

Linked Data for the Masses: Η προσέγγιση και το λογισμικό

Γιώργος Αναδιώτης, Πάνος Ανδριόπουλος, Πάνος
Αλεξόπουλος, Δημήτρης Βεκρής, Αριστοτέλης
Ζωσάκης

IMC Technologies S.A.

15/05/2010

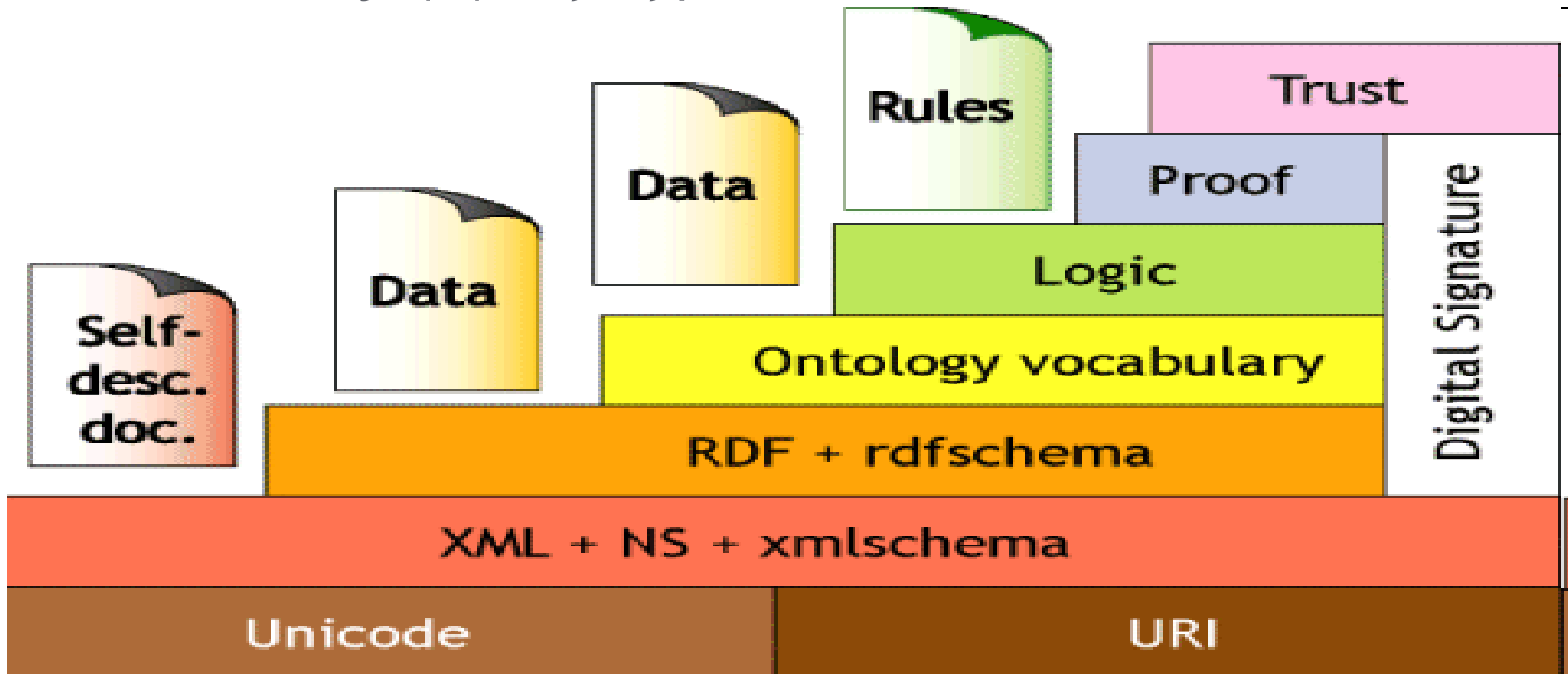
1. Εισαγωγή: Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό και στα Linked Data
2. Η Inbound/Outbound Linked Data Προσέγγιση
3. Υλοποίηση: Πρότυπα και Λογισμικό
4. Εφαρμογές και εξελίξεις





1. Εισαγωγή: Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό και στα Linked Data
2. Η Inbound/Outbound Linked Data Προσέγγιση
3. Υλοποίηση: Πρότυπα και Λογισμικό
4. Εφαρμογές και εξελίξεις

- **Έλλειψη δομής: Πληροφορία \neq Δεδομένα.** Ο Παγκόσμιος Ιστός, με την μορφή που παρέχει την πληροφορία (ιστοσελίδες), περιορίζεται ουσιαστικά στην παρουσίασή της. Δεν μπορεί να δώσει την δομή της πληροφορίας, ακόμα και αν υπάρχει τέτοια (π.χ. η πληροφορία αυτή προέρχεται από κάποια Βάση Δεδομένων): έχουμε ανάμιξη παρουσίασης και περιεχομένου με την μορφή εγγράφων.
- **Έλλειψη σημασιολογικού προσδιορισμού: Τι σημαίνει αυτό?** Ακόμα και αν μπορέσουμε να διαχωρίσουμε την παρουσίαση από το περιεχόμενο, η ερμηνεία του τελευταίου είναι κάτι που εναπόκειται αποκλειστικά στον αναγνώστη: δεν είναι δυνατή η αυτόματη επεξεργασία και ο συνδυασμός των δεδομένων, καθώς δεν προσδιορίζεται η 'σημασία' τους.

Ο Σημασιολογικός Ιστός

- Τις ελλείψεις αυτές υπόσχεται να αντιμετωπίσει ο Σημασιολογικός Ιστός, επιχειρώντας να φέρει στοιχεία **Αναπαράστασης Γνώσης** και **Τεχνητής Νοημοσύνης** στην δομή του Παγκόσμιου Ιστού, αποτελώντας την μετεξέλιξή του.

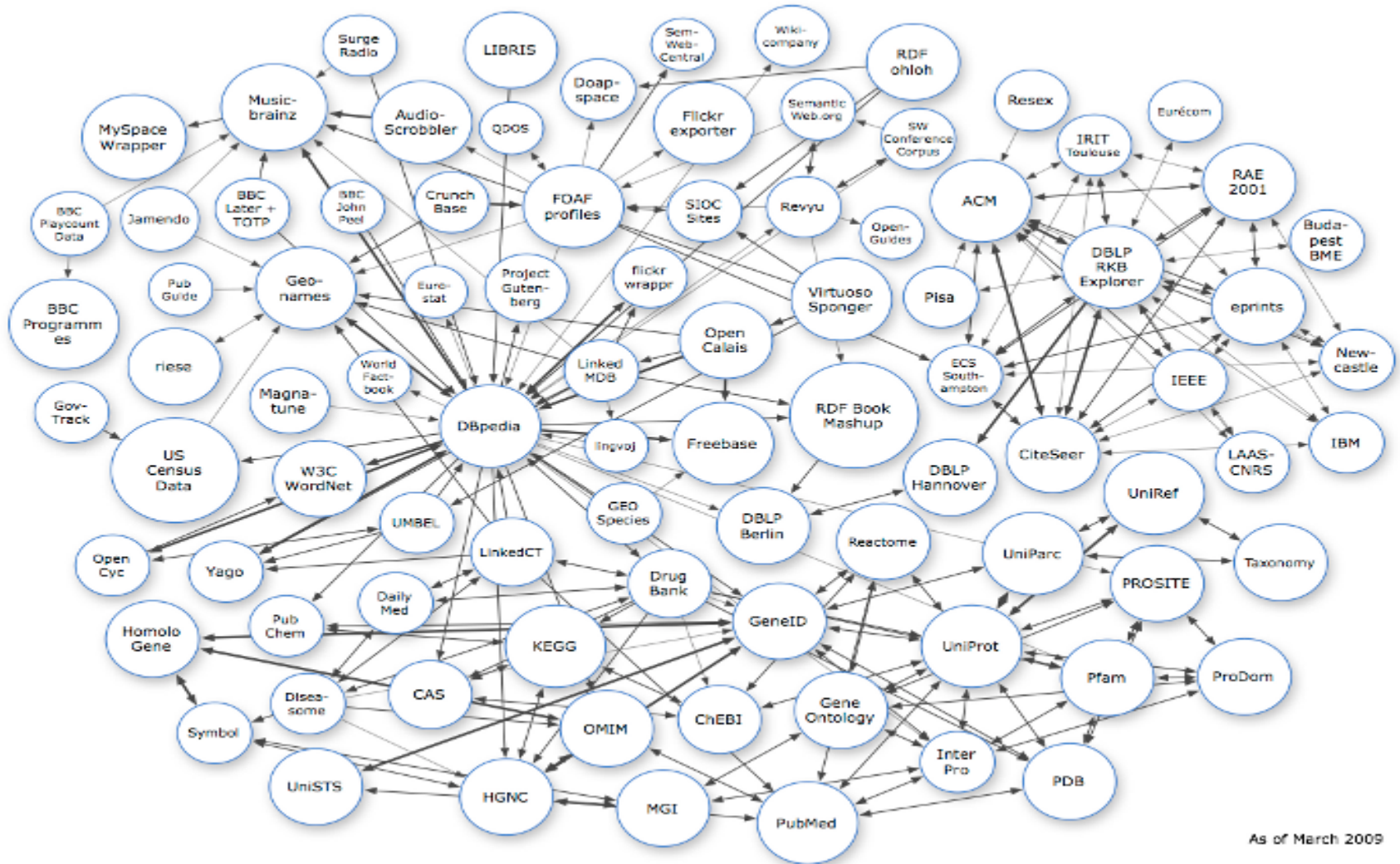


-  Το πρότυπο XML (eXtended Markup Language) είναι το πλέον διαδεδομένο πρότυπο δομημένης αναπαράστασης δεδομένων και η υποστήριξή του εξασφαλίζει την δυνατότητα ανταλλαγής δεδομένων ανάμεσα σε διαφορετικές εφαρμογές και συστήματα.
-  RDF(S). Η γλώσσα RDF (Resource Description Framework) είναι ένα μοντέλο δεδομένων για τα αντικείμενα και τις μεταξύ τους σχέσεις, με αναπαράσταση και σε μορφή XML. Το RDF Schema με τη σειρά του είναι ένα λεξιλόγιο για την περιγραφή των ιδιοτήτων και των κλάσεων RDF αντικειμένων, με μια σημασιολογία για τις γενικευμένες ιεραρχίες τέτοιων ιδιοτήτων και κλάσεων.
-  Η γλώσσα OWL (Web Ontology Language) προσθέτει επιπλέον δυνατότητες για την περιγραφή των ιδιοτήτων και των κλάσεων.
-  Η γλώσσα SPARQL (Simple Protocol and RDF Query Language) είναι η αντίστοιχη της SQL για ερωτήσεις σε σημασιολογικά δεδομένα. Επιπλέον, αποτελεί και πρωτόκολλο πρόσβασης σε αυτά μέσω HTTP.

- Τα RDF δεδομένα διαφέρουν από τα σχεσιακά δεδομένα ως προς το μοντέλο που εκφράζουν: το RDF είναι γράφος
- Τα RDF δεδομένα έχουν την μορφή τριπλετών (triples)
- <subject><predicate><object> : <cat><is-a><mammal>
- Το RDF(S) δίνει ένα πρώτο επίπεδο προσδιοριστικής λογικής: κλάσεις και ταξονομικές σχέσεις (ιεραρχία)
- Η OWL προσθέτει επιπλέον δυνατότητες αξιωματικών προσδιορισμών, περιορισμών και συμπερασμού

- Χρήση του Σημασιολογικού Ιστού για να επιτευχθεί πρόσβαση σε δεδομένα, αλλά και διασύνδεση μεταξύ τους
- Οι παρακάτω βασικές αρχές των Linked Data προσδιορίστηκαν από τον Sir Tim Berners Lee, εφευρέτη του Παγκόσμιου Ιστού, με την προοπτική ‘να κάνουν τον Ιστό να λειτουργήσει σωστά’
- Χρήση των υπαρχόντων προτύπων του Διαδικτύου και υλοποίηση 4 βασικών αρχών:
 1. Χρήση URIs ως προσδιοριστών για αντικείμενα κ έννοιες
 2. Χρήση HTTP URIs ώστε οι προσδιοριστές να μπορούν να αναζητηθούν
 3. Παροχή πληροφοριών μέσω σχετικών προτύπων (RDF, SPARQL) όταν αναζητείται κάποιο URI
 4. Χρήση συνδέσμων προς άλλα URIs ώστε να είναι δυνατή η πρόσβαση σε επιπλέον πληροφορία
- Η σημασιολογία των δεδομένων προσδιορίζεται μέσω οντολογιών/λεξιλογίων

The Linked Data cloud



- Εξάγει δομημένη πληροφορία από την Wikipedia και την δημοσιεύει ως Linked Data.
- Χρησιμοποιεί μια OWL οντολογία για την αναπαράσταση των πληροφοριών και την δημοσίευση τους σε RDF μορφή
 - Places
 - Person
 - Organization
 - ...
- Επιτρέπει την χρήση της γλώσσας SPARQL για την ανάκτηση δεδομένων μέσω ερωτημάτων
- Η πληροφορία που εξάγεται αποθηκεύεται σε μία τεράστια cross-domain knowledge base (479 εκατομμύρια RDF τριάδες)
- Υπάρχουν 2 εκδόσεις:
 - Αγγλική
 - Γερμανική

1. Εισαγωγή: Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό και στα Linked Data
2. Η Inbound/Outbound Linked Data Προσέγγιση
3. Υλοποίηση: Πρότυπα και Λογισμικό
4. Εφαρμογές και εξελίξεις

- Πως μπορούν να φανούν χρήσιμα σε πρακτικές εφαρμογές τα Linked Data?
- Κάθε κόμβος/εφαρμογή στον Σημασιολογικό Ιστό μπορεί να δρα είτε ως καταναλωτής (Inbound Linked Data) είτε ως πάροχος (Outbound Linked Data)
- Ως καταναλωτής, τα πλεονεκτήματα είναι εμφανή: μπορεί να χρησιμοποιήσει εξωτερικά δεδομένα για να εμπλουτίσει το περιεχόμενο και να παρέχει πρόσθετες υπηρεσίες

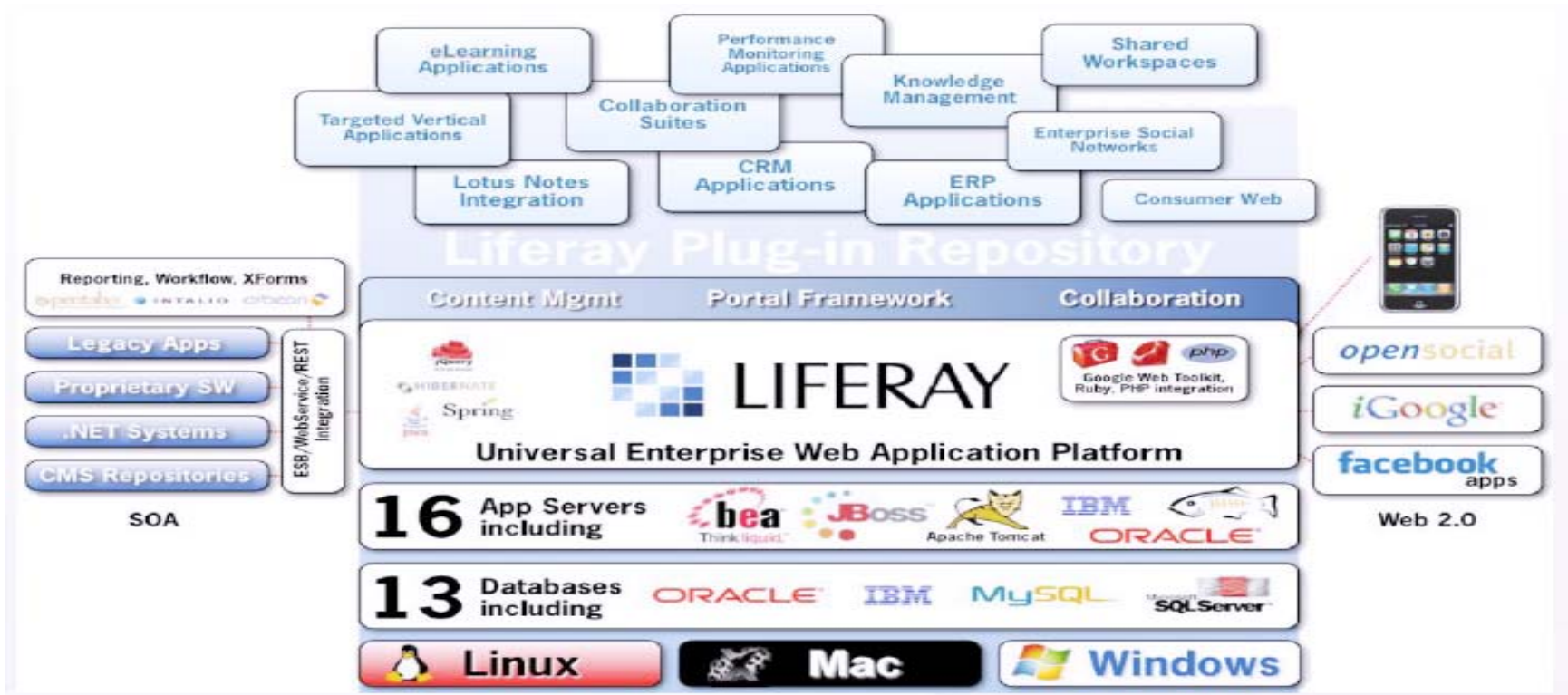
- Ως πάροχος, τα πλεονεκτήματα είναι λιγότερο εμφανή, πλην υπαρκτά:



- Η προσέγγιση παρουσιάστηκε στα πλαίσια του διεθνούς διαγωνισμού **Linking Open Data Triplification Challenge 2009**, ο οποίος διοργανώνεται από διεθνή επιτροπή ειδικών υπό την αιγίδα του Sir Tim Berners Lee με σκοπό την προώθηση των Linked Data διαθέτοντας στο κοινό:
 - Ανοιχτά Datasets σε μορφή Linked Data
 - Ανοιχτό λογισμικό που μπορεί να χρησιμοποιηθεί για την παραγωγή Linked Data
- Outbound Linked Data εφαρμογή: Liferay Linked Data Module
- Inbound Linked Data εφαρμογή: Εφαρμογή Απόδοσης Νοήματος Tags

1. Εισαγωγή: Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό και στα Linked Data
2. Η Inbound/Outbound Linked Data Προσέγγιση
3. Υλοποίηση: Πρότυπα και Λογισμικό
4. Εφαρμογές και εξελίξεις

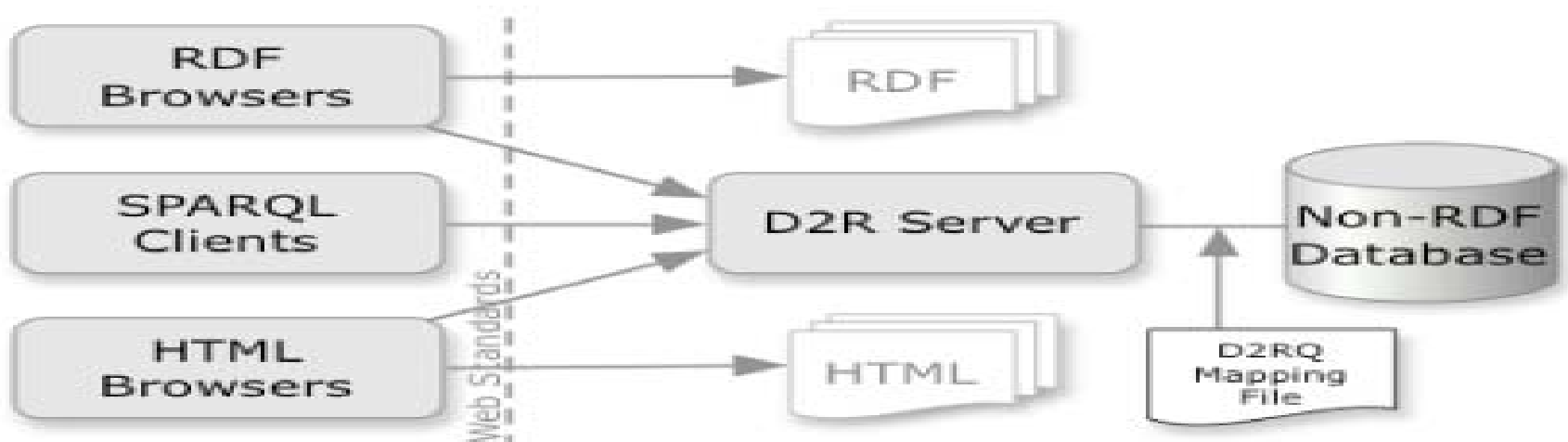
- Liferay: open source Portal/CMS framework (Java, Portlet container)
 - Πάνω από 10 χρόνια ανάπτυξης
 - Μεγάλο μερίδιο αγοράς διεθνώς, με πελάτες όπως UN, Cisco, BMW



- Διάθεση του περιεχομένου που παράγεται από τον Liferay (blogs, web content, forums, wikis...) σε Linked Data μορφή
- Επιπλέον πληροφορία: users, groups, organizations, tags..
- SPARQL endpoint.
- Χρήση ανοιχτού λογισμικού: D2R Server + Mapping language
- Χρήση προτύπων λεξιλογίων για το περιεχόμενο
- Διαθέσιμο στο Sourceforge, LGPL license

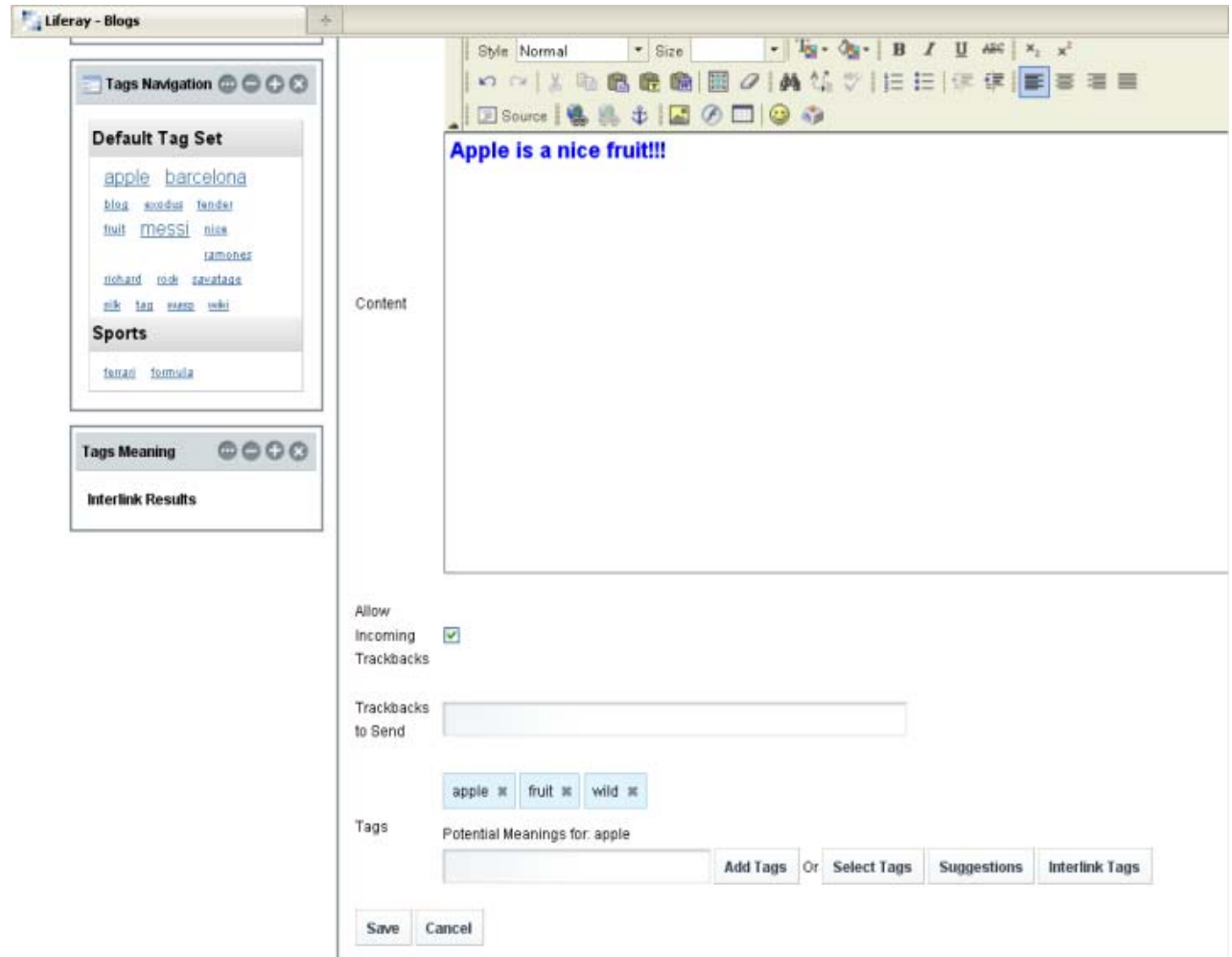
- Η αναπαράσταση των RDF δεδομένων γίνεται με την χρήση κατάλληλων λεξιλογίων, όπως:
 - **FOAF**: Friend-Of-A-Friend –Λεξιλόγιο περιγραφής χρηστών
 - **DC**: Dublin Core - Λεξιλόγιο ορισμού ιδιοτήτων μεταδεδομένων
 - **SIOC**: Semantically Interlinked Online Communities - Λεξιλόγιο αναπαράστασης κοινοτήτων
 - **SKOS**: Simple Knowledge Organization System - Λεξιλόγιο αναπαράστασης ταξονομιών και δομημένης γνώσης
 - **MOAT** (Meaning Of A Tag): Λεξιλόγιο περιγραφής Tags, χρήσης και διαφορετικών νοημάτων τους.
- Η επαναχρησιμοποίηση των λεξιλογίων προωθεί την διασυνδεσιμότητα και βοηθά στην εύκολη και άμεση επεξεργασία των δεδομένων από τις εκάστοτε εφαρμογές.

- Είναι ένα εργαλείο το οποίο αντιστοιχίζει (mapping) μία Σχεσιακή Βάση Δεδομένων σε μορφή RDF και δημοσιεύει το περιεχόμενο της ως Linked Data
- Επιτρέπει την πλοήγηση και την ανάκτηση των RDF δεδομένων
- Πρόσβαση μέσω SPARQL
- Χρησιμοποιείται για την δημοσίευση επιλεγμένων δεδομένων της βάσης περιεχομένου του Liferay Server



- Αναπτύχθηκε στην πλατφόρμα Liferay Portal
- Διασυνδέει δεδομένα τύπου Tag που δημιουργούνται στο περιβάλλον του Liferay και τα οποία χρησιμοποιούνται για την επισημείωση περιεχομένου
- Παρέχει στο χρήστη γραφικό περιβάλλον για τον σημασιολογικό προσδιορισμό της έννοιας κάθε Tag στα πλαίσια της εκάστοτε χρήσης του
- Μπορεί να χρησιμοποιηθεί για
 - ακριβέστερη αναζήτηση
 - εύρεση όμοιων εννοιών
 - αντιστοίχιση tags
 - ...
- Ο προσδιορισμός γίνεται μεταξύ πιθανών εννοιών που ανακτώνται από την DBpedia ασύγχρονα, με την χρήση εξειδικευμένου λογισμικού

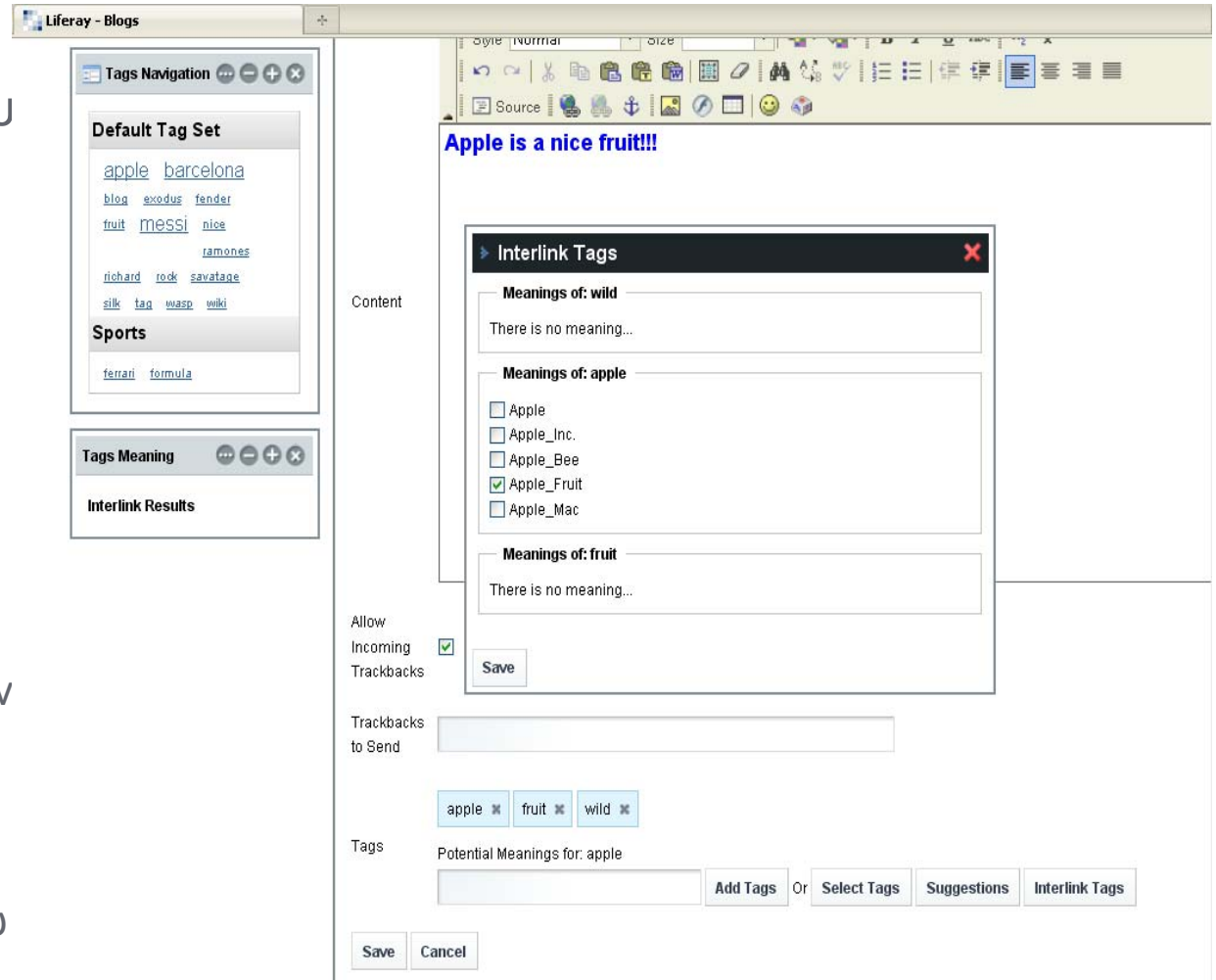
- Δημιουργία Blog “Apple”.
- Προσθήκη Tag στο περιεχόμενο του Blog.
 - Προσθήκη νέου tag “Wild”.
 - Προσθήκη tag “Apple” και “Fruit” από λίστα.



The screenshot shows the Liferay - Blogs editor interface. On the left, there are two panels: "Tags Navigation" and "Tags Meaning". The "Tags Navigation" panel displays a "Default Tag Set" with various tags like "apple", "barcelona", "blog", "wonder", "tender", "fruit", "messi", "nisa", "lamoner", "richard", "rodi", "savatias", "nik", "tan", "wam", "wiki", and a "Sports" section with "ferrari" and "formula". The "Tags Meaning" panel shows "Interlink Results". The main content area on the right displays the text "Apple is a nice fruit!!!" in blue. Below the content area, there are options for "Allow incoming Trackbacks" (checked), a "Trackbacks to Send" field, and a "Tags" section showing "apple", "fruit", and "wild" tags. At the bottom, there are buttons for "Add Tags", "Select Tags", "Suggestions", "Interlink Tags", "Save", and "Cancel".

Inbound Linked Data: Εφαρμογή Απόδοσης Νοήματος Tags

- Διασύνδεση αυτών μέσω του “Interlink Tags”
 - Αναζήτηση πιθανών νοημάτων από προηγούμενες διασυνδέσεις.
 - Ανάκτηση πιθανών νοημάτων για κάθε tag.
 - Συνολική παρουσίαση αποτελεσμάτων
 - Επιλογή κατάλληλου νοήματος
 - Αποθήκευση νοήματος μέσω του “Save”.

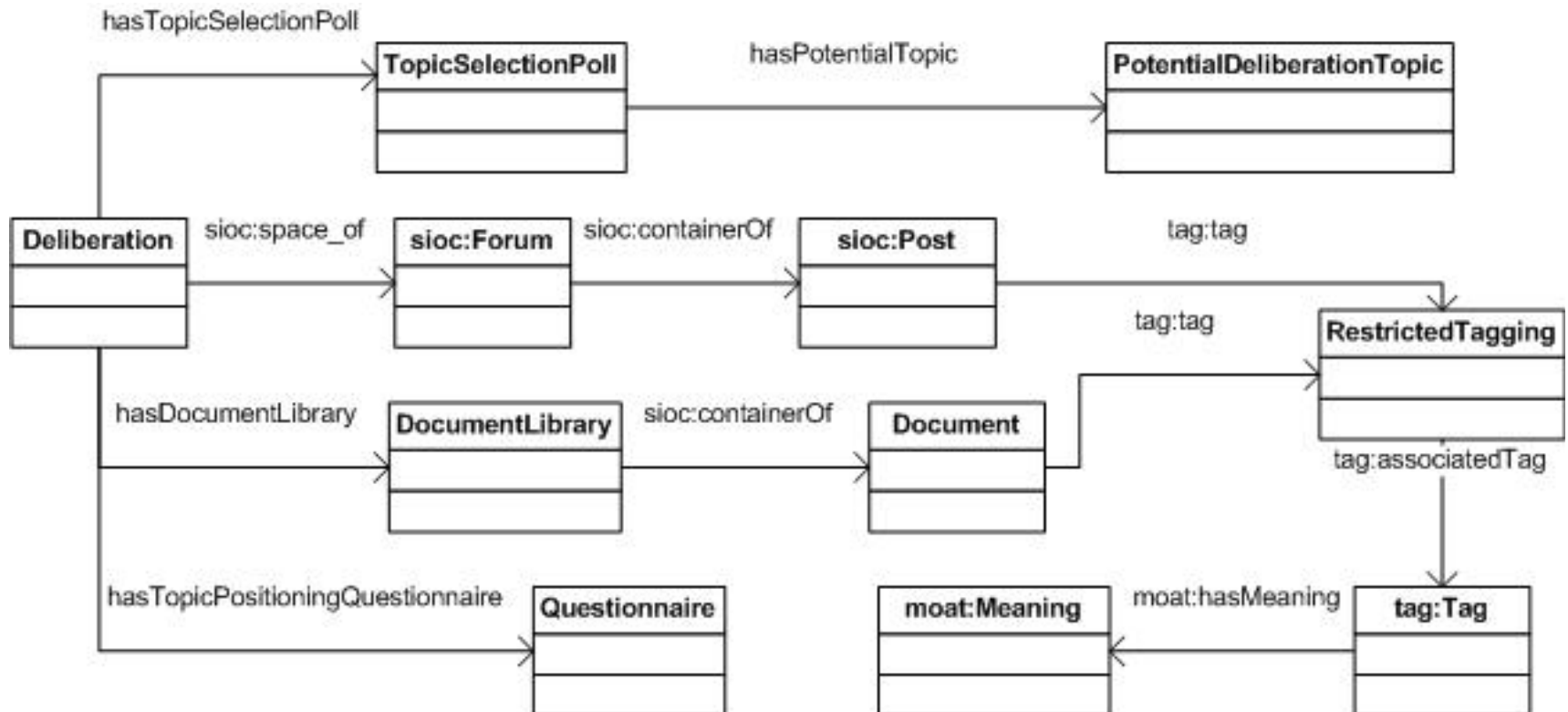


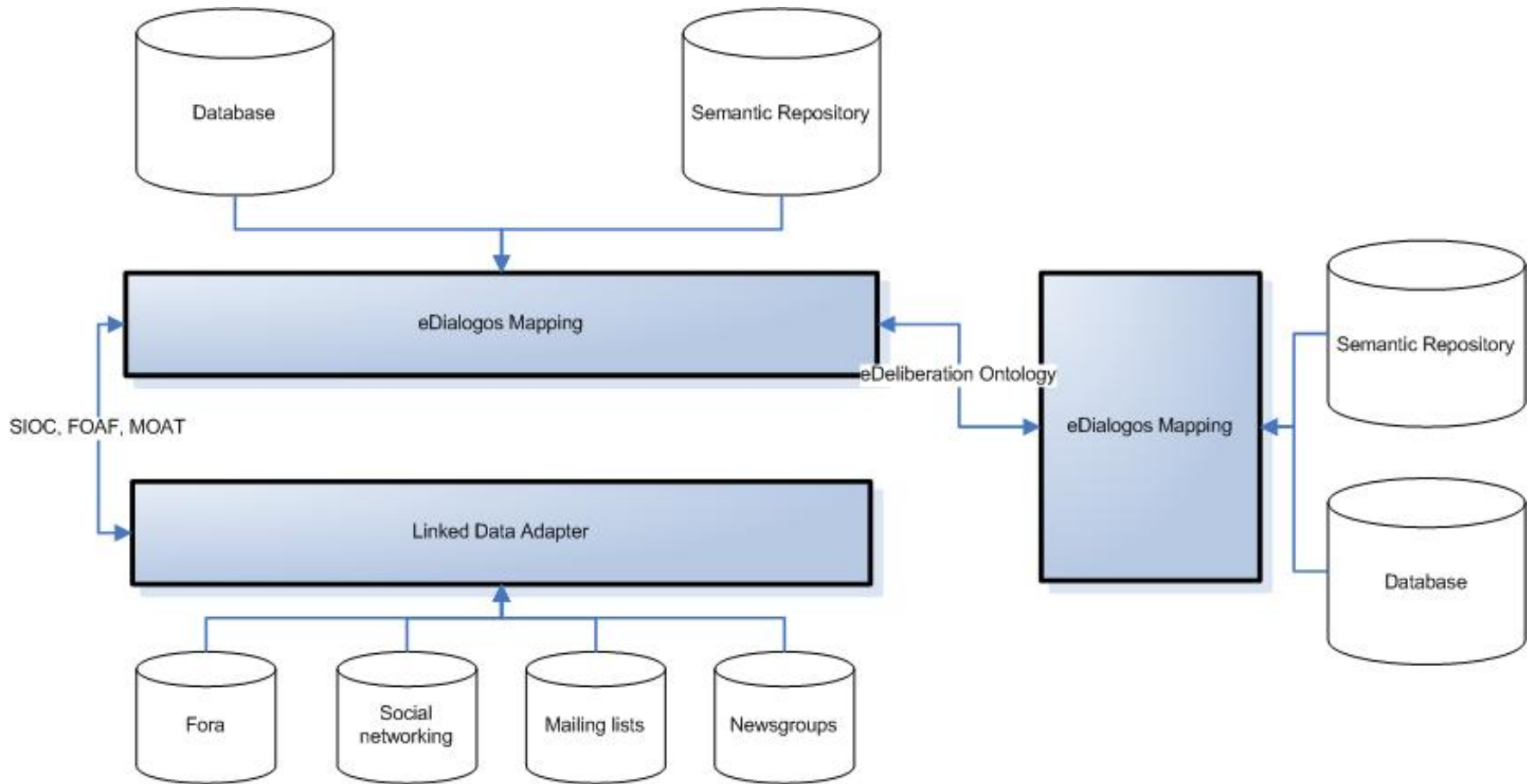
The screenshot shows the Liferay - Blogs interface. On the left, there are two panels: "Tags Navigation" and "Tags Meaning". The "Tags Navigation" panel shows a "Default Tag Set" with tags like apple, barcelona, blog, exodus, fender, fruit, messi, nice, ramones, richard, rook, savatage, silk, tag, wasp, wiki, and a "Sports" section with ferrari and formula. The "Tags Meaning" panel shows "Interlink Results". The main content area displays the text "Apple is a nice fruit!!!". A modal window titled "Interlink Tags" is open, showing "Meanings of: wild" (no meaning), "Meanings of: apple" (Apple, Apple_Inc., Apple_Bee, Apple_Fruit, Apple_Mac), and "Meanings of: fruit" (no meaning). The "Apple_Fruit" option is selected. Below the modal, there are checkboxes for "Allow Incoming Trackbacks" (checked) and a "Save" button. At the bottom, there are "Tags" (apple, fruit, wild) and "Potential Meanings for: apple" with buttons for "Add Tags", "Select Tags", "Suggestions", and "Interlink Tags".

1. Εισαγωγή: Από τον Παγκόσμιο Ιστό στον Σημασιολογικό Ιστό και στα Linked Data
2. Η Inbound/Outbound Linked Data Προσέγγιση
3. Υλοποίηση: Πρότυπα και Λογισμικό
4. Εφαρμογές και εξελίξεις

- Το Liferay Linked Data Module έχει ενσωματωθεί από την IMC ως μέρος της πλατφόρμας Ηλεκτρονικής Συμμετοχής eDialogos, όπως έχει χρησιμοποιηθεί στα αντίστοιχα έργα που έχουν υλοποιηθεί από την Νομαρχία Σάμου και τις πόλεις της Άρτας και της Θεσσαλονίκης
- Δημιουργία ‘οικοσυστήματος διαλόγου’, στο οποίο μπορούν να συμμετέχουν ως πάροχοι ή/και καταναλωτές όλοι όσοι επιθυμούν μέσω απευθείας πρόσβασης στα δεδομένα
 - Διαφάνεια - Προσβασιμότητα: Ανοιχτά Δεδομένα
 - Συμβατότητα: Χρήση προτύπων, πρόσβαση απ’ευθείας στα δεδομένα, χωρίς μεσολάβηση API

- Συμβατότητα με τα (γενικά) λεξιλόγια που χρησιμοποιήθηκαν, επέκταση με επιπλέον έννοιες που αφορούν το πεδίο της Ηλεκτρονικής Συμμετοχής





- Συνεργασία με Liferay
 - Ενσωμάτωση του Linked Data Module στο επίσημο distribution (μετά την επερχόμενη έκδοση)
 - Consulting για την επέκταση και ενσωμάτωση ακόμα περισσότερων χαρακτηριστικών στο μέλλον
- Συνεργασία με DBpedia, Πανεπιστήμια
 - Δημιουργία Ελληνικής DBpedia
- Διάθεση του eDialogos σε Open Source έκδοση
 - Το ακριβές πλαίσιο θα καθορισθεί εντός του 2010

- [1] Anadiotis, G., Andriopoulos, P., Vekris, D. and Zosakis, A. Linked data for the masses – using open source infrastructure and the inbound/outbound linked data approach to bring added value to end user applications. In I-KNOW 09 and I-SEMANTICS 09, 2009. See http://i-semantics.tugraz.at/2009/triplification/04_liferay_TriplificationChallenge2009.pdf
- [2] Liferay Linked Data Module: <http://sourceforge.net/projects/liferayldm/>
- [3] Bryan Cheung (Liferay CEO), “Liferay Linked Data Module”, Liferay Blog, December 12, 2009. See <http://www.liferay.com/web/bryan.cheung/blog/-/blogs/liferay-linked-data-module>
- [4] Anadiotis, G., Alexopoulos, P., Mpaslis, K., Zosakis, A., Kafentzis, K. and Kotis, K. (2010). Facilitating Dialogue – Using Semantic Web Technology for eParticipation. Extended Semantic Web Conference, June 2010, Crete.

????