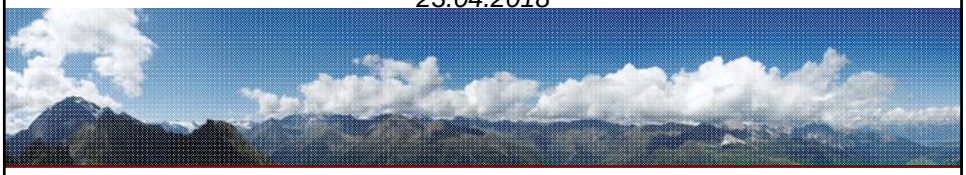
# Semantic Web Services
# SS 2018

## Semantic Web

Anna Fensel

*23.04.2018*





## Where are we?

STI · INNSBRUCK

| # | Title |
|---|---|
| 1 | Introduction |
| 2 | Web Science + Cathy O'Neil's talk: "Weapons of Math Destruction" |
| 3 | Service Science |
| 4 | Web services |
| 5 | Web2.0 services + ONLIM APIs (separate slideset) |
| **6** | **Semantic Web** |
| 7 | Semantic Web Service Stack (WSMO, WSML, WSMX) |
| 8 | OWL-S and the others |
| 9 | Semantic Services as a Part of the Future Internet and Big Data Technology |
| 10 | Lightweight Annotations |
| 11 | Linked Services |
| 12 | Applications |
| 13 | Mobile Services |

## Agenda

STI · INNSBRUCK

1. Motivation
   1. Development of the Web
      1. Internet
      2. Web 1.0
      3. Web2.0
   2. Limitations of the current Web
2. Technical solution
   1. Introduction to Semantic Web
   2. Architecture and languages
3. Semantic Web - Data
4. Extensions
   1. Linked (Open) Data
   2. Schema.org
   3. LOV
5. Summary
6. References

3

STI · INNSBRUCK

# MOTIVATION

4

STI · INNSBRUCK

# DEVELOPMENT OF THE WEB

5

## Development of the Web

STI · INNSBRUCK

1. Internet
2. Web 1.0
3. Web 2.0

6

STI · INNSBRUCK

# INTERNET

7

## Internet

STI · INNSBRUCK

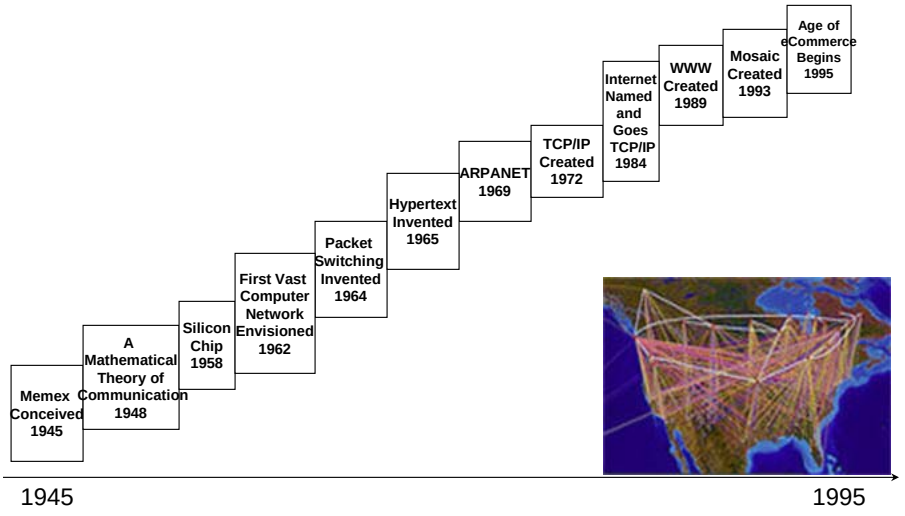
- "The **Internet** is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide. It is a network of networks that consists of millions of private and public, academic, business, and government networks of local to global scope that are linked by a broad array of electronic and optical networking technologies."

http://en.wikipedia.org/wiki/Internet

8

## A brief summary of Internet evolution

STI · INNSBRUCK




Source: http://slidewiki.org/slide/24721

9

STI · INNSBRUCK

# WEB 1.0

10

## Web 1.0

STI · INNSBRUCK

- "The **World Wide Web** ("**WWW**" or simply the "**Web**") is a system of interlinked, hypertext documents that runs over the Internet. With a Web browser, a user views Web pages that may contain text, images, and other multimedia and navigates between them using hyperlinks".

http://en.wikipedia.org/wiki/World_Wide_Web

11

## Web 1.0

STI · INNSBRUCK

- **Netscape**
  - *Netscape* is associated with the breakthrough of the Web.
  - *Netscape* had rapidly a large user community making attractive for others to present their information on the Web.

- **Google**
  - *Google* is the incarnation of Web 1.0 mega grows
  - *Google* indexed already in 2008 more than 1 trillion pages [*]
  - *Google* and other similar search engines turned out that a pie of information can be faster found again on the Web than in the own bookmark list

Google

[*] http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html

12

## Web 1.0 principles

STI · INNSBRUCK

- The success of Web1.0 is based on three simple principles:
  1. A simple and uniform addressing schema to indentify information chunks i.e. ***Uniform Resource Identifiers (URIs)***
  2. A simple and uniform representation formalism to structure information chunks allowing browsers to render them i.e. ***Hyper Text Markup Language (HTML)***
  3. A simple and uniform protocol to access information chunks i.e. ***Hyper Text Transfer Protocol (HTTP)***

13

## 1. Uniform Resource Identifiers (URIs)

STI · INNSBRUCK

- Uniform Resource Identifiers (URIs) are used to name/identify resources on the Web
- URIs are pointers to resources to which request methods can be applied to generate potentially different responses
- Resource can reside anywhere on the Internet
- Most popular form of a URI is the Uniform Resource Locator (URL)

14

## 2. Hyper-Text Markup Language (HTML)

STI · INNSBRUCK

- Hyper-Text Markup Language:
  - A subset of Standardized General Markup Language (SGML)
  - Facilitates a hyper-media environment
- Documents use elements to “mark up” or identify sections of text for different purposes or display characteristics
- HTML markup consists of several types of entities, including: elements, attributes, data types and character references
- Markup elements are not seen by the user when page is displayed
- Documents are rendered by browsers

15

## 3. Hyper-Text Transfer Protocol (HTTP)

STI · INNSBRUCK

- Protocol for client/server communication
  - The heart of the Web
  - Very simple request/response protocol
    - Client sends request message, server replies with response message
  - Provide a way to publish and retrieve HTML pages
  - Stateless
  - Relies on URI naming mechanism

16

STI · INNSBRUCK

# WEB 2.0

17

## Web 2.0

STI · INNSBRUCK

- “The term "**Web 2.0**" (2004–present) is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web”

http://en.wikipedia.org/wiki/Web_2.0

18

## Web 2.0

STI · INNSBRUCK

- Web 2.0 is a vaguely defined phrase referring to various topics such as social networking sites, wikis, communication tools, and folksonomies.
- Tim Berners-Lee is right that all these ideas are already underlying his original web ideas, however, there are differences in emphasis that may cause a qualitative change.
- With Web 1.0 technology a significant amount of software skills and investment in software was necessary to publish information.
- Web 2.0 technology changed this dramatically.

19

## Web 2.0 major breakthroughs

STI · INNSBRUCK

- The four major breakthroughs of Web 2.0 are:
  1. Blurring the distinction between content consumers and content providers.
  2. Moving from media for individuals towards media for communities.
  3. Blurring the distinction between service consumers and service providers
  4. Integrating human and machine computing in a new and innovative way

20

## 1. Blurring the distinction between content consumers and content providers

STI · INNSBRUCK

Wiki, Blogs, and Twiter turned the publication of text in mass phenomena, as flickr and youtube did for multimedia


21

## 2. Moving from a media for individuals towards a media for communities

STI · INNSBRUCK

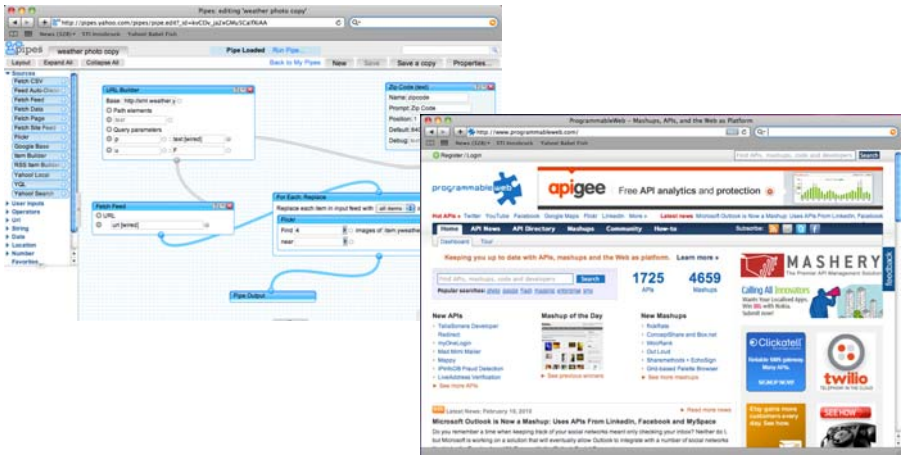
Social web sites such as del.icio.us, facebook, FOAF, linkedin, myspace and Xing allow communities of users to smoothly interweave their information and activities


22

## 3. Blurring the distinction between service consumers and service providers

STI · INNSBRUCK

Mashups allow web users to easy integrate services in their web site that were implemented by third parties

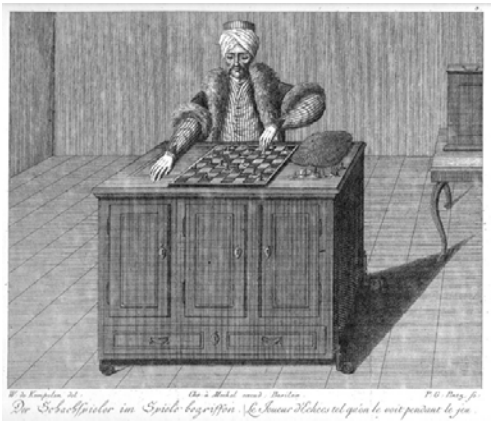



23

## 4. Integrating human and machine computing in a new way

STI · INNSBRUCK

Amazon Mechanical Turk - allows to access human services through a web service interface blurring the distinction between manually and automatically provided services



24

# LIMITATIONS OF THE CURRENT WEB



## Limitations of the current Web

- The current Web has its limitations when it comes to:
  1. finding relevant information
  2. extracting relevant information
  3. combining and reusing information

## Limitations of the current Web
*Finding relevant information*

STI · INNSBRUCK

- Finding information on the current Web is based on keyword search
- Keyword search has a limited recall and precision due to:
  - ***Synonyms***:
    - e.g. Searching information about “Cars” will ignore Web pages that contain the word “Automobiles” even though the information on these pages could be relevant
  - ***Homonyms:***
    - e.g. Searching information about “Jaguar” will bring up pages containing information about both “Jaguar” (the car brand) and “Jaguar” (the animal) even though the user is interested only in one of them


27

## Limitations of the current Web
*Finding relevant information*

STI · INNSBRUCK

- Keyword search has a limited recall and precision due also to:
  - ***Spelling variants:***
    - e.g. “organize” in American English vs. “organise” in British English
  - ***Spelling mistakes***
  - ***Multiple languages***
    - i.e. information about same topics in published on the Web on different languages (English, German, Italian,…)
- Current search engines provide no means to specify the relation between a resource and a term
  - e.g. sell / buy

28

## Limitations of the current Web
*Extracting relevant information*

STI · INNSBRUCK

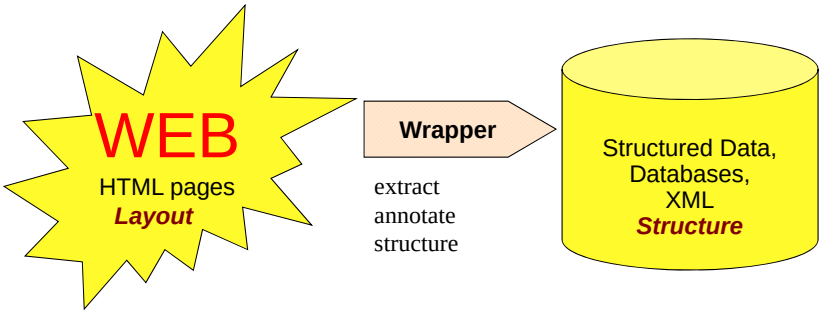
- One-fit-all automatic solution for extracting information from Web pages is not possible due to different formats, different syntaxes
- Even from a single Web page is difficult to extract the relevant information


29

## Limitations of the current Web
*Extracting relevant information*

STI · INNSBRUCK

- Extracting information from current web sites can be done using ***wrappers***




30

## Limitations of the current Web
*Extracting relevant information*

STI · INNSBRUCK

- The actual extraction of information from web sites is specified using standards such as XSL Transformation (XSLT) [1]
- Extracted information can be stored as structured data in XML format or databases.
- However, using wrappers do not really scale because the actual extraction of information depends again on the web site format and layout

[1] http://www.w3.org/TR/xslt

31

## Limitations of the current Web
*Combining and reusing information*

STI · INNSBRUCK

- Tasks often require to combine data on the Web
  1. Searching for the same information in different digital libraries
  2. Information may come from different web sites and needs to be combined

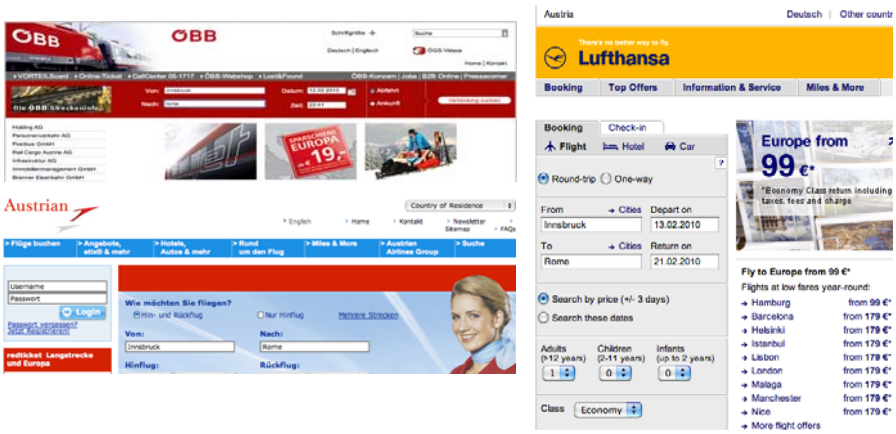
32

## Limitations of the current Web
*Combining and reusing information*

STI · INNSBRUCK

1. Searches for the same information in different digital libraries

**Example: I want travel from Innsbruck to Rome.**



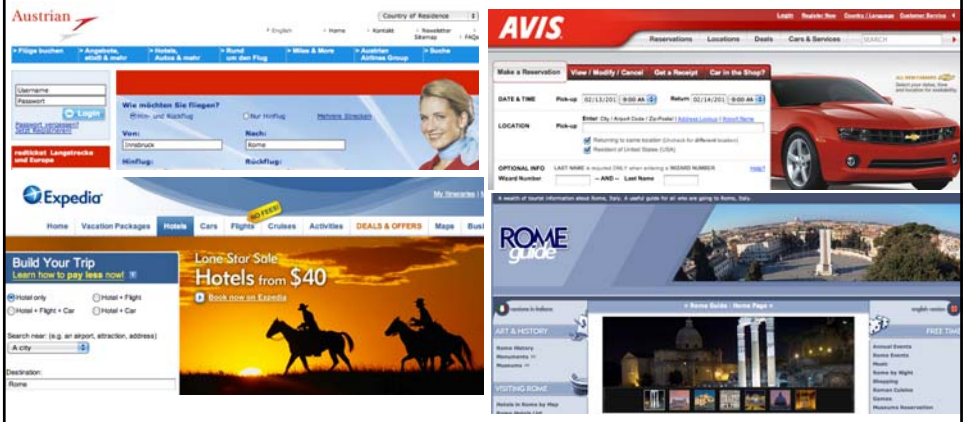



## Limitations of the current Web
*Combining and reusing information*

STI · INNSBRUCK

2. Information may come from different web sites and needs to be combined

**Example: I want to travel from Innsbruck to Rome where I want to stay in a hotel and visit the city**



## How to improve the current Web?

STI · INNSBRUCK

- Increasing automatic linking among data
- Increasing recall and precision in search
- Increasing automation in data integration
- Increasing automation in the service life cycle

- Adding semantics to data and services is the solution!

35

STI · INNSBRUCK

# TECHNICAL SOLUTION

36

STI · INNSBRUCK

# INTRODUCTION TO SEMANTIC WEB



## The Vision

STI · INNSBRUCK

More than 3 billion users,
more than a trillion pages (2016)

Static

**WWW**

**URI, HTML, HTTP**

http://www.internetlivestats.com/internet-users/

## The Vision (contd.)

STI · INNSBRUCK

Serious problems in

- information finding,
- information extracting,
- information representing,
- information interpreting and
- and information maintaining.

Static

**WWW**
**URI, HTML, HTTP**



**Semantic Web**
**RDF, RDF(S), OWL**

39

## What is the Semantic Web?

STI · INNSBRUCK

- *"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."*
  T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web", Scientific American, May 2001.

40

## What is the Semantic Web?

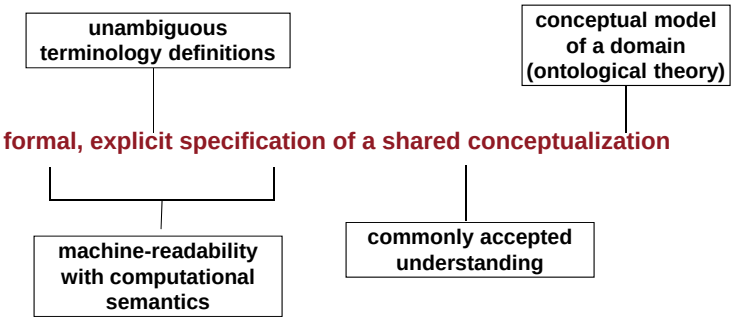
STI · INNSBRUCK

- The next generation of the WWW
- Information has machine-processable and machine-understandable semantics
- Not a separate Web but an augmentation of the current one
- The backbone of Semantic Web are ***ontologies***

41

## Ontology definition

STI · INNSBRUCK




Gruber, "Toward principles for the design of ontologies used or knowledge sharing?" , Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6,1995.

42

## … "well-defined meaning" …

STI · INNSBRUCK

- *"An ontology is an **explicit specification** of a **conceptualization**"*
  Gruber, "Toward principles for the design of ontologies used for knowledge sharing?" , Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6,1995.
- Ontologies are the modeling foundations to Semantic Web
  - They provide the ***well-defined meaning*** for information

43

## … explicit, … specification, … conceptualization, …

STI · INNSBRUCK

An ontology is:

- A conceptualization
  - An ontology is a model of the most relevant concepts of a phenomenon from the real world
- Explicit
  - The model explicitly states the type of the concepts, the relationships between them and the constraints on their use
- Formal
  - The ontology has to be machine readable (the use of the natural language is excluded)
- Shared
  - The knowledge contained in the ontology is consensual, i.e. it has been accepted by a group of people.

Studer, Benjamins, D. Fensel, "Knowledge engineering: Principles and methods", Data Knowledge Engineering, vol. 25, no. 1-2, 1998.

44

## Ontology example

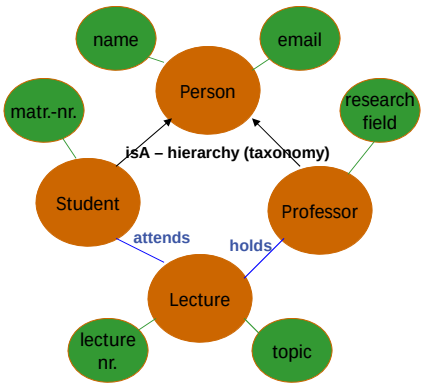
STI · INNSBRUCK

**Concept**
conceptual entity of the domain

**Property**
attribute describing a concept

**Relation**
relationship between concepts or properties

**Axiom**
coherency description between Concepts / Properties / Relations via logical expressions




holds(Professor, Lecture) =>
Lecture.topic = Professor.researchField

45

## Types of ontologies

STI · INNSBRUCK

*describe very **general concepts like space, time, event**, which are independent of a particular problem or domain*

Top Level O., Generic O. Core O., Foundational O., High-level O, Upper O.

*describe the vocabulary related to a generic task or activity **by specializing the top-level ontologies**.*

Domain Ontology

Task & Problem-solving Ontology

*describe the vocabulary related to a generic domain **by specializing the concepts introduced in the top-level ontology**.*

Application Ontology

*the most specific ontologies. Concepts in application ontologies often correspond to roles played by domain entities while performing a certain activity.*

Guarino, N. (1998). *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy* (Vol. 46). IOS press.
http://www.lirmm.fr/~mugnier/DEA/guarino98formal.pdf

46

## The Semantic Web is about…

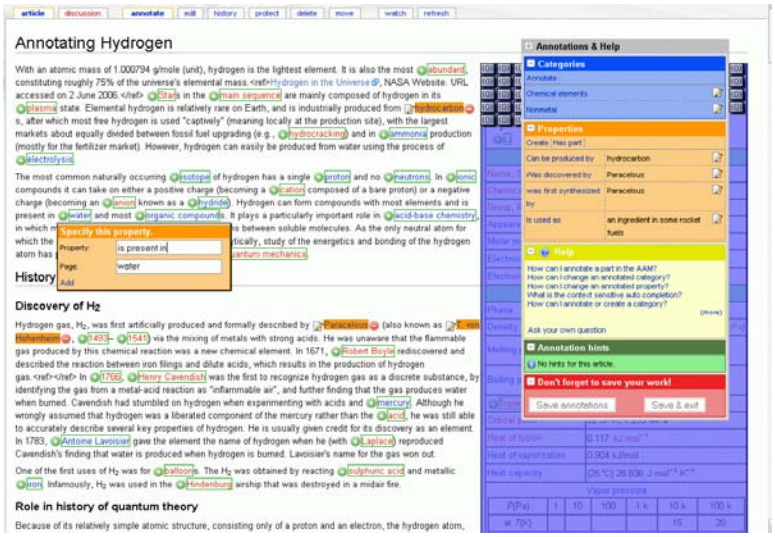
STI · INNSBRUCK

- Web Data Annotation
  - connecting (syntactic) Web objects, like text chunks, images, … to their semantic notion (e.g., this image is about Innsbruck, Anna Fensel is a lecturer)
- Data Linking on the Web (Web of Data)
  - global networking of knowledge through URI, RDF, and SPARQL (e.g., connecting my calendar with my rss feeds, my pictures, ...)
- Data Integration over the Web
  - seamless integration of data based on different conceptual models (e.g., integrating data coming from my two favorite book sellers)

47

## Web Data Annotating

STI · INNSBRUCK




Ontoprise (formerly), now: http://www.semafora-systems.com

48

## Data integration over the Web

STI · INNSBRUCK

- Data integration involves combining data residing in different sources and providing user with a unified view of these data
- Data integration over the Web can be implemented as follows:
  1. **Export the data sets to be integrated as RDF graphs**
  2. **Merge identical resources (i.e. resources having the same URI) from different data sets**
  3. **Start making queries on the integrated data, queries that were not possible on the individual data sets.**

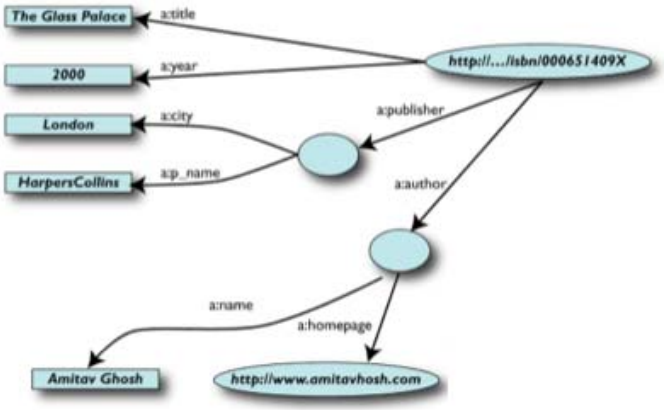
49

## Data integration over the Web

STI · INNSBRUCK

1. **Export first data set as RDF graph**

   For example the following RDF graph contains information about book “The Glass Palace” by Amitav Ghosh



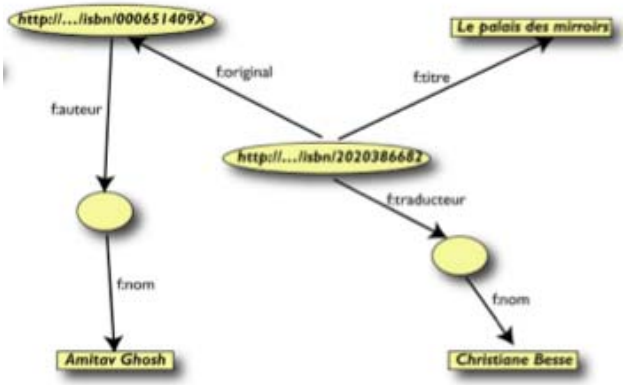

http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/Slides.pdf

50

## Data integration over the Web

STI · INNSBRUCK

1. **Export second data set as RDF graph**
   Information about the same book but in French this time is modeled in RDF graph below



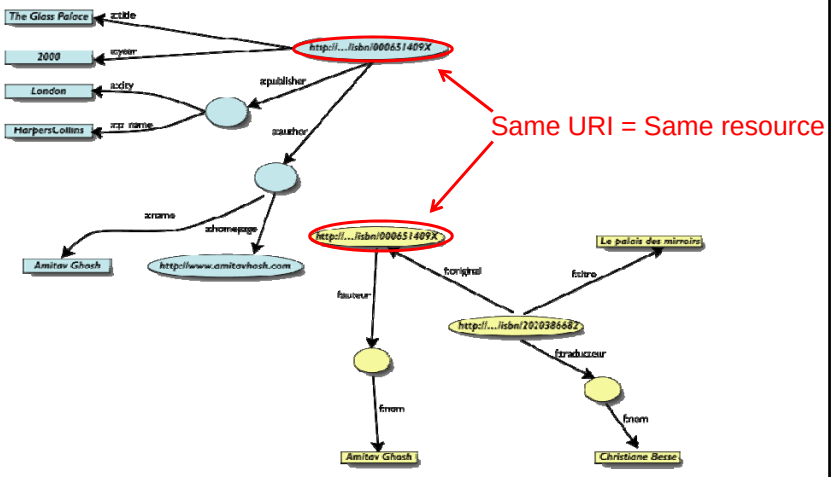

http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/Slides.pdf

51

## Data Integration over the Web

STI · INNSBRUCK

2. **Merge identical resources (i.e. resources having the same URI) from different data sets**



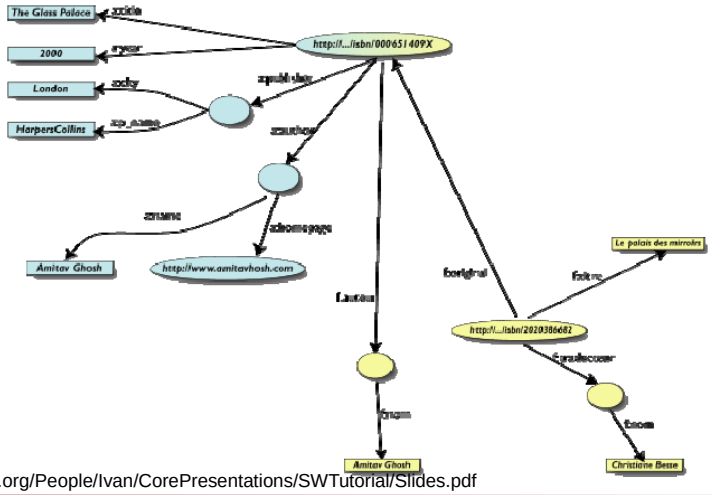

http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/Slides.pdf

52

## Data integration over the Web

STI · INNSBRUCK

**2. Merge identical resources (i.e. resources having the same URI) from different data sets**




http://www.w3.org/People/Ivan/CorePresentations/SWTutorial/Slides.pdf

53

## Data integration over the Web

STI · INNSBRUCK

**3. Start making queries on the integrated data**

- A user of the second dataset may ask queries like*: "give me the title of the original book"*
- This information is not in the second dataset
- This information can be however retrieved from the integrated dataset, in which the second dataset was connected with the the first dataset

54

STI · INNSBRUCK

# ARCHITECTURE AND LANGUAGES



## Web Architecture

STI · INNSBRUCK

- Things are denoted by URIs
- Use them to denote things
- Serve useful information at them
- Dereference them

## Semantic Web Architecture

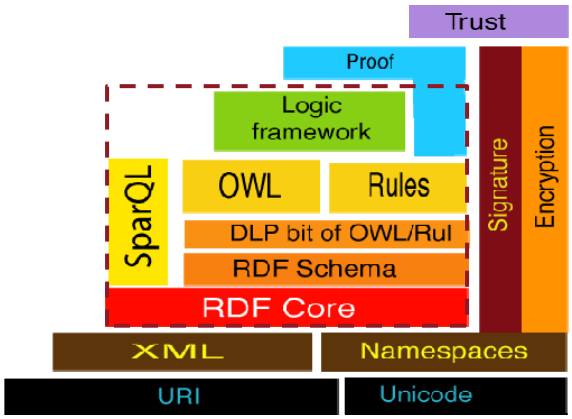
STI · INNSBRUCK

- Give important concepts URIs
- Each URI identifies one concept
- Share these symbols between many languages
- Support URI lookup

57

## Semantic Web - Data

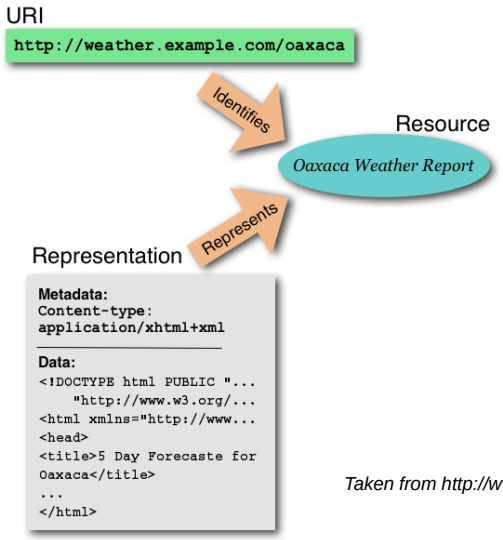
STI · INNSBRUCK




58

## Identifier, Resource, Representation

STI · INNSBRUCK



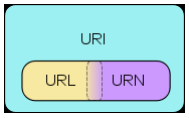

*Taken from http://www.w3.org/TR/webarch/*

59

## URI, URN, URL

STI · INNSBRUCK

- A Uniform Resource Identifier (URI) is a string of characters used to identify a name or a resource on the Internet




- A URI can be a URL or a URN
- A Uniform Resource Name (URN) defines an item's identity
  - the URN *urn:isbn:0-395-36341-1* is a URI that specifies the identifier system, i.e. International Standard Book Number (ISBN), as well as the unique reference within that system and allows one to talk about a book, but doesn't suggest where and how to obtain an actual copy of it
- A Uniform Resource Locator (URL) provides a method for finding it
  - the URL *http://www.sti-innsbruck.at/* identifies a resource (STI's home page) and implies that a representation of that resource (such as the home page's current HTML code, as encoded characters) is obtainable via HTTP from a network host named www.sti-innsbruck.at

60

## eXtensible Markup Language (XML)

STI · INNSBRUCK

- Language for creating languages
  - "Meta-language"
  - XHTML is a language: HTML expressed in XML
- W3C Recommendation (standard)
  - XML is, for the information industry, what the container is for international shipping
  - For structured and semistructured data
- Main plus: wide support, interoperability
  - Platform-independent
- Applying new tools to old data

61

## XML Schema Definition (XSD)

STI · INNSBRUCK

- A grammar definition language
  - Like DTDs but better
    - Uses XML syntax
  - Defined by W3C
- Primary features
  - Datatypes
    - e.g. integer, float, date, etc…
  - More powerful content models
    - e.g. namespace-aware, type derivation, etc…

62

## Resource Description Framework (RDF)

STI · INNSBRUCK

- The Resource Description Framework (RDF) provides a domain independent data model
- Resource (identified by URIs)
  - Correspond to nodes in a graph
  - E.g.:
    `http://www.w3.org/`
    `http://example.org/#john`
    `http://www.w3.org/1999/02/22-rdf-syntax-ns#Property`
- Properties (identified by URIs)
  - Correspond to labels of edges in a graph
  - Binary relation between two resources
  - E.g.:
    `http://www.example.org/#hasName`
    `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`
- Literals
  - Concrete data values
  - E.g.:
    `"John Smith", "1", "2006-03-07"`

63

## Resource Description Framework (RDF) – Triple Data Model

STI · INNSBRUCK

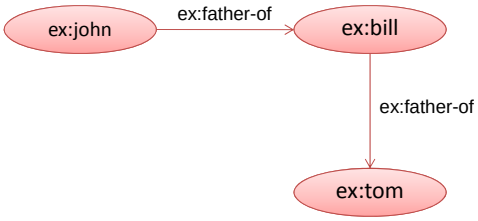
- Triple data model:
  `<subject, predicate, object>`
  - **Subject**: Resource or blank node
  - **Predicate**: Property
  - **Object**: Resource, literal or blank node
- Example:
  `<ex:john, ex:father-of, ex:bill>`
- Statement (or triple) as a logical formula P(x, y), where the binary predicate P relates the object x to the object y.
- RDF offers only binary predicates (properties).

64

## Resource Description Framework (RDF) – Graph Model

STI · INNSBRUCK

- The triple data model can be represented as a graph
- Such graph is called in the Artificial Intelligence community a **semantic net**
- Labeled, directed graphs
  - **Nodes**: resources, literals
  - **Labels**: properties
  - **Edges**: statements




65

## RDF Schema (RDFS)

STI · INNSBRUCK

- RDF Schema (RDFS) is a language for capturing the semantics of a domain, for example:
  - In RDF:
    `<#john, rdf:type, #Student>`
  - What is a "`#Student`"?
- RDFS is a language for defining RDF types:
  - Define classes:
    - "*`#Student` is a class*"
  - Relationships between classes:
    - "*`#Student` is a sub-class of `#Person`*"
  - Properties of classes:
    - "*`#Person` has a property `hasName`*"

66

## RDF Schema (RDFS)

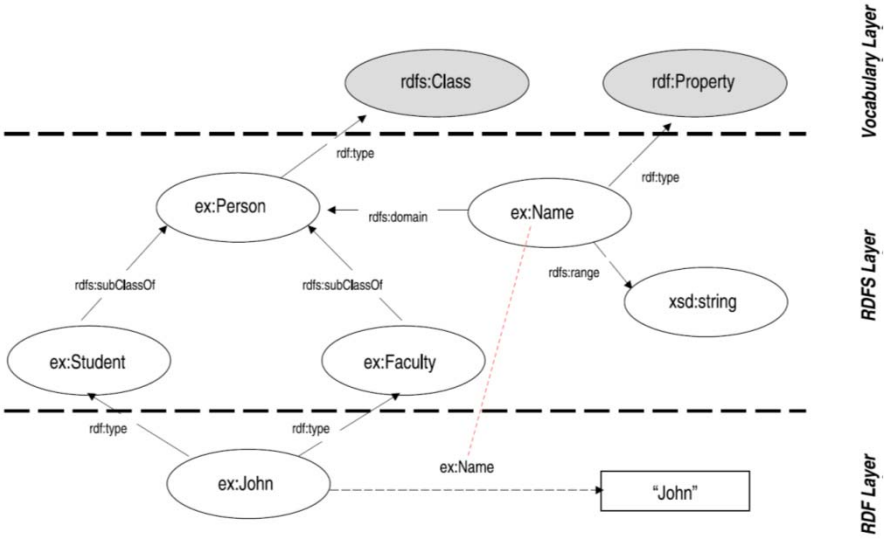
STI · INNSBRUCK

- Classes:
  `<#Student, rdf:type, #rdfs:Class>`
- Class hierarchies:
  `<#Student, rdfs:subClassOf, #Person>`
- Properties:
  `<#hasName, rdf:type, rdf:Property>`
- Property hierarchies:
  `<#hasMother, rdfs:subPropertyOf, #hasParent>`
- Associating properties with classes (a):
  - "The property `#hasName` only applies to `#Person`"
    `<#hasName, rdfs:domain, #Person>`
- Associating properties with classes (b):
  - "The type of the property `#hasName` is `#xsd:string`"
    `<#hasName, rdfs:range, xsd:string>`



## RDF Schema (RDFS) - Example

STI · INNSBRUCK



## Web Ontology Language (OWL)

STI · INNSBRUCK

- RDFS has a number of Limitations:
  - Only binary relations
  - Characteristics of Properties, e.g. *inverse*, *transitive*, *symmetric*
  - Local range restrictions, e.g. for class *Person*, the property *hasName* has range *xsd:string*
  - Complex concept descriptions, e.g. *Person* is defined by *Man* and *Woman*
  - Cardinality restrictions, e.g. a *Person* may have at most 1 *name*
  - Disjointness axioms, e.g. nobody can be both a *Man* and a *Woman*
- The Web Ontology Language (OWL) provides an ontology language, that is a more expressive Vocabulary Definition Language for use with RDF
  - Class membership
  - Equivalance of classes
  - Consistency
  - Classification

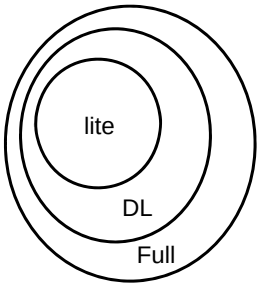
69

## OWL

STI · INNSBRUCK

- OWL is layered into languages of different expressiveness
  - OWL Lite: Classification Hierarchies, Simple Constraints
  - OWL DL: Maximal expressiveness while maintaining tractability
  - OWL Full: Very high expressiveness, loses tractability, all syntactic freedom of RDF
- More expressive means harder to reason with
- Different Syntaxes:
  - RDF/XML (Recommended for Serialization)
  - N3 (Recommended for Human readable Fragments)
  - Abstract Syntax (Clear Human Readable Syntax)
- OWL is a W3C recommendation since 2004.

There is also its 2nd version, OWL 2, a W3C recommendation from 2009.




70

## OWL – Example: The Wine Ontology

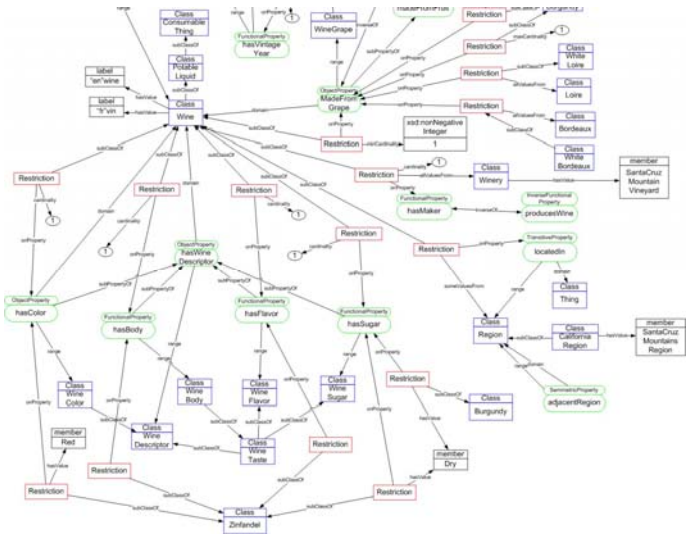
STI · INNSBRUCK

- An Ontology describing wine domain.
- One of the most widely used examples for OWL and referenced by W3C.
- There is also a wine agent associated to this ontology that performs OWL queries using a web-based ontological mark-up language. That is, by combining a logical reasoner with an OWL ontology.
- The agent's operation can be described in three parts: consulting the ontology, performing queries and outputting results.
- Available here: http://www.w3.org/TR/owl-guide/

71

## OWL – Example: The Wine Ontology Schema

STI · INNSBRUCK



https://sites.google.com/site/semanticsimulations2/visioowl

72

## SPARQL – Querying RDF

STI · INNSBRUCK

- SPARQL
  - RDF Query language
  - Based on RDQL
  - Uses SQL-like syntax
- Example:

```
PREFIX uni: <http://example.org/uni/>

SELECT ?name
FROM <http://example.org/personal>
WHERE { ?s uni:name ?name.
?s rdf:type uni:lecturer }
```

73

## SPARQL Queries

STI · INNSBRUCK

```
PREFIX uni: <http://example.org/uni/>
SELECT ?name
FROM <http://example.org/personal>
WHERE { ?s uni:name ?name. ?s rdf:type uni:lecturer }
```

- PREFIX
  - Prefix mechanism for abbreviating URIs
- SELECT
  - Identifies the variables to be returned in the query answer
  - SELECT DISTINCT
  - SELECT REDUCED
- FROM
  - Name of the graph to be queried
  - FROM NAMED
- WHERE
  - Query pattern as a list of triple patterns
- LIMIT
- OFFSET
- ORDER BY

74

## SPARQL Example Query 1

STI · INNSBRUCK

*"Return the full names of all people in the graph"*

```
PREFIX vCard: <http://www.w3.org/2001/vcard-rdf/3.0#>
SELECT ?fullName
WHERE {?x vCard:FN ?fullName}
```

*result:*

```
fullName
=================
"John Smith"
"Mary Smith"
```

```
@prefix ex: <http://example.org/#> .
@prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
ex:john
  vcard:FN "John Smith" ;
  vcard:N [
    vcard:Given "John" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 32 ;
  ex:marriedTo :mary .
ex:mary
  vcard:FN "Mary Smith" ;
  vcard:N [
    vcard:Given "Mary" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 29 .
```

75

## SPARQL Example Query 2

STI · INNSBRUCK

*"Return the relation between John and Mary"*

```
PREFIX ex: <http://example.org/#>
SELECT ?p
WHERE {ex:john ?p ex:mary}
```

*result:*

```
p
=================
<http://example.org/#marriedTo>
```

```
@prefix ex: <http://example.org/#> .
@prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
ex:john
  vcard:FN "John Smith" ;
  vcard:N [
    vcard:Given "John" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 32 ;
  ex:marriedTo :mary .
ex:mary
  vcard:FN "Mary Smith" ;
  vcard:N [
    vcard:Given "Mary" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 29 .
```

76

## SPARQL Example Query 3

STI · INNSBRUCK

*"Return the spouse of a person by the name of John Smith"*

```
PREFIX vCard: <http://www.w3.org/2001/vcard-rdf/3.0#>
PREFIX ex: <http://example.org/#>
SELECT ?y
WHERE {?x vCard:FN "John Smith".
       ?x ex:marriedTo ?y}
```

*result:*

```
y
==================
<http://example.org/#mary>
```

```
@prefix ex: <http://example.org/#> .
@prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#> .
ex:john
  vcard:FN "John Smith" ;
  vcard:N [
    vcard:Given "John" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 32 ;
  ex:marriedTo :mary .
ex:mary
  vcard:FN "Mary Smith" ;
  vcard:N [
    vcard:Given "Mary" ;
    vcard:Family "Smith" ] ;
  ex:hasAge 29 .
```

77

## SPARQL and Rule languages

STI · INNSBRUCK

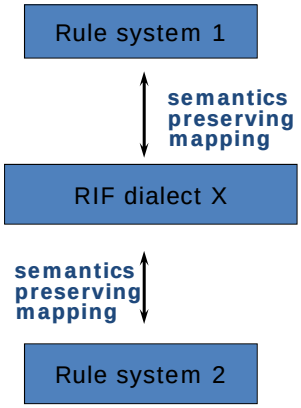
- SPARQL
  - Query language for RDF triples
  - A protocol for querying RDF data over the Web
- Rule languages (e.g. SWRL)
  - Extend basic predicates in ontology languages with proprietary predicates
  - Based on different logics
    - Description Logic
    - Logic Programming

78

## Rule Interchange Format (RIF)

STI · INNSBRUCK

- A set of dialects to enable rule exchange among different rule systems




79

## Rule Interchange Format Goals

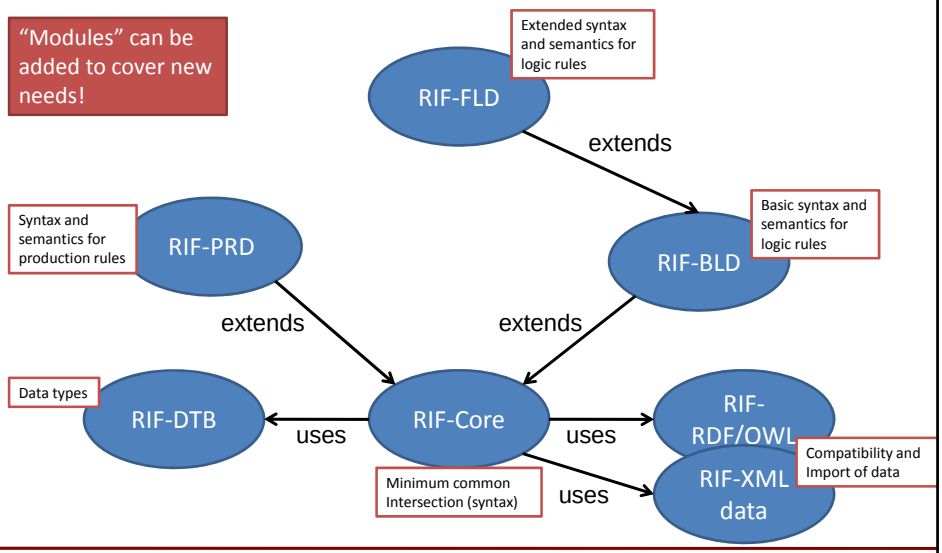
STI · INNSBRUCK

- Exchange of rules
  - The primary goal of RIF is to facilitate the exchange of rules
- Consistency with W3C specifications
  - A W3C specification that builds on and develops the existing range of specifications that have been developed by the W3C
  - Existing W3C technologies should fit well with RIF
- Wide scale adoption
  - Rules interchange becomes more effective the wider is their adoption ("network effect“)

80

## RIF Architecture

STI · INNSBRUCK




81

## RIF Dialects

STI · INNSBRUCK

- RIF Core
  - A language of definite Horn rules without function symbols (~ Datalog)
  - A language of production rules where conclusions are interpreted as assert actions
- RIF BLD
  - A language that lies within the intersection of first-order and logic-programming systems
- RIF FLD
  - A formalism for specifying all logic dialects of RIF
  - Syntax and semantics described mechanisms that are commonly used for various logic languages (but rarely brought all together)
- RIF PRD
  - A formalism for specifying production rules
- Other common specifications
  - RIF DTB – Defines data types and builtins supported by RIF
  - RIF OWL/RDF compatibility – Defines how OWL and RDF can be used within RIF
  - RIF XML data - Defines how XML can be used within RIF

82

STI · INNSBRUCK

# SEMANTIC WEB - DATA



## Semantic Web - Data

STI · INNSBRUCK

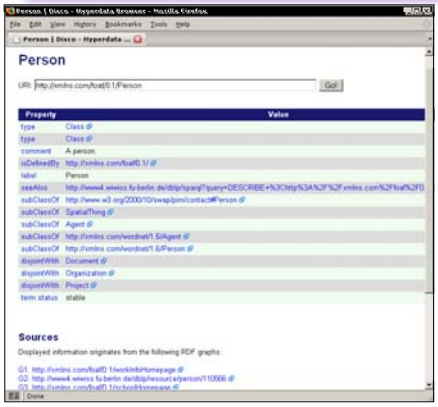
- URIs are used to identify **resources**, not just things that exists on the Web, e.g. *Sir Tim Berners-Lee*
- RDF is used to make statements about resources in the form of **triples**
  *<entity, property, value>*
- With RDFS, resources can belong to **classes** (*my Mercedes belongs to the class of cars*) and classes can be **subclasses** or **superclasses** of other classes (*vehicles are a superclass of cars, cabriolets are a subclass of cars*)

## Semantic Web - Data

STI · INNSBRUCK

Dereferencable URI

Disco Hyperdata Browser
navigating the Semantic Web as an unbound set of data sources

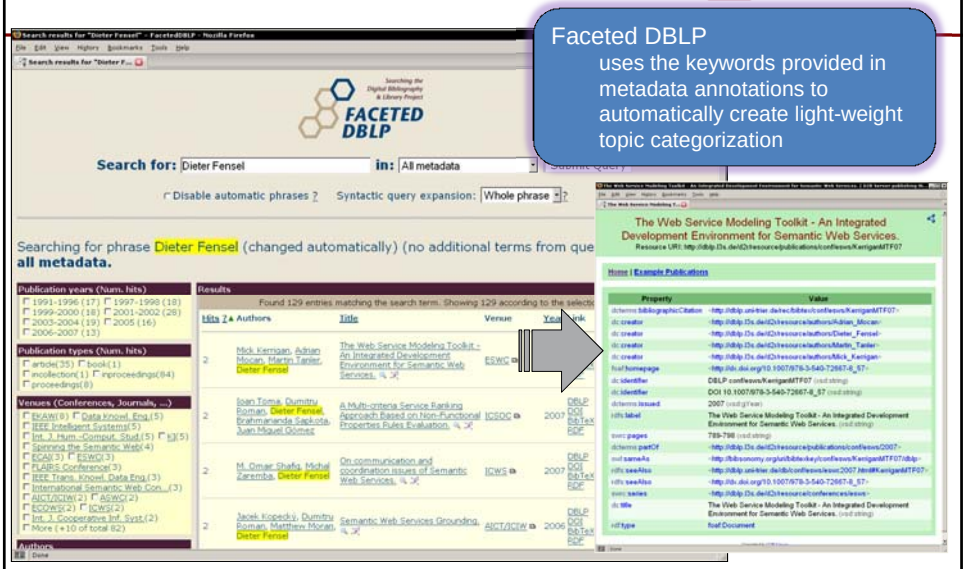




85

## Semantic Web - Data

STI · INNSBRUCK

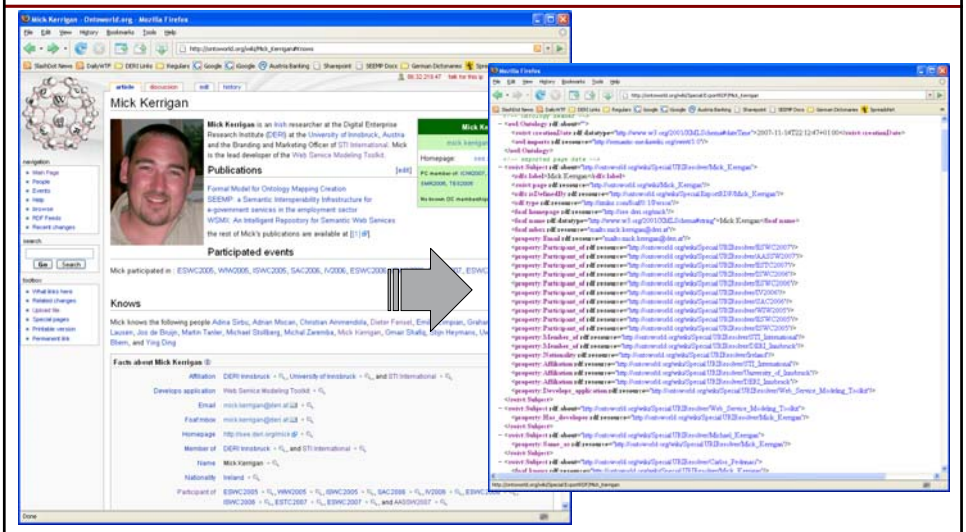
Faceted DBLP
uses the keywords provided in metadata annotations to automatically create light-weight topic categorization




86

## Semantic Web - Data

STI · INNSBRUCK

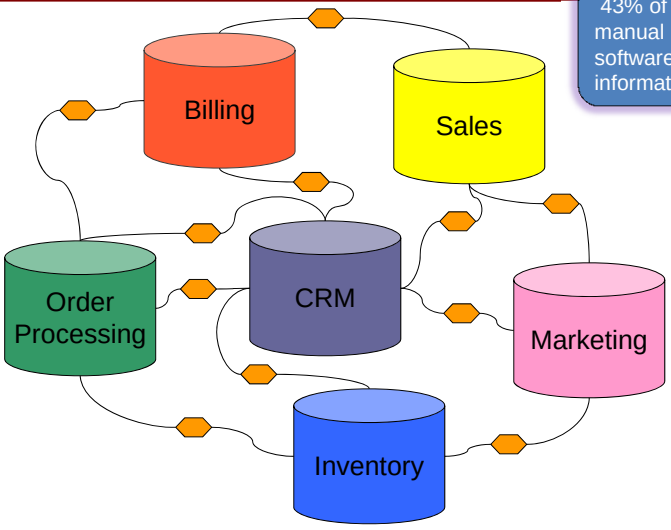



87

## Semantic Web - Data

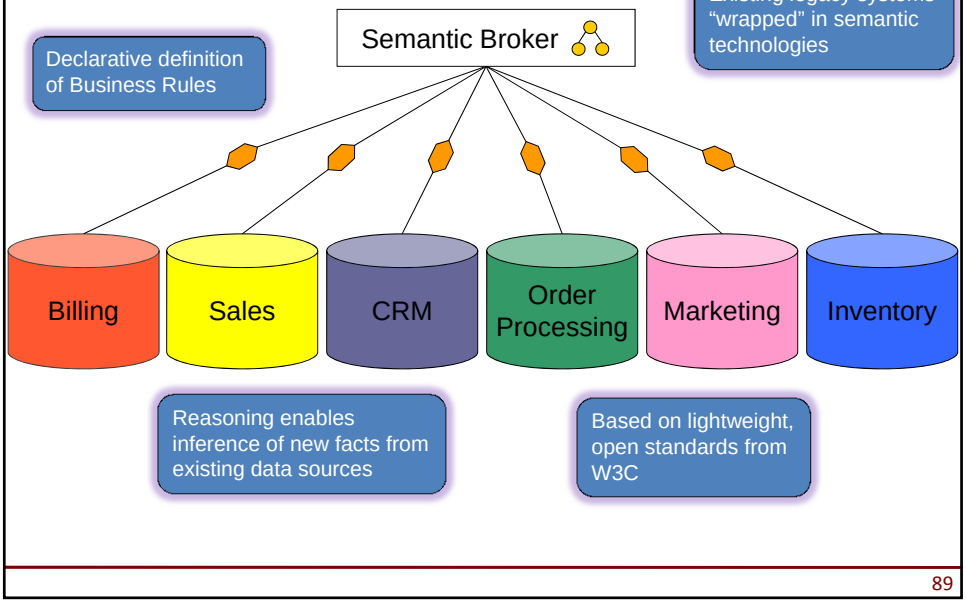
STI · INNSBRUCK

43% of businesses resort to manual processes and/or new software when integrating information for reporting




88

## Semantic Web - Data




89

## KIM platform

The KIM platform provides a novel infrastructure and services for:

– automatic semantic annotation,
– indexing,
– retrieval of unstructured and semi-structured content.

90

## KIM Constituents

STI · INNSBRUCK

The KIM Platform includes:

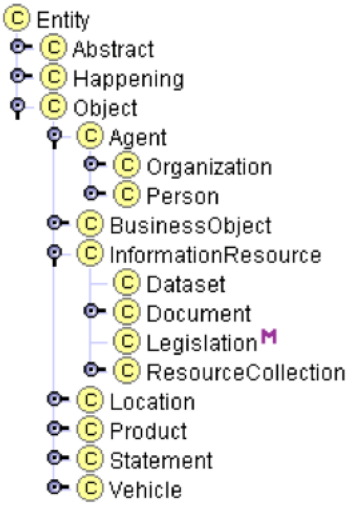
- Ontologies (PROTON + KIMSO + KIMLO) and KIM World KB
- KIM Server – with a set of APIs for remote access and integration
- Front-ends: Web-UI and plug-in for Internet Explorer.

91

## KIM Ontology (KIMO)

STI · INNSBRUCK

- light-weight upper-level ontology
- 250 NE classes
- 100 relations and attributes:
- covers mostly NE classes, and ignores general concepts
- includes classes representing lexical resources




92

## KIM KB

STI · INNSBRUCK

- KIM KB consists of above 80,000 entities (50,000 locations, 8,400 organization instances, etc.)
- Each location has geographic coordinates and several aliases (usually including English, French, Spanish, and sometimes the local transcription of the location name) as well as co-positioning relations (e.g. **subRegionOf.)**
- The organizations have **locatedIn** relations to the corresponding Country instances. The additionally imported information about the companies consists of short description, URL, reference to an industry sector, reported sales, net income,and number of employees.

93

## KIM is Based On…

STI · INNSBRUCK

KIM is based on the following open-source platforms:

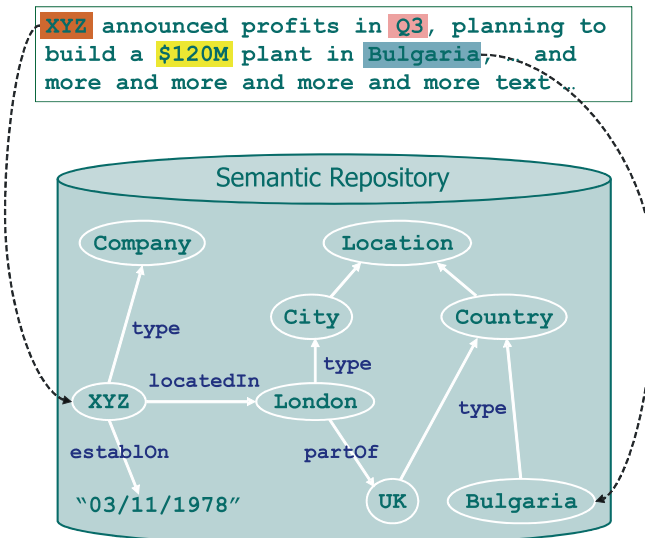
- GATE – the most popular NLP and IE platform in the world, developed at the University of Sheffield. Ontotext is its biggest co-developer. www.gate.ac.uk and www.ontotext.com/gate

- OWLIM – OWL repository, compliant with Sesame RDF database from Aduna B.V. (now OWLIM is called GraphDB) http://ontotext.com/products/graphdb/


- Lucene – an open-source IR engine by Apache. jakarta.apache.org/lucene/


94

## KIM Platform – Semantic Annotation




95

## KIM platform – Semantic Annotation

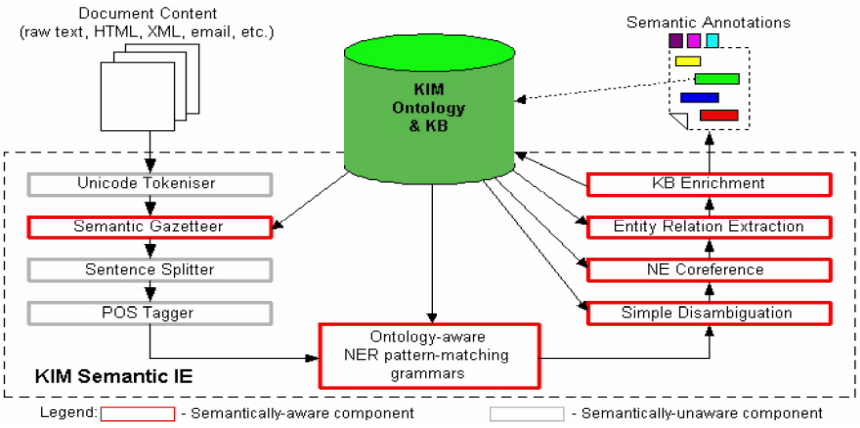
- The automatic semantic annotation is seen as a named-entity recognition (NER) and annotation process.
- The traditional flat NE type sets consist of several general types (such as Organization, Person, Date, Location, Percent, Money). In KIM the NE type is specified by reference to an ontology.
- The semantic descriptions of entities and relations between them are kept in a knowledge base (KB) encoded in the KIM ontology and residing in the same semantic repository. Thus KIM provides for each entity reference in the text (i) a link (URI) to the most specific class in the ontology and (ii) a link to the specific instance in the KB. Each extracted NE is linked to its specific type information (thus Arabian Sea would be identified as **Sea**, instead of the traditional – **Location).**

96

## KIM platform – Information Extraction

STI · INNSBRUCK

- KIM performs IE based on an ontology and a massive knowledge base.

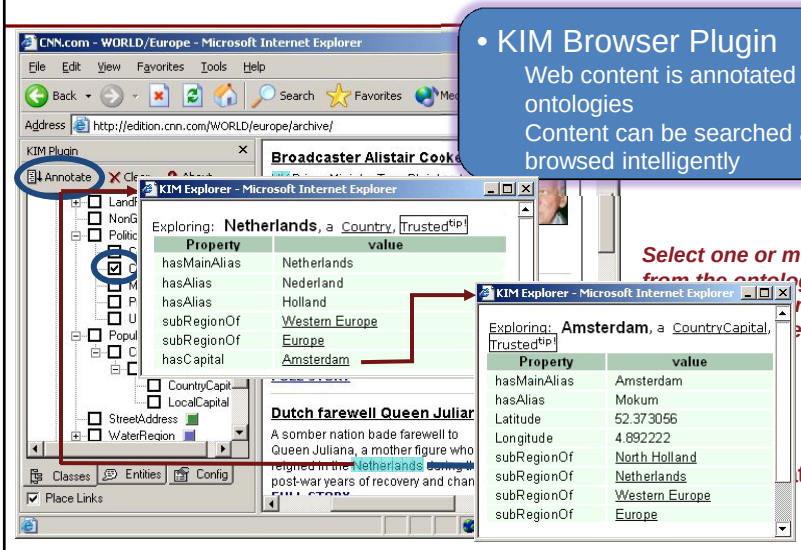



97

## KIM platform - Browser Plug-in

STI · INNSBRUCK

- KIM Browser Plugin
  - Web content is annotated using ontologies
  - Content can be searched and browsed intelligently




*Select one or more concepts from the ontology…*
*…rently loaded*
*…e Annotation*

*…ted Content*

98

STI · INNSBRUCK

# EXTENSIONS



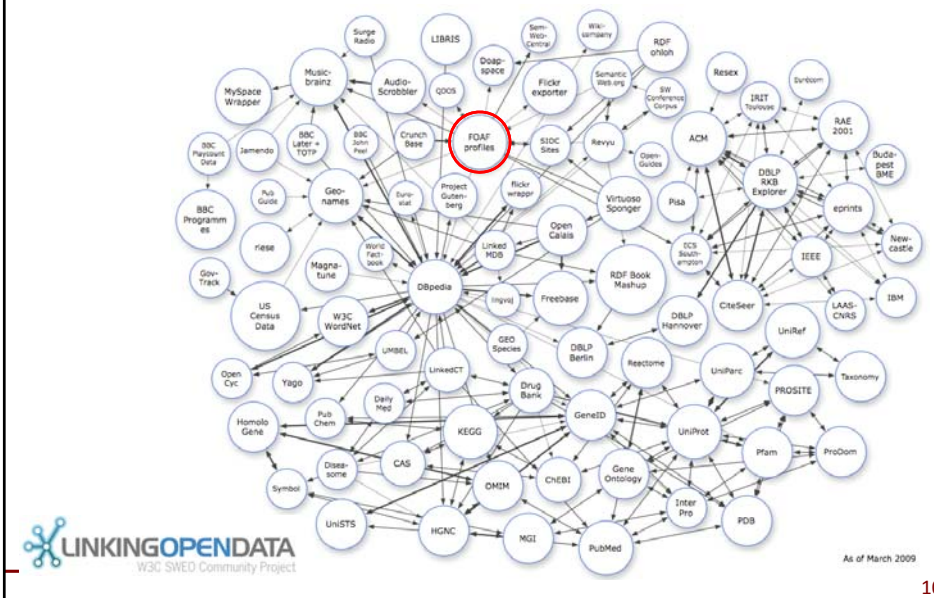
## Extensions: Linked Open Data

STI · INNSBRUCK

- Linked Data is a method for exposing and sharing connected data via dereferenceable URI's on the Web
  - Use URIs to identify things that you expose to the Web as resources
  - Use HTTP URIs so that people can locate and look up (dereference) these things
  - Provide useful information about the resource when its URI is dereferenced
  - Include links to other, related URIs in the exposed data as a means of improving information discovery on the Web
- Linked Open Data is an initiative to interlink open data sources
  - Open: Publicly available data sets that are accessible to everyone
  - Interlinked: Datasets have references to one another allowing them to be used together

## Extensions: Linked Open Data

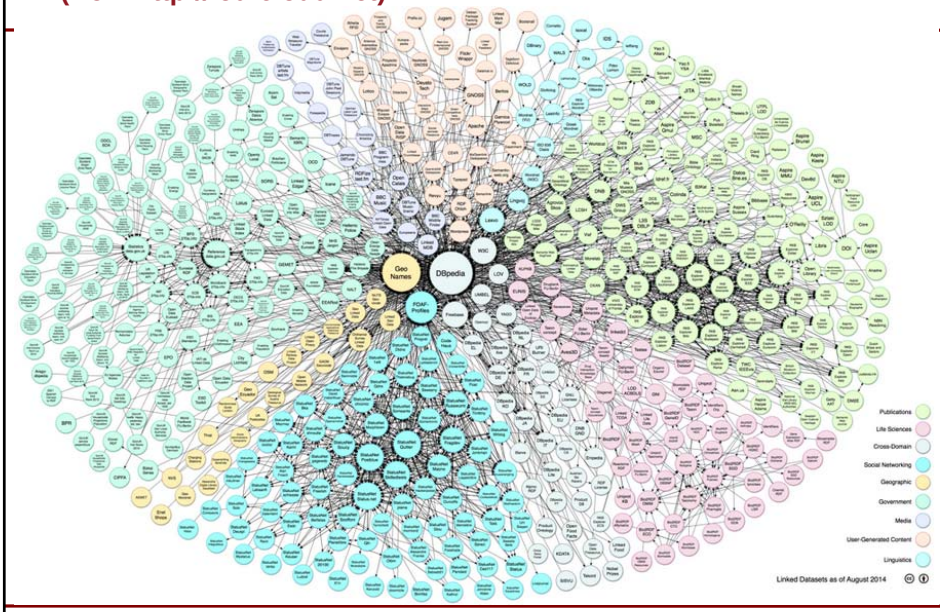
STI · INNSBRUCK






## Linked Open Data Cloud – August 2014 (from http://lod-cloud.net)

STI · INNSBRUCK



## Extensions: Linked Open Data

STI · INNSBRUCK

- Linked Open Data statistics:
  - 2009: 121 data sets, total number of triples: 13.112.409.691, total number of links between data sets: 142.605.717
  - 2014: 1014 data sets, 900,129 documents describing 8,038,396 resources, most - 18,05% from government sector
  - 2017: 1163 datasets

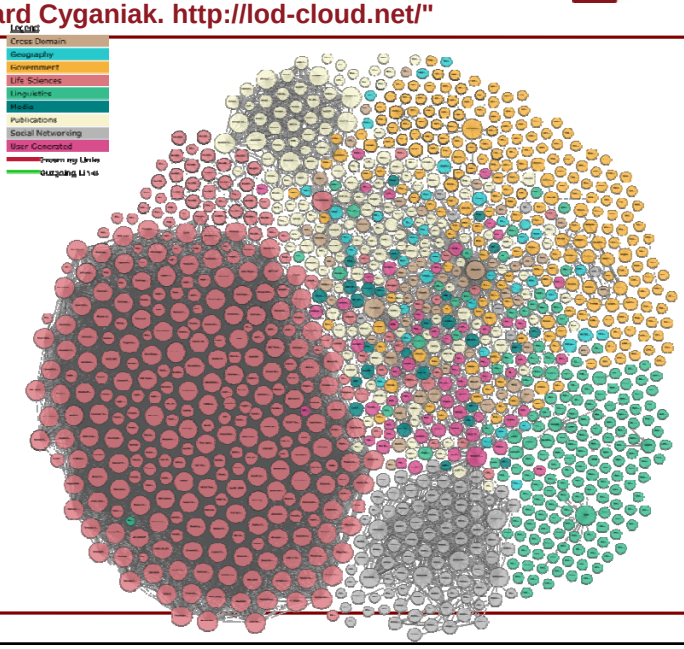
Statistics are available at:

- For 2009:
  - http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics
  - http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/LinkStatistics
- For 2014: http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/
- For 2017: http://lod-cloud.net

103

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

STI · INNSBRUCK




104

## Extensions: Linked Open Data principles

STI · INNSBRUCK

- Use URIs as names for things
  - anything, not just documents
  - you are not your homepage
  - information resources and non-information resources
- Use HTTP URIs
  - globally unique names, distributed ownership
  - allows people to look up those names
- Provide useful information in RDF
  - when someone looks up a URI
- Include RDF links to other URIs
  - to enable discovery of related information

105

## Extensions: Linked Open Data - FOAF

STI · INNSBRUCK

- Friend Of A Friend (FOAF) provides a way to create machine-readable pages about:
  - People
  - The links between them
  - The things they do and create
- Anyone can publish a FOAF file on the web about themselves and this data becomes part of the Web of Data

```
<foaf:Person>
        <foaf:name>Dieter Fensel</foaf:name>
        <foaf:homepage rdf:resource="http://www.fensel.com"/>
</foaf:Person>
```

- FOAF is connected to many other data sets, including
  - Data sets describing music and musicians (Audio Scrobbler, MusicBrainz)
  - Data sets describing photographs and who took them (Flickr)
  - Data sets describing places and their relationship (GeoNames)

106

## Extensions: Linked Open Data - GeoNames

STI · INNSBRUCK

- The GeoNames Ontology makes it possible to add geospatial semantic information to the Web of Data
- We can utilize GeoNames location within the FOAF profile

```
<foaf:Person>
        <foaf:name>Dieter Fensel</foaf:name>
        <foaf:homepage rdf:resource="http://www.fensel.com"/>
        <foaf:based_near " http://ws.geonames.org/rdf?geonameId=2775220"/>
</foaf:Person>
```

- GeoNames is also linked to more datasets
  - US Census Data
  - Movie Database (Linked MDB)
  - Extracted data from Wikipedia (DBpedia)

107

## Extensions: Linked Open Data - DBpedia

STI · INNSBRUCK

- Dbpedia (www.dbpedia.org) is a community effort to:
  - Extract structured information from Wikipedia
  - Make the information available on the Web under an open license
  - Interlink the DBpedia dataset with other open datasets on the Web
- DBpedia is one of the central interlinking-hubs of the emerging Web of Data
- Formally, it is also a non-profit association

Content on this slide adapted from Anja Jentzsch and Chris Bizer

108

## Extensions: Linked Open Data - DBpedia

STI · INNSBRUCK

- As our FOAF profile has been linked to GeoNames, and GeoNames is linked to DBpedia, we can ask some interesting queries over the Web of Data
  - What is the population of the city in which Anna Fensel lives?
    => 124,579 people
  - At which elevation does Anna Fensel live?
    => 574m
  - Who is the mayor of the city in which Anna Fensel lives
    => Christine Oppitz-Plörer



## Extensions: Linked Open Data – Dbpedia Dataset

STI · INNSBRUCK

- 125 languages
- Describes 4.58 million things, out of which 4.22 million are classified in a consistent ontology (http://wiki.dbpedia.org/Ontology2014), including
  - 1,445,000 persons,
  - 735,000 places (including 478,000 populated places),
  - 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games),
  - 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and
  - 6,000 diseases.
- DBpedia 2014 release consists of 3 billion pieces of information (RDF triples) out of which 580 million were extracted from the English edition of Wikipedia, 2.46 billion were extracted from other language editions.

Source: http://wiki.dbpedia.org/about/about-dbpedia/facts-figures (April 2016)

## Linked Open Data and Mobiles

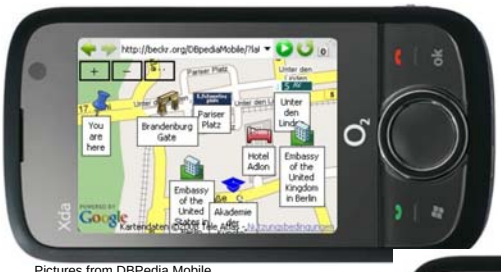
STI · INNSBRUCK

- Combination of Linked Open Data and Mobiles has trigger the emergence of new applications
- One example is **DBpedia Mobile** that based on the current GPS position of a mobile device renders a map containing information about nearby locations from the DBpedia dataset.
- It exploits information coming from DBpedia, Revyu and Flickr data.
- It provides a way to explore maps of cities and gives pointers to more information which can be explored

111

## Linked Open Data and Mobiles

STI · INNSBRUCK




Pictures from DBPedia Mobile


Try yourself: http://wiki.dbpedia.org/projects/dbpedia-mobile

112

## What is Schema.org?

STI · INNSBRUCK

- **Schema.org** provides a collection of shared vocabularies.
- Launched in June 2011 by Bing, Google and Yahoo
- Yandex joins in November 2011
- Purpose:

Create a common set of schemas for webmasters to mark-up with structured data their websites.

bing Google YAHOO! Yandex

113

## How to mark-up with schema.org?

STI · INNSBRUCK

- Schema.org can be used to enrich the web sites with the following formats:
  - Microdata (most popular)
    - Tags introduced within HTML 5
    - Based on Item descriptions
    - Itemscope, Itemtype, Itemprop
  - RDFa
  - JSON-LD



114

```
{
 "@context": {
  "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
  "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
  "schema": "http://schema.org/",
  "xsd": "http://www.w3.org/2001/XMLSchema#"
 },
 "@graph": [
  {
   "@id": "http://tourpack.redlink.io/seekda/2015/S000128",
   "schema:makesOffer": {
    "@id": "_:ub70bL5C28"
   }
  },
  {
   "@id": "_:ub70bL12C36",
   "@type": "schema:Service",
   "schema:name": {
    "@language": "en",
    "@value": "Massage"
   }
  },
  {
   "@id": "_:ub70bL26C53",
   "@type": "schema:BuyAction",
   "schema:target": [
    "http://www.fancyhotel.com/offer/buy/12345",
    {
     "@id": "_:ub70bL28C22"
    }
   ]
  },
```

STI · INNSBRUCK

**Schema.org example in JSON-LD for a touristic service package**

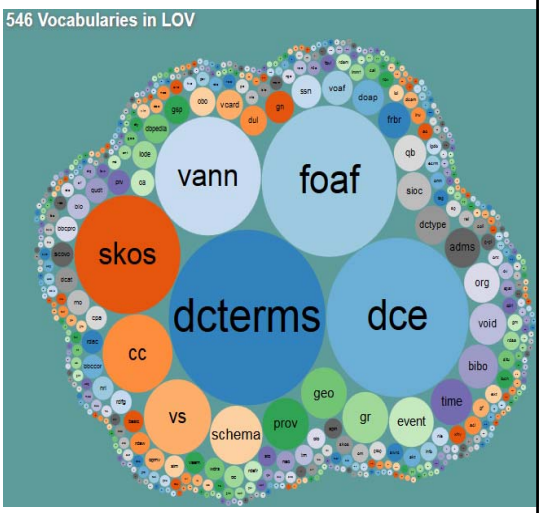
**see more details about the use case in this slideset: https://de.slideshare.net/annafensel/tour-pack-projectpresentationiwost**

115

## Linked Open Vocabularies (LOV)

STI · INNSBRUCK

- In order to describe a piece of information, we need a vocabulary that includes the relevant terms
- LOV (http://lov.okfn.org/dataset/lov) is a directory with all the existing vocabularies for various domains, where we can search for terms to use in our annotations.




State: April 2016

116

STI · INNSBRUCK

# SUMMARY



## Summary

STI · INNSBRUCK

- Semantic Web is not a ***replacement*** of the current Web, it's an ***evolution*** of it
- Semantic Web is about:
  - annotation of data on the Web
  - data linking on the Web
  - data Integration over the Web
- Semantic Web aims at ***automating*** tasks currently carried out by humans
- Semantic Web is ***real*** now (maybe not as we originally envisioned it, but it is)

# REFERENCES



## References


- Mandatory reading:
  - T. Berners-Lee, J. Hendler, O. Lassila. *The Semantic Web*, Scientific American, 2001.
- Further reading:
  - D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd Edition, Springer 2003.
  - G. Antoniou and F. van Harmelen. *A Semantic Web Primer*, (2nd edition), The MIT Press 2008.
  - H. Stuckenschmidt and F. van Harmelen. *Information Sharing on the Semantic Web*, Springer 2004.
  - T. Berners-Lee. *Weaving the Web*, HarperCollins, 2000.
  - T.R. Gruber, *Toward principles for the design of ontologies used or knowledge sharing? ,* Int. J. Hum.-Comput. Stud., vol. 43, no. 5-6,1995.
  - **Fensel, A., Kärle, E., Toma, I. "**TourPack: Packaging and Disseminating Touristic Services with Linked Data and Semantics**"**. In Proceedings of the 1st International Workshop on Semantic Technologies (IWOST), CEUR Workshop Proceedings, Vol-1339, ISSN 1613-0073, pp. 43-54, 11-12 March 2015, Changchun, China.

## References

STI · INNSBRUCK

- Wikipedia and other links:
  - http://en.wikipedia.org/wiki/Semantic_Web
  - http://en.wikipedia.org/wiki/Resource_Description_Framework
  - http://en.wikipedia.org/wiki/Linked_Data
  - http://www.w3.org/TR/rdf-primer/
  - http://www.w3.org/TR/rdf-mt/
  - https://www.w3.org/TR/owl-ref/
  - http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial
  - http://linkeddata.org/
  - http://www.opengeospatial.org/projects/groups/sensorweb
  - http://www.data.gov.uk/
  - http://schema.org
  - http://lov.okfn.org/dataset/lov



## Next Lecture

STI · INNSBRUCK

| # | Title |
|---|---|
| 1 | Introduction |
| 2 | Web Science + Cathy O'Neil's talk: "Weapons of Math Destruction" |
| 3 | Service Science |
| 4 | Web services |
| 5 | Web2.0 services + ONLIM APIs (separate slideset) |
| 6 | Semantic Web |
| 7 | **Semantic Web Service Stack (WSMO, WSML, WSMX)** |
| 8 | OWL-S and the others |
| 9 | Semantic Services as a Part of the Future Internet and Big Data Technology |
| 10 | Lightweight Annotations |
| 11 | Linked Services |
| 12 | Applications |
| 13 | Mobile Services |

## Questions?

STI · INNSBRUCK