Matteo Palmonari

matteo.palmonari@disco.unimib.it

Making Sense of Big Data & Data Linking

The web of data and the open linked data cloud

ITIS Lab – Innovative Technologies for Interaction and Services

Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca

Outline

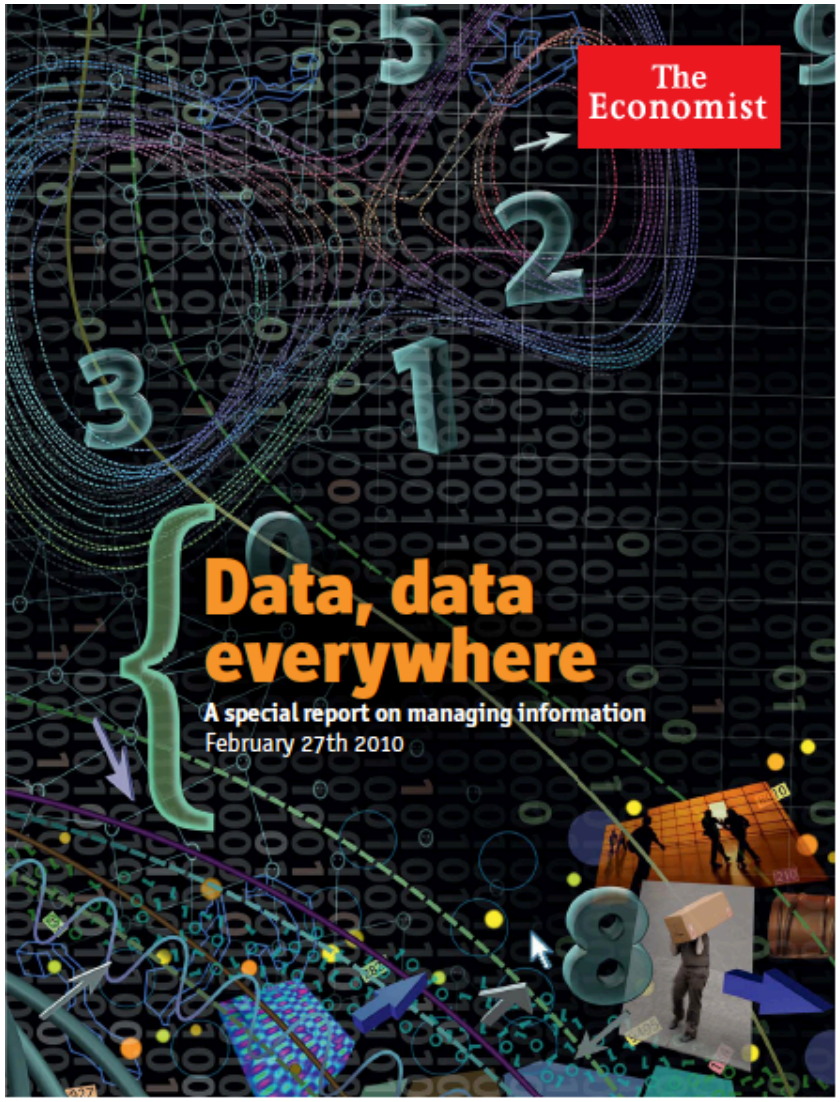
- Big Data
- Linking Web Data
- Complex Questions & Decision Making
- Open Data
- Linked Data

Big Data

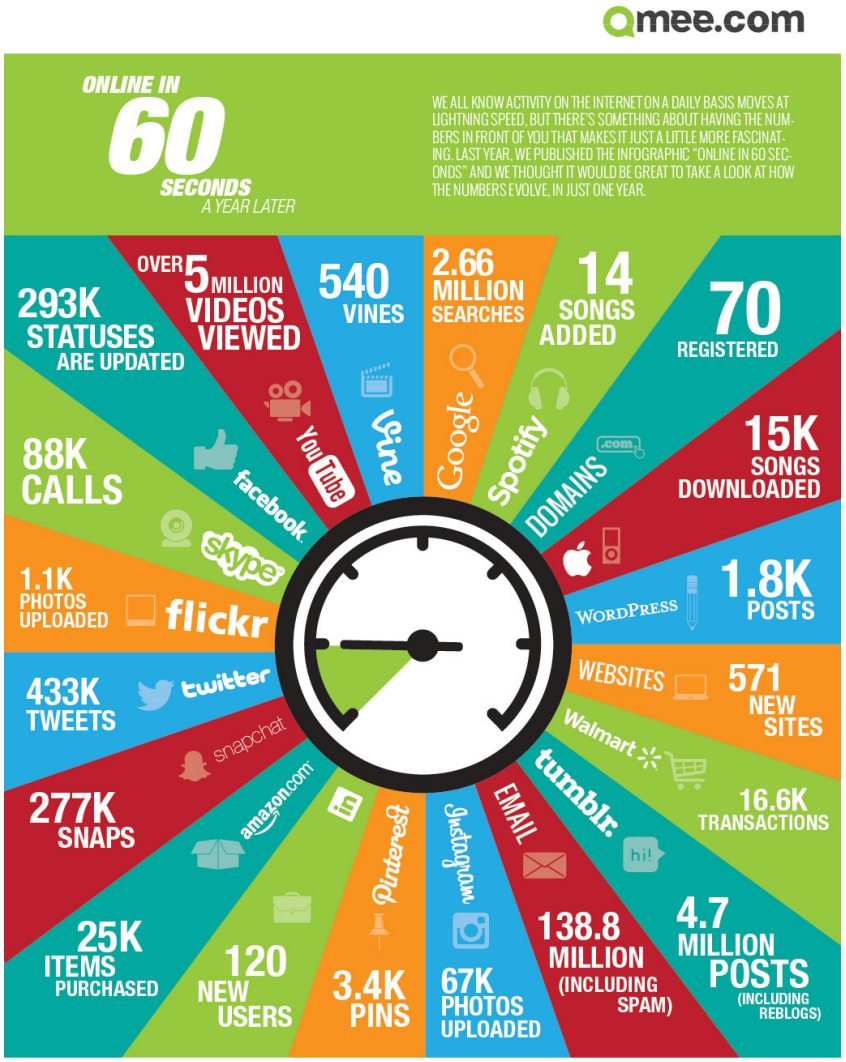
Volume Velocity Variety

In 2010

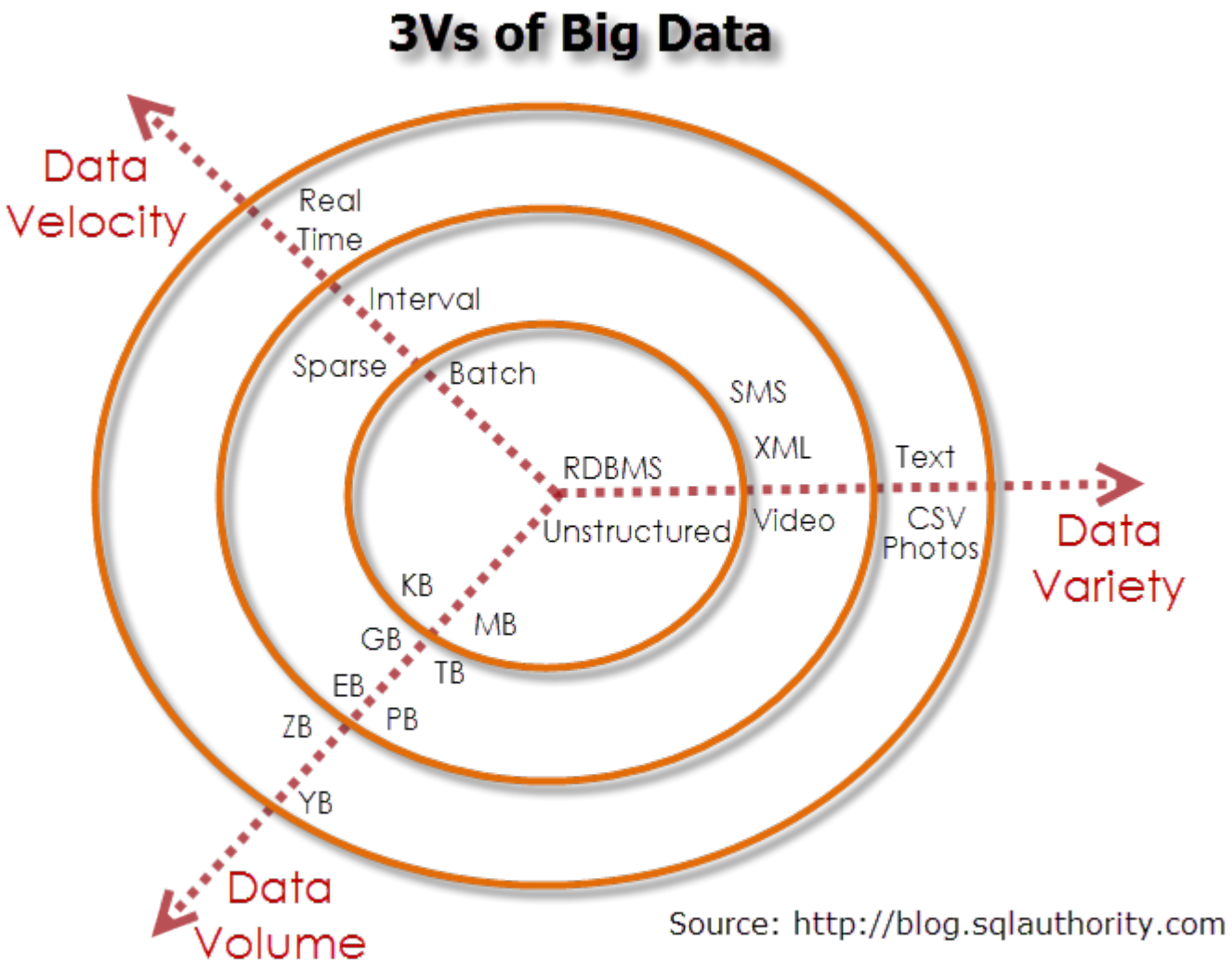
THE “BIG DATA” PHENOMENON



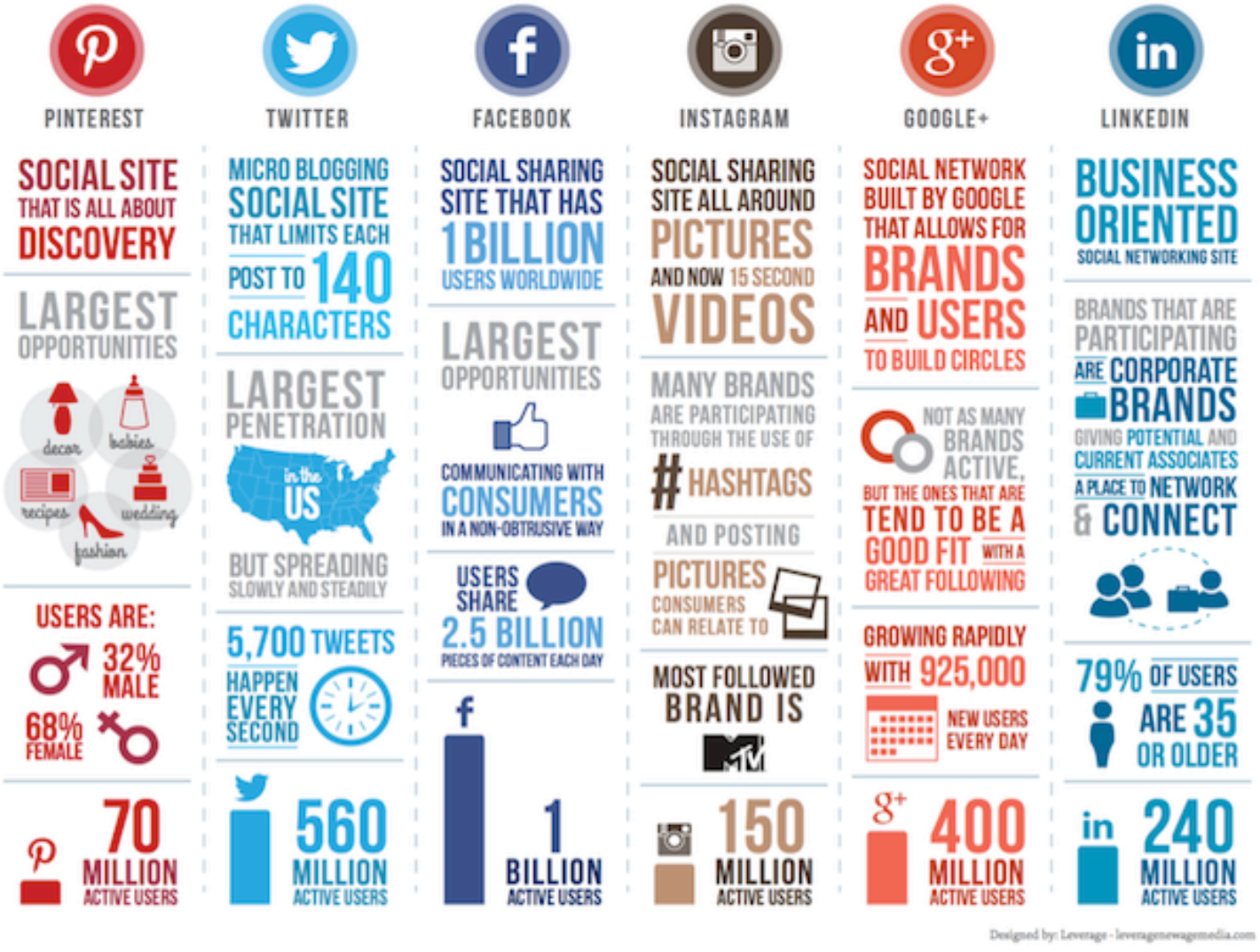
Every 60 secs in 2014



Big Data: Volume Velocity Variety



Variety (Social Media)

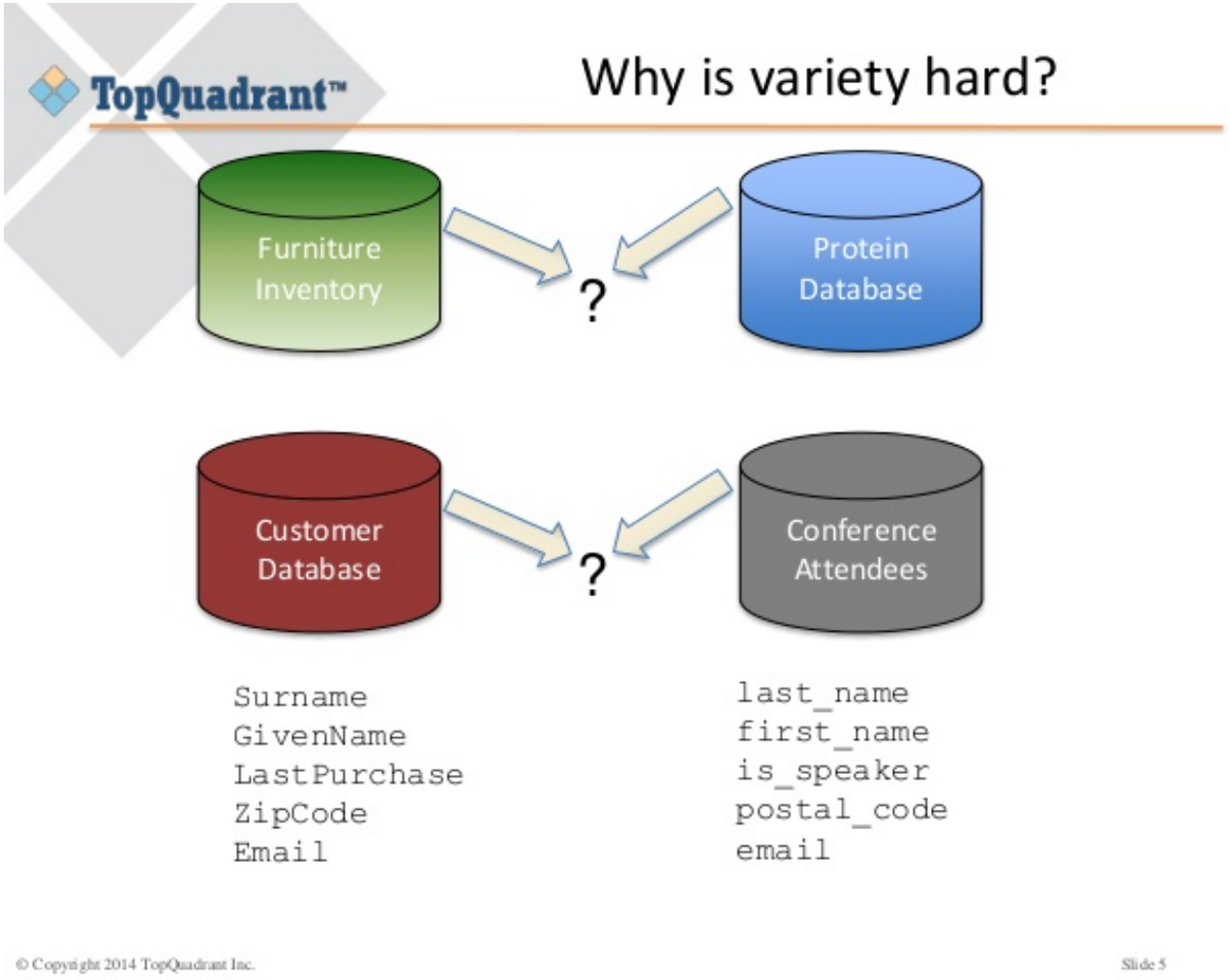


Variety

According to Kapow Software, a comprehensive list of typical Big Data sources comprises these nine categories: archives, docs, media, business apps, social media, public Web, data storage, machine log data, and sensor data.
Yellow means external data (public Web), red is internal (archives and data storage), and orange designates both (docs, media, business apps, social media, sensor data).

Source: http://labs.sogeti.com/three-data-scientists-share-six-insights-big-data-analytics/

Variety & Semantics (First Dive)



Web Data

Stadium	City
Eden Gardens	Kolkata
Feroz Shah Kotla Ground	Delhi
M. A. Chidambaram Stadium	Chepauk, Chennai
Wankhede Stadium	Mumbai
Green Park	Kanpur

1. Structured Data

- Organized in attributes
- Well defined schema, follow some order

e.g., Relational tables, tables in Wikipedia

2. Semi-Structured Data

- Hierarchical organization
- May have no schema

e.g., HTML Tables and spreadsheets with nested headers

	Temperate climate		Semi-arid climate		Arid climate	
	%	mm	%	mm	%	mm
Total precipitation	100	500–1,500	100	200–500	100	0–200
Evaporation /Evapotranspiration	~ 33	160–500	~ 50	100–250	~ 70	0–140
Groundwater recharge	~ 33	160–500	~ 20	40–100	~ 1	0–2
Surface runoff	~ 33	160–500	~ 30	60–150	~ 29	0–60

IBM's Watson Computer Made A BBQ Sauce, And It's Delicious

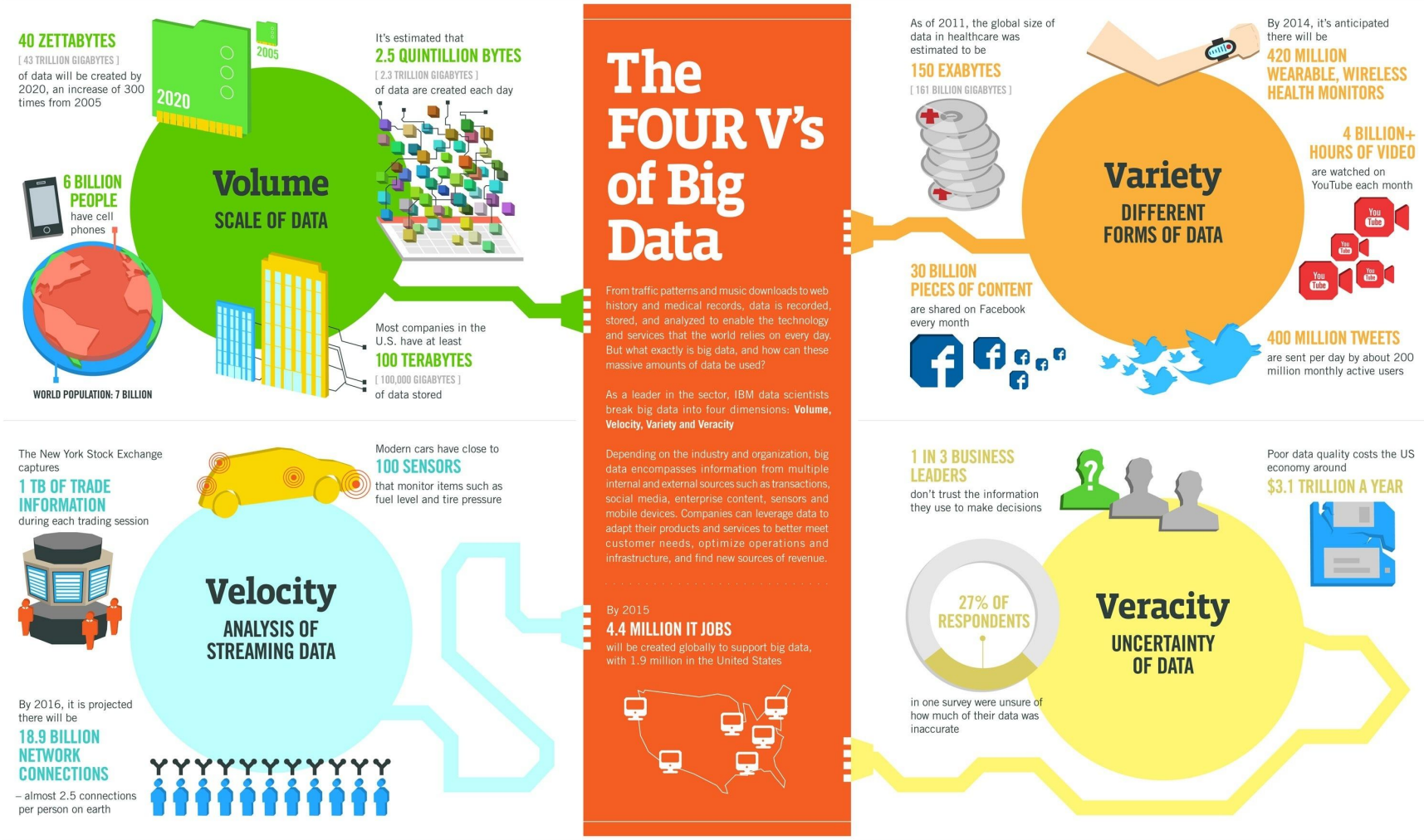
Watson, a cognitive computing system that can learn and process natural human language, has been one of IBM's most exciting projects of the last decade. Over the past few years, Watson has learned a variety of

3. Unstructured Data

- Natural language text

e.g., News articles, Wikipedia content

VVV + Veracity



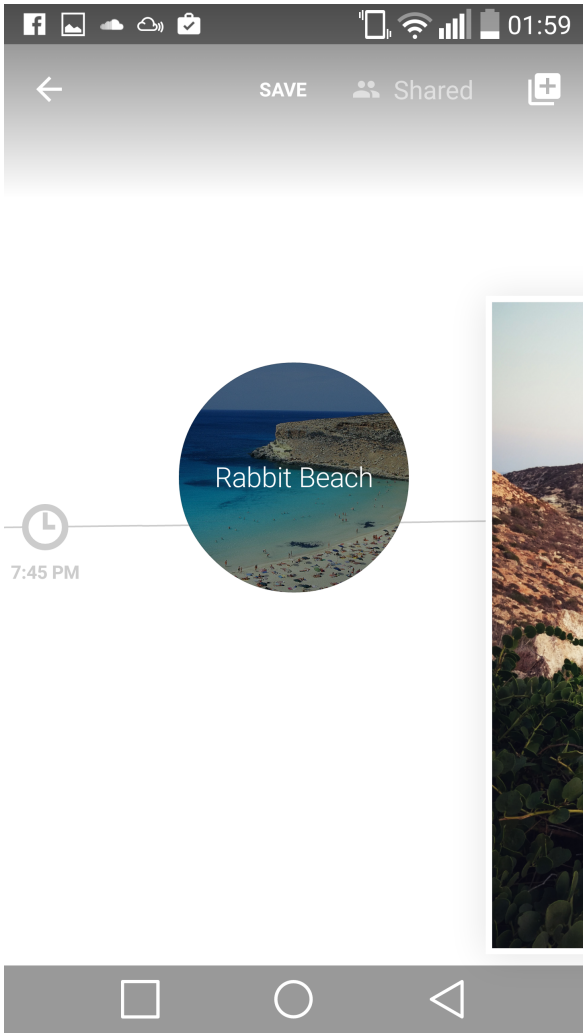
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

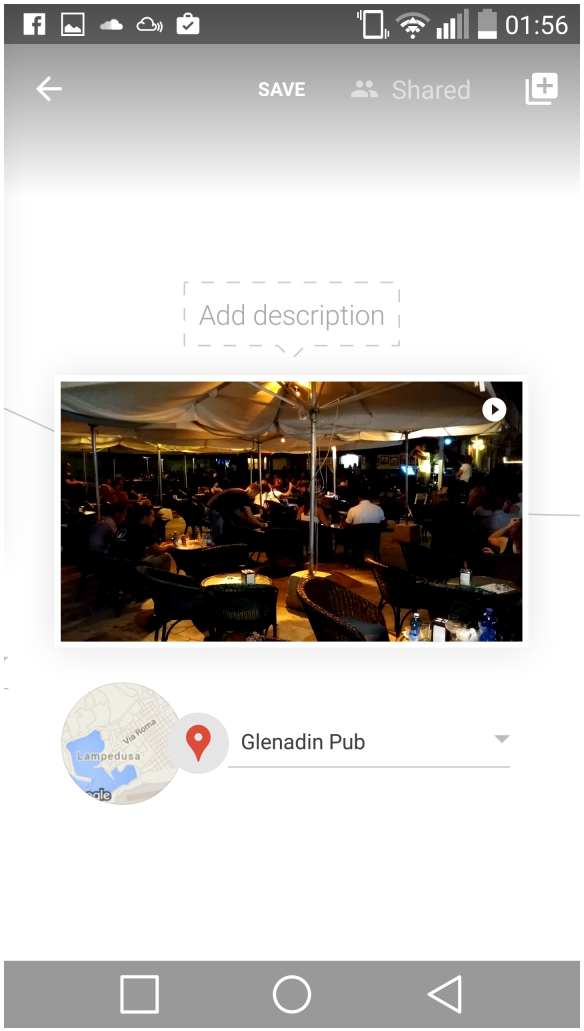
IBM

Linking Web Data

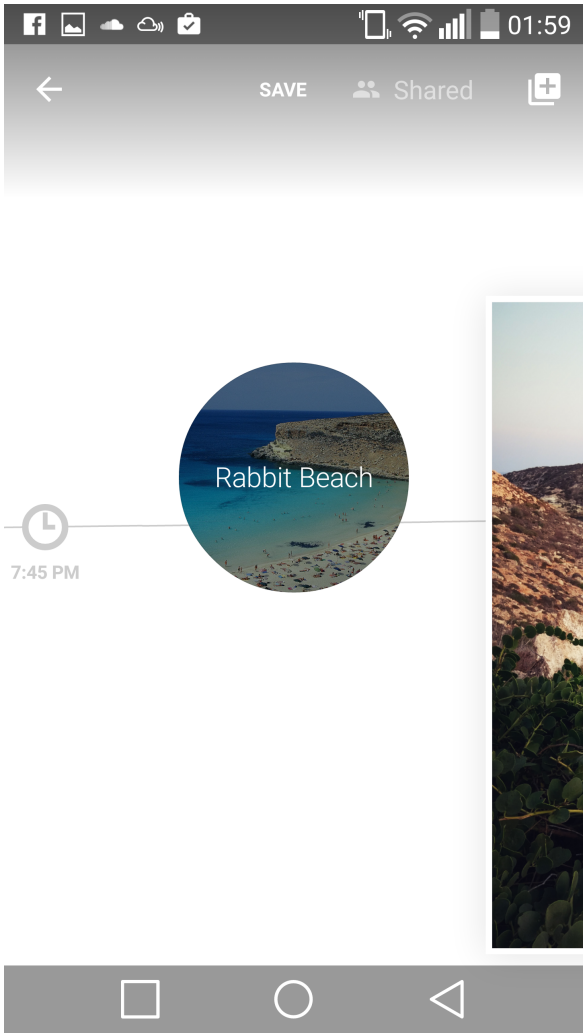
Ovvero perché le connessioni sono importanti per passare da un web di documenti a un web di dati

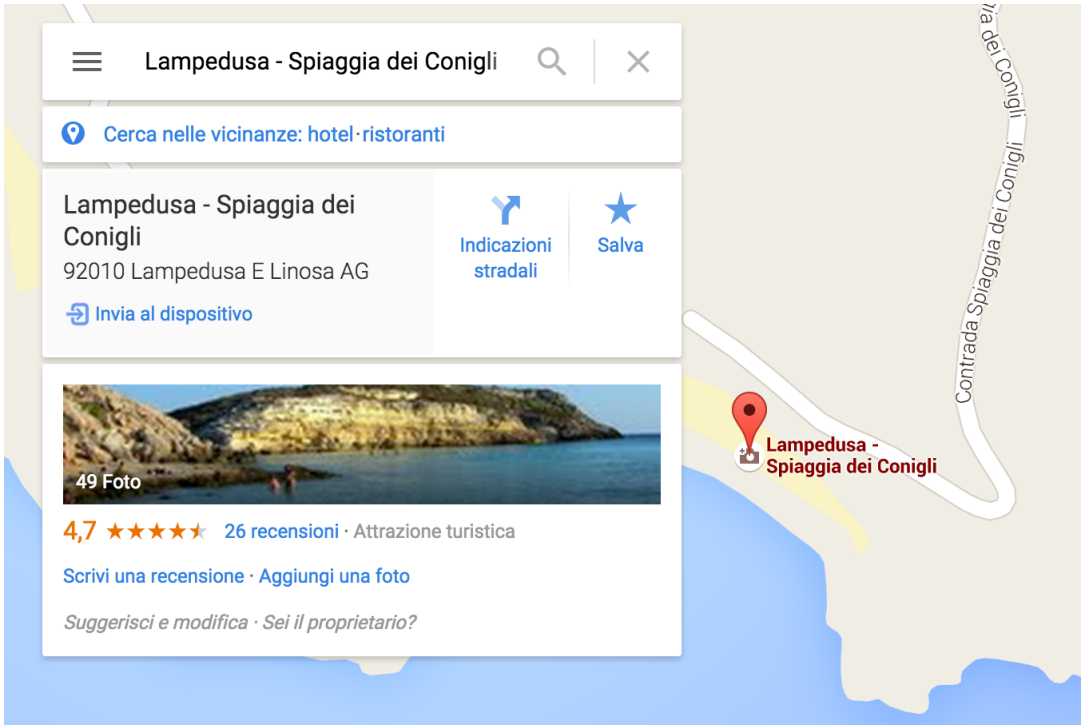
Connettere esperienze personali a luoghi



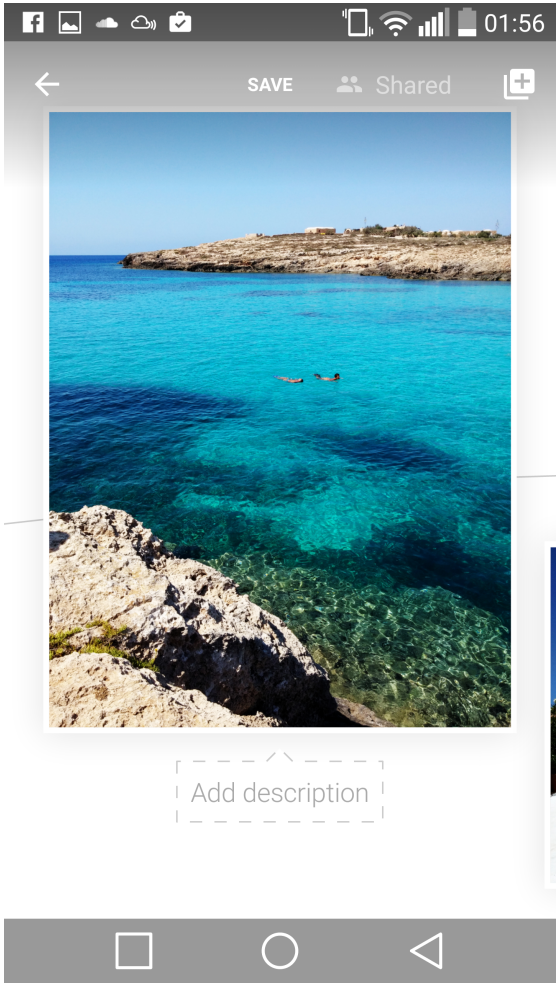


Connettere esperienze personali a luoghi

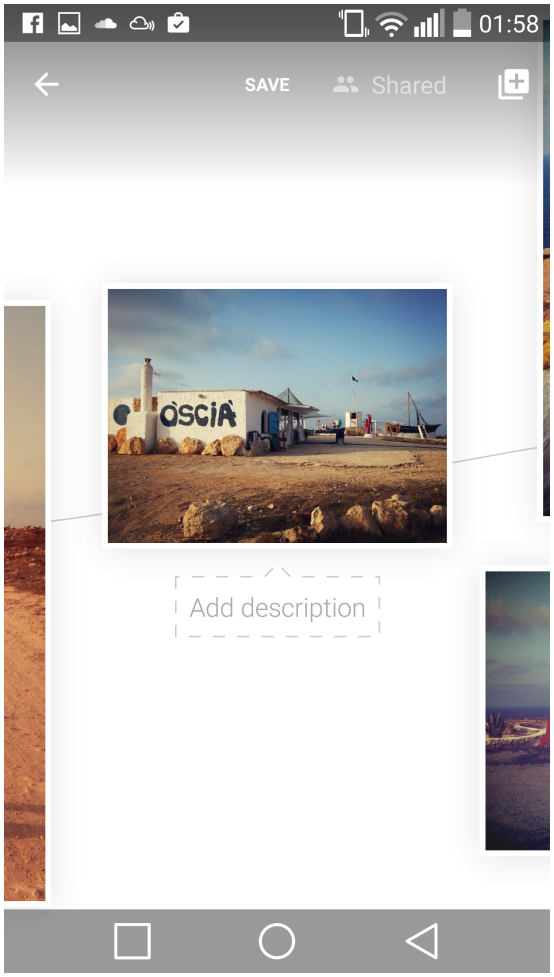


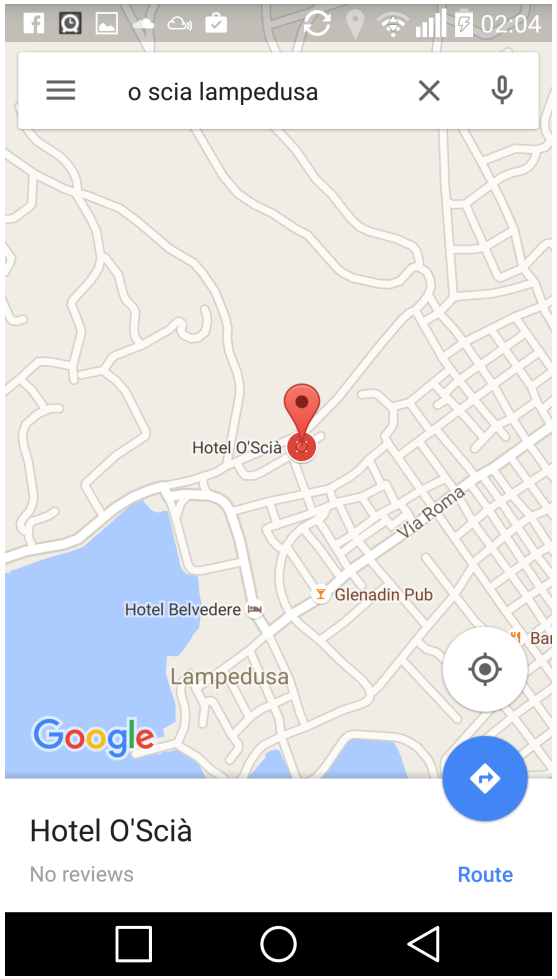


Il valore del dato per le applicazioni



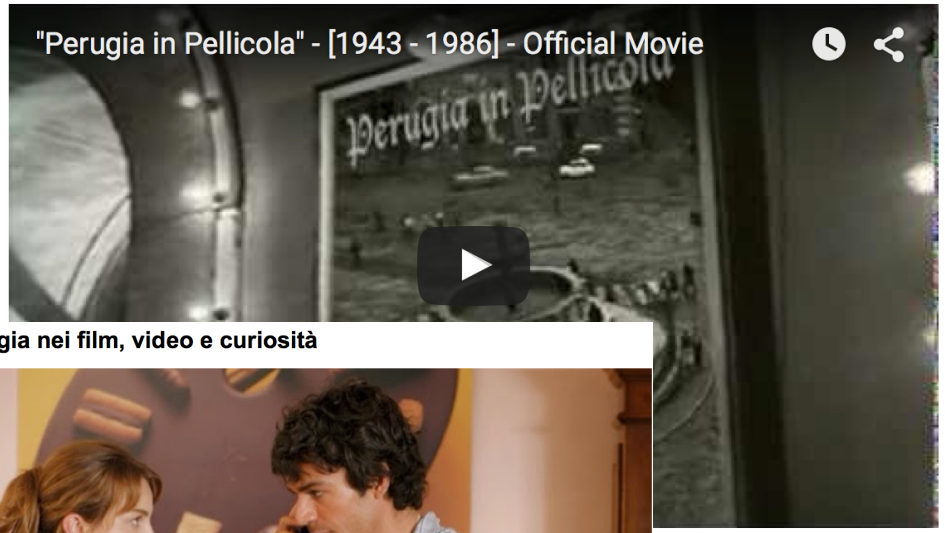
Il valore del dato per le applicazioni





Perugia Film: Documenti vs Dati

- Un appassionato di cinema vuole fare un tour in giro per l'Italia visitando luoghi di film a lui cari…
- **Quali film sono stati girati in Umbria?**



Perugia nei film, video e curiosità

In questi giorni gira in rete un video caricato su Youtube contenente gli spezzoni di film girati a Perugia, come *Fumo di Londra* (1966), con Alberto Sordi, o *I corpi presentano tracce di violenza carnale* (1973), con Luc Merenda, antesignano del sottogenere horror "Slasher", poi reso celebre dal serial cult Venerdì 13.

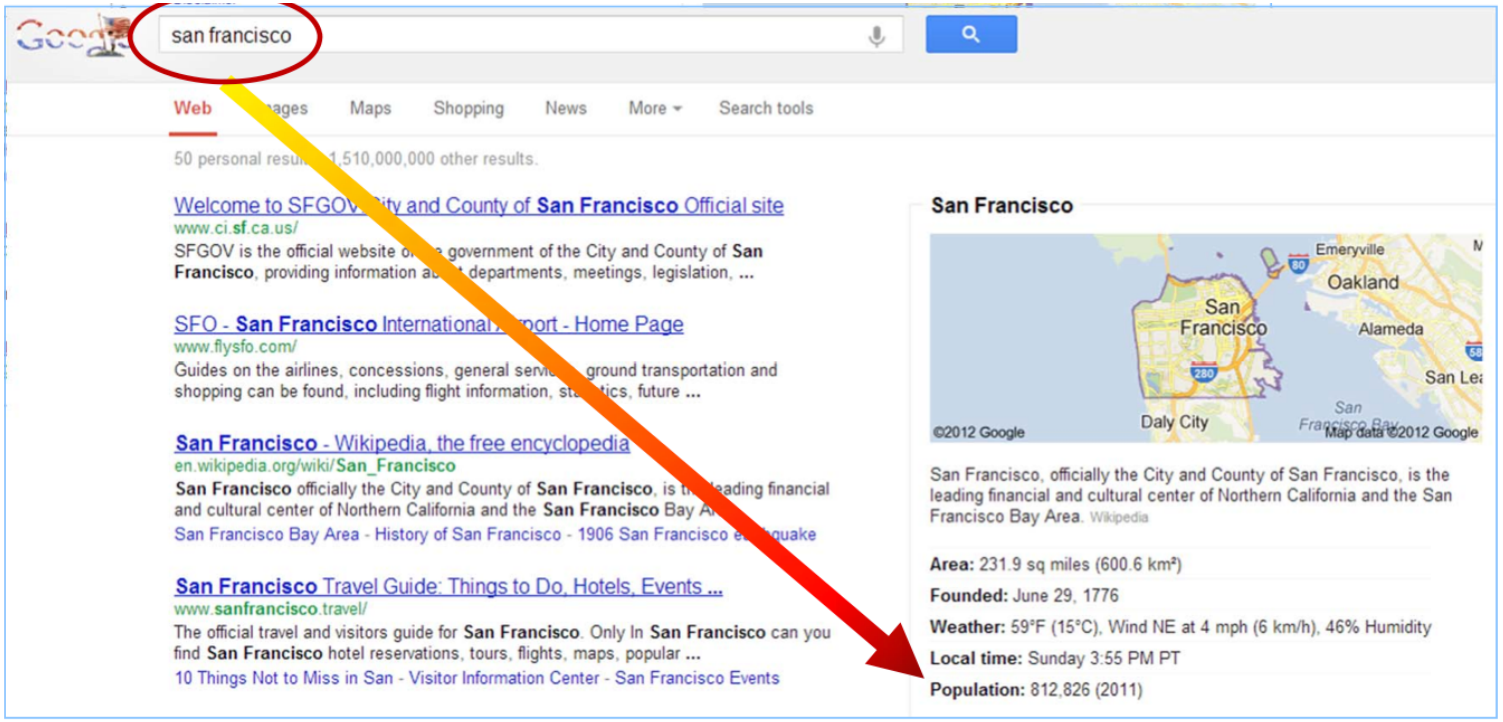
http://www.perugiacity.com/2014/01/perugia-nei-film-video-e-curiosita.html

Perugia Film: Documenti vs Dati

- Uno studioso di cinema vuole studiare il rapporto tra cinema e territorio umbro
- **Quali film sono stati girati in Umbria fuori da Perugia?**
- **Quali film sono stati girati a Perugia tra il 1970 e il 1979?**
- **Quali film recitati da Alberto Sordi sono stati girati a Perugia tra il 1960 e il 1979?**
- **…**

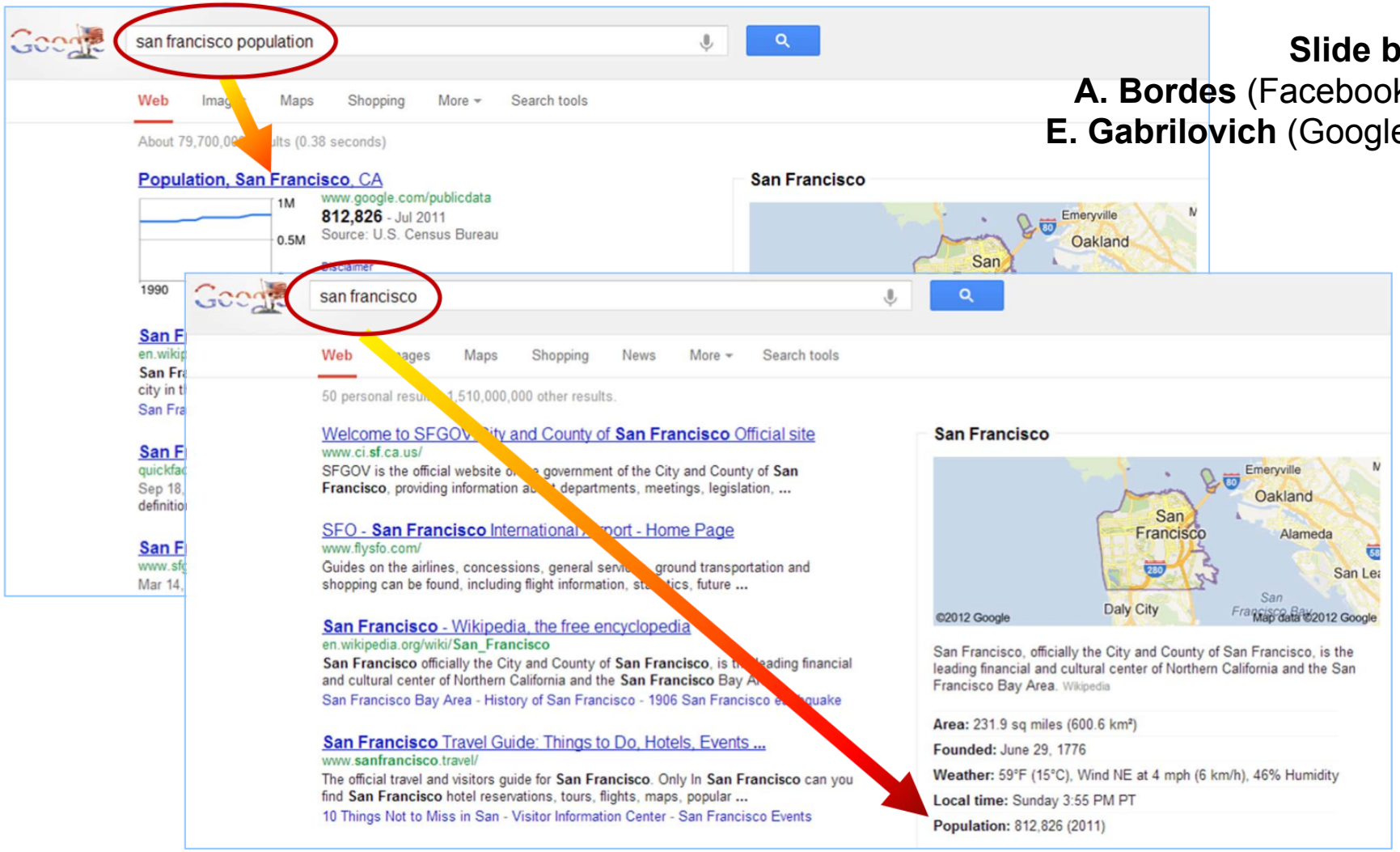
Documenti vs dati

Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)



source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Informazioni fattuali nei risultati di ricerca



Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Rappresentare informazioni (e connessioni) mediante *knowledge graph*

Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)

Why (knowledge) graphs?

- We're surrounded by **entities**, which are connected by **relations**
- We need to store them somehow, e.g., using a **DB** or a **graph**
- **Graphs** can be processed **efficiently** and offer a convenient **abstraction**

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Rappresentare informazioni (e connessioni) mediante *knowledge graph*

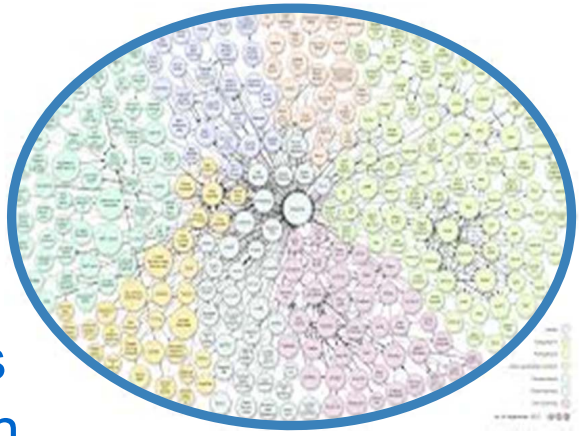
Knowledge graphs

Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)

Freebase

yago
select knowledge

DBpedia



NELL: Never-Ending Language Learning

OpenIE (Reverb, OLLIE)

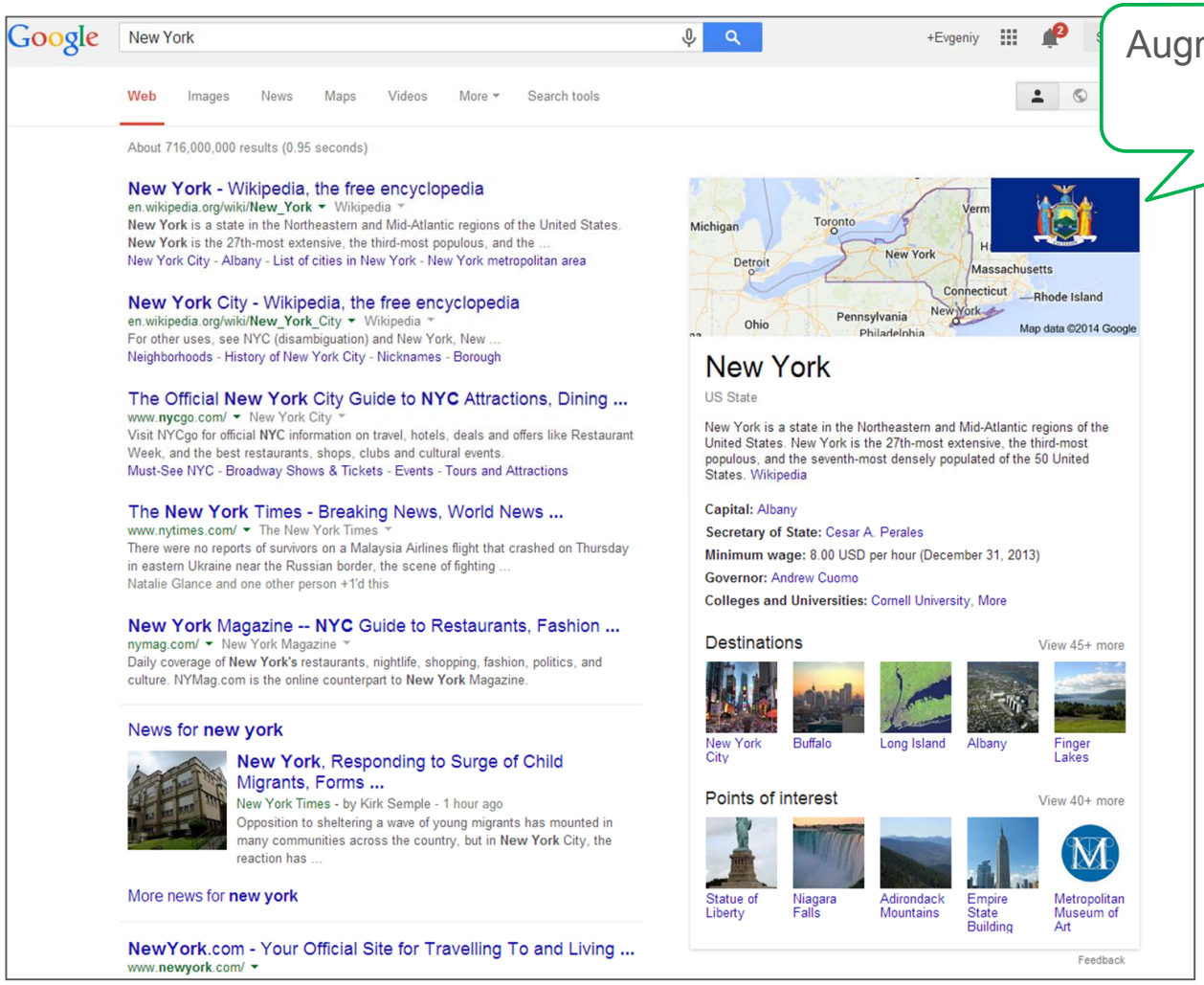
Facebook's Entity Graph

Microsoft's Satori

Google's Knowledge Graph

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Documenti vs dati

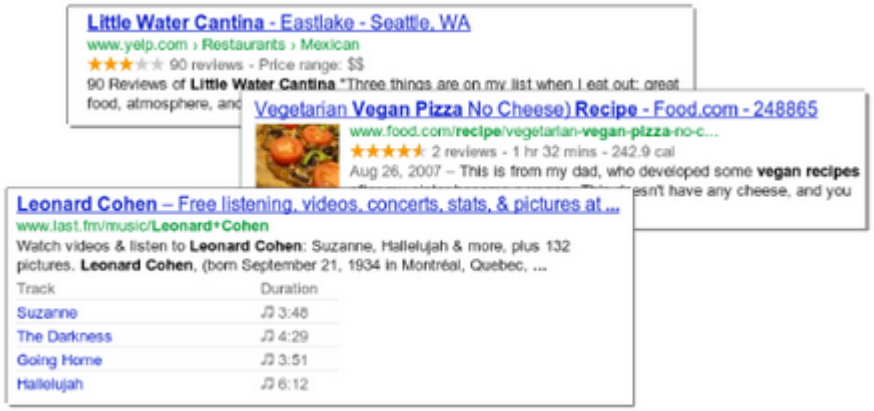


Augmenting the presentation with relevant facts

Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Rich Snippets in Search Results



Also useful for
Search Engine Optimization

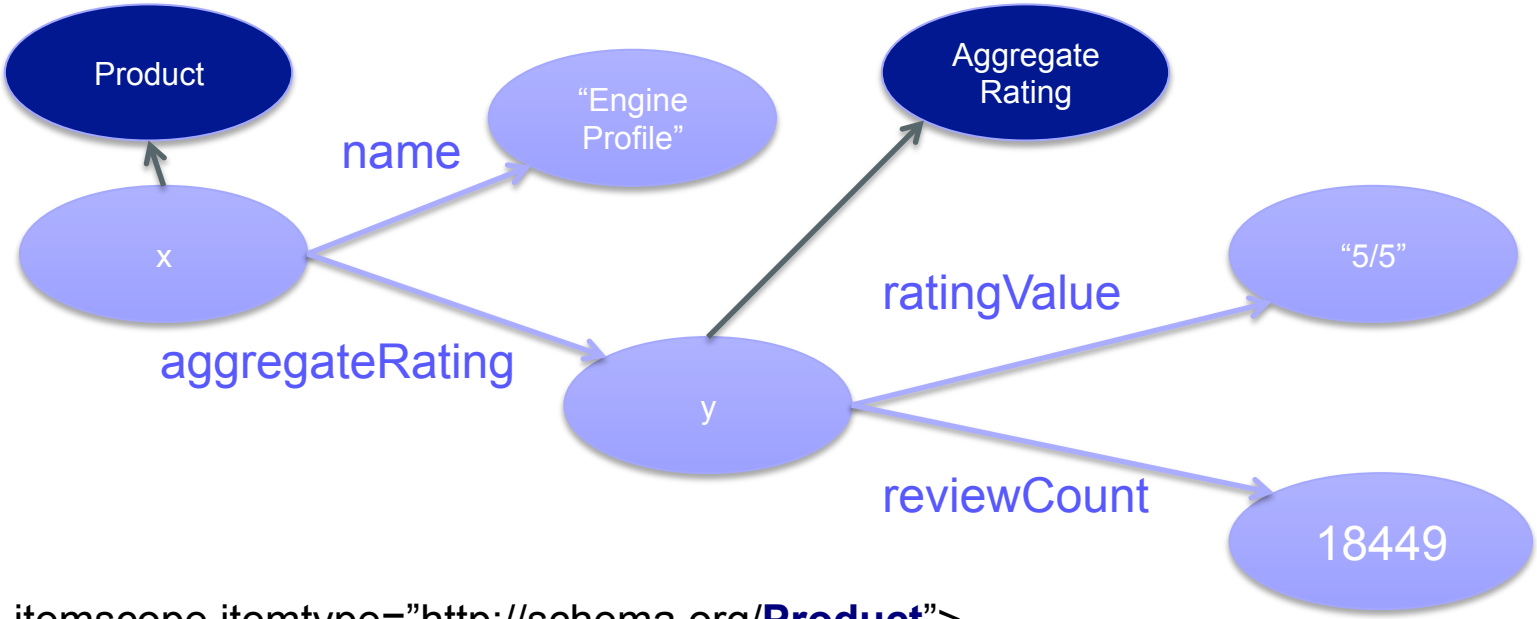
```
<span itemscope itemtype="http://schema.org/Product">
<span itemprop="name"><strong>Engine Profiles</strong></span>
<span itemprop="aggregateRating" itemscope
        itemtype="http://schema.org/AggregateRating">
rated <span itemprop="ratingValue">5</span> / 5<br />
based on <span itemprop="reviewCount">18449</span></span>
```

→

Engine Profiles
rated **5 / 5**
based on **18449**

http://www.cpcstrategy.com/blog/2012/04/google-rich-snippets-schema-org/

Rich Snippets in Search Results

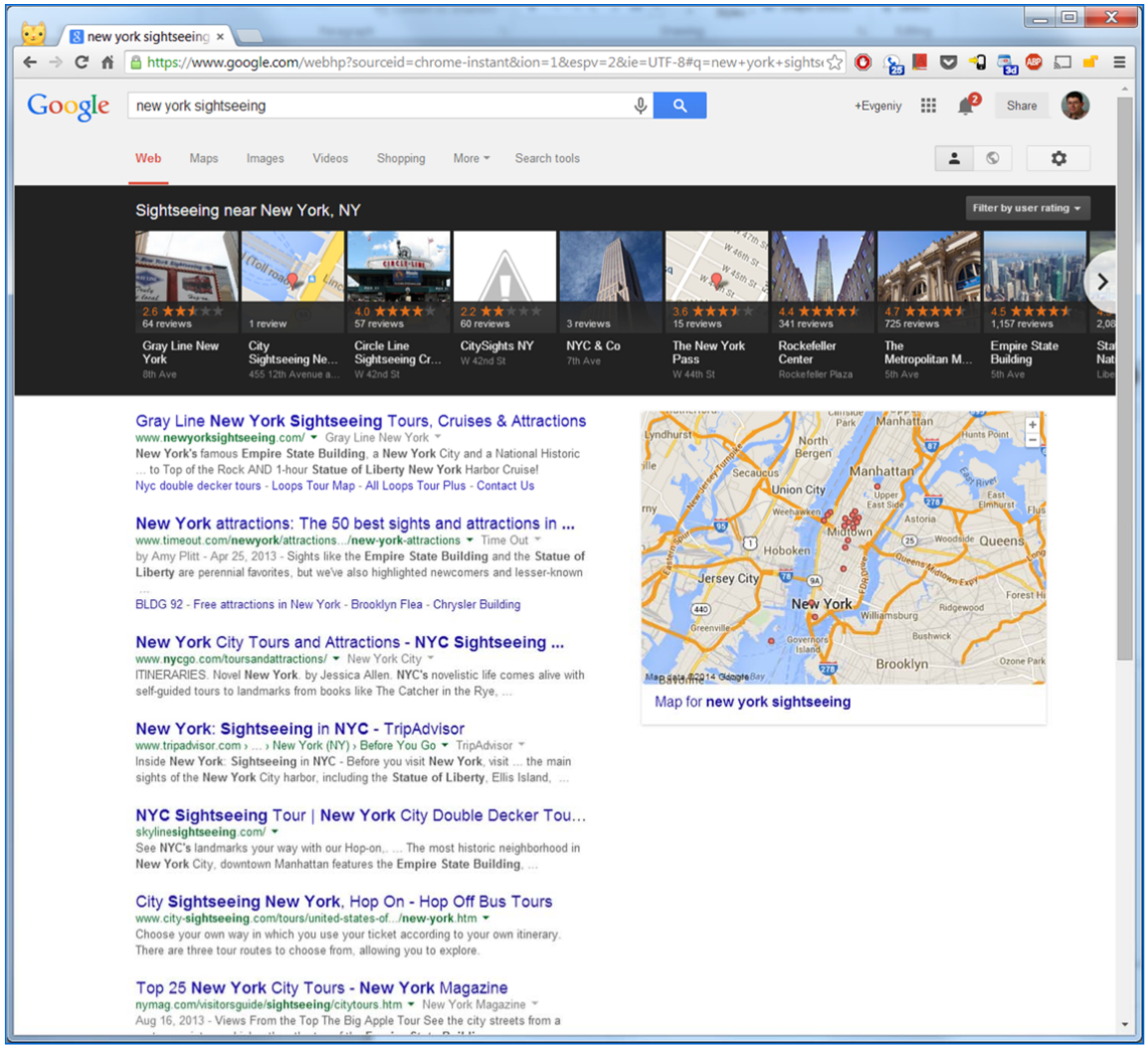



```
<span itemscope itemtype="http://schema.org/Product">
<span itemprop="name"><strong>Engine Profiles</strong></span>
<span itemprop="aggregateRating" itemscope
        itemtype="http://schema.org/AggregateRating">
rated <span itemprop="ratingValue">5</span> / 5<br />
based on <span itemprop="reviewCount">18449</span></span>
```

Engine Profiles
rated **5 / 5**
based on **18449**

http://www.cpcstrategy.com/blog/2012/04/google-rich-snippets-schema-org/

Ricerche Esplorative

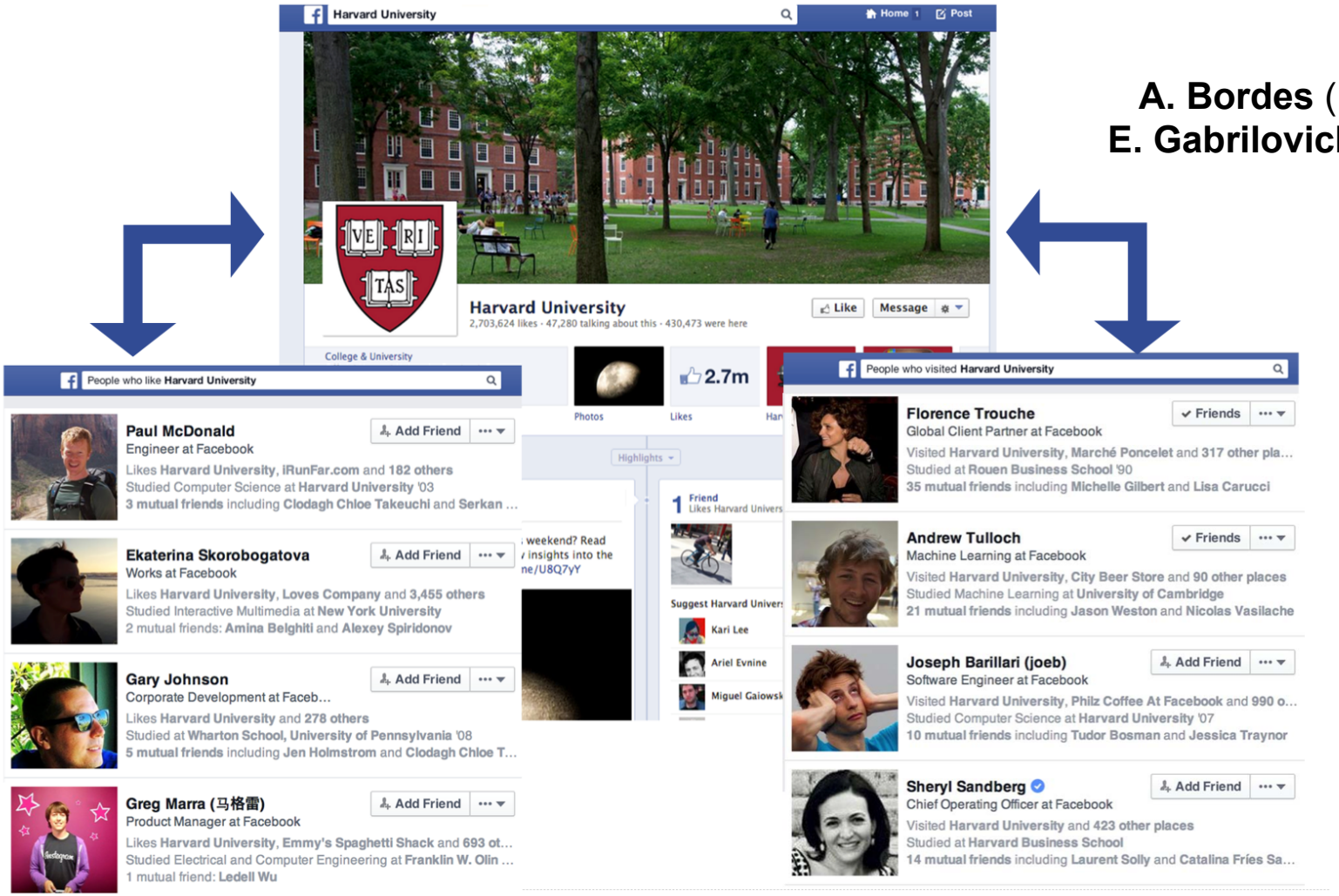


Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Connettere Persone e Luoghi

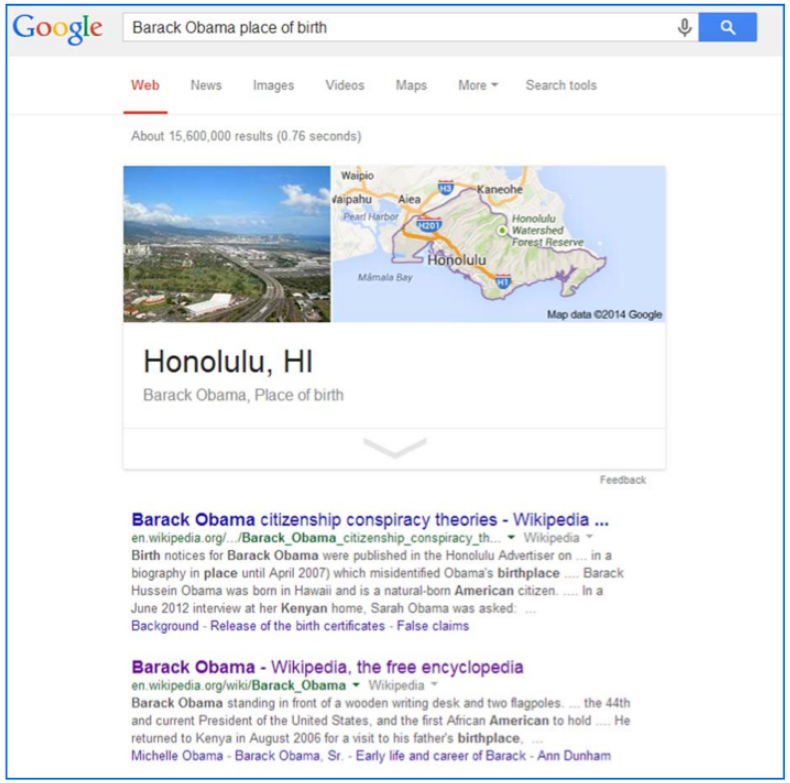
Slide by
A. Bordes (Facebook)
E. Gabrilovich (Google)



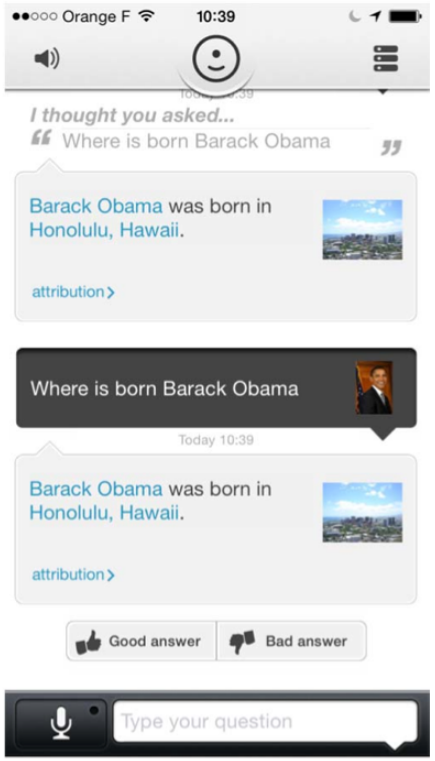
source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Question Answering: risposte dirette a domande espresse in linguaggio naturale

Slide by A. Bordes (Facebook) & **E. Gabrilovich** (Google)



Google



EVI

(Amazon)

Siri

(Apple)

Online Marketing: Digital Data Trends

Search, Social, & Content Fusion (by Andy Betts, April 10, 2014)

- "One of the biggest challenges marketers face is understanding what to do with mass data at hand. This was a key theme from the recent ClickZ Live New York."
- "Identifying what content engages with your digital audience is the key to driving efficient, and scalable, integrated marketing campaigns. The panel was in agreement that content is the catalyst that should be placed at the center of all digital campaigns."
- key trends that focus on breaking down big data into small pieces to help marketers understand how content, search, and social (and subsequent measurement) are morphing together.

Online Marketing: Digital Data Trends

Search, Social, & Content Fusion (by Andy Betts, April 10, 2014)

- "top factors that impact conversion and how best to optimize campaigns around data."

Complex Questions & Decision Making

Complex Questions over Social Networks

Structured search within the graph

source: http://www.cs.technion.ac.il/~gabr/publications/papers/KDD14-T2-Bordes-Gabrilovich.pdf

Questions over Web information

Qual è il cantante rock di maggior successo in questo momento in UK?

Questions over Web information

Quali sono gli ultimi 3 eventi che hanno determinato l'aumento di ascolto di canali televisivi stranieri?

Questions over Web information

Qual è il cantante rock di maggior successo in questo momento in UK?

Data Analysis Scenario: School Closings in Chicago

- Utilization Crisis (May 2013)
  - 403,000 students enrolled in Chicago Public Schools (CPS)
  - 511,000 available seats
- Proposed solution
  - 47 schools closed (affecting 12,000 students)
  - Displaced students assigned to a *welcoming school* that is rated higher than the closed school
  - Safe passage routes implemented
- Problems
  - Potential to destabilize fragile neighborhoods (South and West sides) already grappling with high poverty, crime, and unemployment
  - Disproportionately affect the most vulnerable students

Welcoming School Choice

Current Observations

- 93% of displaced students attended schools with higher performance ratings than the closed schools
- One-quarter of students attended schools that were lower performance than their designated welcoming schools
- Only a small portion of displaced students from these past closings attended substantially higher-performing schools

Evaluation of Educational Outcomes

- Displaced students have (on average) better academic outcomes
- Displaced students who attend substantially higher-performing schools have better academic outcomes
- Displaced students have (on average) the same academic outcomes
- Displaced students have (on average) worse academic outcomes

Analytic Method

(2009)

- Has access to individual student data
- Uses two methods
  - Longitudinal survey (temporal), by matching expected student performance trajectory with actual trajectory after displacement
  - Performance comparison, by matching performances of displaced students with that of students from “similar” schools.
- Links to poverty data from Census
- Ignores other Census variables, and economical, societal, and transportation data
- Concludes: Displaced students who attended substantially higher-performing schools improve their academic outcomes. For others no change after being displaced (some change prior).

M. de la Torre and Julia Gwynne. When Schools Close. Effects on Displaced Students in Chicago Public Schools. Consortium on Chicago School Research at the University of Chicago Education Institute, 2009.

Conclusions on Scenarios

- Use of many datasets
- Use of specific domain analytic methods
- However, domain scientists lack
  - Automatic methods for data extraction/ cleaning/integration
  - Support for interacting with data and analytics
- In general, complex real-world problems are difficult to tackle

School Closing Datasets

- Academic: CPS schools (S), School Quality Rating (SQR), demographic variables of students (D), e.g., gender, race, age, attendance/preference boundaries
- Community Profile: Chicago census data (U), e.g., population, income, poverty, infrastructure (I), e.g., roads, transportation.
- Safety: crime statistics (C), police districts (P), sex offenders and their blocks (SO), safety passages (SP).
- Background information: geospatial ontologies (location ontology, transportation ontology), time ontology, academic institution ontology

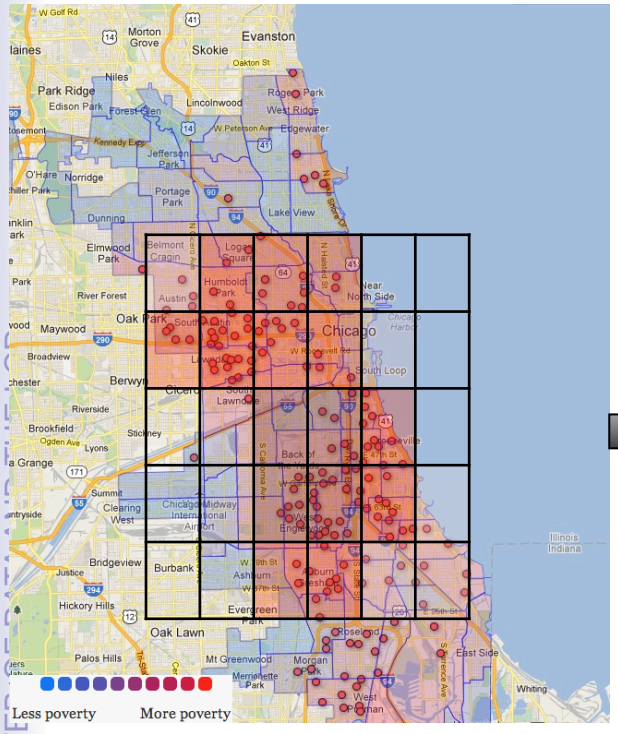
Information Processing Methods

- Visualization (V)
  - The answer is not given in terms of a quantitative analysis but information is displayed so as to provide insight into the addressed question
- Query Answering (QA)
  - The answer can be computed by queries over the integrated dataset (e.g., group by + count), including geospatial queries (points, regions, trajectories)
- Descriptive Statistical Analysis (DSA)
  - The answer is given by descriptive statistical analysis, which may include methods such as correlation analysis, multivariate analysis, spatial correlation, and so on.
- Predictive Statistical Analysis (PSA)
  - The answer is given by making predictions about unknown variables of a problem, e.g., using machine learning methods

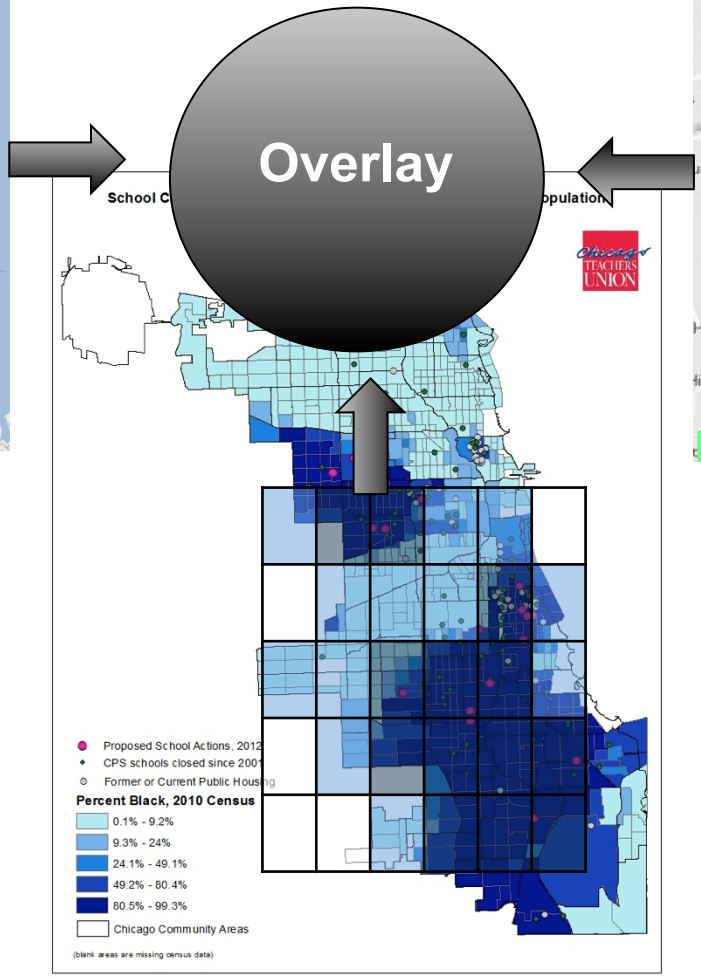
Process Complexity

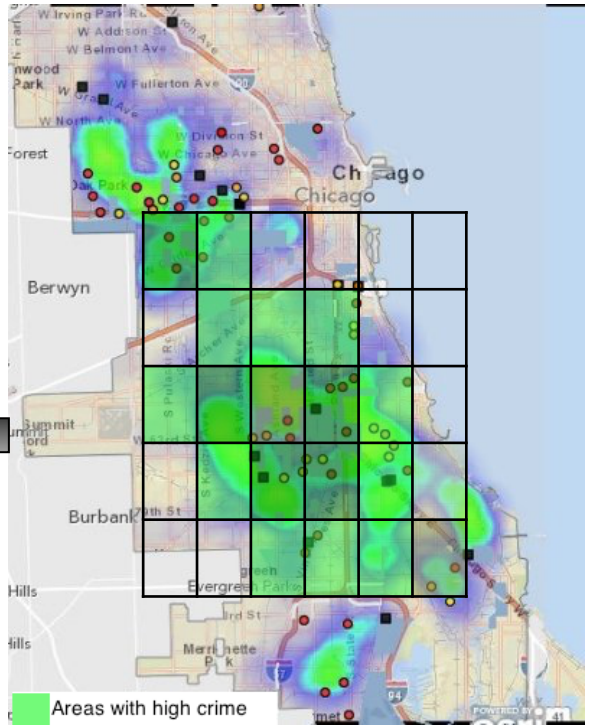
Question	Data	Geospatial Integration	Information Processing	Difficulty
Show school performance, poverty and crime rate on one map	S,SQR,C	Map Overlaying	V	
What are the characteristics of displaced students and of underperforming schools?	S,SQR	-	QA	
Is there a correlation between imbalanced school demography and school performance?	S,D	-	DSA	
Is there a correlation between poverty and school performance?	S,SQR, C	Transformation + Join	DSA	
Is there a correlation between violent crime rate and performance of school quality indicators?	S,SQR,C	Transformation + Join	DSA	
Do charter schools perform better than district schools under similar vulnerability conditions?	S,SQR, D, U	Transformation + Join	PSA	
Are safe passages close to crime areas and sex offenders?	S,SO,SP	Transformation + Join	QA, PSA	
Is there optimal student relocation in terms of safe passage coverage, performance, and demographic balance of the hosting school?	S,SQR, D, I	Transformation + Join	PSA	

Visualization



Underutilized schools in the CPS system based on poverty in the neighborhood



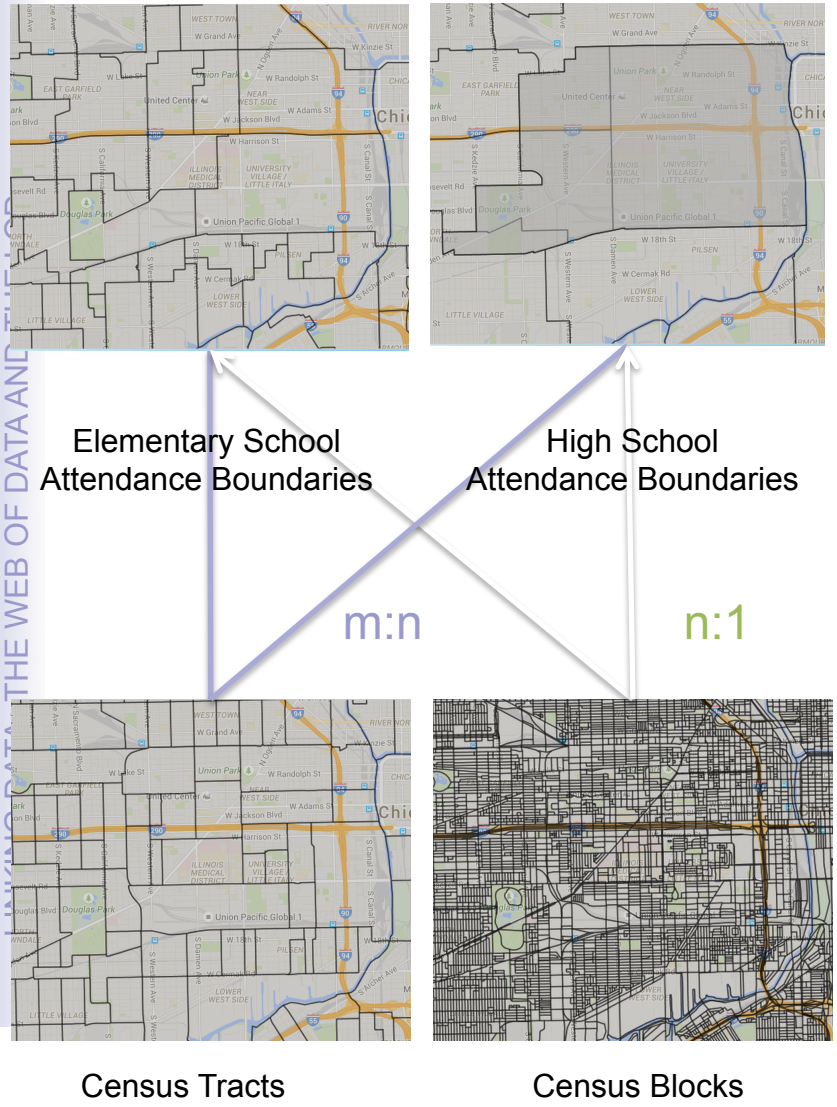


School closings in neighborhoods with high crime rate

LINKING DATA: THE WE

Geospatial Integration with Transformation

"Is there a correlation between poverty and school performance?"



1. Transform format (to obtain harmonized Coordinate Reference System (CRS) or geometry serialization)

   E.g., Convert every point in the geometries of School Boundaries to WGS 84 CRS

2. Compute geometric similarity for 1:1 mappings (e.g., with Hausdorff Distance) or discover spatial relations for n:1/1:n/m:n mappings using RCC8 or DE-9IM (e.g., contains, within, intersects)

   E.g., compute containment between blocks and attendance boundaries

Mapping specification Spatial matching and execution

LINKING DATA THE WEB OF DATA AND THE

Open Data

Il valore del dato integrato e integrabile

Open Data: Tabelle e Spreadshit

Query?

Integrazione?

Agenzie Turistiche Umbria

	A	B	C	E	F	G	H	I
1	tipo	lingua	id contenuto	denominazione	indirizzi	telefoni	fax	email
2	Agenzia viaggio	it_IT	35362	YELLOW RABBIT TRAVEL	[06012 | CITTA' DI CASTELLO | PG | Via E. Kant 29/G |]	075/8511766	075/8511486	turismo@autonoleggivaltiberina.it
3	Agenzia viaggio	it_IT	35372	HERMITAGE TRAVEL FILIALE	[05015 | FABRO | 55011 | TR | Via del Casermone,54 |] | [8	0763/832149 | 081/8706966	0763/832149 | 08	contabilita@hermitagetravelsrl.it | booking@h
4	Agenzia viaggio	it_IT	35382	ADMIRE	[06122 | PERUGIA | 54039 | PG | Via del Sole, 6 |]	075/5730808 | 3356247240	075/5730303	admire.italy@libero.it
5	Agenzia viaggio	it_IT	35392	SPAZIO E TEMPO LIBERO VIAGGI	[05100 | TERNI | 55032 | TR | Via I Maggio, 13 |]	0744/59679	0744/406943	info@spaziotempolibero.it
6	Agenzia viaggio	it_IT	35402	STRAVAGARIO VIAGGI E TURISMO	[05018 | ORVIETO | 55023 | TR | C. Cavour, 97 |]	0763/340890	0763/343422	stravagario@tiscali.it
7	Agenzia viaggio	it_IT	35412	INVIAGGI	[05100 | TERNI | 55032 | TR | V.le della Stazione 27 |]	0744/803711	0744/817404	contabilita@inviaggi.it | info@inviaggi.it
8	Agenzia viaggio	it_IT	35422	AMBASSADOR TRAVEL SERVICE FILIALE	[05021 | ACQUASPARTA | 55001 | TR | Via Roma, 9 |]	0744/930831 | 0541/605509	0744/930831 | 05	ambassador@ambassadortravel.it
9	Agenzia viaggio	it_IT	35432	MELPIDEA VIAGGI	[05015 | FABRO | 55011 | TR | P.le C.Levi, 17 |]	0761/556566 | 3493818827	0761/1760298	melpidea@sensidiviaggio.it
10	Agenzia viaggio	it_IT	34522	IMPAEKT	[06062 | CITTA' DELLA PIEVE | PG | Via G. Marconi, 1 |]	0578/298446 | 3461380007		hariet@impaekt.com | info@impaekt.com
11	Agenzia viaggio	it_IT	34532	LOCCI VIAGGI	[06049 | SPOLETO | 54051 | PG | Piazza Garibaldi, 2 |]	0743/54415	0743/202677	fabio@locciviaggi.it
12	Agenzia viaggio	it_IT	34542	ITALSPRING	[06121 | PERUGIA | 54039 | PG | Via Baglioni, 12 |]	075/5731732	075/5090929	info@italspring.it
13	Agenzia viaggio	it_IT	34552	VENTO DI VACANZE	[05022 | AMELIA | 55004 | TR | Via Europa, 48 |]	0744/978673	0744/060064	info@ventodivacanze.it
14	Agenzia viaggio	it_IT	34562	ANTHOS VIAGGI	[06126 | PERUGIA | 54039 | PG | Via della Pallotta, 18/c |]	075/30465	075/36194	anthos@anthosviaggi.it
15	Agenzia viaggio	it_IT	34572	I VIAGGI DI MACOS	[06121 | PERUGIA | 54039 | PG | Via dei Filosofi, 25/C |]	075/36396	075/34015	macosviaggi@tiscali.it
16	Agenzia viaggio	it_IT	34582	MENIGATTI	[06132 | PERUGIA | 54039 | PG | Via del Pettirosso, 4 |]	075/5149707	075/774611	amministrazione@menigattiviaggi.it

Significato dei dati?

Eventi

	A	B	C	D	E	F	G	H	I	J	K	L
1	tipo	lingua	id contenuto	id contenu	url risorsa	titolo	descrizione	data inizio	ora inizio	data fine	ora fine	keywords
2	Evento	it_IT	2916883	90470	http://www.umbriatourism.it/-/it-fits	IT FITS	Il 25 settembre, ad Assisi, la terza edizione del Forum Italiano sul Turismo e la SostenibilitÝ†.	25/09/15		26/09/15		
3	Evento	it_IT	2032108	90422	http://www.umbriatourism.it/-/il-lusso-del-s	Il Lusso del So	Fino al 31 agosto all'Antiquarium di Corciano la mostra dedicata ai letti etruschi in bronzo rinvenuti	21/08/15		31/08/16		
4	Evento	it_IT	2032245		http://www.umbriatourism.it/-/ecobike	Ecobike	Fino a ottobre escursioni e visite guidate attraverso l'Umbria in sella a una bicicletta elettrica.	21/08/15		31/10/15		

	M	P	Q	T	U	V	W	X	Y	Z	AA	AB
1	titolo testo	caregorie evento	immagini evento	immagine copertina evento	testo alternativo immagi	latitudine	longitudine	comune	codice IS	via o piazza	localitÝ†	cap
2	Forum Italiano sul Turismo e la SostenibilitÝ†	Mostre e rassegne | Assisi	http://www.umbriatourism	http://www.umbriatourism.ithttp:	IT FITS	430.707.017	12.619.596.600.000.000	Assisi	54001			
3	I letti in bronzo di Strozzacapponi	Mostre e rassegne | Corciano	http://www.umbriatourism	http://www.umbriatourism.ithttp:	Il lusso del Sonno	431.237.872	12.289.259.600.000.000	Corciano	54015			
4	Alla scoperta dell'Umbria con la bici elettrica	Eventi sportivi	http://www.umbriatourism	http://www.umbriatourism.ithttp:	Ecobike							

Open Data: Tabelle e Spreadsheet

- Query & Integrazione
  - CSV non interrogabili cosi come sono
  - Esplorazione e manipolazione richiede applicazioni
    - Locali / Non web friendly
  - Integrazione manuale (vedi Open Refine)
- Semantica
  - Ambiguità: tipo (evento) vs tipo (agenzia)
  - Tipizzazione debole: e.g., codici
  - Vocabolari locali e non interpretabili da terze parti

REMARK

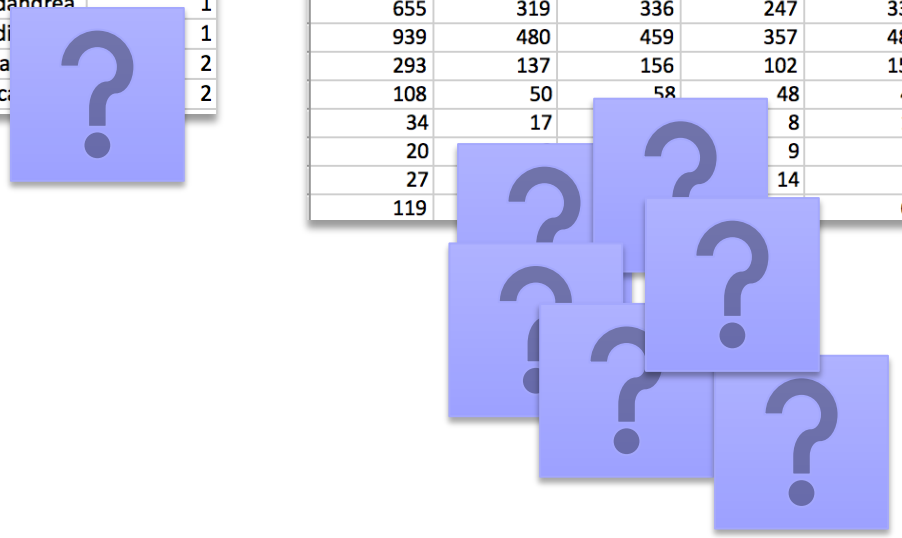
La finalità degli Open Data è proprio quella di poter essere usati da *terze parti*, tipicamente con conoscenza approfondita dell'organizzazione che ha prodotto i dati

ISTAT Open Data: CSV file

COMUNE	PROCOM	LOC2011	CODLOC	LOCALITA	TIPOLOC
Assisi	54001	5400110001	10001	Armenzano	1
Assisi	54001	5400110002	10002	Assisi	1
Assisi	54001	5400110003	10003	Capodacqua	1
Assisi	54001	5400110004	10004	Case Nuove	1
Assisi	54001	5400110005	10005	Castelnuovo	1
Assisi	54001	5400110006	10006	Palazzo	1
Assisi	54001	5400110007	10007	Petrignano	1
Assisi	54001	5400110008	10008	Pianello	1
Assisi	54001	5400110009	10009	Rivotorto	1
Assisi	54001	5400110010	10010	Santa Maria	1
Assisi	54001	5400110011	10011	San Vitale	1
Assisi	54001	5400110012	10012	Sterpeto	1
Assisi	54001	5400110013	10013	Torchiagina	1
Assisi	54001	5400110014	10014	Tordandrea	1
Assisi	54001	5400110015	10015	Tordi	1
Assisi	54001	5400120007	20007	Rena	2
Assisi	54001	5400120008	20008	Rocc	2

...

P1	P2	P3	P4	P5	P6	P7
33	21	12	8	16	1	5
3732	1690	2042	1664	1526	77	378
142	68	74	49	69	2	17
634	316	318	237	306	17	62
791	385	406	307	394	11	68
1817	866	951	721	890	41	137
3184	1538	1646	1200	1624	68	229
107	52	55	40	55	1	11
1425	687	738	513	734	27	122
7719	3710	4009	3336	3442	164	597
603	287	316	254	285	9	46
13	8	5	6	3	2	1
655	319	336	247	332	12	60
939	480	459	357	483	17	64
293	137	156	102	151	10	23
108	50	58	48	48	0	10
34	17	19	8	15	0	10
20	9	9	9	7	0	1
27	14	14	14	8	1	4
119	69	69	69	69	2	6

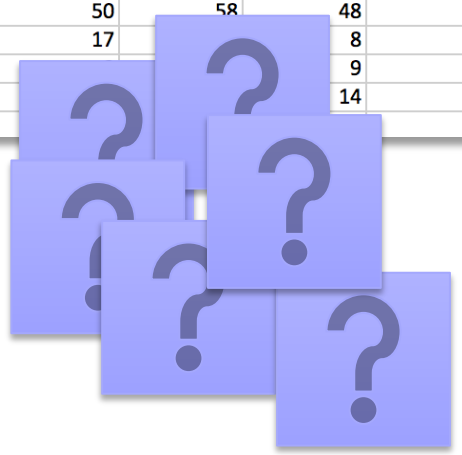


ISTAT Open Data: CSV file

COMUNE	PROCOM	LOC2011	CODLOC	LOCALITA	TIPOLOC
Assisi	54001	5400110001	10001	Armenzano	1
Assisi	54001	5400110002	10002	Assisi	1

Nome Camp	Definizione
CODREG	Codice numerico che identifica univocamente la regione nell'ambi
REGIONE	Denominazione della regione
CODPRO	Codice numerico che identifica univocamente la provincia nell'aml
PROVINCIA	Denominazione della provincia
CODCOM	Codice numerico che identifica univocamente il comune nell'ambi
COMUNE	Denominazione del comune
PROCOM	Codice numerico che identifica univocamente il comune nell'ambi
LOC2011	Codice numerico che identifica univocamente la località 2011 nell'
CODLOC	Codice numerico che identifica la località 2011 nell'ambito del terr
LOCALITA	Denominazione della località 2011
TIPOLOC	Codice numerico che identifica la tipologia della località 2011. Il ca
AMPLOC	Codice numerico che identifica l'ampiezza demografica della locali
CAPOLUOGO	Codice numerico valorizzato a 1 nel caso di centro capoluogo e a 0
ALTITUDINE	Altitudine della località 2011
P1	Popolazione residente - Totale
P2	Popolazione residente - Maschi
P3	Popolazione residente - Femmine
P4	Popolazione residente - Celibi/nubili
P5	Popolazione residente - Coniugati/e (+ separati/e di fatto)
P6	Popolazione residente - Separati/e legalmente
P7	Popolazione residente - Vedovi/e
P8	Popolazione residente - Divorziati/e
P9	Popolazione residente - Maschi celibi
P10	Popolazione residente - Maschi coniugati o separati di fatto

P1	P2	P3	P4	P5	P6	P7
33	21	12	8	16	1	5
	1690	2042	1664	1526	77	378
	68	74	49	69	2	17
	316	318	237	306	17	62
	385	406	307	394	11	68
	866	951	721	890	41	137
	1538	1646	1200	1624	68	229
	52	55	40	55	1	11
	687	738	513	734	27	122
	3710	4009	3336	3442	164	597
	287	316	254	285	9	46
	8	5	6	3	2	1
	319	336	247	332	12	60
	480	459	357	483	17	64
	137	156	102	151	10	23
	50	58	48	48	0	10
	17		8	15	0	10
			9	7	0	1
			14	8	1	4
				69	2	6

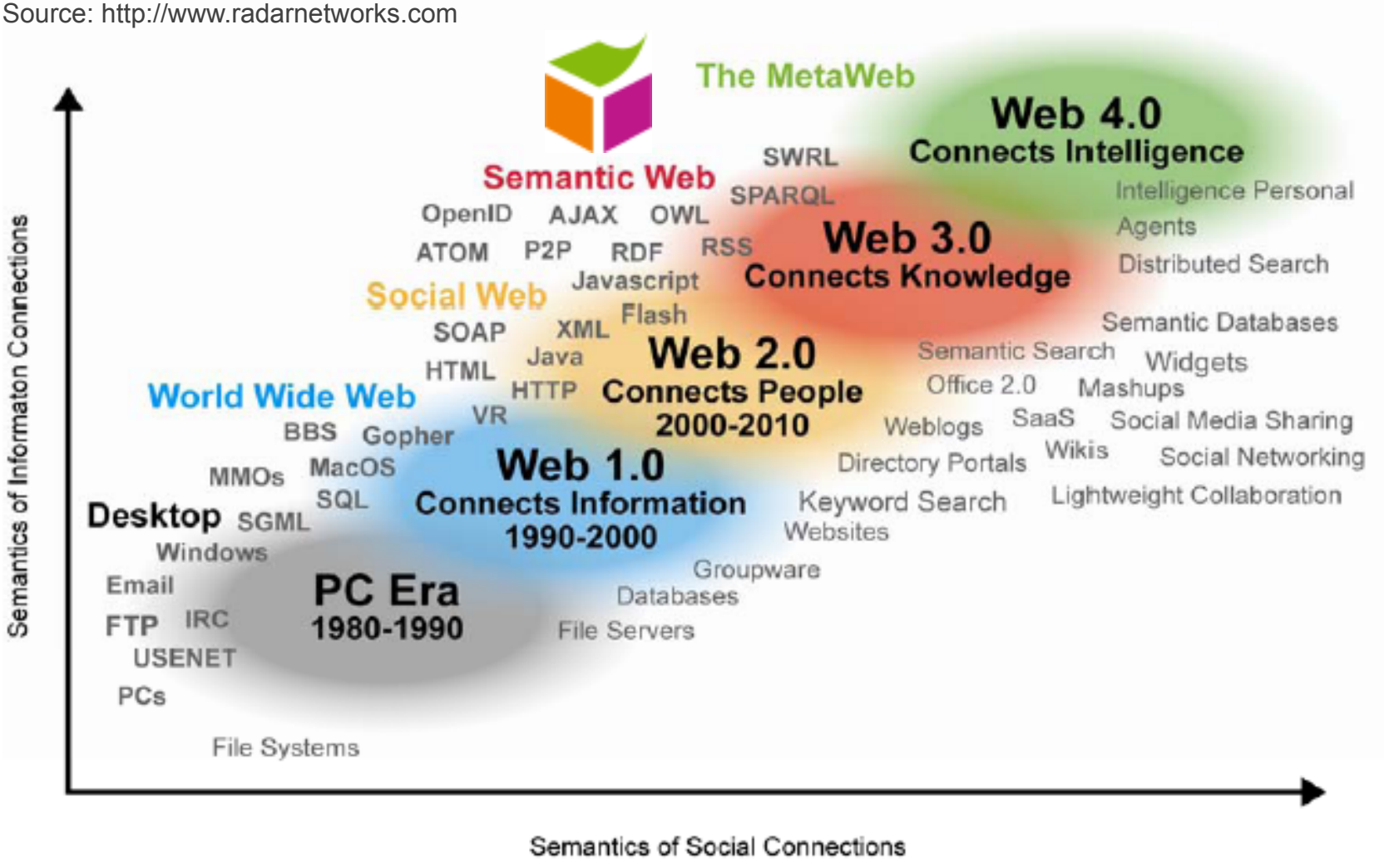


Spiegazione in file separato

The Web: Evolution

EUCLID
EdUcational Curriculum for the usage of LInked Data

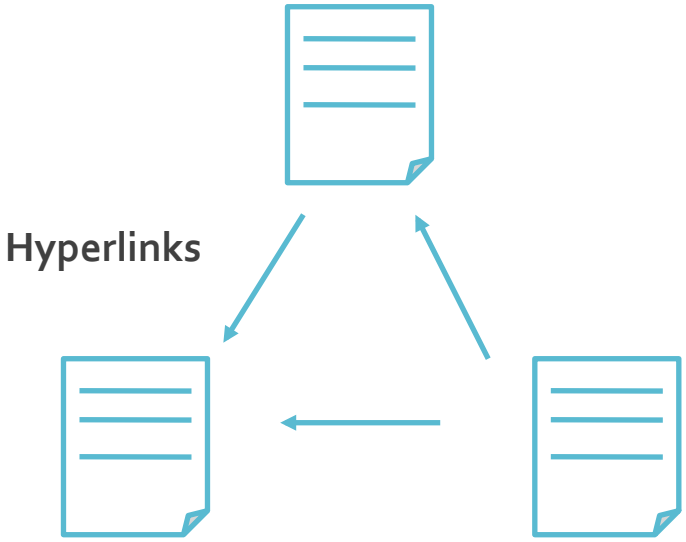
Source: http://www.radarnetworks.com

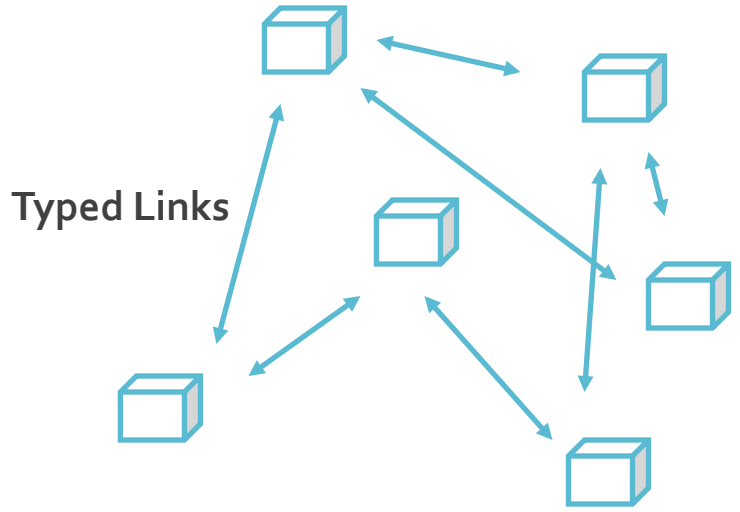


The Web: Evolution

EUCLID
EdUcational Curriculum for the usage of LInked Data

Web of Documents → Web of Data





"Documents"

"Things"

Dai documenti ai dati

- Entità e relazioni
- Entità e attributi
- Processabili da una macchina, supporto a interrogazioni (complesse)
- Integrazione di testi, immagini, video con dati strutturati
- Valore del dato di qualità
- Politiche sui dati di ampio respiro (costruzione di knowledge graph): grandi investimenti per alti ricavi

Linked Data

Linked Data

EUCLID
EdUcational Curriculum for the usage of LInked Data

- Set of best practices for **publishing** data on the Web
- Data from different knowledge domains, self-described, linked and accessible
- Follows 4 simple principles…