

# 1 Introduction

## 1.1 About Course

### Course Organization

<https://cw.fel.cvut.cz/wiki/courses/b4m36osw>

## 1.2 Motivation

### 1.2.1 Scenario 1: What is a dataset about?

What is inside a dataset?

_id	building name	...
3	Atrium Flora	...
...	...	...

?



Are there **bus stops** or **large stores** in this dataset?

See OpenData portal of Prague OpenData portal of Prague

What is a building?

Building is a construction

- both above and below ground
- spatially compact
- with walls and roof
- with heating

Act 406/2000 Coll., on Energy Management

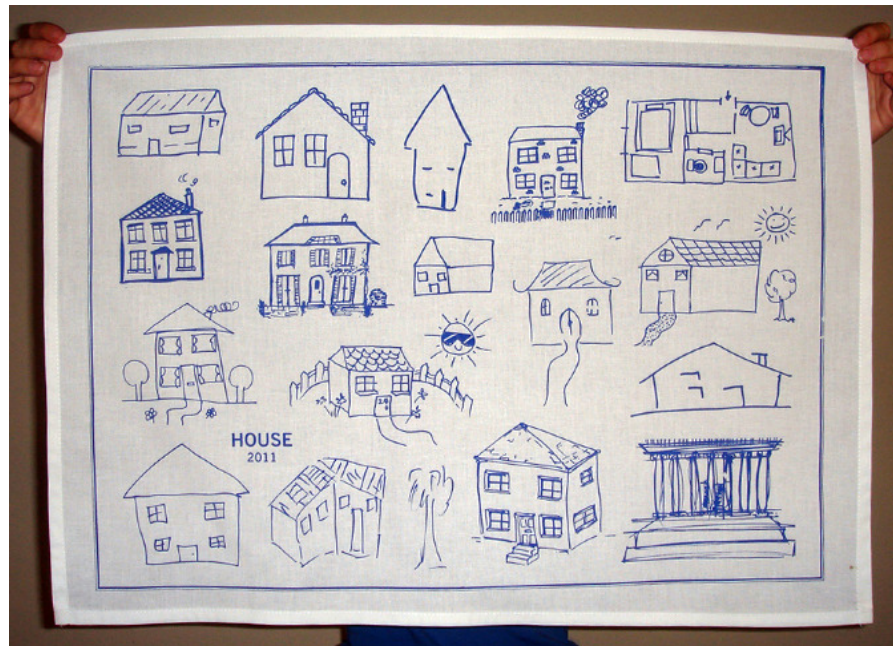
Building is a construction

- above ground
- with solid foundations
- spatially compact
- with walls and roof

Act 256/2013 Coll, Cadastral Law

## 1 Introduction

But things are worse ...



What is a building?



1. ... is a **construction** which is **heated**.

2. ... is a **construction** to provide **protection** to their users or internal equipment and is typically **closed** and has a **permanent position**.

ČSN EN 15643-5 -Sustainability of construction works

# Building

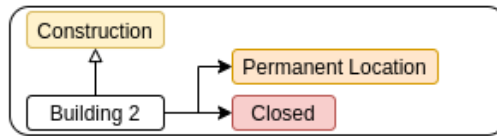
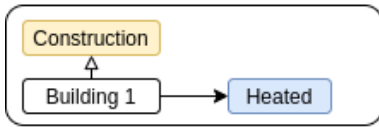
3. ... is a **construction above ground** which is **spatially-compact** and **closed by walls and roof**.

Act 256/2013 Coll., Cadastral Law

4. ... is a **construction above and below ground** which is **spatially-compact** and **closed by walls and roof** and is **heated or cooled**.

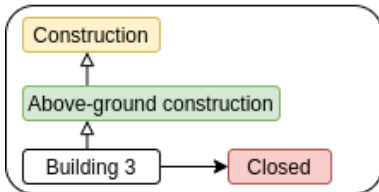
Act 406/2000 Coll., on Energy Management

What is a building?

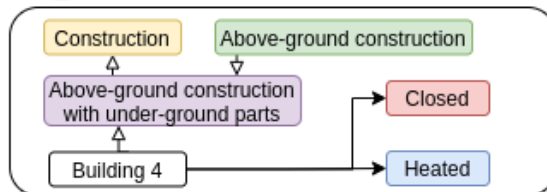


ČSN EN 15643-5 - Sustainability of construction works

# Building

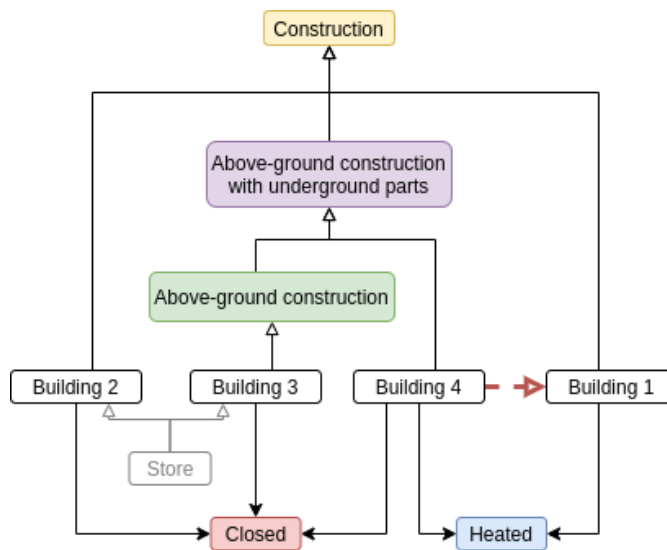


Act 256/2013 Coll., Cadastral Law



Act 406/2000 Coll., on Energy Management

New knowledge can be inferred



## 1.2.2 Scenarios: Examples of misunderstanding

One event or two events?

# DID YOU KNOW



Awaken the mind.

Just months before 9/11, the World Trade Center's lease was privatized and sold to Larry Silverstein.

Silverstein took out an insurance plan that 'fortuitously' covered terrorism.

After 9/11, Silverstein took the insurance company to court, claiming he should be paid double because there were 2 attacks.

Silverstein won, and was awarded \$4,550,000,000.

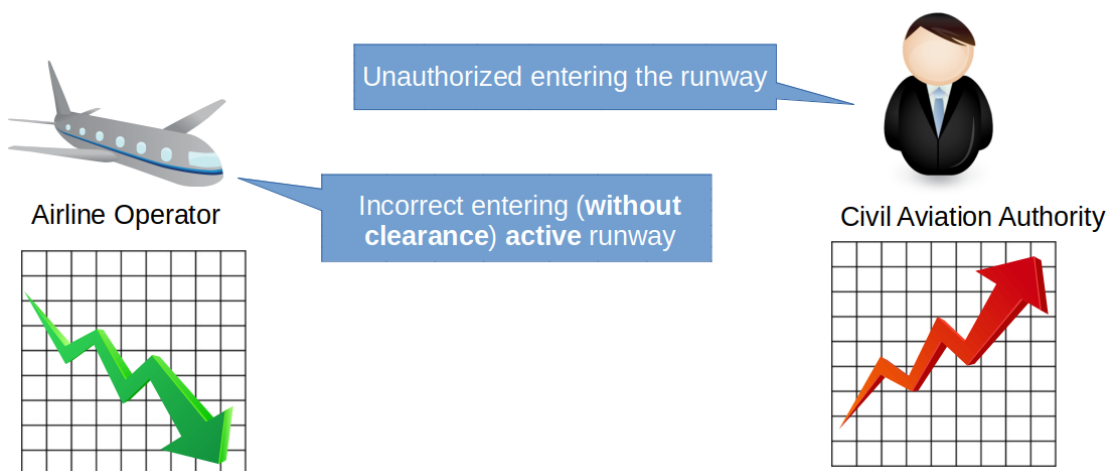
What is an event ? How many events occurred at 9/11 – One or Two ?

## Knowledge Management

9/11 ... matter of billions of USD

source:<https://www.metabunk.org/larry-silversteins-9-11-insurance.t2375>

What is the trend of Runway Incursion incidents at an airline operator ?



### 1.2.3 From Conceptual Models to Ontologies

#### Ontological Conceptual Modeling

- a way to **capture** and **explain** meaning.
- the language must be understandable to non-experts (UML max)
- the language must be computable – we want to use the models to infer new knowledge or validate data

#### About ontologies

Ontologies are **formal specifications of conceptualization**.

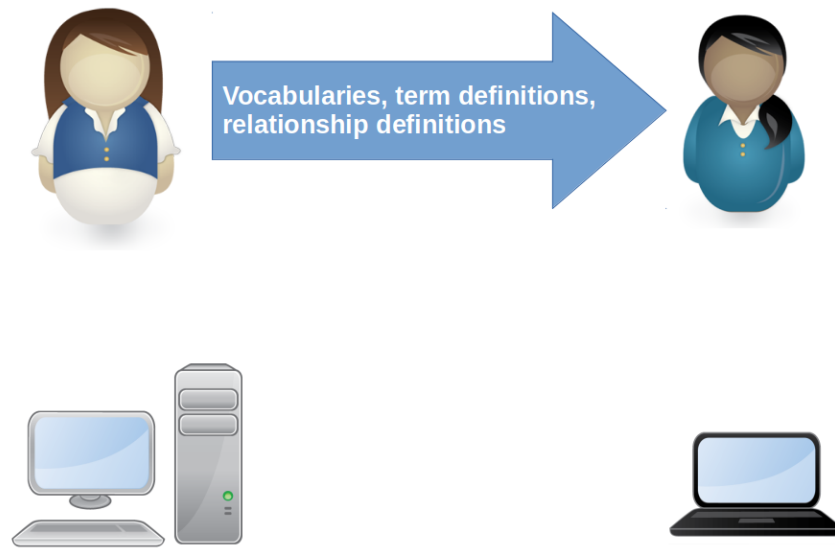
Ontologies help to stabilize the knowledge, to share meaning both among computers and among people. Use-cases include

- Data Integration
- Semantic Web
- Open (Linked) Data

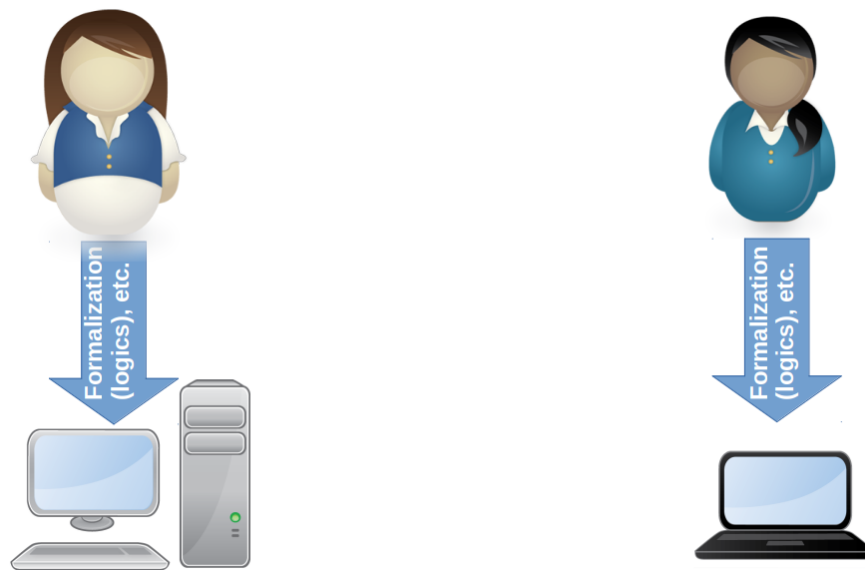
## 1.3 Overview of Ontologies

First, People Need to Understand Each Other

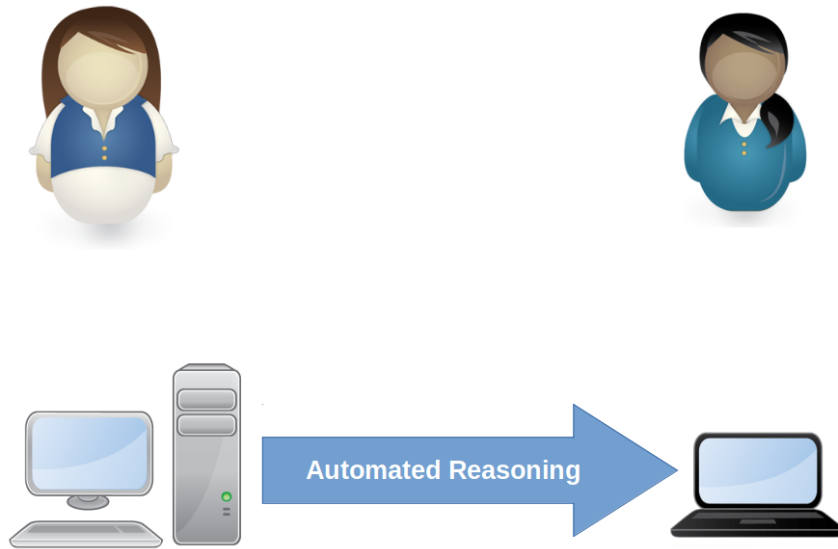
1 Introduction



Second, People Need to Explain Things to Computers

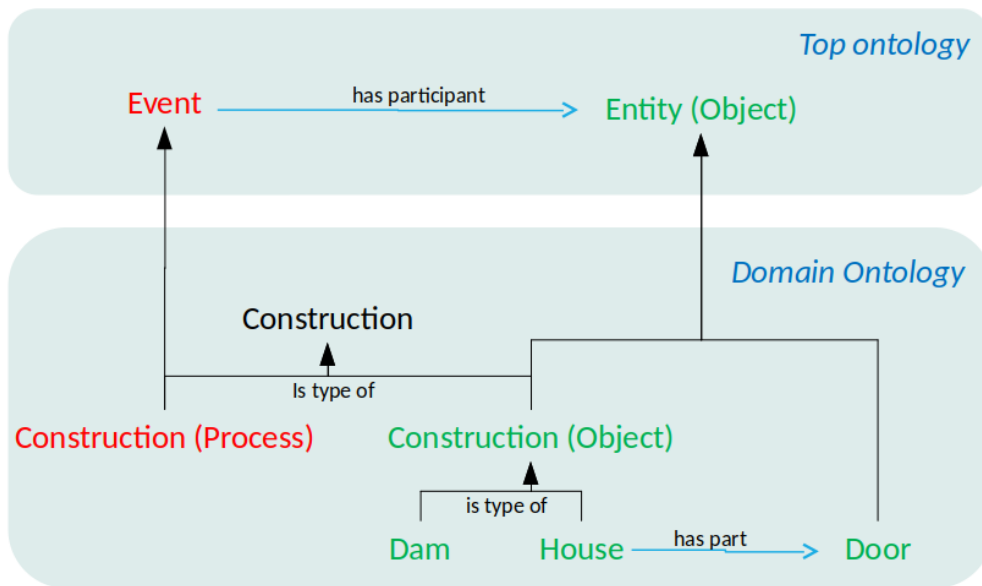


Third, Computers Can Understand One Another



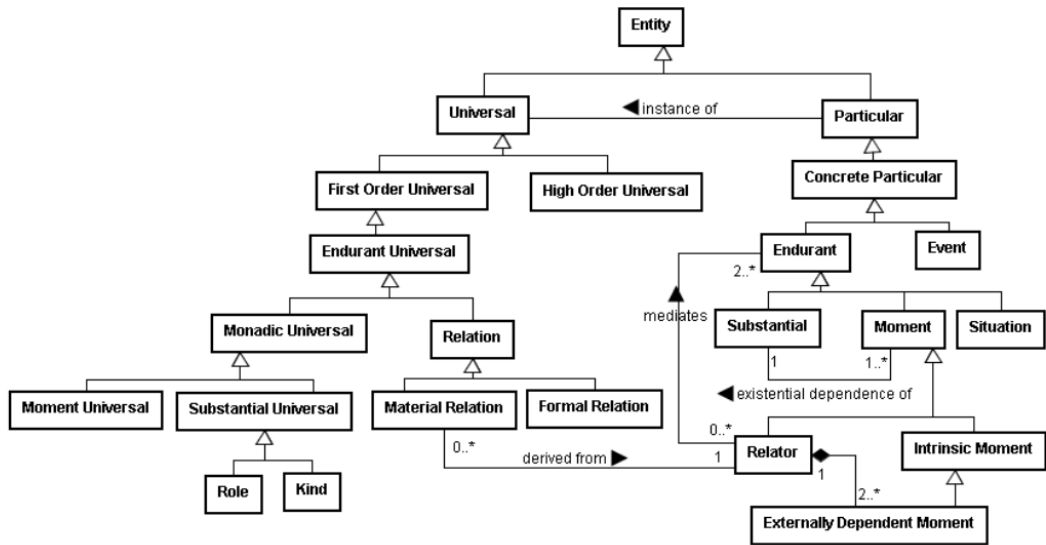
**Solution = Ontology**

*Explicit Conceptualization of Shared Meaning*



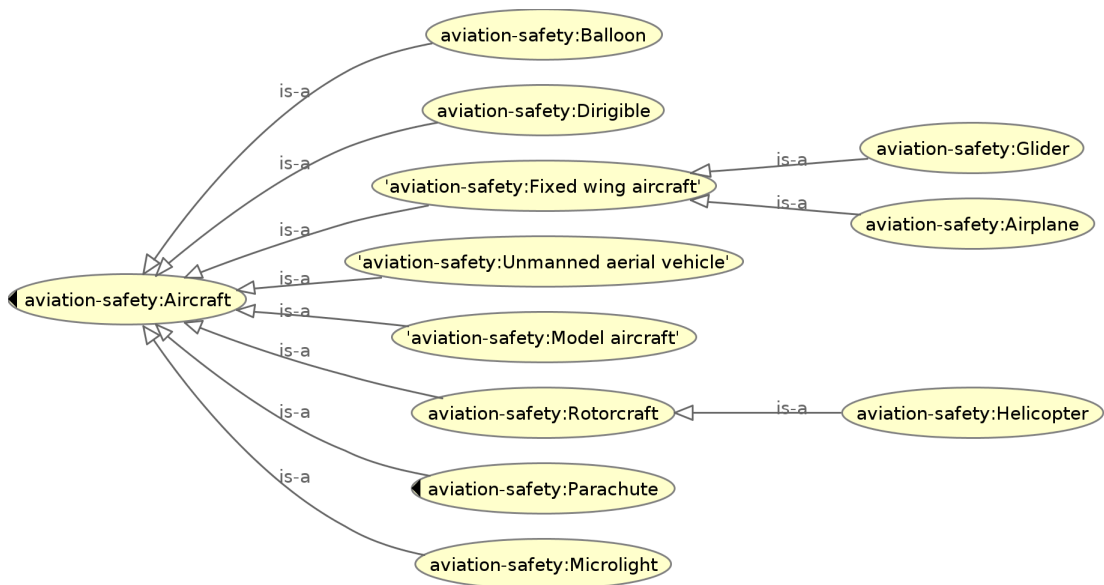
**Example Top-Level Ontology**

*Small part of Unified Foundational Ontology (UFO)*



**Example Ontology Hierarchy**

*Each helicopter is also an aircraft.*



**Ontologies ≠ Taxonomies**

Taxonomies = just a single type of relationship.

**Construction**

→ broad meaning (object, construction site, process)

**Dam**

**House**

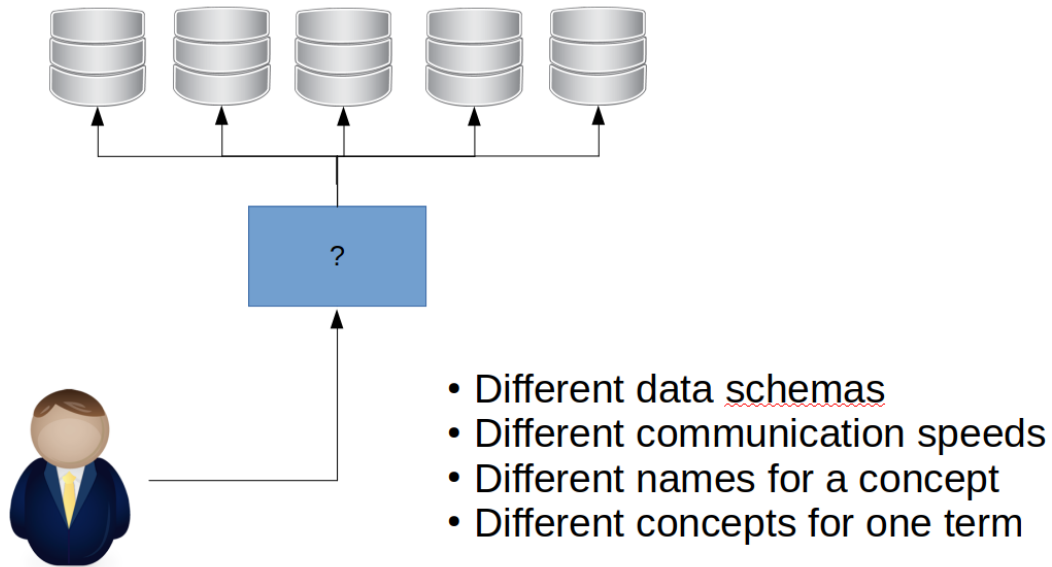
→ broad meaning (dwelling, construction)



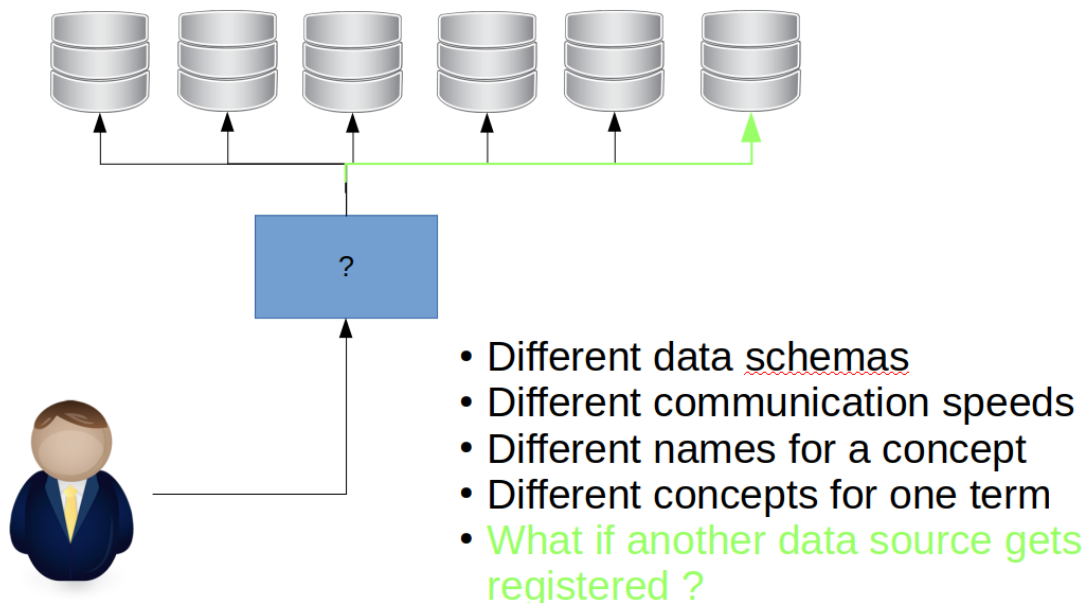
Door → specific meaning (not type of house, but its part)

## 1.4 Use-case: Data Integration

### Data Integration Scenario

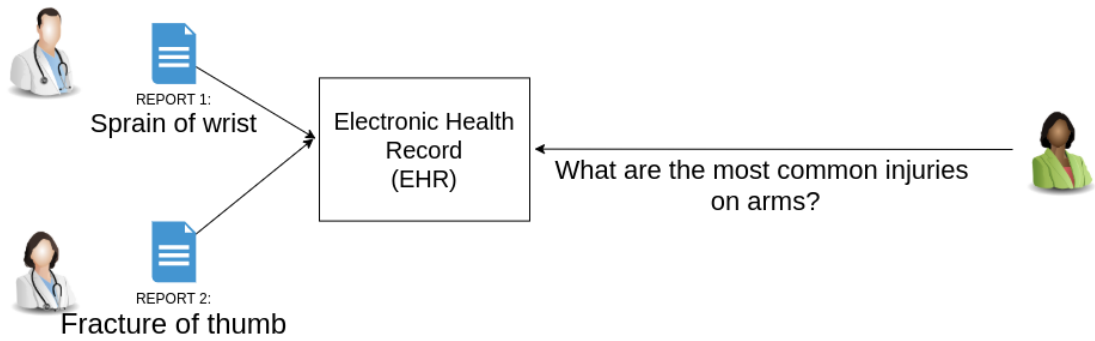


### Data Integration Scenario



## 1 Introduction

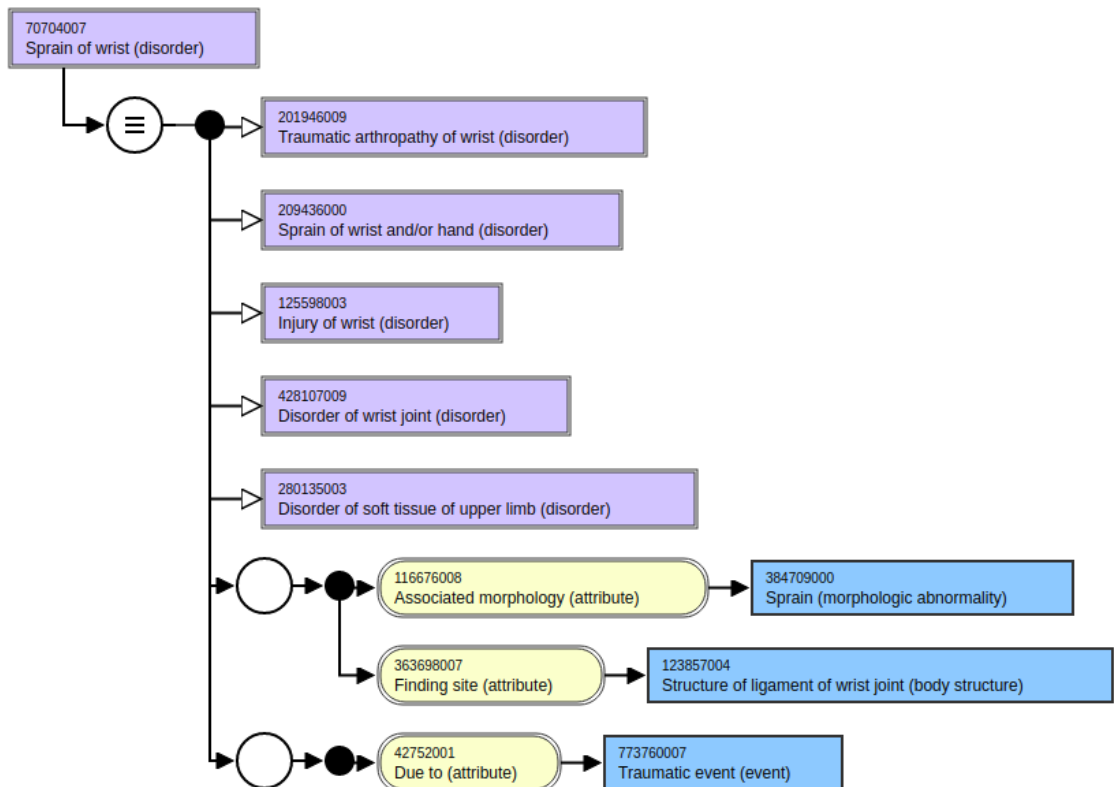
### Use-case – HealthCare Data Integration



### SNOMED-CT

*Systematized Nomenclature of Medicine - Clinical Terms*

- ~ 300k clinical concepts
- international standard – adopted e.g. in UK, USA, Australia
- uses ontology reasoning to classify/query the concepts



## SNOMED-CT

*Systematized Nomenclature of Medicine - Clinical Terms*

<https://browser.ihtsdo.org/?perspective=full&conceptId1=70704007&edition=MAIN/2020-07-31&release=&languages=en>

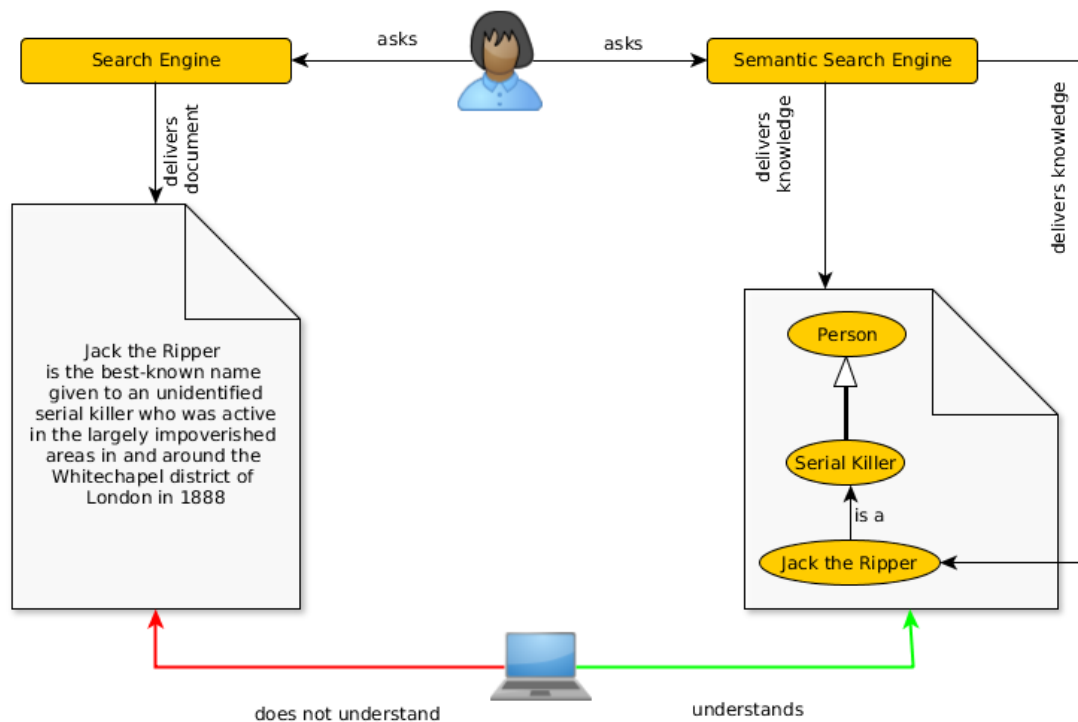
## 1.5 Semantic Web

### Current Web vs. Semantic Web

- SoA – semistructured HTML or XML data. There is vast amount of search engines like Google, Yahoo, MSN, etc. Many of them are invaluable, but as the engines use just keywords and/or some natural language preprocessing methods, the search results contain lots of irrelevant results that need to be processed manually.
- How to make web search more efficient ?
  - more expressive power for web designers to capture complexities – SW languages (RDF(S), OWL),
  - more efficient search engines to handle SW languages – new inference techniques for these languages,
  - better search engines interfaces – more expressive query languages
- **the amount of (unstructured) data is steadily growing**

### Semantic search

## 1 Introduction



### Ontologies and Semantic Web

**ontology** has many definitions, but let's consider it a **formal representation of a complex domain knowledge that is shared with others to ensure intelligent system interoperability,**

**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. (cit. Semantic Web. Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, 2001)

### Idea of Semantic Web

- W3C web page - <http://www.w3.org/2001/sw>
- The data format will be either RDF(S) or OWL,
- Reasoners for RDF(S) can be used for partial derivation in OWL,
- Reasoners for OWL can be used for derivation in RDF(S)

## Unique Data Identification – URIs

Semantic web speaks about resources.

URI is a unique identifier for addressing web resources in the form

```
<scheme name> : <hier. part> [ ? <query> ] [ # <fragment> ]
```

. HTTP scheme is used typically.

URN a URI with *scheme name* equal to 'urn'; used e.g. in SWRL atom identification,

URL a URI that can be resolved to a content using the protocol (e.g. HTTP),

IRI generalization of URIs allowing non-ascii characters. IRI is the standard identifier for OWL.

## Open World Assumption

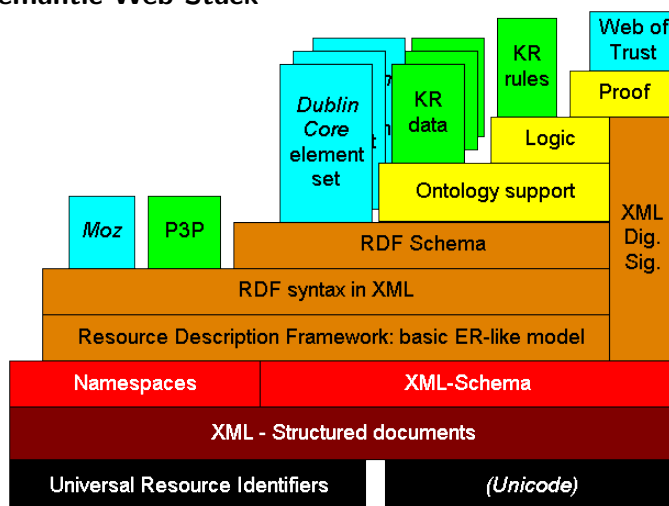
The semantic web inference must take into account that we handle *incomplete knowledge*.

## Description

Open world (OWA): Everything that cannot be proven is unknown, Closed world (CWA): Everything that cannot be proven is false.

Statement : “John is a Man.” Query: “Is Jack a Man ?” OWA Answer: “I don’t know.” CWA Answer: “No.”

## Semantic Web Stack



Taken from <http://www.w3.org/2000/Talks/0906-xmlweb-tbl/slide9-0.html>, by Tim Berners Lee.

## 1 Introduction

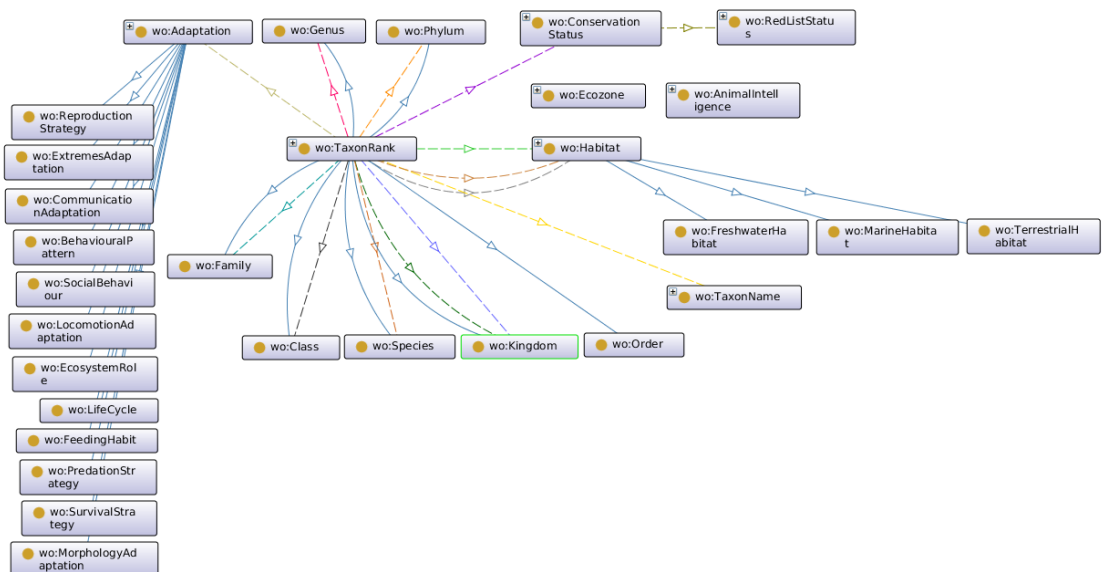
### 1.5.1 Semantic Web Adopters

#### Who is Using Semantic Web Technologies

Let's name a few:

- Google – *Knowledge Graph* (although they do not name it Semantic web – [http://semanticweb.com/google-just-hi-jacked-the-semantic-web-vocabulary\\_b29092](http://semanticweb.com/google-just-hi-jacked-the-semantic-web-vocabulary_b29092))
- Microsoft – Satori, <http://research.microsoft.com/en-us/projects/trinity/query.aspx>
- Facebook – Open Graph Protocol <http://ogp.me/>
- BBC – various datasets in RDF – <http://www.bbc.co.uk/developer/technology/apis.html>
- Ordnance Survey – geographic datasets in RDF – <http://data.ordnancesurvey.co.uk>

#### BBC Wildlife Ontology



#### Ordnance Survey Linked Data

## Kents Hill, Monkston and Brinklow



Kents Hill, Monkston and Brinklow is a Parish in Milton Keynes.		<b>Core facts about "Kents Hill, Monkston and Brinklow"</b>	
<b>Objects related to "Kents Hill, Monkston and Brinklow"</b>		<b>Type</b>	Parish
<b>Extent</b>	41649-49	<b>Label</b>	Kents Hill, Monkston and Brinklow
<b>In European Region</b>	South East	<b>Pref Label</b>	Kents Hill, Monkston and Brinklow
<b>Within</b>	Milton Keynes	<b>Alt Label</b>	Kents Hill, Monkston and Brinklow CP
<b>In District</b>	Milton Keynes	<b>Northing</b>	238013.803835
<b>Touches</b>	Walton Broughton Old Woughton Milton Keynes Wavendon	<b>Easting</b>	489602.596729
<b>Same As</b>	E04001285	<b>Lat</b>	52.0333028515
		<b>Long</b>	-0.695254366017
		<b>Area Code</b>	CPC
		<b>Gss Code</b>	E04001285

## 1.6 Linked Data

## How to publish data related to other ?

Based on semantic web principles, Linked Data provide means to efficiently connect data created by different publishers.

- Web of Documents – WWW
  - webpage – readable by human
  - identifiers – IRI
  - transfer protocol – HTTP
  - unified language – HTML
- Web of Data – Linked Data
  - webpage – readable by machine
  - identifiers – IRI
  - transfer protocol – HTTP
  - unified language – RDF

## 1 Introduction

*Linked Data* [Heath2011] is a method for publishing structured and interlinked data on the web, building up on URIs, HTTP and RDF technologies.

### Linked Data Principles

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

(Tim Berners-Lee, 2009 – <http://www.w3.org/DesignIssues/LinkedData.html>)

URIs satisfying the third point are **dereferencable**.

### Document vs. its Content

When designing a URI scheme it is necessary to ensure proper distinction between a **document** and its **content**

#### Example

```
@prefix people: <http://example.com/people/>
people:John people:likes people:Mary
```

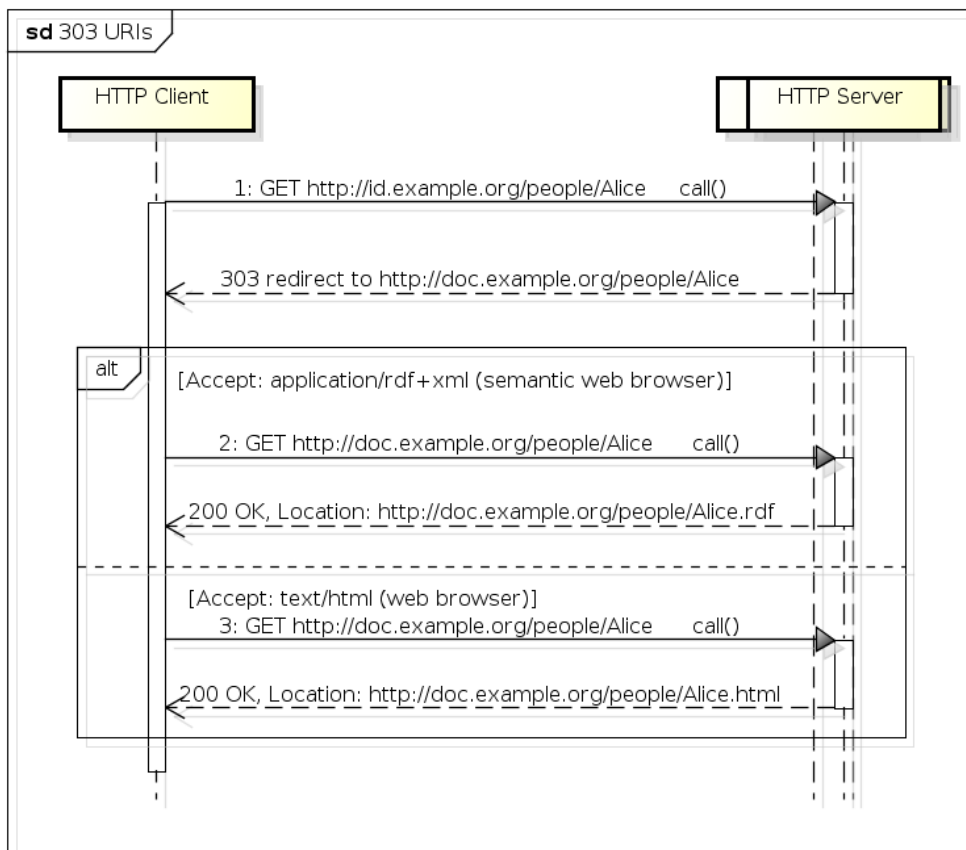
Is <http://example.com/people/Mary> a web document or a resource ? (Consider semantic consequences of each option).

This is handled by two strategies – 303 URIs and Hash URIs, each being suitable for different scenarios.

### 303 URIs

- 303 URIs are of the form <http://id.example.org/people/Alice>
- HTTP server sends 303 redirect to the corresponding **document** of the requested **resource**.
- HTTP client makes another request, based on Accept headers, the RDF/HTML version is delivered.

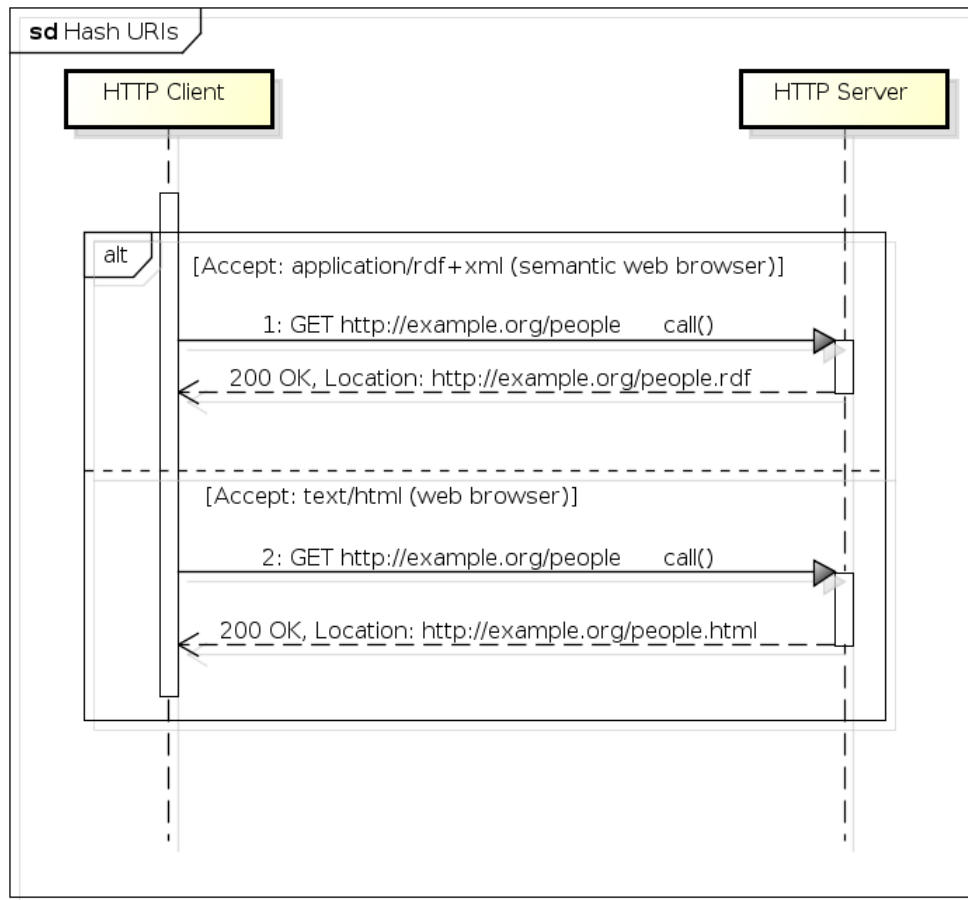




## Hash URIs

- Hash URIs are of the form `http://example.org/people#Alice`
- HTTP server sends the whole **document** of either RDF or HTML type based on Accept headers.
- Within the document, the HTTP client gets the particular entity after the hash symbol.

## 1 Introduction



powered by Astah

### 303 URIs vs. Hash URIs

**Hash URIs** are suitable for small datasets that will hardly grow up,

**303 URIs** are suitable for large datasets for the sake of good performance.

#### Reason

The fragment part of an URL (after #) is evaluated on the HTTP client (not the HTTP server), so the HTTP client must fetch all data first and then filter them for the subsequent use locally.

#### Linked Data Platforms

**Pubby** is a simple Linked Data publication server connectable to SPARQL endpoints,

**Callimachus** is an application server for linked data applications. To be explored in the tutorials,

**Marmotta** is a platform for publishing Linked Data (contributed from Linked Media Framework),

**D2R** is a platform for publishing relational database data in the form of Linked Data.

## 1.7 Use-case: Open Data

### CKAN and DataHub

CKAN (<http://ckan.org/>) is an open-source data portal for publishing, sharing and search of datasets.

It is prominently hosted at <http://datahub.io>. Datasets on DataHub can be sub-

The screenshot shows the DataHub website interface. At the top, there is a navigation bar with the DataHub logo and the tagline 'The easy way to get, use and share data'. The main content area is titled '/ Datasets'. On the left, there are two filter sections: 'Organizations' and 'Tags'. The 'Organizations' section lists various categories like 'Global (5)', 'Linking Open Data C... (2)', etc. The 'Tags' section lists tags like 'lod (6)', 'culturalheritage (6)', etc. The main search results area shows a search for 'cultural heritage' with 14 datasets found. The top result is 'Swedish Open Cultural Heritage' with a description and format links. Other results include 'Culture Grid', 'Flickr - The Commons', 'Amsterdam Museum as Linked Open Data in the Europeana Data Model', and 'British Museum Collection'.

mitted to the Linked Data Cloud.

### Datasets search

<https://datahub.io/search?q=coronavirus>

## Národní katalog otevřených dat (NKOD)

**OTEVŘENÁ DATA** Datové sady [Poskytovatelé](#) [Klíčová slova](#) [Další](#) ▼

**Poskytovatelé (1)**

- HLAVNÍ MĚSTO PRAHA (136)

**Klíčová slova (18)**

- Praha (136)**
- Česká republika (3)
- Digitální mapa Prahy (1)
- Lítačka (1)
- budovy (1)
- district (1)
- děti (1)
- [Zobrazit další](#)

**Formáty (10)**

- Esri Shape (98)
- Zipped GML (95)
- GeoJSON (80)

Vyhledat:

[Zobrazit pokročilé filtry](#) [Smaž filtry](#) [Název vzestupně ▼](#)

136 datových sad nalezeno

**Praha**

**Absolutní výšky budov**  
 HLAVNÍ MĚSTO PRAHA  
 Klasifikovaný rastr vytvořený z digitálního modelu zástavby zobrazuje absolutní nadmořské výšky budov.

[TIFF](#) [Plain text](#)

**Bonita klimatu**  
 HLAVNÍ MĚSTO PRAHA  
 Bonita klimatu - komplexní charakteristika dle všech hodnocených klimatologických hledisek Data byla vytvořena pomocí prostředku ArcGIS 9.2, Spatial Analyst. Vrstva byla převedena z rastrové vrstvy bonita, s horizontálním rozlišením 25m. Pro realizaci této mapy byla využita tato data: Digitální referenční mapa Praha-bloková mapa budo...

[GeoJSON](#) [Zipped GML](#) [Esri Shape](#) [ZIP](#)

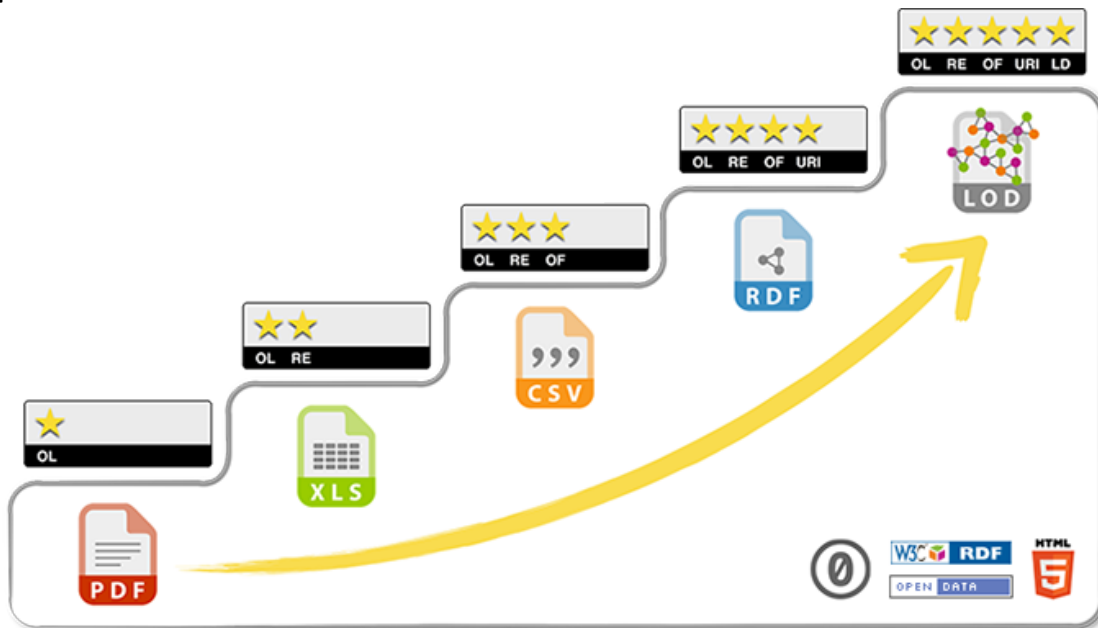
**Bonita klimatu z hlediska míry zastavěnosti území**  
 HLAVNÍ MĚSTO PRAHA  
 Data byla vytvořena pomocí prostředku ArcGIS 9.2, Spatial Analyst. Vrstva byla převedena z rastrové vrstvy bonita, s horizontálním rozlišením 25m. Pro realizaci této mapy byla využita tato data: Digitální referenční mapa Praha-bloková mapa budovy Liniová vrstva uličních úseku Vektorová data tématické vrstvy Úpn-doprava-liniová vrstva...

[GeoJSON](#) [Zipped GML](#) [Esri Shape](#) [ZIP](#)

gov.cz/

https://data.

## Open Data Levels



Taken from <http://5stardata.info/cs/>.

## Open Data Levels – description

- ★ 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)
- ★★★ All the above, 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

(Tim Berners-Lee, 2009 – <http://www.w3.org/DesignIssues/LinkedData.html>)

### From Open Data to Linked Data

★★★

★★★★

Aircraft (CAA)

s/n	type	operator_ic
1	Boeing 737	1234567
2	Airbus 319	9876543

→ ?

Companies (Business Registry)

company_ic	company_name
1234567	Best Airlines
9876543	Funny Flight School

### From Open Data to Linked Data

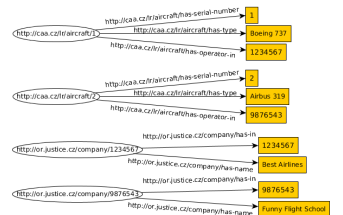
★★★

★★★★

Aircraft (CAA)

s/n	type	operator_ic
1	Boeing 737	1234567
2	Airbus 319	9876543

→

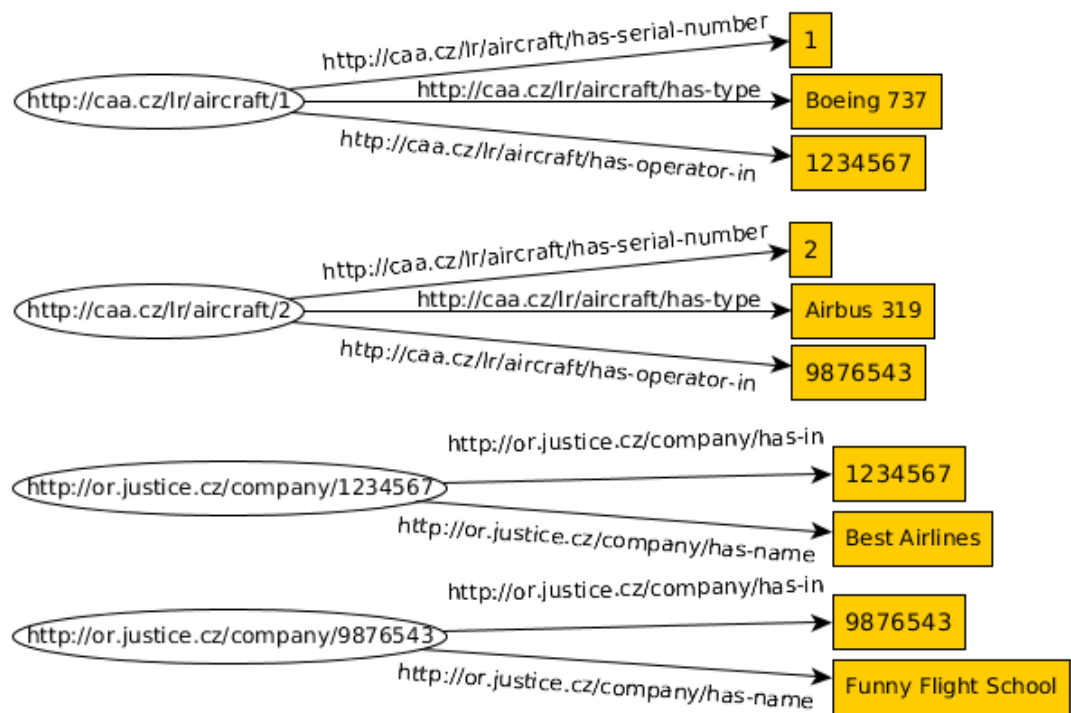


Companies (Business Registry)

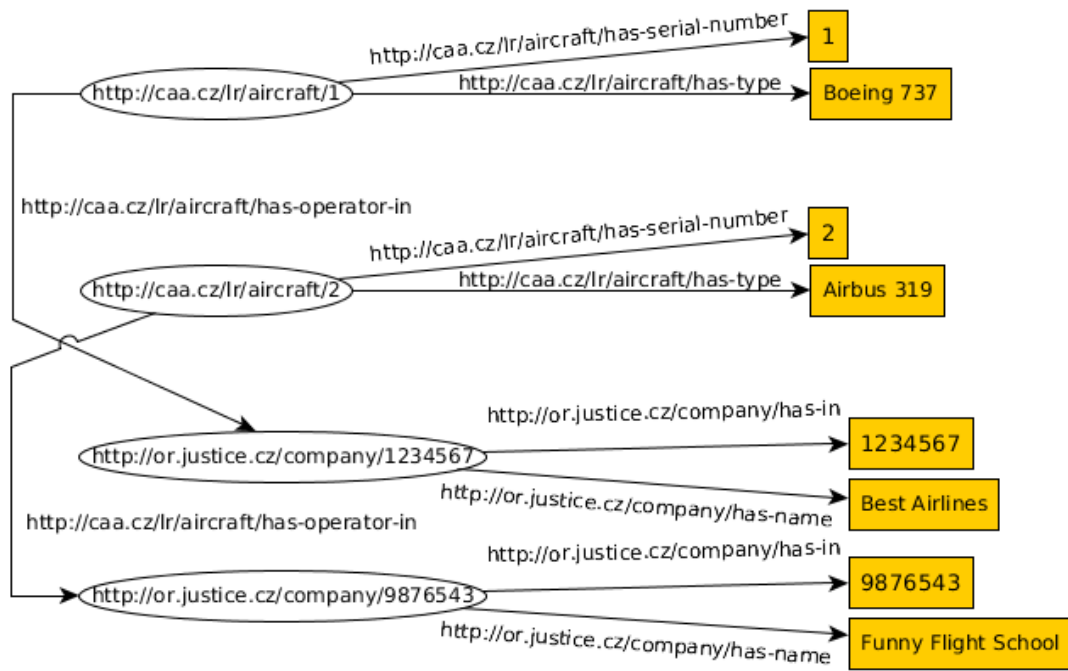
company_ic	company_name
1234567	Best Airlines
9876543	Funny Flight School

### From Open Data to Linked Data (4\*)

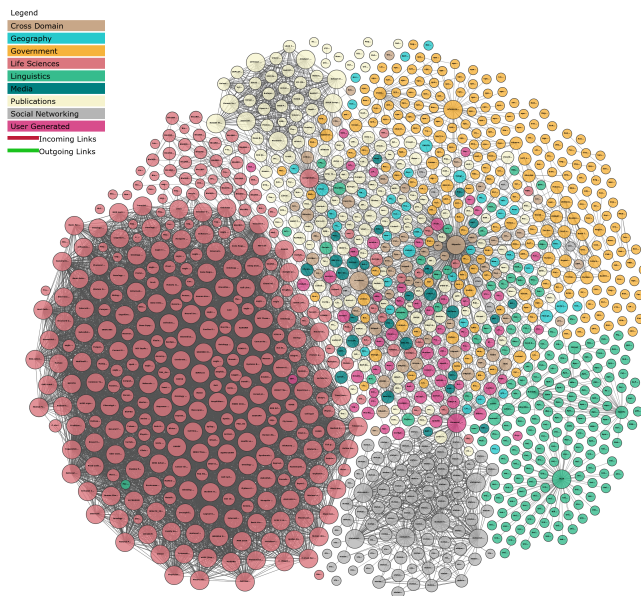
## 1 Introduction



## From Open Data to Linked Data (5\*)



## Linked Open Data Cloud



<http://lod-cloud.net/>, 2018

## 1 Introduction

### Linked Data vs. Open Data

**linked, not open** – enterprise data, master data

**linked, open** – 5\* data

**not linked, open** – typical case in OpenData

**not linked, not open** – we do not care

### Selected Materials

- OSW pages – <https://cw.fel.cvut.cz/wiki/courses/osw>
- RDF Primer – <https://www.w3.org/TR/rdf11-primer/>
- SPARQL Query Language Spec – <https://www.w3.org/TR/2013/REC-sparql11-query/>
- OWL Primer – <https://www.w3.org/TR/owl2-primer/>
- SKOS Primer – <https://www.w3.org/TR/skos-primer/>
- Description Logic Reasoning – P. Křemen, Ontologie a Deskripční logiky. In Umělá inteligence VI., Academia, 2013.
- Linked Data – <http://linkeddata.org>
- Nice supplementary tutorial on RDF/OWL – <https://www.obitko.com/tutorials/ontologies-semantic-web/>