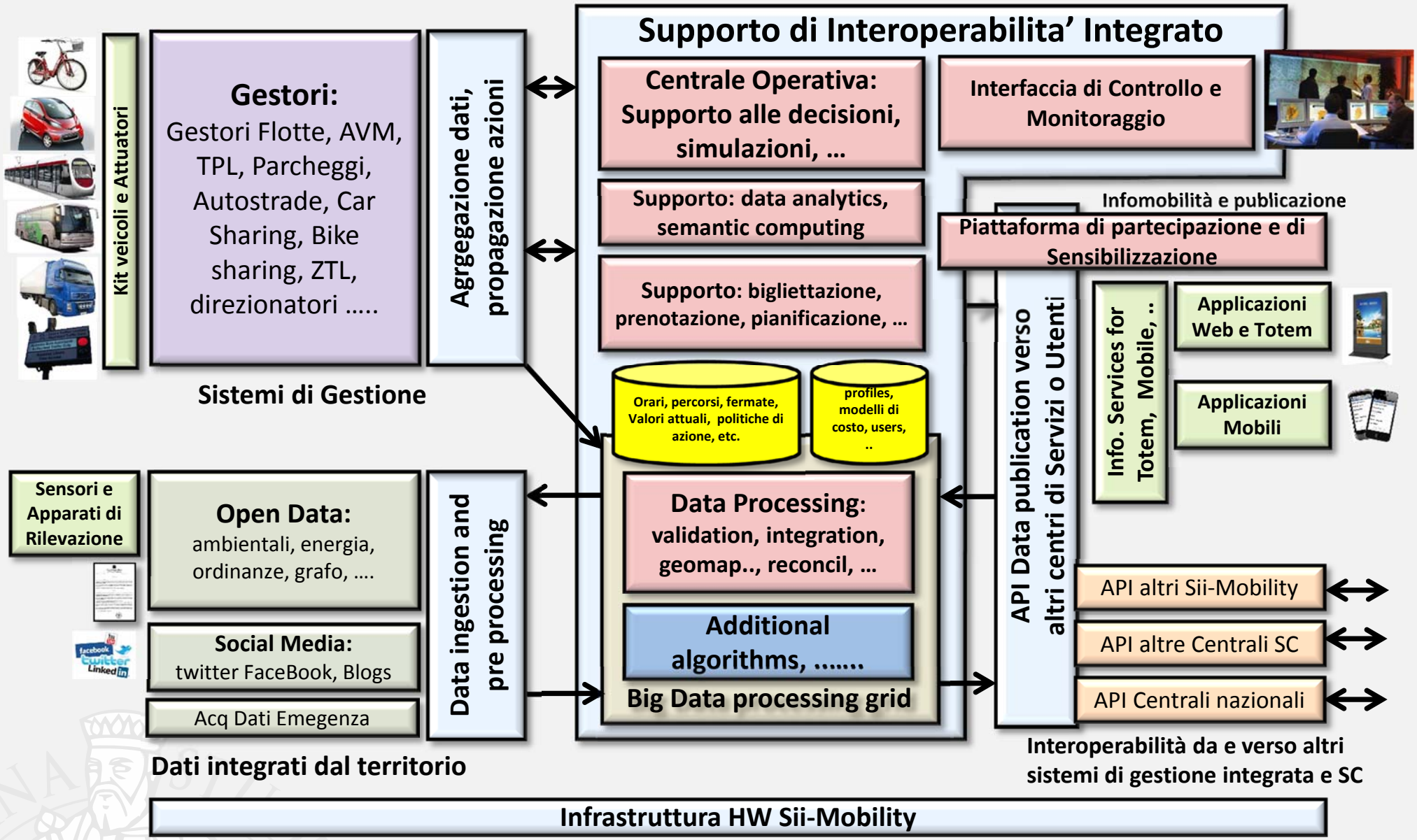
- Sperimentazioni principalmente in Toscana
- Sperimentazioni piu' complete in aree primarie ad alta integrazione dati
- Integrazione con i sistemi presenti

Regione Toscana

# Sii-Mobility: Scenari principali

- **soluzioni di guida/percorso connessa/o**
  - servizi personalizzati, segnalazioni, il veicolo/la persona riceve comandi e informazioni in tempo reale ma modo personalizzato e contestualizzato;
- **Piattaforma di partecipazione e sensibilizzazione**
  - per ricevere dal cittadino informazioni, il cittadino come sensore intelligente, informare e formare il cittadino, tramite totem, applicazioni mobili, web applications, etc.;
- **gestione personalizzata delle politiche di accesso**
  - Politiche di incentivazione e di dissuasione dell'uso del veicolo, Crediti di mobilità, monitoraggio flussi;
- **interoperabilità ed integrazione dei sistemi di gestione**
  - contribuzione a standard, verifiche e validazione dei dati, riconciliazione dei dati, etc.;
- **integrazione di metodi di pagamento e di identificazione**
  - Politiche pay-per-use, monitoraggio comportamento degli utenti;
- **gestione dinamica dei confini delle aree a traffico controllato**
  - tariffazione dinamica e per categoria di veicoli;
- **gestione rete condivisa di scambio dati fra servizi (PA e privati)**
  - affidabilità dei dati e separazione delle responsabilità, Integrazione di open data, riconciliazione, ….;
- **monitoraggio della domanda e dell'offerta di trasporto pubblico in tempo reale**
  - soluzioni per l'integrazione e l'elaborazione dei dati.

# Architettura Sii-Mobility



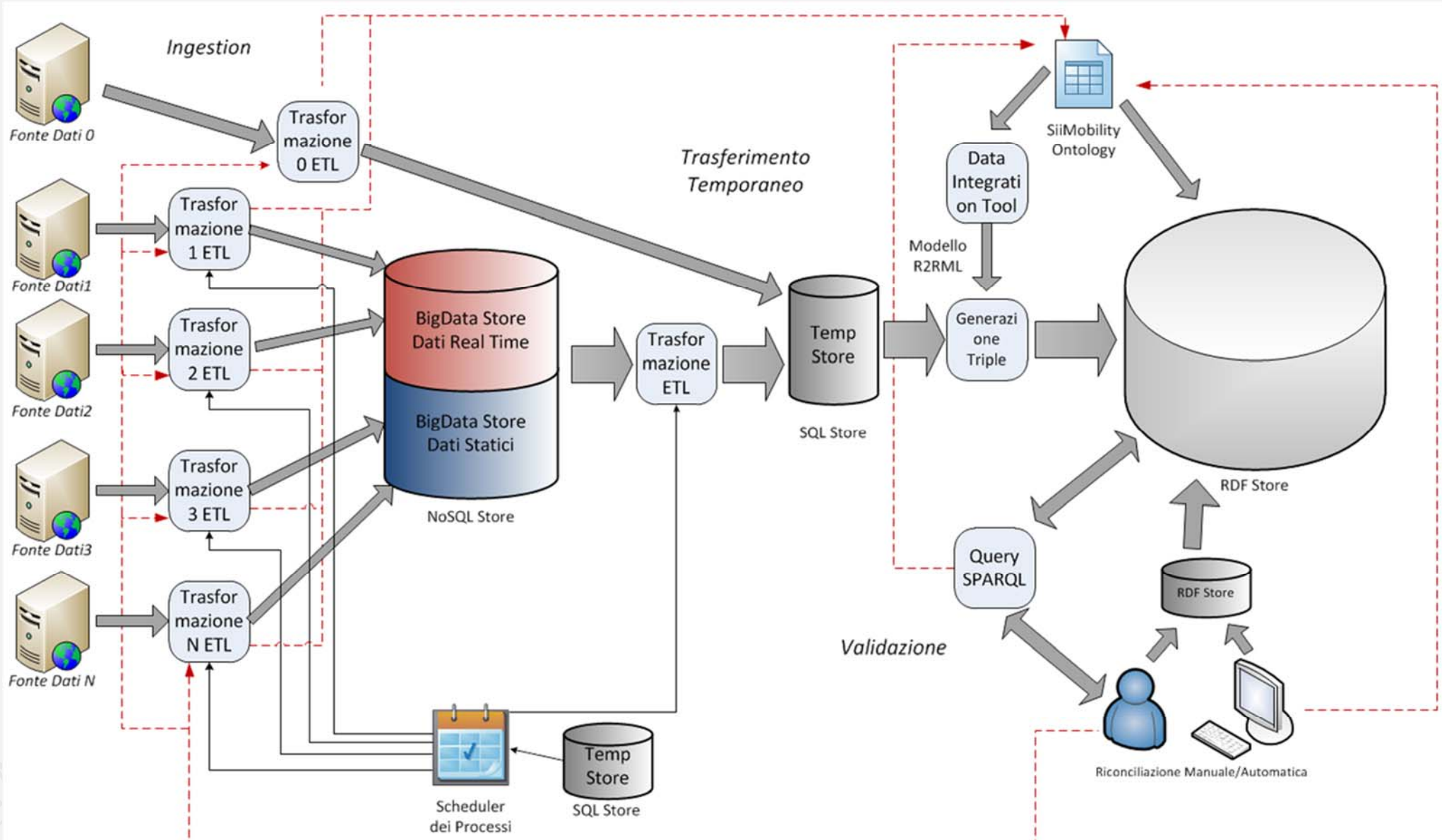
# Problematiche di Ricerca "Smart"

- ***Modellazione della conoscenza con semantica coerente per effettuare deduzioni corrette sfruttando informazioni numeriche e simboliche, su grandi volumi e flussi di dati, automazione***
  - ***Problematiche*** derivate da oltre **1000 OD e PD**, servizi con modelli diversi e formati diversi: resilienza, qualità, misura, accesso, integrazione real time, …
  - ***Tecniche di:*** modellazione, semantic computing, scheduling, …
- **Ricerca su integrazione e modellazione dati**:
  - ***Alto livello***: predizione su servizi e comportamenti, correlazioni inattese, situazioni critiche, flusso dei viaggiatori, …
  - ***Problemi e algoritmi*** di riconciliazione del dato, tracciamento e versionamento, reputazione, filtraggio, integrazione OD e LOD internazionali/nazionali, validazione e verifica formale, …. …. ….

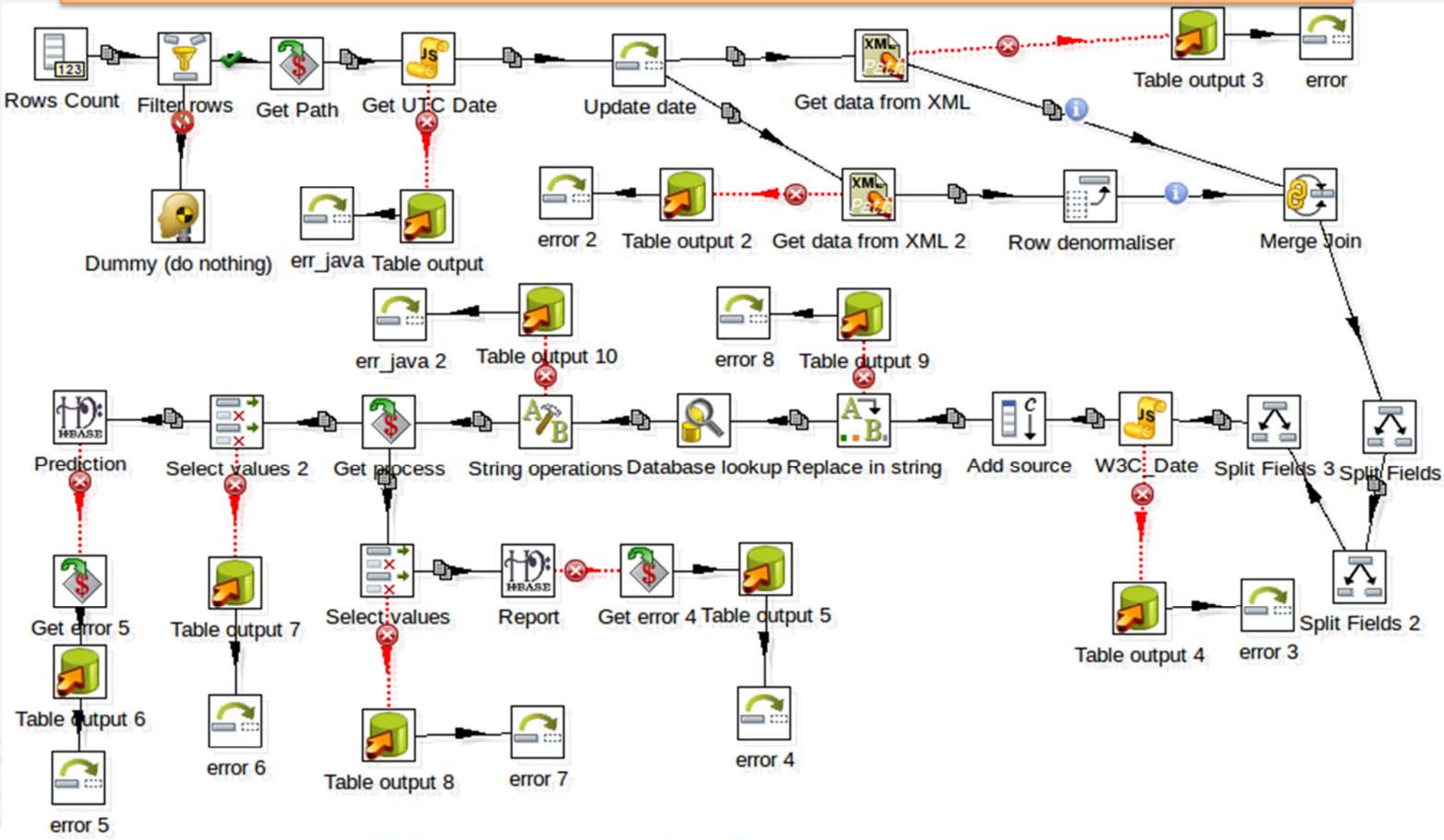
# Riconciliazione servizi

| Service_Key | Road_Key | R_address | extendName | Cod_ toponimo |
|---|---|---|---|---|
| BELLAVISTA-Largo_F.lli_Alinari_15 | 048017LargoF.lliAlinari | Largo F.lli Alinari | LARGO FRATELLI ALINARI | RT04801701866TO |
| CASA_DEL_LAGO-Lungarno__A._Vespucci_58 | 048017LungarnoA.Vespucci | Lungarno A. Vespucci | LUNGARNO AMERIGO VESPUCCI | RT04801701874TO |
| COSMOPOLITAN-Via_F._Baracca187 | 048017ViaF.Baracca | Via F. Baracca | VIA FRANCESCO BARACCA | RT04801702987TO |
| SAN_PAOLO_IMI-VIA_DE'_VECCHIETTI_22/R | 048017VIADE'VECCHIETTI | VIA DE' VECCHIETTI | VIA DEI VECCHIETTI | RT04801702383TO |
| CREDITO_ARTIGIANO-VIA_DE'_BONI_1 | 048017VIADE'BONI | VIA DE‘ BONI | VIA DEI BONI | RT04801702326TO |
| Auditorium_al_Duomo-Via_De'_Cerretani_54/r | 048017ViaDe'Cerretani | Via De' Cerretani | VIA DEI CERRETANI | RT04801702317TO |

# Architettura del sistema



# Ingestion (un esempio)



# Validazione e Verifica

- Query in linguaggio SPARQL
- Query di verifica del collegamento tra gli oggetti creati e quelli del grafo stradale
  - Si seleziona un elemento del grafo stradale e si richiedono tutti gli elementi ad esso collegati
  - Nel caso di grandi quantità di dati risulta necessario contare il numero di elementi collegati

| Oggetto | Totale | Riconciliabili (7 comuni di Firenze) | Riconciliati |
|---|---|---|---|
| Previsioni meteo | 286 | 7 | 7 |
| Statistiche del comune | 115 | 115 | 115 |
| Uffici Pubblici | 752 | 176 | 176 |
| Servizi | 28560 | 3559 | 3502 |
| Statistiche sulle vie di Firenze | 7987 | 7987 | 7987 |

# Coll@bora

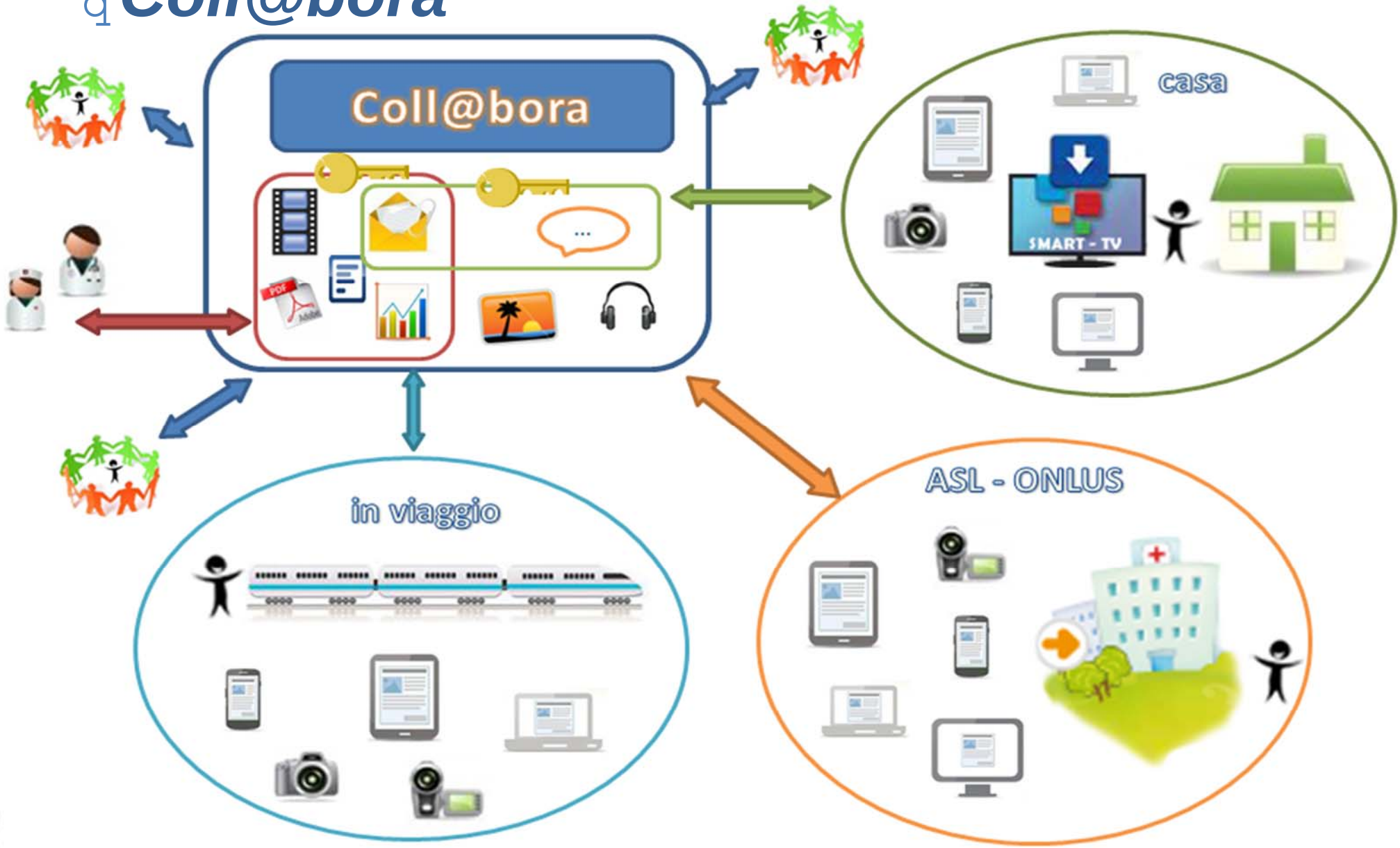
- ***Title***: *Collaborative Support for Parents and Operators of Disabled*
- ***Duration***: *24 months*
- ***Cost***: *1 Meuro*
- ***Objectives***: *providing strong advantages for*
  1. *Relatives interested in facilitating relations with the management team;*
  2. *Associations in order to offer a better service to the families and people with disabilities by providing a collaborative support to the involved teams, but also to manage the wealth of knowledge, to support the training of the staff, etc.*

- *Coll@bora provides a secure collaboration tool for the teams and for the association to support the families and the disabled people.*
- ***Link***: *http://www.disit.dinfo.unifi.it/collabora.html*

# Coll@bora

- **Objective 1:** *Study and development of a platform for collaboration and management for team (consisting of parents, family members, physicians, physician assistants, volunteers, etc..) for disable support in privacy*
- **Objective 2:** *Study and development of web applications, and mobile smartTv to support the activities of assistance and service*
- **Objective 3:** *validation of the solution*

# Coll@bora



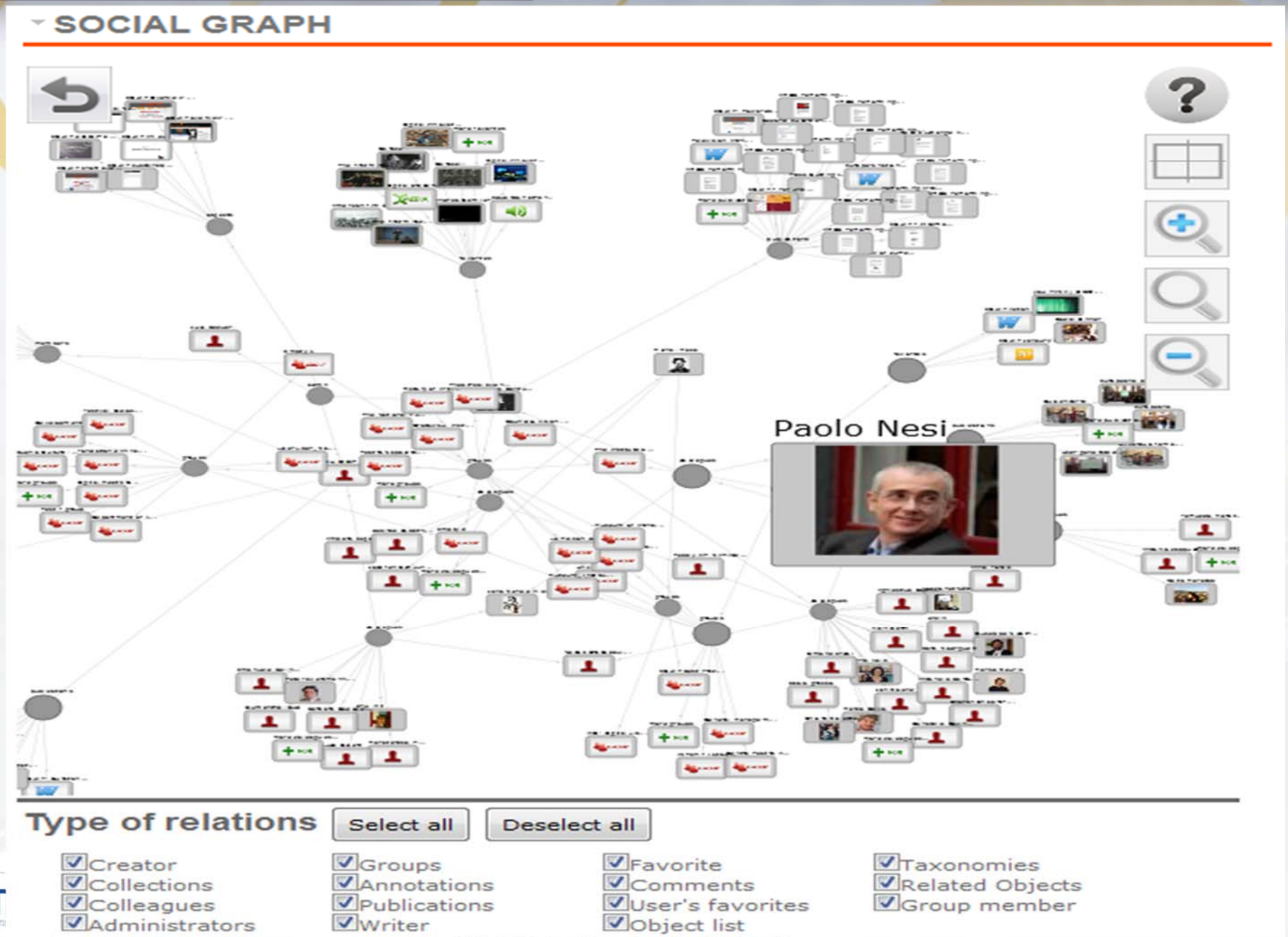
# SACVAR/OSIM: Open Mind Innovative Space

- **Title**: Open Mind Innovative Space, and SACVAR
- **Duration**: 36 months, 12 as OSIM e 24 as SACVAR
- **Objectives**:
  1. Knowledge mining construction from people competence
  2. Semantic search engine on mined knowledge
  3. Web pages and Blog mining and analysis in Natural Language
  4. Match making and affective computing analysis

– We are using these tools for social media analysis, to analyze citizens appreciation and comments on Smart City services.

- **Link**: http://openmind.disit.org
  - http://www.disit.dinfo.unifi.it/osim.html
  - http://www.disit.dinfo.unifi.it/sacvar.html

# Social Graph



## LOCATION MAP


## Federal Republic of Germany

(1-10 of 26 in 4163 ms)

### Tutta casa, letto e chiesa -1977, Monologhi satirici sull...

Manifesto dello spettacolo "Tutta casa, letto e chiesa" Nur Kinder, Küche, Kirche

1 Hits Rating ★★★★★ Actions

### Disegni, bozzetti e dipinti non collegati a opere particol...

Dario Fo: serie di disegni raffiguranti le sorelle Monteros, note ballerine di flamenco. Le sorelle Monteros.

1 Hits Rating ★★★★★ Actions


### rejestracja akcji Miasta Równoległe/ Ciudades Paralela...

rejestacja akcji Miasta Równoległe/ Ciudades Paralelas/ Parallele Staedte. Edycja: Berlin. Koncepcja projektu: Lola Arias / Stefan Kaegi. Opis projektu: ...

1 Hits Rating ★★★★★ Actions

### Estero - Documenti vari e corrispondenza relativa a Pa...

Dario Fo e Franca Rame incontrano il pubblico del teatro tedesco Berliner Ensamble in una conversazione con il regista Peymann Claus.

1 Hits Rating ★★★★★ Actions

## CONTENT

- *Featured*
- *Popular*
- *Popular in the period*
- *Last Posted*
- *Top Rated*
- *Location*
- *Upload new content*

ORGANIZE PERSONAL COLLECTION
ROOT
KEYWORD CLOUD
QUERY CLOUD
CLASSIFICATION
GROUPS

## POTENTIAL COLLEAGUES

**boloska**
stefania bolognesi, Female
ITALY
Add to your colleagues Details

**Alessandra**
Alessandra MARFOGLIA, Female, 38
ITALY
Add to your colleagues Details

**carolinamadriz**
Ana Carolina Madriz Zuniga
COSTA RICA
Add to your colleagues Details

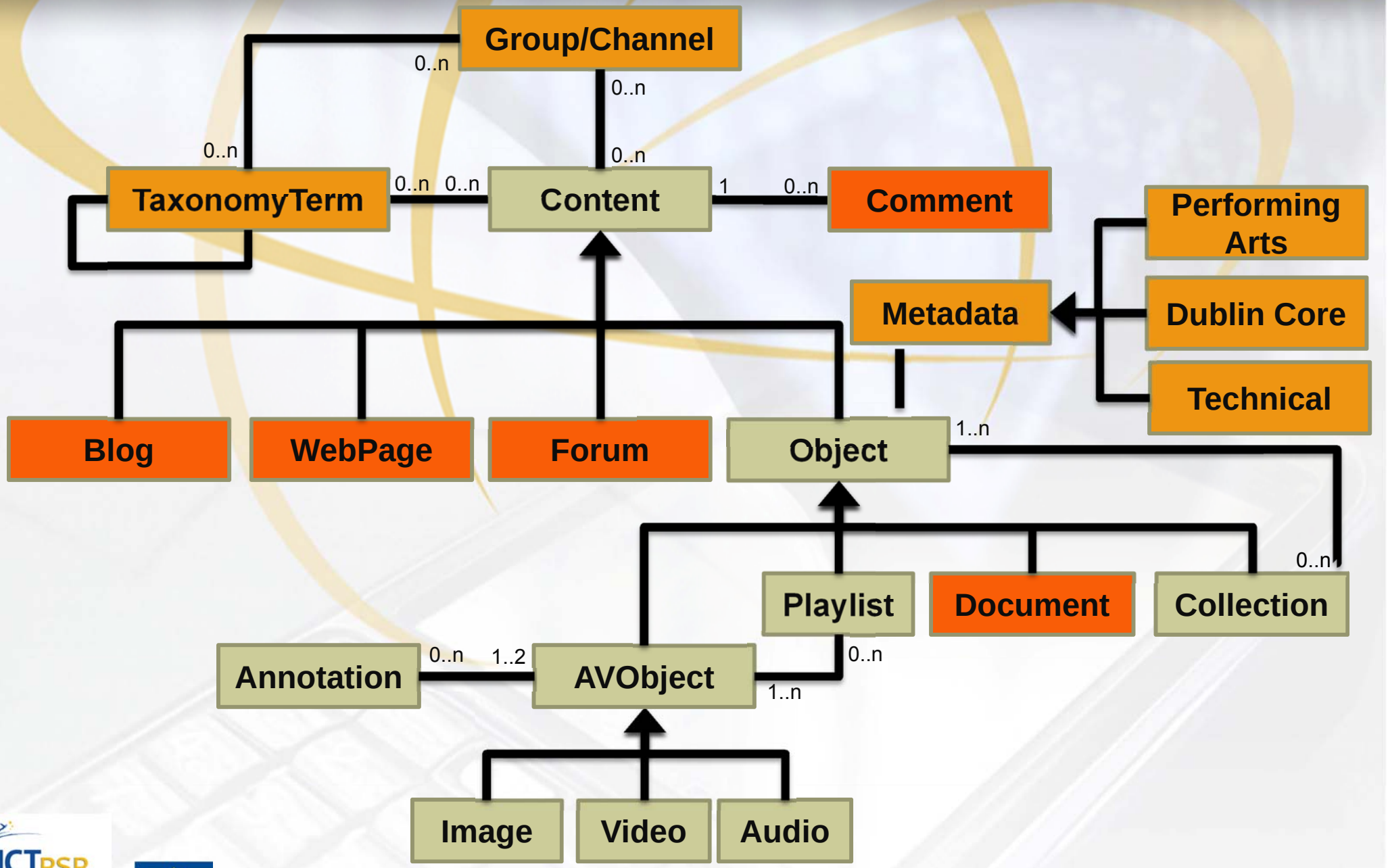
**©borja**
Carol Borja, 33
MEXICO, Michoacán
Add to your colleagues Details

**Valentina Valentini**
Valentina Valentini, Female
ITALY
Add to your colleagues Details

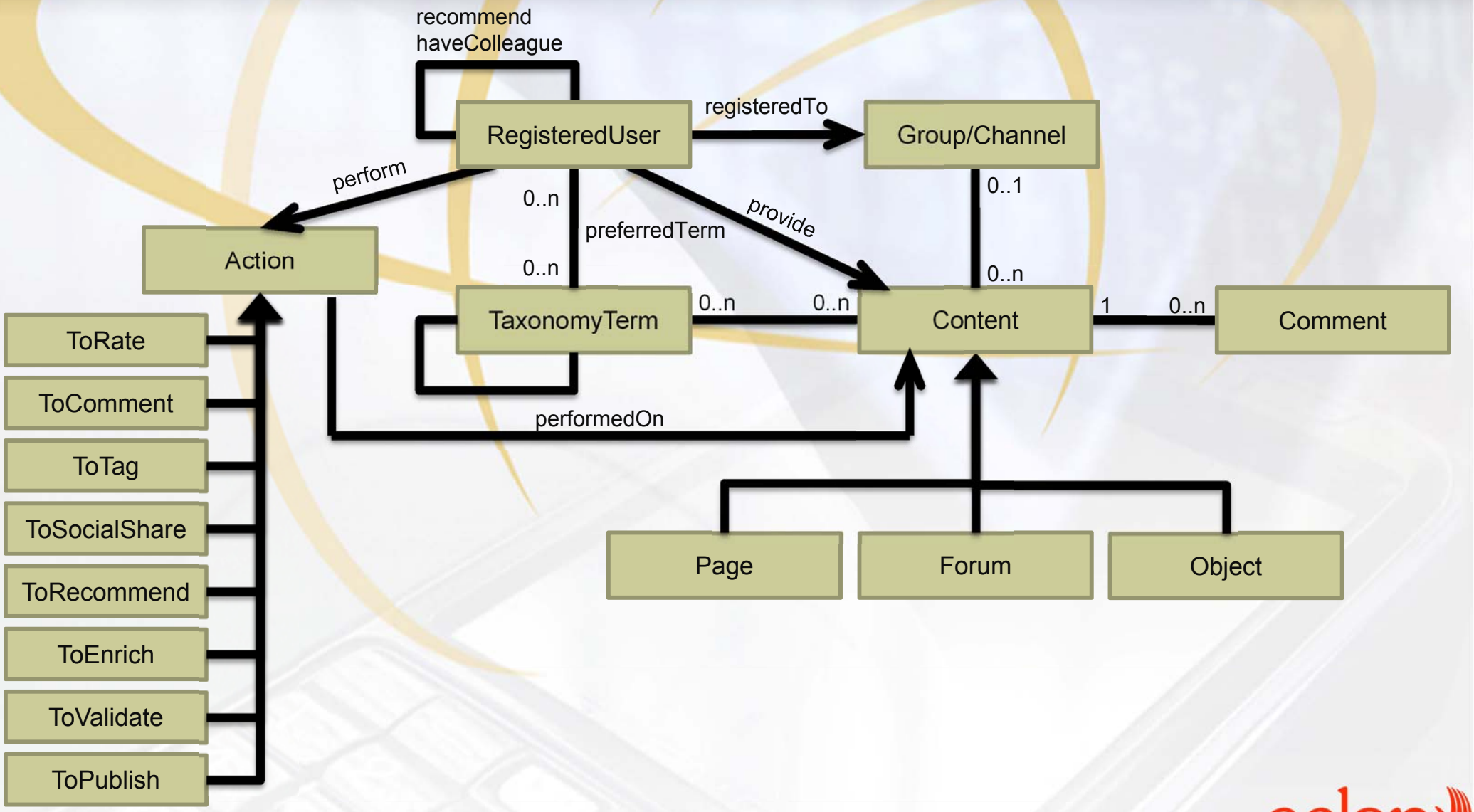
1 2 3 4 5 6 7 8 9 next > last »

STATISTICS ON YOUR ACTIVITIES
STATISTICS ON YOUR GROUPS
STATISTICS ON DOWNLOADS
STATISTICS ON QUERIES

# ECLAP semantic model

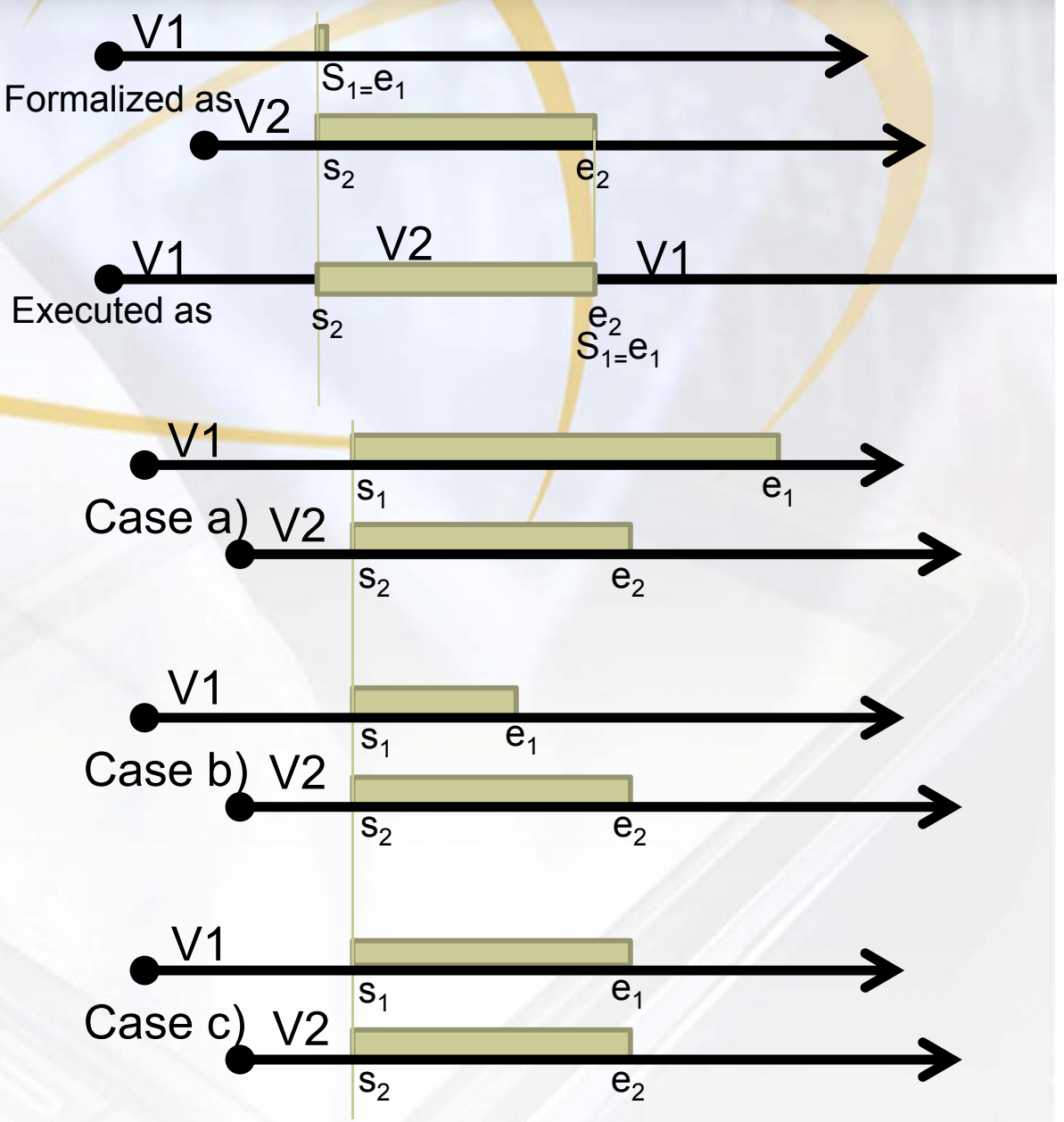


# BPN: User based semantic model



# Relazioni fra media

- Sincronizzazioni
- Sequenze
- Dirette
- Esplosive

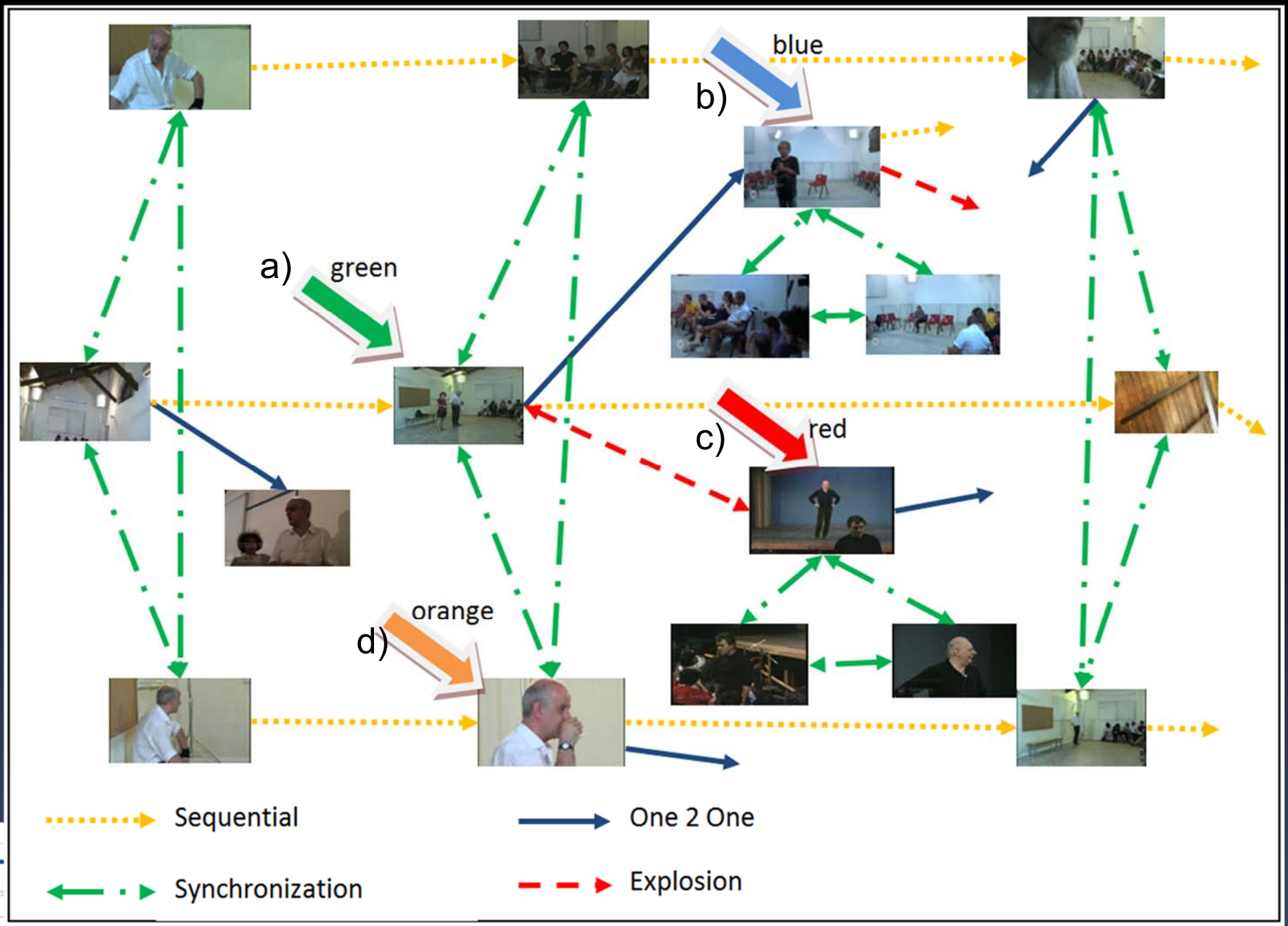


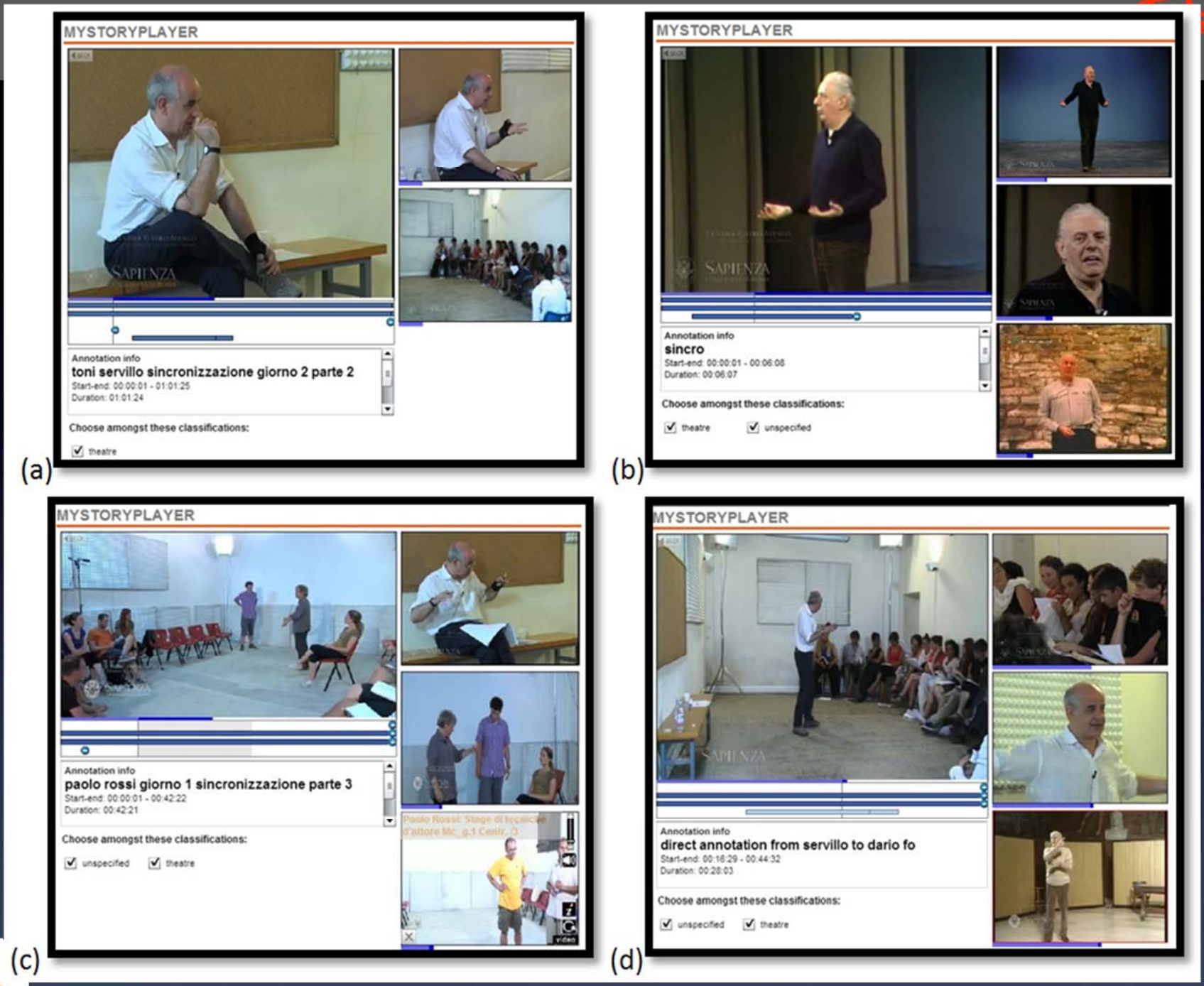
# MyStoryPlayer

# Networks of Annotations

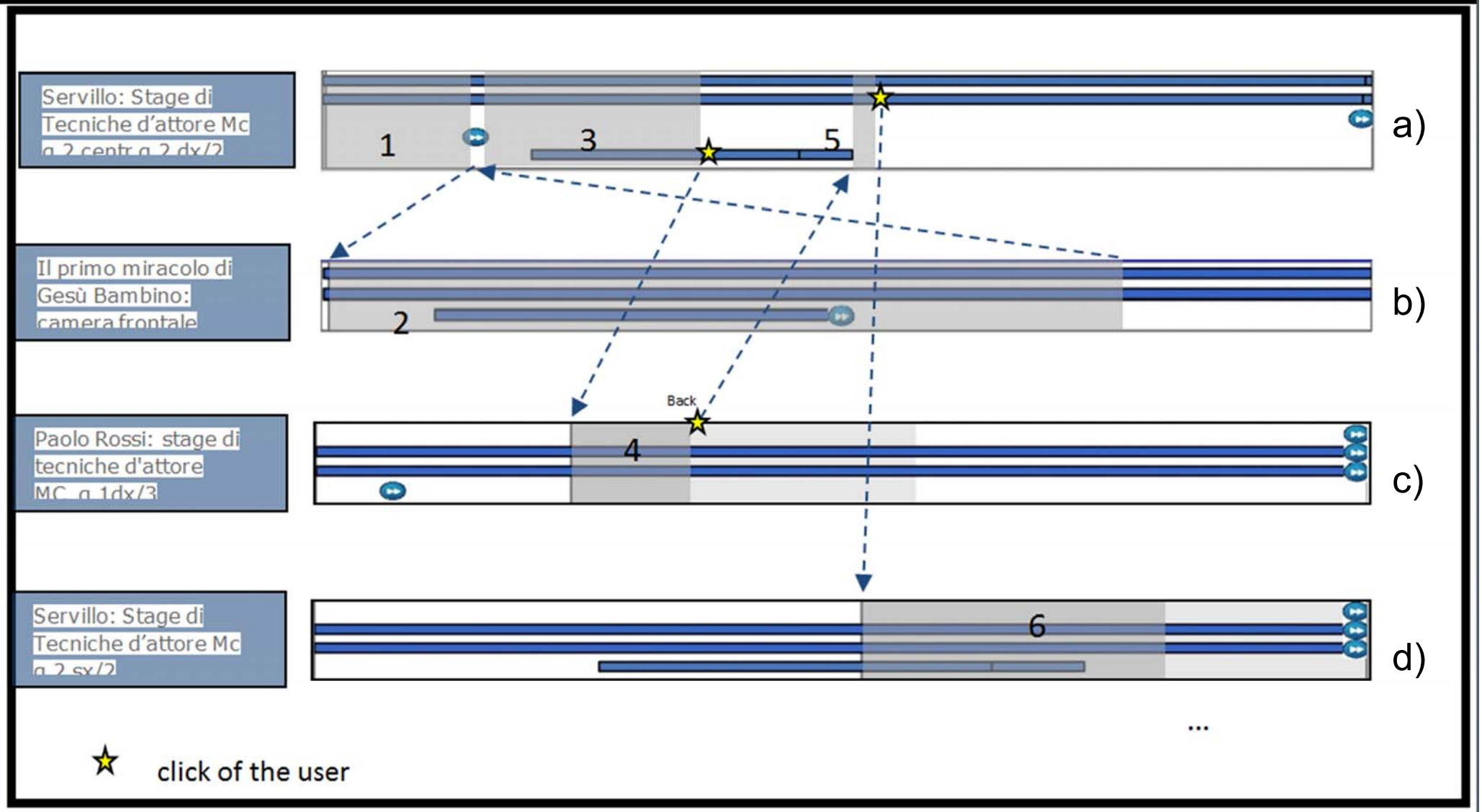
- http://www.eclap.eu/drupal/?q=en-US/msp&axoid=urn%3Aaxmedis%3A00000%3Aobj%3A04e0caef-b33b-4f4a-ba50-a80d96766192&axMd=1&axHd=1

# Modeling relationships

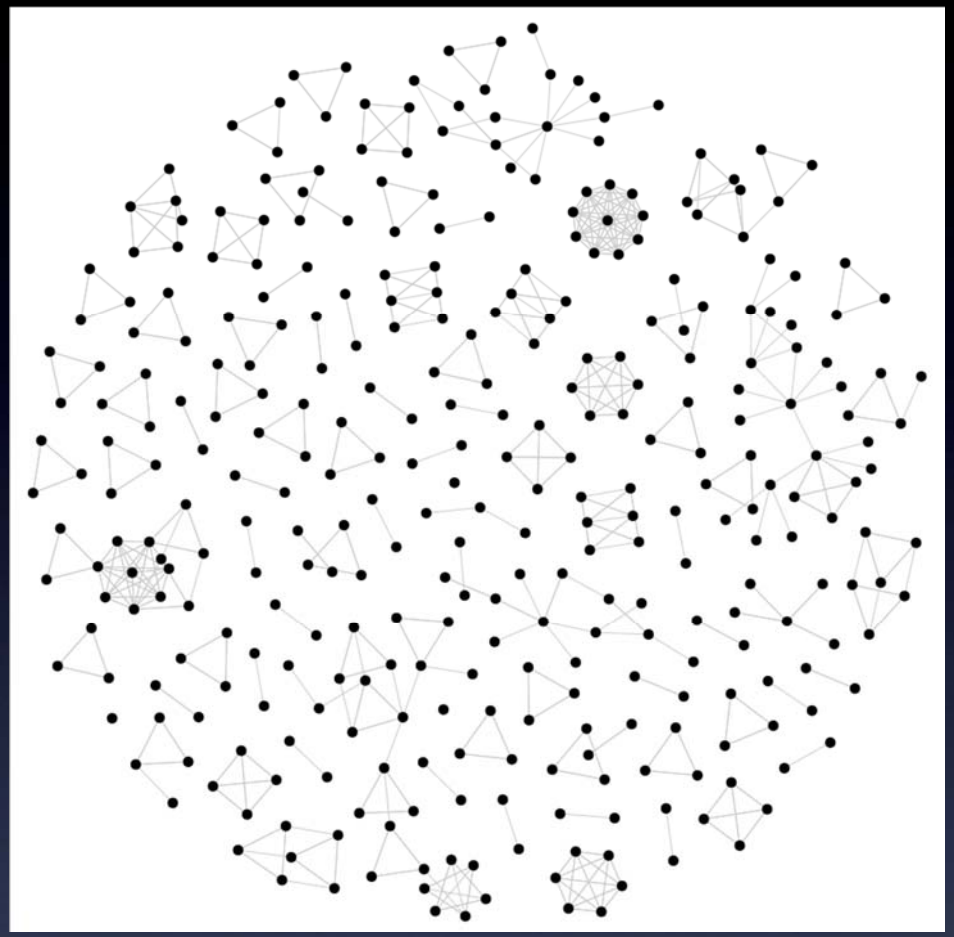


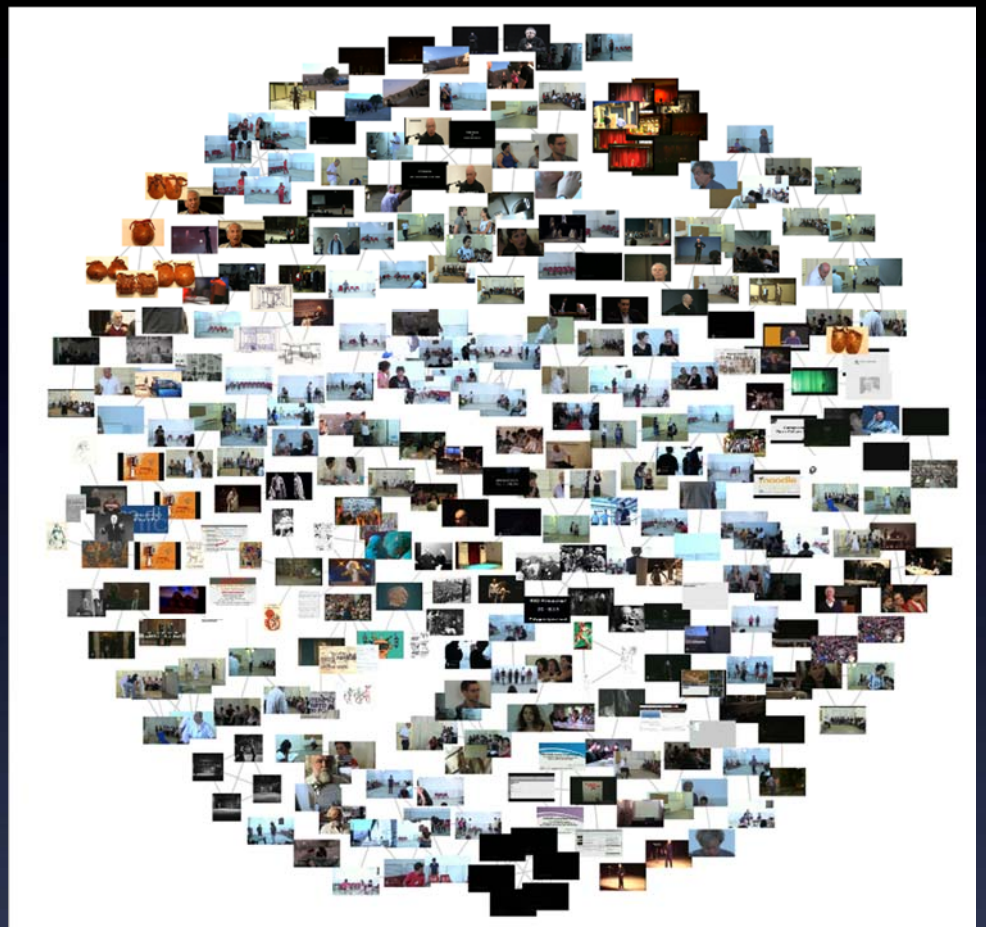

MYSTORYPLAYER
Annotation info
toni servillo sincronizzazione giorno 2 parte 2
Start-end: 00:00:01 - 01:01:25
Duration: 01:01:24
Choose amongst these classifications:
theatre
(a)
MYSTORYPLAYER
Annotation info
sincro
Start-end: 00:00:01 - 00:06:08
Duration: 00:06:07
Choose amongst these classifications:
theatre
unspecified
(b)
MYSTORYPLAYER
Annotation info
paolo rossi giorno 1 sincronizzazione parte 3
Start-end: 00:00:01 - 00:42:22
Duration: 00:42:21
Choose amongst these classifications:
unspecified
theatre
(c)
MYSTORYPLAYER
Annotation info
direct annotation from servillo to dario fo
Start-end: 00:16:29 - 00:44:32
Duration: 00:28:03
Choose amongst these classifications:
unspecified
theatre
(d)

# Navigating on media: a,b,c,d



# Network of Relationships

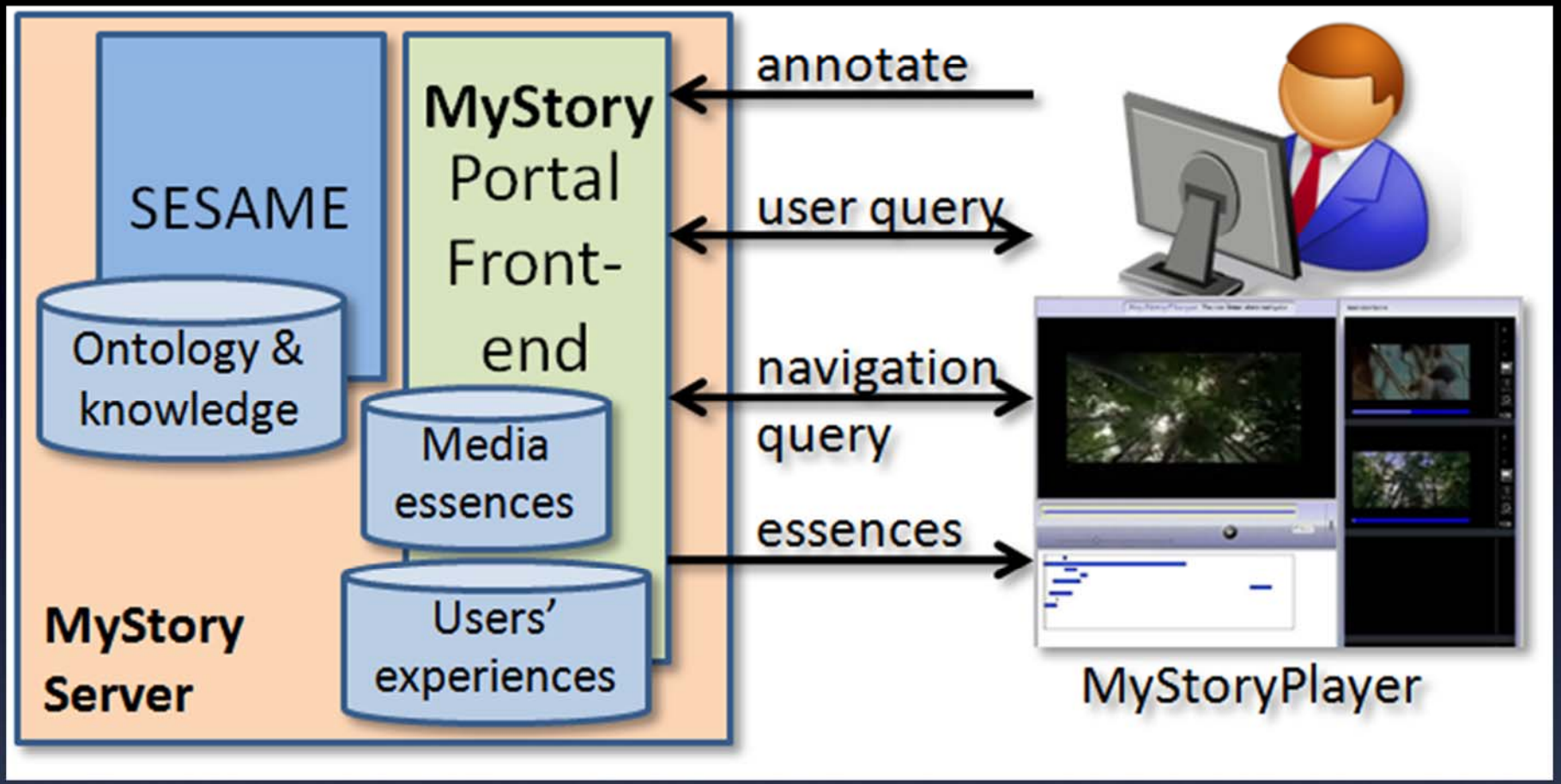




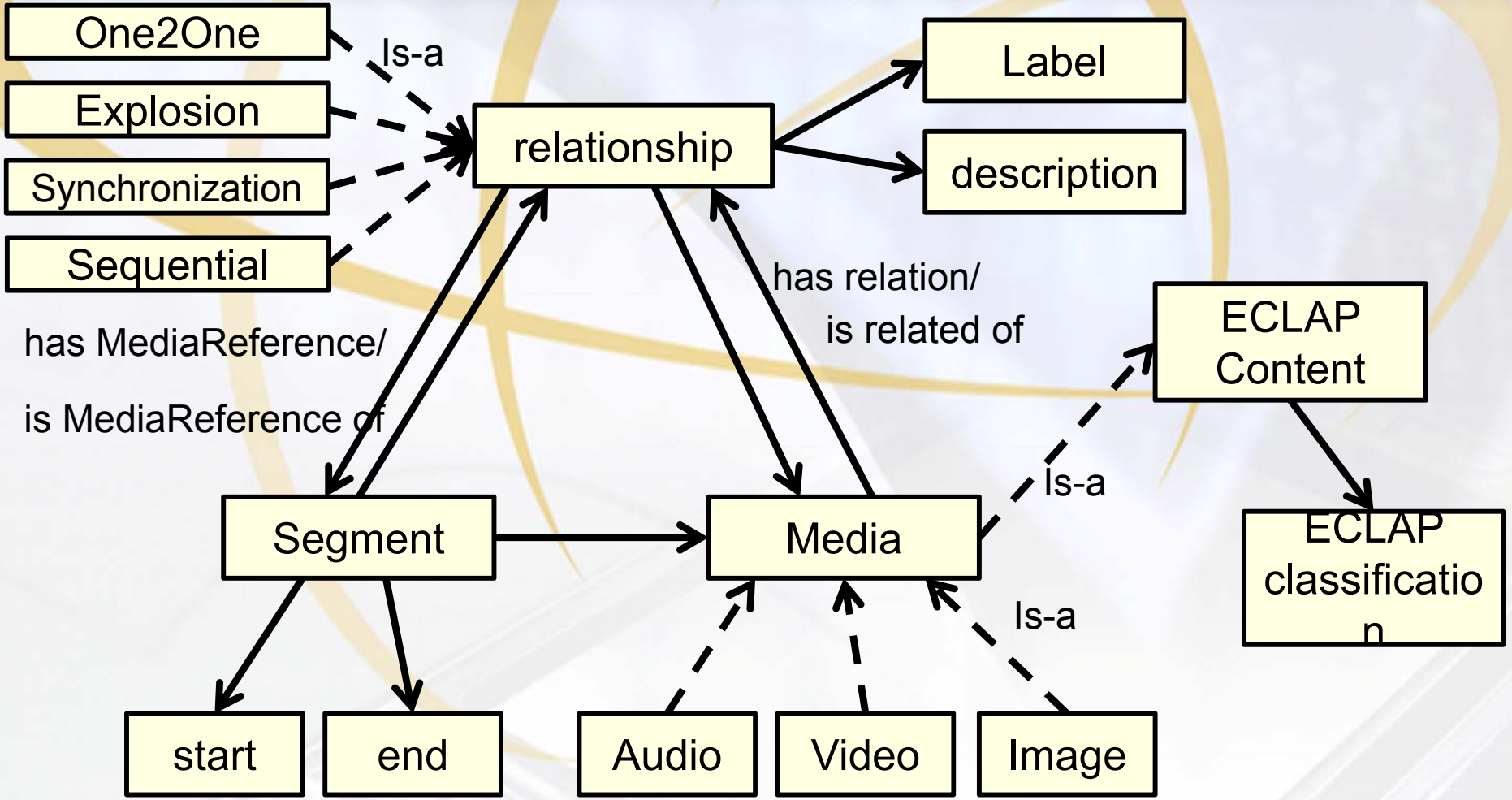
http://www.eclap.eu/d3/graph.html

http://www.eclap.eu/d3/graph2.html

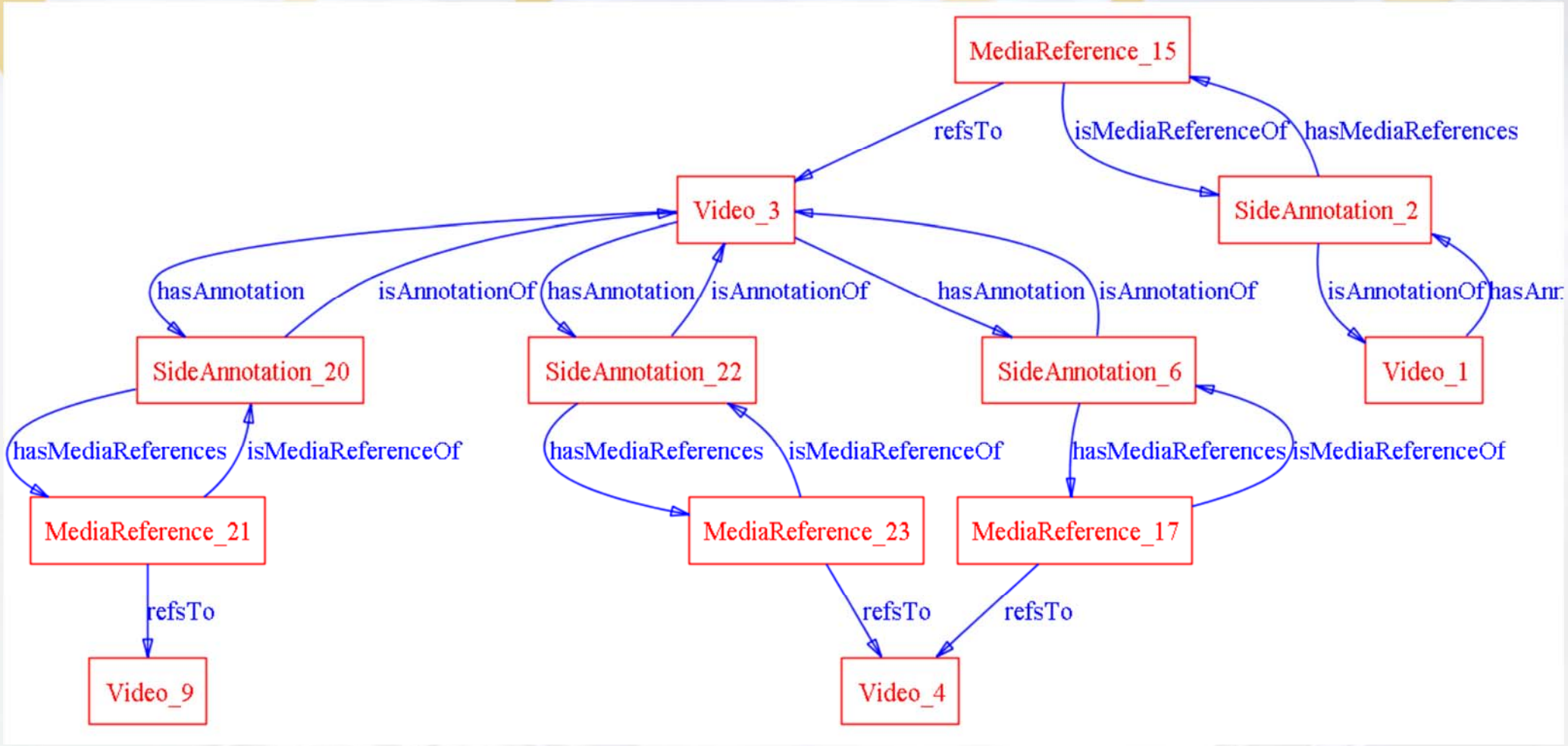
# MyStoryPlayer Architecture



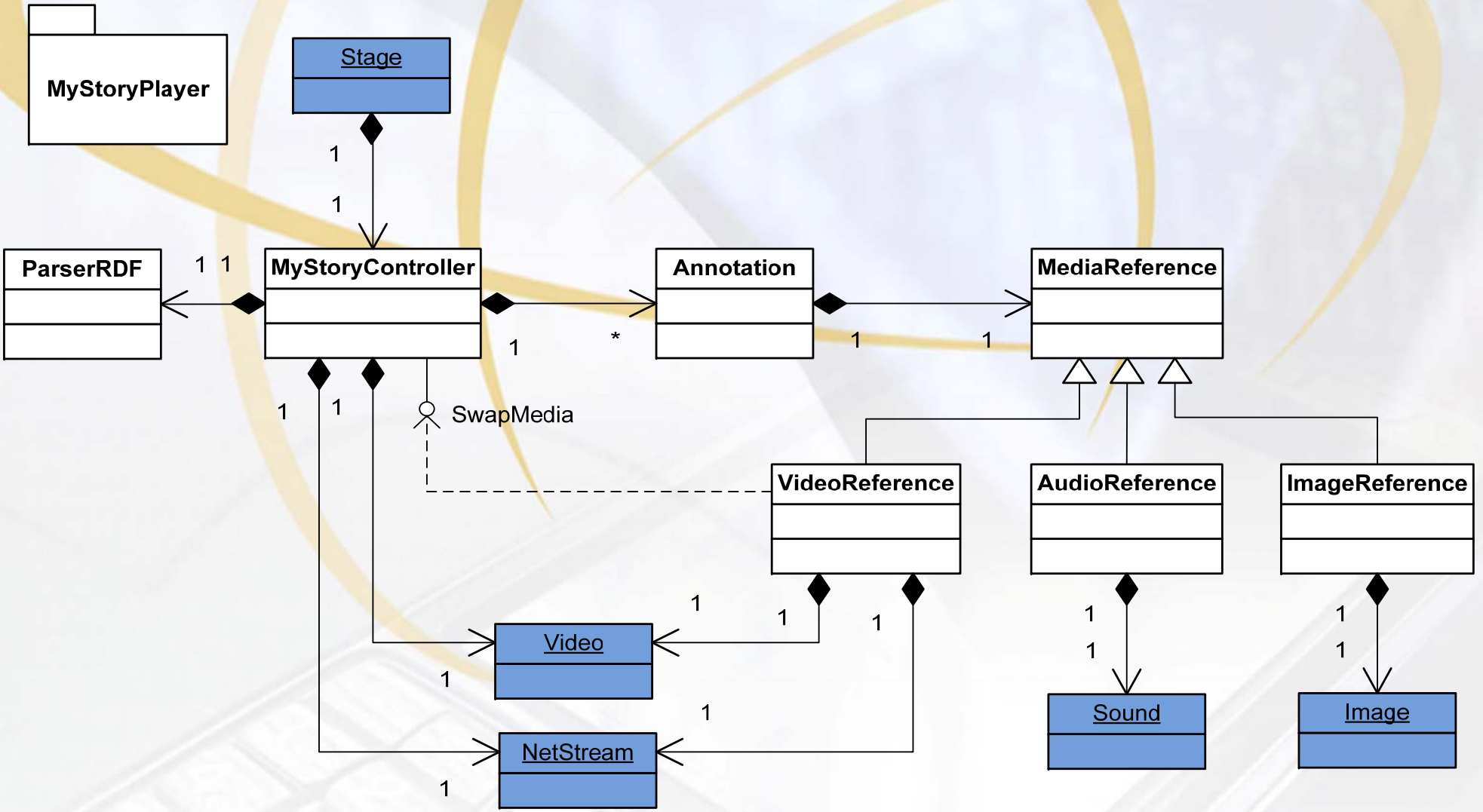
# Modello Semantico MSP



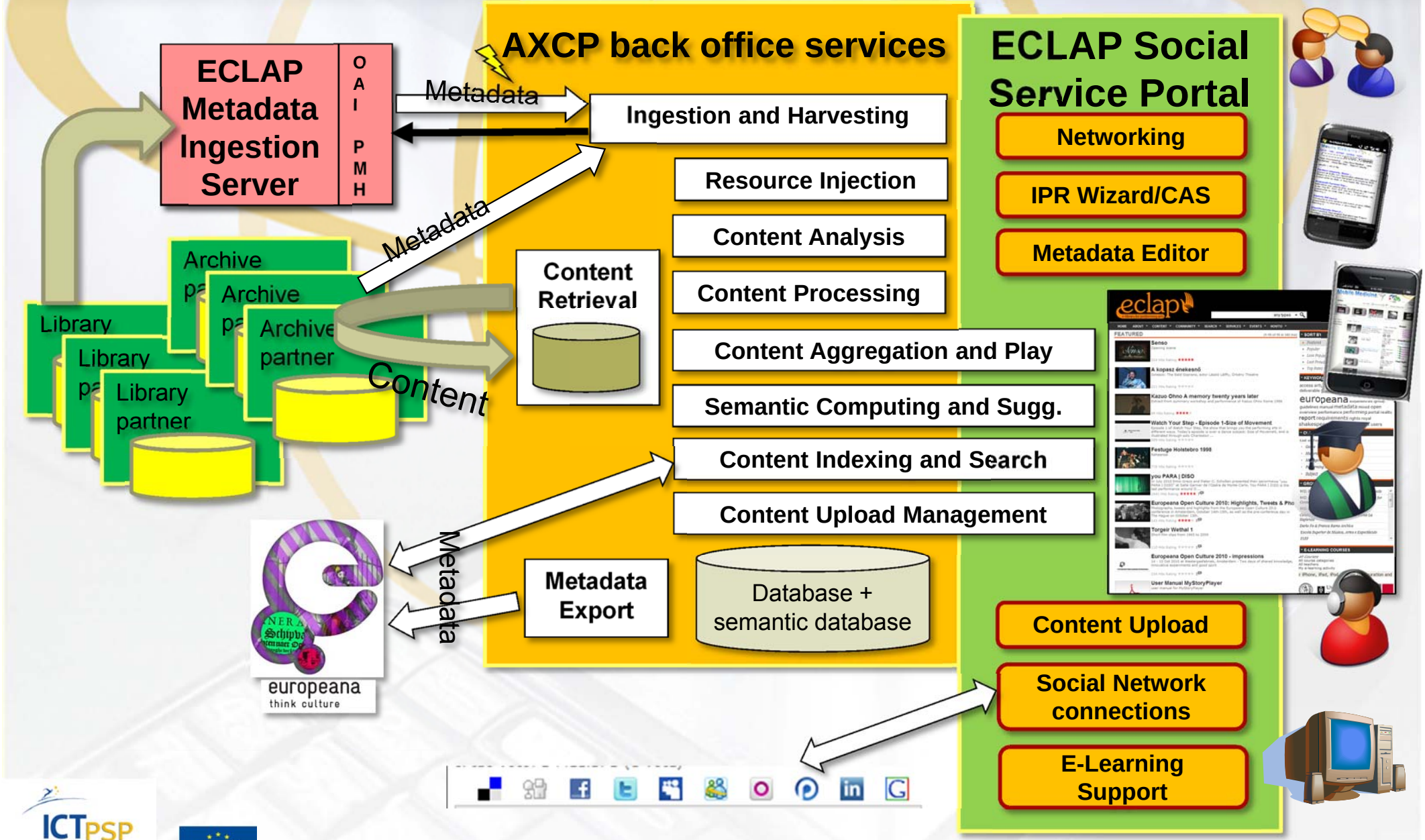
# An instance of an annotation



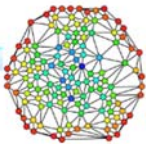
# Class diagram of MyStoryPlayer



# Architecture and Back-office





# Estrazione e Deduzione della Conoscenza via Modelli Semantici: From Social Network to Smart City

***seminario per il** Corso di Dottorato 2014*


***Prof. Paolo Nesi***

Dipartimento di Ingegneria dell'Informazione

University of Florence

Via S. Marta 3, 50139, Firenze, Italy

tel: +39-055-4796523, fax: +39-055-4796363

**DISIT Lab**

Paolo.nesi@unifi.it

**http://www.disit.dinfo.unifi.it/**