

# Semantic Matching: KG Integration and Construction

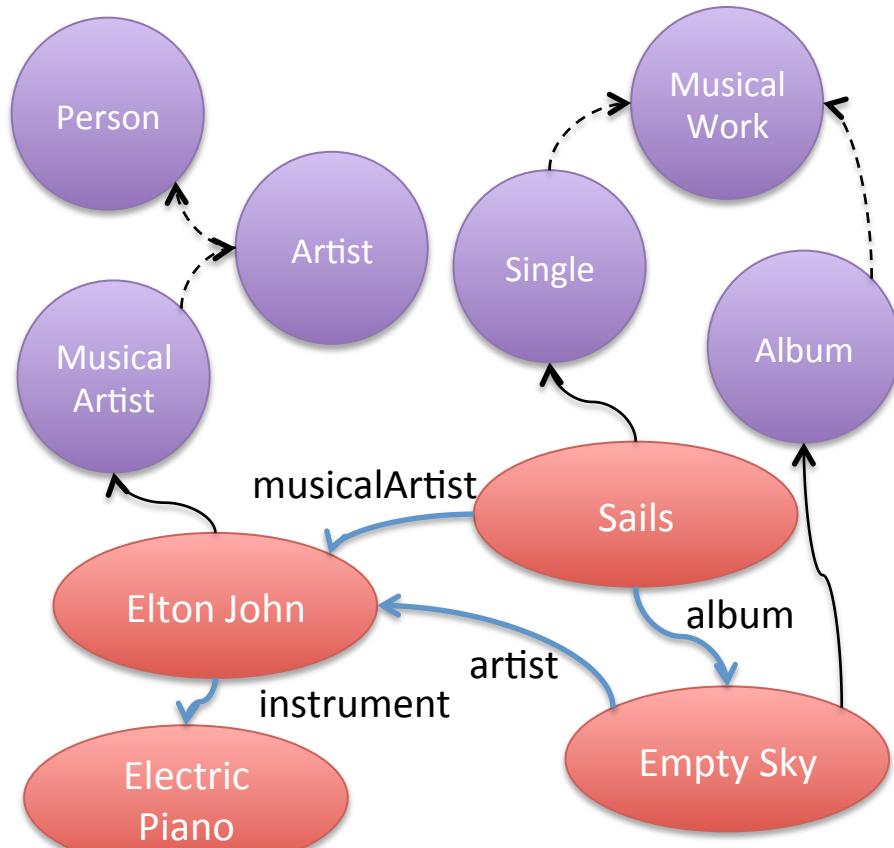
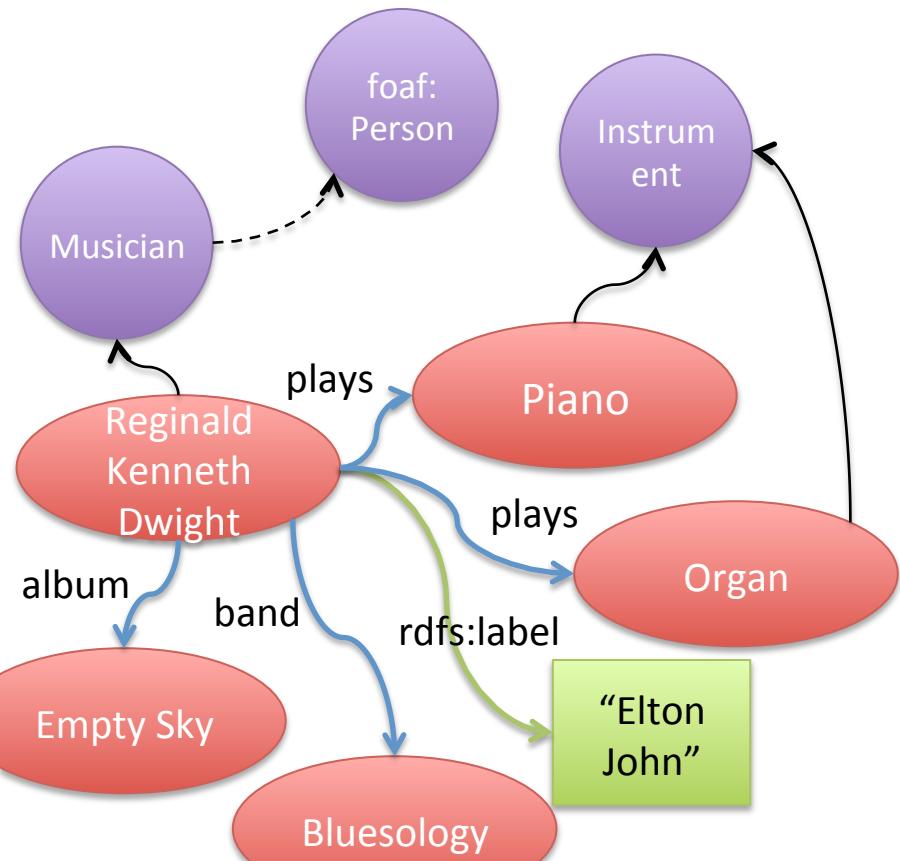
Part I: Introduction

# How to make knowledge grow?

- We-scale KGs:
  - Excellent tutorial from Gabrilovich (Google) and Usunier (Facebook)
    - Constructing and Mining Web-scale Knowledge Graphs
    - Slides at  
[http://www.cs.technion.ac.il/~gabr/publications/papers/  
SIGIR-2016-KG-tutorial.pdf](http://www.cs.technion.ac.il/~gabr/publications/papers/SIGIR-2016-KG-tutorial.pdf)
- Specific (focused) KGs:
  - UCSB group
    - Szekely et al (too many authors ☺): Building and Using a Knowledge Graph to Combat Human Trafficking. International Semantic Web Conference (2) 2015: 205-221
    - Craig A. Knoblock, Pedro A. Szekely: Exploiting Semantics for Big Data Integration. AI Magazine 36(1): 25-38 (2015)

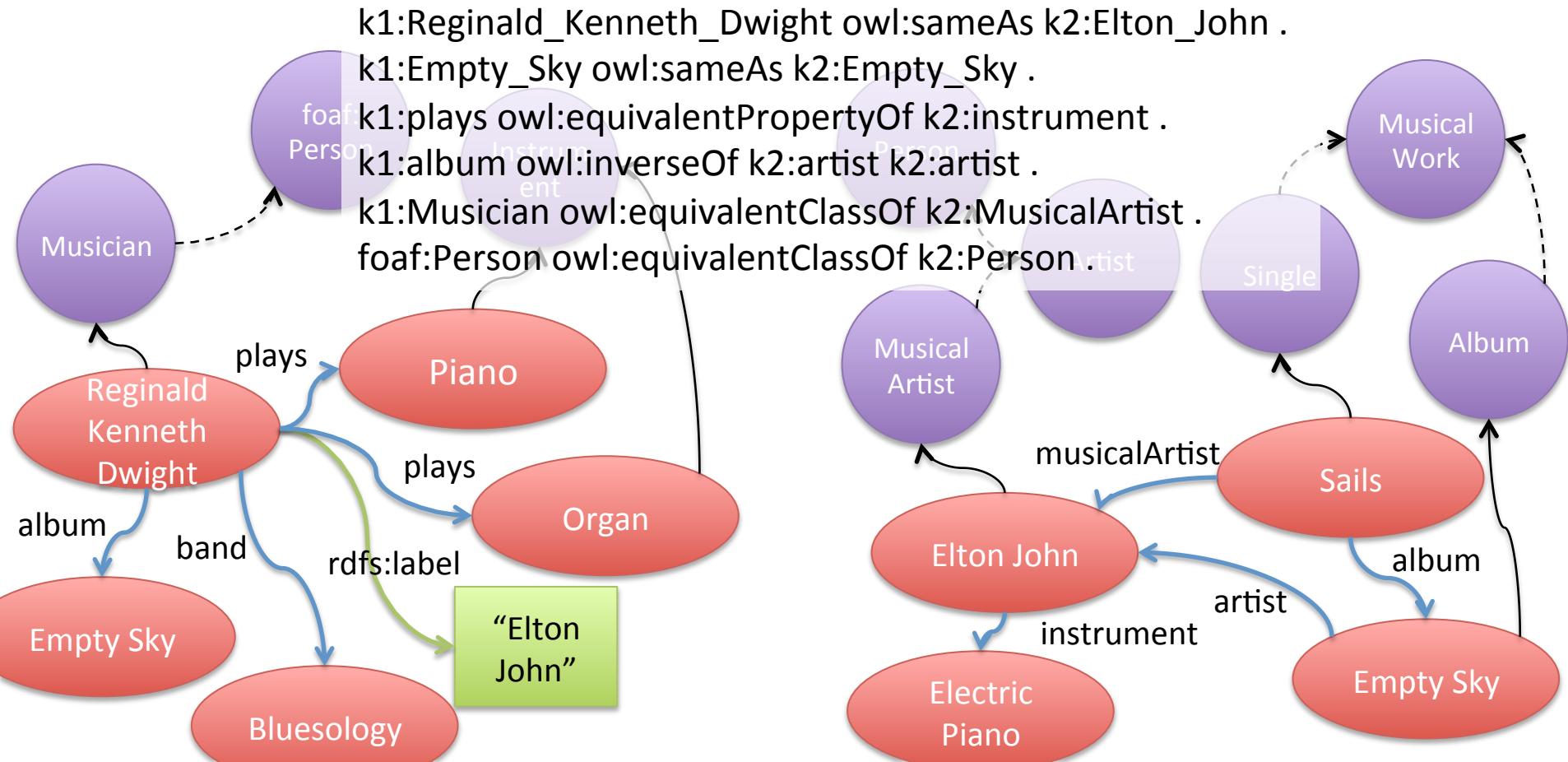
# How to make knowledge grow?

- Let's consider two KGs that we want to integrate



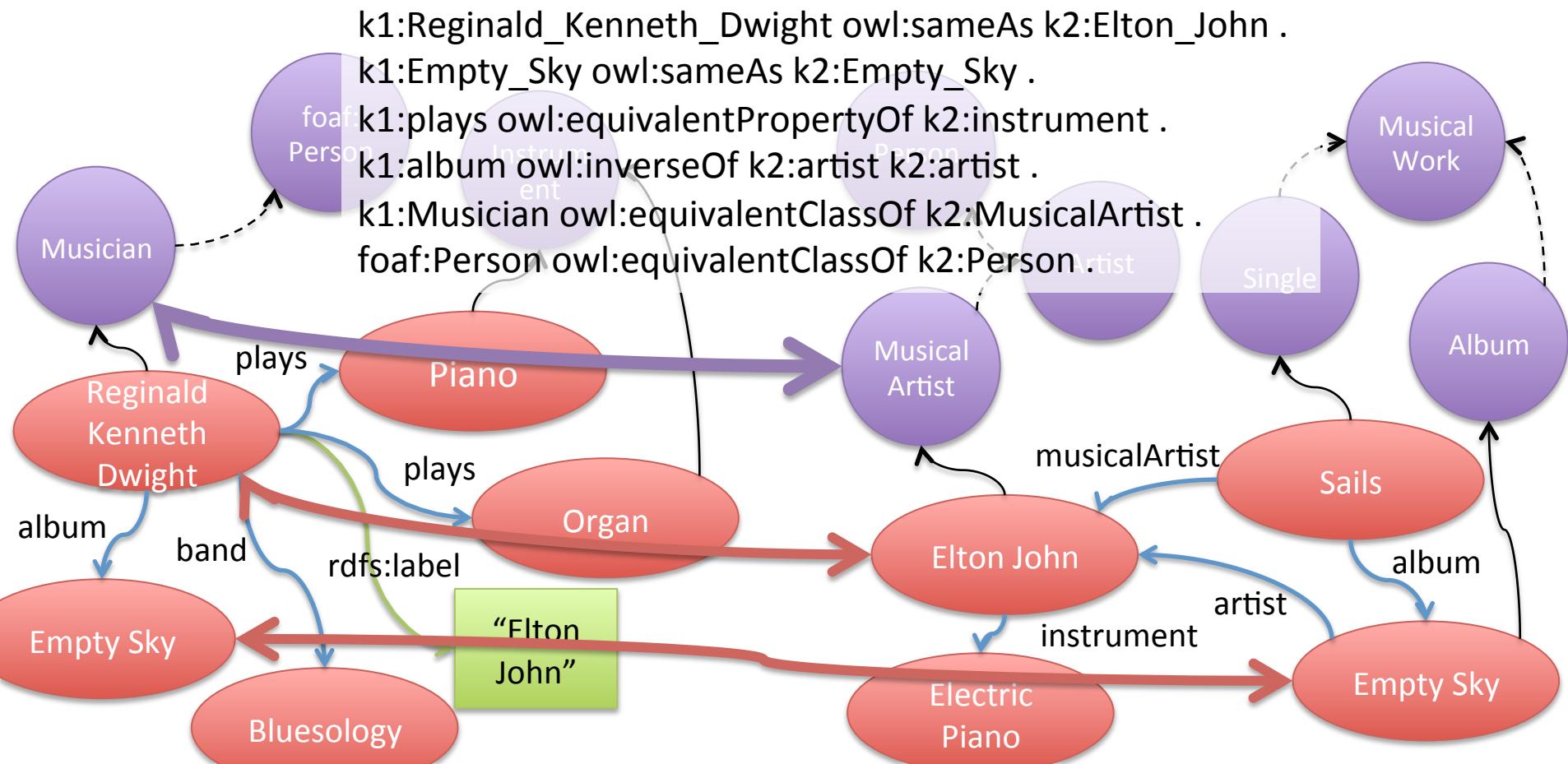
# How to make knowledge grow?

- **Map** two KGs so as to make one connected KG
  - Discover and establish mappings, e.g., (simple) links



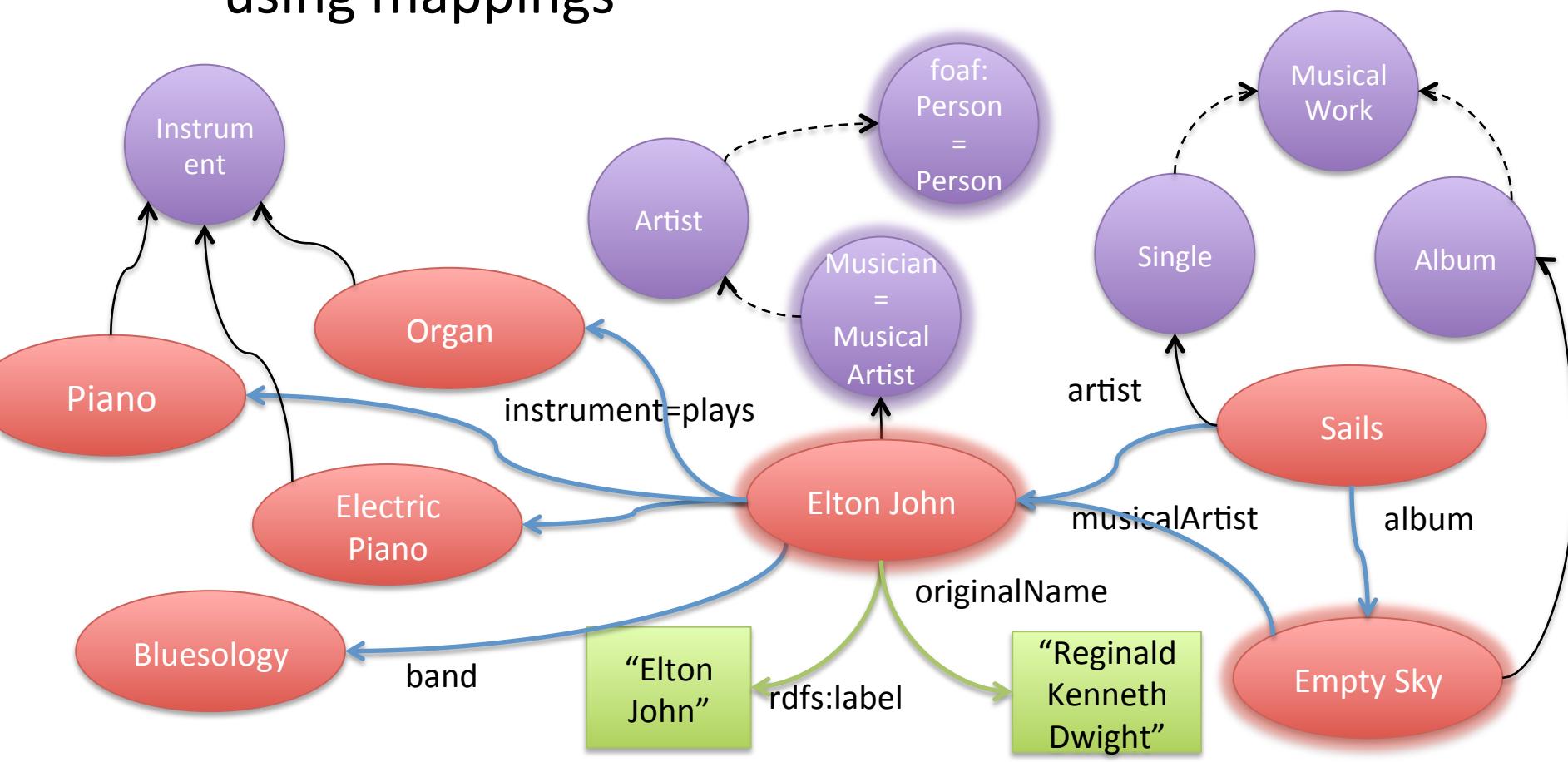
# How to make knowledge grow?

- Map two KGs so as to make one connected KG



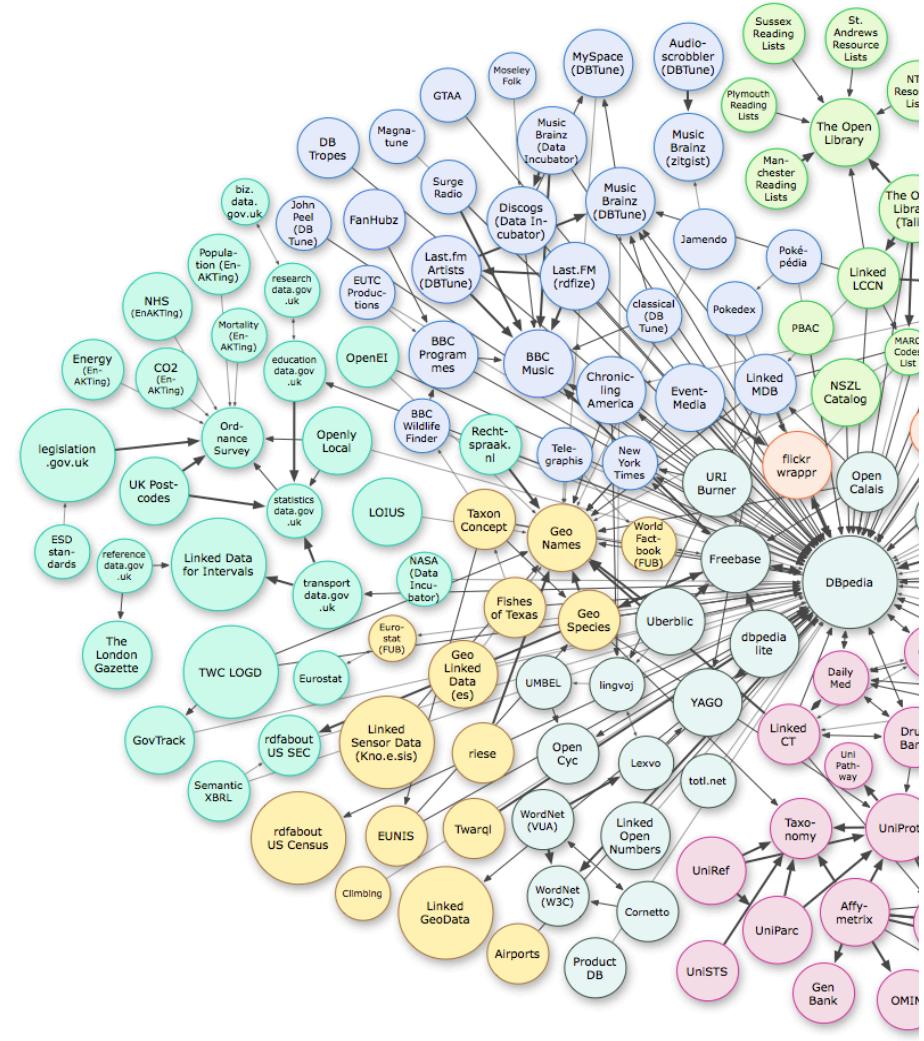
# How to make knowledge grow?

- **Merge** two KGs into one larger KG
  - using mappings



# Linking Data: two KGs

- Schema-level vs instance-level mappings
  - Schema-level: mappings between concepts and properties
    - Ontology matching
  - Instance-level: mappings between entity descriptors in heterogeneous sources
    - Entity co-resolution (sameAs)
    - Link discovery
    - Instance matching
    - ≈ Record Linkage (from RDBs)

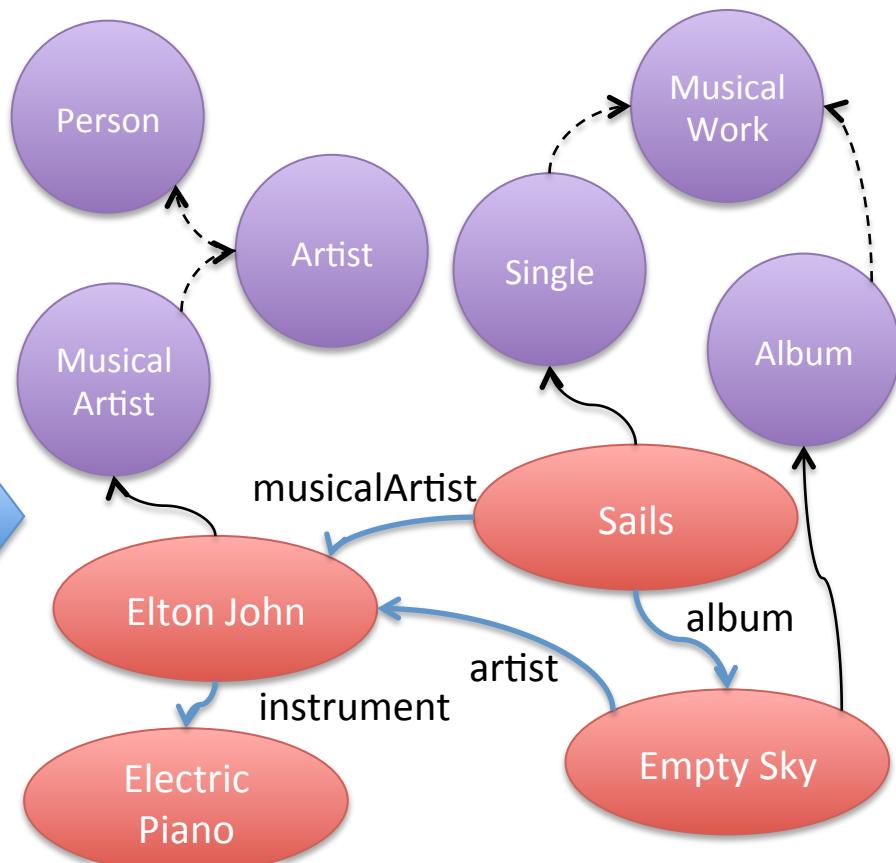


# How to make knowledge grow?

- **Ingest** data from structured or weakly structured sources, e.g., tables

The screenshot shows a search result for 'Elton John – Empty Sky'. It includes a thumbnail image of the album cover, basic metadata (Genre: Rock, Style: Pop Rock, Year: 1969), and a link to 'More Images'. Below this is a table titled 'Versions (55)' with columns: Title (Format), Label, Cat#, Country, and Year. The table lists seven entries for 'Empty Sky' in various formats (LP, Album, Mono, Red, Gat) from different years (1969, 1972) and countries (UK, Germany, Italy, New Zealand). A blue arrow points from this table towards the knowledge graph on the right.

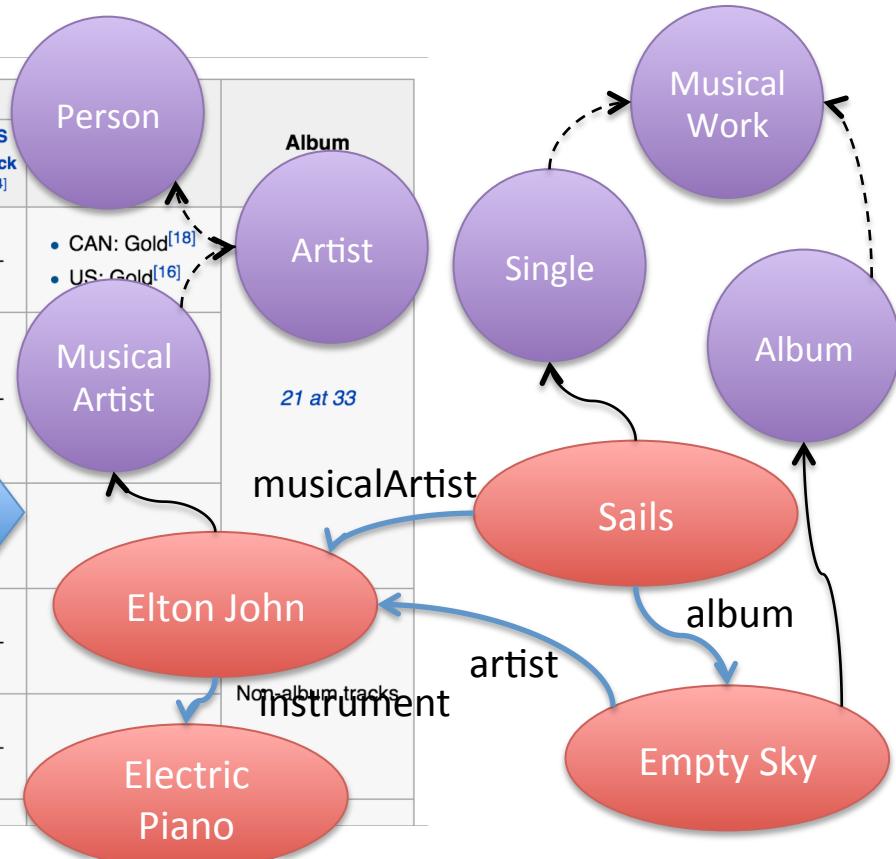
Title (Format)	Label	Cat#	Country	Year
Empty Sky (LP, Album, Gat)	DJM Records (2), DJM Records (2)	DJLPS 403, DJLPS.403	UK	1969
Empty Sky (LP, Album)	DJM Records (2)	0064 211	Germany	1969
Empty Sky (LP, Album, Mon)	DJM Records (2), DJM Records (2)	DJLPS.403, DJLP 403	UK	1969
Empty Sky (LP, Album, Mono)	DJM Records (2)	DJLP 403	UK & Europe	1969
Empty Sky (LP, Album, Red)	DJM Records (2), DJM Records (2)	DJLPS 403, DJLPS.403	UK	1969
Empty Sky (LP, Album)	DJM Records (2), DJM Records (2)	SNIR-DJ 25034, (S/25034/1-2)	Italy	1972
Empty Sky (LP, Album, Gat)	DJM Records (2)	SDJL 934074	New Zealand	1972



# How to make knowledge grow?

- Ingest** data from structured or weakly structured sources, e.g., tables

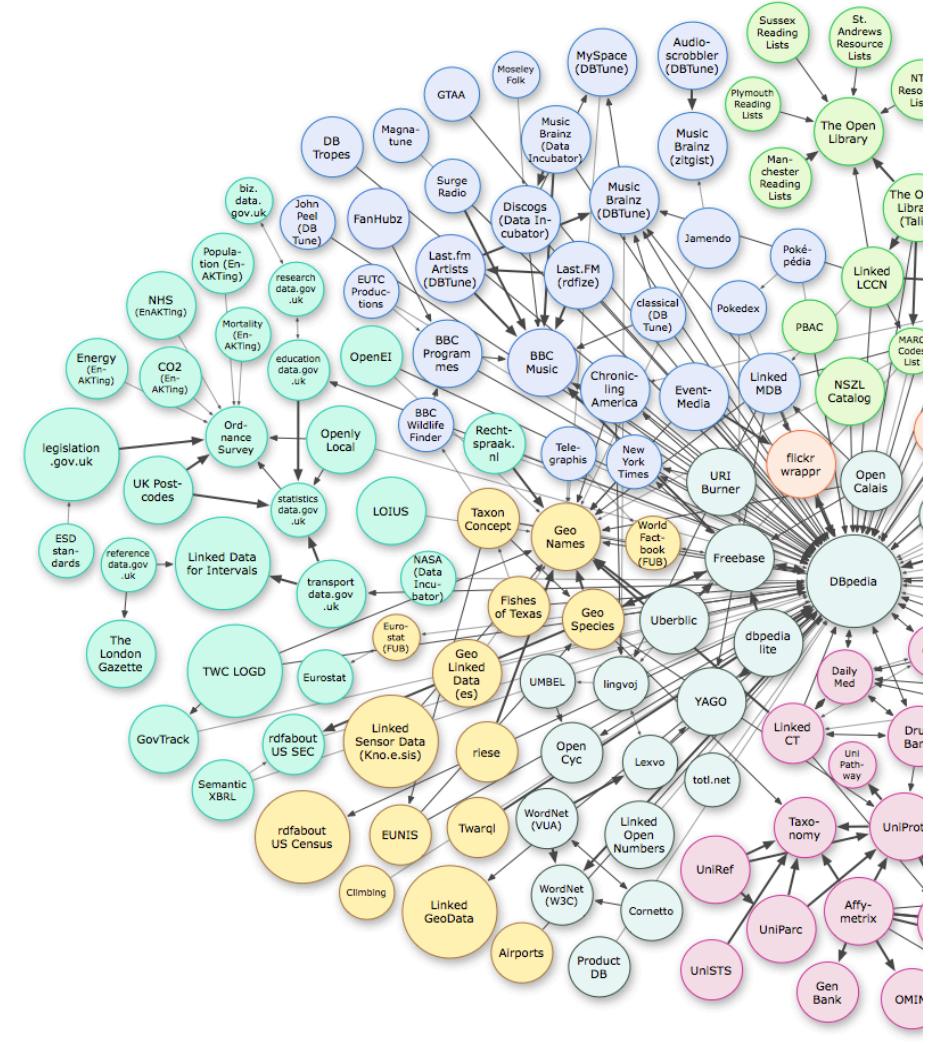
Year	Single	Peak chart positions													
		UK [4]	AUS [5]	BEL [6]	CAN [7]	GER [8]	IRE [9]	NL [15]	NOR [11]	NZ [12]	SWI [13]	US [14]	US AC [14]	US Rock [14]	
1980	"Little Jeannie" b/w "Conquer the Sun" (Non-album track)	33	9	20	1	23	16	—	—	5	4	3	1	—	
	"Sartorial Eloquence (Don't Ya Wanna Play This Game No More?)" b/w "Cartier"/"White Man Danger" (Non-album tracks)	44	91	—	57	—	—	—	—	—	39	45	—	—	
	"Dear God" b/w "Tactics" (Non-album track)	—	82	—	—	—	—	—	—	—	—	—	—	—	
	"Les Aveux" b/w "Donner Pour Donner" (with France Gall)	—	—	—	—	—	—	—	—	—	—	—	—	—	
	"I Saw Her Standing There" (with John Lennon)	40	81	—	—	—	13	—	—	—	—	—	—	—	
	"Nobody Wins"	—	—	—	—	—	—	—	—	—	—	—	—	—	
		—	—	—	—	—	—	—	—	—	—	—	—	—	



# Linking Data: from tables to KGs

- Normalize tables
- Annotate components of the table with components of the KG
  - Schema-level: assign concepts and properties to columns
  - Instance-level: assign entities to values of the table
- Use annotations to define (complex) mappings, e.g., using R2ML
- Execute the mappings and represent the table content as KG

Seamantic Table Annotation

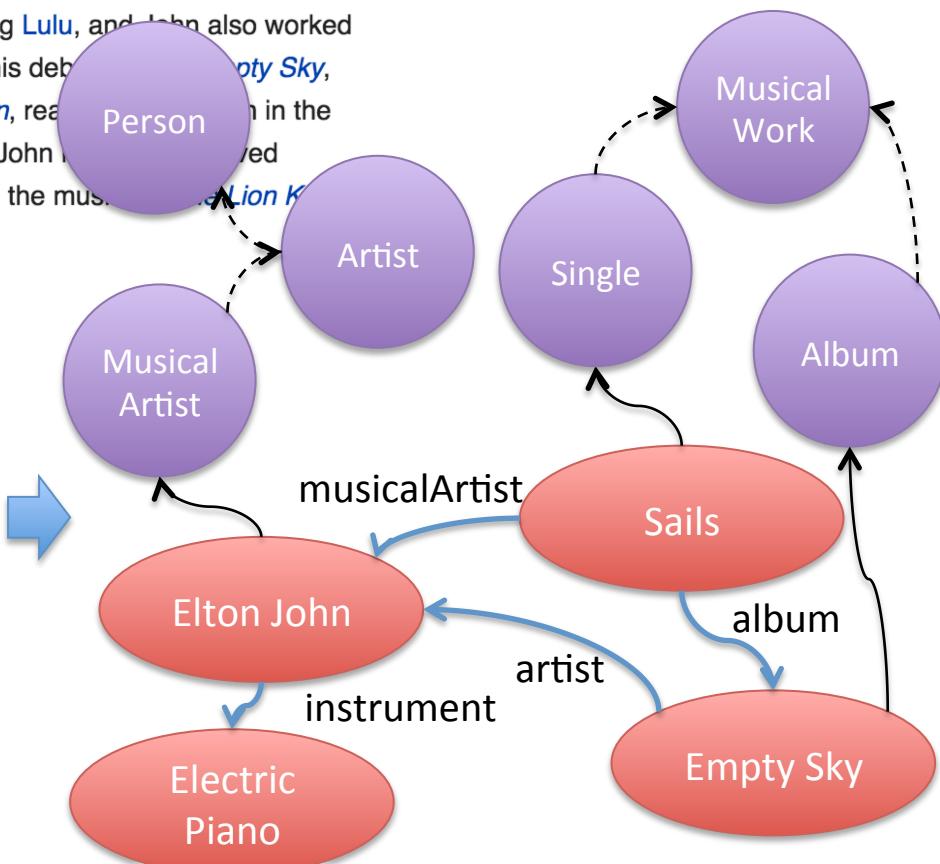


# How to make knowledge grow?

- **Ingest** data from texts (long or short)

Raised in the [Pinner](#) area of London, John learned to play piano at an early age, and by 1962 had formed [Bluesology](#). John met his songwriting partner, Bernie Taupin, in 1967, after they had both answered an advert for songwriters. For two years they wrote songs for other artists, including [Lulu](#), and John also worked as a session musician for artists such as [the Hollies](#) and [the Scaffold](#). In 1969 his debut single, [Empty Sky](#), was released. In 1970 a single, "Your Song", from his second album, [Elton John](#), reached number one in the UK and the US, his first hit single. After decades of commercial chart success, John has enjoyed success in musical theatre, both in the [West End](#) and on [Broadway](#), composing the music for [The Lion King](#) (film and musical), [Aida](#) and [Billy Elliot the Musical](#).

("Elton John") "raised in" "Pinner"  
"Pinner" "area of" "London"  
("Elton John") "learn to play" "piano"  
...



# How to make knowledge grow?

- Ingest data from texts (long or short)



Eminem & Elton John Live at 43rd Grammy Awards in 2001 where they performed "Stan".  
2 Legends of MUSIC!  
#Eminem #EltonJohn

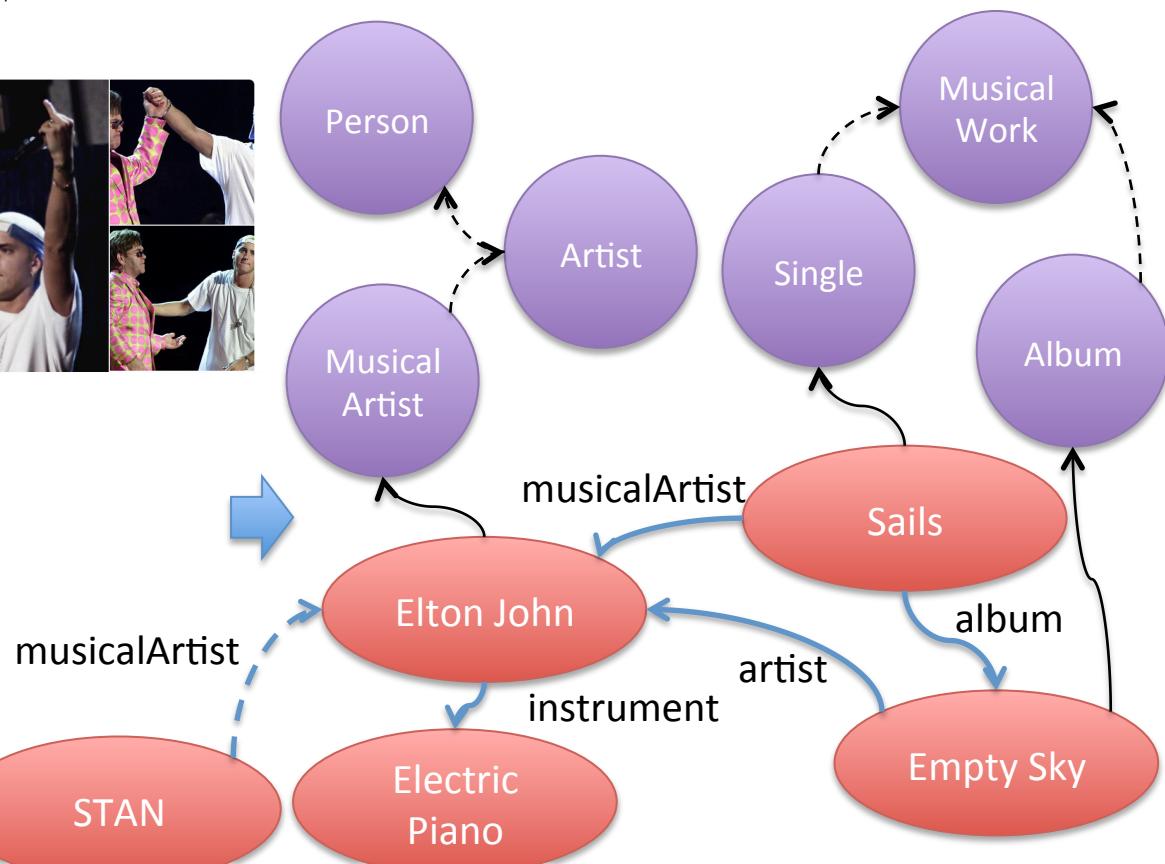
The Voice @NBCTheVoice

Impossible to watch @aarondgibson singing "Rocket Man" by @eltonofficial without the biggest smile. #VoiceTop10

RETWEETS 487 LIKES 1,243

5:40 PM - 28 Nov 2016

"Elton John" "performed" "STAN"  
"Elton John" "Live at" "43<sup>rd</sup> Grammy Awards"  
...  
("Aaron Gibson") "singing" "Rocket Man"  
...



# How to make knowledge grow?

- Ingest data from texts (long or short)



Eminem & Elton John Live at 43rd Grammy Awards in 2001 where they performed "Stan".  
2 Legends of MUSIC!  
#Eminem #EltonJohn

The Voice @NBCTheVoice

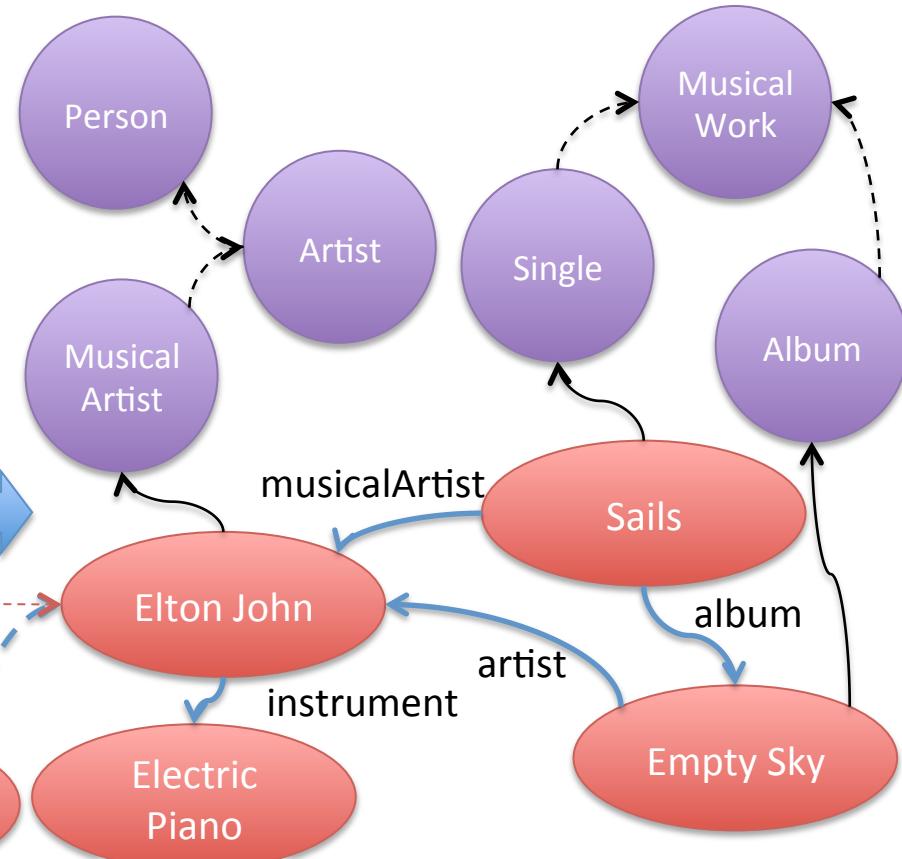
Impossible to watch @aarondgibson singing "Rocket Man" by @eltonofficial without the biggest smile. #VoiceTop10

RETWEETS 487 LIKES 1,243

5:40 PM - 28 Nov 2016  
“Elton John” “performed” “STAN”  
“Elton John” “Live at” “43<sup>rd</sup> Grammy Awards”

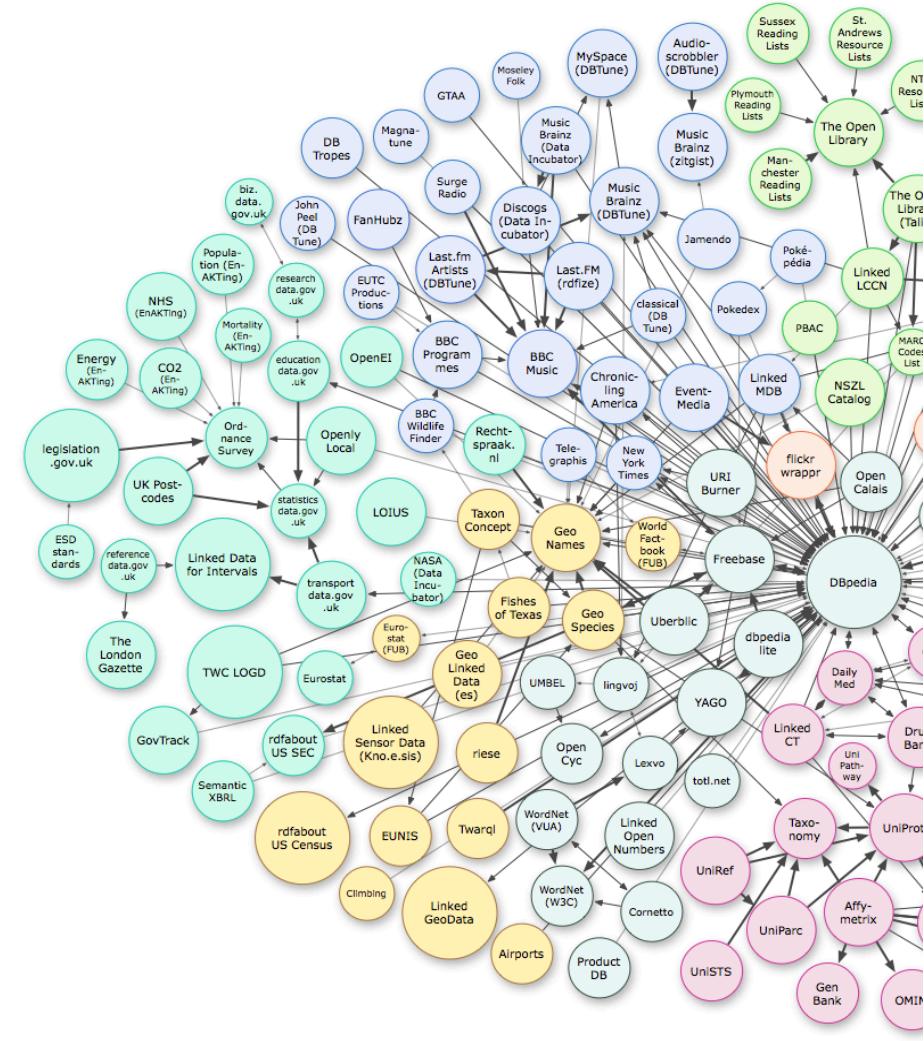
(“Aaron Gibson”) “singing” “Rocket Man”

...



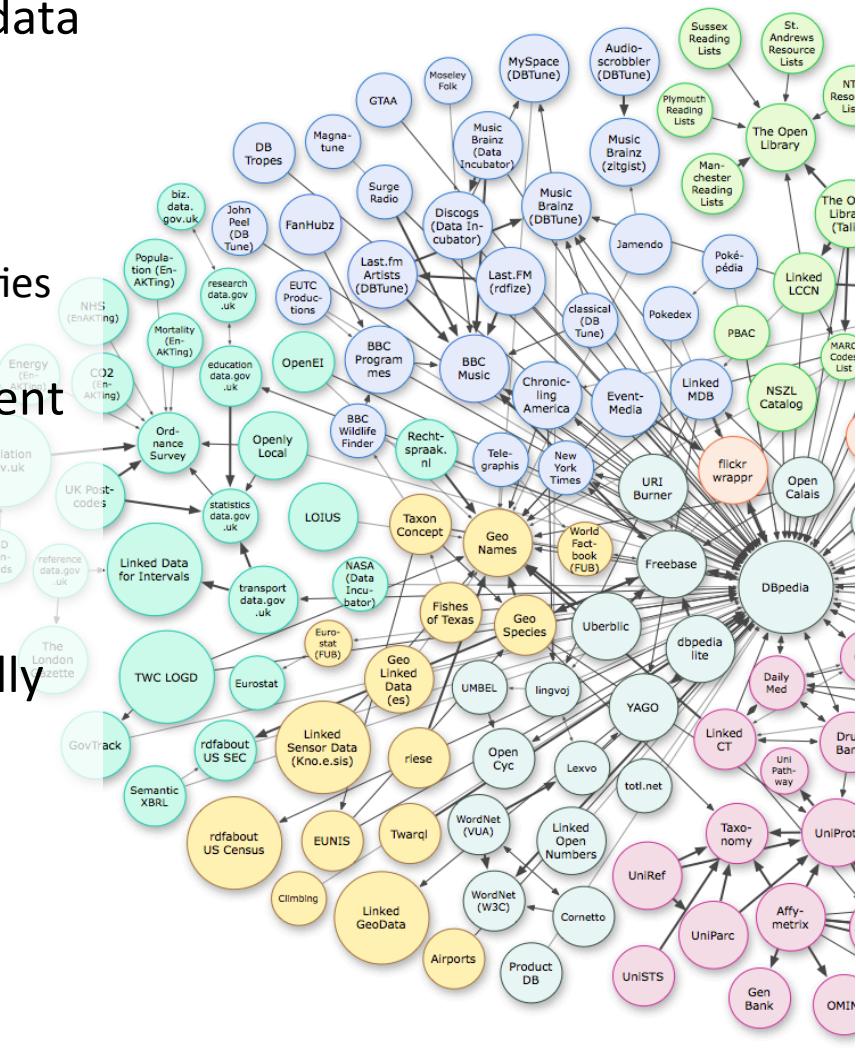
# Linking Data: from entities to KGs

- NLP tasks: Named Entity Recognition, Relation Extraction
- Link named entities
  - Entity linking
- Align properties and facts
  - ???



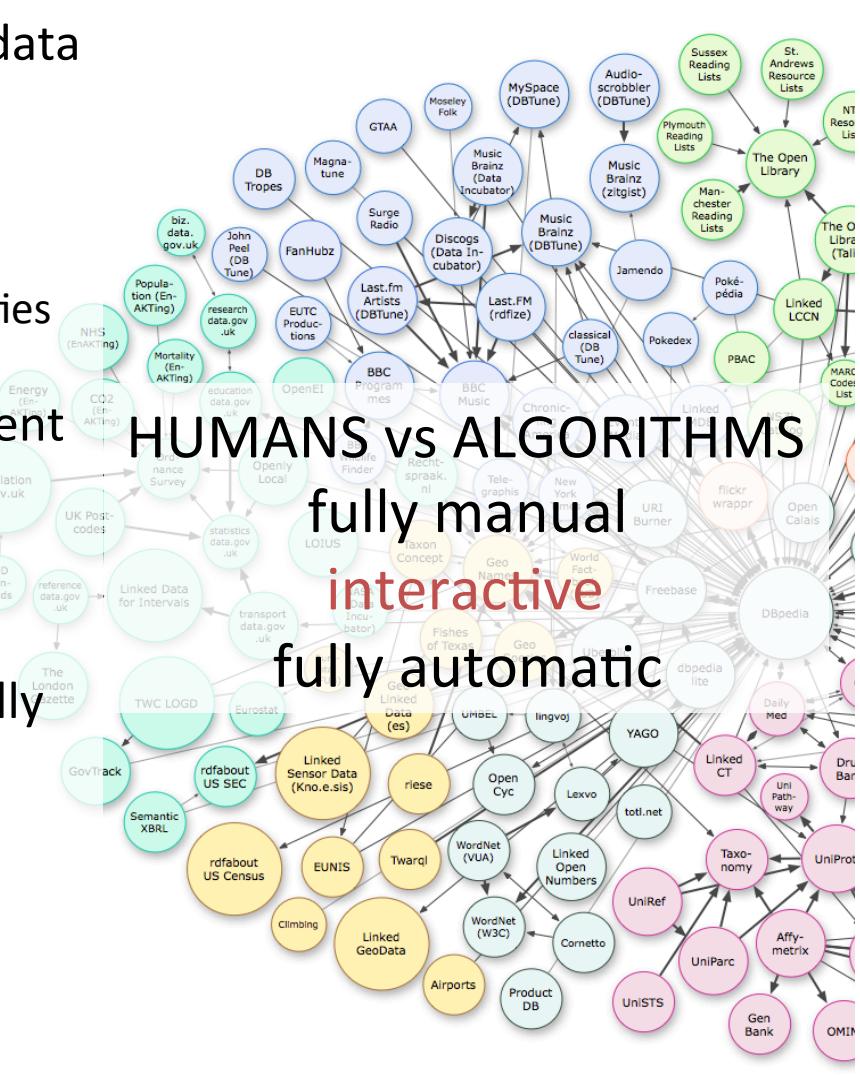
# How to map/link ?

- Evaluate **similarity** between the input data and KG using multiple criteria, e.g., semantics, structure of the KG, and between different kinds of input:
  - Instance vs instance
  - Concepts/properties vs Concepts/properties
  - Text fragment vs instance
- Combine evidence acquired with different criteria
  - With manual settings
  - Based on the quality of the evidence
  - Learning to combine
- Decide about the mappings to eventually establish (by analyzing the collected evidence)
  - Threshold-based
  - Optimization-based
  - Learning to...



# How to map/link ?

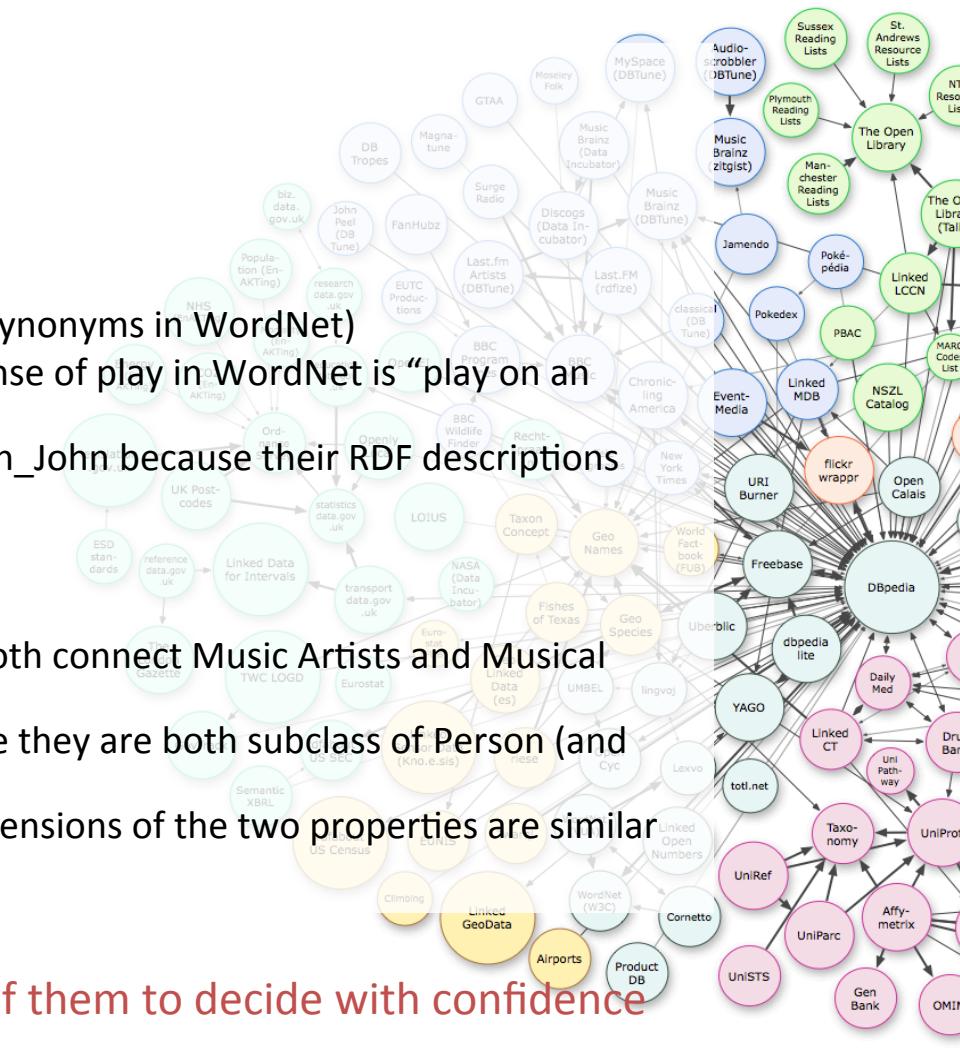
- Evaluate **similarity** between the input data and KG using multiple criteria, e.g., semantics, structure of the KG, and between different kinds of input:
  - Instance vs instance
  - Concepts/properties vs Concepts/properties
  - Text fragment vs instance
- Combine evidence acquired with different criteria
  - With manual settings
  - Based on the quality of the evidence
  - Learning to combine
- Decide about the mappings to eventually establish (by analyzing the collected evidence)
  - Threshold-based
  - Optimization-based
  - Learning to...



# When are things similar?

- Syntactic similarity
  - E.g., “Person” vs “Person”
  - E.g., “Elton John” vs “Sir Elton John” – but
- Lexical similarity
  - E.g., Television vs TV (because they are synonyms in WordNet)
  - E.g., play vs instrument (because one sense of play in WordNet is “play on an instrument”)
  - E.g., Reginald\_Kenneth\_Dwight and Elton\_John because their RDF descriptions share many words
- Structural similarity
  - E.g., plays vs instrument because they both connect Music Artists and Musical Instruments
  - E.g., Musician and Musical Artist because they are both subclass of Person (and have similar properties)
  - E.g., plays vs instrument because the extensions of the two properties are similar

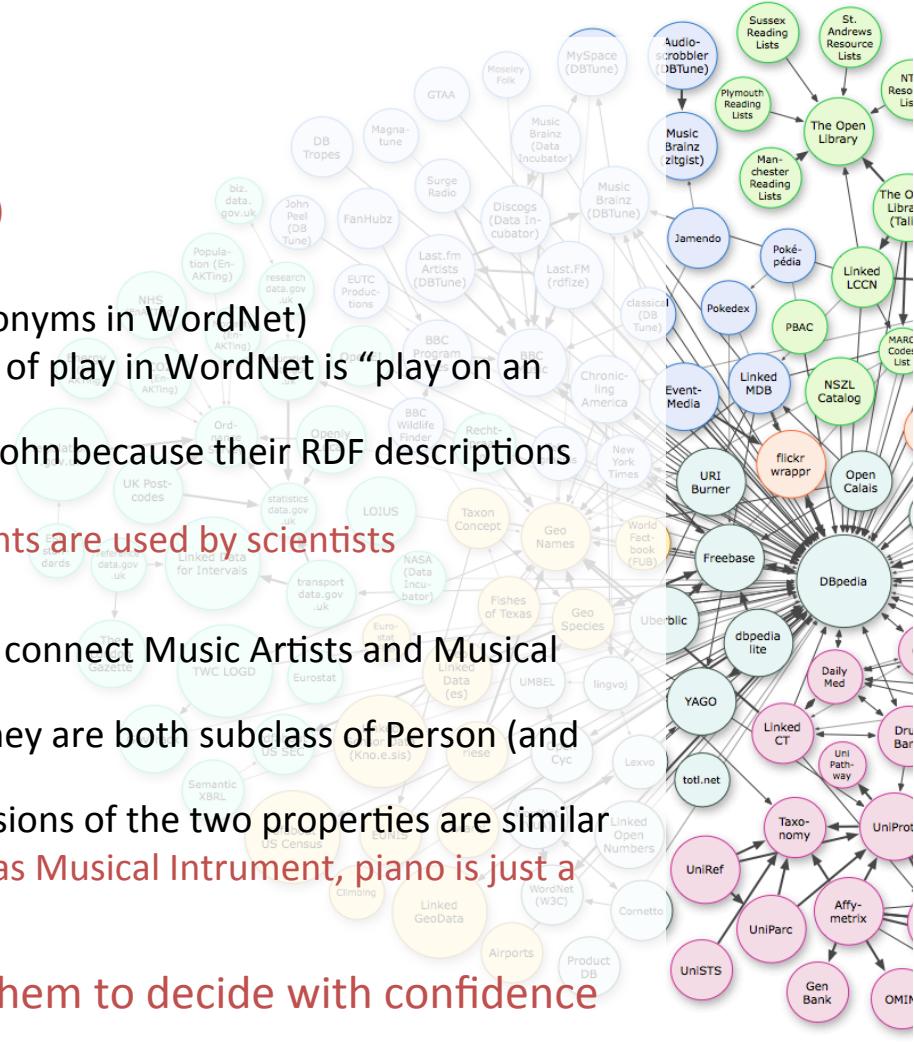
In practice, you really need several of them to decide with confidence



# When are things similar?

- Syntactic similarity
  - E.g., “Person” vs “Person”
  - E.g., “Elton John” vs “Sir Elton John” – but
  - But “trump”(card game) vs “Trump” (the devil)
- Lexical similarity
  - E.g., Television vs TV (because they are synonyms in WordNet)
  - E.g., play vs instrument (because one sense of play in WordNet is “play on an instrument”)
  - E.g., Reginald\_Kenneth\_Dwight and Elton\_John because their RDF descriptions share many words
  - But persons also play games, and instruments are used by scientists
- Structural similarity
  - E.g., plays vs instrument because they both connect Music Artists and Musical Instruments
  - E.g., Musician and Musical Artist because they are both subclass of Person (and have similar properties)
  - E.g., plays vs instrument because the extensions of the two properties are similar
  - But in DBpedia there is not a concept such as Musical Instrument, piano is just a Thing

In practice, you really need several of them to decide with confidence

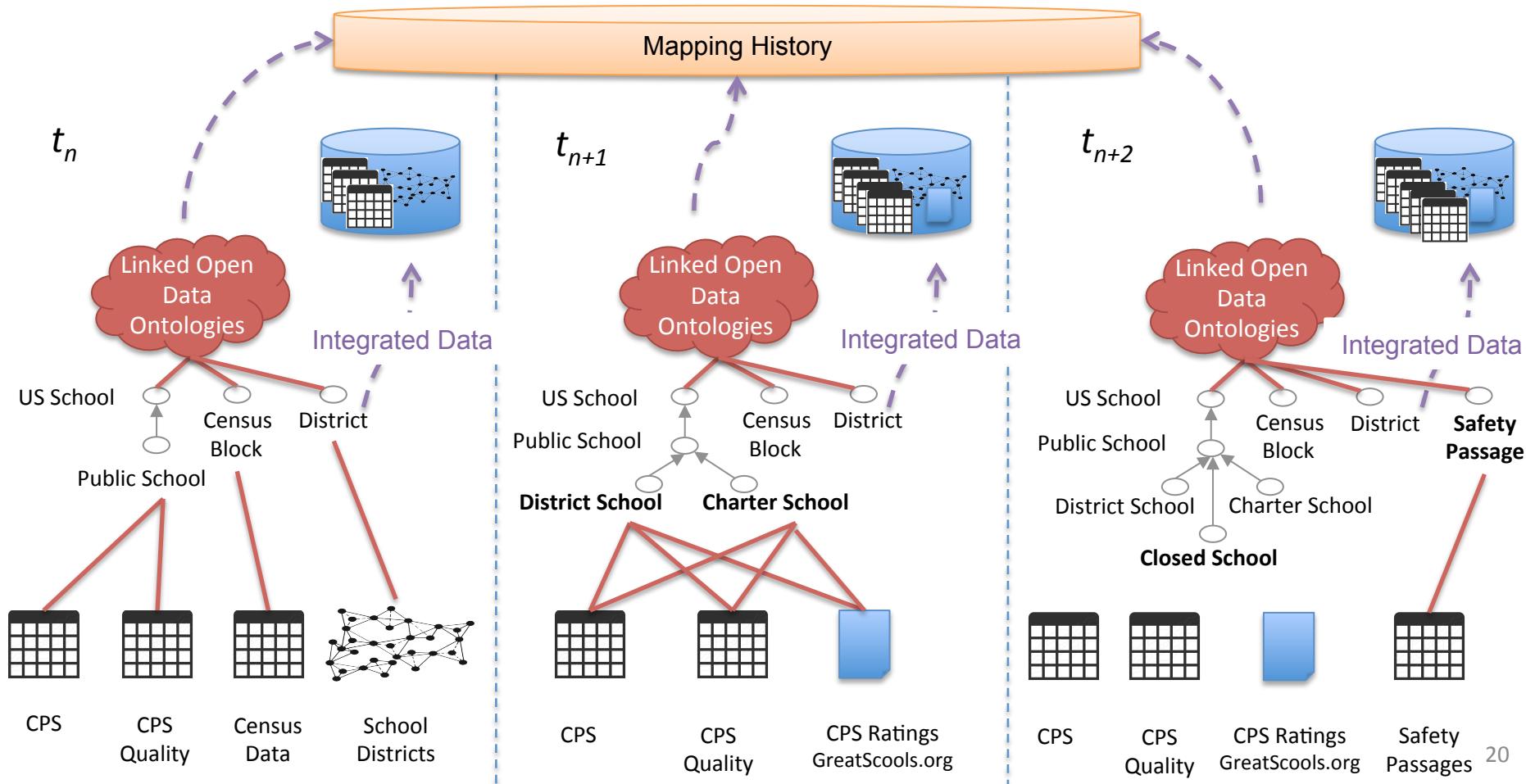


# KG Integration and Construction for data analytics

Part II: Data Integration in Analytics

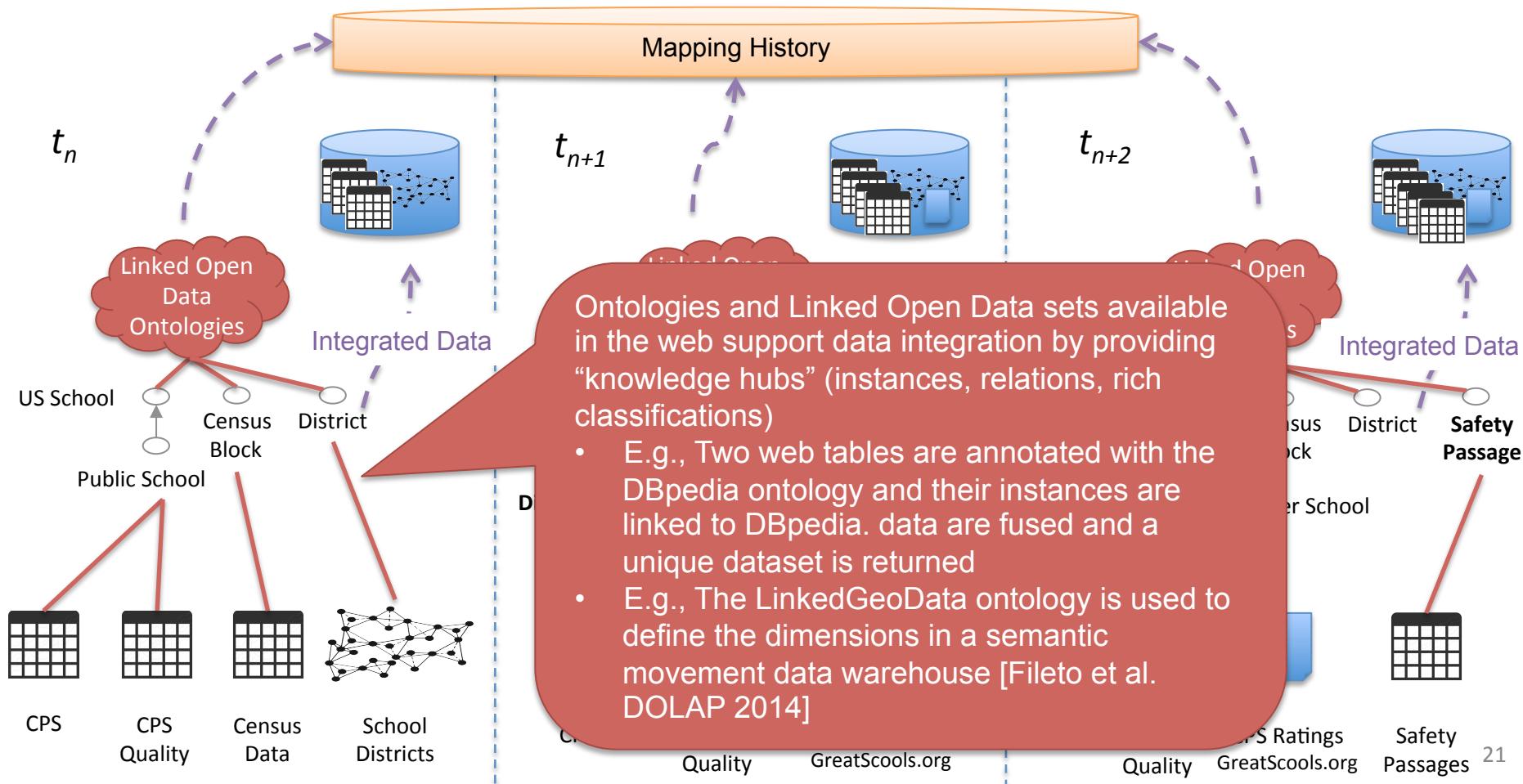
# Pay-as-you-go Data Integration with Matching Tools

Matching, Mapping Definition, Data Transformation



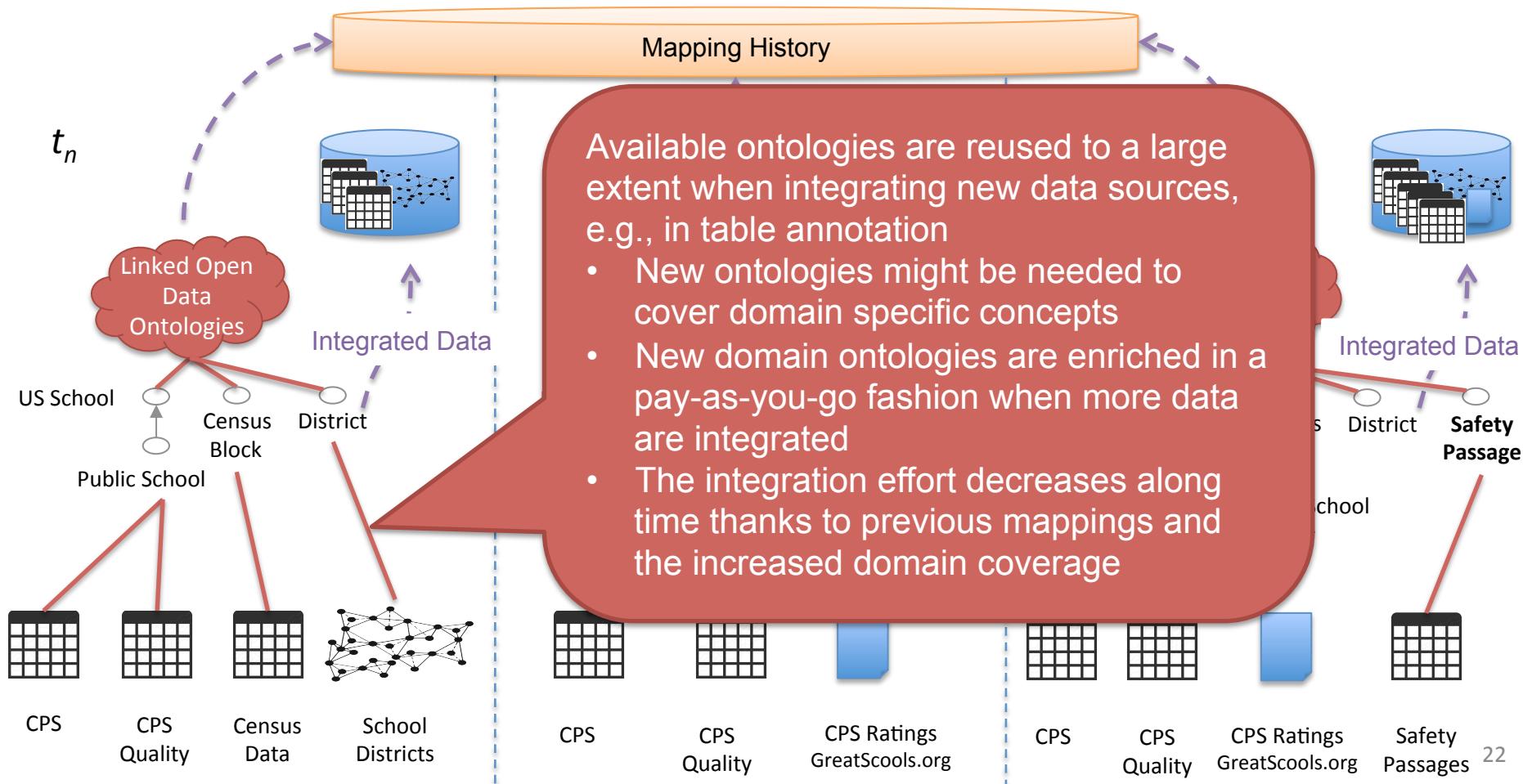
# Pay-as-you-go Data Integration with Matching Tools

## Matching, Mapping Definition, Data Transformation



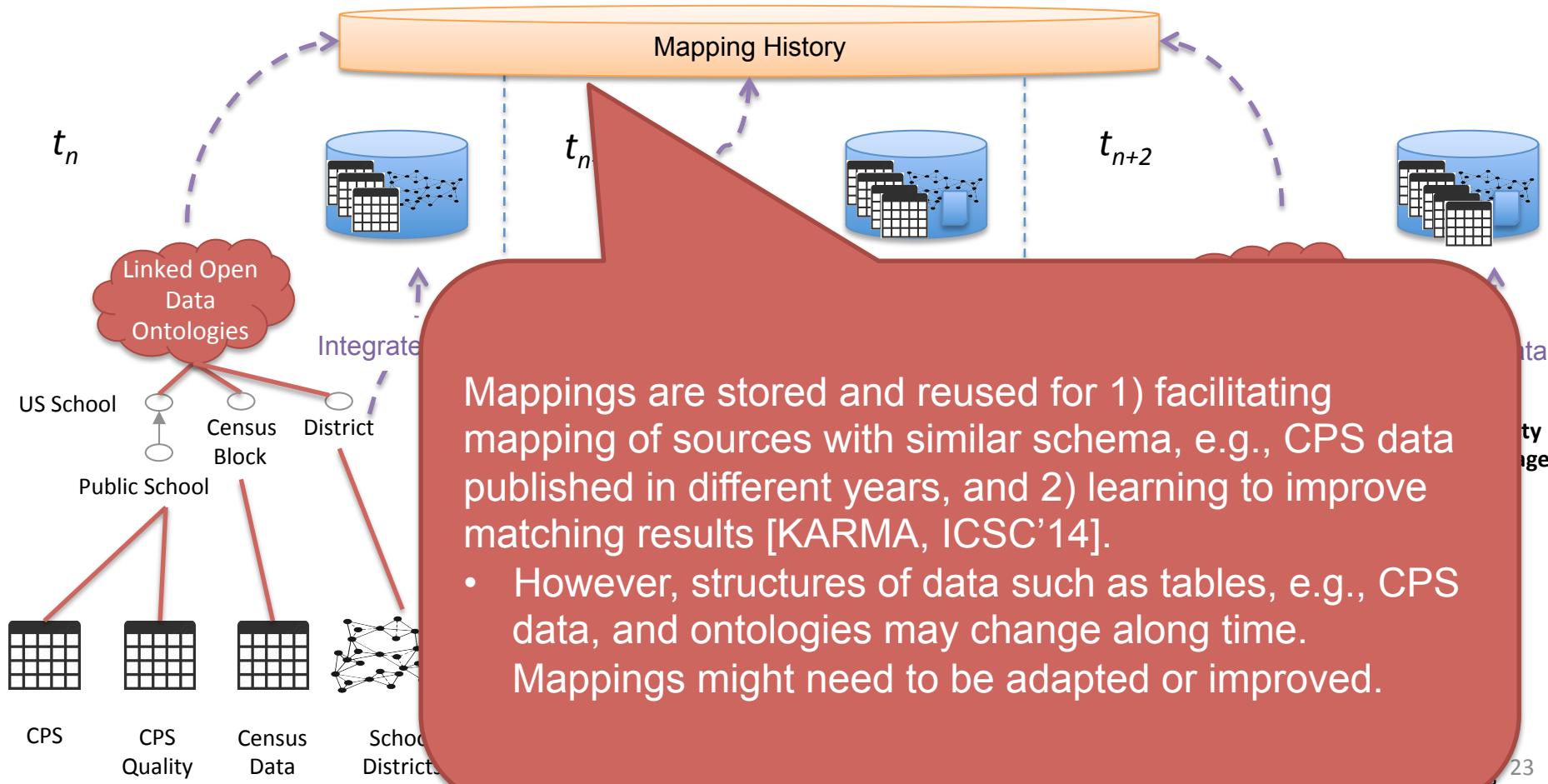
# Pay-as-you-go Data Integration with Matching Tools

## Matching, Mapping Definition, Data Transformation



# Pay-as-you-go Data Integration with Matching Tools

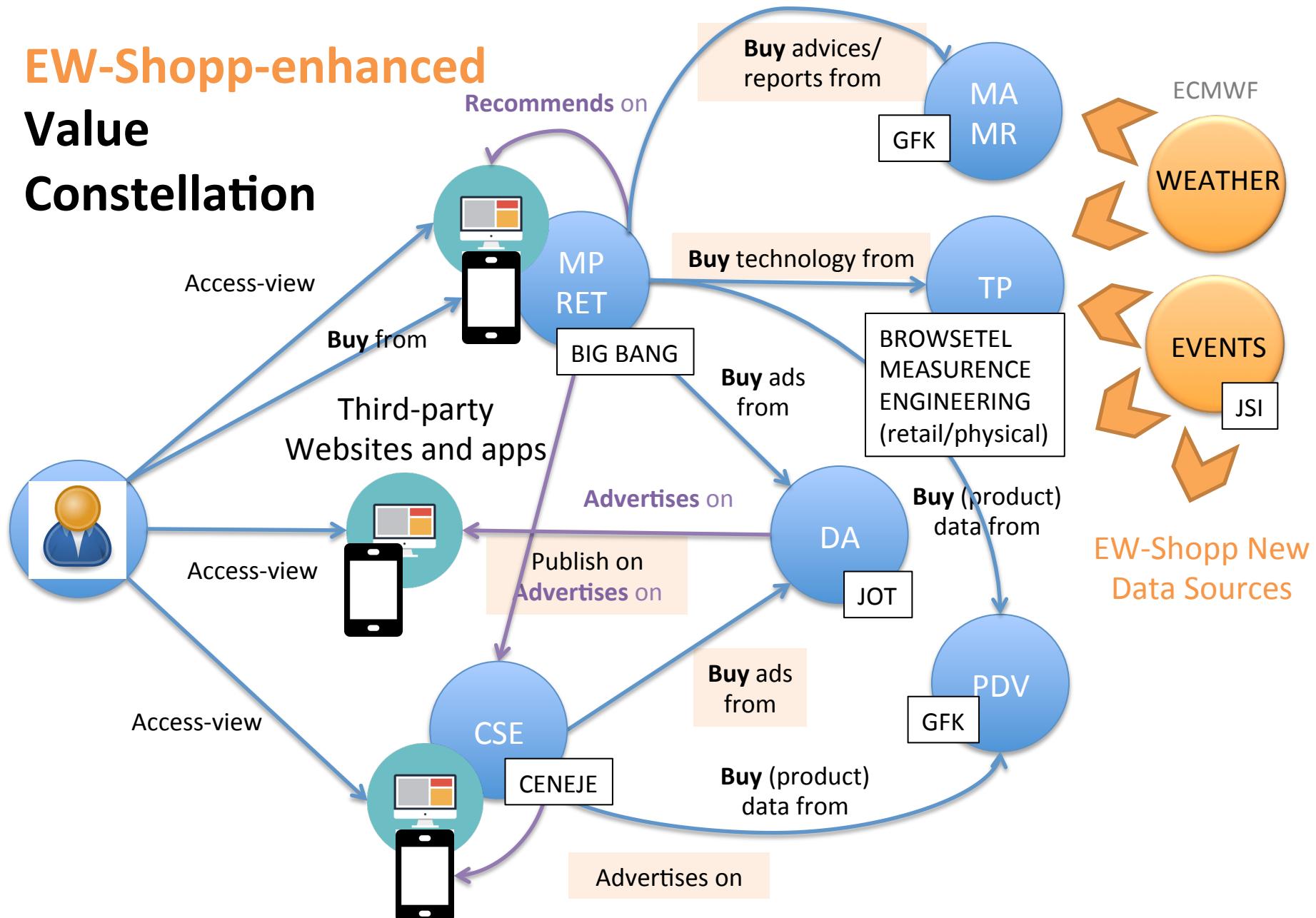
Matching, Mapping Definition, Data Transformation



# EW-Shopp-enhanced

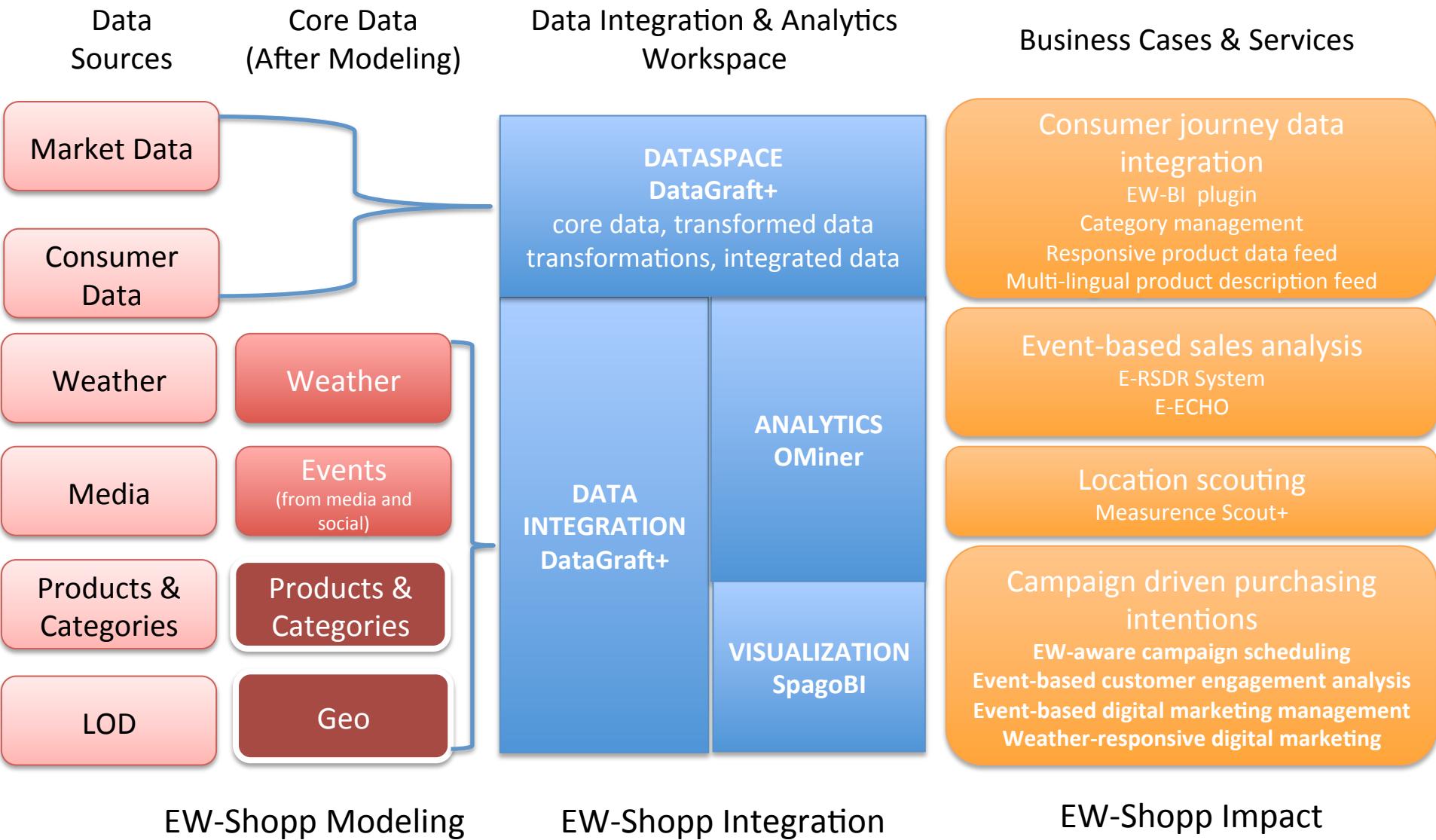
## Value

## Constellation



EW-Shopp: Supporting Event and Weather-based Data Analytics and Marketing along the Shopper Journey (H2020 EU innovation project, 10 partners - 7 IND., ≈3M€, 36 months)

# EW-Shopp Key Components



# Data Value Chain

## For JOT MEDIA (SP)

