

Primene Web of Data tehnologija

(u kompanijama, državnim organima,
bibliotekama, muzejima, klinikama,...)

Jelena Jovanovic

jeljov@gmail.com

Tema 1:

- Kada koristiti Web of Data (WoD) tehnologije?
- Za koje tipove problema se njihova primena preporučuje?

Slabo definisani (open ended) problemi

- **Model podataka nije u potpunosti poznat/definisan**
 - vrlo je moguće da će biti potrebno menjati/proširivati model u (bliskoj) budućnosti i da će novi zahtevi za promenama biti učestani
- **Model (slučajevi) korišćenja nisu u potpunosti poznati/definisani**
 - vrlo je moguće da će korisnici zahtevati neke nove i/ili drugačije oblike analiza, izveštaja, i sl., kao i da će ti novi zahtevi učestano stizati
- **Mogući korisnici sistema nisu u potpunosti definisani**
 - vrlo je moguće da će biti potrebno da se sistem proširi na dodatne kategorije korisnika u odnosu na one inicijalno anticipirane
 - dodavanje novih korisnika povlači za sobom i nove vrste podataka i nove oblike korišćenja podataka

Slabo definisani (open ended) problemi

Ukratko:

što je problem slabije definisan,
to će primena Web of Data tehnologija biti
korisnija tj doneti veće benefite u odnosu na
'tradicionalna' rešenja

Još neke vrste problema pogodne za primenu Web of Data (WoD) tehnologija

- Kada je potrebno integrisati podatke iz različitih izvora korišćenjem otvorenih standarda i tehnologija
 - tipičan slučaj u kompanijama nakon merdžera i akvizicija
 - WoD tehnologije omogućuju eksplicitno definisanje konteksta mapiranja podataka i razloga zašto je mapiranje realizovano baš na taj način

“Whenever you integrate content, you have the same problem: the integration entails mapping data to data and does not capture any intrinsic understanding of the process or context of how the data is related. It's essentially dumb mapping.” [1]

Još neke vrste problema pogodne za primenu WoD tehnologija

- Kada je potrebno raditi sa nestrukturiranim sadržajima
 - Izveštaji i drugi oblici poslovne dokumentacije,
 - Web stranice,
 - novinski članci,
 - stručni tekstovi,
 - ...

Problemi za koje WoD tehnologije *nisu* najbolje rešenje

- Rad sa ogromnim količinama podataka (Big Data)
 - još uvek su manje efikasne od tradicionalnih rešenja
 - kombinovana primena sa NoSQL bazama podataka i njima srodnim tehnologijama vodi unapređenju efikasnosti
- Visoka frekvencija transakcija
 - više hiljada read/write zahteva u sekundi po serveru
- Numeričke operacije nad ogromnom količinom (terabajti) numeričkih podataka

Optimalno rešenje: kombinovanje prednosti različitih tehnologija

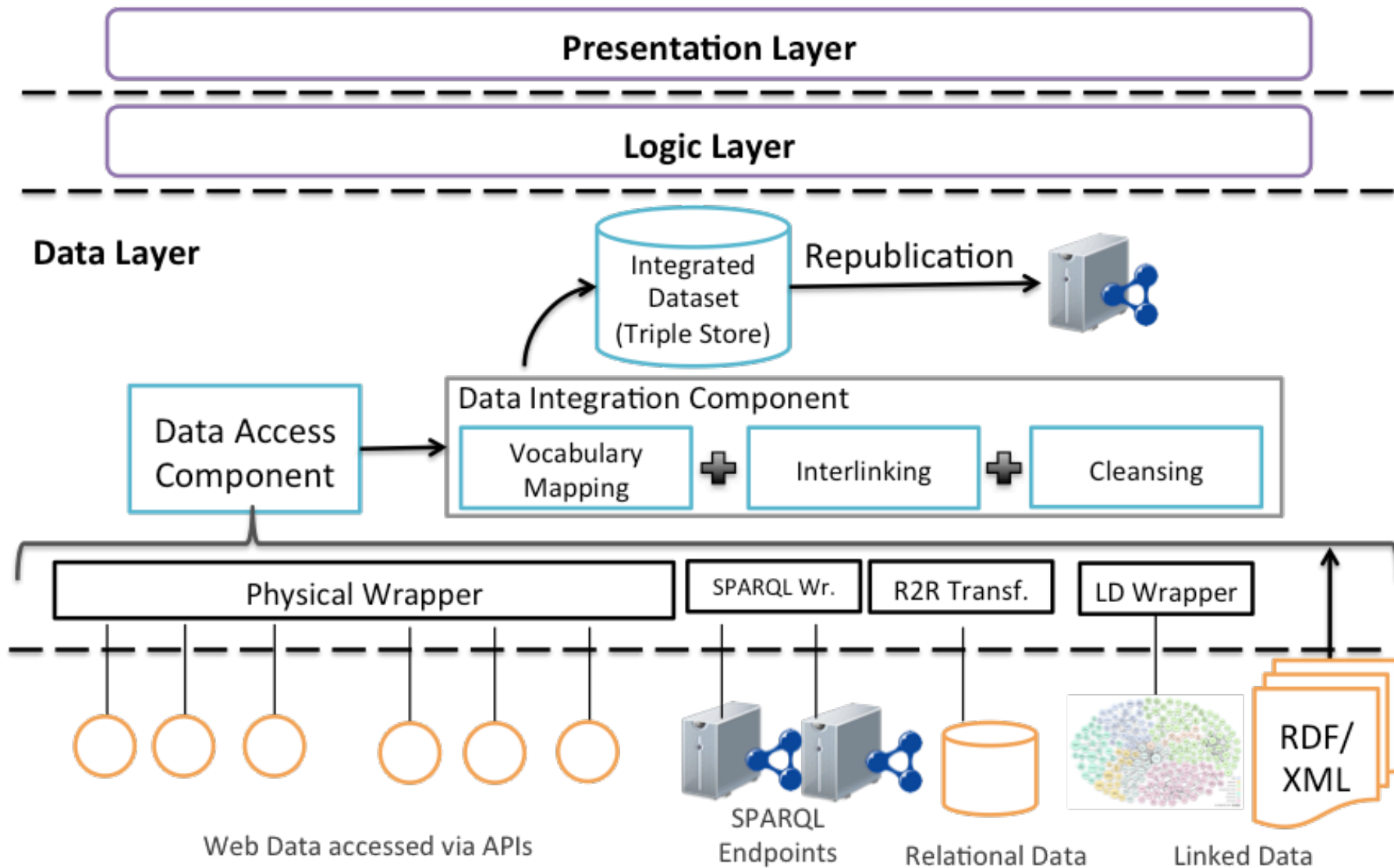
- WoD tehnologije se mogu jednostavno kombinovati sa drugim tehnologijama kako bi se došlo do optimalnog rešenja za dati problem
 - Npr., nad relacionom bazom podataka koja obrađuje visokofrekventne transakcije može se postaviti SPARQL endpoint kako bi podaci kojima raspolaže bili dostupni za federativne upite
 - Ili, relevantni segment podataka iz nekog velikog data warehouse-a bi se mogao mapirati u odgovarajući vokabular/ontologiju i koristiti kao osnova sistema preporuke ili neke druge ‘inteligentne’ aplikacije koja koristi ontologiju za rezonovanje nad podacima

Zaključak:

kada koristiti Web of Data tehnologije

- Kada je potrebno je raditi sa dokumentima i drugim oblicima nestrukturiranih podataka
- Kada se očekuje dodavanje novih ili izmena postojećih podataka sa kojima se radi
- Kada se očekuje dodavanje novih ili izmena postojećih načina korišćenja podataka
- Kada se očekuje pojava novih korisnika sa novim/drugačijim zahtevima
- Kada se obim podataka sa kojima se radi *ne* meri u petabajtima
- Kada učestnost transakcija nije prevelika
- Kada nisu potrebna zahtevna numerička izračunavanja

Opšta arhitektura WoD aplikacija



Tema 2: tipični slučajevi primene WoD tehnologija

Agilna integracija podataka u velikim poslovnim /državnim (eko)sistemima

- [Ministarstvo odbrane SAD-a](#): povezivanje više različitih informacionih sistema koji su pod jurisdikcijom ovog Ministarstva
- [Farmaceutska kompanija Biogen Idec](#): integracija podataka u sistemu upravljanja lancem snabdevanja (> 30 kompanija)
- [Chevron](#): integracija podataka (vezanih za gas i naftu) koji dolaze iz različitih izvora i različitih su vrsta i formata
- [Amdocs](#): bolje opsluživanje korisnika kroz integraciju i analizu raspoloživih podataka o korisnicima; posebna primena u domenima sa velikim brojem korisnika (npr., telekomunikacije, zdravstvo)

Anotacija, klasifikacija i pretraga nestrukturiranih sadržaja

- Biblioteke, muzeji, arhive i druge institucije sa velikim kolekcijama umetničkih, naučnih, stručnih radova
 - [Library of Congress](#)
 - [British Library](#)
 - [ZBW German National Library of Economics](#)
 - [Amsterdam Museum](#)
 - [Europeana](#)
 - ...

Dinamičko publikovanje semantički povezanih sadržaja

- **BBC**

- Od 2009. god, BBC razvija i aktivno koristi softversku platformu za dinamičko publikovanje Web sadržaja zasnovanu na WoD tehnologijama ([slajdovi o tome](#))
- Kompletan BBC web sajt posvećen 2010 World Cup-u je zasnovan na WoD tehnologijama ([članak o tome](#))
- BBC je dalje unapredio svoju WoD platformu za potrebe Web sajta za 2012 Summer Olympics ([članak o tome](#))
- BBC i dalje razvija i aktivno koristi svoju WoD platformu ([članak na tu temu](#))

Dinamičko publikovanje semantički povezanih sadržaja

- WoD tehnologije su osnova soft. platformi za kreiranje i publikovanje sadržaja i drugih poznatih medijskih kuća:
 - Time Inc., Elsevier, the Library of Congress, Financial Times
- Osnovni razlozi za primenu WoD tehnologija u ovim organizacijama:
 - rad sa ogromnim količinama nestrukturiranih sadržaja
 - brojne i raznovrsne semantičke relacije između pojedinačnih (multi-medijalnih) sadržaja
 - veoma česte izmene sadržaja koji se publikuju i linkova koji povezuju te sadržaje

Semantička pretraga, otkrivanje (discovery), i preporuka sadržaja

- Seevl

- omogućuje pretraživanje i otkrivanje novih/nepoznatih muzičara, muzičkih albuma, pesama i njihovih međusobnih relacija
- povlači podatke iz različitih izvora na Web-u: MusicBrainz, Wikipedia, Freebase, BBC, NY Times
- objedinjuje podatke korišćenjem Music Ontology kao modela podataka, čime praktično formira ogroman (RDF) graf različitih entiteta iz domena muzike
- koristi tako kreirani graf podataka za preporuku muzičara i muz. numera, kao i za kreiranje objašnjenja za generisane preporuke
- obezbeđuje RESTful pristup svom grafu muzičkih entiteta - pogledati: <https://developer.seevl.fm/>