Student Online Teaching Advice Notice

The materials and content presented within this session are intended solely for use in a context of teaching and learning at Trinity.

Any session recorded for subsequent review is made available solely for the purpose of enhancing student learning.

Students should not edit or modify the recording in any way, nor disseminate it for use outside of a context of teaching and learning at Trinity.

Please be mindful of your physical environment and conscious of what may be captured by the device camera and microphone during videoconferencing calls.

Recorded materials will be handled in compliance with Trinity's statutory duties under the Universities Act, 1997 and in accordance with the University's policies and procedures.

Further information on data protection and best practice when using videoconferencing software is available at https://www.tcd.ie/info_compliance/data-protection/.

University of Dublin
Trinity College

CSU22041: Information Management I

Concept of the Semantic Web

2020-2021

Gaye Stephens gaye.stephens@tcd.ie

https://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=en#t-960912

Some Core Concepts from Module Introduction

ORGANISATION

How data represented/associated

UML- Class Diagram
XML

METADATA

Data about what the data is

XML
DTD
XML Schema

ACCESS

How get at the data efficiently

Xpath
XQuery

Express
UML in XML

Document Type Definition(DTD)
Element Declarations
Element Occurances
Entity Declarations
Attribute List Declarations
……..

XML Schemas Definitions (XSD)
Simple Type/Complex Types
Structure
Attributes/Attribute Groups
Mixed Content
Empty Elements
……..Lots more

UML Diagrams
Class Diagram
Use Case
….
Ethics Canvas

Validated Using DTDs

XML Schemas are also used for validation

BaseX
XML Database
and processor

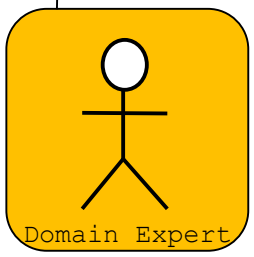
XML
User Defined Tags
hierarchical Structure
Prolog
Document Type Declaration
Elements
Attributes
Entities
Cdata Sections
Processing Instructions
Comments

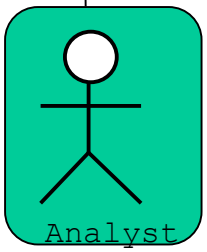
NameSpaces

Not supported in DTDs

XPath

XQuery





Well-Formed

"XML namespaces provide a simple method for qualifying element and attribute names used in Extensible Markup Language documents by associating them with namespaces identified by URI references." W3.org

Information Retrieval Systems

at their most basic…

Information Need or Query

IR System

Highly Relevant Results (Hopefully!!)

Information Objects

Information Retrieval – Past and Present

The Fundamental Paradox of Information Retrieval

- "The need to describe that which you do not know in order to find it" - Roland Hjerrpe

circa 1955-1975

- Foundational research
- Fundamental IR concepts advanced in research
- Database research ongoing in parallel

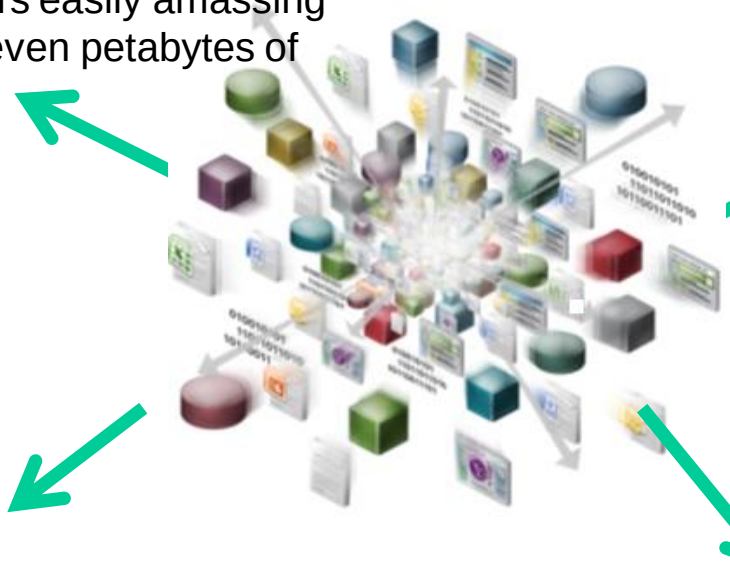
1975 to present

- Slow adoption of IR research into operational systems
- Accelerated in mid-1990's with emergence search engines
- Accelerated in mid-2000's with growth web
- Accelerating again now .. Due to changing nature WWW

Common Challenges managing data for Enterprises and Individuals

Volume

Awash with data, consumers easily amassing terabytes and enterprises even petabytes of information.



Velocity

Often time-sensitive, data must be processed as it is streaming in order to maximize its value

Validity

Data protection – consent and compliance;
Data privacy – what data an individual willing to share;
Data ethics – consideration of ethical issues when processing data.

Variety

Data extends beyond structured data, including semi-structured and unstructured data of all varieties: text, audio, video, click streams, log files and more.

Metadata will solve all problems?

Classes of Metadata

- **Domain specific – e.g. based on some domain like "University"**
- Direct content-based – e.g. inverted tree index of a document
- Content Descriptive – e.g. textual annotations descibing an image
- Content dependent – e.g. size of the document
- Content independent – e.g. modification date of document
- Domain independent – e.g HTML document type definitions

- Relies on common understanding of syntax/terms

XML will solve all problems?

- Domain Semantics are complex:
  - implicit assumptions, hidden semantics

  ⇒ sources seem unrelated to the non-expert

  XML DTDs and XML Schemas are sufficient for exchanging data between parties who have agreed to meaning of terms beforehand

- XML is Syntax too
  - DTDs talk about element nesting
  - XML Schema schemas give you data type definitions

- Need Structure and Semantics beyond XML trees!

- What we need is some model that can be used as reference between applications from diverse parties in order to ensure terms used are semantically equivalent

Representing Information: Data and Relationships

So far, for representation of data and their relationships, you have already covered :

- As *objects*, using the well-accepted techniques of object-oriented analysis and design to capture a model
- As *UML*, but not easily machine processable
- As *a tree-like* representation (e.g. Binary, AVL, XML etc.)

For the semantic web, we are looking for infinitely extensible, mergeable and scalable solutions. We need the above plus something more….

University of Dublin
Trinity College

Heading toThe Semantic Web

Next Generation information representation, retrieval and processing?

Web has made Data Available

Web

- Easy publication
- An infrastructure for retrieving and representing documents
- An infrastructure for accessing data

Next step is *semantic interoperation*:

- Understanding what the data means
- Linking in insightful ways
- Automated support for data integration
- Smart applications

$$\textit{Sharing data} \Rightarrow \textit{Sharing meaning}$$

Web of Documents vs. Web of Data

The Web of Documents were created by humans for humans; the links between documents bore little meaning for machines and documents provided little structured information.

Structured information can be found on the Web such as XML, CSV, etc. – but, …

How do we link data rather than documents, and create a global “database” of information?

Web of Documents vs. Web of Data

	Web of Documents	Web of Data
Analogy	Global file system	Global database
Primary objects	Documents	(Descriptions of) Things
Links between	(Parts of) Documents	Things
Degree of structure	Low	High
Semantics between links and content	Implicit	Explicit
Designed for	Human consumption	Both human and computer-based agents

Compiled from http://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf

“ The Semantic Web is a Web of Data —

The collection of Semantic Web technologies provides an environment where applications can query that data, draw inferences using vocabularies, etc.

However, to make the Web of Data a reality, it is important to have the **huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools**.

Furthermore, not only does the Semantic Web need access to data, but ***relationships among data* should be made available, too, to create a *Web* of Data** (as opposed to a sheer collection of datasets). This collection of interrelated datasets on the Web can also be referred to as **Linked Data**.”

Enabling persons or machines to explore the Web of Data.

https://www.w3.org/standards/semanticweb/data.html

University of Dublin
Trinity College

Linked Data Initiative

Connect distributed data across the Web → http://linkeddata.org/

What is Linked Data?

Linked Data started off as a initiative called the Linking Open Data (LOD) project.

Linked Data is a global initiative to publish and interlink structured data on the Web using a clever combination of simple, standardized technologies.

- Uniform Resource Identifiers (URI)– to name things;
- Resource Description Framework(RDF) – to represent things;
- HTTP infrastructure – to obtain those representations.

Linked Data

Linked Data is also a community effort to publish *(open)* data sets as Linked Data on the Web (to which anyone can refer to)

According to some “protocol” and ...

Interlink these data sets and ...

Develop clients that consume Linked Data from the Web

Linked Data Principles

Formulated by Tim Berners-Lee (2006):

1. Use URIs as names for things
2. Use HTTP URIs so that people/apps can look up these names
3. When someone/an app looks up a URI, provide useful information
4. Include links to other URIs so that they can discover more things

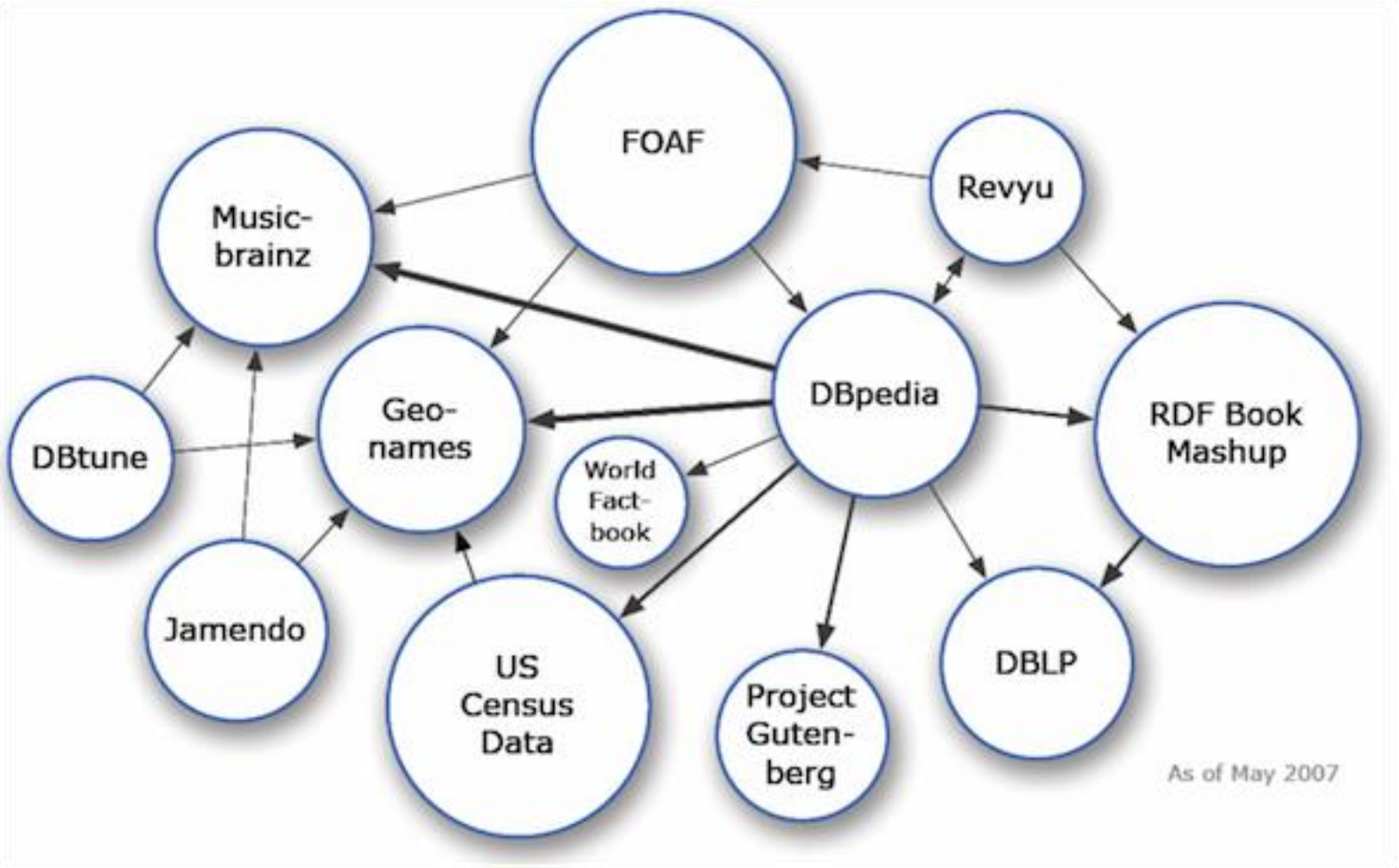
This is **not** an unambiguous specification, just a set of principles.

DBpedia

Initiative to make available Wikipedia knowledge as
- Linked Open Data

Structure from
- Infoboxes
- HTML (titles)
- Categories
- Links
  - other languages
  - redirects
  - disambiguations, etc

Linked Open Data 2007



LOD Cloud 2014 Statistics

By Max Schmachtenberg, Christian Bizer, and Heiko Paulheim in the context of the EU PlanetData project.

Also provides information about vocabularies used as well as popular predicates for interlinking.

Datasets by topical domain.

Topic	Datasets	%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Social web	520	51.28%
Total	**1014**	

Categorization by number of linked datasets

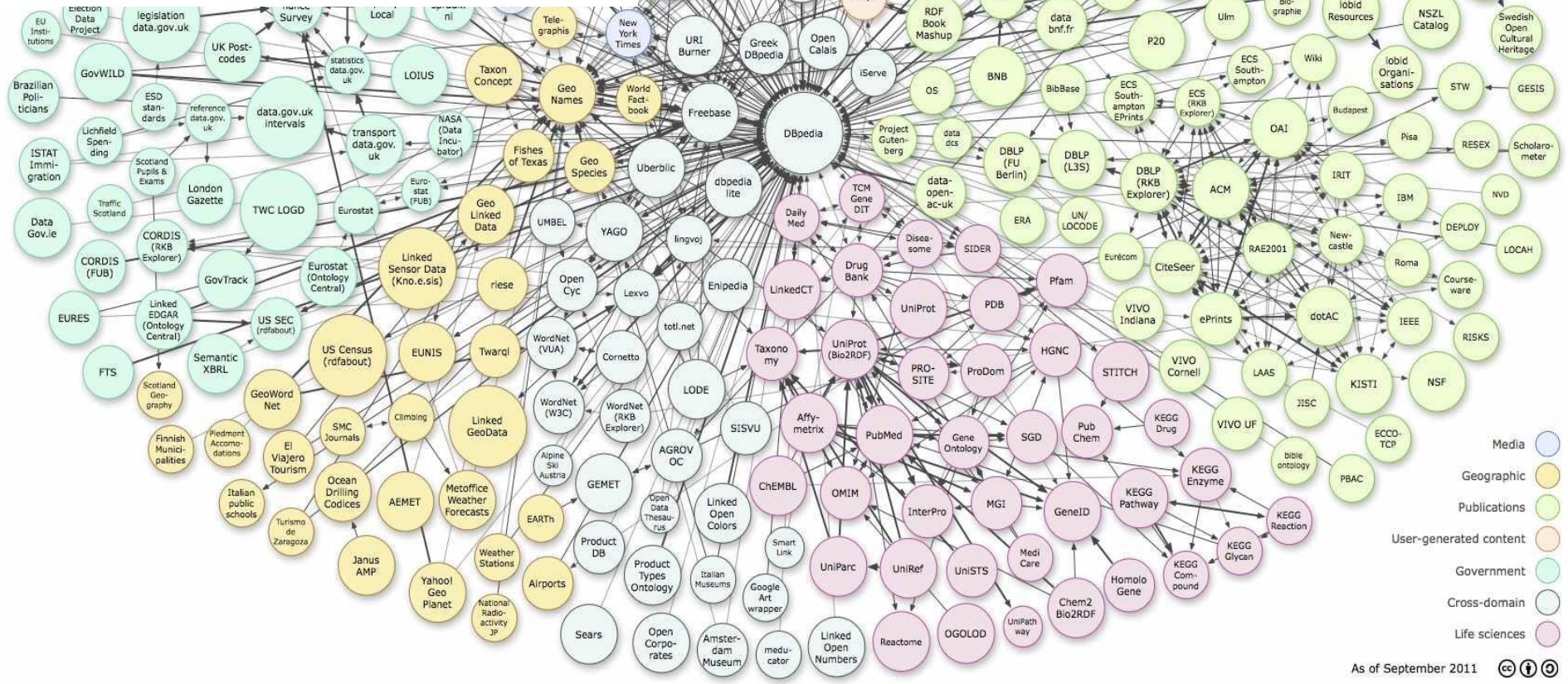
Number of linked datasets	Number of datasets
more than 10	79 (7.79%)
6 to 10	81 (7.99%)
5	31 (3.06%)
4	42 (4.14%)
3	54 (5.33%)
2	106 (10.45%)
1	176 (17.36%)
0	445 (43.89%)

Circle size **Triple Count**

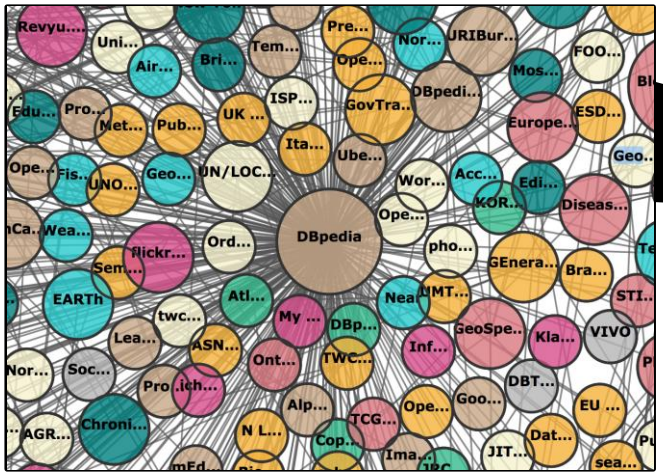
Very large >1B
Large 1B-10M
Medium 10M-500k
Small 500k-10k
Very small <10k

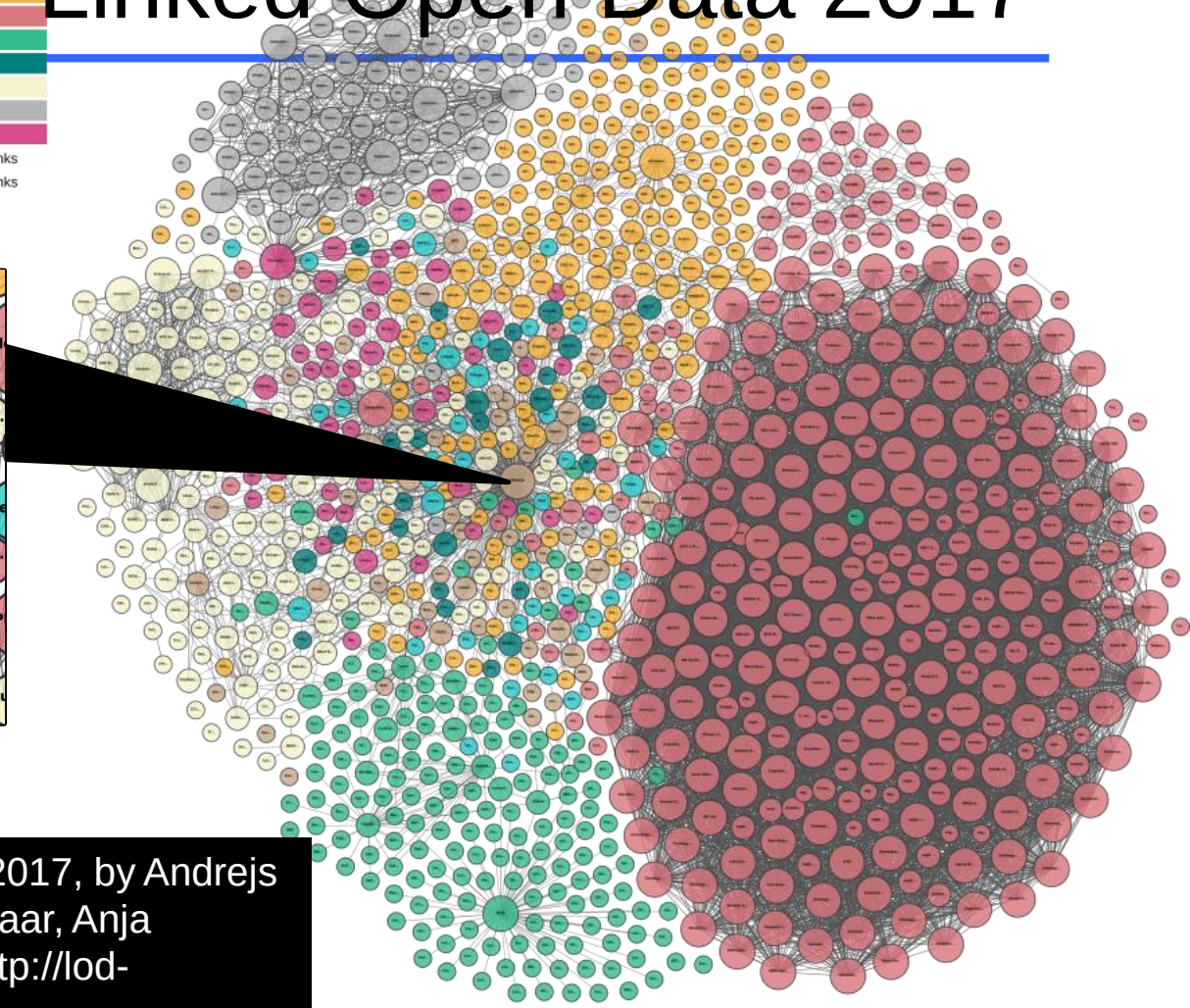
The arrows indicate the existence of at least 50 links between two datasets.

Linked Open Data 2012



Linked Open Data 2017





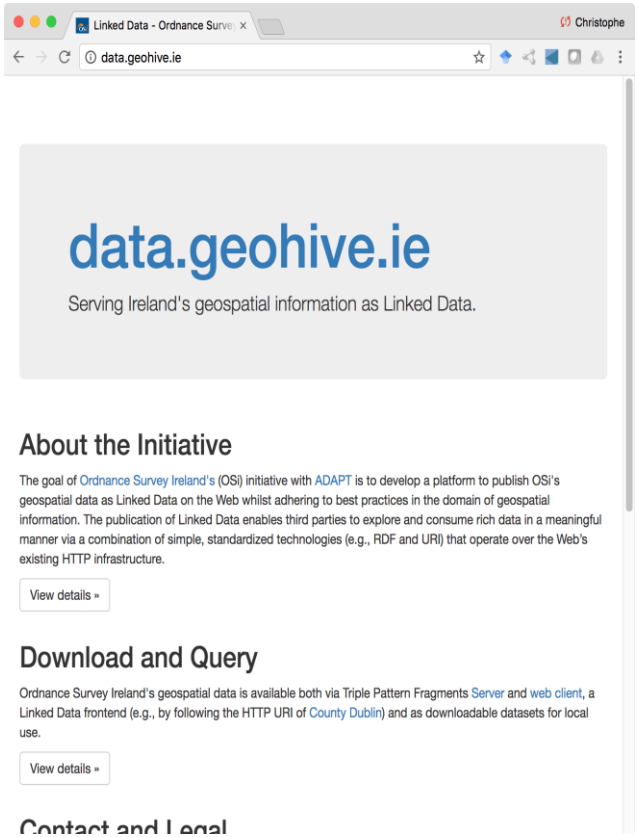
Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

Comon Vocabs

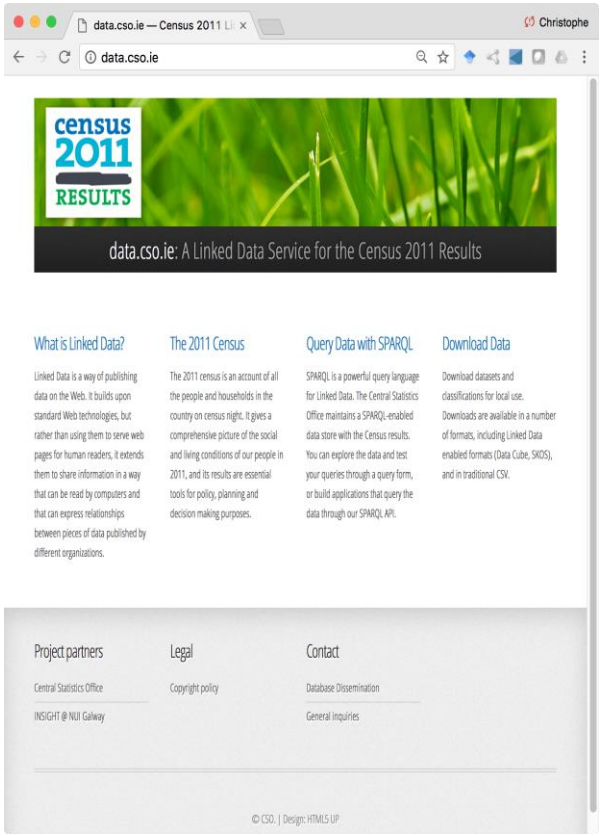
bibo -- Bibilographic ontology
cc -- Creative Commons ontology
damltime -- Time Zone ontology
doap -- Description of a Project ontology
event -- Event ontology
foaf -- Friend-of-a-Friend ontology
frbr -- Functional Requirements for Bibliographic Records
geo -- Geo wgs84 ontology
geonames -- GeoNames ontology
mo -- Music Ontology
opencyc -- OpenCyc knowledge base
owl -- Web Ontology Language
pim_contact -- PIM (personal information management) Contacts ontology
po -- Programmes Ontology (BBC)
rss -- Really Simple Syndicate (1.0) ontology
sioc -- Socially Interlinked Online Communities ontology
sioc_types -- SIOC extension
skos -- Simple Knowledge Organization System
umbel -- Upper Mapping and Binding Exchange Layer ontology
wordnet -- WordNet lexical ontology
yandex_foaf -- FOAF (Friend-of-a-Friend) Yandex extension ontology

Examples

http://data.geohive.ie/

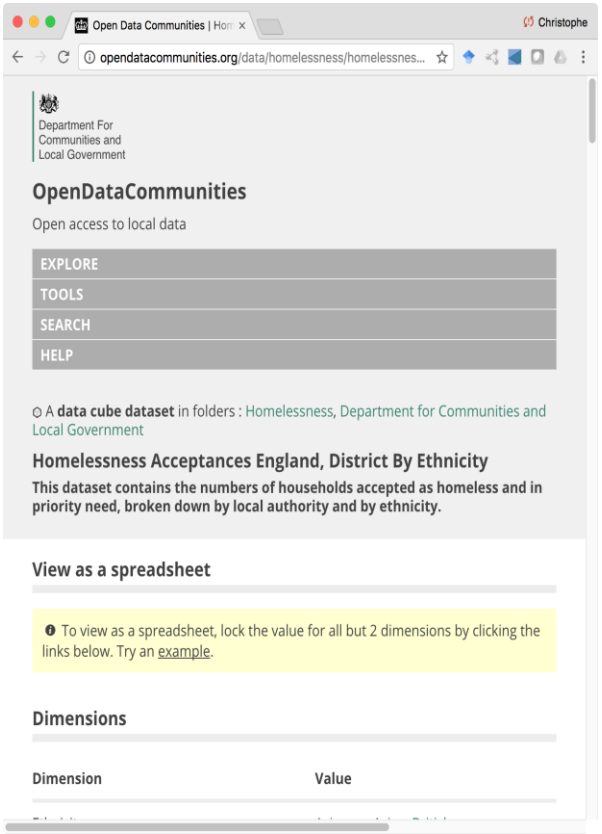


http://data.cso.ie/

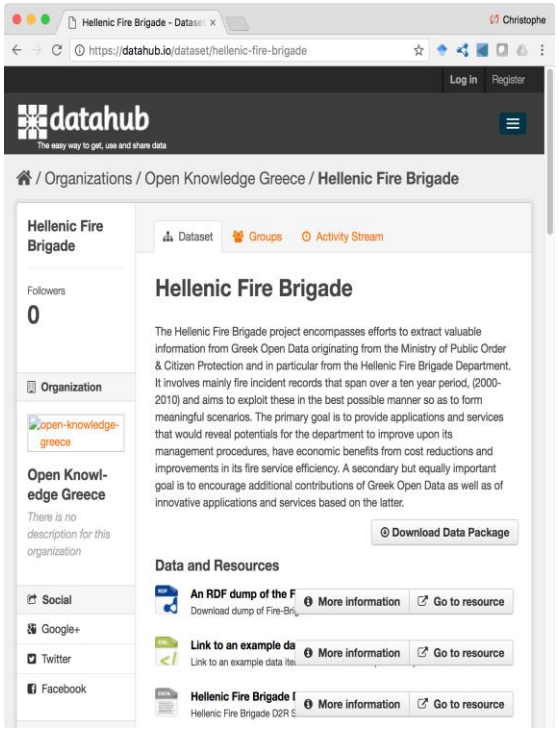


Examples

http://opendatacommunities.org/

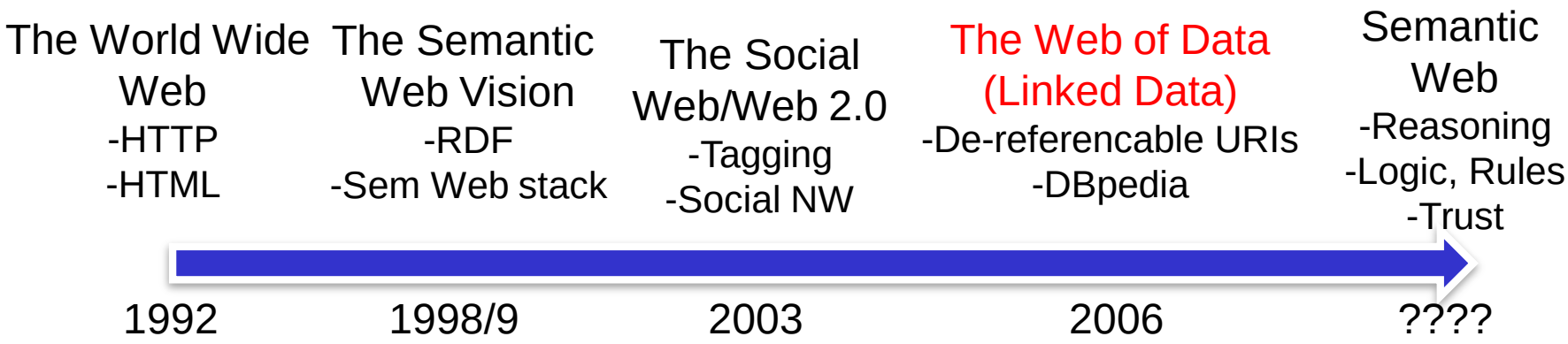


https://datahub.io/



https://www.slideshare.net/echo4ngel/open-data-and-linked-data-59118102

The Long Road for Semantic Web



- Linked Data Movement
  - Tim Berners Lee driven
  - Treat schemas as vocabularies
  - Reuse existing schemas

What?

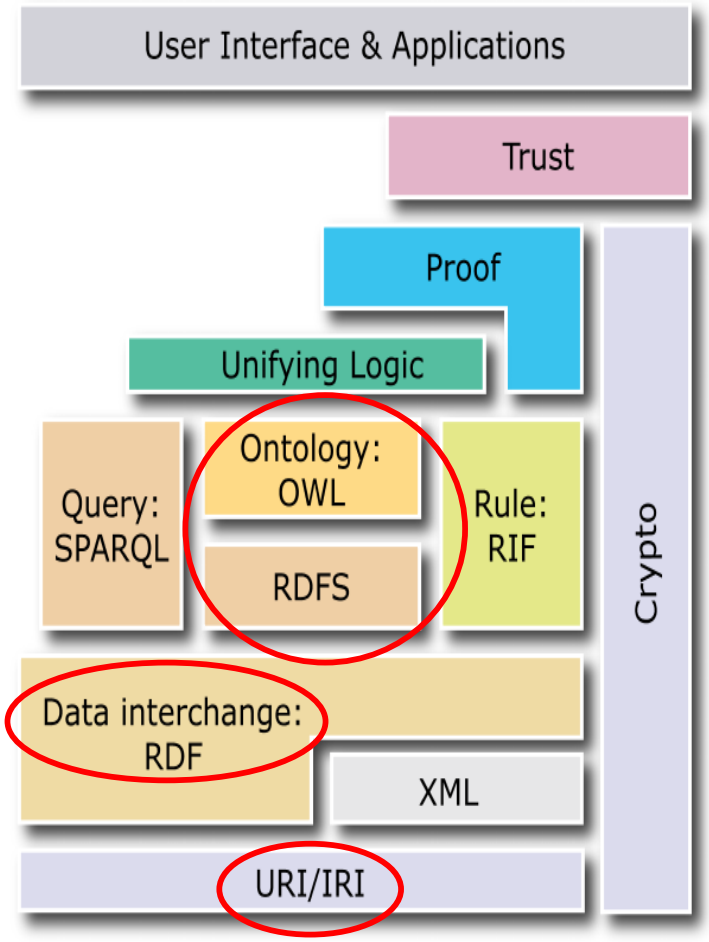
In contrast to the full-fledged Semantic Web vision, linked data is mainly about **publishing structured data** in **RDF** using **URIs** rather than focusing on the ontological level or inference. This simplification—just as the Web simplified the established academic approaches of Hypertext systems—lowers the entry barrier for data provider, hence fosters a wide-spread adoption.

[EXPL]

The Semantic Web Stack

Traditional Web
URI/IRI
XML
Cryptography

Semantic Web
Resource Description Framework (RDF)
RDF Schema
Web Ontology Language
SPARQL
Rules: RIF (and SWRL)



(from https://www.w3.org/)

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

“AFTER WE GIVE a presentation about the Semantic Web, we’re often asked, “Okay, so what is the killer application of the Semantic Web?” The “killer app” of any technology, of course, is the application that brings a user to investigate the technology and start using it. The transistor radio was a killer app of transistors, and the cell phone is a killer app of wireless technology. So what do we answer? “The Semantic Web is the killer app.” At this point we’re likely to be told we’re crazy, so we ask a question in turn: **“Well, what’s the killer app of the World Wide Web?” Now we’re being stared at kind of fisheyed, so we answer ourselves: “The Web is the killer app of the Internet. The Semantic Web is another killer app of that magnitude.” The point here is that the abilities of the Semantic Web are too general to be thought about in terms of solving one key problem or creating one essential gizmo. It will have uses we haven’t dreamed of.”**

THE SEMANTIC WEB Author(s): TIM BERNERS-LEE, JAMES HENDLER and ORA LASSILA Source: Scientific American , Vol. 284, No. 5 (MAY 2001), pp. 34-43 Published by: Scientific American, a division of Nature America, Inc.

Sample Linked Data Exam Question

1. The move to Linked Data (and eventually the Semantic Web) will bring benefits for application developers, compared to how data is currently available on the web.

   Discuss the statement above. Diagrams can be included to support or illustrate points made in your discussion.
   Include at least the following points in your answer.
   - Describe the benefits that Linked Data could bring;
   - Explain the concept of Linked Data;
   - Explain the concept of the Semantic Web;
   - Describe the Semantic Web Stack;
   - Explain in what way OWL builds on RDF and what benefits this brings.

   [Total 50 Marks]

Essay based answer

That’s All Folks
Thank You for Listening