USC Viterbi
School of Engineering

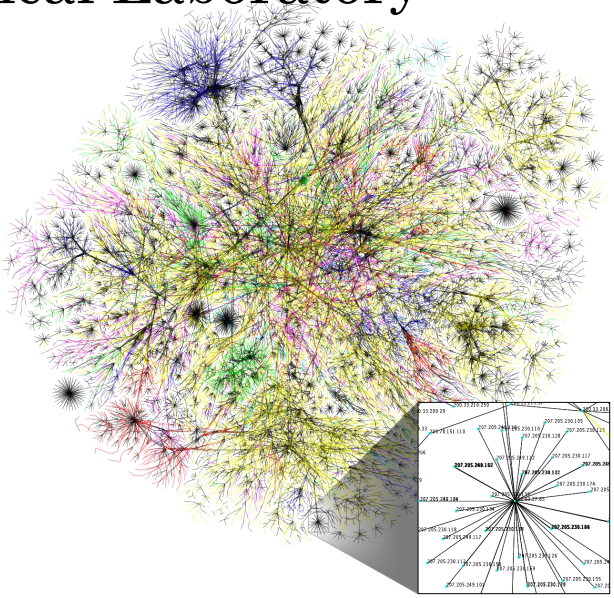
Linked Data

Craig Knoblock

DSCI 558: Building Knowledge Graphs

The Internet

- 1962, Rand Corporation, communications systems for military
- 1965, Packet switching at the National Physical Laboratory
- 1969, ARPANET
- 1971, First Email
- 1981, CSNET & BITNET are born
- 1983, DNS is born
- 1983, TCP/IP is born
- 1985, FTP is standardized

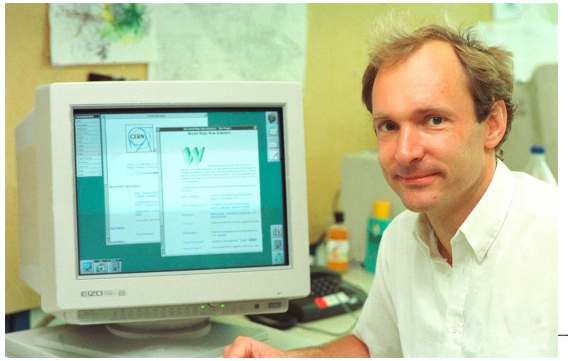


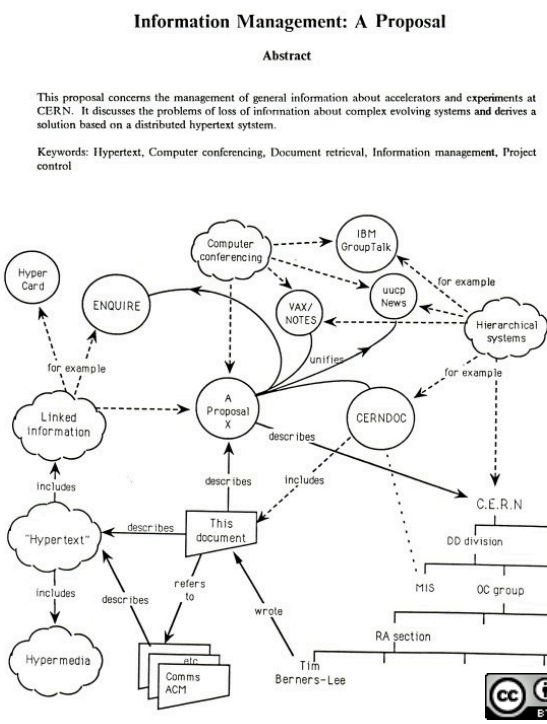
slide by Basel Shbita, picture by the Opte Project https://en.wikipedia.org/wiki/History_of_the_Internet#/media/File:Internet_map_1024_-_transparent,_inverted.png

The World Wide Web

- 1980s, Tim Berners-Lee @ CERN
- 1989, HTTP
- 1990, WWW Proposal
- 1991, HTML
- 1993, Mosaic (NCSA), 1st Web Browser

- *Since*, Web of Documents
- *Now*, Web of Data





Tim Berners-Lee, CERN/DD

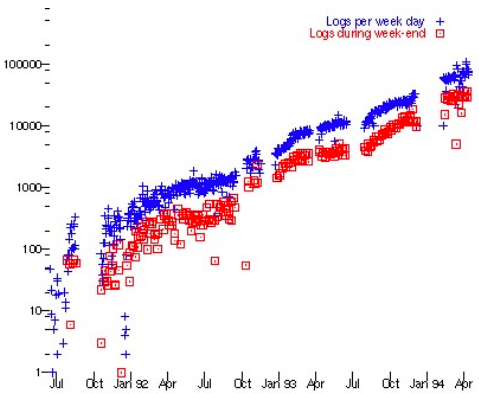
March 1989

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 sytstem.

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

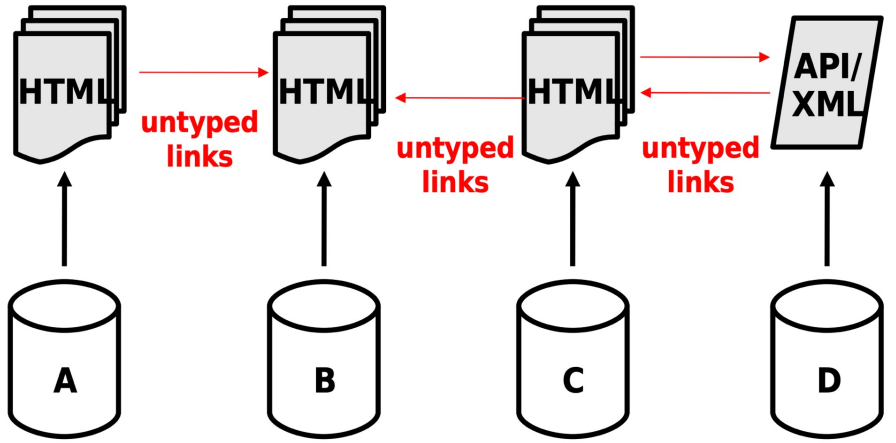


https://www.w3.org/History/1989/proposal.html

slide by Basel Shbita, content from https://www.w3.org/History.html

The Web of Documents

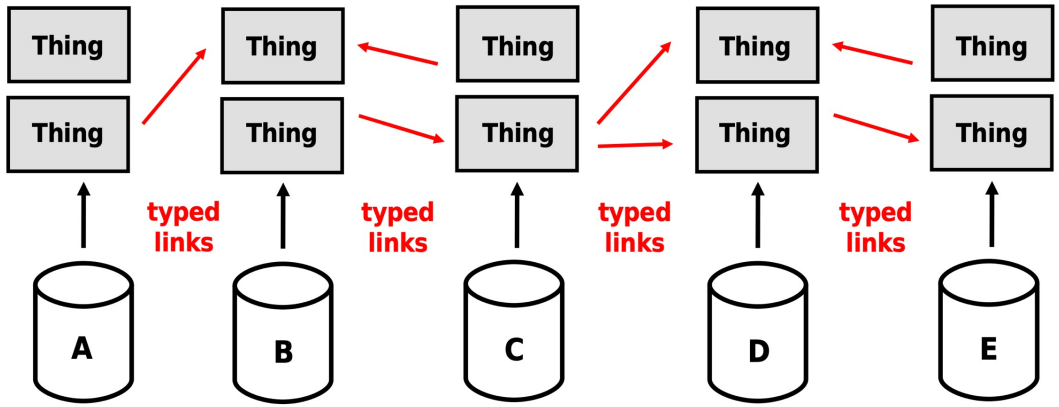
- Analogy
  - a global filesystem
- Primary objects
  - documents
- Links between
  - documents (or sub-parts of)
- Degree of structure in objects
  - fairly low
- Semantics of content and links
  - implicit
- Designed for
  - human consumption



Disconnected Data ☹

The Web of Data

- Analogy
  - a global database
- Primary objects
  - things (or descriptions of things)
- Links between
  - things (including documents)



- Degree of structure in (descriptions of) things
  - high
- Semantics of content and links
  - explicit
- Designed for
  - machines first, humans later

slide by Christian Bizer, Tom Heath, Tim Berners-Lee: Linking Open Data (2/19/2008)

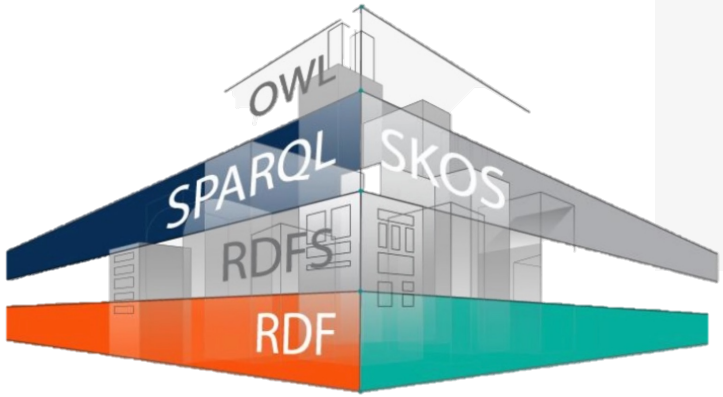
Linked Data

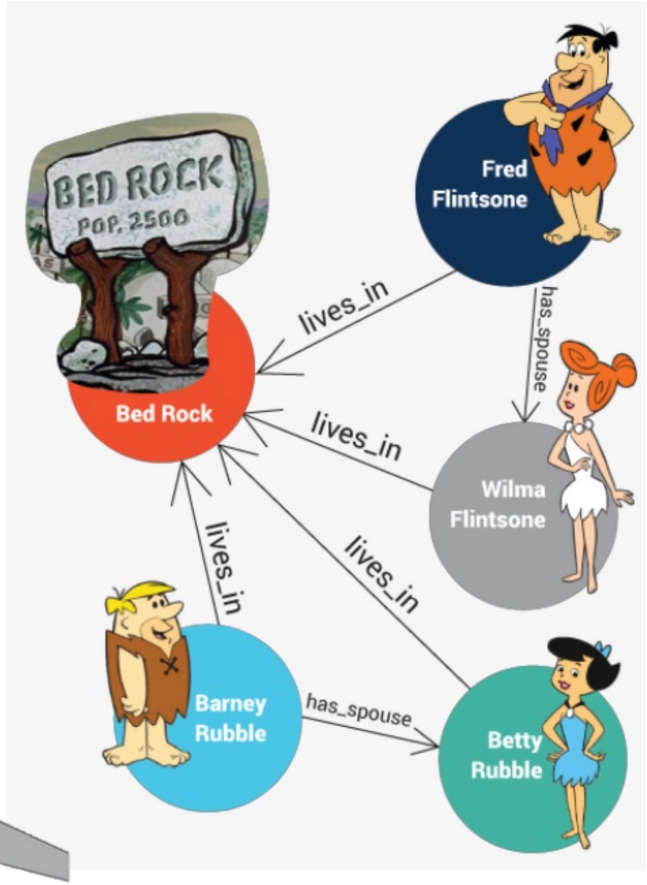
- A method of publishing structured data so that it can be interlinked and become more useful
- Web technologies:
  - HTTP
  - URIs
  - RDF

  to share information → processed automatically by computers

Semantic Web

- W3C, extension of the WWW through new standards
- Encoding of semantics with the data
- Emergence of Ontologies
  - FOAF
  - OWL

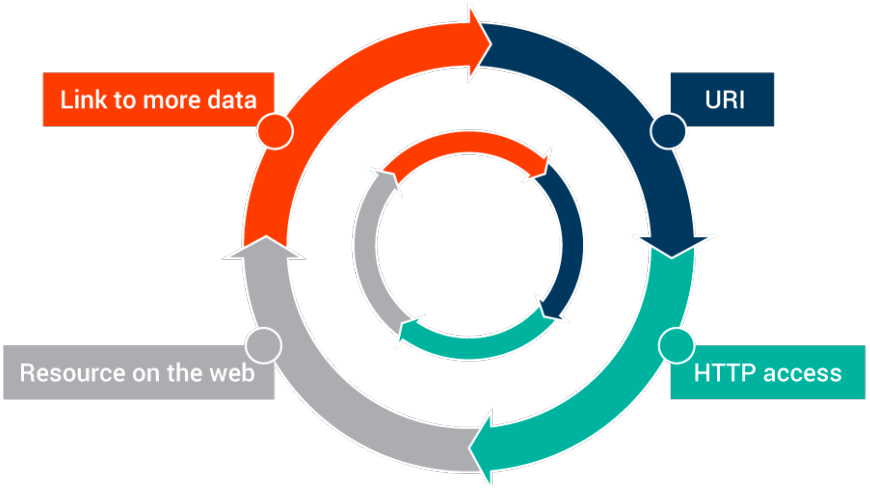




slide by Basel Shbita, images from https://www.ontotext.com/knowledgehub/fundamentals/what-is-the-semantic-web/

Linked Open Data

- Linked Data + Open Data
  - DBpedia
  - GeoNames
  - Wikidata
- Present day "**Knowledge Graphs**"
  - across vast amount of general importance
  - alive LOD
  - graph-computing techniques and algorithms



slide by Basel Shbita, image from https://www.ontotext.com/knowledgehub/fundamentals/what-is-the-semantic-web/

Linked Data Principles

- Use URIs as names for things
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
- Include links to other URIs so that they can discover more things

http://youtu.be/OM6XIICm_qo

Tim Berners-Lee
http://www.w3.org/DesignIssues/LinkedData.html

slide by Pedro Szekely

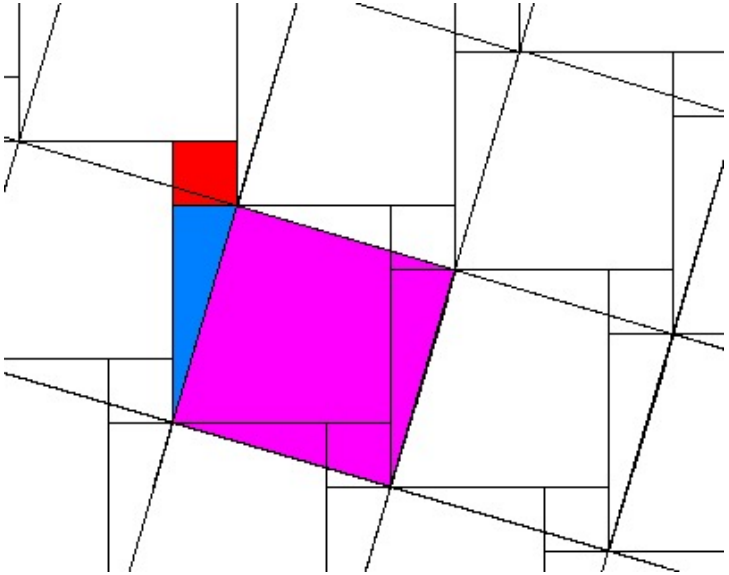
Principle 1:
Use URIs as names for things

Principle 2:
Use HTTP URIs so that people can look up those names

slide by Pedro Szekely

Can USC Have a URI?

Can the Pythagoras Theorem Have a URI?



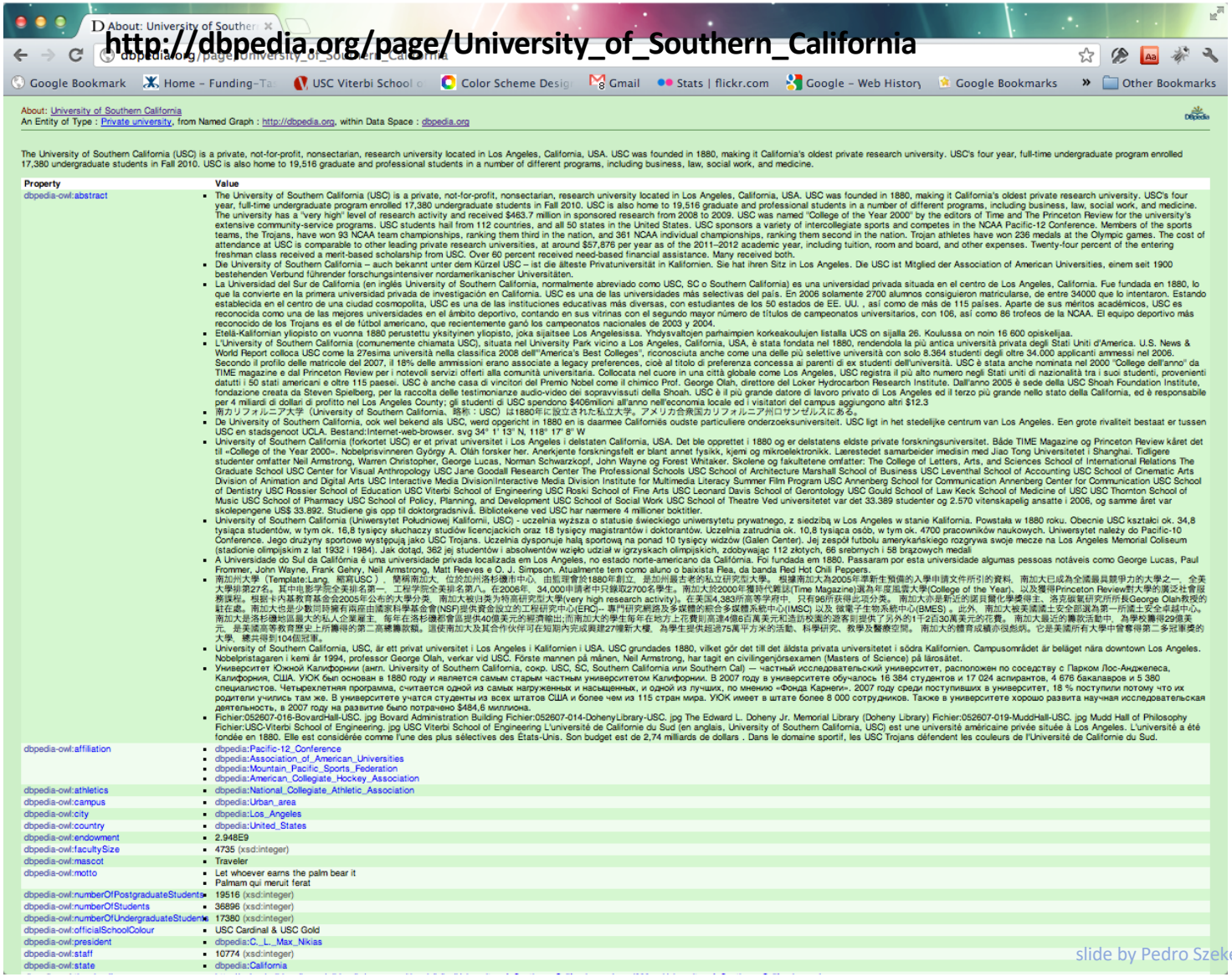
My Dog: Can He Have a URI?

http://szekelys.com/diego

slide by Pedro Szekely

Principle 3:
When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

slide by Pedro Szekely



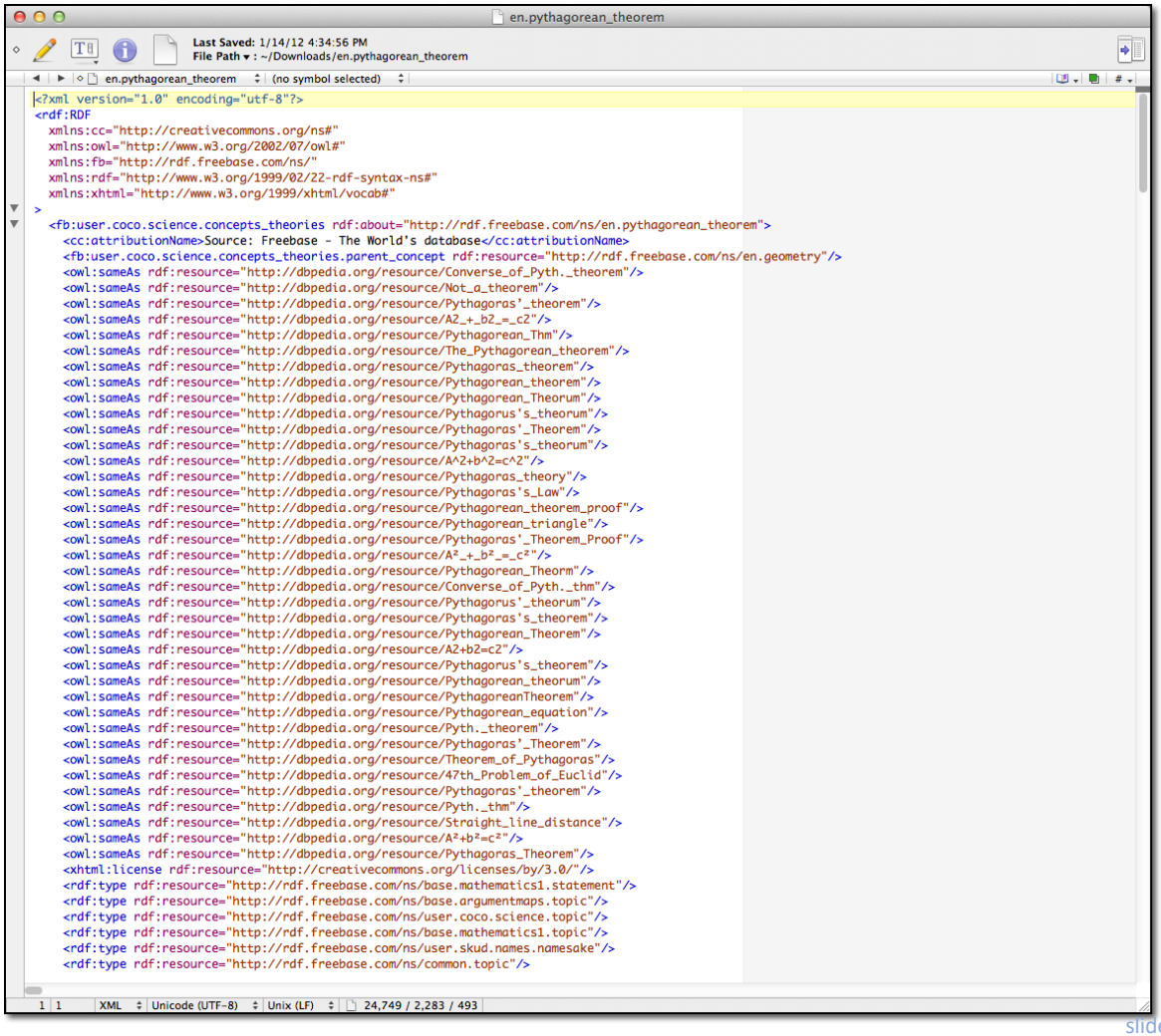
About: University of Southern California
An Entity of Type : Private university, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

The University of Southern California (USC) is a private, not-for-profit, nonsectarian, research university located in Los Angeles, California, USA. USC was founded in 1880, making it California's oldest private research university. USC's four year, full-time undergraduate program enrolled 17,380 undergraduate students in Fall 2010. USC is also home to 19,516 graduate and professional students in a number of different programs, including business, law, social work, and medicine.

Property	Value
dbpedia-owl:abstract	• The University of Southern California (USC) is a private, not-for-profit, nonsectarian, research university located in Los Angeles, California, USA. USC was founded in 1880, making it California's oldest private research university. USC's four year, full-time undergraduate program enrolled 17,380 undergraduate students in Fall 2010. USC is also home to 19,516 graduate and professional students in a number of different programs, including business, law, social work, and medicine. The university has a "very high" level of research activity and received $463.7 million in sponsored research from 2008 to 2009. USC was named "College of the Year 2000" by the editors of Time and The Princeton Review for the university's extensive community-service programs. USC students hail from 112 countries, and all 50 states in the United States. USC sponsors a variety of intercollegiate sports and competes in the NCAA Pacific-12 Conference. Members of the sports teams, the Trojans, have won 93 NCAA team championships, ranking them third in the nation, and 361 NCAA individual championships, ranking them second in the nation. Trojan athletes have won 236 medals at the Olympic games. The cost of attendance at USC is comparable to other leading private research universities, at around $57,876 per year as of the 2011–2012 academic year, including tuition, room and board, and other expenses. Twenty-four percent of the entering freshman class received a merit-based scholarship from USC. Over 60 percent received need-based financial assistance. Many received both. • Die University of Southern California – auch bekannt unter dem Kürzel USC – ist die älteste Privatuniversität in Kalifornien. Sie hat ihren Sitz in Los Angeles. Die USC ist Mitglied der Association of American Universities, einem seit 1900 bestehenden Verbund führender forschungsintensiver nordamerikanischer Universitäten. • La Universidad del Sur de California (en inglés University of Southern California, normalmente abreviado como USC, SC o Southern California) es una universidad privada situada en el centro de Los Angeles, California. Fue fundada en 1880, lo que la convierte en la primera universidad privada de investigación en California. USC es una de las universidades más selectivas del país. En 2006 solamente 2700 alumnos consiguieron matricularse, de entre 34000 que lo intentaron. Estando establecida en el centro de una ciudad cosmopolita, USC es una de las instituciones educativas más diversas, con estudiantes de los 50 estados de EE. UU. , así como de más de 115 países. Aparte de sus méritos académicos, USC es reconocida como una de las mejores universidades en el ámbito deportivo, contando en sus vitrinas con el segundo mayor número de títulos de campeonatos universitarios, con 106, así como 86 trofeos de la NCAA. El equipo deportivo más reconocido de los Trojans es el de fútbol americano, que recientemente ganó los campeonatos nacionales de 2003 y 2004. • Etelä-Kalifornian yliopisto on vuonna 1880 perustettu yksityinen yliopisto, joka sijaitsee Los Angelesissa. Yhdysvaltojen parhaimpien korkeakoulujen listalla UCS on sijalla 26. Koulussa on noin 16 600 opiskelijaa. • L'University of Southern California (comunemente chiamata USC), situata nel University Park vicino a Los Angeles, California, USA, è stata fondata nel 1880, rendendola la più antica università privata degli Stati Uniti d'America. U.S. News & World Report colloca USC come la 27esima università nella classifica 2008 dell'"America's Best Colleges", riconosciuta anche come una delle più selettive università con solo 8.364 studenti degli oltre 34.000 applicanti ammessi nel 2006. Secondo il profilo delle matricole del 2007, il 18% delle ammissioni erano associate a legacy preferences, cioè al titolo di preferenza concessa ai parenti di ex studenti dell'università. USC è stata anche nominata nel 2000 "College dell'anno" da TIME magazine e dal Princeton Review per i notevoli servizi offerti alla comunità universitaria. Collocata nel cuore in una città globale come Los Angeles, USC registra il più alto numero negli Stati uniti di nazionalità tra i suoi studenti, provenienti datutti i 50 stati americani e oltre 115 paesi. USC è anche casa di vincitori del Premio Nobel come il chimico Prof. George Olah, direttore del Loker Hydrocarbon Research Institute. Dall'anno 2005 è sede della USC Shoah Foundation Institute, fondazione creata da Steven Spielberg, per la raccolta delle testimonianze audio-video dei sopravvissuti della Shoah. USC è il più grande datore di lavoro privato di Los Angeles ed il terzo più grande nello stato della California, ed è responsabile per 4 miliardi di dollari di profitto nel Los Angeles County; gli studenti di USC spendono $406milioni all'anno nell'economia locale ed i visitatori del campus aggiungono altri $12.3 • 南カリフォルニア大学（University of Southern California、略称：USC）は1880年に設立された私立大学。アメリカ合衆国カリフォルニア州ロサンゼルスにある。 • De University of Southern California, ook wel bekend als USC, werd opgericht in 1880 en is daarmee Californiës oudste particuliere onderzoeksuniversiteit. USC ligt in het stedelijke centrum van Los Angeles. Een grote rivaliteit bestaat er tussen USC en stadsgenoot UCLA. Bestand:Internet-web-browser. svg 34° 1′ 13″ N, 118° 17′ 8″ W • University of Southern California (forkortet USC) er et privat universitet i Los Angeles i delstaten California, USA. Det ble opprettet i 1880 og er delstatens eldste private forskningsuniversitet. Både TIME Magazine og Princeton Review kåret det til «College of the Year 2000». Nobelprisvinneren György A. Oláh forsker her. Anerkjente forskningsfelt er blant annet fysikk, kjemi og mikroelektronikk. Læresteder samarbeider imedisin med Jiao Tong Universitetet i Shanghai. Tidligere studenter omfatter Neil Armstrong, Warren Christopher, George Lucas, Norman Schwarzkopf, John Wayne og Forest Whitaker. Skolene og fakultetene omfatter: The College of Letters, Arts, and Sciences School of International Relations The Graduate School USC Center for Visual Anthropology USC Jane Goodall Research Center The Professional Schools USC School of Architecture Marshall School of Business USC Leventhal School of Accounting USC School of Cinematic Arts Division of Animation and Digital Arts USC Interactive Media DivisionInteractive Media Division Institute for Multimedia Literacy Summer Film Program USC Annenberg School for Communication Annenberg Center for Communication USC School of Dentistry USC Rossier School of Education USC Viterbi School of Engineering USC Roski School of Fine Arts USC Leonard Davis School of Gerontology USC Gould School of Law Keck School of Medicine of USC USC Thornton School of Music USC School of Pharmacy USC School of Policy, Planning, and Development USC School of Social Work USC School of Theatre Ved universitetet var det 33.389 studenter og 2.570 vitenskapelig ansatte i 2006, og samme året var skolepengene US$ 33.892. Studiene gis opp til doktorgradsnivå. Bibliotekene ved USC har nærmere 4 millioner boktitler. • University of Southern California (Uniwersytet Południowej Kalifornii, USC) - uczelnia wyższa o statusie świeckiego uniwersytetu prywatnego, z siedzibą w Los Angeles w stanie Kalifornia. Powstała w 1880 roku. Obecnie USC kształci ok. 34,8 tysiąca studentów, w tym ok. 16,8 tysięcy słuchaczy studiów licencjackich oraz 18 tysięcy magistrantów i doktorantów. Uczelnia zatrudnia ok. 10,8 tysiąca osób, w tym ok. 4700 pracowników naukowych. Uniwersytet należy do Pacific-10 Conference. Jego drużyny sportowe występują jako USC Trojans. Uczelnia dysponuje halą sportową na ponad 10 tysięcy widzów (Galen Center). Jej zespół futbolu amerykańskiego rozgrywa swoje mecze na Los Angeles Memorial Coliseum (stadionie olimpijskim z lat 1932 i 1984). Jak dotąd, 362 jej studentów i absolwentów wzięło udział w igrzyskach olimpijskich, zdobywając 112 złotych, 66 srebrnych i 58 brązowych medali • A Universidade do Sul da Califórnia é uma universidade privada localizada em Los Angeles, no estado norte-americano da Califórnia. Foi fundada em 1880. Passaram por esta universidade algumas pessoas notáveis como George Lucas, Paul Frommer, John Wayne, Frank Gehry, Neil Armstrong, Matt Reeves e O. J. Simpson. Atualmente tem como aluno o baixista Flea, da banda Red Hot Chili Peppers. • 南加州大學（Template:Lang，簡稱南加大，位於加州洛杉磯市中心，由監理會於1880年創立，是加州最古老的私立研究型大學。根據南加大為2005年準新生預備的入學申請文件所引的資料，南加大已成為全國最具競爭力的大學之一，全美大學排第27名。其中电影学院全美排名第一，工程学院全美排名第八。在2006年，34,000申請者中只錄取2700名學生。南加大於2000年獲時代雜誌(Time Magazine)選為年度風雲大學(College of the Year)、以及獲得Princeton Review對大學的選舉社會服務課程。根据卡内基教育基金会2005年公布的大学分类，南加大被归类为特高研究型大学(very high research activity)。在美国4,383所高等学府中，只有96所获得此项分类。 南加大亦是新近的諾貝爾化學獎得主、洛克坡氧研究所所長George Olah教授的駐在處。南加大也是少數同時擁有兩座由國家科學基金會(NSF)提供資金設立的工程研究中心(ERC)-- 專門研究網路及多媒體的綜合多媒體系統中心(IMSC) 以及 微電子生物系統中心(BMES) 。此外，南加大被美國國土安全部選為第一所國土安全卓越中心。南加大是洛杉磯地區最大的私人企業雇主，每年在洛杉磯都會直接供40億美元的經濟輸出;而南加大的學生每年在地方上花費則高達4億6百萬美元和造訪校園的遊客則提供了另外的1千2百30萬美元的花費。 南加大最近的籌款活動中，為學校籌得29億美元，是美國高等教育歷史上所籌得的第二高總籌款額。這使南加大及其合作伙伴可在短期內完成興建27幢新大樓，為學生提供超過75萬平方米的活動、科學研究、教學及醫療空間。 南加大的體育成績亦很彪炳。它是美國所有大學中曾奪得第二多冠軍獎的大學，總共得到104個冠軍。 • University of Southern California, USC, är ett privat universitet i Los Angeles i Kalifornien i USA. USC grundades 1880, vilket gör det till det äldsta privata universitetet i södra Kalifornien. Campusområdet är beläget nära downtown Los Angeles. Nobelpristagaren i kemi år 1994, professor George Olah, verkar vid USC. Förste mannen på månen, Neil Armstrong, har tagit en civilingenjörsexamen (Masters of Science) på lärosätet. • Университет Южной Калифорнии (англ. University of Southern California, сокр. USC, SC, Southern California или Southern Cal) — частный исследовательский университет, расположен по соседству с Парком Лос-Анджелеса, Калифорния, США. УЮК был основан в 1880 году и является самым старым частным университетом Калифорнии. В 2007 году в университете обучалось 16 384 студентов и 17 024 аспирантов, 4 676 бакалавров и 5 380 специалистов. Четырехлетняя программа, считается одной из самых нагруженных и насыщенных, и одной из лучших, по мнению «Фонда Карнеги». 2007 году среди поступивших в университет, 18 % поступили потому что их родители учились там же. В университете учатся студенты из всех штатов США и более чем из 115 стран мира. УЮК имеет в штате более 8 000 сотрудников. Также в университете хорошо развита научная исследовательская деятельность, в 2007 году на развитие было потрачено $484,6 миллиона. • Fichier:052607-016-BovardHall-USC. jpg Bovard Administration Building Fichier:052607-014-DohenyLibrary-USC. jpg The Edward L. Doheny Jr. Memorial Library (Doheny Library) Fichier:052607-019-MuddHall-USC. jpg Mudd Hall of Philosophy Fichier:USC-Viterbi School of Engineering. jpg USC Viterbi School of Engineering L'université de Californie du Sud (en anglais, University of Southern California, USC) est une université américaine privée située à Los Angeles. L'université a été fondée en 1880. Elle est considérée comme l'une des plus sélectives des États-Unis. Son budget est de 2,74 milliards de dollars . Dans le domaine sportif, les USC Trojans défendent les couleurs de l'Université de Californie du Sud.
dbpedia-owl:affiliation	• dbpedia:Pacific-12_Conference • dbpedia:Association_of_American_Universities • dbpedia:Mountain_Pacific_Sports_Federation • dbpedia:American_Collegiate_Hockey_Association
dbpedia-owl:athletics	• dbpedia:National_Collegiate_Athletic_Association
dbpedia-owl:campus	• dbpedia:Urban_area
dbpedia-owl:city	• dbpedia:Los_Angeles
dbpedia-owl:country	• dbpedia:United_States
dbpedia-owl:endowment	• 2.948E9
dbpedia-owl:facultySize	• 4735 (xsd:integer)
dbpedia-owl:mascot	• Traveler
dbpedia-owl:motto	• Let whoever earns the palm bear it • Palmam qui meruit ferat
dbpedia-owl:numberOfPostgraduateStudents	• 19516 (xsd:integer)
dbpedia-owl:numberOfStudents	• 36896 (xsd:integer)
dbpedia-owl:numberOfUndergraduateStudents	• 17380 (xsd:integer)
dbpedia-owl:officialSchoolColour	• USC Cardinal & USC Gold
dbpedia-owl:president	• dbpedia:C._L._Max_Nikias
dbpedia-owl:staff	• 10774 (xsd:integer)
dbpedia-owl:state	• dbpedia:California

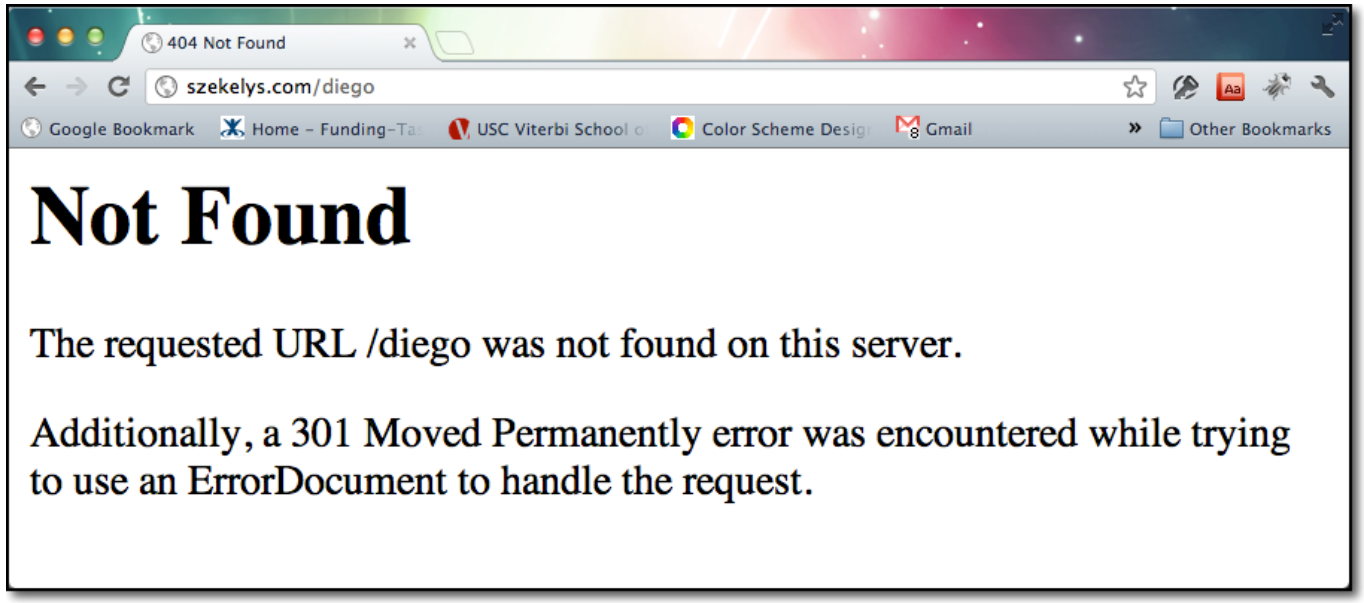
slide by Pedro Szekely

http://www.freebase.com/view/en/pythagorean_theorem



slide by Pedro Szekely

http://szekelys.com/diego



~~Principle 3:~~
~~When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)~~

slide by Pedro Szekely

Principle 4:
Include links to other URIs so that they can discover more things

slide by Pedro Szekely

http://szekelys.com/diego

```
@prefix rdf:          <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpprop:<http://dbpedia.org/property/> .
@prefix dbpedia:      <http://dbpedia.org/resource/> .
@prefix dbpedia-owl:  <http://dbpedia.org/ontology/> .
@prefix fb:           <http://rdf.freebase.com/ns/> .

http://szekelys.com/diego
        rdf:type                              “Dog” ;
        http://szekelys.com/name              ”Diego" ;
        dbpedia-owl:species                   “Labrador Retriever” ;
        dbprop:country                        “Canada” ;
        dbprop:color                          “Yellow” ;
        fb:base.petbreeds.dog_size.gender     “Male” .
```

Linked Data?

slide by Pedro Szekely

http://szekelys.com/diego

```
@prefix rdf:            <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpprop:<http://dbpedia.org/property/> .
@prefix dbpedia:        <http://dbpedia.org/resource/> .
@prefix dbpedia-owl:    <http://dbpedia.org/ontology/> .
@prefix fb:             <http://rdf.freebase.com/ns/> .

http://szekelys.com/diego
        rdf:type                                "Dog" ;
        http://szekelys.com/name                "Diego" ;
        dbpedia-owl:species                     "Labrador Retriever" ;
        dbprop:country                          "Canada" ;
        dbprop:color                            "Yellow" ;
        fb:base.petbreeds.dog_size.gender       "Male" .
```

Not Linked Data

slide by Pedro Szekely

http://szekelys.com/diego

```
@prefix rdf:            <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpprop:<http://dbpedia.org/property/> .
@prefix dbpedia:        <http://dbpedia.org/resource/> .
@prefix dbpedia-owl:    <http://dbpedia.org/ontology/> .
@prefix fb:             <http://rdf.freebase.com/ns/> .

http://szekelys.com/diego
        rdf:type                                dbpedia:Dog ;
        http://szekelys.com/name                ”Diego" ;
        dbpedia-owl:species                     dbpedia:Labrador Retriever ;
        dbprop:country                          dbpedia:Canada ;
        dbprop:color                            dbpedia:Yellow ;
        fb:base.petbreeds.dog_size.gender fb:en.male"/> .
```

Principle 4:
Include links to other URIs so that they can discover more things

slide by Pedro Szekely

http://szekelys.com/diego

```
@prefix rdf:                <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpprop:<http://dbpedia.org/property/> .
@prefix dbpedia:            <http://dbpedia.org/resource/> .
@prefix dbpedia-owl:        <http://dbpedia.org/ontology/> .
@prefix fb:                 <http://rdf.freebase.com/ns/> .

http://szekelys.com/diego
        rdf:type                                dbpedia:Dog ;
        http://szekelys.com/name                ”Diego" ;
        dbpedia-owl:species                     dbpedia:Labrador Retriever ;
        dbprop:country                          dbpedia:Canada ;
        dbprop:color                            dbpedia:Yellow ;
        fb:base.petbreeds.dog_size.gender fb:en.male"/> .
```

Principle 3:
When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)

slide by Pedro Szekely

http://szekelys.com/diego

```
@prefix rdf:          <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpprop:<http://dbpedia.org/property/> .
@prefix dbpedia:      <http://dbpedia.org/resource/> .
@prefix dbpedia-owl:  <http://dbpedia.org/ontology/> .
@prefix fb:           <http://rdf.freebase.com/ns/> .
@prefix foaf:         <http://xmlns.com/foaf/0.1/> .

http://szekelys.com/diego
        rdf:type                              dbpedia:Dog ;
        foaf:name                             ”Diego" ;
        dbpedia-owl:species                   dbpedia:Labrador Retriever ;
        dbprop:country                        dbpedia:Canada ;
        dbprop:color                          dbpedia:Yellow ;
        fb:base.petbreeds.dog_size.gender fb:en.male"/> .
```

foaf is a widely used vocabulary

friend of a friend

slide by Pedro Szekely

Now we know what linked data is

What can go wrong?

Different URIs For the Same Thing

https://www.wikidata.org/wiki/Q36107

http://dbpedia.org/resource/Muhammad_Ali

https://yago-knowledge.org/resource/Muhammad_Ali

http://data.nytimes.com/N13611972026987463463

Linked Data Challenges

- Different URIs for the same thing
  - … makes it harder to link the data
- Timeliness
  - … not up to date
- Provenance
  - … not only a linked data problem
- Tools
  - … slow performance compared to traditional data
  - … search engines not yet mature
  - … many RDF formats, not supported by all tools

slide by Pedro Szekely

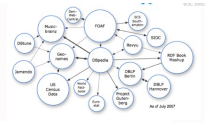
Linked Data Challenges

- Different URIs for the same thing
  - … makes it harder to link the data
- Timeliness
  - … not up to date
- Provenance
  - … not only a linked data problem
- Tools
  - … slow performance compared to traditional data
  - … search engines not yet mature
  - … many RDF formats, not supported by all tools

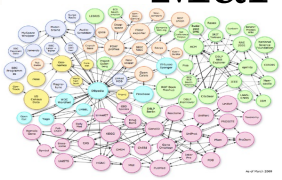
slide by Pedro Szekely

Getting Bigger Every Day

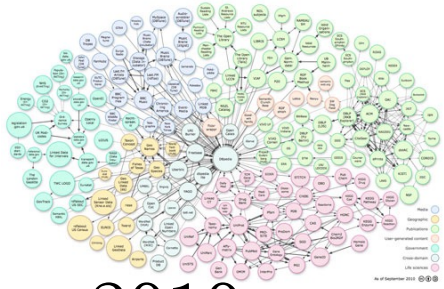
July 2007



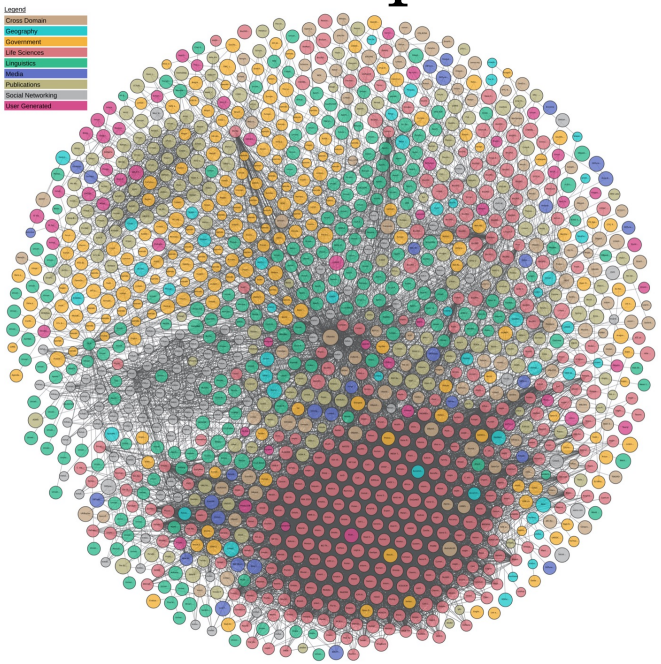
March 2009



September 2010



April 2020



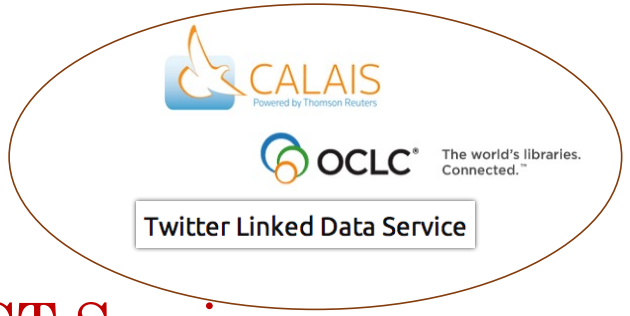
slide by Pedro Szekely

Working with Linked Data

RDF Dump

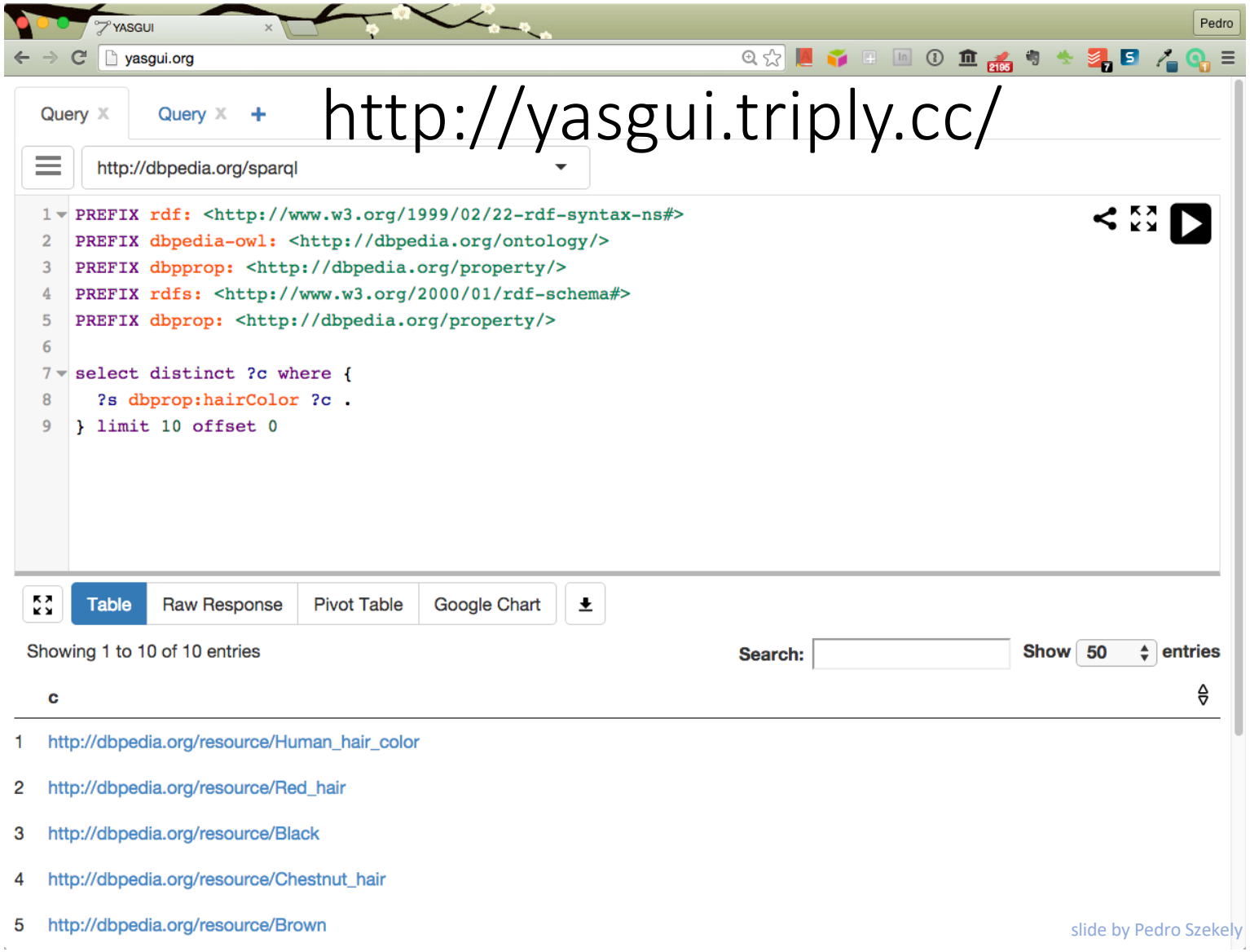
SPARQL Endpoint

URI Derefencing



REST Service

slide by Pedro Szekely


YASGUI
Pedro
yasgui.org
Query
Query
http://yasgui.triply.cc/
http://dbpedia.org/sparql
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbprop: <http://dbpedia.org/property/>
select distinct ?c where {
?s dbprop:hairColor ?c .
} limit 10 offset 0
Table
Raw Response
Pivot Table
Google Chart
Showing 1 to 10 of 10 entries
Search:
Show 50 entries
c
1 http://dbpedia.org/resource/Human_hair_color
2 http://dbpedia.org/resource/Red_hair
3 http://dbpedia.org/resource/Black
4 http://dbpedia.org/resource/Chestnut_hair
5 http://dbpedia.org/resource/Brown
slide by Pedro Szekely

Is your Linked Open Data 5 Star?

Available on the web (whatever format) *but with an open licence, to be Open Data*

Available as machine-readable structured data (e.g. excel instead of image scan of a table)

as (2) plus non-proprietary format (e.g. CSV instead of excel)

All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

All the above, plus: Link your data to other people's data to provide context

Best Practices for Data on the Web

https://www.w3.org/TR/dwbp

Best Practice 1: Provide metadata

Best Practice 2: Provide descriptive metadata

Best Practice 3: Provide structural metadata

Best Practice 4: Provide data license information

Best Practice 5: Provide data provenance information

Best Practice 6: Provide data quality information

Best Practice 7: Provide a version indicator

Best Practice 8: Provide version history

Best Practice 9: Use persistent URIs as identifiers of datasets

Best Practice 10: Use persistent URIs as identifiers within datasets

Best Practice 11: Assign URIs to dataset versions and series

Best Practice 12: Use machine-readable standardized data formats

Best Practice 13: Use locale-neutral data representations

Best Practice 14: Provide data in multiple formats

Best Practice 15: Reuse vocabularies, preferably standardized ones

Best Practice 16: Choose the right formalization level

Best Practice 17: Provide bulk download

Best Practice 18: Provide Subsets for Large Datasets

Best Practice 19: Use content negotiation for serving data available in multiple formats

Best Practice 20: Provide real-time access

Best Practice 21: Provide data up to date

Best Practice 22: Provide an explanation for data that is not available

Best Practice 23: Make data available through an API

Best Practice 24: Use Web Standards as the foundation of APIs

Best Practice 25: Provide complete documentation for your API

Best Practice 26: Avoid Breaking Changes to Your API

Best Practice 27: Preserve identifiers

Best Practice 28: Assess dataset coverage

Best Practice 29: Gather feedback from data consumers

Best Practice 30: Make feedback available

Best Practice 31: Enrich data by generating new data

Best Practice 32: Provide Complementary Presentations

Best Practice 33: Provide Feedback to the Original Publisher

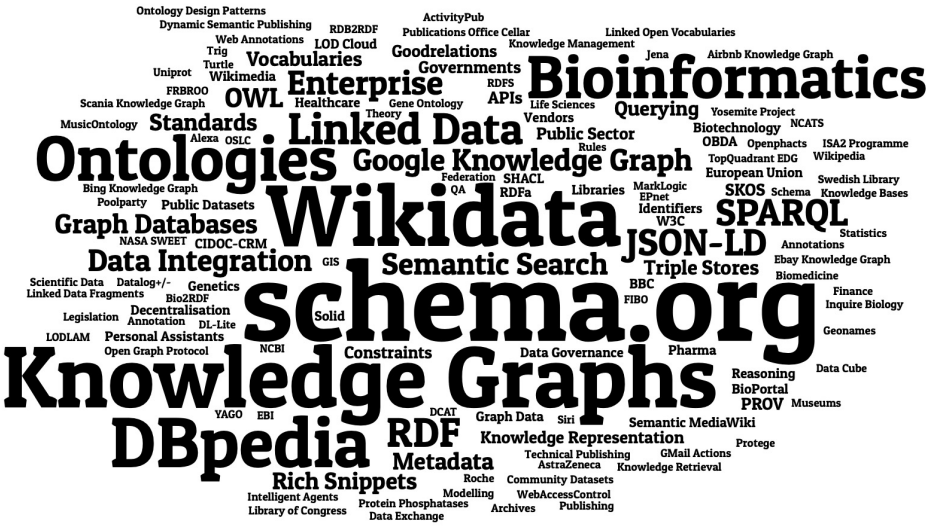
Best Practice 34: Follow Licensing Terms

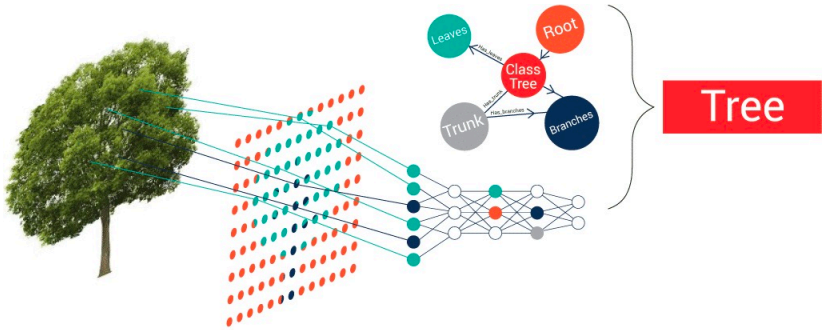
Best Practice 35: Cite the Original Publication

slide by Basel Shbita

What's next?

- Players
  - Academia
  - Industry
- Criticisms
  - lack of usable tools
  - lack of incentives
  - lack of robustness for unreliable publishers
  - overly verbose standards
- Vision
  - bridge the gap, in particular, to developers
  - inject Semantic Web interoperability into already broadly-adopted technologies





slide by Basel Shbita, images from https://www.ontotext.com/knowledgehub/fundamentals/what-is-the-semantic-web/