

Semantic Web: Vision, Reality, Revision

James Hendler

Tetherless World Professor of Computer, Web and Cognitive Sciences

Rensselaer Polytechnic Institute (RPI)

<http://www.cs.rpi.edu/~hendler>

@jahendler (twitter)

2012: Google introduces the knowledge graph

The image is a screenshot of a blog post on the Google Official Blog. At the top left, it says 'Google Official Blog' with the tagline 'Insights from Googlers into our products, technology and the Google culture'. The main heading is 'Introducing the Knowledge Graph: things, not strings', dated May 16, 2012. The post text explains that search is about discovery and introduces the Knowledge Graph, which helps discover new information quickly and easily. It uses the example of 'Taj Mahal' to show how the graph understands context. A sidebar on the right offers social media links for FeedBurner, RSS Feed, and @Google, along with a 'Follow' button. The bottom of the page lists 'More blogs from Google'.

Google Official Blog
Insights from Googlers into our products, technology and the Google culture

Introducing the Knowledge Graph: things, not strings

May 16, 2012

Cross-posted on the [Inside Search Blog](#)

Search is a lot about discovery—the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I'm really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

Take a query like [taj mahal]. For more than four decades, search has essentially been about matching keywords to queries. To a search engine the words [taj mahal] have been just that—two words.

But we all know that [taj mahal] has a much richer meaning. You might think of one of the world's most beautiful monuments, or a Grammy Award-winning musician, or possibly even a casino in Atlantic City, NJ. Or, depending on when you last ate, the

More blogs from Google

The Knowledge Graph enables you to search for things, people or places that Google knows about—landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.

Google's Knowledge Graph isn't just rooted in public sources such as Freebase, Wikipedia and the CIA World Factbook. It's also augmented at a much larger scale—because we're focused on comprehensive breadth and depth. It currently contains more than 500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects. And it's tuned based on what people search for, and what we find out on the web.

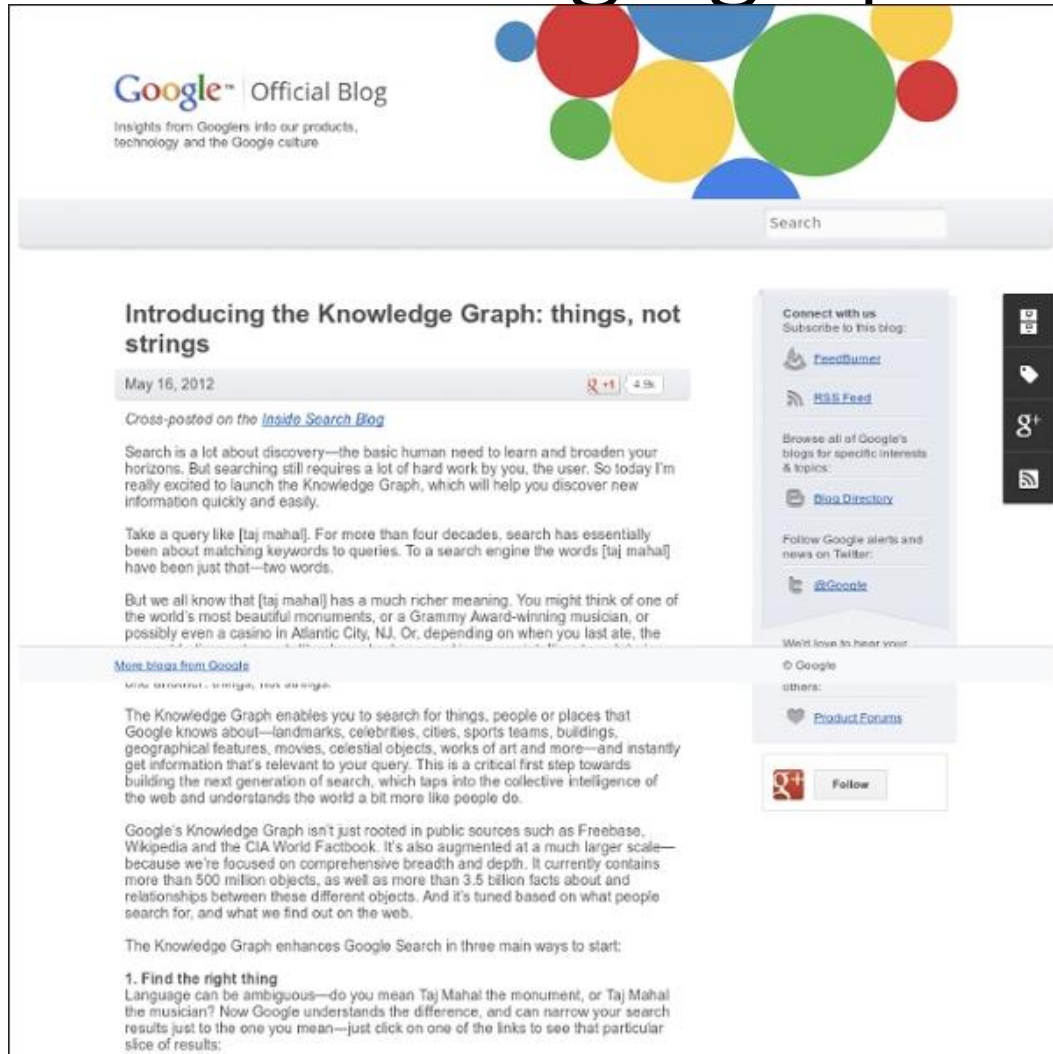
The Knowledge Graph enhances Google Search in three main ways to start:

- 1. Find the right thing**
Language can be ambiguous—do you mean Taj Mahal the monument, or Taj Mahal the musician? Now Google understands the difference, and can narrow your search results just to the one you mean—just click on one of the links to see that particular slice of results:

Search is a lot about discovery – the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I'm really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

Amit Singhal, **2012**

Reality 2012: Google introduces the phrase “knowledge graph”



The image is a screenshot of a Google Official Blog post. At the top left, it says "Google Official Blog" with the tagline "Insights from Googlers into our products, technology and the Google culture". The main heading of the post is "Introducing the Knowledge Graph: things, not strings", dated May 16, 2012. The post text explains that search is about discovery and introduces the Knowledge Graph, which helps discover new information quickly and easily. It uses the example of "Taj Mahal" to show how the graph understands context. The post also mentions that the Knowledge Graph is built from public sources like Freebase and Wikipedia, and contains over 500 million objects and 3.5 billion facts. It lists three ways the Knowledge Graph enhances Google Search, starting with "1. Find the right thing". On the right side of the page, there are social media links for FeedBurner, RSS Feed, Blog Directory, and Twitter (@Google), along with a "Follow" button for Google+.

Google Official Blog
Insights from Googlers into our products, technology and the Google culture

Introducing the Knowledge Graph: things, not strings

May 16, 2012

Cross-posted on the [Inside Search Blog](#)

Search is a lot about discovery—the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I'm really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

Take a query like [taj mahal]. For more than four decades, search has essentially been about matching keywords to queries. To a search engine the words [taj mahal] have been just that—two words.

But we all know that [taj mahal] has a much richer meaning. You might think of one of the world's most beautiful monuments, or a Grammy Award-winning musician, or possibly even a casino in Atlantic City, NJ. Or, depending on when you last ate, the

More blogs from Google

View original - things, not strings

The Knowledge Graph enables you to search for things, people or places that Google knows about—landmarks, celebrities, cities, sports teams, buildings, geographical features, movies, celestial objects, works of art and more—and instantly get information that's relevant to your query. This is a critical first step towards building the next generation of search, which taps into the collective intelligence of the web and understands the world a bit more like people do.

Google's Knowledge Graph isn't just rooted in public sources such as Freebase, Wikipedia and the CIA World Factbook. It's also augmented at a much larger scale—because we're focused on comprehensive breadth and depth. It currently contains more than 500 million objects, as well as more than 3.5 billion facts about and relationships between these different objects. And it's tuned based on what people search for, and what we find out on the web.

The Knowledge Graph enhances Google Search in three main ways to start:

- 1. Find the right thing**
Language can be ambiguous—do you mean Taj Mahal the monument, or Taj Mahal the musician? Now Google understands the difference, and can narrow your search results just to the one you mean—just click on one of the links to see that particular slice of results:

Search is a lot about discovery – the basic human need to learn and broaden your horizons. But searching still requires a lot of hard work by you, the user. So today I'm really excited to launch the Knowledge Graph, which will help you discover new information quickly and easily.

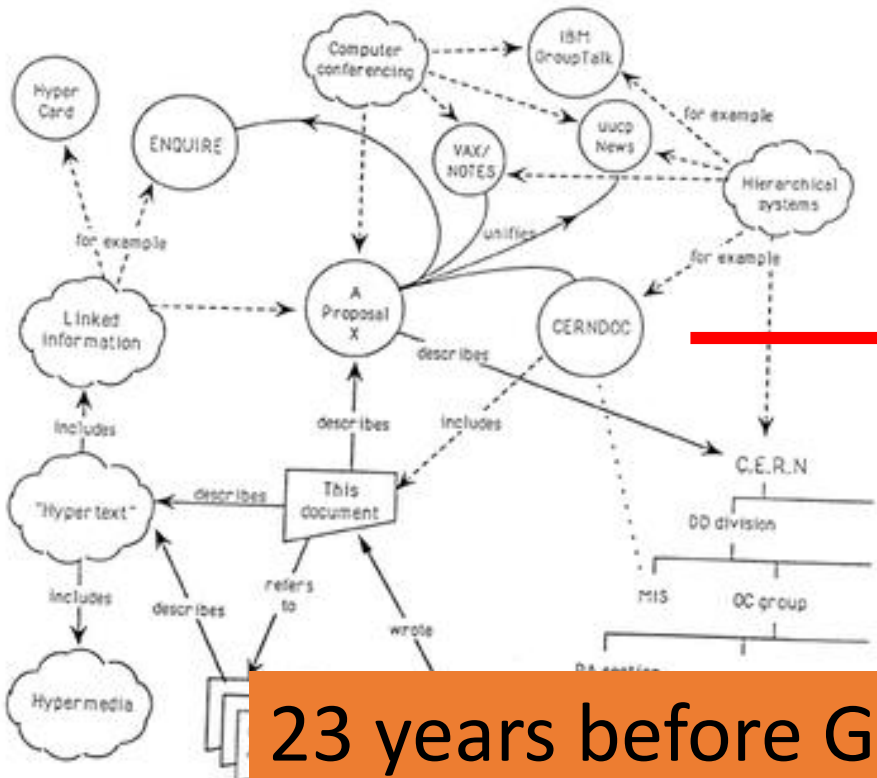
Amit Singhal, 2012

Information Management: A Proposal

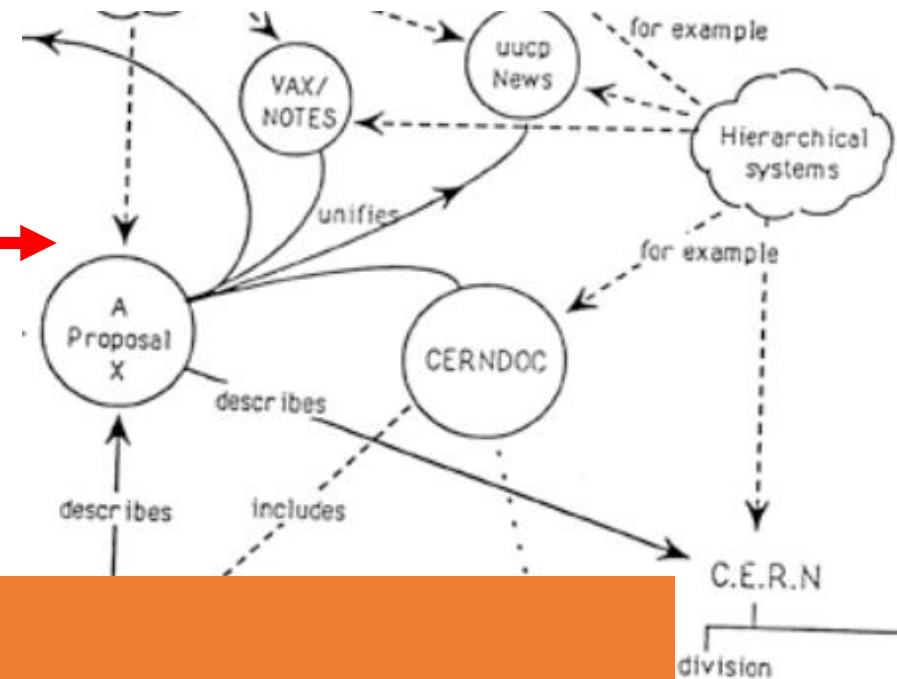
Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

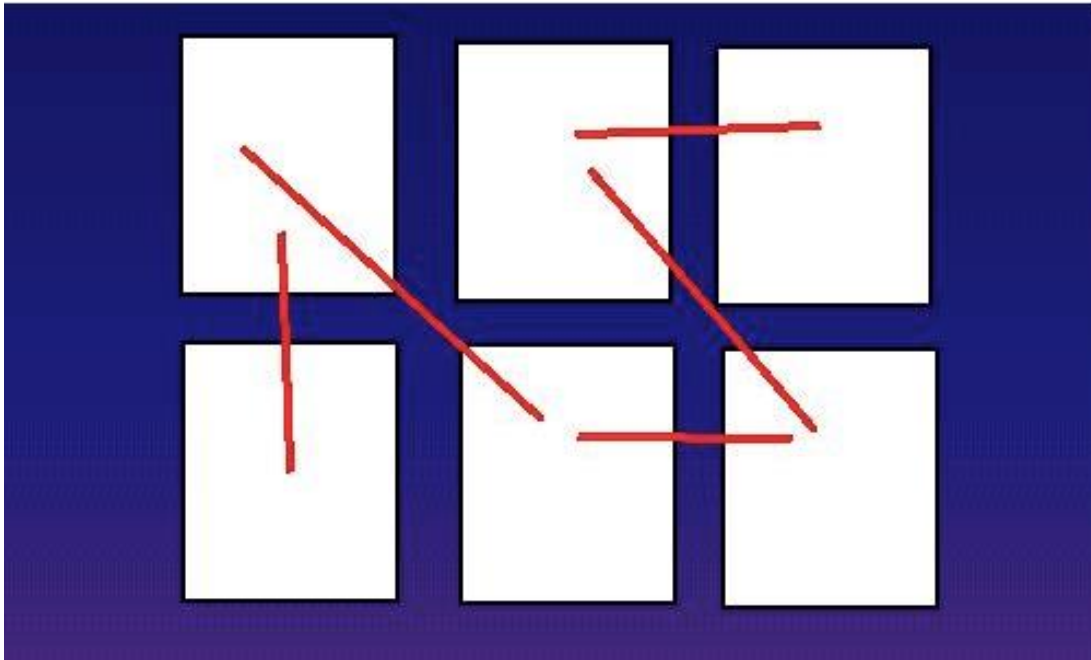


The original web proposal (1989) was actually a knowledge graph!



23 years before Google!!

1994 presentation: What the Web is
Tim Berners-Lee, Geneva, WWW
The Need for Semantics in the Web

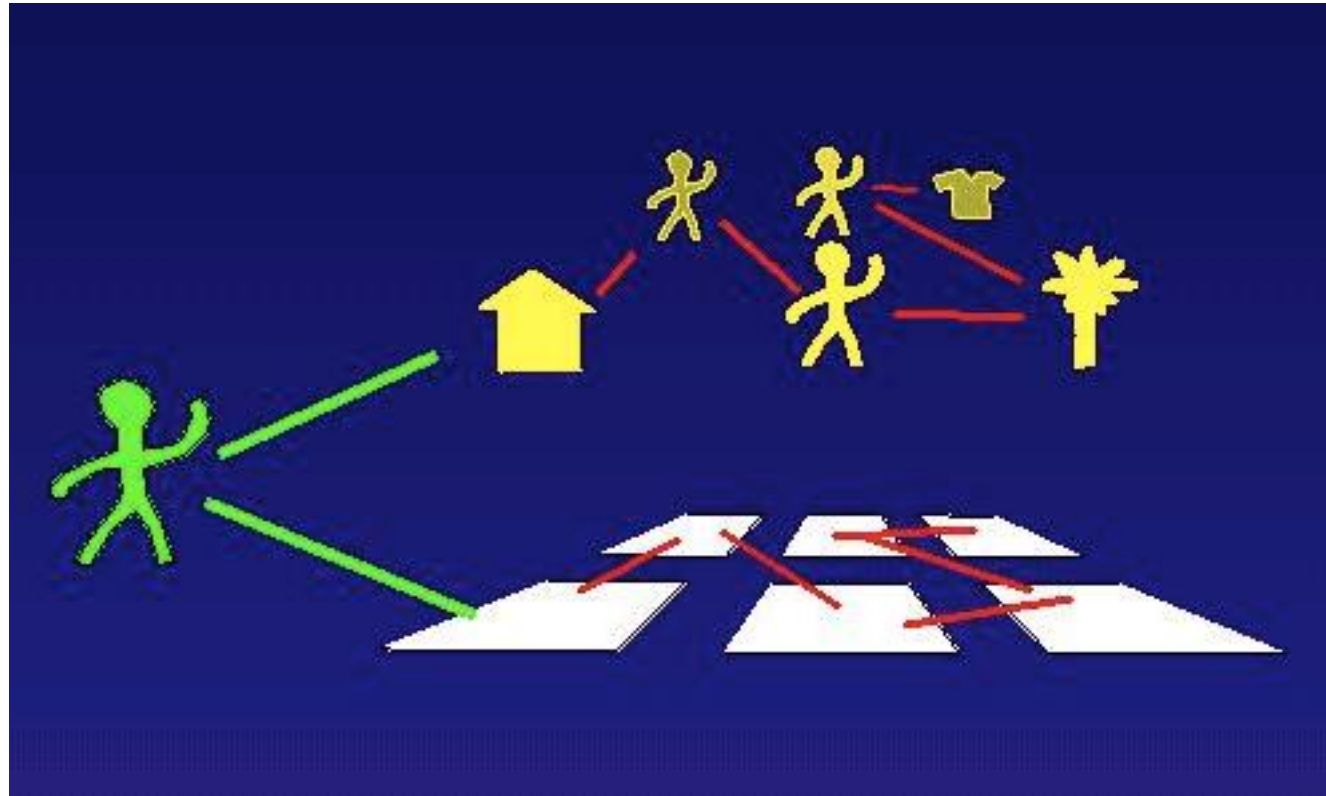


<https://www.w3.org/Talks/WWW94Tim/>

1994 presentation: What the Web is

Tim Berners-Lee, Geneva, WWW

The Need for Semantics in the Web



- “..documents on the web describe real objects and imaginary concepts, and give particular relationships between them...”

The Semantic Web vision article, which attempted to capture Tim's approach, appeared in **2001**



THE SEMANTIC WEB

A new form of Web content
that is meaningful to computers
will unleash a revolution of new abilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

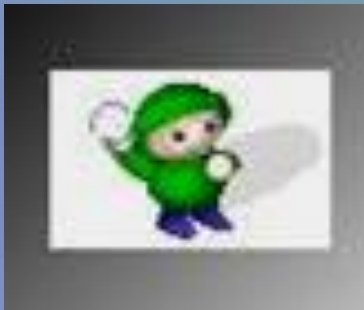
PHOTO CREDIT: NASA



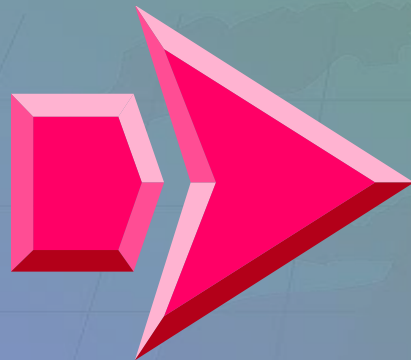
DAML

Notional Schedule

Tetherless World Constellation



Now



Later

2001: We will change the world!



Example – Semantic Search ca. 2010

Tetherless World Constellation





Semantic Search ca. 2015

Tetherless World Constellation

Google Jim hendler

[All](#) [News](#) [Images](#) [Videos](#) [Maps](#) [More](#) [Settings](#) [Tools](#)

About 360.000 results (0,27 seconds)

Prof. James A. Hendler
<https://www.cs.rpi.edu/~hendler/>

Professor James A. Hendler. Tetherless World Professor of Computer, Web and Cognitive Sciences
Director, Rensselaer Institute for Data Exploration and ...

James Hendler - Wikipedia
https://en.wikipedia.org/wiki/James_Hendler

James Alexander Hendler (born April 2, 1957) is an artificial intelligence researcher at Rensselaer Polytechnic Institute, United States, and one of the originators ...
Institutions: Rensselaer Polytechnic Institute U... **Spouse(s):** Terry Horowitz
Known for: Significant Contributions to the Se...
[Research](#) · [Books](#) · [Honors](#)

Jim Hendler - Google Scholar Citations
scholar.google.com/citations?user=JNPbTdIAAAAJ&hl=en

HTN planning for web service composition using SHOP2. E Sirin, B Parsia, D Wu, J Hendler, D Nau.
Web Semantics: Science, Services and Agents on the World ...

jahendler (@jahendler) | Twitter
<https://twitter.com/jahendler?lang=en>

Semantic Web geek, Web Science evangelist, Open Data nerd and (somewhat reluctantly) Big Data lead. usually USA. cs.rpi.edu/~hendler. Joined July 2008.

People of ACM - Jim Hendler - Association for Computing Machinery
<https://www.acm.org/articles/people-of-acm/2018/jim-hendler>

People of ACM - Jim Hendler. August 9, 2018. You are credited with being an originator of the semantic web, an initiative to extend the World Wide Web to ...

dblp: James A. Hendler
<https://dblp.uni-trier.de/Persons>

List of computer science publications by James A. Hendler.

James Hendler | IBM Center for The Business of Government
www.businessofgovernment.org/bio/james-hendler

Apr 20, 2018 - James Hendler is the Director of the Institute for Data Exploration and Applications and the Tetherless World Professor of Computer, Web and ...

Jim Hendler - Rensselaer Polytechnic Institute - VideoLectures.NET

James Hendler

Researcher

James Alexander Hendler is an artificial intelligence researcher at Rensselaer Polytechnic Institute, United States, and one of the originators of the Semantic Web. [Wikipedia](#)

Born: April 2, 1957 (age 61 years), Queens, New York City, New York, United States
Spouse: Terry Horowitz
Education: Southern Methodist University, Yale University, Brown University
Influenced by: Tim Berners-Lee, Dana S. Nau, Ben Shneiderman, Eugene Charniak, Edward Feigenbaum, Jack Minker
Fields: Artificial intelligence, Semantic Web

Books

Social Machines: The Comi... 2016	Robots for Kids: Exploring... 2000	Integrating marker-p... and probl... 1985	Web Science: Understa... 2013

People also search for [View 5+ more](#)

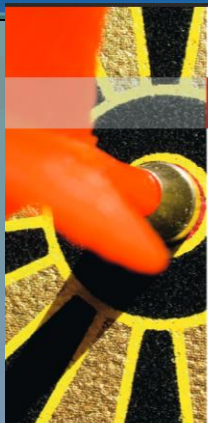
Tim Berners-Lee	Nigel Shadbolt	Dean Allemang	Wendy Hall	Jen Golbeck

[Claim this knowledge panel](#) [Feedback](#)



Google 2009

Tetherless World Constellation



EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

For many tasks, words and word combinations provide all the representational machinery we need to learn from text.

Choose a representation that can use unsupervised learning on unlabeled data, which is so much more plentiful than labeled data.



Google 2012

Tetherless World Constellation



Peter Norvig 4. Towards a semantic web: using the data that's available

So at Google we try to connect our customers with the information they want, and we do that using all the tools available. Now we hear a lot of talk about the Semantic Web and that's coming along, but it's not quite here yet in that there is some markup on the Web that's explicit declaration: I'm an object of this type and I'm related to this other type of object. And wherever that's available, we'll use that. But most of the information that's on the Web is not in that formal type of markup, most of it is designed for human consumption, not designed for computer databases to deal with, and so we have to go with what we're given, and if most of what we're given is text, then we're going to have to figure out what the words in that text mean. And if they aren't in a formal database language and if some of it is in images or videos or tables or other type of information, we're going to use all of that. So we'll take as much as is provided to us and where it's missing, where there is no formal computer representation, we'll go with what was given to the humans and make as much sense of that. So we're creating a sort of semantic interpretation of texts or images that weren't intended in that form.



Google 2013

Tetherless World Constellation

Keynote - Ramanathan V. Guha



Light at the End of the Tunnel

A significant fraction of the pages on the web are generated from structured databases. A longstanding goal of the semantic web initiative is to get webmasters to make this structured data directly available on the web. The path towards this objective has been rocky at best. While there have been some notable wins (such as RSS and FOAF), many of the other initiatives have seen little industry adoption. Learning from these earlier attempts has guided the development of schema.org, which appears to have altered the trajectory. Two years after its launch over 4 million Internet domains are using schema.org markup. In this talk, we recount the history behind the early efforts and try to understand why some of them succeeded while others failed. We will then give an update on Schema.org, its goals, accomplishments and where it is headed. We will also discuss some of the interesting research problems being addressed in the context of this effort.



TW/C

Impressive results

Tetherless World Constellation

The screenshot shows a Google search for "Alces Alces". The search bar at the top contains the text "Alces Alces" and a search icon. Below the search bar are navigation tabs for "All", "Images", "Videos", "News", "Maps", "More", and "Search tools". The search results section shows "About 581,000 results (0.58 seconds)". The first result is from Wikipedia, titled "Moose - Wikipedia, the free encyclopedia", with a URL and a brief description. Below this are two columns of related terms: "Alaska moose" and "Elk". To the right of the search results is a knowledge panel for "Moose", categorized as an "Animal". The panel includes a grid of images, a description of the species, its scientific name, height, lifespan, conservation status, trophic level, and mass. At the bottom of the panel, there are links to "People also search for" with images and labels for Deer, Caribou, Elk, Roe deer, and Bear. A "Feedback" link is located at the bottom right of the panel.

Google “knowledge vault” reported to have over 3.5 billion “facts” (links)

Google finds embedded metadata on >40% of its crawl – Norvig, 2016



Not just Google...

Tetherless World Constellation

Baidu 百度 tom cruise 百度一下

网页 图片 贴吧 视频 知道 新闻 文库 音乐 地图 更多»


百度为您找到相关结果约6,650,000个 搜索工具

您可以仅查看：[英文结果](#)

TOM CRUISE 百度百科

 职业：演员、制片人
生日：1962年7月3日
个人信息：173cm/巨蟹座/O
代表作品：《壮志凌云》《雨人》《生于七月四日》《夜访...
简介：汤姆·克鲁斯 (Tom Cruise)，1962年7月3日出生于美...
[早年经历](#) [演艺经历](#) [个人生活](#) [主要作品](#) [获奖记录](#) [更多>>](#)
<https://baike.baidu.com/>

Tom Cruise 百度翻译

Tom Cruise 

[词典] [人名] 汤姆克鲁斯;

[例句] S: I also heard that **Tom Cruise** and Katie Holmes made an appearance at the Oscars. 
我还听说汤姆克鲁斯和凯蒂·霍尔姆斯在奥斯卡颁奖现场露面了。

[进行更多翻译](#)

fanyi.baidu.com

汤姆·克鲁斯 Tom Cruise - 汤姆·克鲁斯 (豆瓣) 豆瓣电影

 汤姆·克鲁斯简介、图片写真、获奖情况及电影作品一览... 官方网站：
<http://www.tomcruise.com/> 分享到 ...
<https://movie.douban.com/celeb...> - 百度快照

Semantic Search

Google's knowledge graph has been big news this year, but Baidu claims to have been integrating its own version - 'box computing' - since 2009. It works in the same way as the Knowledge Graph, displaying relevant data and information directly on the SERPS above the main results.


国内版 实际版


tom cruise


All Images Videos 翻译或中文 关联网页

1,330,000 Results Any time

Comprehensive coverage on Tom Cruise

 30 Unforgettable Tom Cruise Moments That Changed Hollywood Forever
There are few working actors today as polarizing as Tom Cruise. From his insane movie stunts to his devo...
The Cheat Sheet · 21h

 Tom Cruise and Tropic
Take Surf Back — But LHM
Katie Holmes Loves To Fight!

 LOCAL SOURCE
Have You Spotted Tom Cruise in Clearwater Yet?
USA Patch · 1d

Observer · 1d Radar Online · 6h

Official Tom Cruise Website
www.tomcruise.com
The Official Tom Cruise Website. Featuring Tom Cruise's biography, filmography, links to social media accounts, and information about his latest films.

Tom Cruise - IMDb
<https://www.imdb.com/name/nm0000129>
Tom Cruise, Actor: Top Gun. In 1976, if you had told fourteen year-old Franciscan seminary student Thomas Cruise Mapother IV that one day in the not too distant future he would be Tom Cruise, one of the top 100 movie stars of all time, he would have probably grinned and told you that his ambition was to join the priesthood.



Tom Cruise - Wikipedia
https://en.wikipedia.org/wiki/Tom_Cruise
Relatives: William Mapother (cousin) Children: 3
Occupation: Actor, producer Years active: 1981-present



Overview Contents Early life Career Relationships Scientology >

Thomas Cruise Mapother IV is an American actor and producer. He started his career at age 19 in the film Endless Love, before making his breakthrough in the comedy Risky Business and receiving widespread attention for starring in the action drama Top Gun as Lieutenant Pete "Maverick" Mitchell. After starring in The Color of Money and Cocktail, Cruise starred opposite Dustin Hoffman in the Academy Award for Best Picture-winning drama Rain Man. For his role as anti-war advisor Ron Kovic, ...
See more on en.wikipedia.org · Text under CC-BY-SA license






Twitter Tweet Button
platform.twitter.com/widgets/tweet_button.html?text=A Tom Cruise ...
Twitter Tweet Button ... Tweet

Videos of tom cruise
bing.com/videos


 2:31  FEEL THE NEED.






 **Tom Cruise**  Share
American Actor


Thomas Cruise Mapother IV, best known by his stage name Tom Cruise, is an American actor and producer. He started his career at age 19 in the film Endless Love, before making his breakthrough in the comedy Risky Business and receiving widespread attention for starring in the action drama Top Gun as Lieutenant Pete "Maverick" Mitchell. After starrin...
+






    
Wikipedia IMDb Twitter Facebook Tumblr


Born: Jul 03, 1962 (age 56) Syracuse, NY
Height: 5'7"
Net worth: \$570 million USD (2017)
Spouse: Katie Holmes (m. 2006 - 2012) · Nicole Kidman (m. 1990 - 2001) · Mimi Rogers (m. 1987 - 1990)
Partner: Penelope Cruz (2001 - 2004)
Upcoming movies: Top Gun: Maverick






Movies  See all (20+)

    
Mission: Impossible - Fallout 2018 Edge of Tomorrow 2014 American Made 2017 The Mummy 2017 Oblivion 2013

Romance  See all (5+)

    
Katie Holmes 2006 - 2012 Penelope Cruz 2006 - 2012 Nicole Kidman 1990 - 2001 Mimi Rogers 1987 - 1990 Cher

People also search for  See all (20+)

    
Brad Pitt Emily Blunt Leonardo DiCaprio Kelly McGillis Johnny Depp



2010-2012 saw a major uptake in use of Semantic Web

Tetherless World Constellation

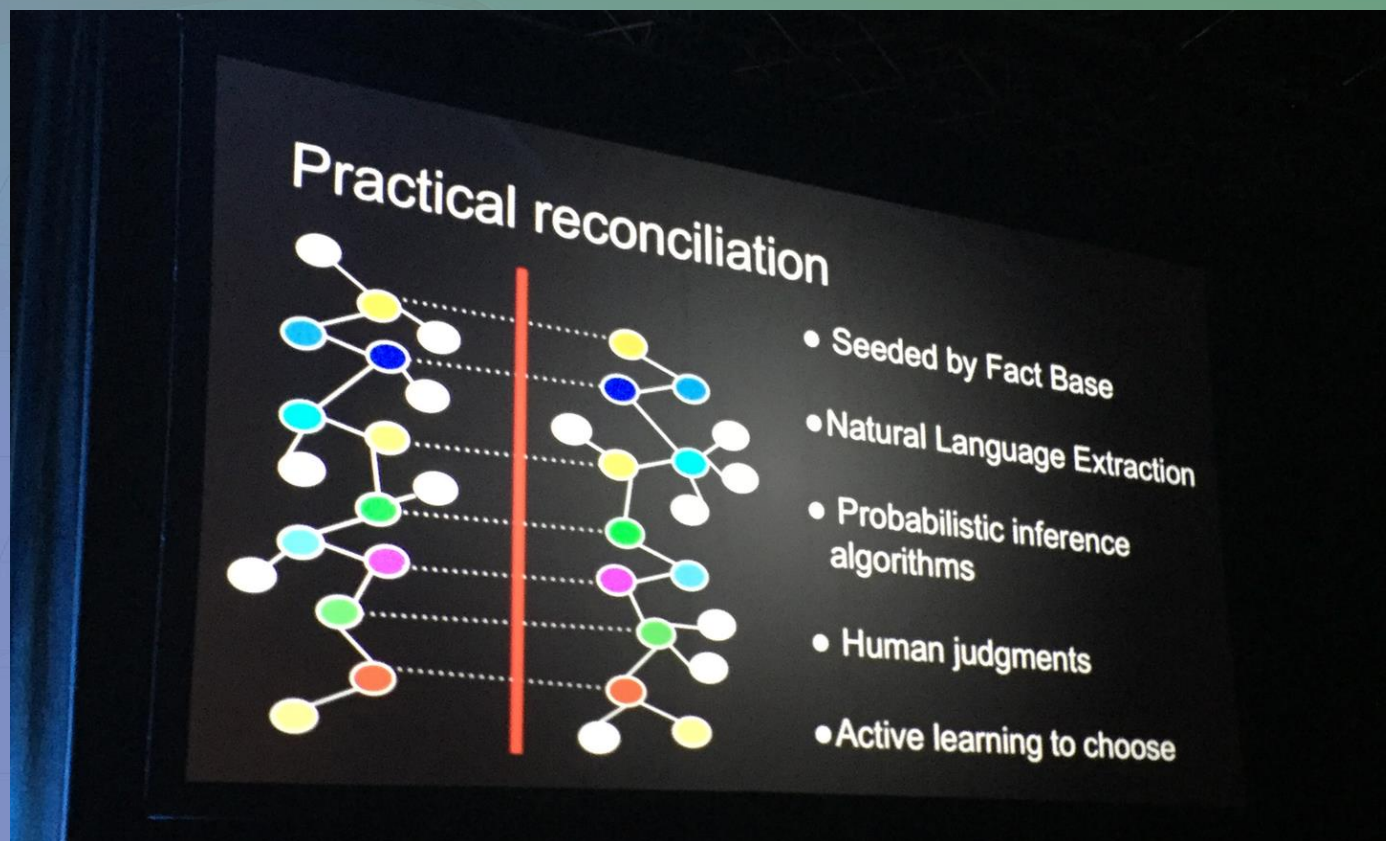
- What drove the change?
 - Maturation of Linked Data technologies
 - Embedded markup
 - RDFa (and schema.org)
 - SPARQL endpoints
 - graph database standard
 - Lightweight Knowledge
 - A little semantics goes a long way



What is driving the main research in industry?

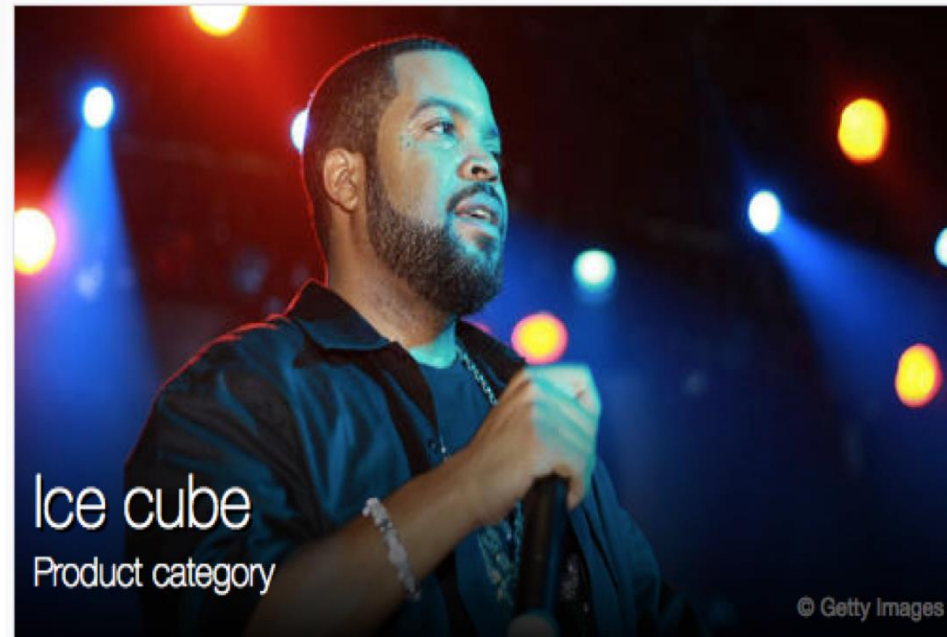
ENTITY SEARCH for advertising

Tetherless World Constellation



Knowledge graphs...

- ... are not perfect
- Or: the importance of human editors



An **ice cube** is a small, roughly **cube**-shaped piece of **ice** (frozen water), conventionally used to cool beverages. **Ice** cubes are sometimes preferred over crushed **ice** because they melt more slowly; they are standard in... [wikipedia.org](https://en.wikipedia.org)

Born: June 15, 1969 (age 44), [Los Angeles, California, USA](#)

Height: 5' 7" (1.73m)

Spouse: [Kimberly Woodruff](#) (m. 1992-present)

Parents: [Doris Benjamin](#), [Hosea Jackson](#)

Children: [Darrel Jackson](#), [O'Shea Jackson Jr.](#), [Shareef Jackson](#), [Karima Jackson](#), [Deja Jackson](#)

[Feedback](#)

Knowledge graphs...

- ... are not perfect
- Or: the importance of human editors

Michelangelo

Artist



Michelangelo di Lodovico Buonarroti Simoni , commonly known as Michelangelo, was an Italian sculptor, painter, architect, poet, and engineer of the High Renaissance who exerted an unparalleled influence on the... [wikipedia.org](https://en.wikipedia.org)

Born: March 6, 1475, [Caprese Michelangelo](#)

Died: February 18, 1564, [Rome](#)

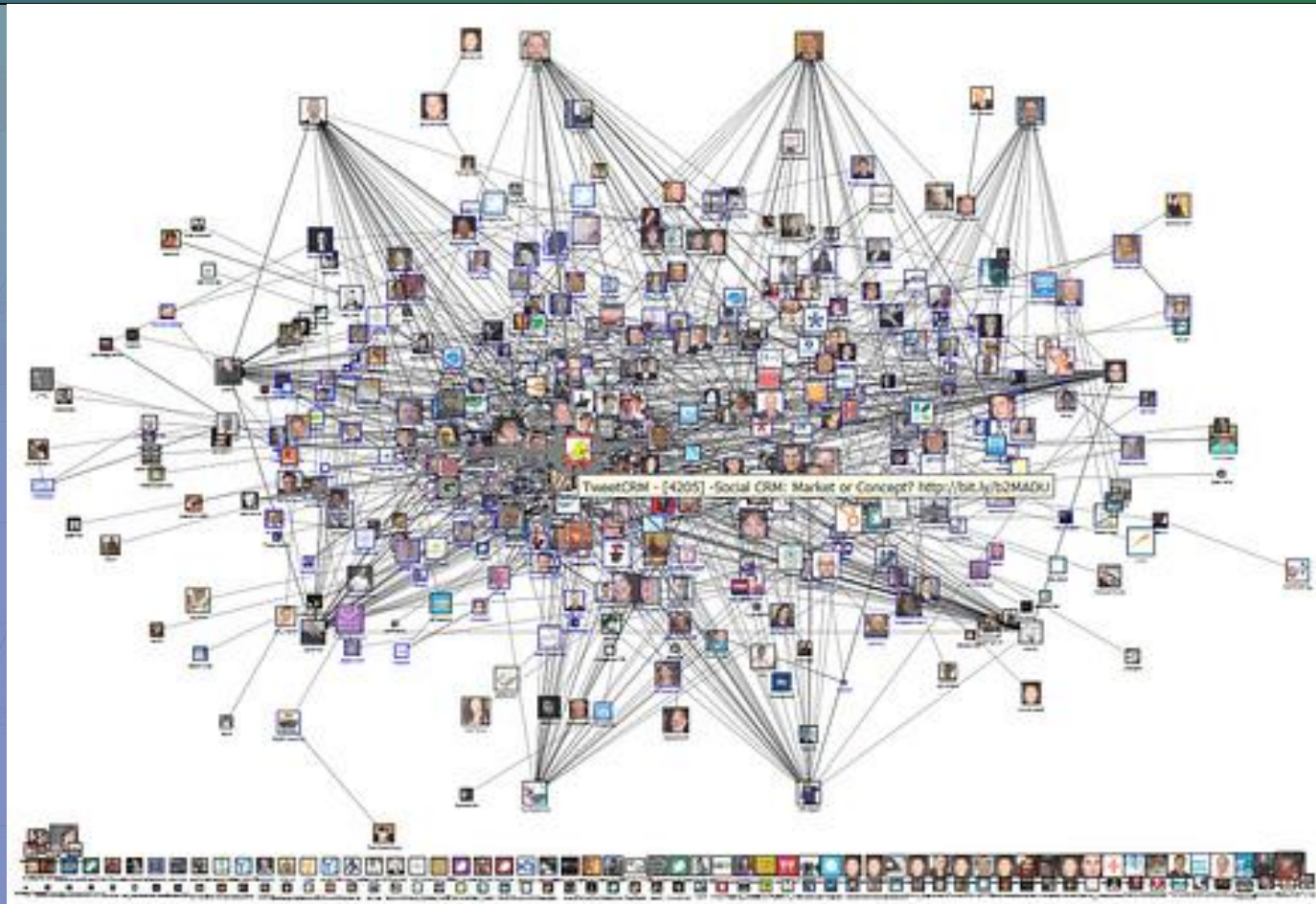
Parents: [Ludovico di Leonardo di Buonarotto Simoni](#), [Francesca di Neri del Miniato di Siena](#)

Feedback



Facebook's open graph protocol (OGP)

Tetherless World Constellation



2012: Extending the Social Graph



TWC

Some others

Tetherless World Constellation

It's not so hard to find out, there are already lots of companies using SW technologies, some of which are:

- New York Times (data.nytimes.com)
- Facebook ([The Open Graph protocol](#))
- Google ([The Knowledge Graph](#))
- IBM ([Watson](#), see also [this interview](#))
- BBC ([World Cup 2010](#))
- Boeing ([Testimonials](#))
- O'Reilly ([Semantic SEO](#))
- Rotten Tomatoes ([Toy Story 3](#))
- Monster.com ([Semantic Search](#))
- Adobe ([XMP](#))
- eBay ([Headphones](#))

link

edited **03 Nov '12, 13:00**

answered **03 Nov '12, 08:08**



fadirra

1.2k ● 3 ● 10

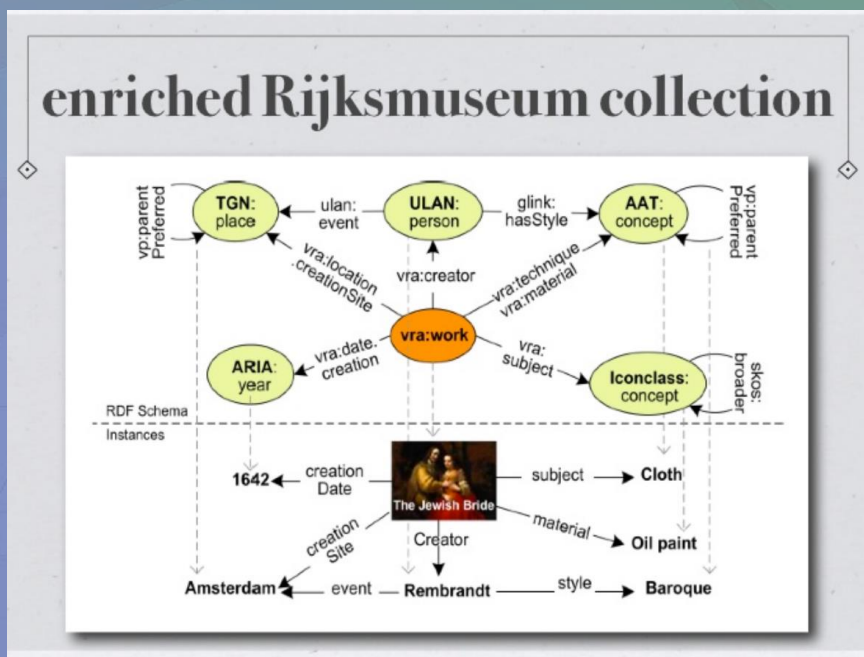
accept rate: 21%

- 1 here is a recent about Semantic Web technologies at BBC: <http://www.cmswire.com/cms/information-management/bbcs-adoption-of-semantic-web-technologies-an-interview-017981.php>

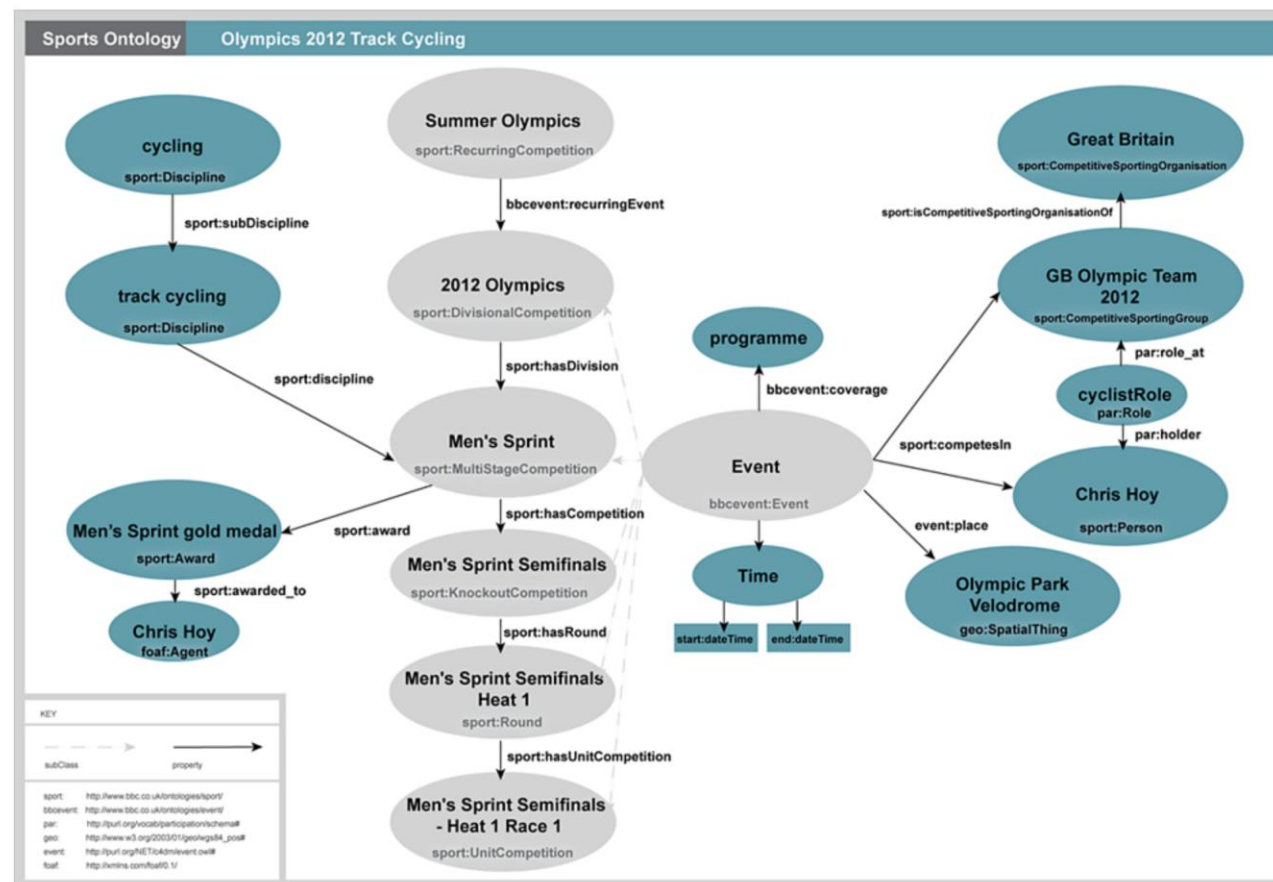


SW technology is deployed around the world in libraries and museums

Tetherless World Constellation



Lora Aroyo, 2011



Many demos 2012 Olympics



But what about rest of the original vision?

Tetherless World Constellation

- Machine-readable markup on the Web exceeds anything we predicted in 2001...
 - And linked-data/KGs without semantics are too weak for many applications
 - “Linked data semantics” and “KG inferencing” are new names for the old issue – we need semantics in data linking!
- ... but the **interoperability** issue that drove the original vision is still challenging
 - Because the story is about METADATA
 - Not just linking!!!

DIVE into data

Discover
Integrate
Validate
Explain

Thinking outside the (database)
box...



Discover

Find datasets and/or content (incl. outside your own organization)

Integrate

Link the relations using meaningful labels: “reconciliation”

Validate

Provide inputs to modeling and simulation systems

Explore

Develop (multimodal) approaches to turn data into actionable knowledge

Metadata lets us DIVE INTO DATA

Discover

Find datasets and/or content (incl. outside your own organization)

Integrate

Link the relations using meaningful labels: “reconciliation”

Validate

Provide inputs to modeling and simulation systems

Explore

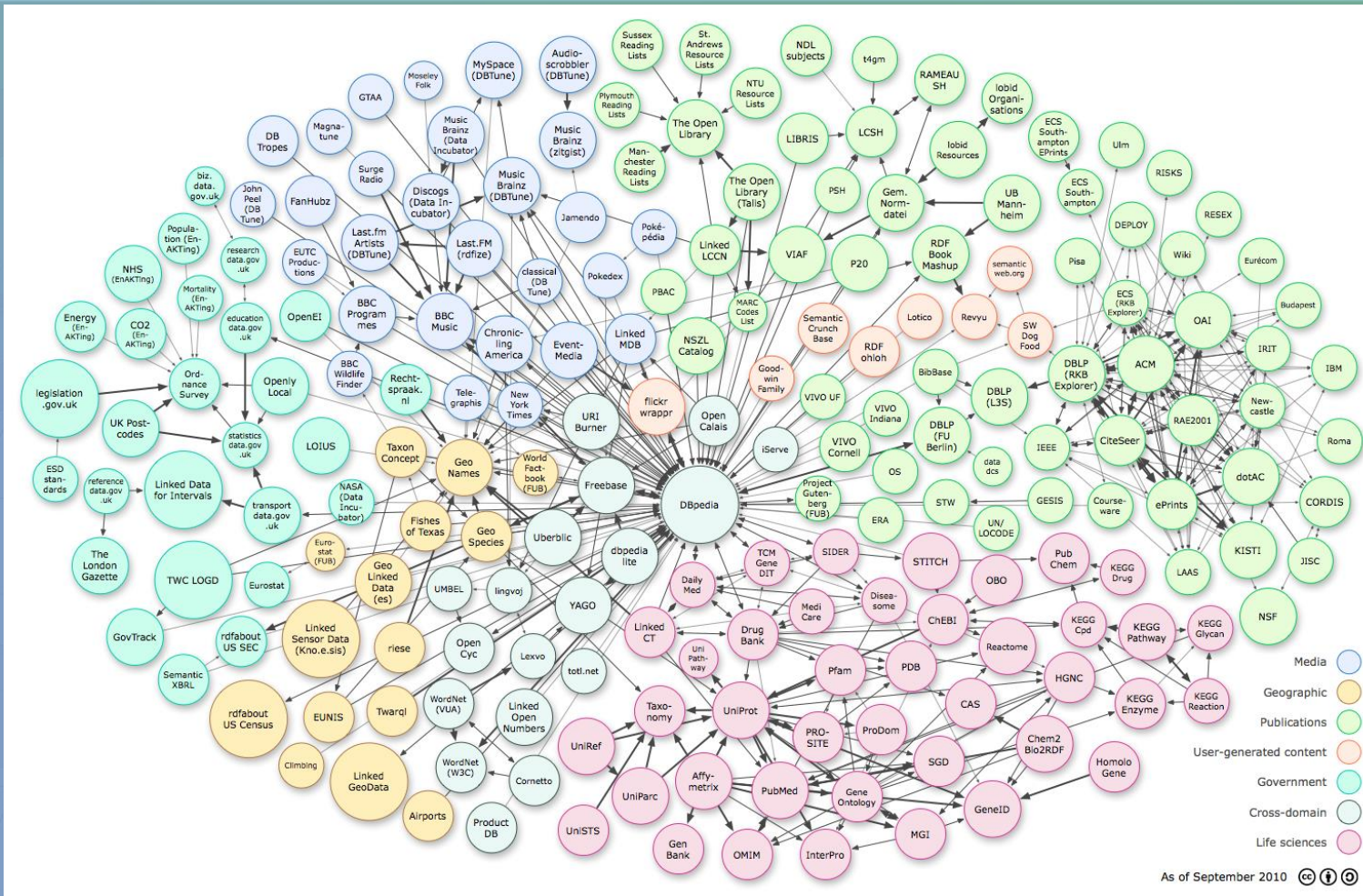
Develop (multimodal) approaches to turn data into actionable knowledge





How does one discover resources in linked data? With great difficulty

Tetherless World Constellation, RPI



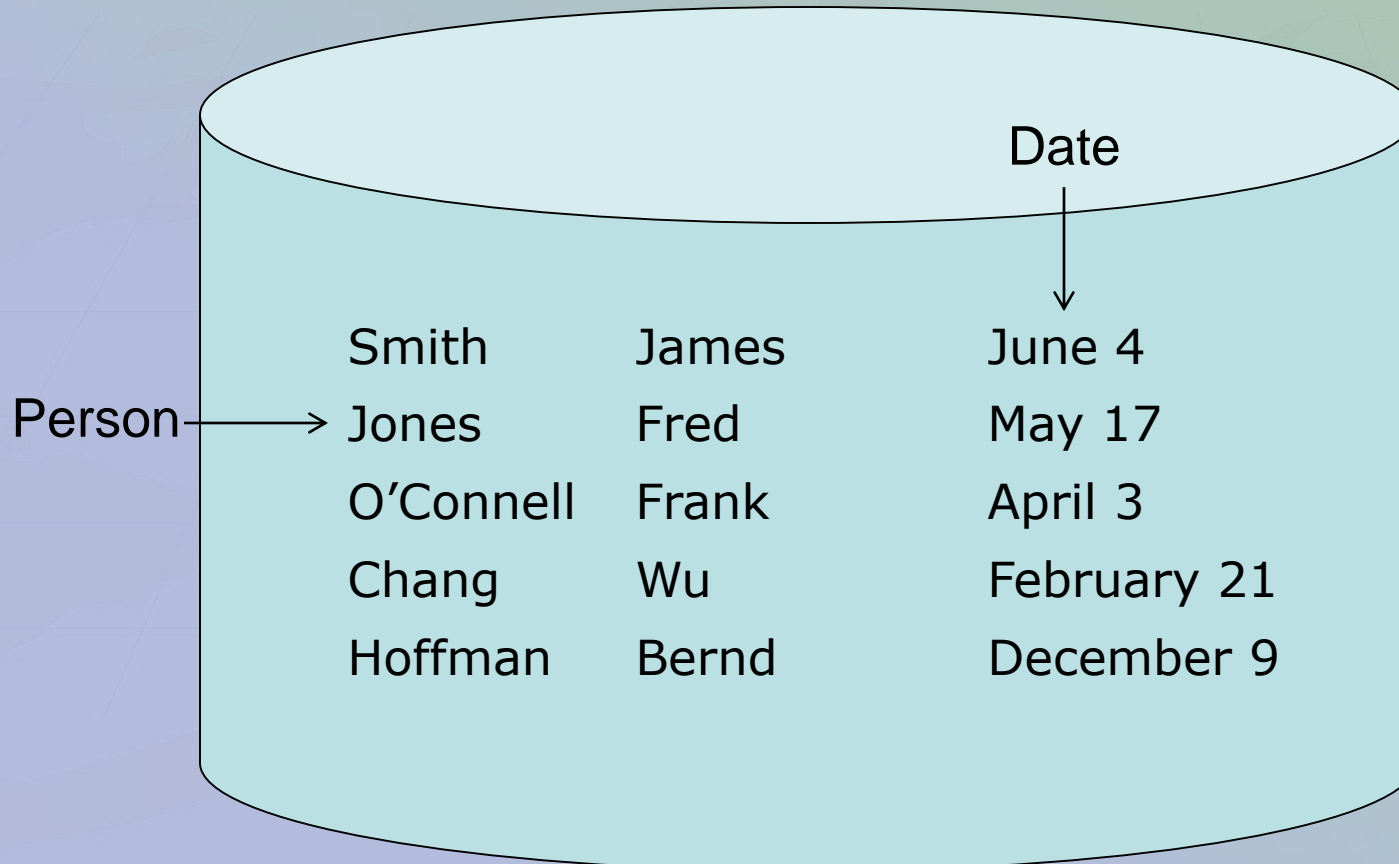


TW/C

Why?

Tetherless World Constellation

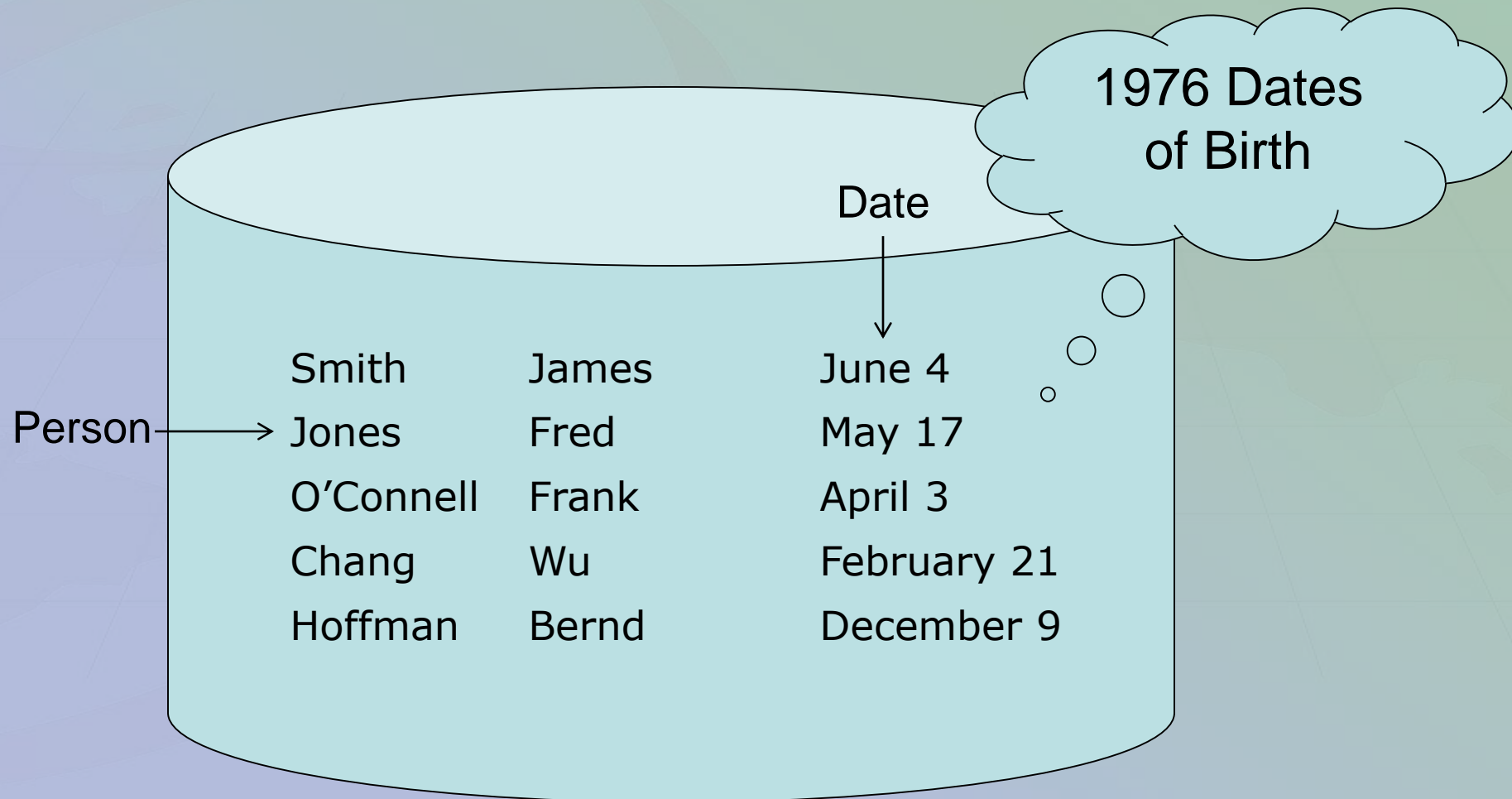
It's not enough just to describe the data elements...





Describing a dataset ... requires a context

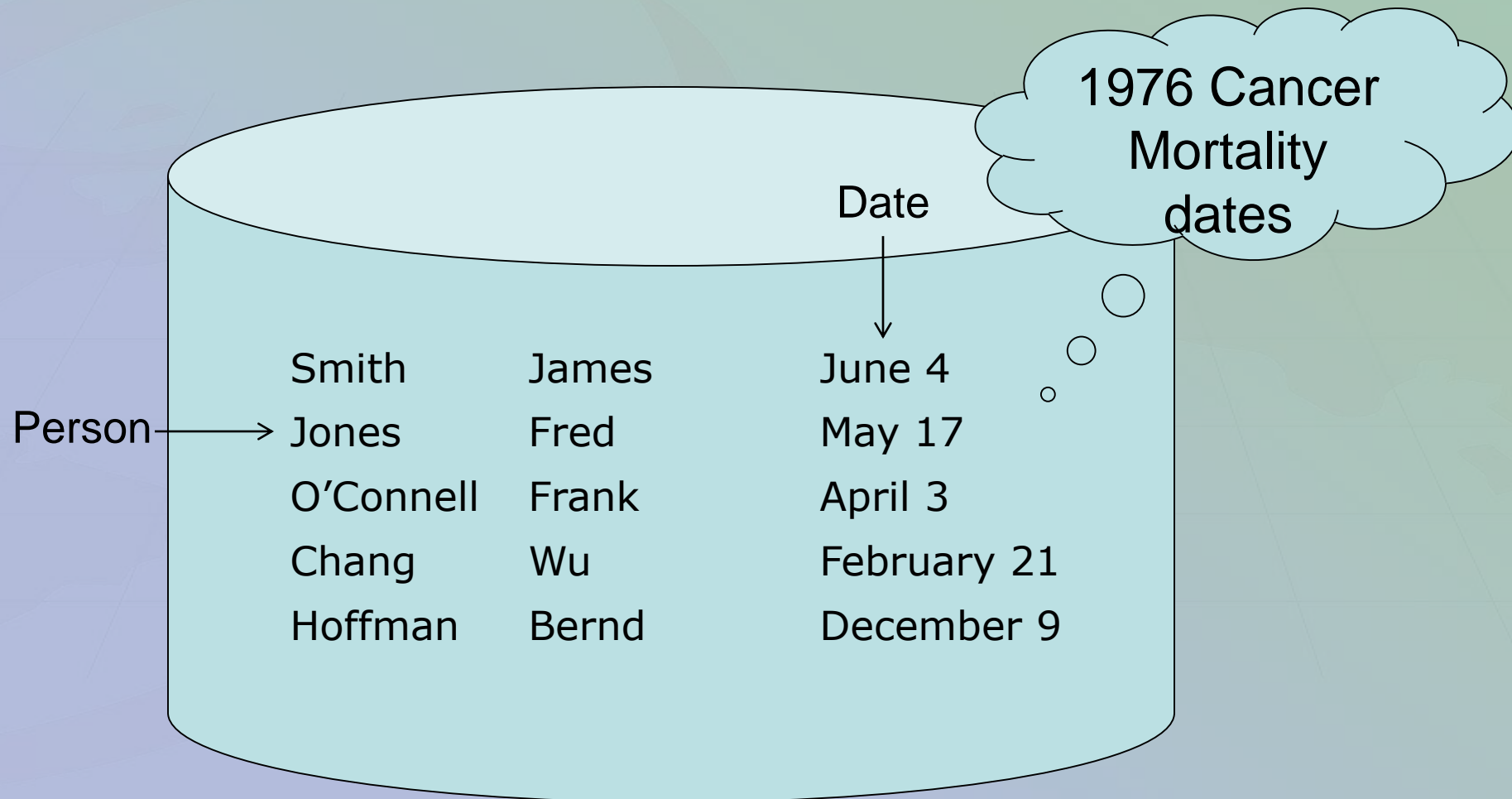
Tetherless World Constellation





Describing a dataset ... requires a context
How do we capture more of this information?

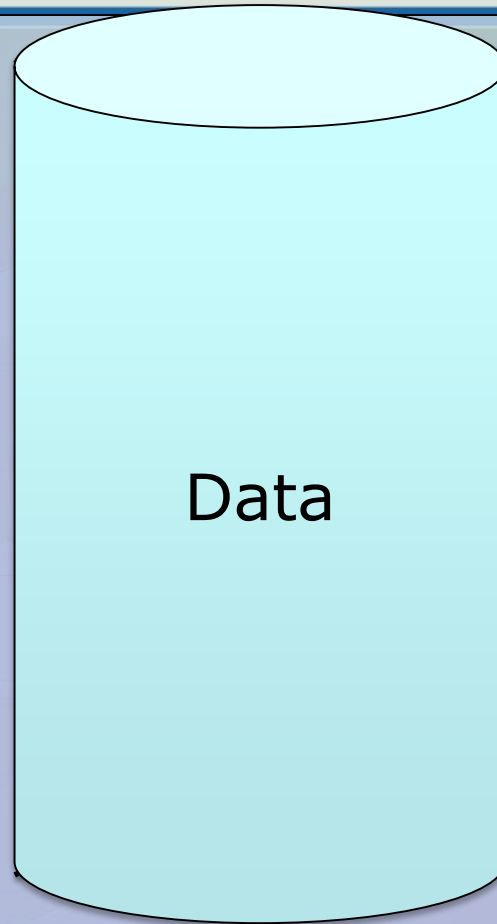
Tetherless World Constellation



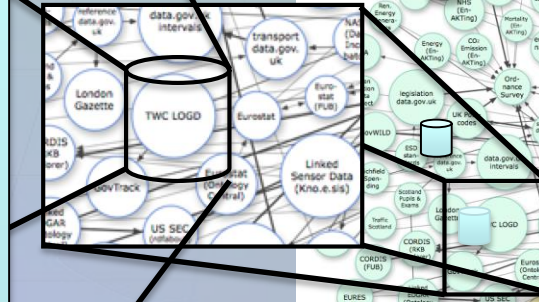


Challenge: Linked Metadata “A Small World of Big Data”

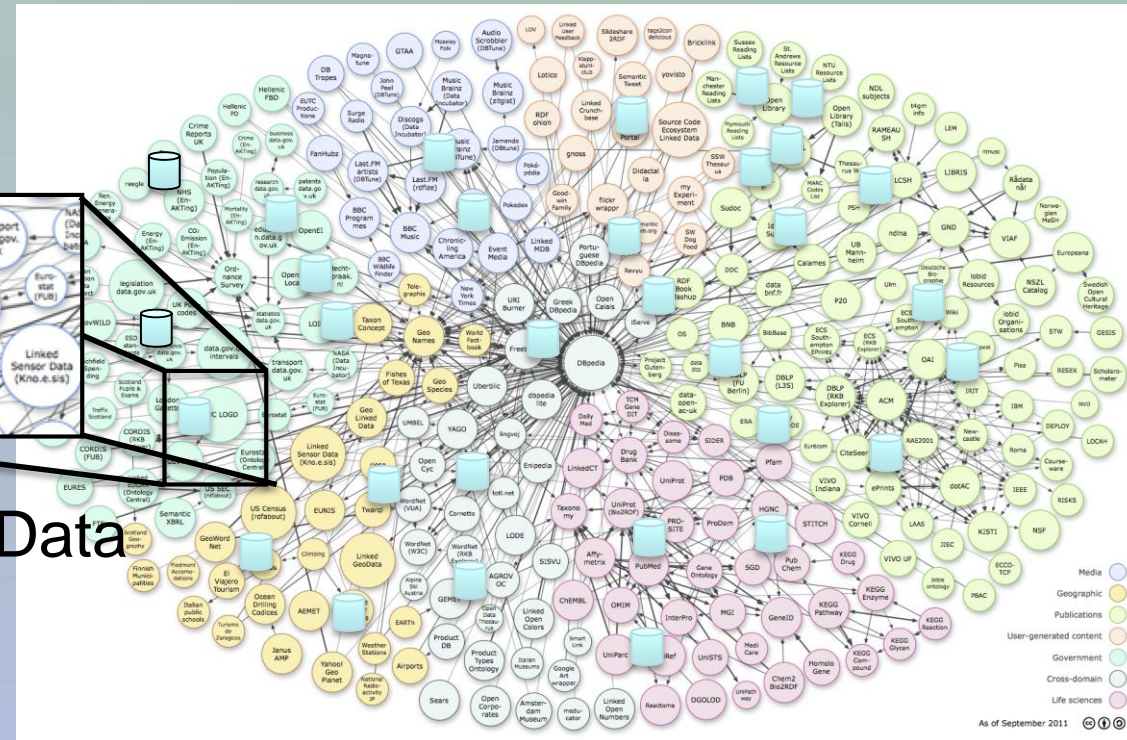
Tetherless World Constellation



Data



MetaData



Linked Semantic Web “metadata” documents can be used to link very large databases in distributed data systems. This leads to orders of magnitude reduction in information flow for large-scale distributed data problems.

Metadata lets us DIVE INTO DATA

Discover

Find datasets and/or content (incl. outside your own organization)

Integrate

Link the relations using meaningful labels: “reconciliation”

Validate

Provide inputs to modeling and simulation systems

Explore

Develop (multimodal) approaches to turn data into actionable knowledge





- Data is converted into RDF and SPARQL'ed
 - creates huge graph DBs less efficient than the original DB
- Data is converted from DB into SPARQL return on demand
 - much better, but you must know the mapping
- owl:sameAs is (ab)used to map data to data
 - but that only lets you map equals – which is an easy mapping to express in many ways
 - defining equality correctly in a model theory is much harder, and thus the abuse, but let's leave that for another talk
- SKOS provides thesaurus-like linking terms
 - But too often it is only used within a single entity

Making Metadata work for integration

- **OWL was based on a model having to do with inference and reasoning**
 - **Not "Big Data" and integration**
- **Proposed solution: rework OWL**
 - **Add key features for the world of linked data**
 - Part-Whole
 - Procedural Attachment
 - Simple temporal relations
 - Ability to close (named) graphs
 - **SEE: <https://www.slideshare.net/jahendler/on-beyond-owl-challenges-for-ontologies-on-the-web>**

Metadata lets us DIVE INTO DATA

Discover

Find datasets and/or content (incl. outside your own organization)

Integrate

Link the relations using meaningful labels: “reconciliation”

Validate

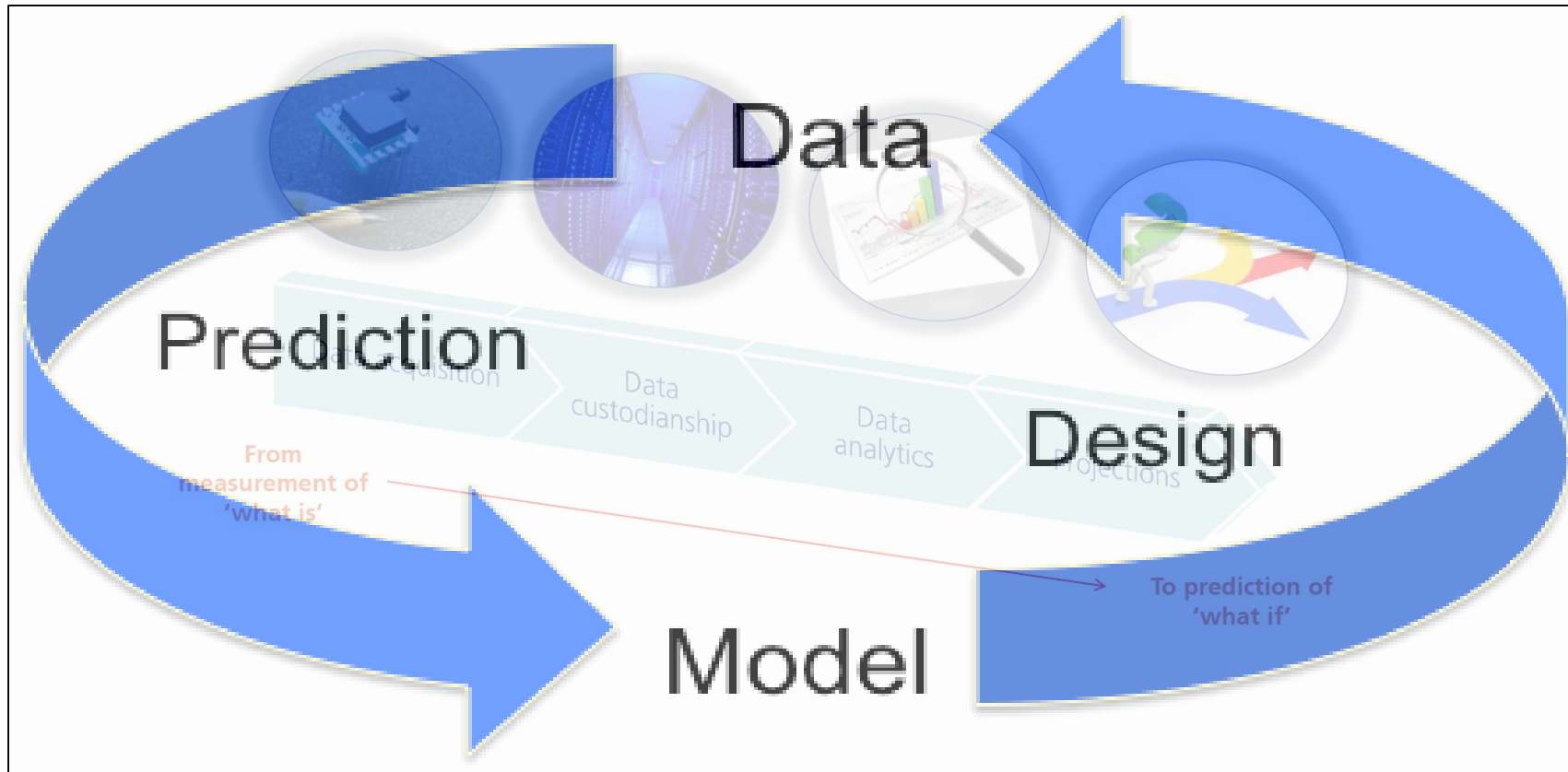
Provide inputs to modeling and simulation systems

Explore

Develop (multimodal) approaches to turn data into actionable knowledge



Metadata can help to "Close the loop"



We have to close the loop between the correlational nature of predictive data analytics (machine learning) and the “causal” models we need for science (natural & social) and engineering!

Metadata lets us DIVE INTO DATA

Discover

Find datasets and/or content (incl. outside your own organization)

Integrate

Link the relations using meaningful labels: “reconciliation”

Validate

Provide inputs to modeling and simulation systems

Explore

Develop (multimodal) approaches to turn data into actionable knowledge





TW/C

Data Exploration

Tetherless World Constellation, RPI



Challenge: Can metadata help us explore/understand the data (structured or unstructured) in archives, collections, datasets, etc?



Agent: The 2008 Summer Olympic Games, officially known as the Games of the XXIX Olympiad and commonly known as Beijing 2008, was a major international multi-sport event that took place in Beijing, China, from August 8 to 24, 2008.

User: What can you tell me about the Beijing National Aquatics Center?

Agent: The Beijing National Aquatics Center, also officially known as the National Aquatics Center, and colloquially known as the Water Cube, is an aquatic center that was built alongside Beijing National Stadium in the Olympic Green for the swimming competitions of the 2008 Summer Olympic Games.

User: Tell me more about the Beijing National Stadium.

Agent: Beijing National Aquatics Center is east of Beijing National Stadium.

User: What is Beijing National Stadium?

Agent: The Beijing National Stadium, officially the National Stadium, also known as the Nest of Bird, is a stadium in Beijing, China. The stadium was designed for use throughout the 2008 Summer Olympic Games and Paralympics. The stadium is currently used mostly for football matches.

User: |

Submit

- Agent-based guide tells user about information on a tourist site
- User can ask questions about it

Using game-based story-telling approach and limited language

First demo covered ~500 topics re: 2008 Beijing Olympics

2008 Summer Olympic Games	Olympic Green Convention Center	Venues outside Beijing	(Fencing) Women's Sabre	(Shooting) Men's skeet	(Track and Field) Women's 100 m hurdles	(Track and Field) Men's 4x400 m Relay	Carmelo Anthony
Aquatics	Olympic Green Hockey Stadium	Promenade	Michael Phelps	Ian Crocker*	Misty May-Treanor	Alyson Felix	Carlos Boozer
Archery	Beijing Science and Technology University Gymnasium	Beijing Olympic Village	(Swimming) Men's 400 m Individual Medley	Mark Gangloff*	Keri Walsh	Natasha Hastings*	Chris Bosh
Athletics	Beijing University of Technology Gymnasium	Digital Beijing Building	Nathan Adrian*	Brendan Hansen	(Beach Volleyball) Women's Tournament	Monique Henderson	Kobe Bryant
Badminton	China Agricultural University Gymnasium	Ling Long Pagoda	Matt Grevers*	Ganett Weber-Gale*	LaShawn Merritt	Sanya Richards	Dwight Howard
Baseball	Peking University Gymnasium	Olympic Park Observation Tower	Cullen Jones	(Swimming) Men's 4x100 m Medley Relay	(Track and Field) Men's 400 m	Mary Wineberg	LeBron James
Basketball	Beijing University of Aeronautics and Astronautics Gymnasium	Olympic Green Venues	Jason Lezak	Serena Williams	Nicole Barnhart	(Track and Field) Women's 4x400 m Relay	Jason Kidd
Bowling	Beijing Institute of Technology Gymnasium	Beihai	Ganett Weber-Gale	Venus Williams	Shannon Boxx	Simone Augustus	Chris Paul
Canoe/Kayak	Beijing Shooting Range Hall	Jingling	Ben Wildman-Tobiner*	(Tennis) Women's Doubles	Rachel Buehler	Sue Bird	Tayshaun Prince
Cycling	Laoshan Velodrome	Huanhuan	(Swimming) Men's 4x100 m Freestyle Relay	Erin Cafaro	Lori Chalupny	Tamika Catchings	Michael Redd
Equestrian	Shunyi Olympic Rowing-Canoeing Park	Yingying	(Swimming) Men's 200 m Freestyle	Anna Cummins	Lauren Cheney	Sylvia Fowles	Dayane Wade
Fencing	Wukesong Indoor Stadium	Nini	Natalie Coughlin	Caryn Davies	Stephane Cox	Kara Lawson	Deron Williams
Football	Beijing Shooting Range Clay Target Field	Fuwa	(Swimming) Women's 100 m Backstroke	Susan Francia	Tobin Heath	Lisa Leslie	United States Men's National Basketball Team
Gymnastics	Capital Indoor Stadium	Mascots	Aaron Peirsol	Anna Goodale	Angela Hucles	DeLisha Milton-Jones	(Basketball) Men's Tournament
Handball	Fengtai Sports Center Softball Field	Beijing	(Swimming) Women's 100 m Backstroke	Caroline Lind	Natasha Kai	Candace Parker	Gold Medalists
Hockey	Laoshan Mountain Bike Course	Shanghai	Walton Eller	Elio Logan	Carl Lloyd	Cappie Pondexter	Aug. 9th, 2008
Judo	Olympic Sports Center Stadium	Qingdao	(Shooting) Men's Double Trap	Lindsay Shoop	Kate Markgraf	Katie Smith	Aug. 10th, 2008
Modern Pentathlon	Olympic Sports Center Gymnasium	Hong Kong	(Swimming) Men's 200 m Butterfly	Mary Whipple	Heather Mills	Diana Taurasi	Aug. 11th, 2008
Rowing	Worker's Stadium	Qinghuangdao	Ricky Berens	(Rowing) Women's Eight	Heather O'Reilly	Tina Thompson	Aug. 12th, 2008
Sailing	Worker's Gymnasium	Tianjin	Kristie Keller*	Stephanie Brown Traflet	Christie Rampone	United States Women's National Basketball Team	Aug. 13th, 2008
Shooting	Ying Tung Natatorium of National Olympic Sports Center	Shenyang	Ryan Lochte	(Track and Field) Women's Discus Throw	Amy Rodriguez	(Basketball) Women's Tournament	Aug. 14th, 2008
Softball	Chaoyang Park Beach Volleyball Ground	Host City	Peter Vanderkaay	Angelo Taylor	Hope Solo	Lloy Ball	Aug. 15th, 2008
Table Tennis	Laoshan Bicycle Moto Cross Venue	Go Host Cities	Erk Vendt*	(Track and Field) Men's 400 m Hurdles	Lindsay Tarpley	Gabe Gardner	Aug. 16th, 2008
Taekwondo	Cities	Venues	David Walters*	Laura Kraut	Aly Wagner	Kevin Hansen	Aug. 17th, 2008
Tennis	Urban Road Cycling Course	You and Me	(Swimming) Women's 4x200 m Freestyle Relay	Beezie Madden	United States Women's National Soccer Team	Tom Hoff	Aug. 18th, 2008
Triathlon	Wukesong Sports Center Baseball Field	Artihem	Kristin Armstrong	Will Simpson	(Football) Women's Tournament	Rich Lambourne	Aug. 19th, 2008
Volleyball	Hong Kong Equestrian Venues	One World, One Dream	(Cycling) Women's Time Trial	McLain Ward	Phil Dahsauer	David Lee	Aug. 20th, 2008
Weightlifting	Qingdao International Sailing Center	Motto	Rebecca Soni	(Equestrian) Team Jumping	Todd Rogers	Ryan Miller	Aug. 21st, 2008
Wrestling	Qinghuangdao Olympic Sports Center Stadium	Aug. 8th, 2008	(Swimming) Women's 200 m Breaststroke	Anna Tunnicliffe	(Beach Volleyball) Men's Tournament	Reid Priddy	Aug. 22nd, 2008
Sports	Shanghai Stadium	Aug. 24th, 2008	(Swimming) Men's 200 m Backstroke	Henry Cejudo	Bryan Clay	Sean Rooney	Aug. 23rd, 2008
Beijing National Aquatics Center	Shenyang Olympic Sports Center Stadium	Opening Ceremony	(Swimming) Men's 200 m Individual Medley	(Sailing) Women's Laser Radial Class	(Track and Field) Men's Decathlon	Riley Salmon	Gold Sports
Beijing National Indoor Stadium	Tianjin Olympic Center Stadium	Closing Ceremony	Nastia Lukin	(Wrestling) Men's Freestyle 55 kg	Keron Clement*	Clay Stanley	United States Olympic Committee (USOC)
Beijing National Stadium	Other Venues	Time	(Gymnastics) Women's Artistic Individual All-Around	Shawn Johnson	David Neville	Scott Touzinsky	Male
Olympic Green Tennis Center	Olympic Green	Martel Zagunis	(Swimming) Men's 100 m Butterfly	(Gymnastics) Women's Balance Beam	Jeremy Wariner	United States Men's National Volleyball Team	Female
Olympic Green Archery Field	Universities Gymnasiums	Vincent Hancock	Dawn Harper	Reggie Witherspoon*	(Volleyball) Men's Tournament		

- These include venues, sports, athletes, special events, and date information
 - topics can be linked along these dimensions (*ie.* same sport, same date, same place, etc.)

Agent: The 2008 Summer Olympic Games, officially known as the Games of the XXIX Olympiad and commonly known as Beijing 2008, was a major international multi-sport event that took place in Beijing, China, from August 8 to 24, 2008.

User: I remember watching the 2008 Swimming events. Tell me about Michael Phelps

Agent: Michael Fred Phelps II is a competition swimmer and the most decorated Olympian of all time, with a total of 22 medals. Phelps also holds the all-time records for Olympic gold medals, Olympic gold medals in individual events, and Olympic medals in individual events for a male. In winning eight gold medals at the 2008 Summer Olympic Games, Phelps took the record for the most first-place finishes at any single Olympic Games. Five of those victories were in individual events, tying the single Games record. In the 2012 Summer Olympic Games in London, Phelps won four golds and two silver medals, making him the most successful athlete of the Games for the third Olympics in a row.

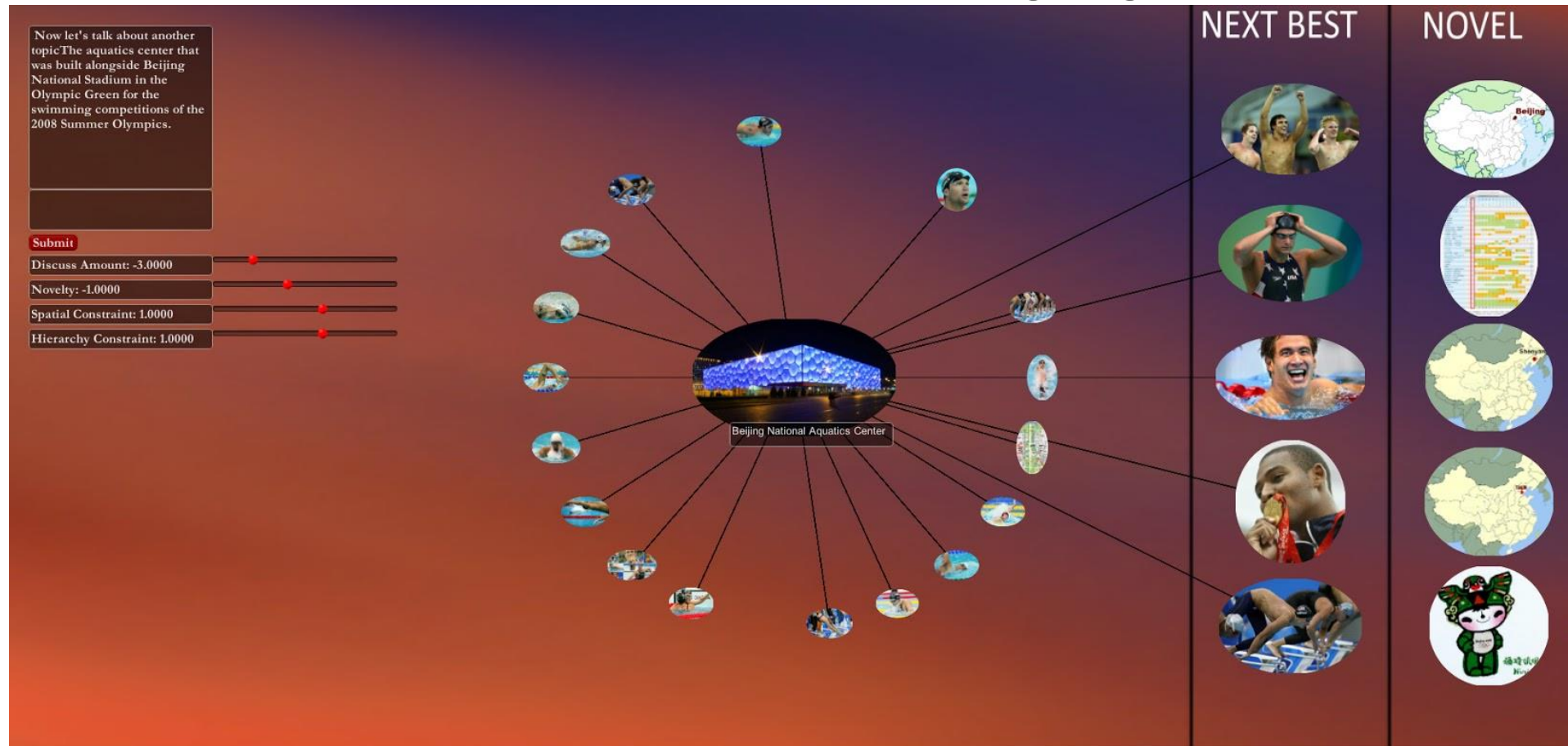
User:

Submit



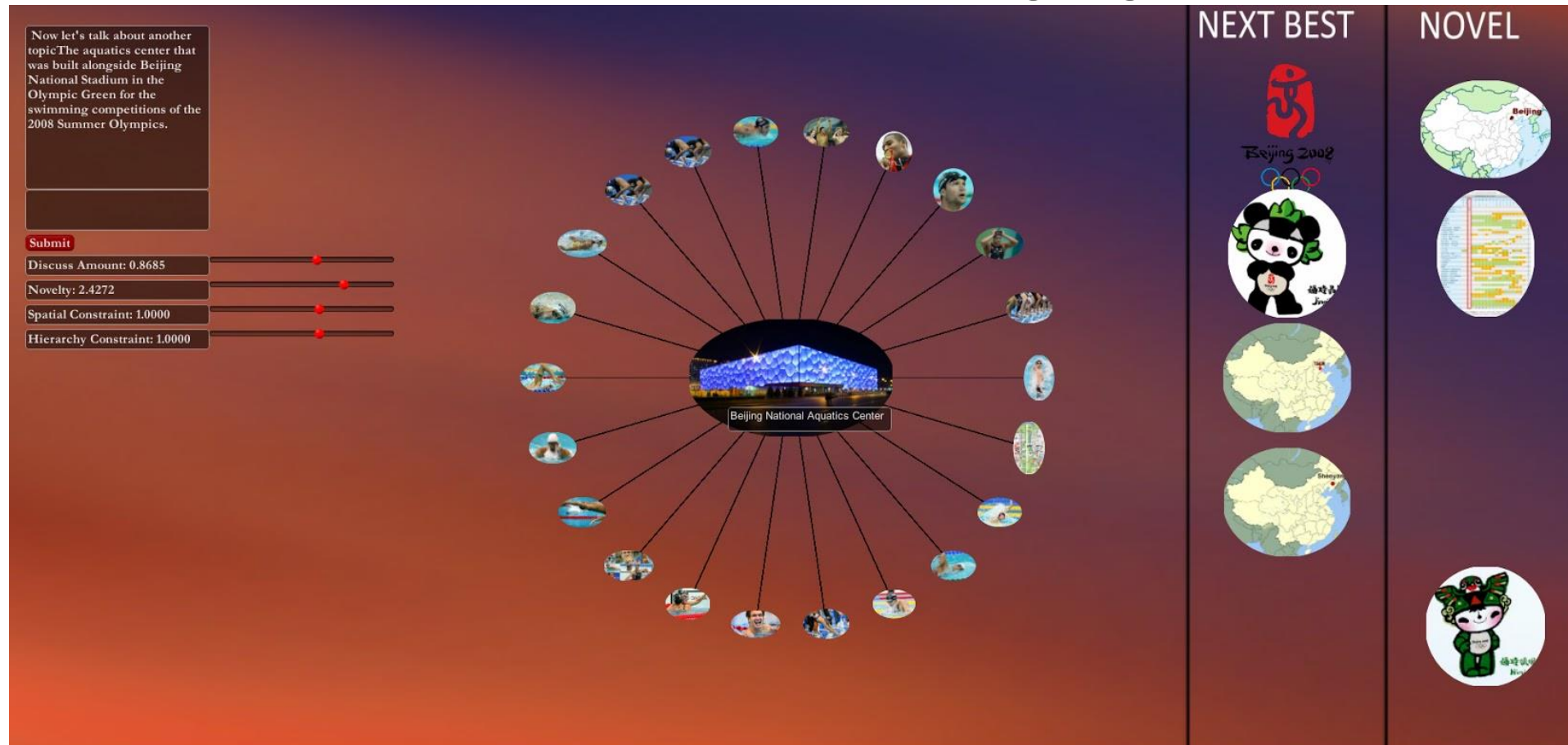
- Example: Michael Phelps

Under the hood – a knowledge graph



- Agent reasons about a network of connected entities and chooses the best next one to discuss based on a combination of factors
 - Consistency, Novelty and Ontological relationships

Under the hood – a knowledge graph



- Different settings of these factors can lead to different presentation order.
 - Settings can be pre-determined for pedagogical or application specific needs
 - Agent can learn users' interests and preferences over time, adjusting the settings

- *Automate Agent Creation* combining:

Information Extraction: Using anew "living information extraction" technique, we will be able to create a "never-ending extractor" which will be pulling from web documents information about entities and events, and the relationships between them.

The new system can work in a dynamic node, and does not need human annotated samples for training, but it works best if there are a number of known relationships between pages to build off of.

- *Automate Agent Creation* combining:

Information Extraction: Using a new "living information extraction" technique we will extract documents, events, them.

The new node, a sample there a between

Semantic Web: The Semantic Web provides a number of known relationships between pages on the Web in a number of domains. Using general knowledge sources, like dbpedia and Yago, and specialized knowledge sources, like the data from musicbrainz, the reviews from Yelp (which have semantic annotations) and even the Open Graph of Facebook (which is available in a semantic web format), provides a jumpstart for the language extraction.

However, the Semantic Web relates pages, but doesn't have any sort of "understanding" of what is on the pages.

- *Automate Agent Creation* combining:

Information Extraction
"living information"
we will be able to
extractor" which
documents information
events, and the
them.

The new system
node, and does
samples for training
there are a number
between pages,
language
However
but does
"unders

Semantic
provides
between
domains
sources
specializ
data from
Yelp (wh
and eve
(which i
format)
language
However
but does
"unders

Cognitive Computing : Cognitive Computing, can allow us to have a better way of accessing information about the entities found on the Web and finding other information about the same entities using various kinds of search and language heuristics. This allows us to have more organized information, rapidly generated, about the entities being explored.

However, given a large graph of entities (even the organized linked-open data cloud has information about billions of things), how do we choose what to display next? If the best we can do is provide links, all of the above isn't much better than choosing a page and clicking from there.

- *Automate Agent Creation* combining:

Information Extraction: Us
"living information extraction
we will be able to create a
extractor" which will be pulling
documents information about
events, and the relationship
them.

The new system can work in
node, and does not need huge
samples for training, but it works
there are a number of knowl
between pages to build off of
language extraction
However, the Semantic
but doesn't have an
"understanding" of what is on the pages.

Story-telling Technology: Interactive
storytelling techniques are being explored
to take information in the kind of
"knowledge graph" resulting from the
above, and tailoring the presentation to a
user using storytelling techniques. It is
aimed at presenting the information as an
interesting and meaningful story by taking
into consideration a combination of factors
ranging from topic consistency and novelty,
to learned user interests and even a user's
emotional reactions. The system can
essentially determine "where to go next"
and what to do there in the organized
information as processed above..

e.g. Turning University Archive Metadata into narrative “tours” (Mei Si, 2017)

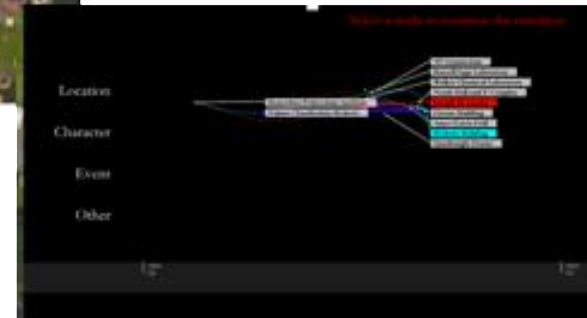
Archive Metadata includes:



People



Places & Events



And links between them

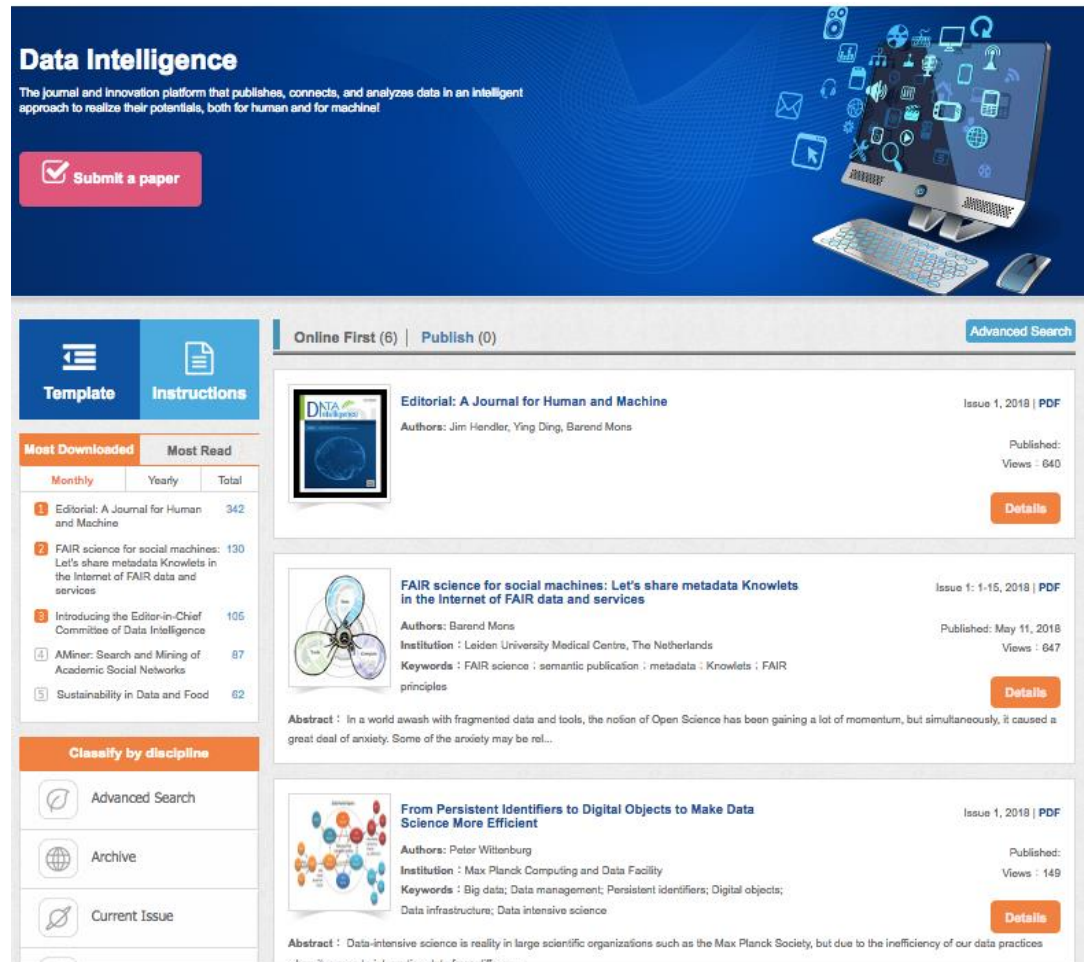




- The long-standing goal of the “Semantic Web initiative” is to create metadata
 - this was a key part of the original vision
- But more and more we need to start to focus more on enriching the metadata
 - To help provide more semantics for the data linking
 - For discovery
 - For integration
- And to use the metadata in new and exciting ways
 - Using the Semantics to power new technologies
 - Closing the loop between correlation and causation research
 - Providing new ways of interacting with the metadata collections

You can help: share your solutions

New Journal: Data Intelligence



Data Intelligence
The journal and innovation platform that publishes, connects, and analyzes data in an intelligent approach to realize their potentials, both for human and for machine!

[Submit a paper](#)

Online First (6) | Publish (0) | [Advanced Search](#)

Editorial: A Journal for Human and Machine
Issue 1, 2018 | PDF
Authors: Jim Hendler, Ying Ding, Barend Mons
Published: Views : 640
[Details](#)

FAIR science for social machines: Let's share metadata Knowlets in the internet of FAIR data and services
Issue 1: 1-15, 2018 | PDF
Authors: Barend Mons
Institution : Leiden University Medical Centre, The Netherlands
Keywords : FAIR science : semantic publication : metadata : Knowlets : FAIR principles
Published: May 11, 2018
Views : 647
[Details](#)
Abstract : In a world awash with fragmented data and tools, the notion of Open Science has been gaining a lot of momentum, but simultaneously, it caused a great deal of anxiety. Some of the anxiety may be rel...

From Persistent Identifiers to Digital Objects to Make Data Science More Efficient
Issue 1, 2018 | PDF
Authors: Peter Wittenburg
Institution : Max Planck Computing and Data Facility
Keywords : Big data; Data management; Persistent identifiers; Digital objects; Data infrastructure; Data intensive science
Published: Views : 149
[Details](#)
Abstract : Data-intensive science is reality in large scientific organizations such as the Max Planck Society, but due to the inefficiency of our data practices when it comes to integrating data from differen...

Most Downloaded | **Most Read**

	Monthly	Yearly	Total
1	Editorial: A Journal for Human and Machine		342
2	FAIR science for social machines: Let's share metadata Knowlets in the internet of FAIR data and services		130
3	Introducing the Editor-in-Chief Committee of Data Intelligence		105
4	AMiner: Search and Mining of Academic Social Networks		87
5	Sustainability in Data and Food		62

Classify by discipline

- [Advanced Search](#)
- [Archive](#)
- [Current Issue](#)

Joint publication of MIT Press and Chinese Academy of Science

Metadata for libraries and collections is one of the topics we are interested in, **please consider submitting a paper!**

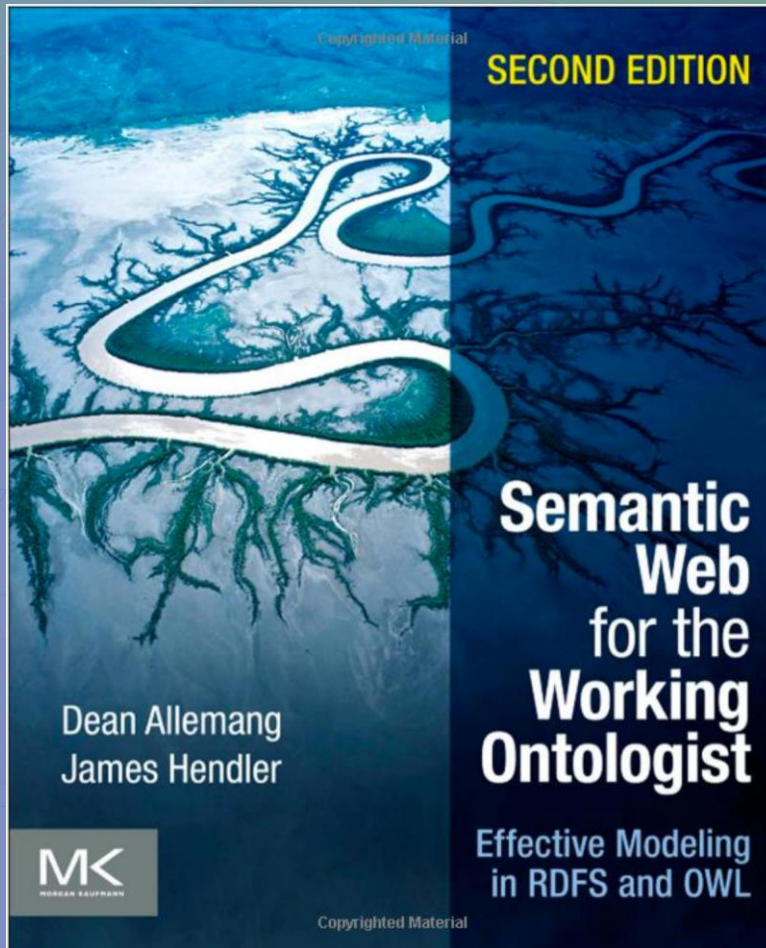
<http://www.data-intelligence-journal.org/>



TWOC

Questions?

Tetherless World Constellation



By the way:

We are working on a 3rd edition (different publisher) with more explicit linked data coverage (and a third author, Fabien Gandon)

+ please send me comments/thoughts/etc.

(email address at <http://www.cs.rpi.edu/~hendler>)