Introduction to the Semantic Web

G. Falquet

Semantic Web Technologies

Main ideas (2001)

A web

readable/understandable by software agents

- pages on the web would be meaningful to programs

encompassing not just documents but every kind of data one could imagine

- interconnecting data (stored in different servers)

EXCLUSIVE: WARP DRIVE UNDERWATER • ARCTIC OIL VS. WILDLIFE

SCIENTIFIC AMERICAN
MAY 2001 $4.95
WWW.SCIAM.COM

Get the Idea?
(TOMORROW'S WEB WILL)

i know what you mean ...

PLUS:
Antibiotics' Dim Future
Rorschach: A Waste of Ink
The Oldest Stars

Copyright 2001 Scientific American, Inc.

A use case: organizing Mom's therapy

[…]At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved the information about Mom's prescribed treatment within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete's and Lucy's busy schedules.

Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." Scientific American, May 2001,

Required technologies

knowledge representation

    formally represent the information/knowledge content of a web site

data representation

    data representation framework for semi-structured data

interconnection

    global/shared object identification technique (for cross-server links)

    shared 'vocabularies' and concept description

reasoning/computing services

    logical inferences; computation (spatial, temporal, …); decision making; …

decentralized web services

Knowledge representation

a typical web page

Human understanding

event descriptions

event place

presenter

presenter's attribute

special announcement

Machine understanding

text text text text text text

    text text text   text

    text text text   text

    text text text

              text

text text

UNIVERSITÉ DE GENÈVE

UNIVERSITÉ   FACULTÉS   ÉTUDIANTS   COLLABORATEURS   SERVICES   ALUMNI

Vie de l'UNIGE   Recherche   Événements   Analyse   Trajectoires   Parutions   Distinctions   Vu dans les médias   Archives

Événements

Leçon d'ouverture d'automne: Exoplanètes

Les exoplanètes révolutionnent notre conception du monde et de la vie dans l'Univers

EXOPLANÈTES
CONFÉRENCE DU
PROF. DIDIER QUELOZ
PRIX NOBEL DE PHYSIQUE 2019
Lundi 21 septembre 2020
18h30 | Uni Dufour

AUDITOIRE U600 COMPLET

L'auditoire auquel se tient la conférence étant complet, nous vous proposons de vous

Événements à venir

Bourse aux vélos - Jeudi 17 septembre

Table ronde sur l'initiative de "limitation" - Vendredi 18 septembre

Exposition «Figures de la peur en Grèce antique» - Dès le 21 septembre

a logic-based approach

- define a logical language (vocabulary)

- represent the page content with logical formulae

- represent more general knowledge

the page in Description logic

```
Event(c1)
Conference(c1)
title(c1, "Exoplanètes …'')
speaker(c1, dq)
...
Person(dq)
name(dq, "Didier Queloz")
...
Event(bav)
title(bav, "Bourse aux vélos")
```

Requires some background knowledge.
What is a conference?

SW - Introduction

formalized in description logics

```
Conference
    subclass-of Event
    subclass-of ∃ title . String
    subclass-of ≥₁ speaker . Person

Person
    subclass-of HumanBeing
    subclass-of ∃ birthPlace . Place
    subclass-of ∀ employer . (Organization or Person)
```

Data representation

- multiple data models

  - relational databases, spreadsheets, graphs, …

- different levels of structure

  - text → unstructured
  - …
  - database → fully structured

not a new problem

The World Wide Web and databases : International Workshop WebDB'98 : Valencia, Spain, March 27-28, 1998 : selected papers

Author:	Paolo Atzeni; Alberto O Mendelzon; Giansalvatore Mecca
Publisher:	Berlin ; New York : Springer, 1999.
Series:	Lecture notes in computer science, 1590.
Edition/Format:	eBook : Document : Conference publication : English View all editions and formats
Database:	WorldCat
Summary:	This book presents the thoroughly refereed post-workshop proceedings of the International Workshop on the Web and Databases, WebDB'98, held in conjunction with EDBT'98 in Valencia, Spain, in March 1998. The 13 revised full papers presented were selected during two rounds of reviewing from initially 37 submissions. The book is divided into sections on Internet programming: tools and applications, integration and Read more...
Rating:	☆☆☆☆☆ (not yet rated) 0 with reviews - Be the first.
Subjects	Database management -- Congresses.
	World Wide Web -- Congresses.
	Database management.
	View all subjects
More like this	Similar Items

Lecture Notes in Computer Science 1590

Paolo Atzeni Alberto Mendelzon
Giansalvatore Mecca (Eds.)

The World Wide Web and Databases

International Workshop WebDB'98
Valencia, Spain, March 1998
Selected Papers

Springer

Preview this item

Semi-structured Data

"Roughly speaking, semi-structured data is data that is neither raw data nor very strictly typed as in conventional database systems" (Abiteboul 1997)

Exemples

- Web pages about restaurants
- BibTeX files
- ...

Serge Abiteboul, "Querying Semi-structured data," in *International Conference on Data Base Theory (ICDT)*, pp. 1 – 18, Delphi, Greece, 1997. http://dbpubs.stanford.edu:8090/pub/1996-19.

@article{**miller1995wordnet**,
    Author = {Miller, George A},
    Journal = {Communications of the ACM},
    Number = {11},
    Pages = {39--41},
    Publisher = {ACM},
    Title = {WordNet: a lexical database for English},
    Volume = {38},
    Year = {1995}}

@techreport{**masolo2003wonderweb**,
    Author = {Masolo, Claudio and Borgo, Stefano and Gangemi, Aldo and Guarino, Nicola and Oltramari, Alessandro},
    Institution = {LOA-ISTC-CNR},
    Title = {The WonderWeb library of foundational ontologies and the DOLCE ontology. WonderWeb (EU IST project 2001-33052) deliverable D18},
    Year = {2003}}

@inproceedings{**niles2001towards**,
    Author = {Niles, Ian and Pease, Adam},
    Booktitle = {Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001},
    Organization = {ACM},
    Pages = {2--9},
    Title = {Towards a standard upper ontology},
    Year = {2001}}

Main aspects

Irregular structure

    heterogeneous, incomplete elements

Implicit structure

    structure in textual parts => parsing

Partial structure

    unstructured parts: plain text, images, external data

Indicative structure vs. constraining structure

    schema adds information

Main aspects

A-posteriori schema/data guide

> created to structure existing data (from the data)

Large schema

> e.g. wikidata

Schema ignored

> in discovery/navigation queries the schema must be ignored

Rapidly evolving schema

> e.g. in scientific databases (new techniques/knowledge)

Guide

SW - Introduction

in JSON

```
[    {"type": "public-conference", "title": "Exoplanètes …", …},
     {"type": "exhibition", "title": " …", ",
         time-period: {"from": "…", "to": …"d}, …
     },
     …
]
```


SW - Introduction

"Standard" Solutions

old style: XML

- + XML Schema, XSL transformations, XML APIs

new style: JSON

- + JSON APIs, JSON Schemas

Interconnection

Problem: different databases use different identifiers for the same entity

Part	Origin	...
Motor	DE	
Windows	FR	
Wheels	USA	

Company	Headquarters	...
IBM	Unites States	
Telefónica	Spain	
Orange	France	

Database 2

Database 1

The Semantic Web isn't just about putting data on the Web. It is about **making links**, so that a person or machine can explore the Web of Data.

With Linked Data, when you have some of it, you can find other, related, data.

Tim Berners-Lee

The Linked-data Solution

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

A resource is the main information building block

Anything that can be named is a resource.

- **Information resources** entities that convey information and can be completely represented in binary code:
    - documents, images, video, software …

- **Non-information resources** cannot be represented as bits:
    - people, phenomena, concepts, ideas …

Web resources are conceptual relations uniquely identified by HTTP URLs

- An HTTP URL points to at most one resource.
- If it is an information resource, HTTP allows clients to retrieve a representation of it.

  - The **concept** pointed to by an URL shouldn't change.

  - The **value** and **representations** retrieved when looking up an URL might change over time.

Using HTTP URIs ensures that anybody can look up the resource

An HTTP URI of a resource can be dereferenced:
use an HTTP client to retrieve a representation.

- Information resources result in a representation.
- Non-information resources result in a 303 redirect.

- Relies on the double role of an HTTP URI as identifier and locator.

- Principle: If you don't know something, look it up. Follow your nose.

Dereferencing a URI should lead to useful information about that resource

"Useful" means the information is available using standard technologies.
(RDF and SPARQL)

"Useful" also means the information provides explanations and/or context for the resource

Define the resource in terms of concepts the client already knows or can look up.

By including links to other resources, we create a Web of Data

- Links connect a resource to known concepts.
  - <u>Alberto</u> is a <u>researcher</u> at <u>U. of Toronto</u>

- Links give meaning to data.
  - These <u>temperatures</u> are <u>measured in</u> <u>degrees Celsius</u>.

- Links allow exploration of related data.
  - Find more by the same author.

Basic information unit: the link
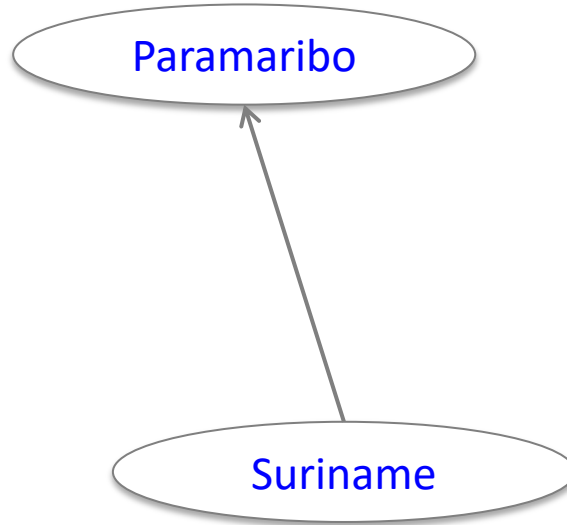
A link connects two resources

The resources are identified by URIs

http://dbpedia.org/resource/Paramaribo

http://dbpedia.org/resource/Suriname
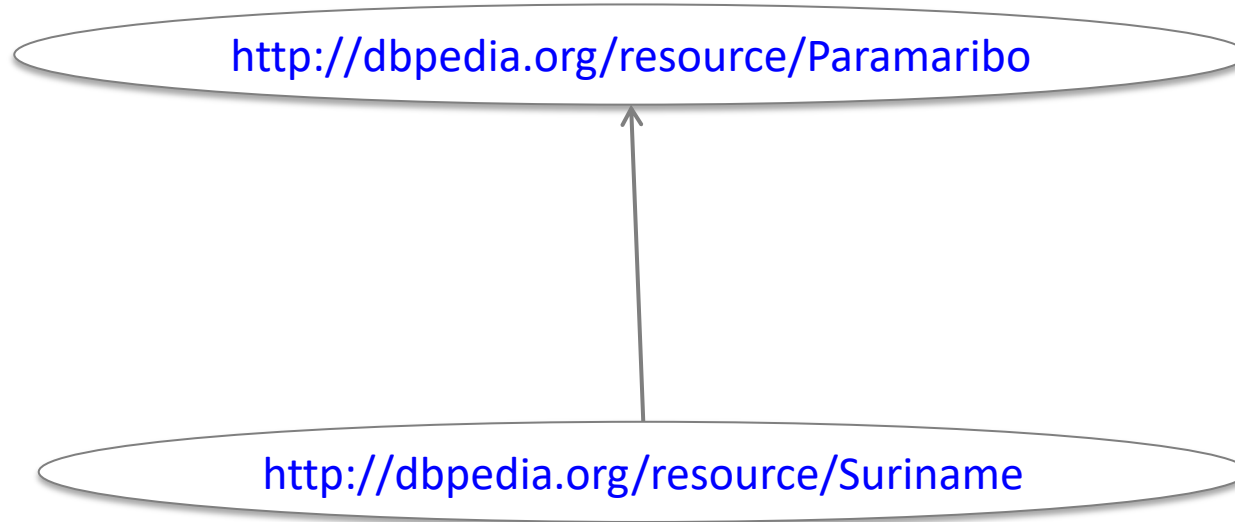
Using prefixes to abbreviate the URIs

dbr = http://dbpedia.org/resource/

dbr:Paramaribo
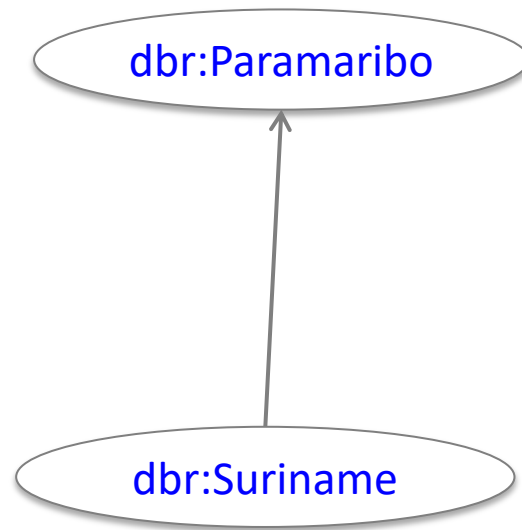
dbr:Suriname

The links are typed (unlike Web links)

dbr:Paramaribo

dbo:capital
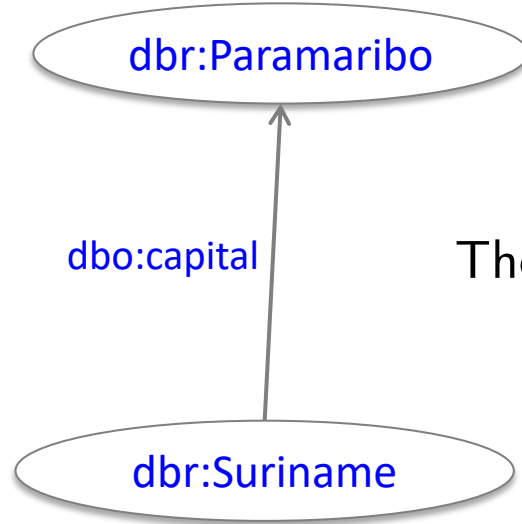
dbr:Suriname

The link type is also identified by a URI

… so the link type can be described

dbr = http://dbpedia.org/resource/
dbo = http://dbpedia.org/ontology/
rdf = http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs = http://www.w3.org/2000/01/rdf-schema#
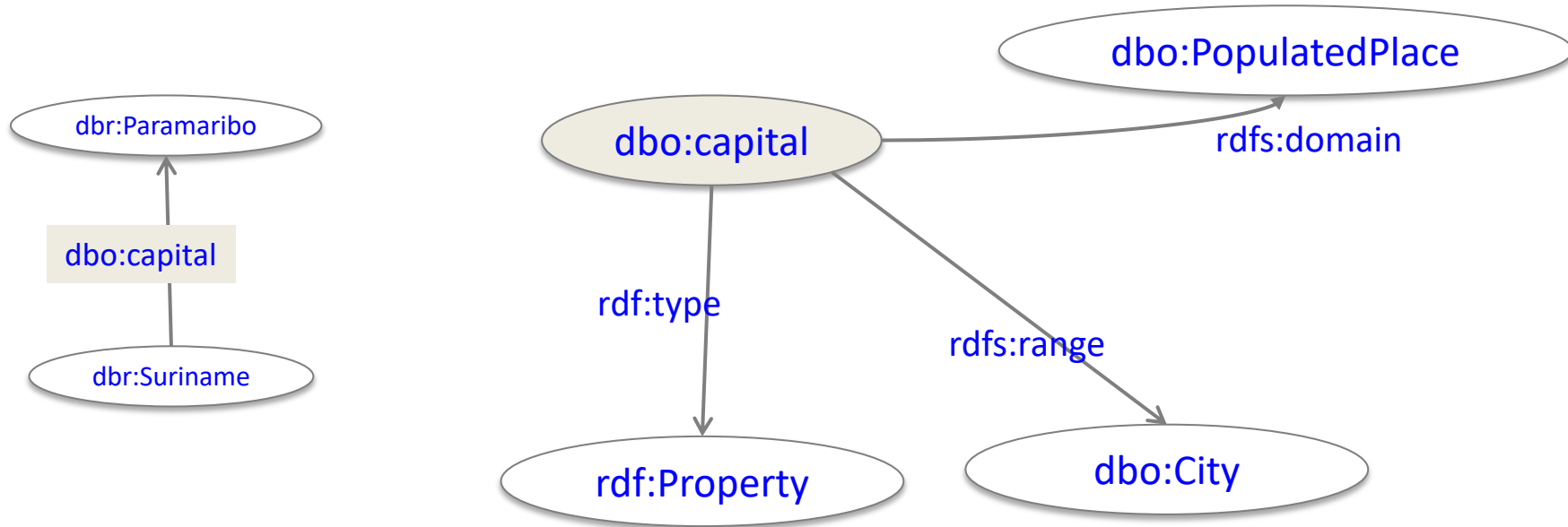
Links can point to typed literal values

Creating a web of data

SW - Introduction
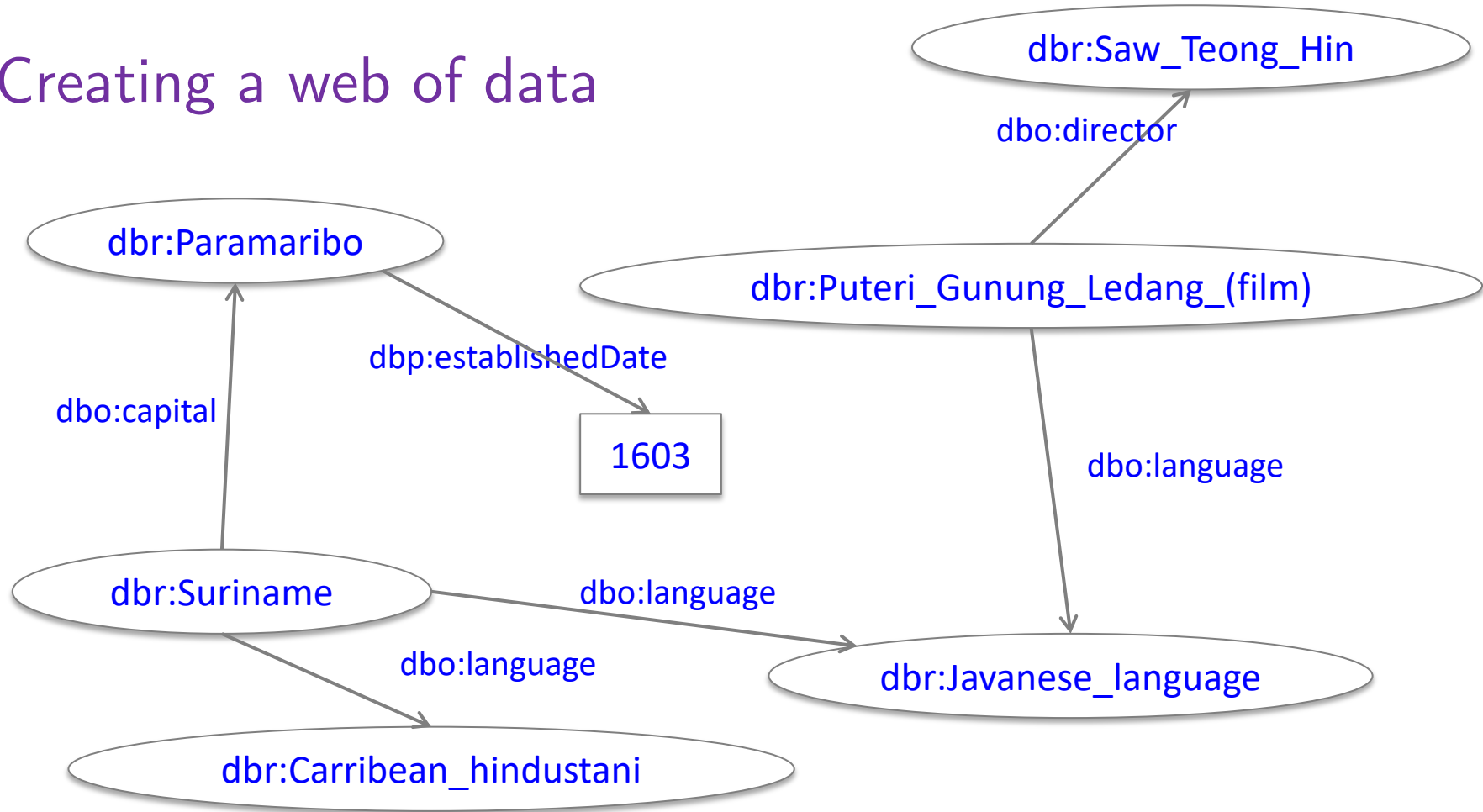
In a machine readable form

```
@prefix dbr: <http://dbpedia.org/resource/>
@prefix dbo: <http://dbpedia.org/ontology/>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>


dbr:Suriname dbo:language dbr:Carribean_hindustani.
dbr:Suriname dbo:language  dbr:Javanese_language.
dbr:Suriname dbo:capital dbr:Parmaribo.

dbr:Parmaribo dbp:establishedDate 1603.

dbr:Puteri_Gunung_Ledang_(film) dbo:dirctor dbr:Saw_Teons_Hin.
dbr:Puteri_Gunung_Ledang_(film) dbo:language  dbr:Javanese_language.
```
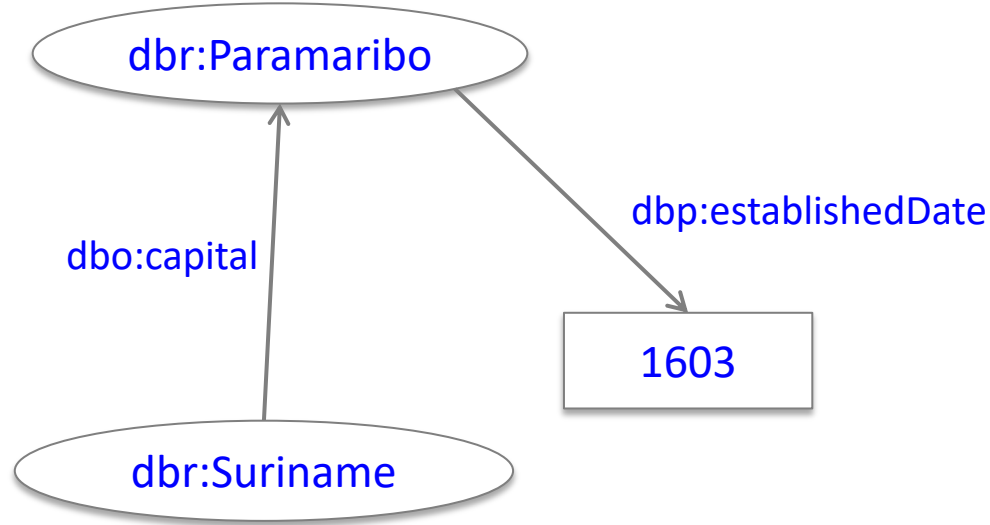
Compared to relational databases

No constraining database schema

- DB: putting data in predefined boxes (tables, rows, columns)
- SW: linking data

Open world

- DB: what is not in the database is *true*, what is absent is *false*, (closed world)
- SW: what is described is *true*, what is absent is *unknown*
  - but we may have negative descriptions

Global vocabulary (identifiers)

- the resource and property names (URIs) are globally visible

RDF

- https://rubenverborgh.github.io/WebFundamentals/semantic-web/#rdf-model

Interconnection

Problem: the same term may have different meaning in different databases

Ontology solution: Create shared concept descriptions

- schema.org, Linked open vocabularies, …
- use common concept description languages (RDFS, OWL, …)

Reasoning

- Make logical inferences
    - find the logical consequences of facts and rules
    - test the consistency of a set of logical formulae

State of the SW

- the Semantic Web does not exist
  - not as imagined by TBL et al.
  - big companies have created their SW (e.g. Apple Siri, Google services, …)

- Many SW technologies are operational
  - Resource description framework, Ontology languages
  - Querying and reasoning software
  - Semi-structured databases (RDF triple stores, graph databases, …)
  - Knowledge graphs

Content of the course

- resource description with RDF graphs
- linked data
- ontologies and logical reasoning for description logic and logic programming
- representing time and space
- interoperability
- knowledge graphs