



WEB SEMÂNTICA

UMA INTRODUÇÃO (BEM GENTIL)

PROF. DR. SERGIO SERRA
SERGIOSERRA@GMAIL.COM

we are

**OPEN
DATA**

Você pode copiar essas transparências à vontade, parte das transparências são do autor, outras vieram de diversas fontes da Web e de diversos colegas

Basedo em alguns slides de Ely Edison Matos e de outros autores, #Obrigado



Missão da aula de hoje...

1. O que é?
2. Qual o problema que trata?
3. Quais as padrões/soluções mais atuais?

**Toda linguagem tem sua própria sintaxe
e semântica e pragmática**

Sintaxe - estuda a gramática → Como se escreve

Semântica - estuda o significado → O que significa

Pragmática - estuda o uso → Como se usa

Sintaxe é
como se
escreve algo

Semântica é
o que algo
significa

syntaxes diferentes
mesma semântica

$x += y$

$x = x + y$

Internet é uma grande teia...



Onde a sintaxe, semântica e a pragmática
estão relacionadas à
troca de dados



Então, as pessoas e as empresas...





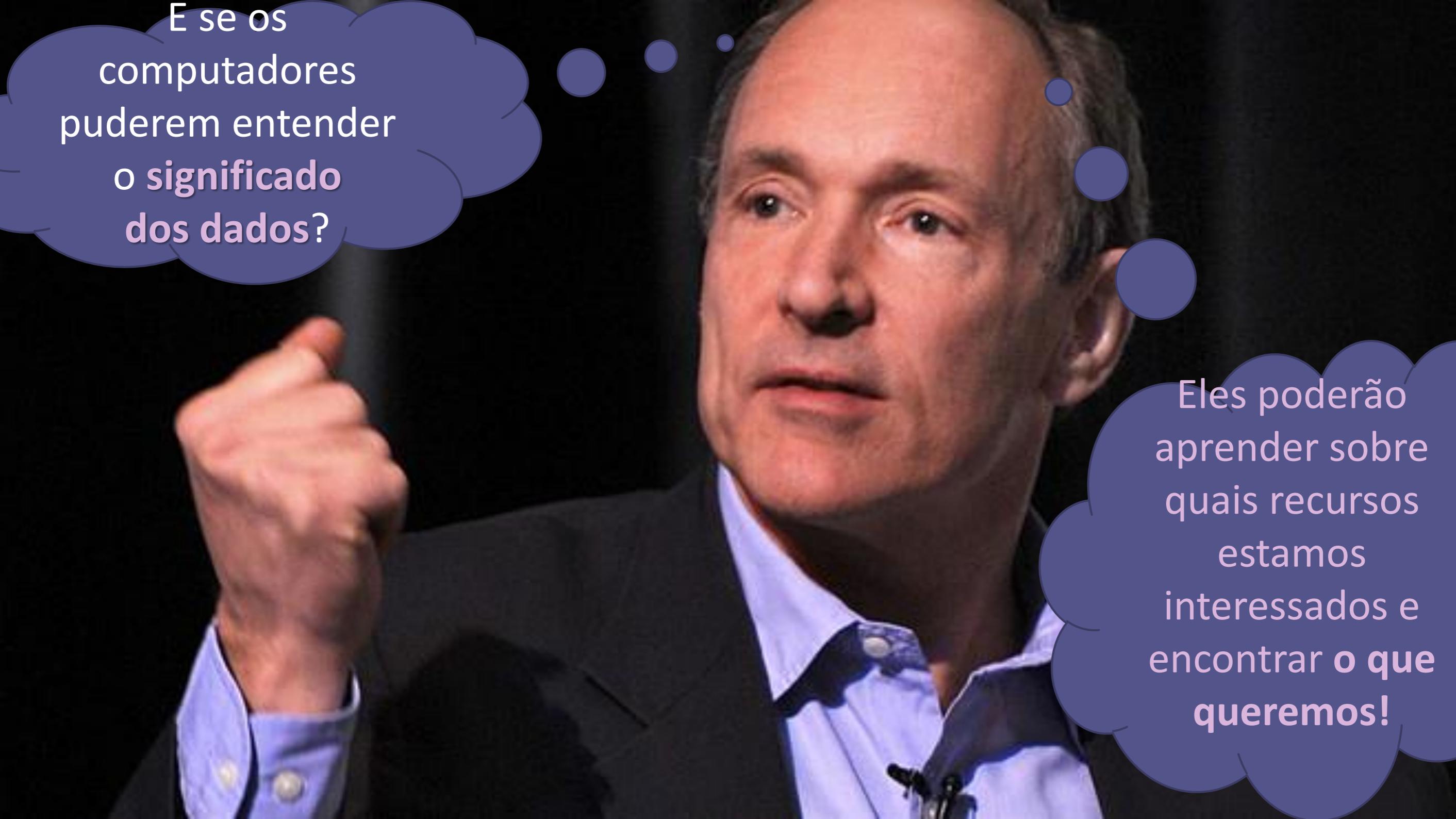
Começaram a se afogar no vasto oceano dos
dados semi-estruturados,
mal anotados,
mal compartilhados, pouco acessíveis e
fracamente reutilizáveis, pior ... tudo isso que
não para de crescer!



A close-up, high-contrast photograph of a man's face. He has a shocked or screaming expression, with wide eyes and an open mouth. His hands are pressed against his cheeks, covering his mouth. The lighting is dramatic, with strong highlights and shadows.

Qual o problema professor?

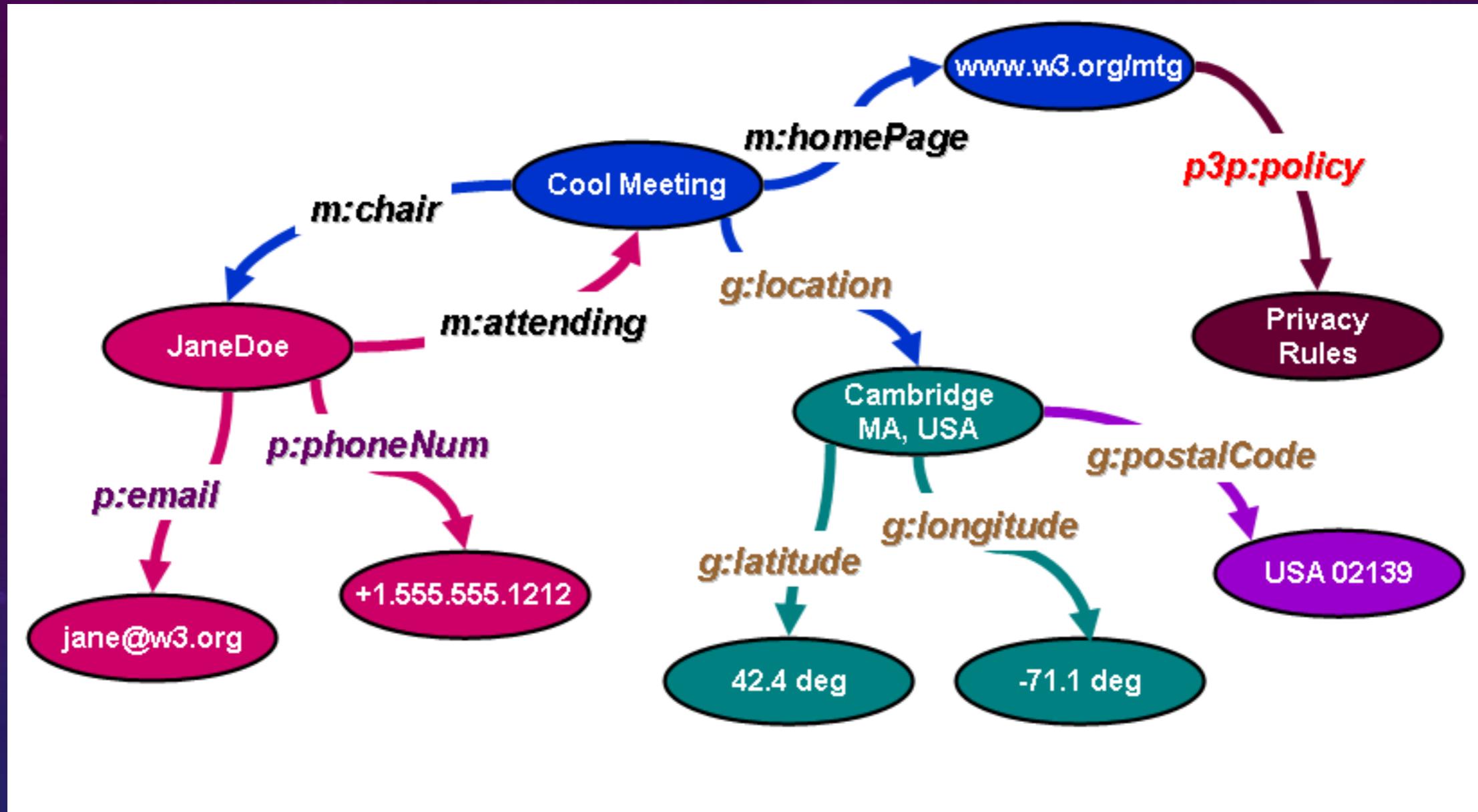
Lembra-se que o HTML
descreve estrutura e
apresentação da
informação, ou seja, sintaxe
mas não a semântica?

A close-up photograph of a middle-aged man with short brown hair, wearing a dark suit jacket over a light blue button-down shirt. He is looking slightly upwards and to his right with a thoughtful expression. His right hand is raised, with his index finger pointing towards the upper left corner of the frame. The background is dark and out of focus.

E se os
computadores
puderem entender
o **significado**
dos dados?

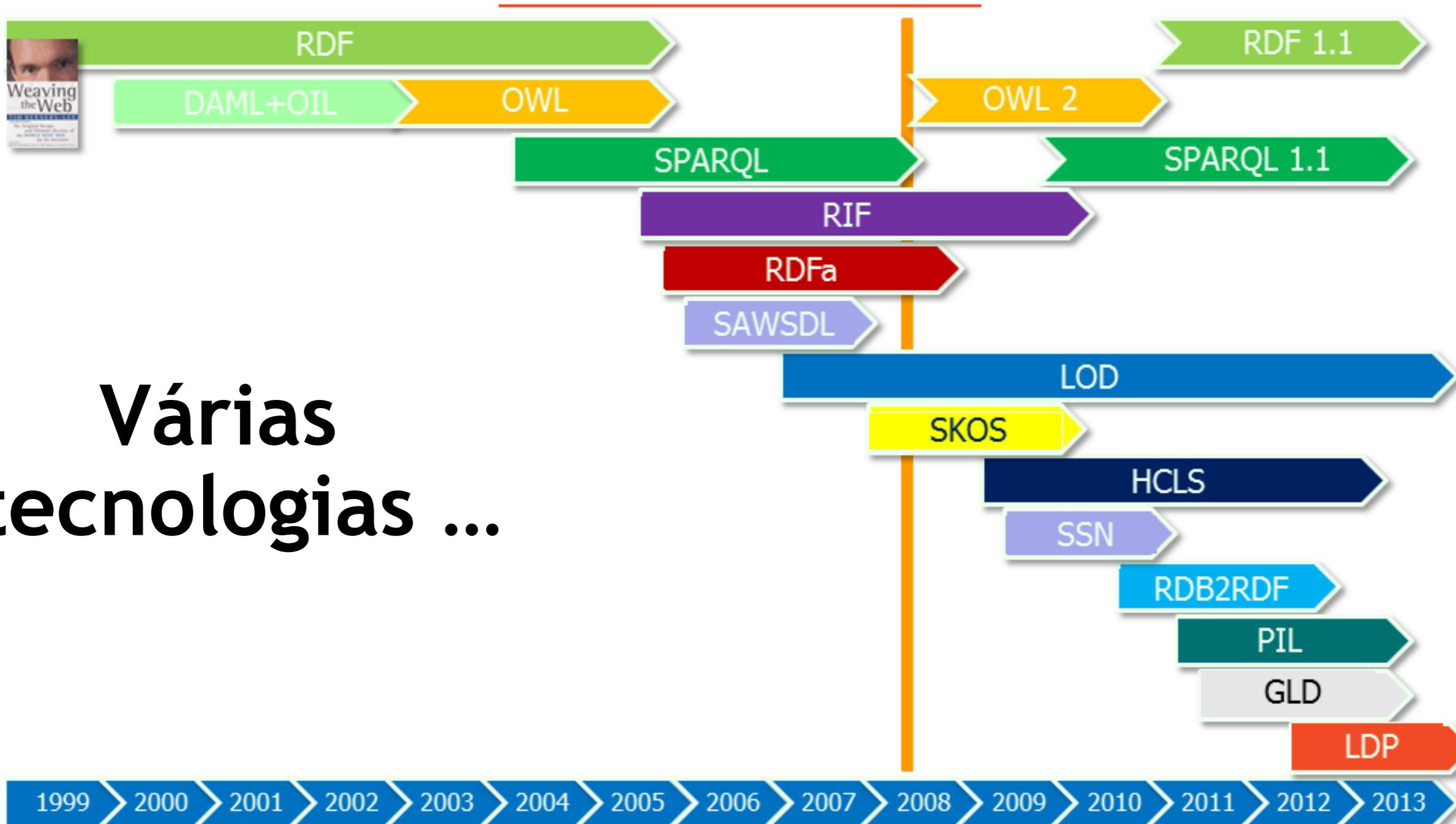
Eles poderão
aprender sobre
quais recursos
estamos
interessados e
encontrar o que
queremos!

+ em 2001 nasceu a
Web
Semântica



A Web Semântica é uma teia de recursos
(dados + relações + metadados)

The Semantic Web timeline



Web Semântica

Adicionar
estrutura e semântica
(RDFS,OWL,RIF)

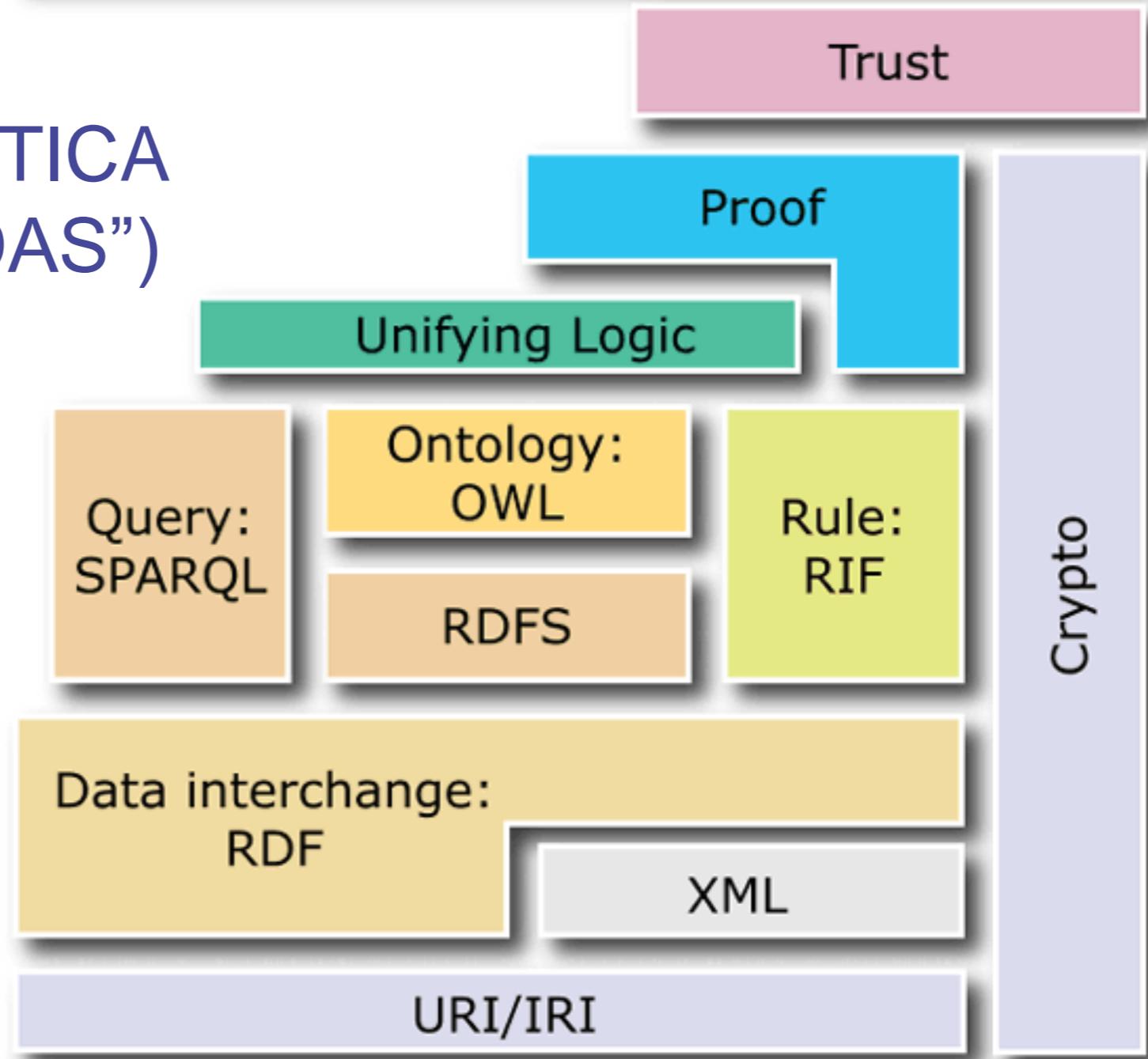
Definir um
Modelo de Dados
(RDF)

Definir uma
Linguagem de Consulta
(SPARQL)

Trabalhar na
Web “real”
(Linked Data, RDFa)

Trust

A VISÃO DA WEB SEMÂNTICA (OU “O BOLO EM CAMADAS”)



A Web Semântica pode ajudar a...

Mapear novos datasets

Integrar e combinar datasets

Inferir novas relações e saberes

Consultar datasets

muito mais...

Vamos estudar um “exemplo visual” bem simples...

1. Mapear os vários dados em uma **representação abstrata de dados**
 - . Tornar os dados independentes de sua representação interna...
2. **Combinar** as representações resultantes
3. Fazer **consultas** no conjunto total.
4. **Novas consultas** que não são possíveis nos datasets individuais

DATASET “A”

Books

ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

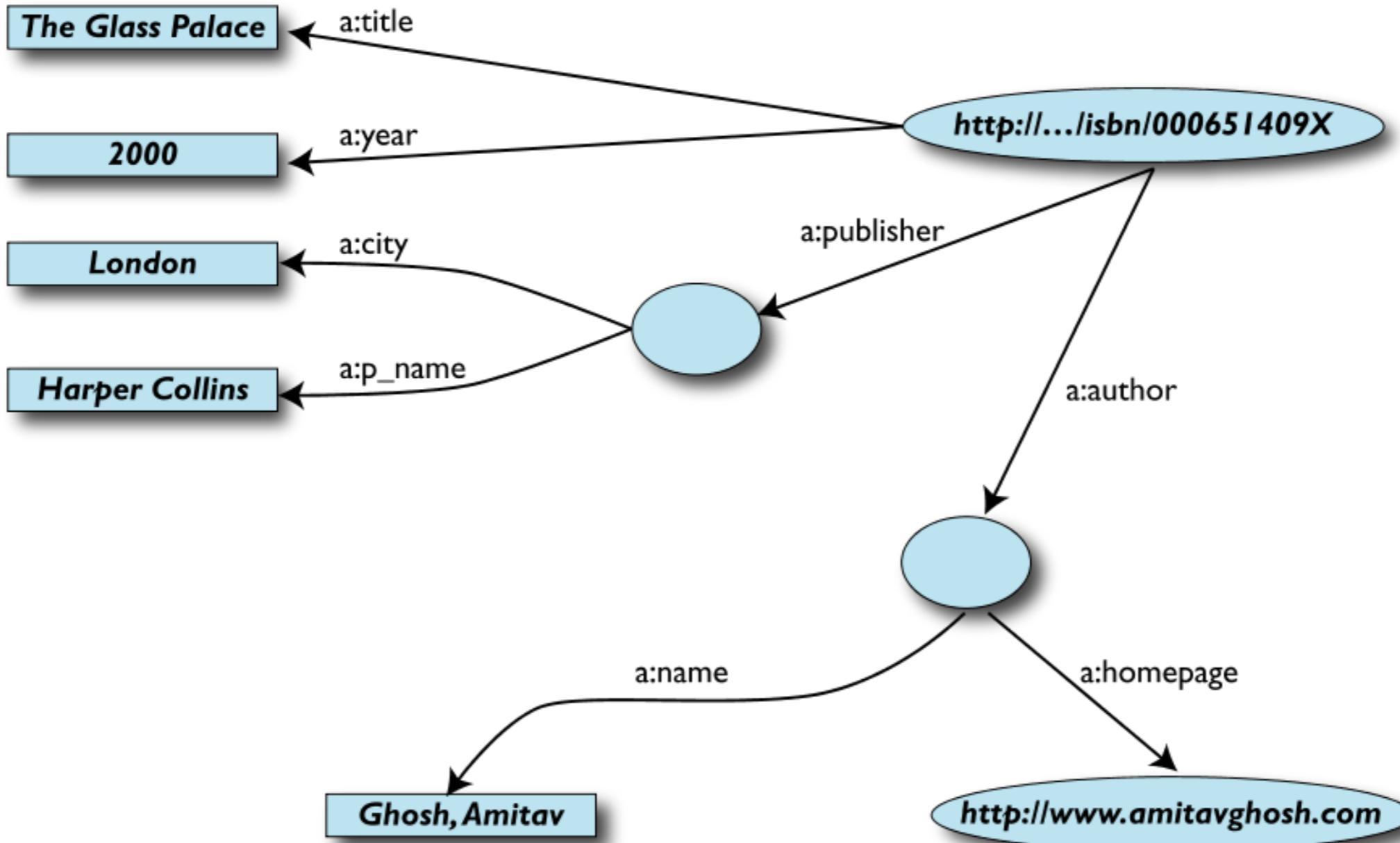
Authors

ID	Name	Home page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

Publishers

ID	Publisher Name	City
id_qpr	Harper Collins	London

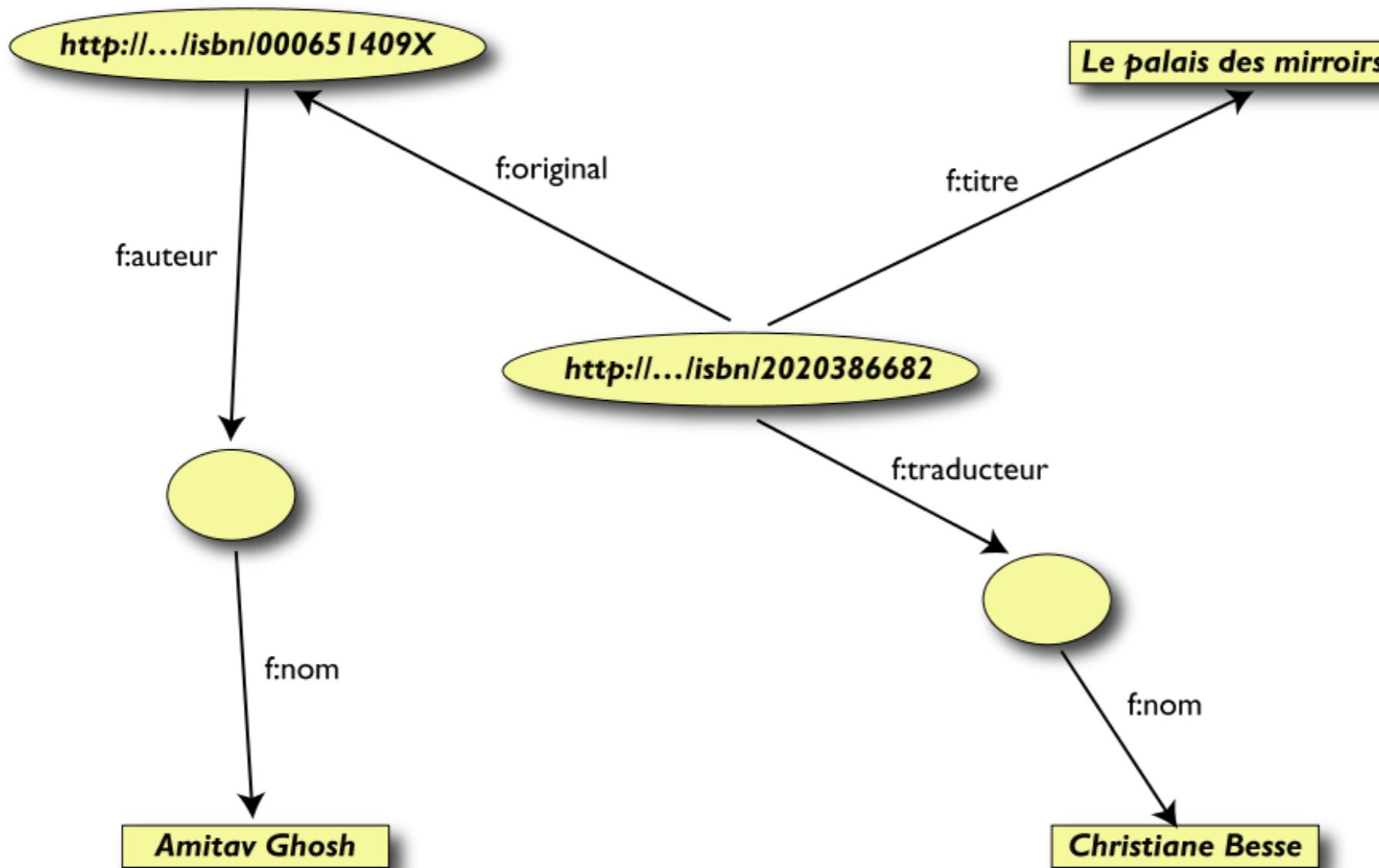
1: EXPORTAR OS DADOS COMO UM CONJUNTO DE RELAÇÕES



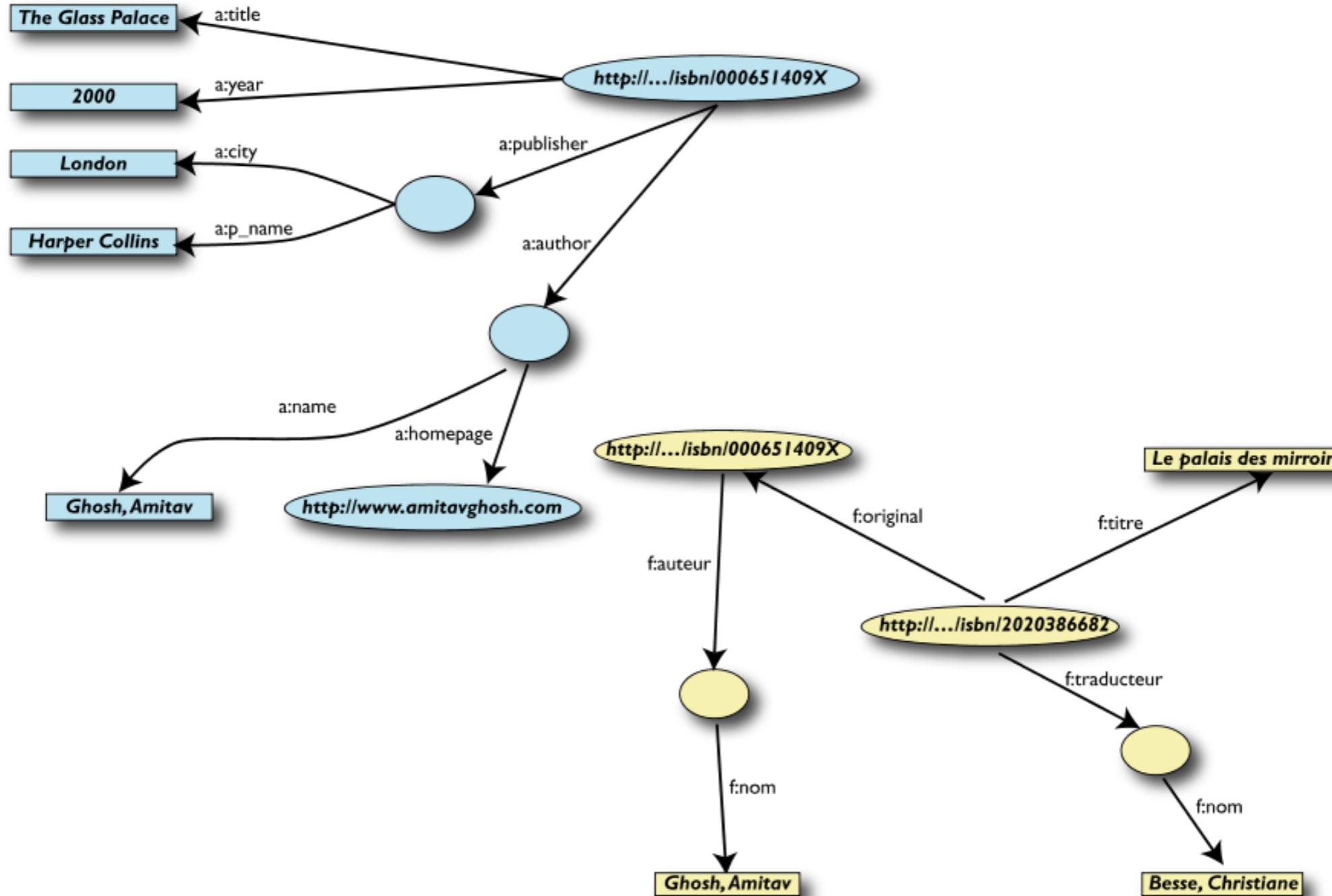
DATASET “F”

	A	B	D	E
1	ID	Titre	Traducteur	Original
2	ISBN0 2020386682	Le Palais des miroirs	A13	ISBN-0-00-651409-X
3				
6	ID	Auteur		
7	ISBN-0-00-651409-X	A12		
11	Nom			
12	Ghosh, Amitav			
13	Besse, Christianne			

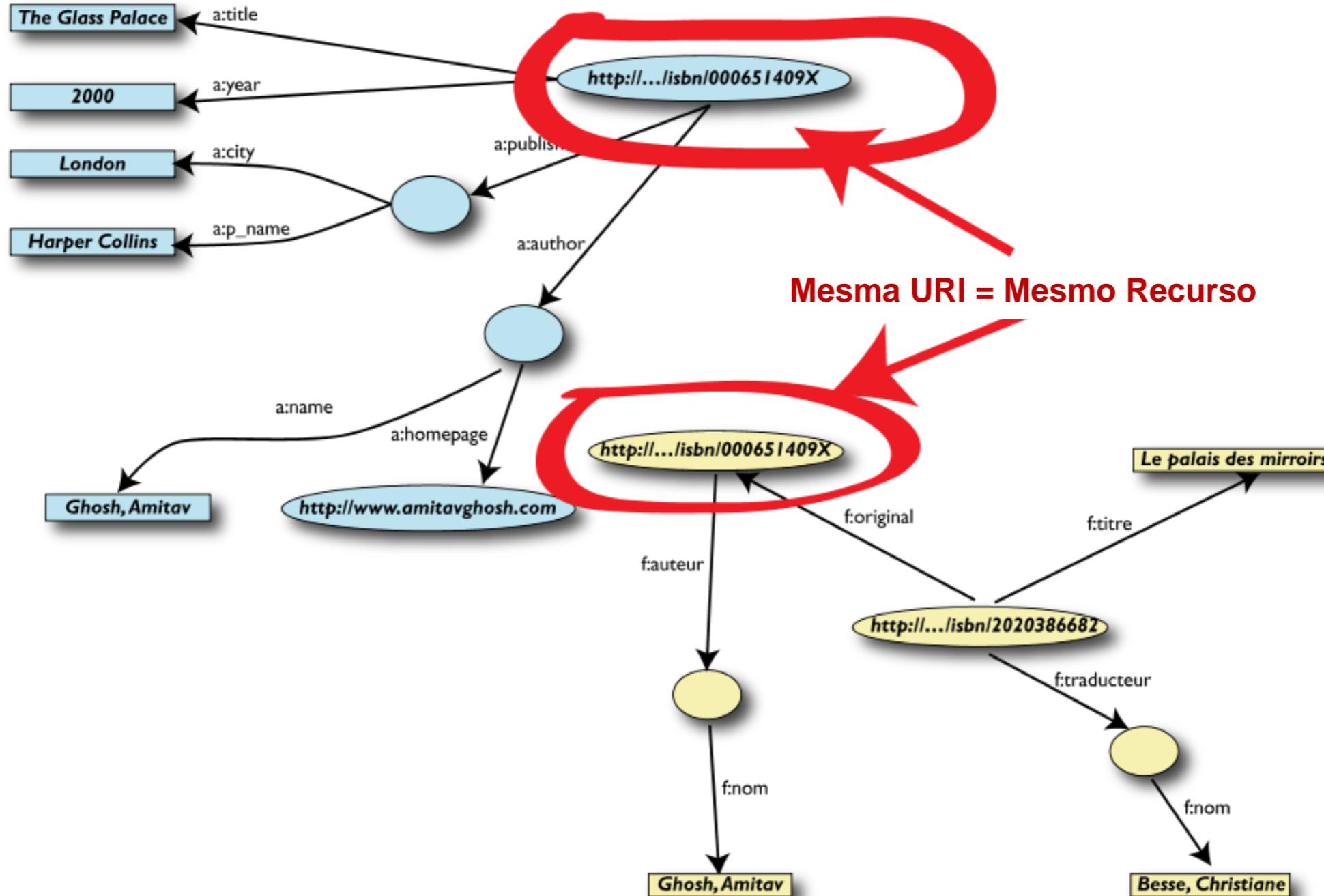
1: EXPORTAR O SEGUNDO CONJUNTO DE DADOS



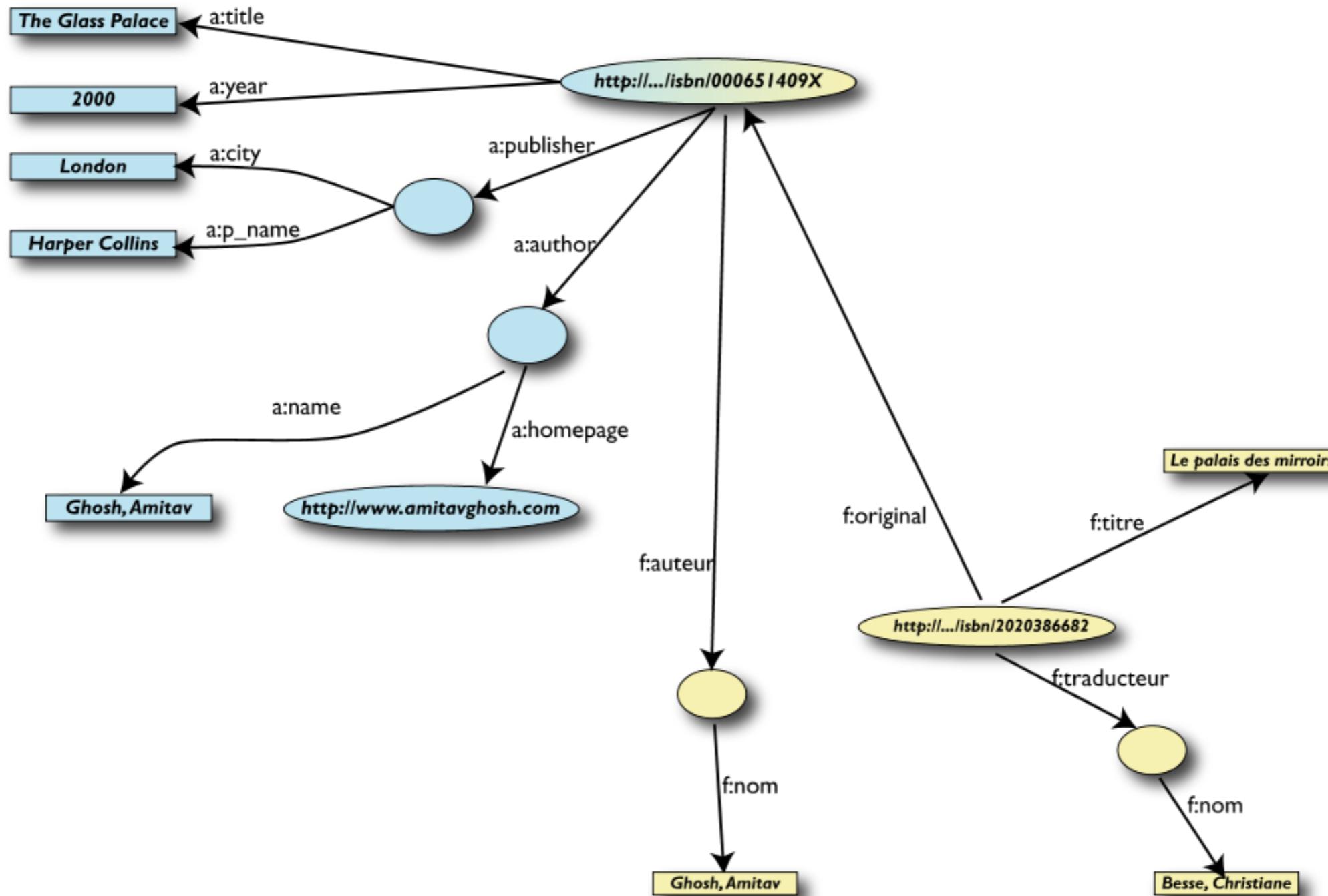
2: COMBINANDO OS DADOS



2: COMBINANDO OS DADOS



2: FUNDIR OS RECURSOS IDÊNTICOS



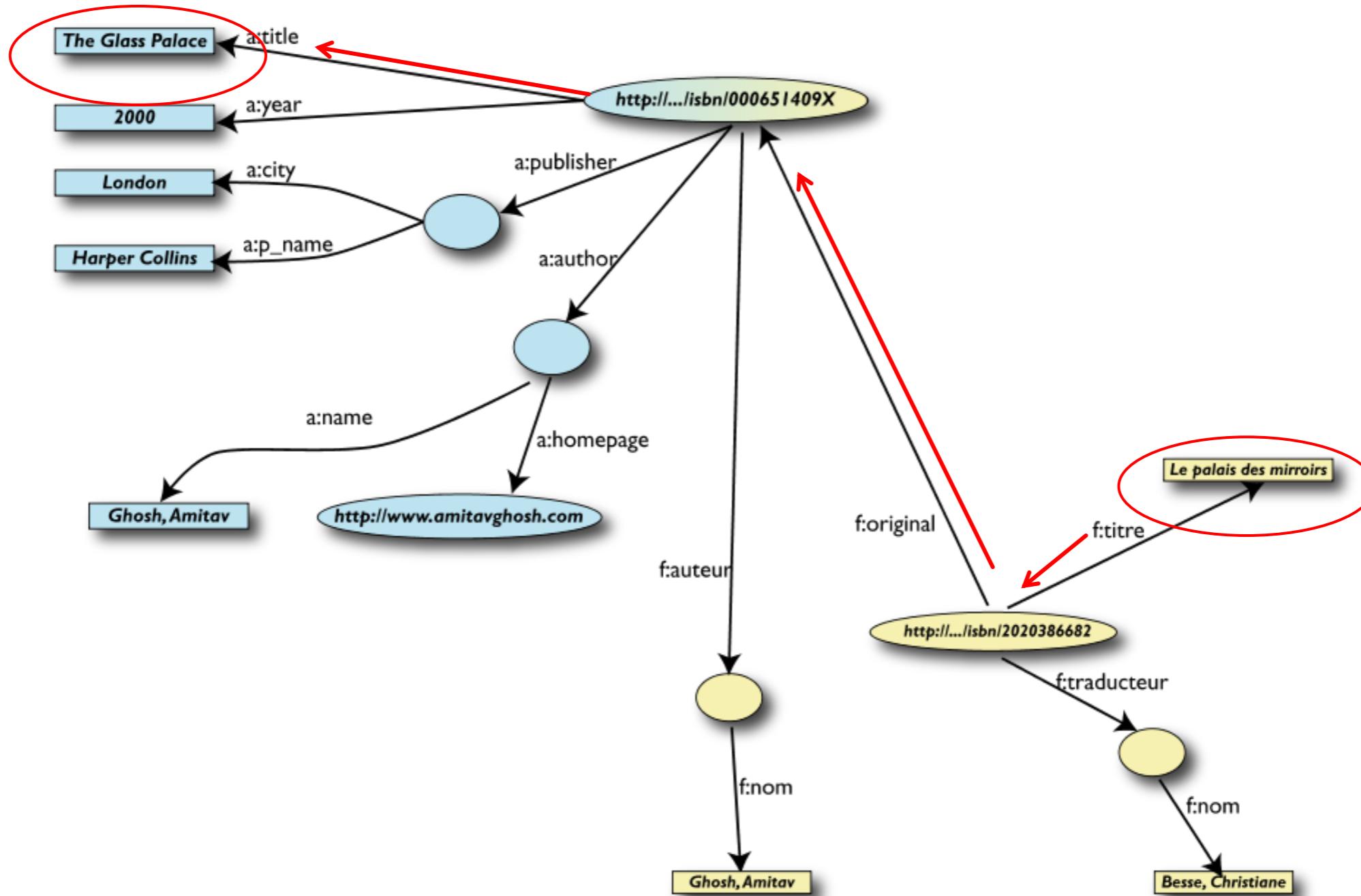
ELABORANDO CONSULTAS NA WEB...

Um usuário do dataset “F” pode agora perguntar algo como:

“Qual o título da versão original de *Le Palais des miroirs*? ”

Essa informação NÃO ESTÁ no dataset “F”, mas pode ser recuperada depois da fusão com o dataset “A”!

3: CONSULTAR O DATASET RESULTANTE



ENTRETANTO, TEMOS MAIS COISAS...

Nós, os humanos “sabemos” que
a:author e *f:auteur*
são a mesma coisa

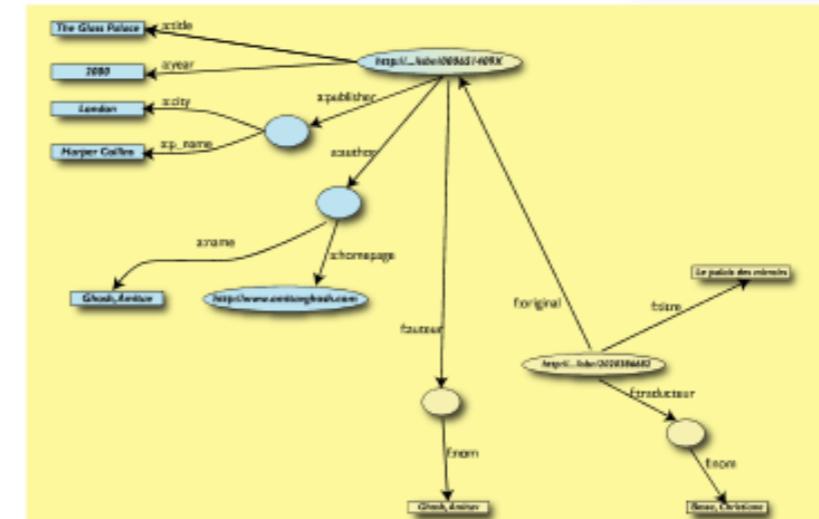
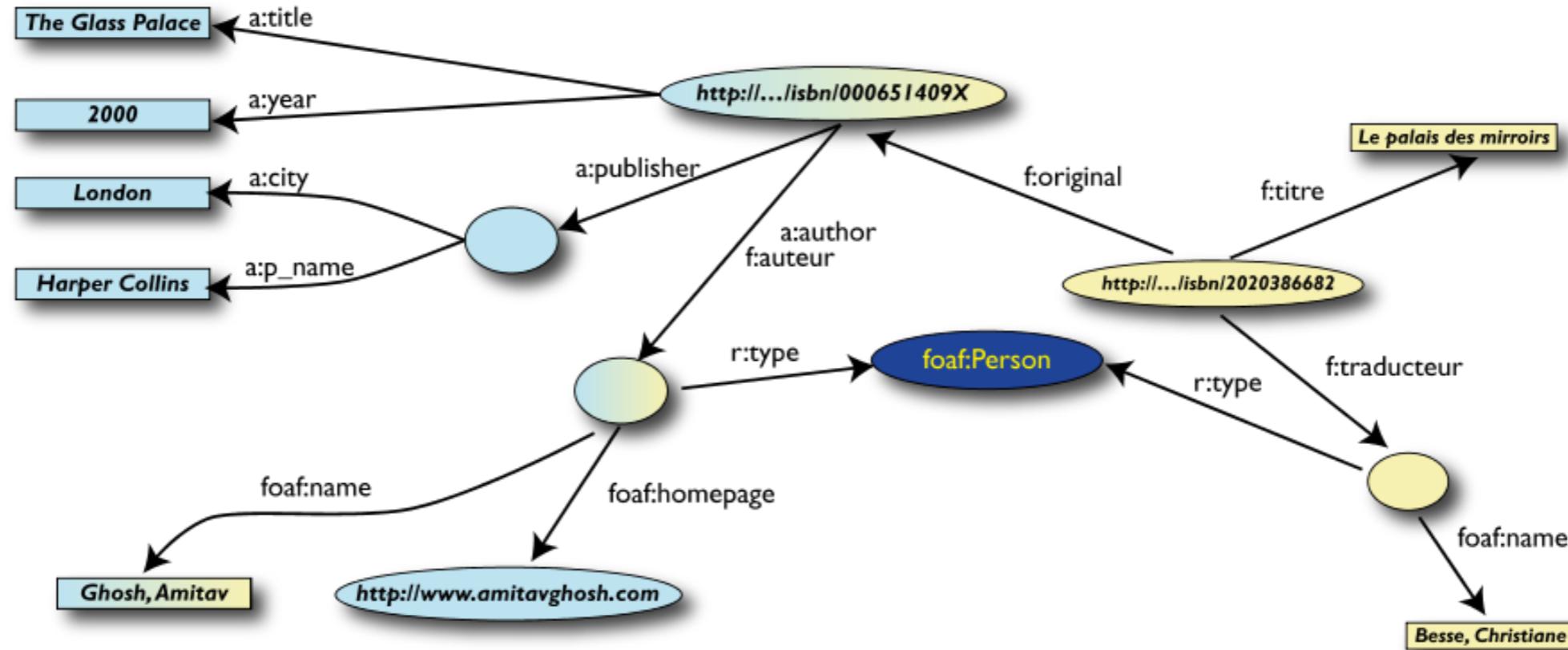
Mas, computador, através da fusão automática NÃO SABE
DISTO!

Que tal acrescentar alguma informação extra
(ANOTAÇÃO) sobre dados fundidos?

a:author same as *f:auteur*

Ambos identificam uma *Person*

3: REVISTO: USO DE CONHECIMENTO EXTRA



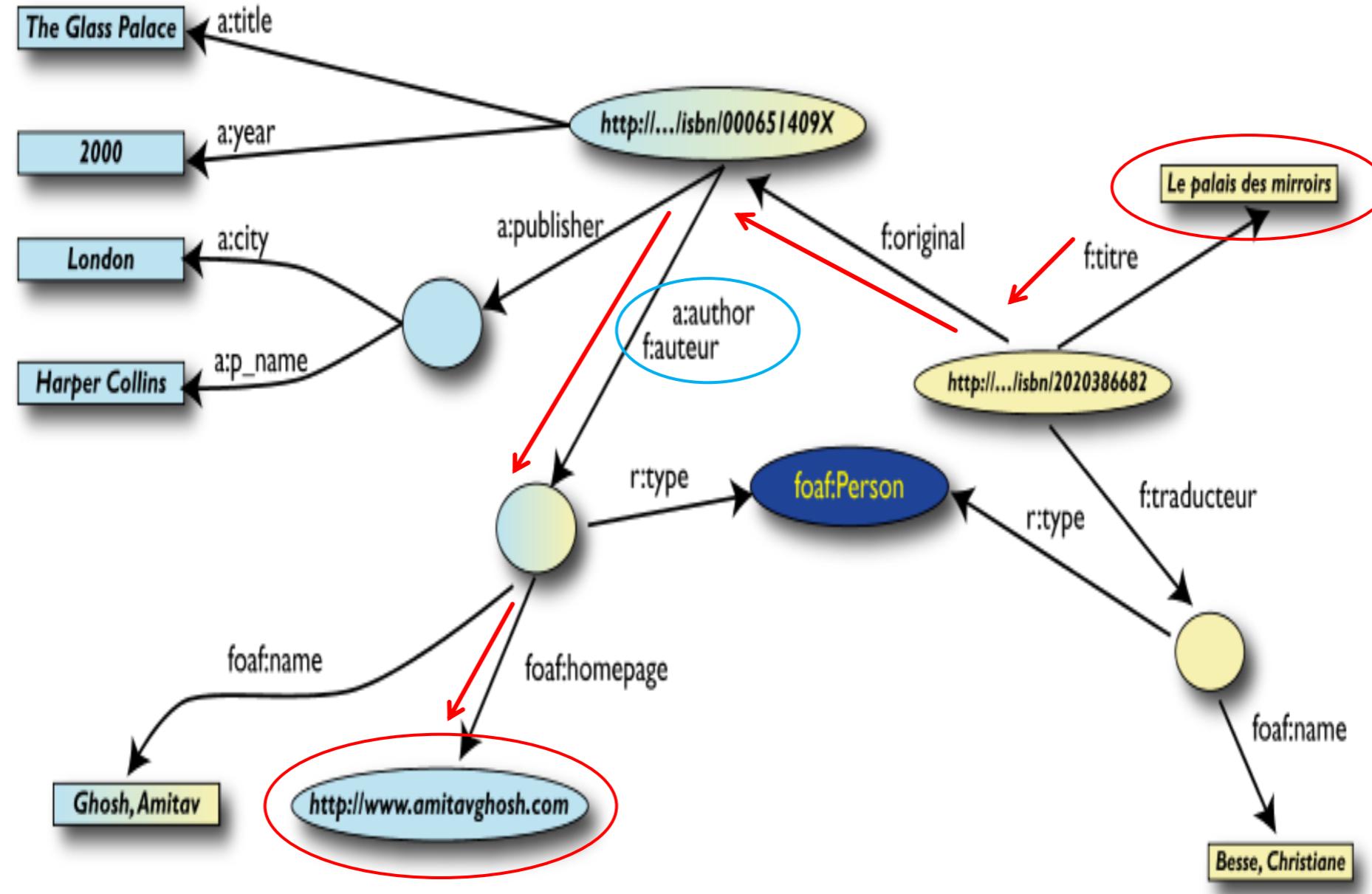
3: CONSULTAS (MUITO) MAIS RICAS!

Um usuário do dataset “F” pode perguntar agora:

“Qual é a homepage do autor de *Le Palais des miroirs*? ”

A informação NÃO está no dataset “F” nem está no dataset “A”, mas se tornou disponível através da fusão (relação) dos datasets “A” e “F” com a adição de 3 declarações simples de “junção”

3: CONSULTAS (MUITO) MAIS RICAS!



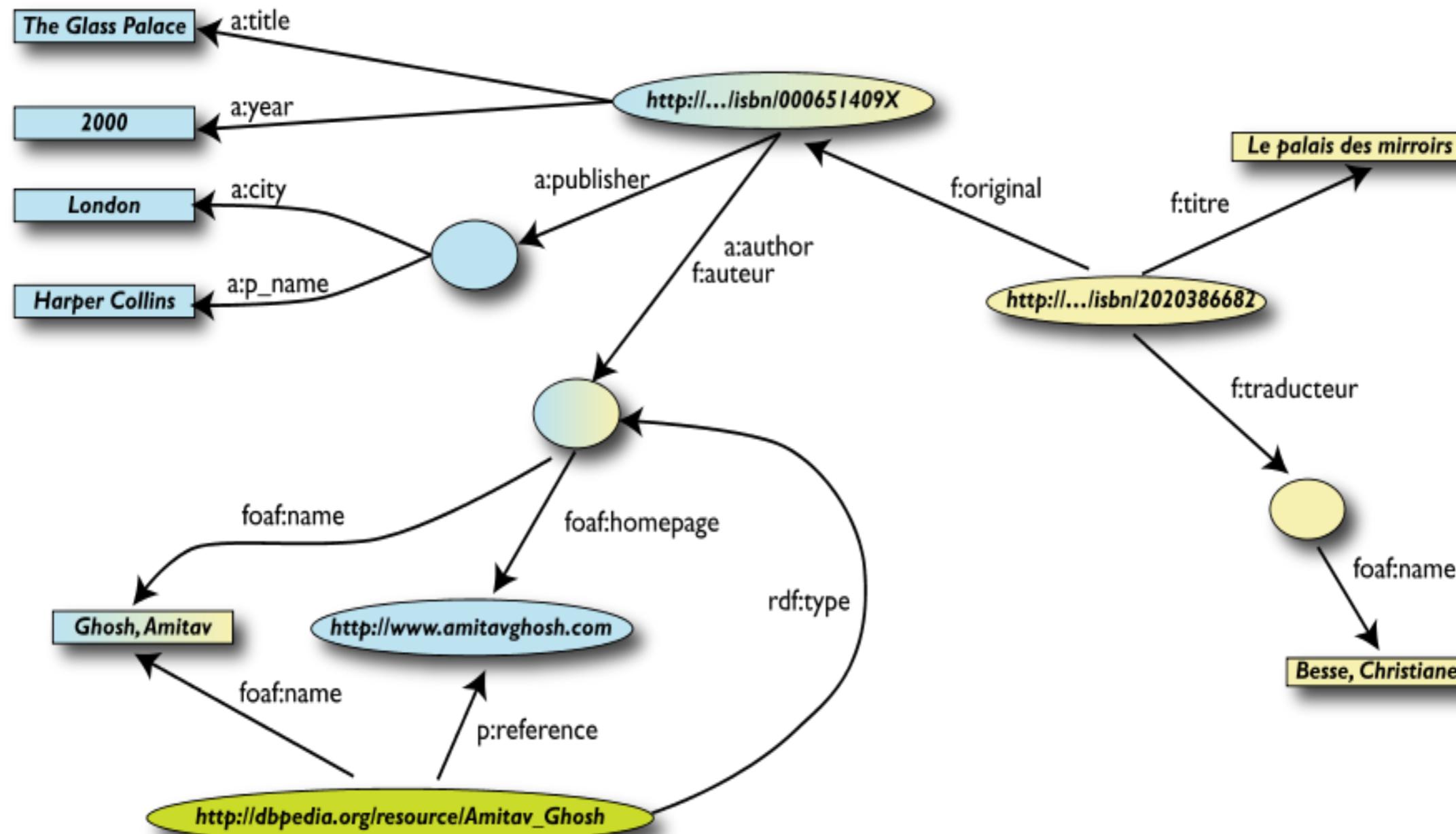
OUTRAS FONTES DE DADOS

Podemos integrar novos dados de outras fontes em nosso dataset

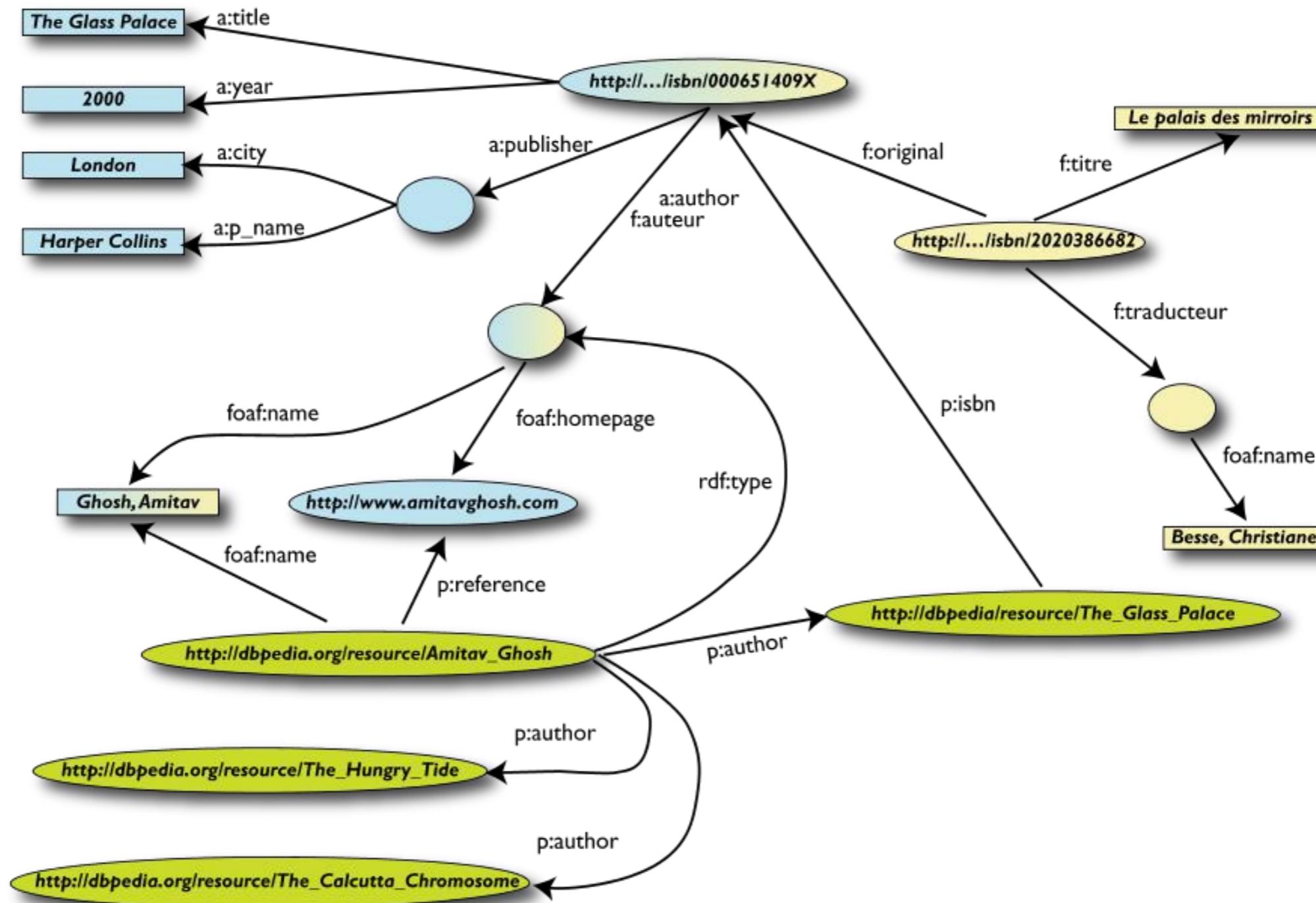
e.g. informação adicional sobre o autor Amitav Ghosh

Talvez a maior fonte de conhecimento público e aberto seja a **Dbpedia***

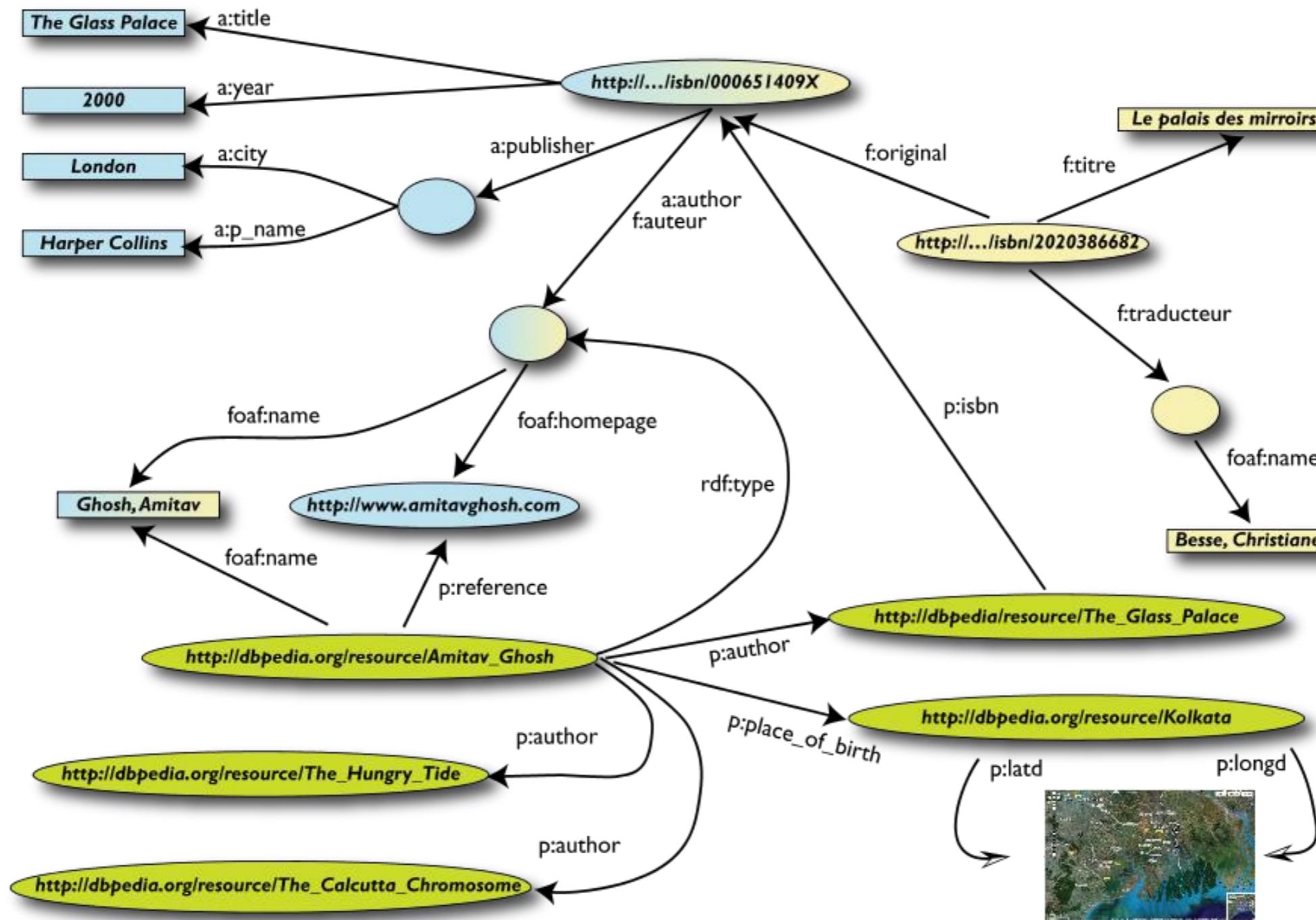
4: FUNDIR COM OS DADOS DA DBPEDIA

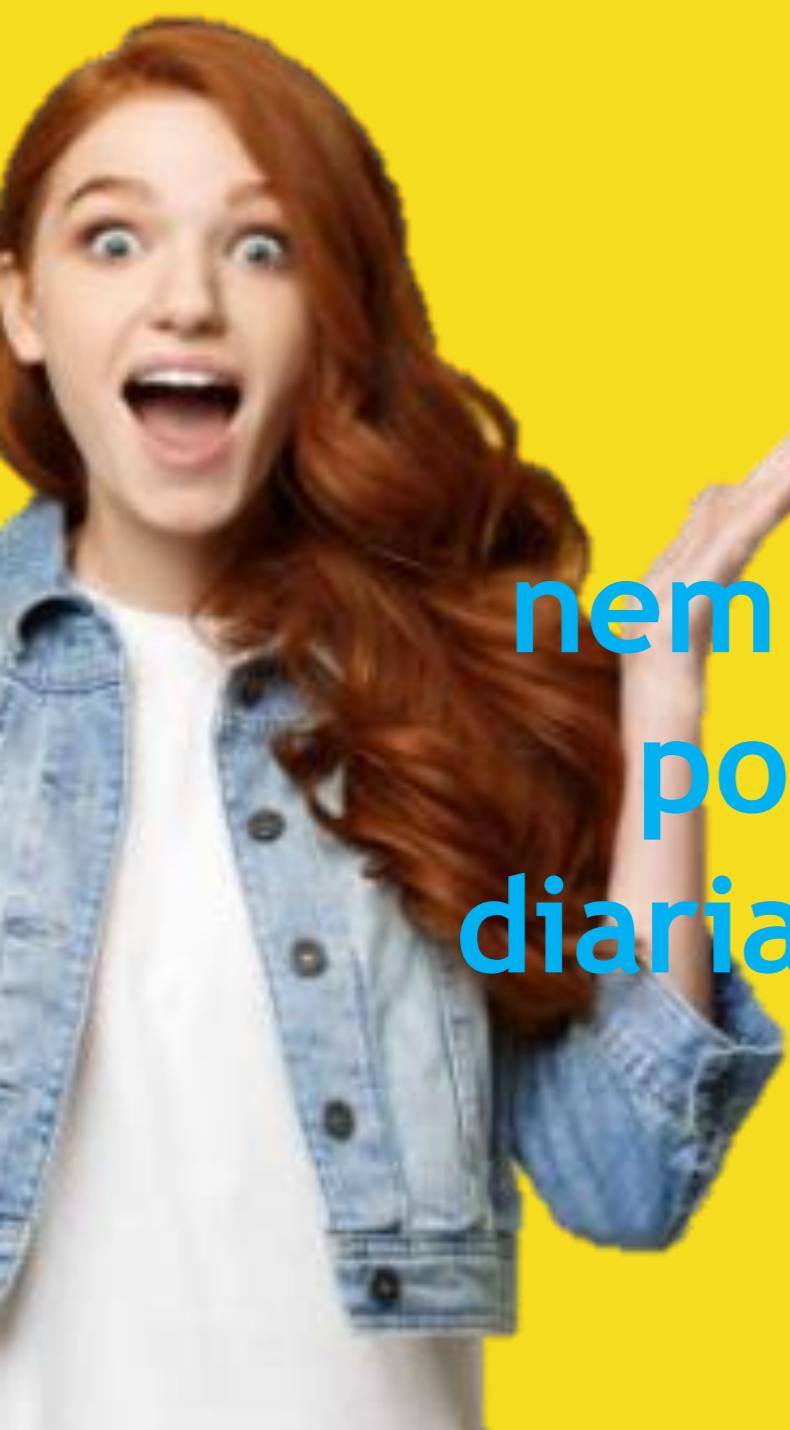


4: FUNDIR COM OS DADOS DA DBPEDIA



4: FUNDIR COM OS DADOS DA DBPEDIA





Isto é surpreendente?

Pode ser, mas de fato,
nem deveria ser, o que aconteceu
por meios automáticos é feito
diariamente pelos usuários na Web!

Qual a diferença?

Houve um pouco mais de rigor,
para que as máquinas
fizessem isso
automaticamente!

Agora vamos apresentar cada
uma das partes...

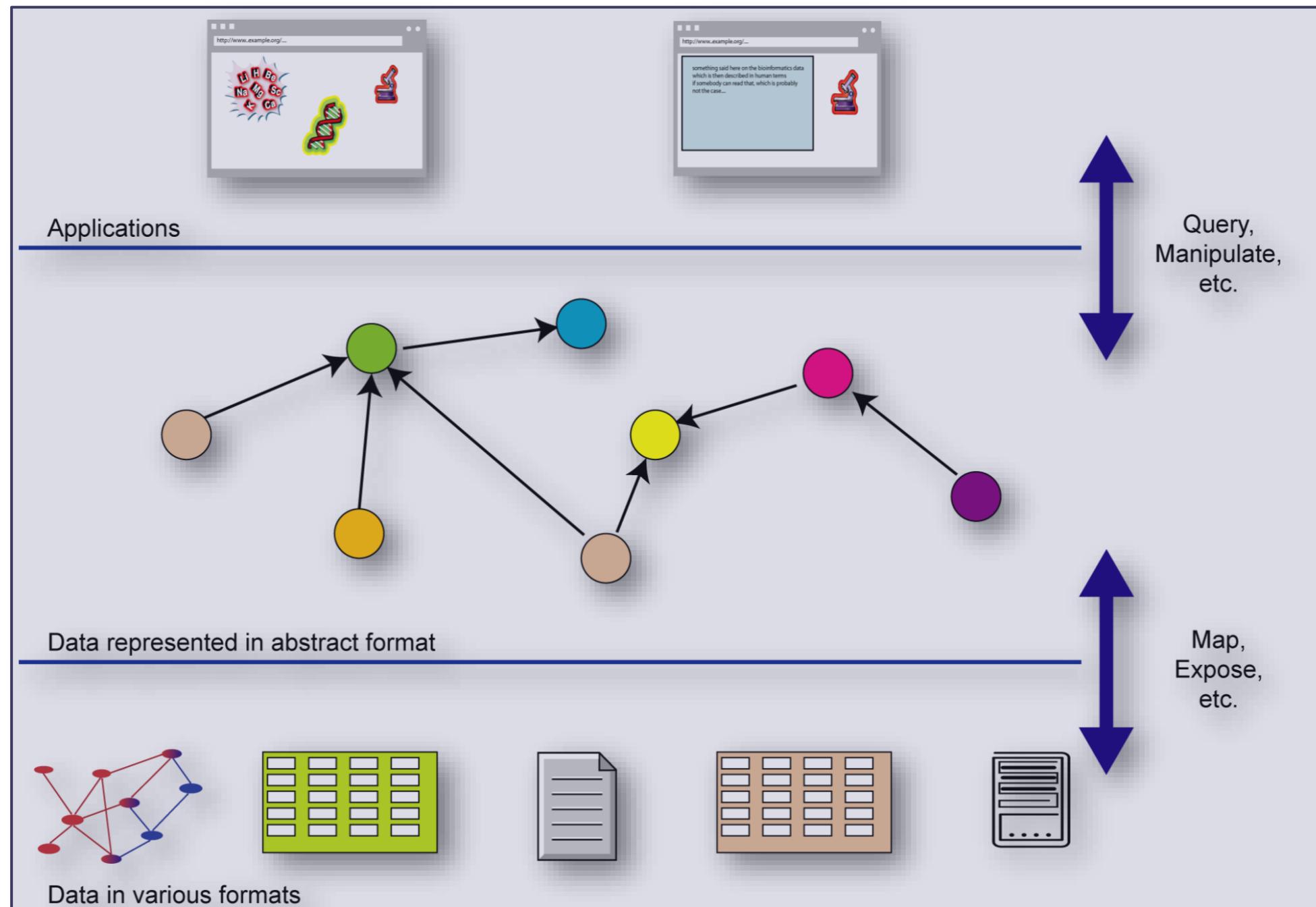


O QUE NÓS FIZEMOS ATÉ AGORA?



- Combinamos diferentes datasets que
 - ...podem ser internos ou estar em algum lugar da Web
 - ...podem ser de formatos diferentes (RDBMS, Excel, CSV, (X)HTML, XML, JSON etc)
 - ...têm diferentes nomes para as mesmas relações
- Combinamos datasets porque algumas URIs são idênticas
 - i.e. o ISBNs no nosso caso
- Adicionamos algumas informações adicionais simples (relações) para ajudar a fundir os datasets
- **O resultado são NOVAS respostas para perguntas que NÃO PODERIAM SER FEITAS ANTES!**

O QUE NÓS FIZEMOS FOI APENAS ISSO...





São as
tecnologias
da
Web
Semântica
que tornam
esta
integração
possível!

As tecnologias da **Web Semântica** são muitas...



RDF é



Resource

Description

Framework

RDF é o modelo de dados
da Web Semântica

Estudar material das aulas de RDF

(se você é aluno(a) do PPGI vá ao final para ver o resumo de RDF/Turtle)

Web Semântica

Adicionar
estrutura e semântica
(RDFS,OWL,RIF)

Definir um
Modelo de Dados
(RDF)

Definir uma
Linguagem de Consulta
(SPARQL)

Trabalhar na
Web “real”
(Linked Data, RDFa)

SPARQL é

SPARQL
Protocol
And
RDF
Query
Language



SPARQL é

A linguagem de consulta
da Web Semântica

Por que SPARQL?

- 1 Obter informação de dados estruturados e semi-estruturados
- 2 Explorar dados através da descoberta de relacionamentos desconhecidos
- 3 Consultar e pesquisar com uma visão integrada de diferentes fontes de dados
- 4 Reunir diferentes aplicações através da transformação de um vocabulário em outro



	A	B	C	D	E	F	G	H	I
CLASS	MPFR	CAR LINE	DISPLACENUMB	CYL TRANS	DRIVE SVINDEX	NUN CITY	MPG		
1	TMO SEAT	SEAT	MAV8	VANTAGE	4. 3	8 Auto(S6) R	12	13	
2	TMO SEAT	SEAT	ASTON	MAV8 VANTAGE	4. 3	8 Manual(M6R)	11	12	
3	TMO SEAT	SEAT	ASTON	MAV8 VANTAGE	4. 3	8 Auto(L5) R	4	19	13
4	TMO SEAT	SEAT	ASTON	MAV8 VANTAGE	4. 3	8 Manual(M6R)	4	19	13
5	TMO SEAT	SEAT	AUDI	TT ROADSTER	4. 2	6 Auto(S6) F	2	25	23
6	TMO SEAT	SEAT	AUDI	TT ROADSTER	4. 2	6 Auto(S6) F	2	25	23
7	TMO SEAT	SEAT	AUDI	TT ROADSTER	3. 2	6 Auto(S6) F	4	8	18
8	TMO SEAT	SEAT	AUDI	TT ROADSTER	3. 2	6 Manual(M6)	4	8	17
9	TMO SEAT	SEAT	BMW	Z4 3.0i	3	6 Auto(S6) R	30	19	
10	TMO SEAT	SEAT	BMW	Z4 3.0i	3	6 Manual(M6R)	30	18	
11	TMO SEAT	SEAT	BMW	Z4 3.0i	3	6 Auto(S6) R	30	19	
12	TMO SEAT	SEAT	BMW	Z4 COUPE	3	6 Auto(S6) R	30	19	
13	TMO SEAT	SEAT	BMW	Z4 COUPE	3	6 Auto(S6) R	30	18	
14	TMO SEAT	SEAT	BMW	Z4 COUPE	3	6 Auto(S6) R	30	18	
15	TMO SEAT	SEAT	BMW	Z4 M COUPE	3. 2	6 Manual(M6)	32	15	
16	TMO SEAT	SEAT	BMW	Z4 M ROADSTER	3. 2	6 Manual(M6R)	32	15	
17	TMO SEAT	SEAT	BUGATTI	VEYRON	8	16 Auto(S6) R	4	13	8
18	TMO SEAT	SEAT	CADILLAC	XLR	4. 4	8 Auto(S6) R	24	14	
19	TMO SEAT	SEAT	CADILLAC	XLR	4. 4	8 Auto(S6) R	30	15	
20	TMO SEAT	SEAT	CHEVROLET	CORVETTE	6. 2	8 Auto(S6) R	26	15	
21	TMO SEAT	SEAT	CHEVROLET	CORVETTE	6. 2	8 Manual(M6R)	27	16	
22	TMO SEAT	SEAT	CHEVROLET	CORVETTE	7	8 Manual(M6R)	35	15	
23	TMO SEAT	SEAT	CHRYSLER	CROSSFIRE	3. 2	6 Auto(L5) R	995	19	
24	TMO SEAT	SEAT	CHRYSLER	CROSSFIRE	3. 2	6 Manual(M6R)	995	15	
25	TMO SEAT	SEAT	CHRYSLER	CROSSFIRE	3. 2	6 Auto(L5) R	995	15	

EPA Fuel Efficiency
Spreadsheet

Employee
Directory

ERP / Budget
System

SPARQL Query Engine

Quais automóveis fazem mais de 10Km por litro, têm custo dentro do orçamento do meu departamento e podem ser comprados em uma concessionária localizada até 15 Km de um dos meus empregados?

Web dashboard

```

SELECT ?automobile
WHERE {
?automobile a ex:Car ; epa:mpg ?mpg ;
ex:dealer ?dealer .
?employee a ex:Employee ; geo:loc ?loc .
?dealer geo:loc ?dealerloc .
FILTER(?mpg > 25 &&
geo:dist(?loc, ?dealerloc) <= 10) .
}
  
```

SPARQL
query

Estudar material das aulas de SPARQL

Encontrar todos os países sem litoral com uma população maior que 15 milhões de habitantes.

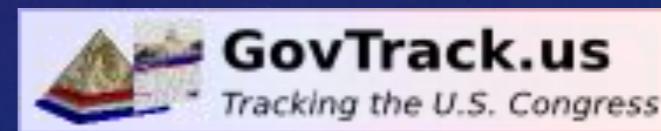
Consultando a Wikipedia...(http://dbpedia.org/sparql)

Exemplo de SPARQL

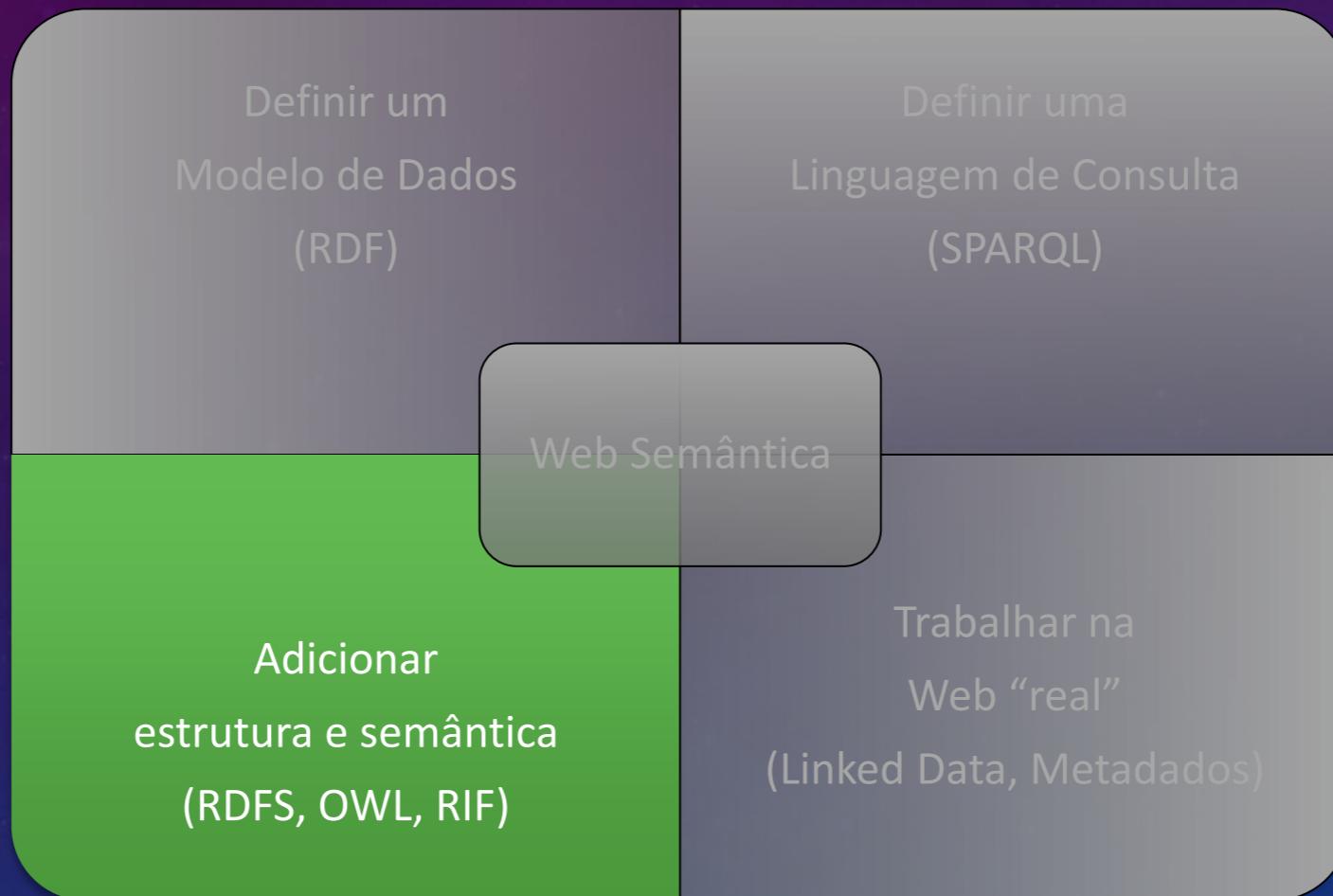
```
PREFIX type: <http://dbpedia.org/class/yago/>
PREFIX prop: <http://dbpedia.org/property/>
SELECT ?country_name ?population
WHERE {
    ?country a type:LandlockedCountries ;
              rdfs:label ?country_name ;
              prop:populationEstimate ?population .
    FILTER (
        ?population > 15000000 &&
        langMatches(lang(?country_name), "EN")
    ) .
}
ORDER BY DESC(?population)
```



(Alguns) Datasets SPARQL'áveis



Mas...
nada do que vimos até
agora se aproxima da
visão a longo prazo da
Web Semântica



**Do conhecimento
explícito para o inferido..
e além!**

RDFS

OWL

RIF

Ontologias

- “Vocabulário de termos e conceitos”
- Novos termos formados a partir dos existentes
- A Semântica é formalmente especificada
- Também registra os relacionamentos entre termos

RDFS é
RDF
Schema

RDF Schema

Vocabulário (definição de termos)

e.g. define um relacionamento “dose prescrita”

Schema (definição de tipos)

“dose prescrita” relaciona “tratamento” e “dosagem”

Taxonomia (definição de hierarquias)

Qualquer “médico” é um “profissional médico”

OWL é
Web
Ontology
Language

OWL

Identidade igual/diferente

“author” e “auteur” são a mesma relação

dois recursos com o mesmo “ISBN” são o mesmo “book”

Definições de tipo mais expressivas

Um “cycle” é um “vehicle” com pelo menos uma “wheel”

Uma “bicycle” é um “cycle” com exatamente duas “wheels”

Definições de relações mais expressivas

“irmãoDe” é uma relação simétrica

*O valor de “diaDaSemana” deve ser um de “segunda”, “terça”, “quarta”,
“quinta”, “sexta”, “sábado”, “domingo”*

O que podemos fazer com OWL?

Consistência

Há alguma contradição neste modelo?

Classificação

Quais são os tipos inferidos deste recurso?

Correção

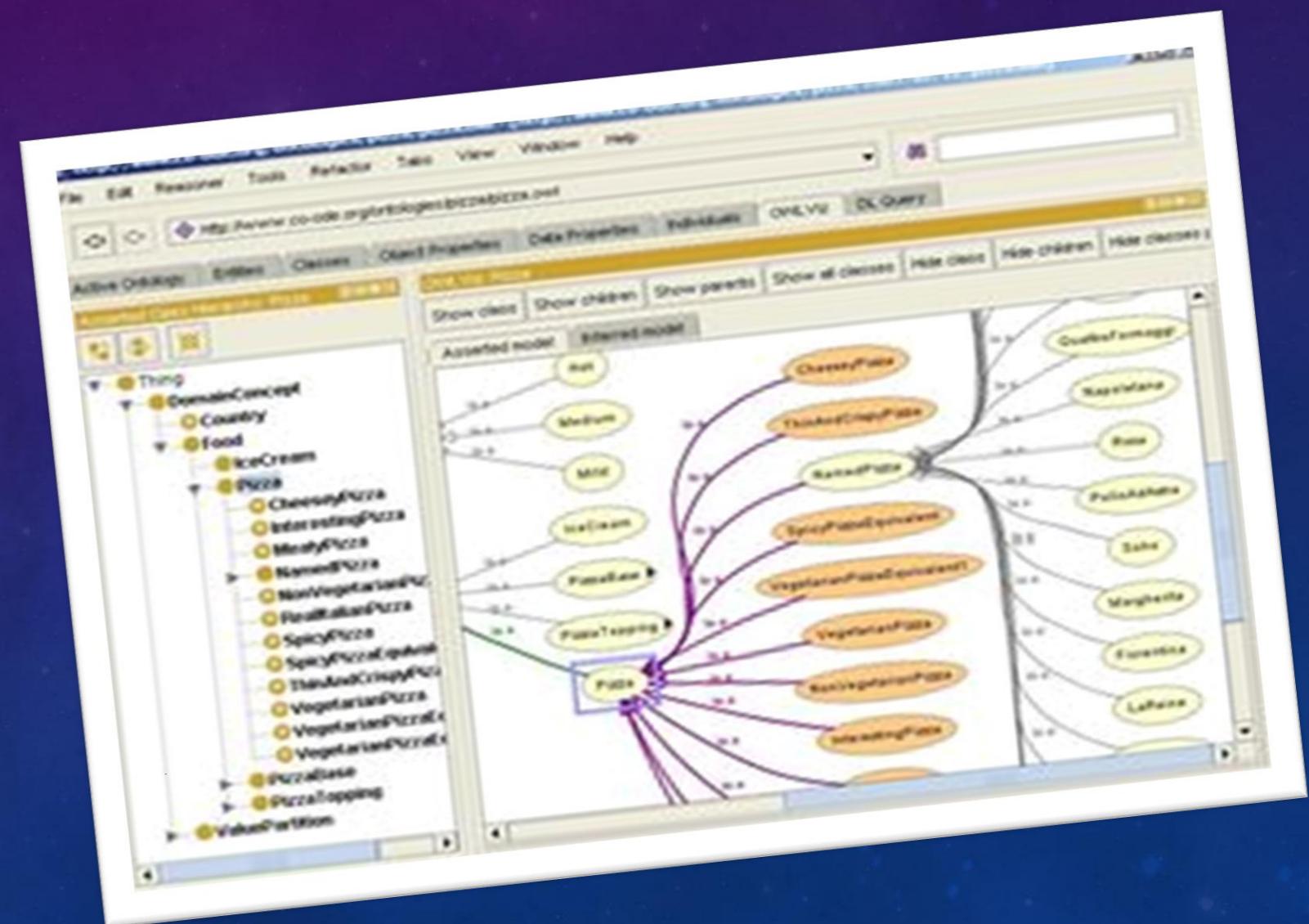
Há alguma classe nesta ontologia que não pode ter membros?

Como podemos criar OWL?

Engenharia de Ontologias + Modelagem Conceitual

Editores de OWL

Etc..



**Estudar materiais das aulas de
Modelagem Ontológica e OWL**



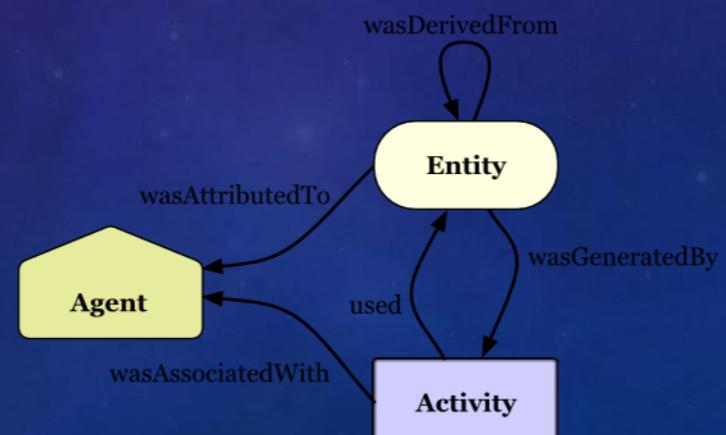
FOAF

Ontologias públicas são reusáveis!



Measurement Units Ontology

Time Ontology in OWL



Provenance Ontology



OBO Foundry



Definir um
Modelo de Dados
(RDF)

Definir uma
Linguagem de Consulta
(SPARQL)

Web
Semântica

Adicionar
estrutura e semântica
(RDFS,OWL,RIF)

**Trabalhar na
Web “real”
(Linked Data,
Metadados)**

Linked Data é

Um conjunto de 4 regras muito simples para publicação de dados RDF na Web

Desenvolvido por Tim Berners-Lee em 2006



Usar URIs como nomes para as coisas,
como forma de uma identidade única
global



Usar URIs HTTP

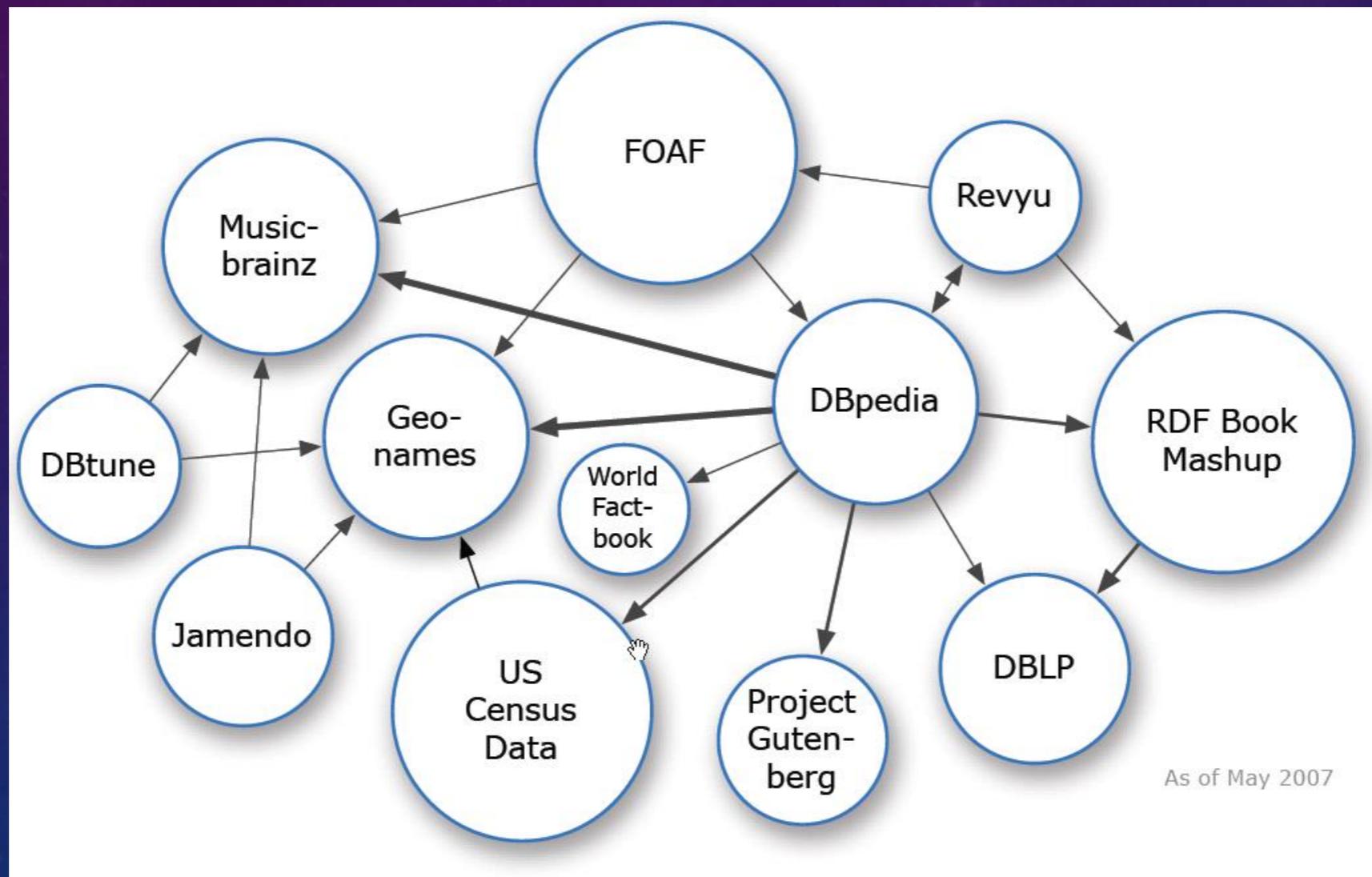


Quando alguém procurar uma URI,
fornecer informação útil na forma de
dados RDF

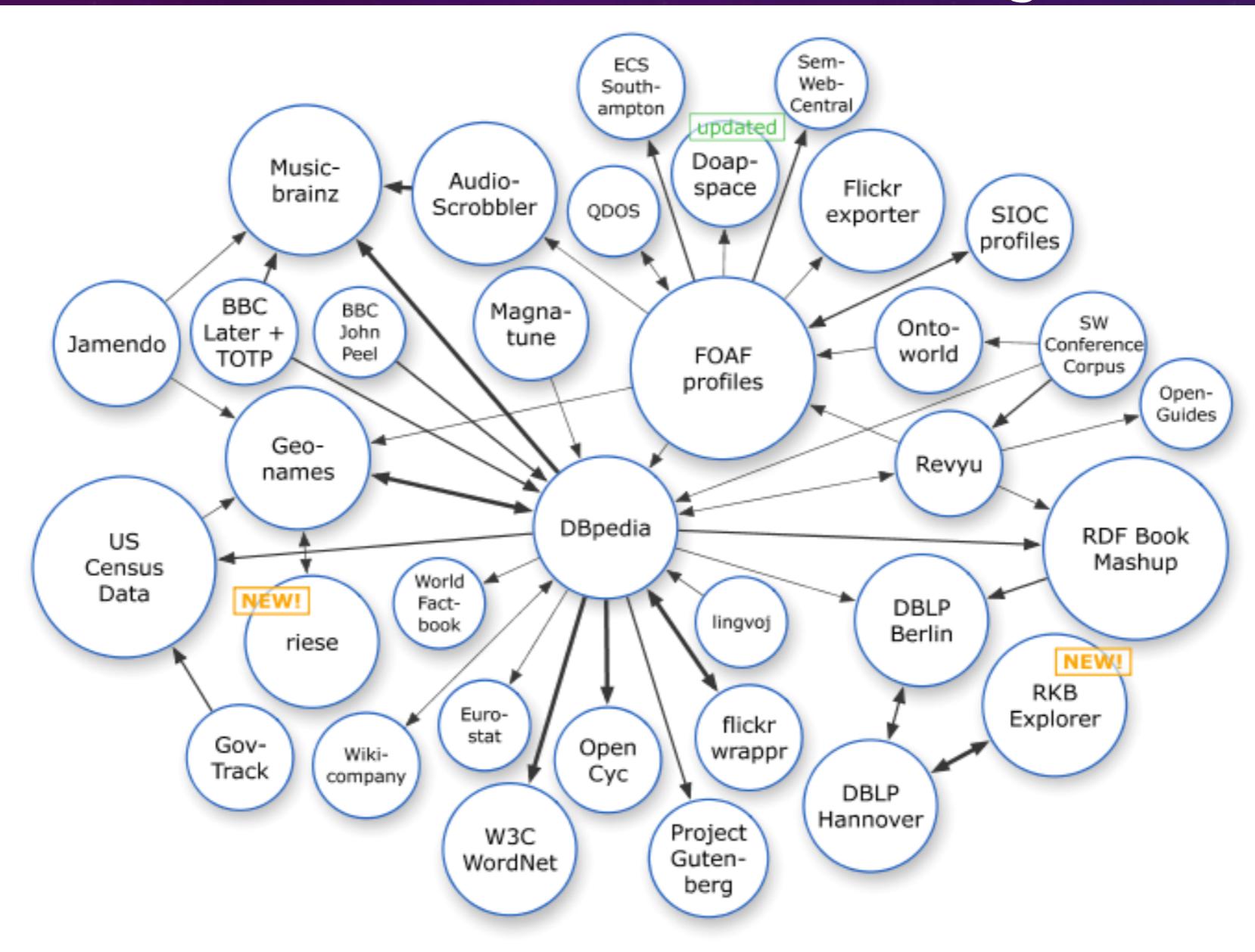


Incluir links para outras URIs

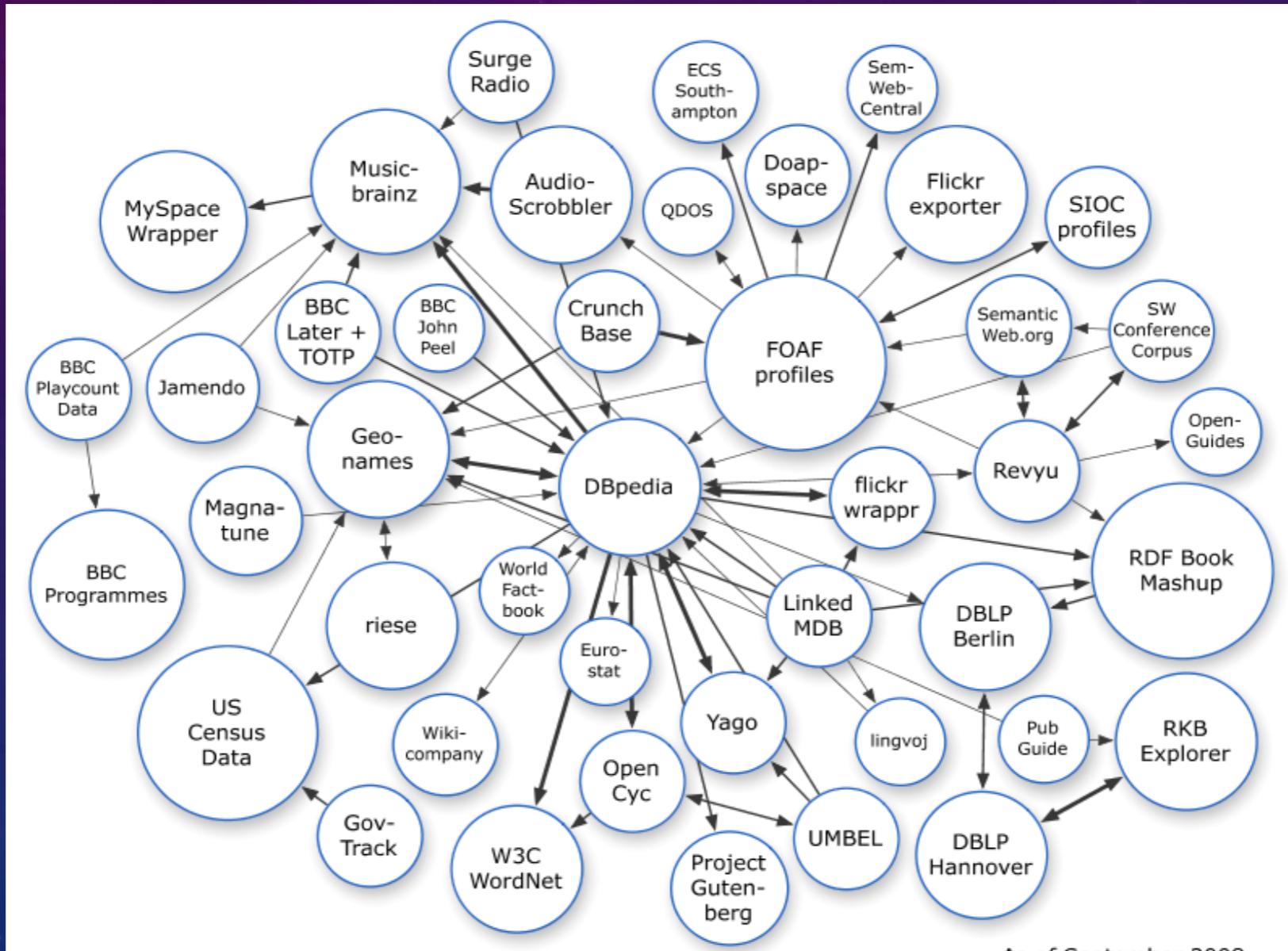
The LOD “cloud”, Maio 2007



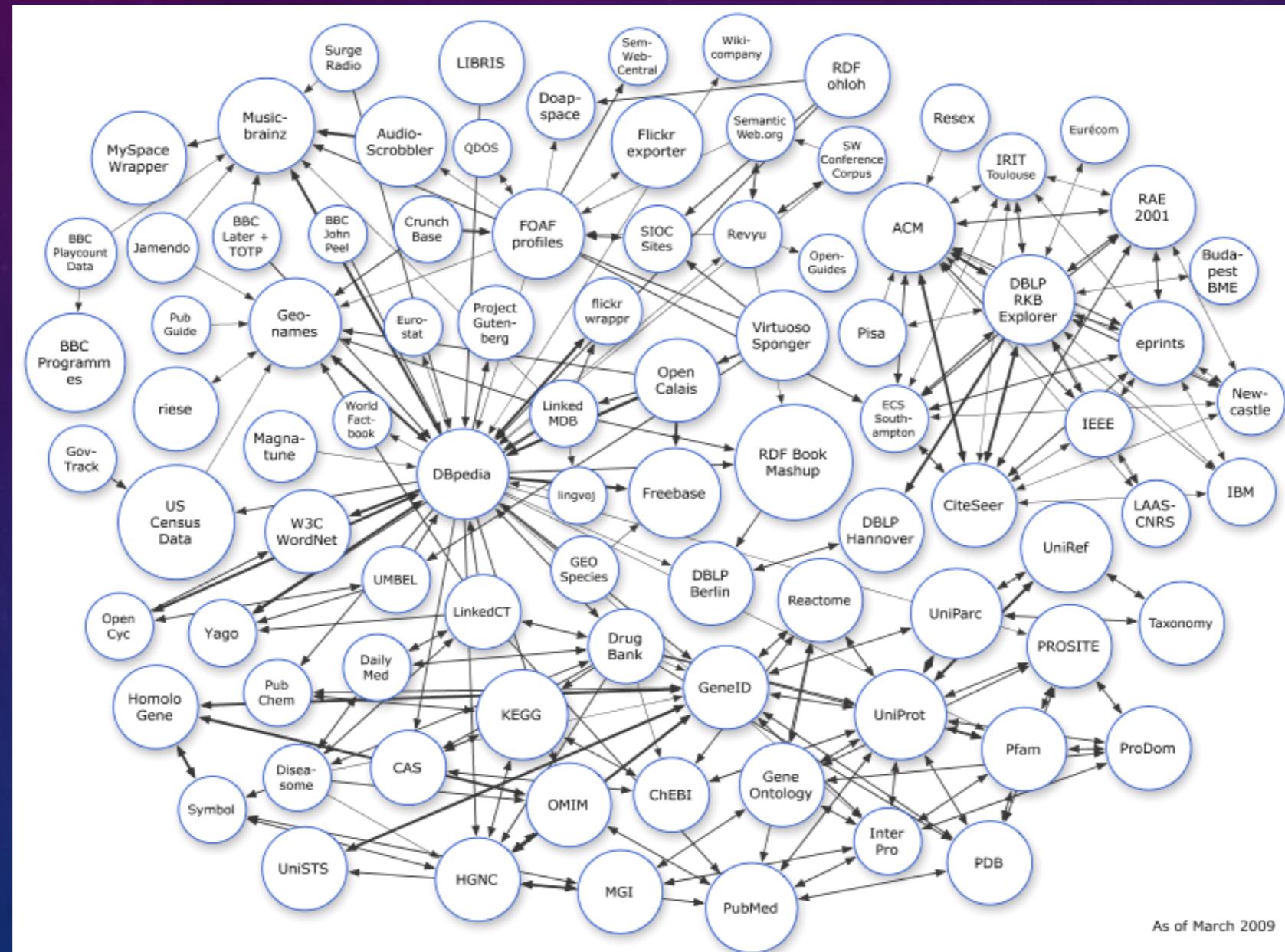
The LOD “cloud”, Março 2008



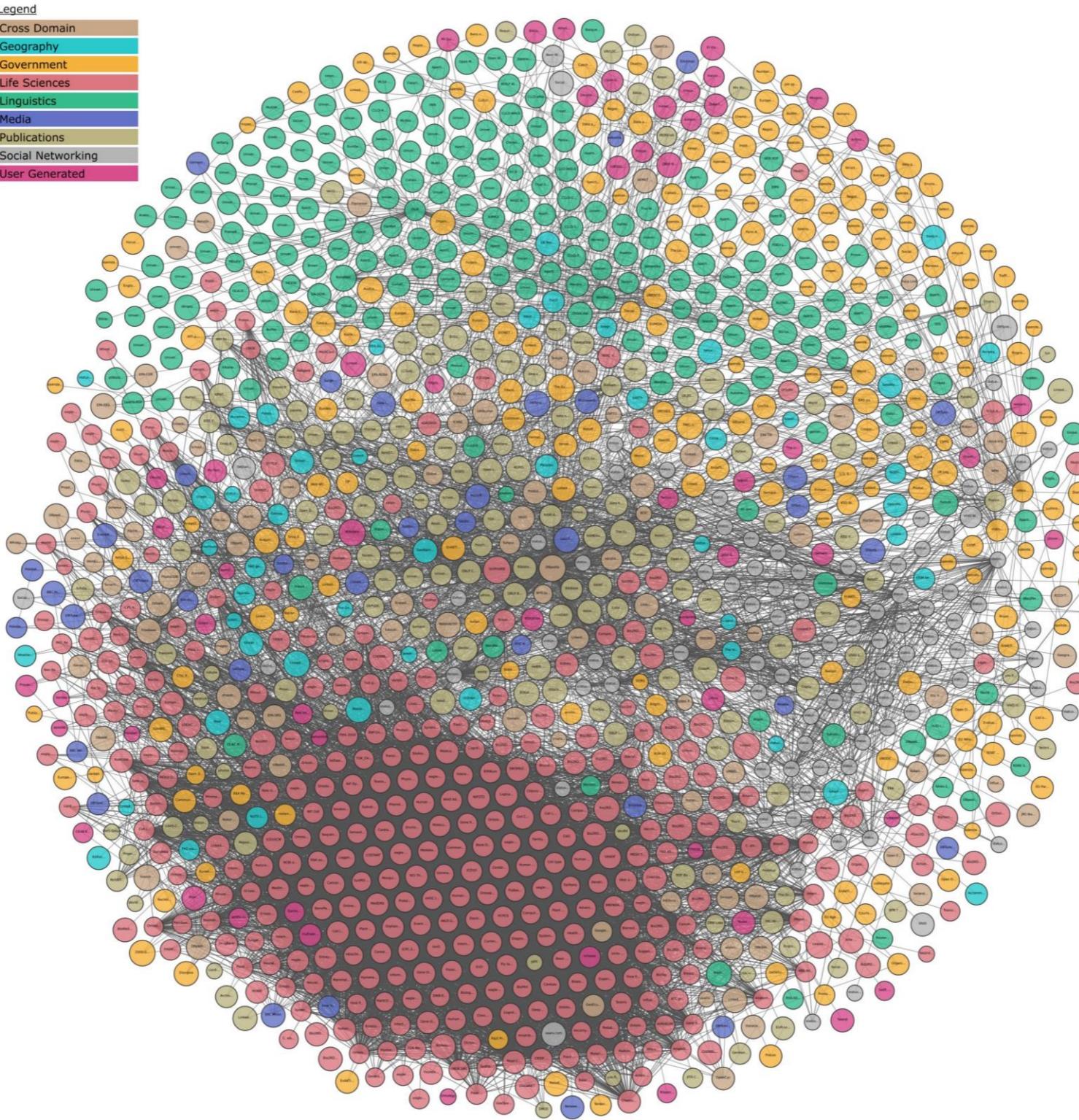
The LOD “cloud”, Setember 2008



The LOD “cloud”, Março 2009



The LOD “cloud”, Fev 2021



<https://lod-cloud.net/>



Estudar slides da aula de Open Data e LOD e Metadados e RDFa

Resumindo o que você viu nessa palestra até agora....

**1- Modelo de Dados
(RDF)**

**2- Linguagem de
Consulta
(SPARQL)**

**Web
Semântica**

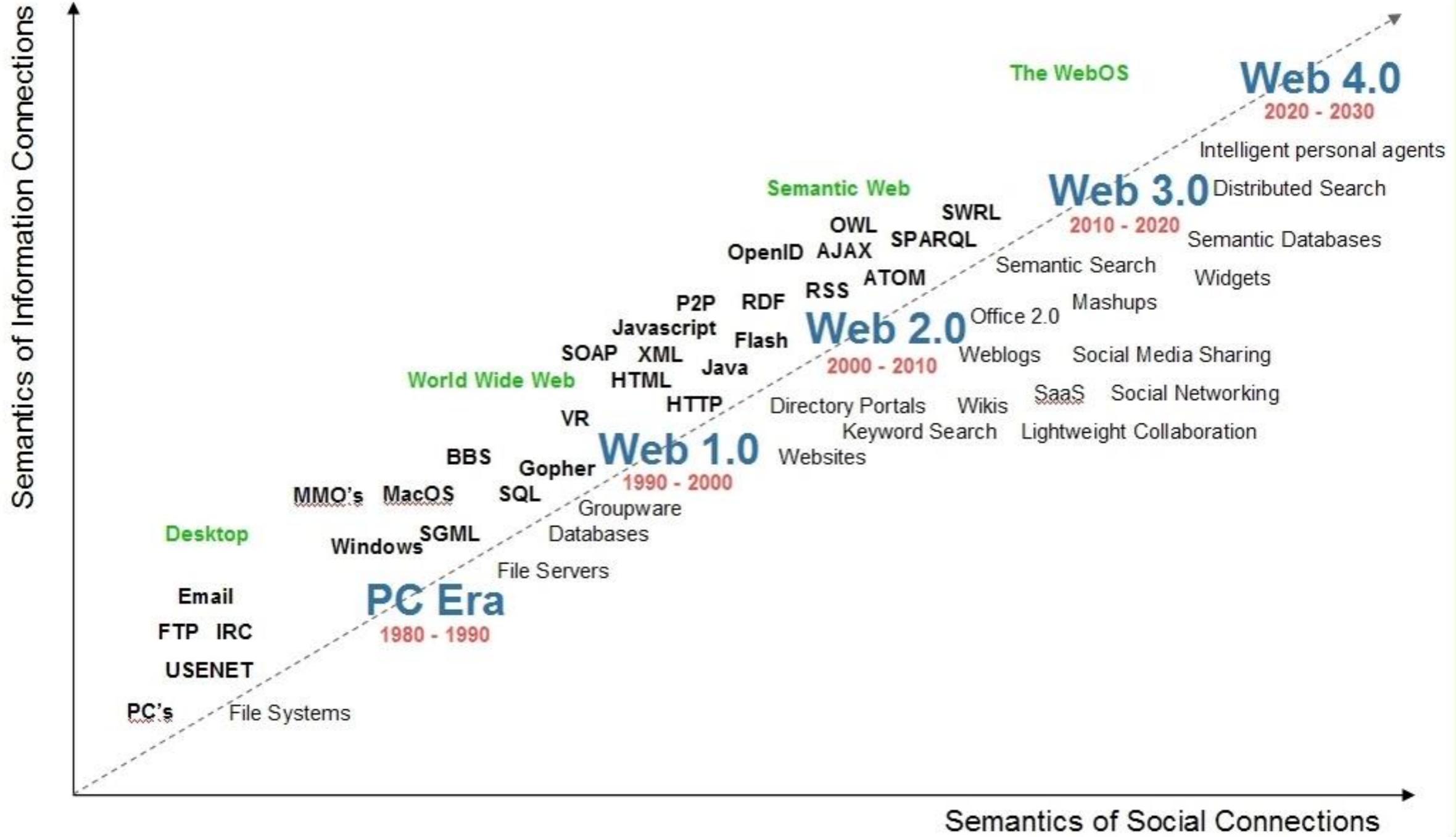
**3- Estrutura
Semântica
(RDFS,OWL,Eng Ont)**

**4- Trabalhar na
Web “real”
(Linked Data,
Metadados)**

*“I have a dream for
the Web in which
computers become
capable of
analyzing all the
data on the Web”*

A close-up portrait of Tim Berners-Lee, the inventor of the World Wide Web. He is a middle-aged man with short, light-colored hair, wearing a dark suit jacket over a light blue shirt and a patterned tie. He is looking slightly upwards and to his right with a thoughtful expression. His right hand is raised, pointing his index finger upwards towards the top left of the frame.

*Tim Berners-Lee, 1999
Inventor of the Web*





A Web Semântica está
Realmente acontecendo no
MERCADO HOJE!

Vamos colocar a
mão na massa?



Por onde
começar?

Quero
saber
como!

Web
Semântica
aulas você deve
participar e
exercícios deve
fazer...

w3.org/standards/semanticweb/  

STANDARDS PARTICIPATE MEMBERSHIP ABOUT W3C 

Semantic Web 

SEMANTIC WEB

On this page → technology topics • news • upcoming events and talks 

In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

Linked Data

The Semantic Web is a Web of data — of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies allow you to embed data in documents (RDFa, GRDDL) or expose what you have in SQL databases, or make it available as RDF files.

Inference

Vocabularies

At times it may be important or valuable to organize data. Using OWL (to build vocabularies, or “ontologies”) and SKOS (for designing knowledge organization systems) it is possible to enrich data with additional meaning, which allows more people (and more machines) to do more with the data.

Vertical Applications

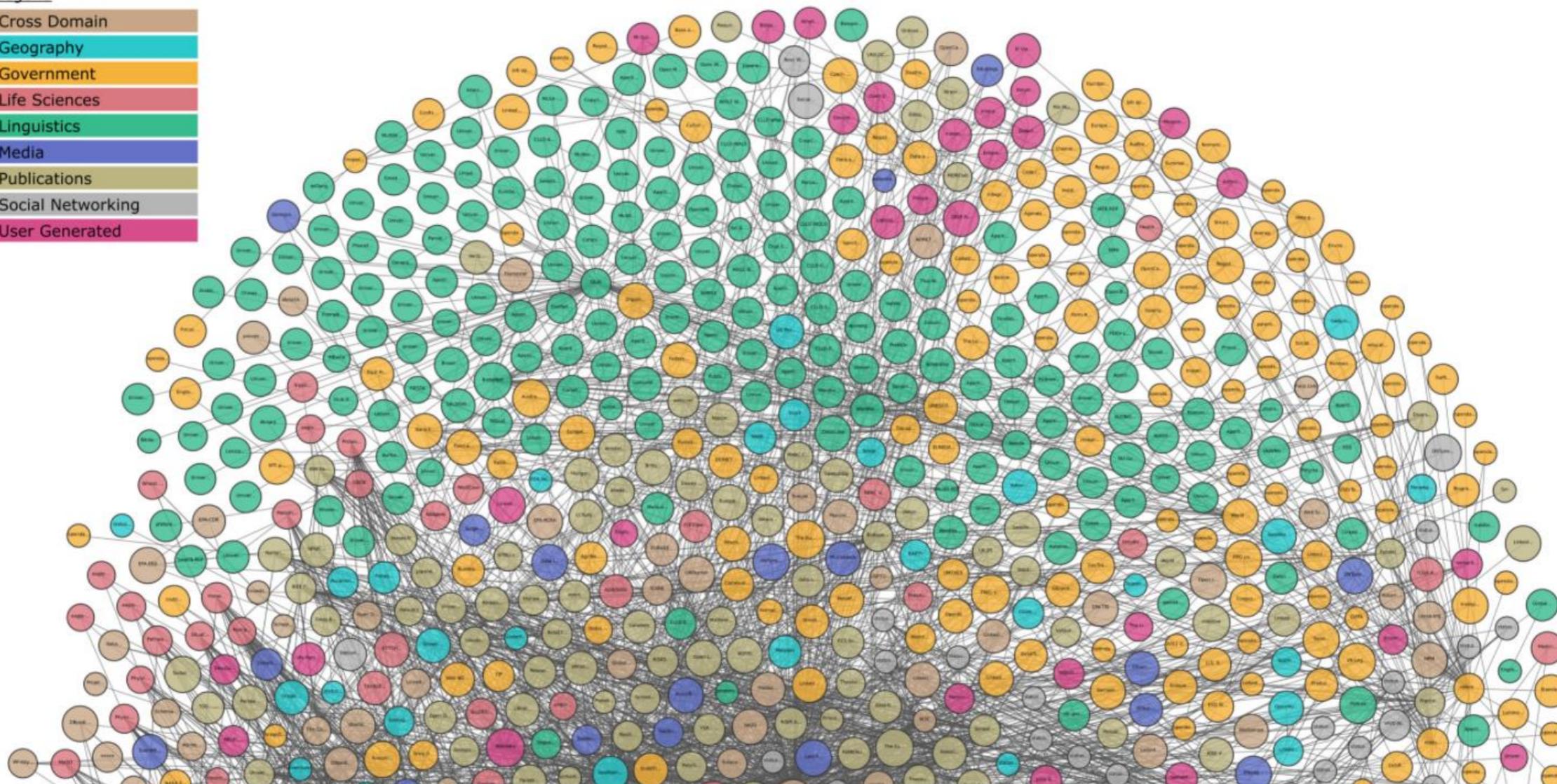
Query

Query languages go hand-in-hand with databases. If the Semantic Web is viewed as a global database, then it is easy to understand why one would need a query language for that data. SPARQL is the query language for the Semantic Web.

The Linked Open Data Cloud

Legend

- Cross Domain
 - Geography
 - Government
 - Life Sciences
 - Linguistics
 - Media
 - Publications
 - Social Networking
 - User Generated



[Page](#) [Discussion](#)[Read](#) [View source](#) [View history](#)[Search](#)

RDF

Resource Description Framework (RDF)



Overview

RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

- Publication date: 2014-02-25 (with a previous version published at: 2004-02-10)
- Created by: [RDF Working Group](#)
- List of documents at: <http://www.w3.org/standards/techs/rdf>

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This *graph view* is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.

Recommended Reading

The RDF 1.1 specification consists of a [suite of W3C Recommendations](#) and Working Group Notes, published in 2014. This suite also includes an [RDF Primer](#). See also Tim Berners-Lee's writings on [Web Design Issues](#), including [Metadata Architecture](#). Other technologies, like [OWL](#) or [SKOS](#), build on RDF and provide language for defining structured, Web-based ontologies which enable richer integration and interoperability of data among descriptive communities.

A number of textbooks have been published on RDF and on Semantic Web in general. Please, refer to a [separate page](#) listing some of those, as maintained by the community. That list also includes references to conference proceedings and article collections that might be of general interest.

Discussions on a possible next version of RDF

W3C has recently set up a new [RDF Working Group](#), whose charter is to make a minor revision of RDF.

[Main Page](#)
[Recent changes](#)
[Tools](#)
[Books](#)
[Validators](#)[Other W3C resources](#)
[Activity news](#)
[Publications](#)
[Logos, buttons](#)
[Activity home page](#)[W3C RSS feeds](#)
[Activity newsfeed](#)
[W3C blogs](#)
[Use cases, case studies](#)[Account request](#)
[W3C Member](#)
[Public](#)[Tools](#)
[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)



Are you looking for the New DBpedia Page? Please browse this way -> www.dbpedia.org

Data Download

Browse the DBpedia Datasets

[Go to Latest Release](#)



Apply

Explore current DBpedia
projects and



Develop

Develop amazing things
with our DBpedia



Research

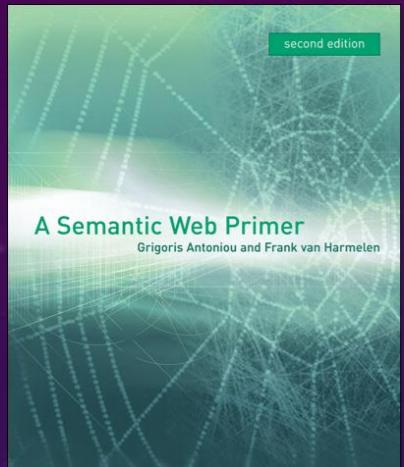
Join



Contribute

Subscribe to our Newsletter

Livros a serem estudos no módulo

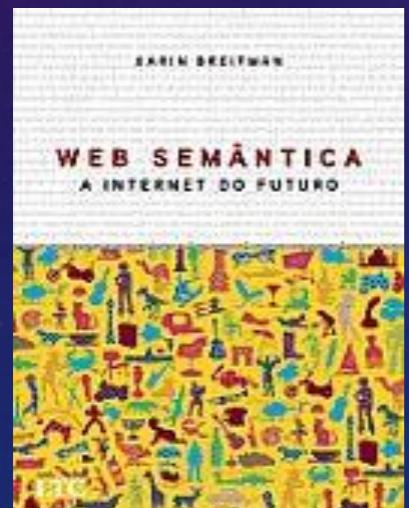


A Semantic Web Primer

Grigoris Antoniou e Frank van Harmelen
Segunda Edição

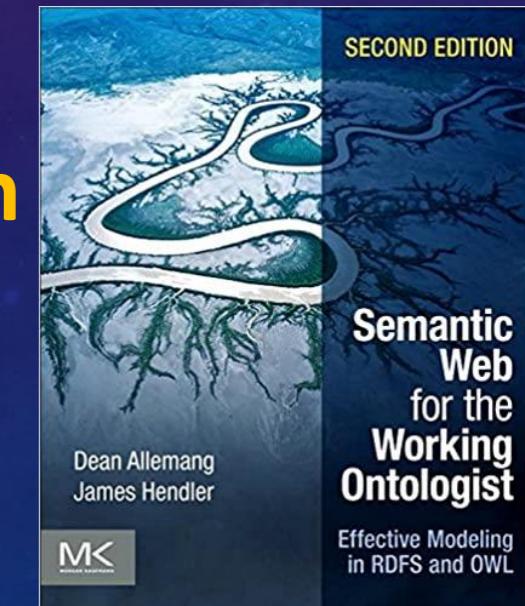
Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL

Dean Allemang , Jim Hendler
Editora MK



Web Semântica A Internet do Futuro

Karin Breitman
Editora LTC



RESUMO DE DADOS ABERTOS E WEB SEMÂNTICA

(RESUMO PARA O PPGI)

PROF. DR. SERGIO SERRA

SERGIOSERRA@GMAIL.COM

we are

OPEN DATA

Você pode copiar essas
transparências à vontade, parte das
transparências são
do autor, outras vieram de diversas fontes
da Web e de diversos colegas



AGENDA ...

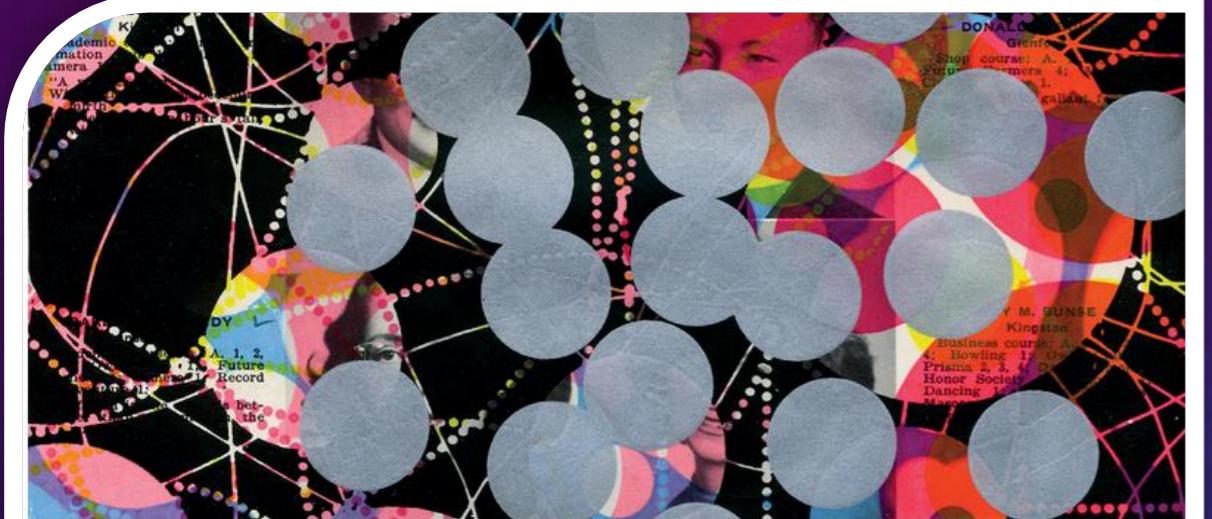
- Motivar, contextualizar e dar exemplos para entender o PROV
 - O que são Dados Abertos?
 - RDF, N3 e Turtle na Web Semântica



MOTIVAÇÃO



MOTIVAÇÃO



Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

SUMMARY SAVE SHARE COMMENT TEXT SIZE PRINT BUY COPIES



Trabalhar com grandes volumes de dados procedentes de diversas fontes e com diferentes formatos é uma das habilidades mais desejadas na última década.

Crescimento exponencial dos dados gerados pela sociedade e a necessidade de processar informações obtidas por meio da análise das conexões semânticas entre conceitos e relações presentes nestes dados.

MOTIVAÇÃO

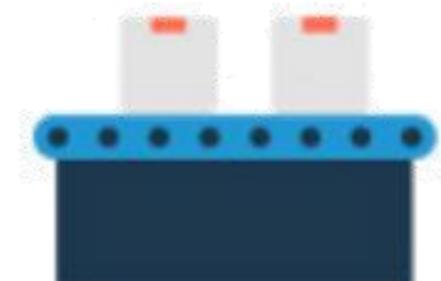
- A sociedade civil e empresas, governos, universidades e institutos de pesquisa realizam esforços para disponibilizar dados e produzir tecnologias Web criem uma **CULTURA DE PRODUÇÃO E CONSUMO DE CENTRADA EM DADOS** para agilizar a **descoberta de conhecimentos novos e agregar valor** aos dados disponibilizados na Web.
- Temos que acompanhar a nova economia!

NOVA ECONOMIA CENTRADA EM... DADOS

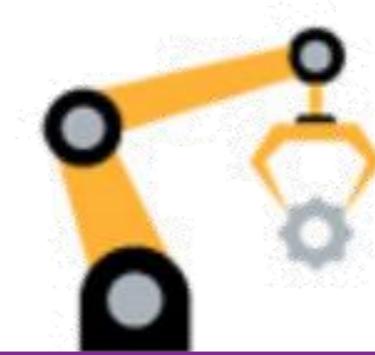
1780



1870



1970



2011



DADOS

1.0

2.0

3.0

4.0

Aprimoramento das máquinas a vapor, criação do tear mecânico.

Utilização do aço, energia elétrica, combustíveis derivados do petróleo.

Avanço da eletrônica, sistemas computadorizados e robóticos para manufatura.

Sistemas CiberFísicos, “Internet das coisas” e processos de manufaturas descentralizados

1- O que são
Dados Abertos?

WHAT
IS
OPEN
DATA
?

DADOS ABERTOS

Dados abertos (Open Definition, 2014)...

Podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição à fonte original e compartilhamento pelas mesmas licenças em que as informações foram apresentadas.

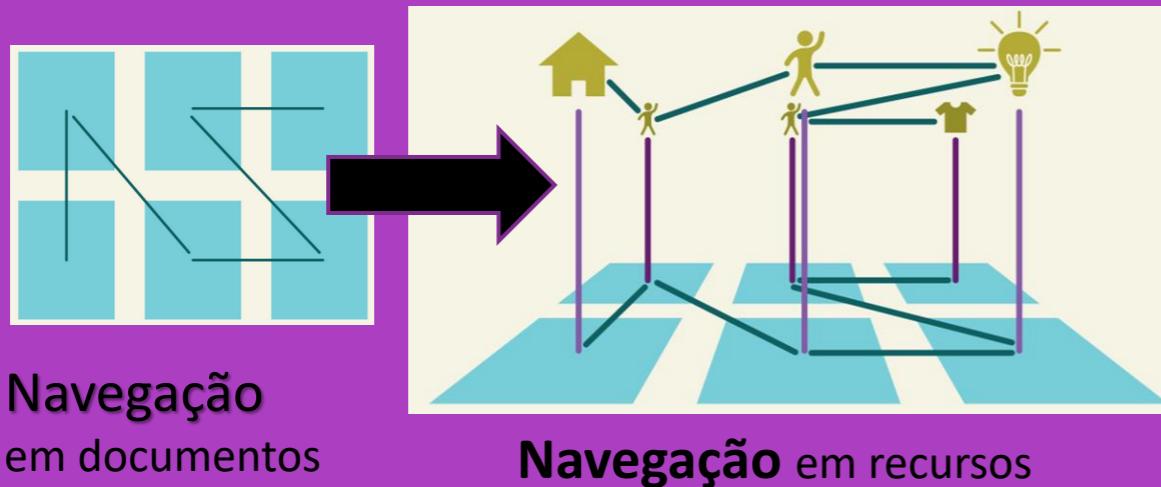
A abertura de dados busca **evitar um mecanismo de controle e restrições sobre os dados que forem publicados**, permitindo que tanto pessoas de qualquer natureza possam explorar estes dados de forma livre.

DADOS ABERTOS

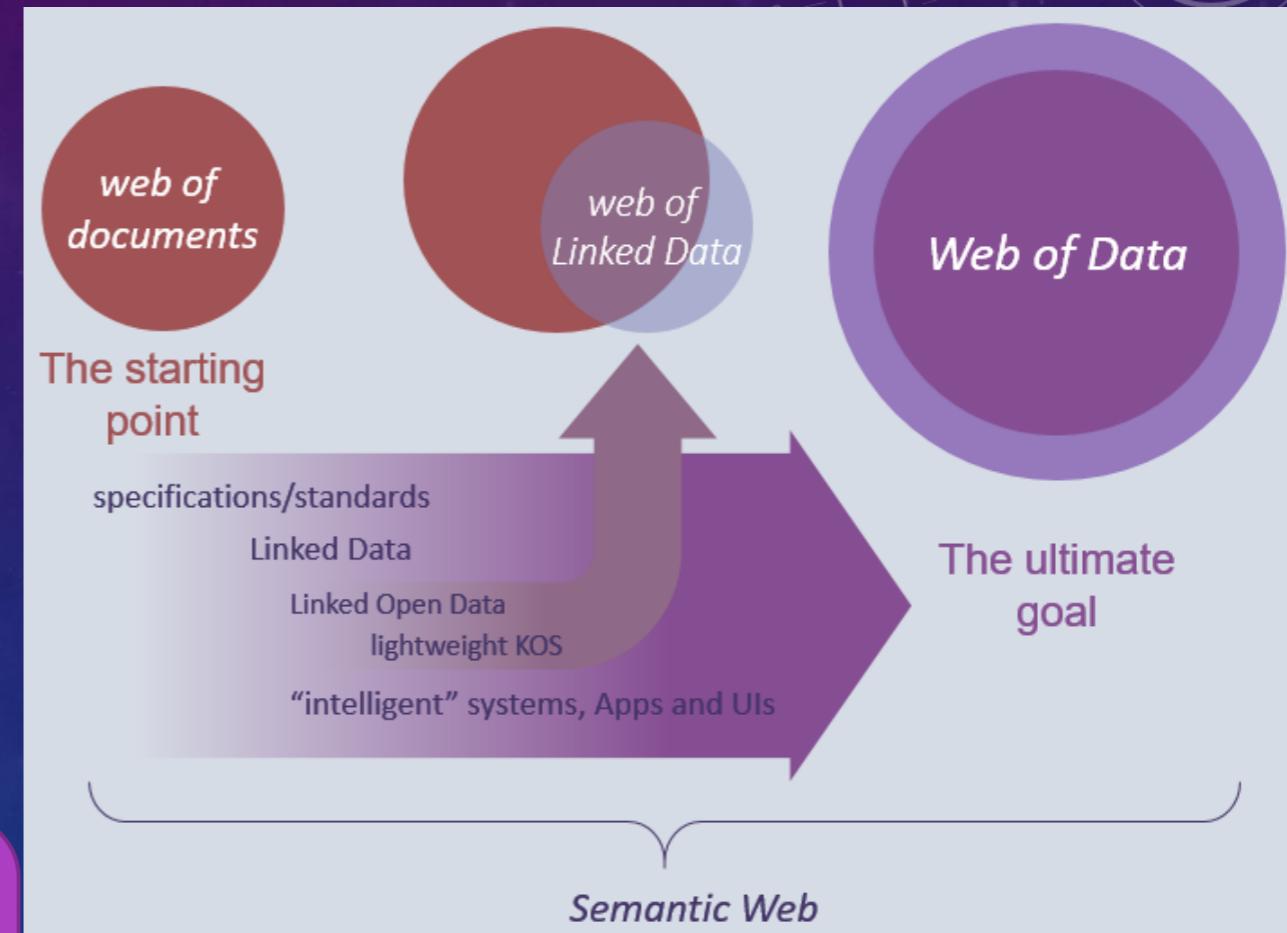
- Termo surgiu em ~1995 em um documento de uma agência científica americana, sobre divulgação de dados ambientais e geofísicos (Chignard, 2013).
- Possuem **3 normas fundamentais**



RESUMO: VISÃO DE TIM BERNERS LEE – WEB SEMÂNTICA

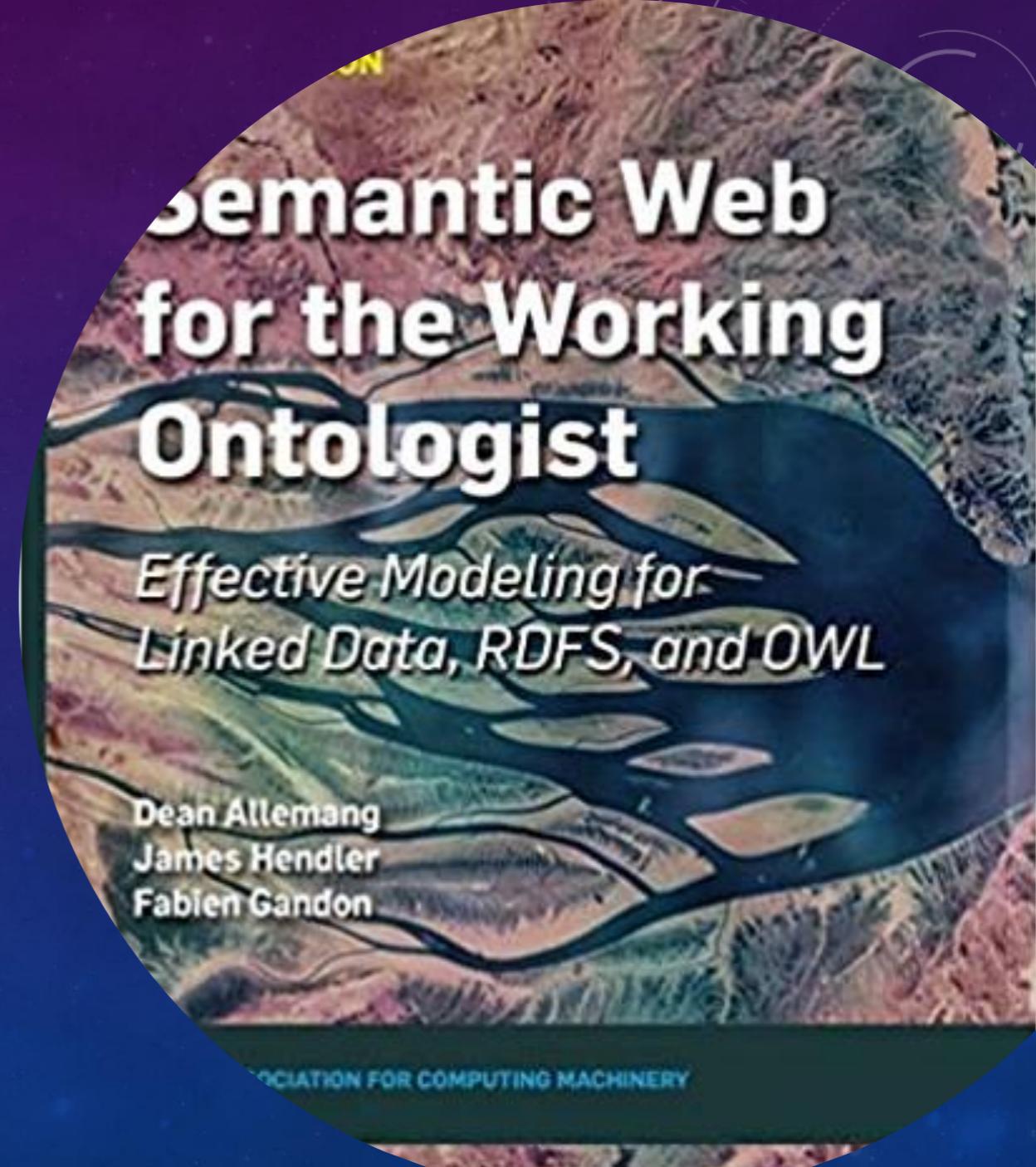


- O conhecimento deverá ser **semanticamente codificado** a partir de **representações logicamente estruturadas e conectadas** !
- Será possível desenvolver **regras de inferência** que direcionem raciocínios automáticos.



WEB SEMÂNTICA

- A Web Semântica é uma extensão da Web atual, aonde à informação é atribuído significado bem definido, permitindo que pessoas e computadores trabalhem em cooperação.
 - Os blocos básicos que definem a Web Semântica são:
 - um **modelo de dados** padrão;
 - um conjunto de **vocabulários** de referência;
 - um **protocolo** padrão de consulta.



WEB SEMÂNTICA NO DIAS DE HOJE

Pilha de Tecnologias da Web Semântica (não tão simples...)

A maioria das aplicações usa somente um subconjunto da pilha

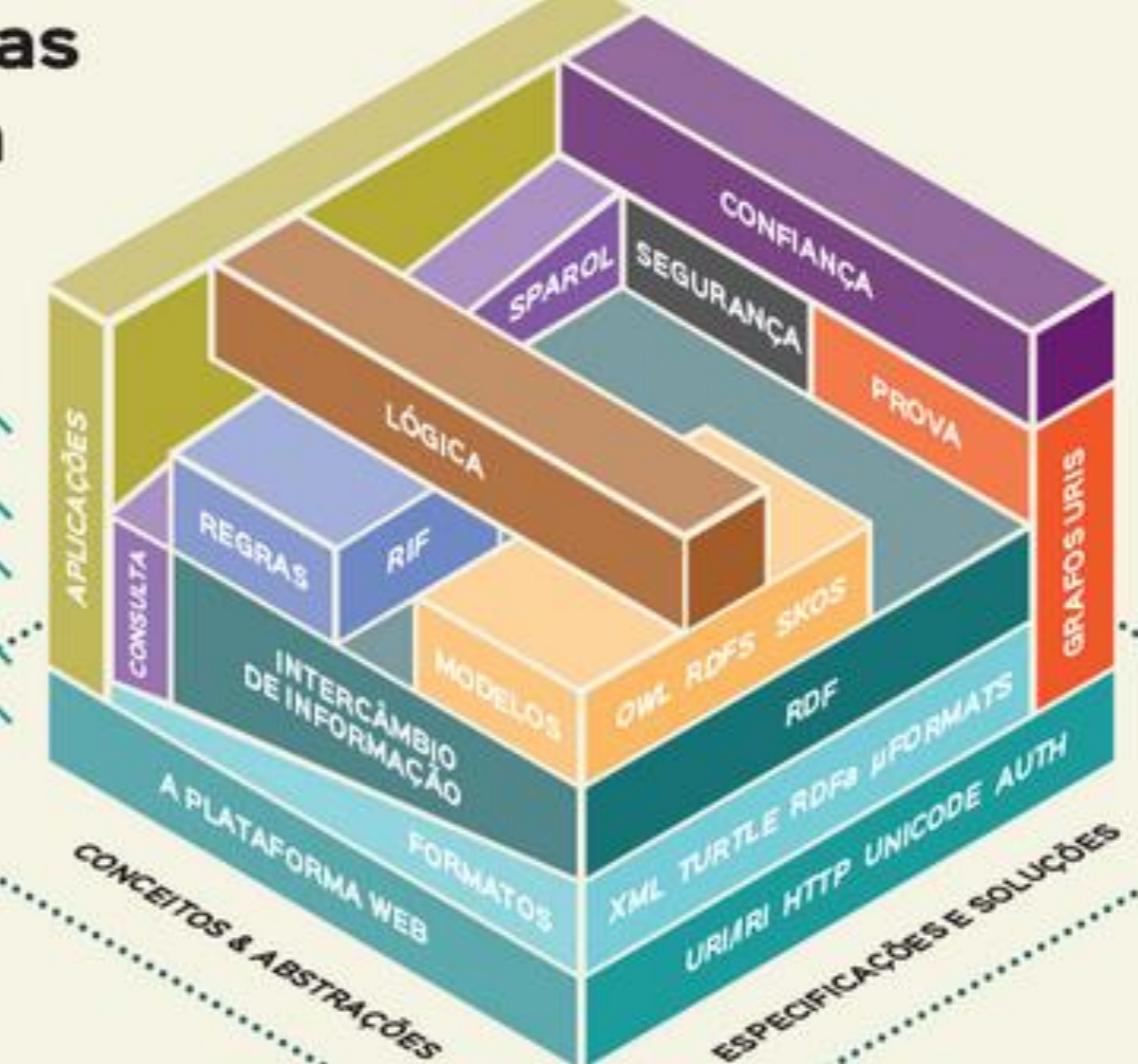
Consultas permitem o acesso a dados de baixa granularidade

Troca de informações padronizadas é a chave

Formatos são necessários, mas não tão importantes

A Web Semântica é baseada na Web

Dados Conectados usa um pequeno grupo de tecnologias.



QUE TAL LER?

GUIA DE WEB SEMÂNTICA

nic.br egibr Embaixada Britânica Brasília GREAT BRITAIN GOVERNO DO ESTADO SÃO PAULO



Editor: Steffen Staab
University of Koblenz-Landau
staab@uni-koblenz.de

The Semantic Web

The Semantic Web Revisited

Nigel Shadbolt and Wendy Hall, University of Southampton
Tim Berners-Lee, Massachusetts Institute of Technology



In the 50 years since the term AI was coined at the Dartmouth Conference, the digital world has evolved at a prodigious rate. It has produced an information infrastructure that few would have anticipated—with the possible exception of Vannevar Bush,¹ although even he might have thought the scale of achievement extraordinary. Today, the World Wide Web links 10 billion pages, and search engines can divine themes embodied in the links to serve useful and relevant content almost instantaneously.

Fifty years ago it might have appeared audacious to build a global web of information, to deploy semantics on such a scale, and to attempt inference over the resulting components. Fifty years ago, even if you could have explained it, a Semantic Web would have seemed as remote as general AI. Yet today we believe that the Semantic Web is attainable. We are seeing its first stirrings, and it will draw on some key insights, tools, and techniques derived from 50 years of AI research.

From documents to data and information

The original *Scientific American* article on the Semantic Web appeared in 2001.² It described the evolution of a Web that consisted largely of documents for humans to read to one that included data and information for computers to manipulate. The Semantic Web is a Web of actionable information—information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of “meaning” in which the logical connection of terms establishes interoperability between systems. This was not a new vision. Tim Berners-Lee articulated it at the very first World Wide Web Conference in 1994. This simple idea, however, remains largely unrealized.

A Web of data and information would look very different from the Web we experience today. It would routinely let us recruit the right data to a particular use context—for example, opening a calendar and seeing business meetings, travel arrangements, photographs, and financial transactions appropriately placed on a time line. The Sci-

entific American article assumed that this would be straightforward, but it’s still difficult to achieve in today’s Web.

The article included many scenarios in which intelligent agents and bots undertook tasks on behalf of their human or corporate owners. Of course, shopbots and auction bots abound on the Web, but these are essentially handcrafted for particular tasks; they have little ability to interact with heterogeneous data and information types. Because we haven’t yet delivered large-scale, agent-based mediation, some commentators argue that the Semantic Web has failed to deliver. We argue that agents can only flourish when standards are well established and that the Web standards for expressing shared meaning have progressed steadily over the past five years. Furthermore, we see the use of ontologies in the e-science community presaging ultimate success for the Semantic Web—just as the use of HTTP within the CERN particle physics community led to the revolutionary success of the original Web.

A growing need for data integration

Meanwhile, the need has increased for shared semantics and a web of data and information derived from it. One major driver—one that this magazine has reported on extensively—has been e-science (*IEEE Intelligent Systems*, special issue on e-science, Jan 2005). Life sciences research demands the integration of heterogeneous data sets that originate in separate communities of scientists in separate laboratories, and regulatory authorities demand the integration of complex pharmacogenomics, clinical drug trials, and epidemiological data. The challenge is how to find a way to integrate these components. This is being done in large part through the adoption of ontologies referred to as *ontologies*. In the argument in favor of using ontologies, numerous initiatives are developing ontologies (for example, see <http://obo.sourceforge.net>) for genomics, and related fields. These communities are defining language standards that can be deployed.

Many other disciplines are adopting web semantics. Environmental science is looking at ways to integrate data from hydrology, climatology, ecology,

WEB
SEMÂNTICA
UMA INTRODUÇÃO
(BEM GENTIL)...

PROF. DR. SÉRGIO SERRA
SERGIOSERRA@GMAIL.COM

UFRRJ

DADOS CONECTADOS (LINKED DATA OU LOD)

- A partir da Web Semântica derivou-se o conceito de dados conectados (LD)
- **Dados Conectados é um conjunto de boas práticas** para publicar e conectar de dados na Web, com o intuito de criar uma “Web de Dados” (Bizer, et al., 2006).
 - Boas práticas são apoiadas em tecnologias Web (HTTP, URI, etc..), objetivo de permitir a **leitura dos dados conectados semanticamente**, de forma automática por agentes de software.
 - A Web de Dados cria oportunidades para a **integração semântica dos dados**, motiva novos tipos de aplicações e ferramentas,
 - Dados Conectados também estão representados uma pequena parcela das tecnologias que compõem a *Web Semântica*.

DADOS CONECTADOS

Pilha de Tecnologias da Web Semântica (não tão simples...)

A maioria das aplicações usa somente um subconjunto da pilha

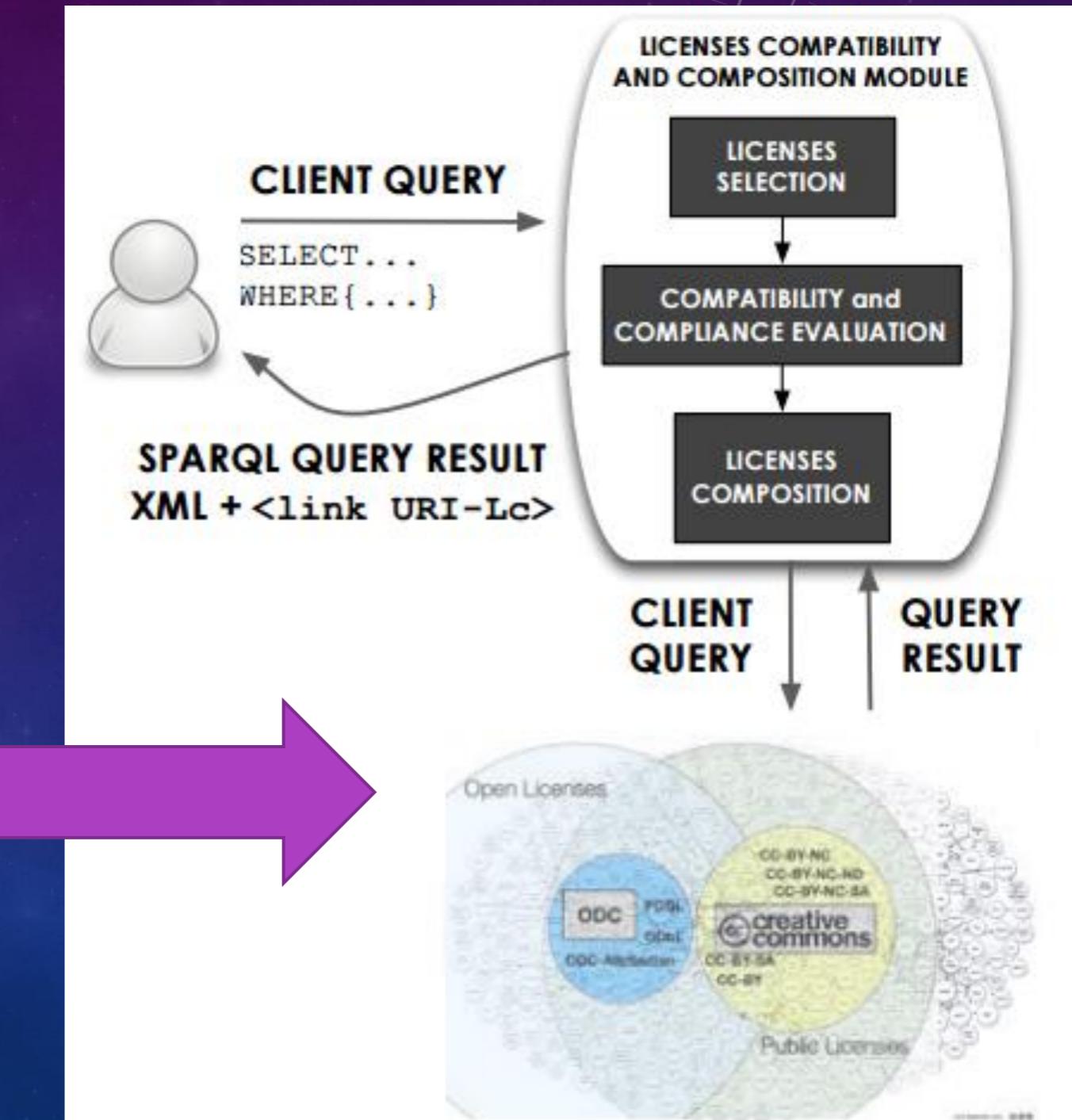
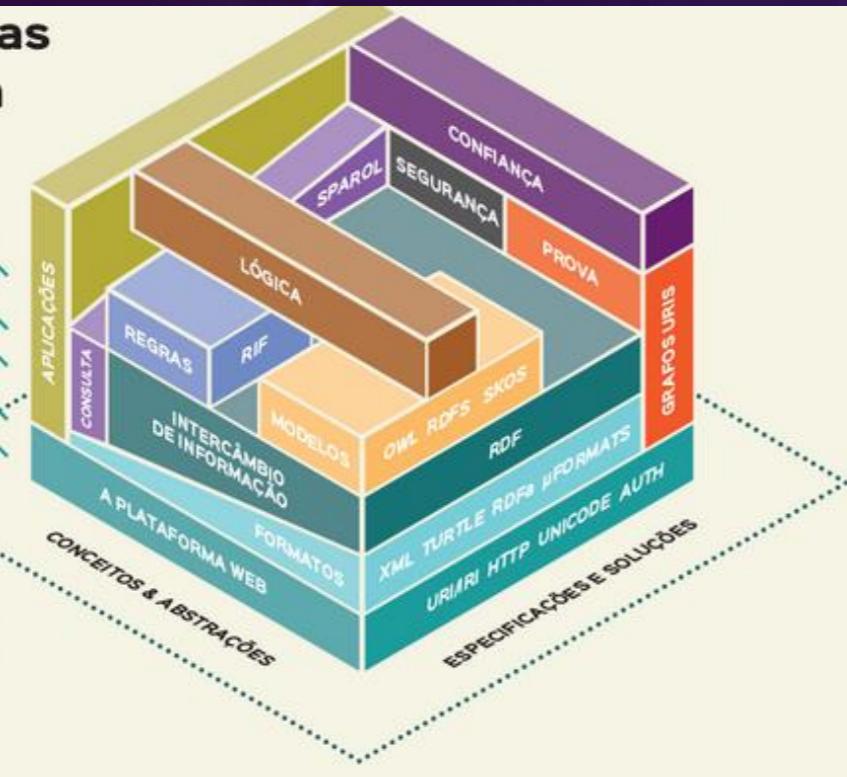
Consultas permitem o acesso a dados de baixa granularidade

Troca de informações padronizadas é a chave

Formatos são necessários, mas não tão importantes

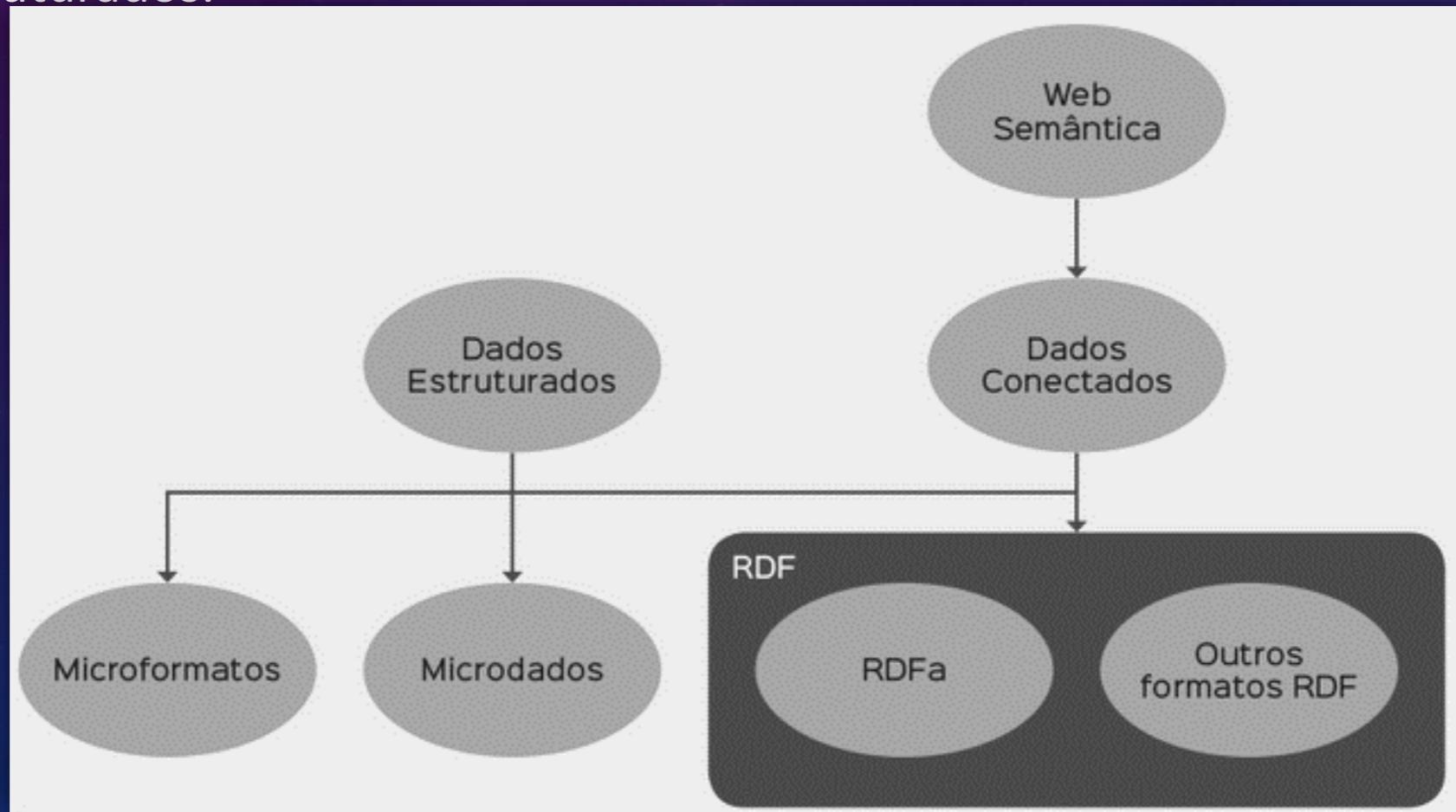
A Web Semântica é baseada na Web

Dados Conectados usa um pequeno grupo de tecnologias.



UM NOVO ECOSSISTEMA DE DADOS

- Relação entre Web Semântica, Dados Conectados, RDF e os diversos formatos de dados estruturados.



UM NOVO ECOSSISTEMA DE DADOS

- Padrões de Dados Conectados permitem que qualquer pessoa publique os dados de uma forma que pode ser lido por pessoas e processado por máquinas.
- Esta conexão de dados permite que todos (homens e máquinas) possam trabalhar conjuntamente de forma mais eficiente.
- Dados Conectados podem proporcionar, desde **negócios até a melhoria do sistema público (governos)**.
- Dados Conectados podem formar novos conhecimentos!

UM NOVO ECOSSISTEMA DE DADOS

- Exemplos da economia do mundo real:
 - Best Buy, melhorou entre 15% a 30% no número de clicks via buscador Google através da utilização do formato de serialização de Dados Conectados (RDFa).
 - Google anunciou a utilização do formato de serialização JSON-LD para o Gmail;
 - IBM anunciou que o DB2 se tornará um servidor de Dados Conectados;
 - Facebook expôs os Dados Conectados via Graph API;
 - Governo Britânico disponibiliza vários de suas fontes de dados em formato RDF (Data.gov, 2014).
 -

PADRÕES DE REPRESENTAÇÃO DE DADOS ABERTOS

- A preparação dos dados para que passem de Dados Abertos para Dados Abertos Conectados se dá por meio de uma série de **transformações intermediárias...**



RDF

- O RDF (*Resource Description Framework*) é uma linguagem de representação de informação na Web, **recursos podem ser descritos formalmente e sejam acessíveis por máquinas.**
 - O objetivo do RDF é prover uma representação minimalista do conhecimento da Web.
 - A versão RDF 1.1 (Fev 2014) estendeu a versão 1.0 (2004) proposta em 2004.
 - Disponível em <http://www.w3.org/TR/rdf11-new/>

W3C Working Group Note



What's New in RDF 1.1

[W3C Working Group Note 25 February 2014](http://www.w3.org/TR/2014/NOTE-rdf11-new-20140225/)

This version:

<http://www.w3.org/TR/2014/NOTE-rdf11-new-20140225/>

Latest published version:

<http://www.w3.org/TR/rdf11-new/>

Previous version:

<http://www.w3.org/TR/2013/WD-rdf11-new-20131217/>

Editor:

[David Wood, 3 Round Stones Inc.](#)

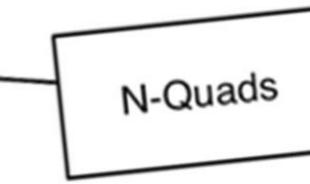
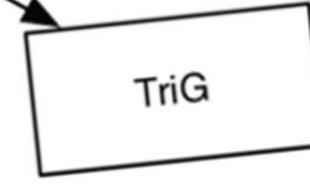
PRINCIPAIS DIFERENÇAS RDF 1.0 X 1.1

RDF 1.1

RDF 1.0



Supports Multiple Graphs



extended by

similar to

RDF

- O modelo RDF é que ele foi **projeto para descrever recursos na Web.**
 - É importante lembrar que a Web é um espaço de informação no qual os **itens de interesse (recursos) precisam ser identificados.**
 - Cada recurso possui um **URI (identificador único e global)** para que o mesmo possa ser identificado na WWW.
- RFC 2396 descreve os URIs
- A arquitetura da Web Semântica é composta por três bases fundamentais



RDF – 1- RECURSOS ÚNICOS

- O **identificador URI** provê uma maneira simples e única de identificar recursos na Web, é caracterizado por três aspectos:
 - a) **Uniformidade**: utilização de recursos tanto no mesmo contexto quanto em contextos diferenciados;
 - b) **Recurso**: qualquer coisa que pode ser identificado por um URI (vídeo, imagem, serviço, documento, etc);
 - c) **Identificador**: informação requerida para identificar e diferenciar um determinado recurso de qualquer outro.

RDF – 1- RECURSOS ÚNICO

- Um URI pode ser classificado como:
 - **URL (*Uniform Resource Locator*)**, define um localizador/endereço para um determinado recurso através de um protocolo existente
 - **URN (*Unified Resource Name*)** representa um nome para um determinado recurso, garantindo unicidade e persistência de forma global mesmo quando o recurso não está disponível.
 - **IRI (*International Resource Identifier*)** é uma generalização do URI. Diferente do URI (baseado nos caracteres ASCII) o IRI amplia o número de caracteres, permitindo que caracteres (Chineses (*kanji*) e Japoneses (*hiragana e katakana*), Cirílicos e Coreanos).

RDF - 2- INTERAÇÃO

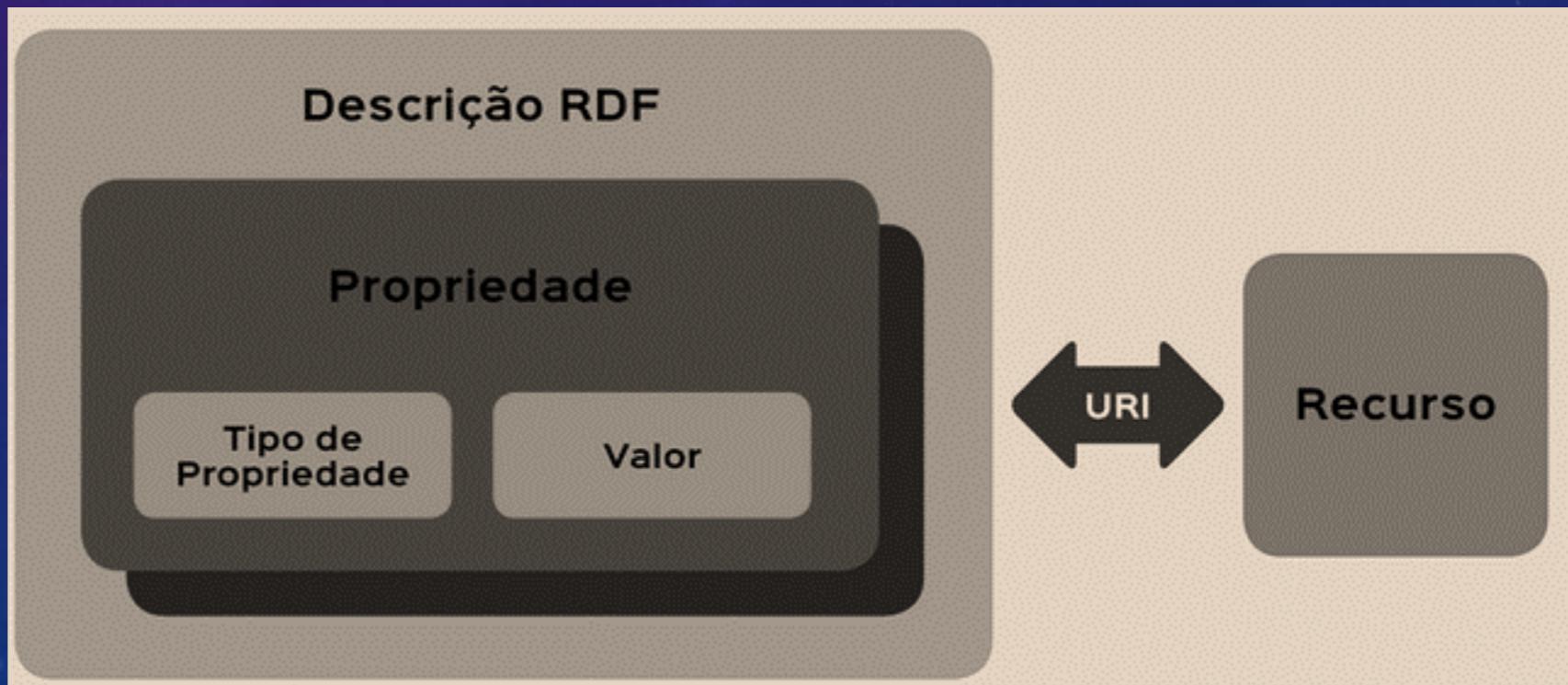
- A Web possui uma arquitetura cliente-servidor (.... e muito mais)
- A comunicação acontece através de **protocolos padrões** que permitem a troca de mensagens entre um Servidor que implementa este protocolo e um navegador (cliente) que envia a solicitação para o servidor Web.
 - O Protocolo padrão utilizado na Web é o HTTP (*Hypertext Transfer Protocol*)
(já visto em outras disciplinas da graduação)
 - Por exemplo, no navegador, ao digitar o URI `http://r1.ufrrj.br/petsi`, o navegador enviará uma requisição do tipo HTTP GET para o Servidor web da UFRRJ, ele retornará uma mensagem contendo a representação do recurso.

RDF – 3 - FORMATOS

- Na comunicação cliente-servidor, o servidor irá retornar para o cliente uma representação em um determinado formato.
- Cada formato de representação retornado para o cliente conterá informação de **metadados e dados**.
 - Neste caso, os metadados são os campos de cabeçalhos utilizados para identificar o formato de representação.
- No RDF temos **diferentes formas de representação/serialização**. Por exemplo,
 - **Turtle** (que possui *Content-Type: text/turtle*)
 - **JSON-LD** (que possui *Content-Type: application/ld+json*).
 - Outros.. Trig, NQuad

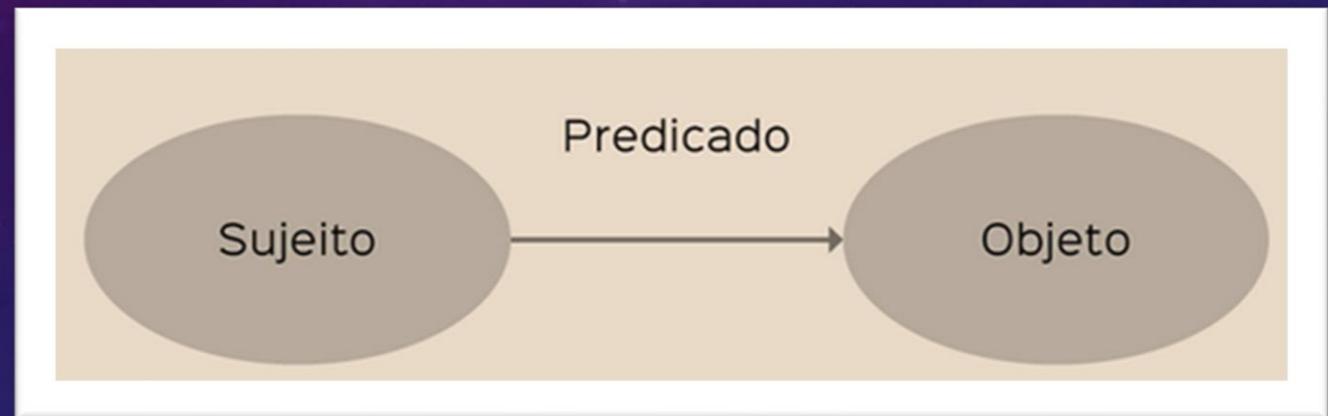
REPRESENTAÇÃO ABSTRATA DE UM RDF

- Além descrever os recursos disponíveis na Web (**Descrição sintática**), o RDF também oferece a possibilidade de descrever a relação e significados entre diversos recursos (**Descrição semântica**). **Graficamente um RDF é isso ai....**



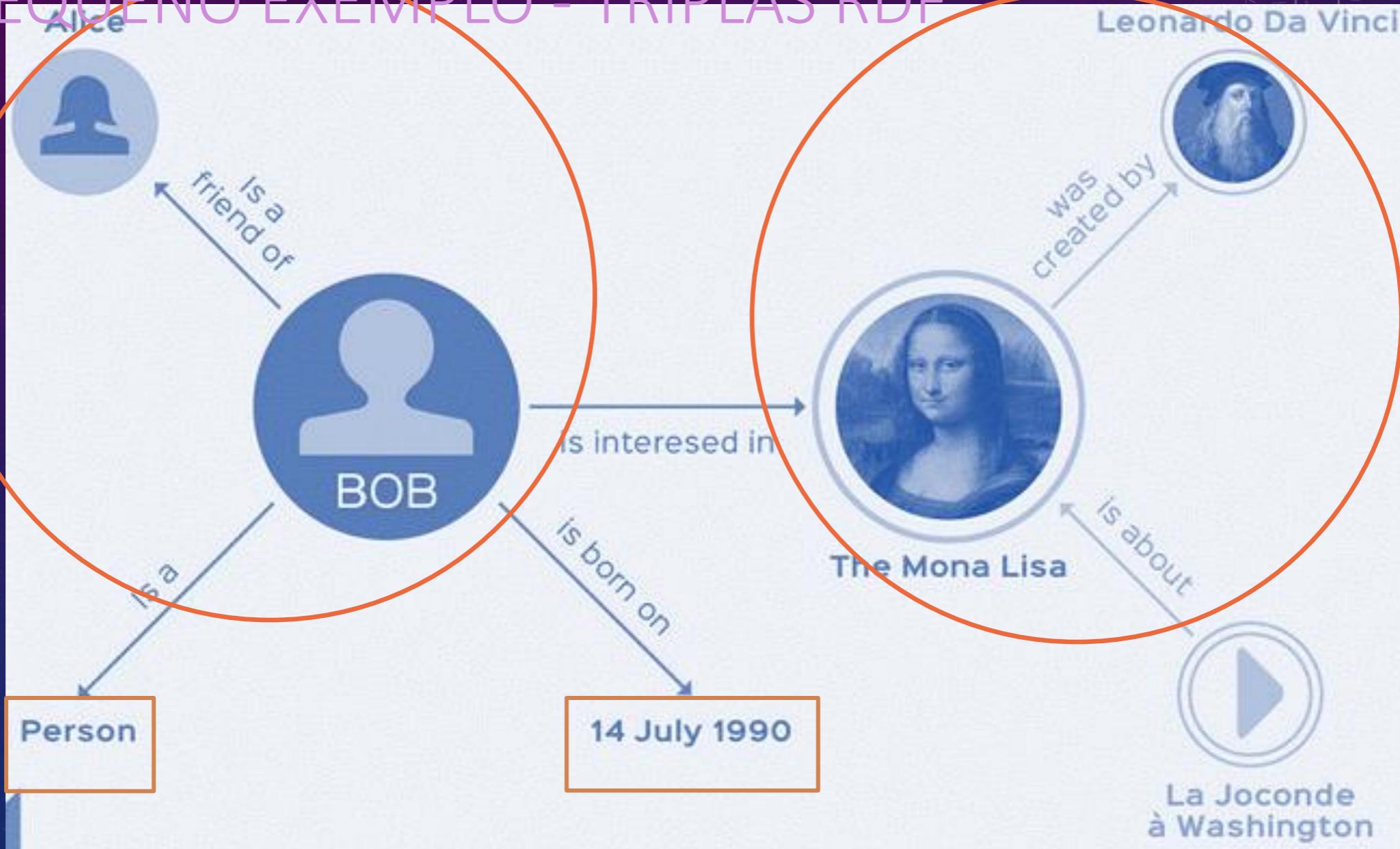
TRIPLAS RDF

- RDF oferece recursos para descrição de triplas composta por uma estrutura do tipo **<sujeito> <predicado> <objeto>** (Cyganiak, et al., 2014).



- O conjunto destas estruturas em uma **tripla** é chamado de Grafo RDF (DAG - grafo dirigido e acíclico).
 - Cada tripla é representada como nó-arco-nó. O grafo RDF normalmente possui múltiplas triplas que irão compor um documento RDF.

UM PEQUENO EXEMPLO - TRIPLAS RDF



UM PEQUENO EXEMPLO - TRIPLAS RDF

- 6 triplas que representam a relação entre conceitos/recursos
 - **<Bob> <is a friend of> <Alice>**
 - <Bob> representa um sujeito, que é um recurso, <Alice> representa o objeto que é outro recurso e <*is a friend of*> representa uma propriedade que relaciona <Bob> à <Alice>.
 - **<A Mona Lisa> <was created by> <Leonardo Da Vinci>**
 - *representa a relação entre o sujeito <A Mona Lisa> e seu criador, o objeto <Leonardo Da Vinci>, por meio da relação/propriedade <was created by>.*

EXEMPLO - TRIPLAS RDF

- A **riqueza do RDF** permite que um computador possa interpretar os dados representados por estas triplas.
- Para que isso seja realizado precisamos garantir que cada um dos elementos do grafo sejam representados e referenciados de maneira única.



IRI- IDENTIFICADOR DE RECURSO INTERNACIONAL

- IRI identifica de maneira única um recurso na Web.
- Como um RDF é um Framework, onde todo recurso deve possuir um URI/IRI, o IRI pode representar **<sujeito>**, **<predicado>** ou **<objeto>**.
- EXEMPLO
 - Na tripla **<Bob> <*is a friend of*> <Alice>** poderia ser representado utilizando IRIs como:
 - **Sujeito:** <<http://example.com/Bob>>
 - **Predicado:** <http://example.com/is_a_friend_of>
 - **Objeto:** <<http://example.com/Alice>>

IRI- IDENTIFICADOR DE RECURSO INTERNACIONAL

- O mesmo poderia ocorrer para os diversos recursos que vimos no exemplo anterior.
 - Por exemplo, de acordo com a Tabela abaixo, podemos ver o reuso de **três vocabulários**, sendo eles presentes na **FOAF**, **DBpedia** e a **própria especificação do RDF**.

Recurso da Figura 2.9	IRI (que podem ser reusados)
<Person>	http://xmlns.com/foaf/0.1/Person
<is a>	http://www.w3.org/1999/02/22-rdf-syntax-ns#type
<The Mona Lisa>	http://dbpedia.org/page/Mona_Lisa
<Leonardo da Vinci>	http://dbpedia.org/page/Leonardo_da_Vinci
<is about>	http://www.w3.org/1999/02/22-rdf-syntax-ns#about

- RDF permite utilizar **literais e tipos de dados** para descrever uma informação.

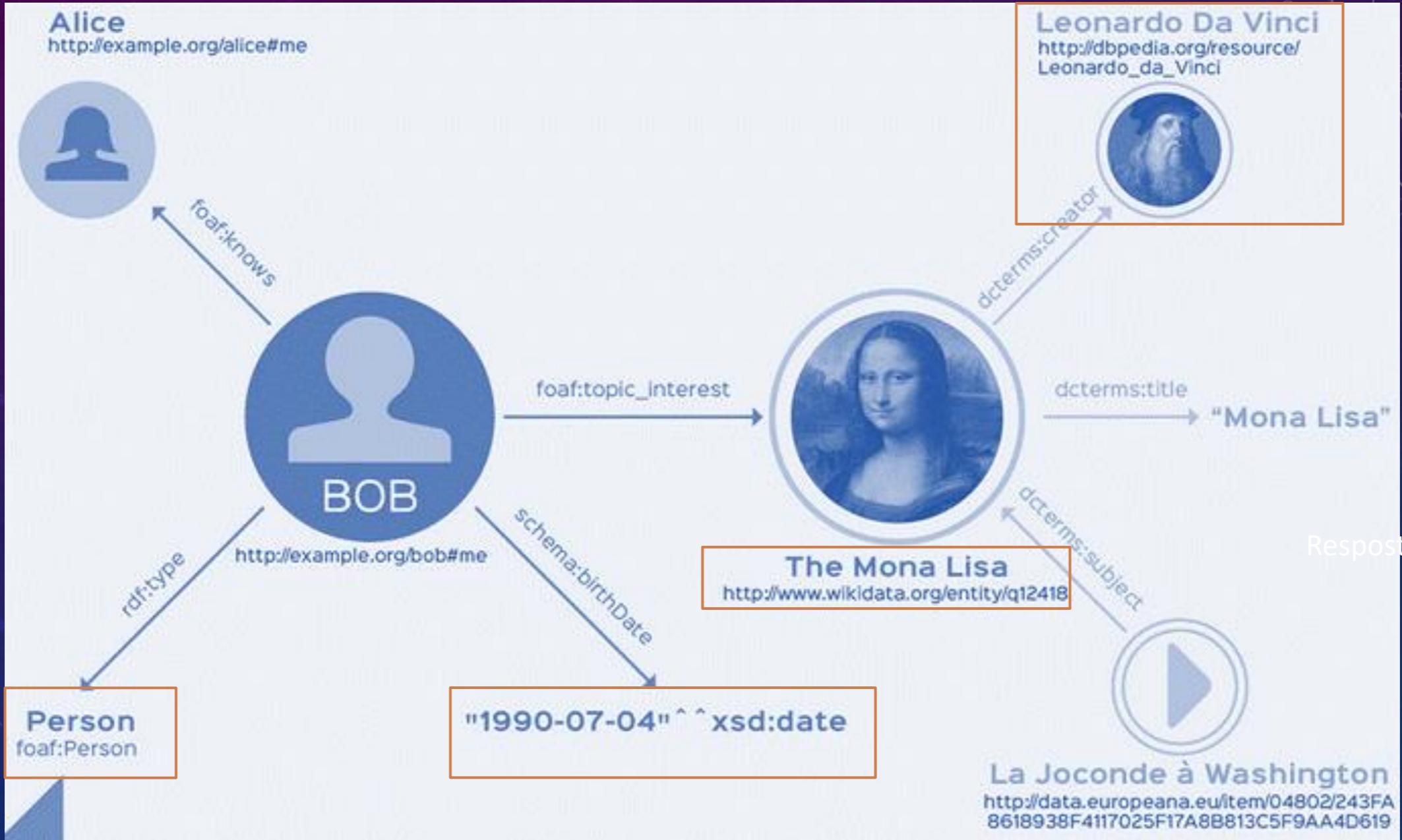
RDF – LITERAIS E TIPOS

- Os **literais** (usados em conjunto com os tipos de dados) podem ser compreendidos como todos os valores identificados num grafo RDF que não possuem um IRI associado.
 - No exemplo, um literal é a data “*14 July 1990*”. Este tipo de literal é um tipo de dado *date*.
 - Podemos associar a expressão “*La Joconde à Washington*” a um literal que representa um texto e seu tipo de dado é *string*.
- Uma das vantagens incorporadas com o IRI é a ampliação dos caracteres além da Tabela ASCII (podemos definir que “*La Joconde à Washington*” é um literal relacionado a um vídeo sobre a Monalisa descrito no idioma Francês (i.e. “fr”), enquanto ワシントンでモナリザ é um literal relacionado ao mesmo vídeo, porém utilizando o idioma Japonês (ex. “jp”).
- Os **literais estão associados a tipos de dados**. Como o RDF foi construído como uma camada acima do XML, todos os tipos de dados disponíveis em XML podem ser reusados em RDF.



NOTA: Os **literais só podem ser aplicados a objetos**, ou seja, nunca podem descrever sujeitos ou predicados.

TRIPLAS RDF E LITERAIS.. QUAIS A NOVIDADES?



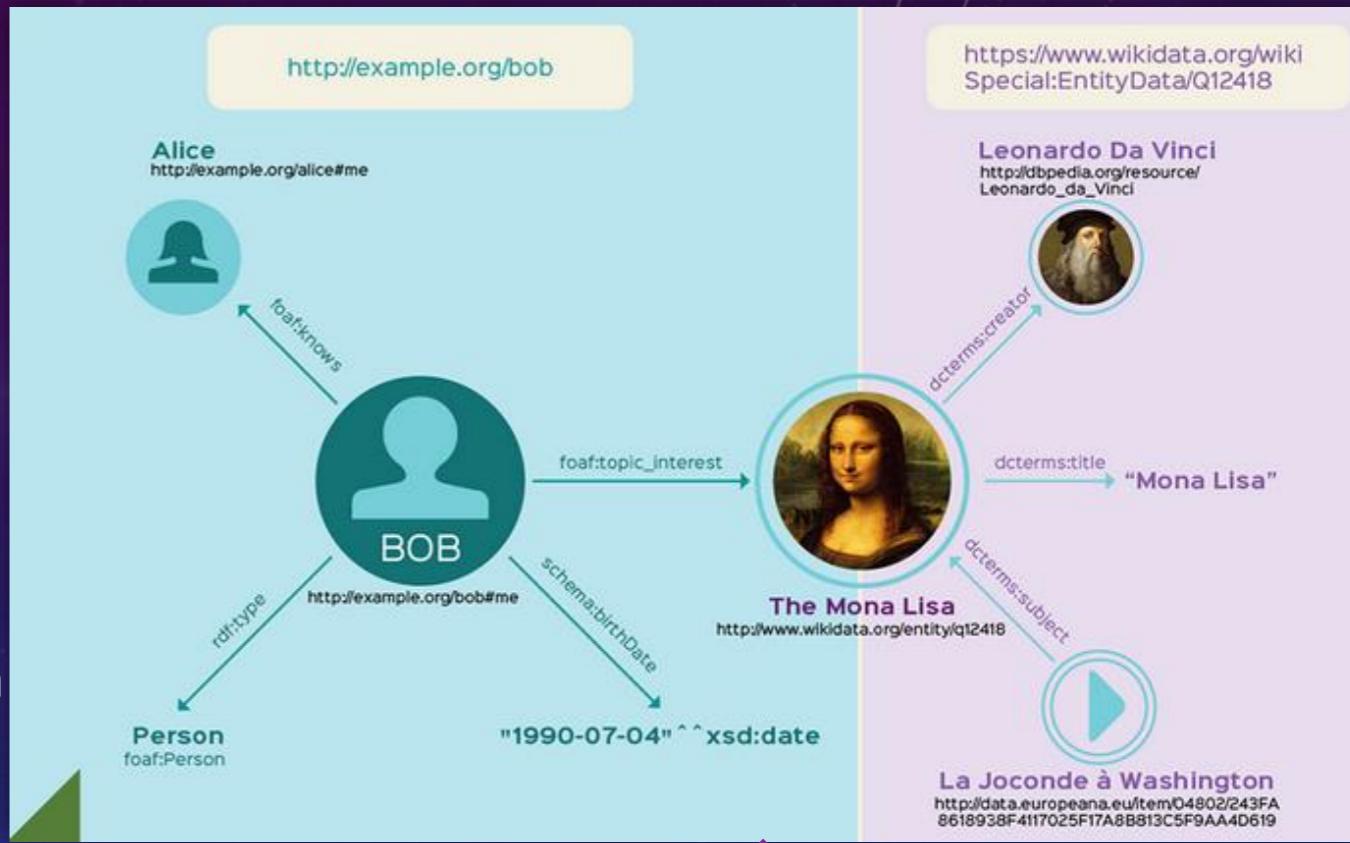
TRIPLAS RDF E LITERAIS... REPOSTAS

- Como resultado do reuso de **diversos vocabulários**, bem como literais, temos um **novo grafo**
 - Observe que em “La Joconde à Washington” foi reutilizado um recurso do catálogo de dados publicado pela **Europeana**, enquanto que o literal de data permaneceu.
 - O literal de data para representar a **data de nascimento de Bob**, apareceu após a data “`^^xsd:date`”. O delimitador “`^`” caracteriza que está sendo descrito um literal, enquanto que o `xsd` é o alias (pseudônimo) criado que faz referência ao *XML Schema*.
 - Finalmente, “date” equivale ao tipo de dado que está sendo definido.
- Outro vocabulário reusado foi o **Schema.org**, provê uma série de esquemas e tem sido usado por *engines* de buscas de empresas como Google, Microsoft, Yahoo! entre outros.



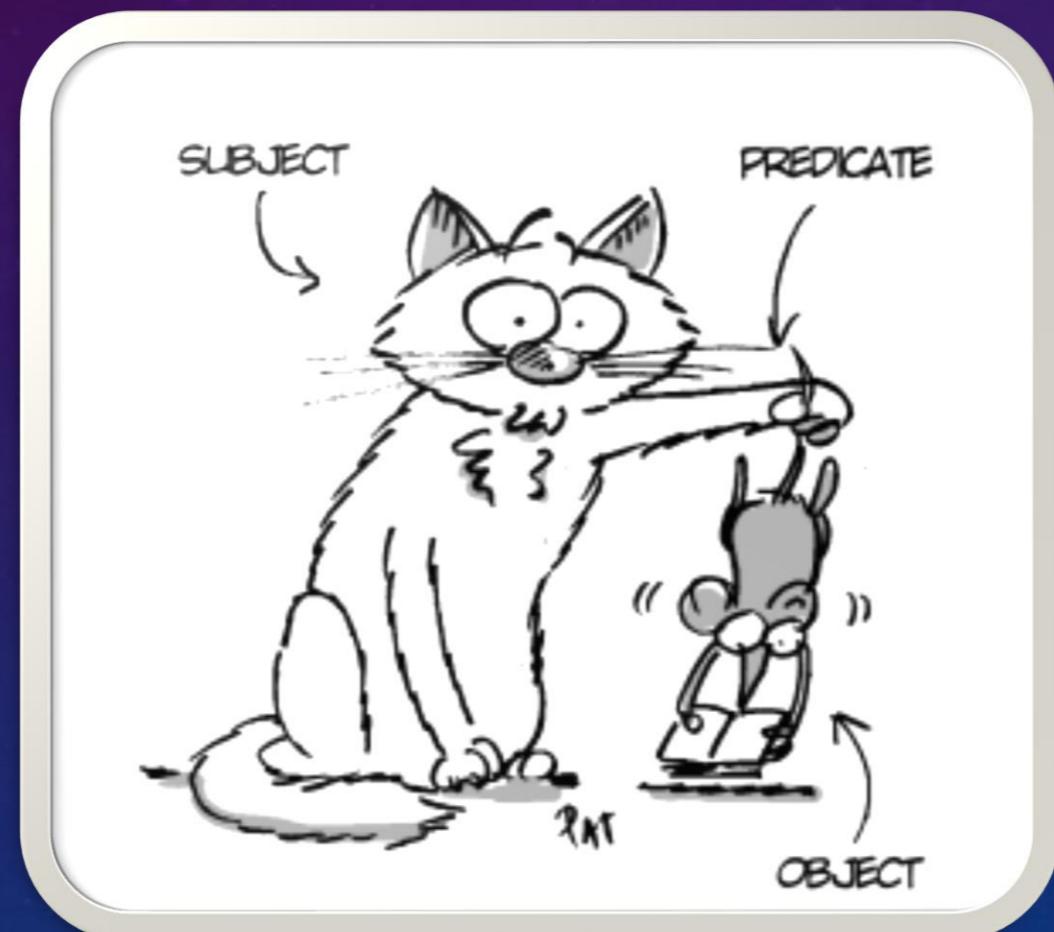
MÚLTIPLOS GRAFOS

- Um RDF pode conter outros grafos conectados ao grafo original (formando *datasets*) conectados.
- Múltiplos grafos podem ser acessados por um único IRI.
 - Todos os dados da Mona Lisa são reutilizados, o Grafo da Mona Lisa pode ser acessado pelo IRI único, disponível em Wikidata,
<https://www.wikidata.org/wiki/Special:EntityData/Q12418>.
 - Temos dois grafos conectados, um para descrever Bob (<http://example.org/bob>) e outro para Mona Lisa (Schreiber, et al., 2014).



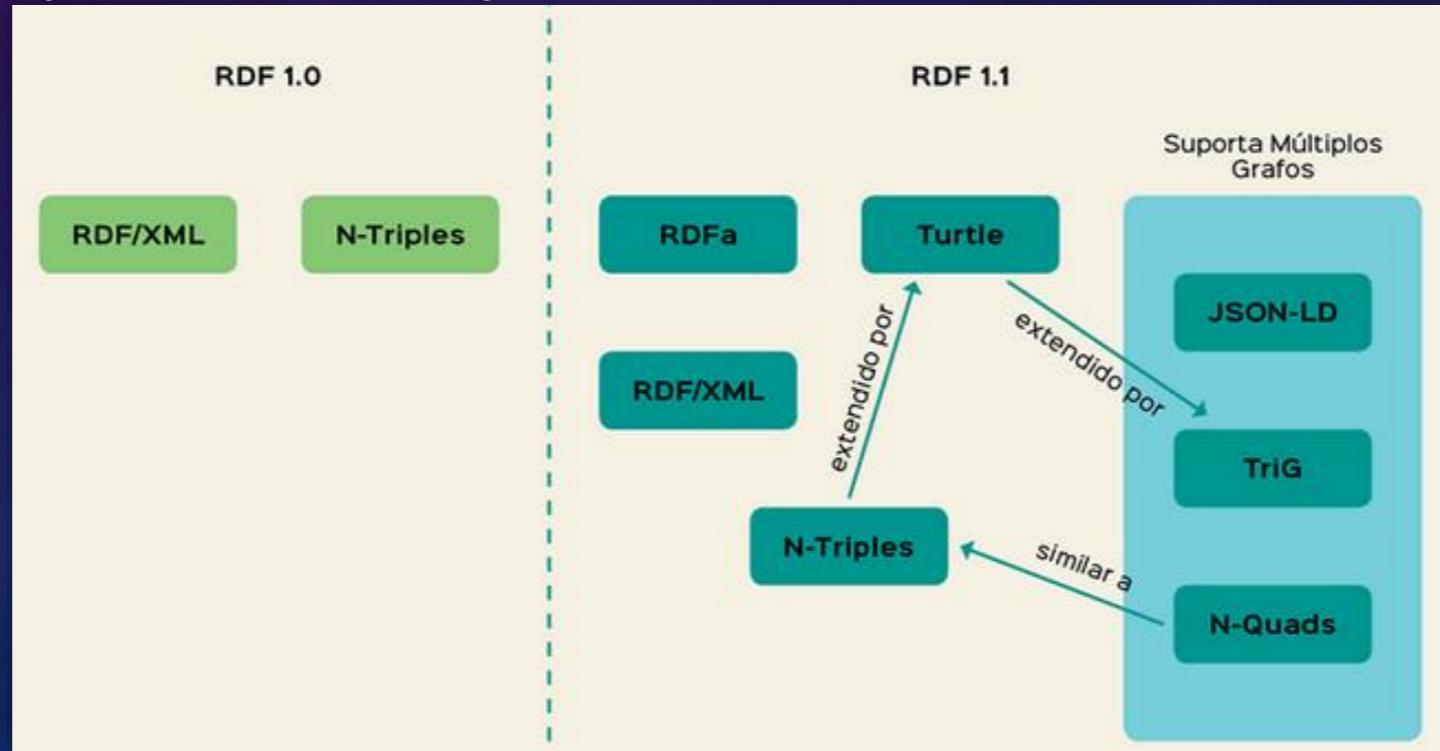
Facilita a reutilização de grandes quantidades de dados em grafos RDF sem a necessidade de ter o conhecimento completo dos mesmos. Facilita a extensão de grafos!!!

EXISTEM DIVERSAS REPRESENTAÇÕES DE DADOS E
DADOS ABERTOS CONECTADOS EM RDF...



SERIALIZAÇÃO EM RDF 1.1

- Serialização (de modo geral e simplificado) é uma forma de representação dos dados.
- No RDF 1.1, o número de formatos de serialização foi bastante ampliado em relação ao 1.0 (Wood, 2014).

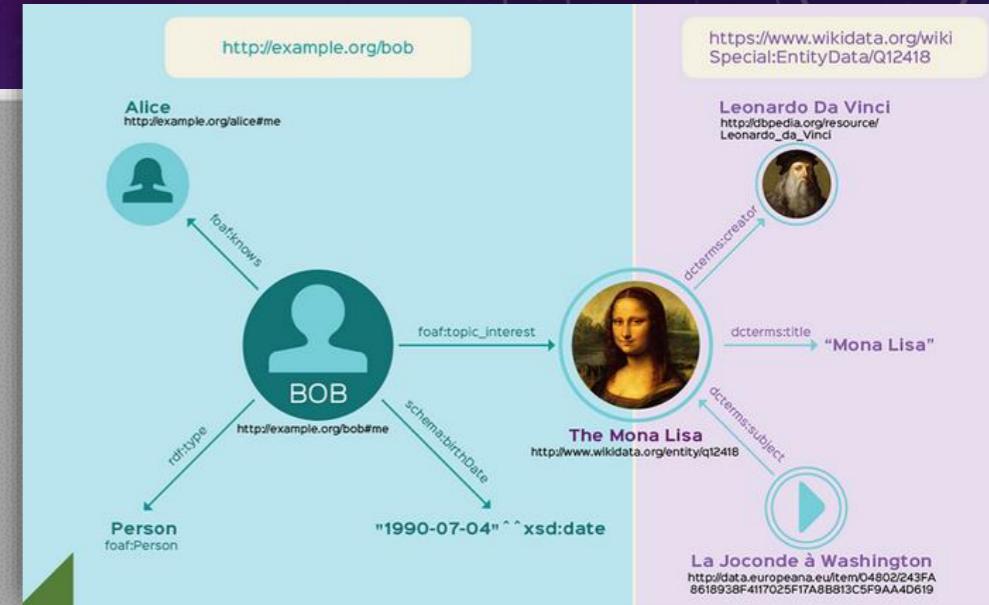


RDFa (*RESOURCE DESCRIPTION FRAMEWORK IN ATTRIBUTES*)

- Tem por objetivo embutir código RDF em estruturas HTML e XML (Herman, et al., 2013).
- Inclusão de significado via atributos dos elementos (RDFa é considerada o RDF em Atributos)
- **Vantagem do RDFa é que máquinas de buscas melhoram seus resultados aumentando a precisão sobre o real significado de um documento.**
 - Máquinas de buscas podem agregar os dados de um documento com dados de outro documento, enriquecendo assim os resultados de buscas.

SERIALIZAÇÃO EM RDFA (VAMOS ESTUDAR O EXEMPLO....)

```
01 <body prefix="foaf: http://xmlns.com/fcaf/0.1/
02           schema: http://schema.org/
03           dcterms: http://purl.org/dc/terms/">
04   <div resource="http://example.org/bob#me" typeof="foaf:Person">
05     <p>
06       Bob knows <a property="foaf:knows" href="http://example.org/alice#me">Alice</a>
07       and was born on the <time property="schema:birthDate">1990-07-04</time>.
08     </p>
09     <p>
10       Bob is interested in <span property="foaf:topic_interest"
11         resource="http://www.wikidata.org/entity/Q12418">the Mona Lisa</span>.
12     </p>
13   </div>
14   <div resource="http://www.wikidata.org/entity/Q12418">
15     <p>
16       The <span property="dcterms:title">Mona Lisa</span> was painted by
17       <a property="dcterms:creator" href="http://dbpedia.org/resource/Leonardo_da_Vinci">Leonardo da Vinci</a>
18       and is the subject of the video
19       <a href="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">'La Joconde à Washington'</a>
20     </p>
21   </div>
22   <div resource="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
23     <link property="dcterms:subject" href="http://www.wikidata.org/entity/Q12418"/>
24   </div>
25 </body>
```



SERIALIZAÇÃO EM RDFA (VAMOS ESTUDAR O EXEMPLO....)

- No código HTML do slide anterior aparecem quatro atributos:
 - **prefix**- descreve os vocabulários reusados no documento HTML. Haverá reuso de **3 vocabulários** (FOAF, Schema.org e Dublin Core);
 - **resource** - descreve um recurso (conforme foi explicado no RDF-S) para a classe *rdfs:Resource*. Ex. na linha 4 há a descrição do recurso *http://example.org/bob#me*;
 - **property** - descreve uma propriedade. O atributo *property* é similar à propriedade *rdf:Property*. Tem o objetivo de relacionar dois elementos, relacionar um sujeito a um objeto uma propriedade relaciona dois recursos. Observe que todas as propriedades apresentadas no código aparecem dentro da estrutura de um elemento *<div>*. Exemplo, das linhas 4 a 13, há 3 triplas relacionadas à Bob
 - *http://example.org/bob#me foaf:knows Alice*
 - *http://example.org/bob#me schema:birthDate 1990-07-04*
 - *http://example.org/bob#me foaf:topic_interest the Mona Lisa*

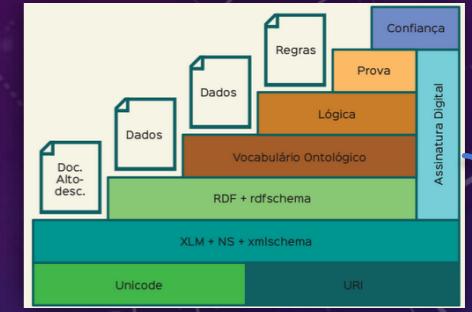
SERIALIZAÇÃO EM RDFA (VAMOS ESTUDAR O EXEMPLO....)

- No código HTML do slide anterior aparecem quatro atributos:
 - *typeof*, que equivale ao atributo para representar o elemento *rdf:type* (tem o mesmo objetivo do *rdf:type*).
 - Exemplo, na linha 04 temos
 - *http://example.org/bob#me rdf:type foaf:Person*

```
01 <body prefix="foaf: http://xmlns.com/fcaf/0.1/
02           schema: http://schema.org/
03           dcterms: http://purl.org/dc/terms/">
04 <div resource="http://example.org/bob#me" typeof="foaf:Person">
05   <p>
06     Bob knows <a property="foaf:knows" href="http://example.org/alice#me">Alice</a>
07     and was born on the <time property="schema:birthDate">1990-07-04</time>.
08   </p>
09   <p>
```

RDF/XML

- Foi a primeira serialização RDF, conhecido pela camadas “bolo de noiva” (Gandon, et al., 2014).
- O formato possui **sintaxe XML para os grafos RDF**.



```
01  <?xml version="1.0" encoding="utf-8"?>
02  <rdf:RDF
03      xmlns:dcterms="http://purl.org/dc/terms/"
04      xmlns:foaf="http://xmlns.com/foaf/0.1/"
05      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
06      xmlns:schema="http://schema.org/"
07      <rdf:Description rdf:about="http://example.org/bob#me">
08          <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
09          <schema:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1990-07-04<schema:birthDate>
10          <foaf:knows rdf:resource="http://example.org/alice#me"/>
11          <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/Q12418">
12      </rdf:Description>
13      <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
14          <dcterms:title>Mona Lisa</dcterms:title>
15          <dcterms:creator rdf:resource="http://dbpedia.org/resource/Leonardo_da_Vinci"/>
16      </rdf:Description>
17      <rdf:Description rdf:about="http://data.europeana.eu/item/04802/243FA861893F4117025F117025F17A8B813C5F9AA4D619">
18          <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
19      </rdf:Description>
20  </rdf:RDF>
```

SERIALIZAÇÃO EM RDF/XML

- A primeira linha do código já apresenta o elemento <?xml>.
 - Diferente do RDFa, que tem todo os seus elementos dentro da estrutura do HTML, no RDF/XML possui toda a sua estrutura dentro do XML e da tag <*rdf:RDF*>.
- O elemento *rdf:Description* é utilizado para identificar sujeitos.
 - O primeiro sujeito apresentado (Linha 07) possui quatro sub-elementos. Isto quer dizer que cada sub-elemento apresentado entre as linhas 08 e 11 representam uma tripla combinada com o sujeito descrito na linha 07.
- No RDF/XML, o sujeito, apesar de ser identificado pelo elemento *rdf:Description*, é descrito pelo *rdf:about*.
 - Diferentemente do código em RDFa, os tipos de dados reusados do esquema XML foram explicitamente definidos para identificar a semântica do *schema:birthDate*.

SERIALIZAÇÃO EM JSON-LD

- É um dos mais recente formatos de serialização, surgiu como uma extensão do JSON (Sporny, et al., 2014)
- Tem como objetivo de transformar código JSON para RDF com o mínimo de esforço possível. Formato é intuitivo para programadores já familiarizados com a sintaxe JSON.

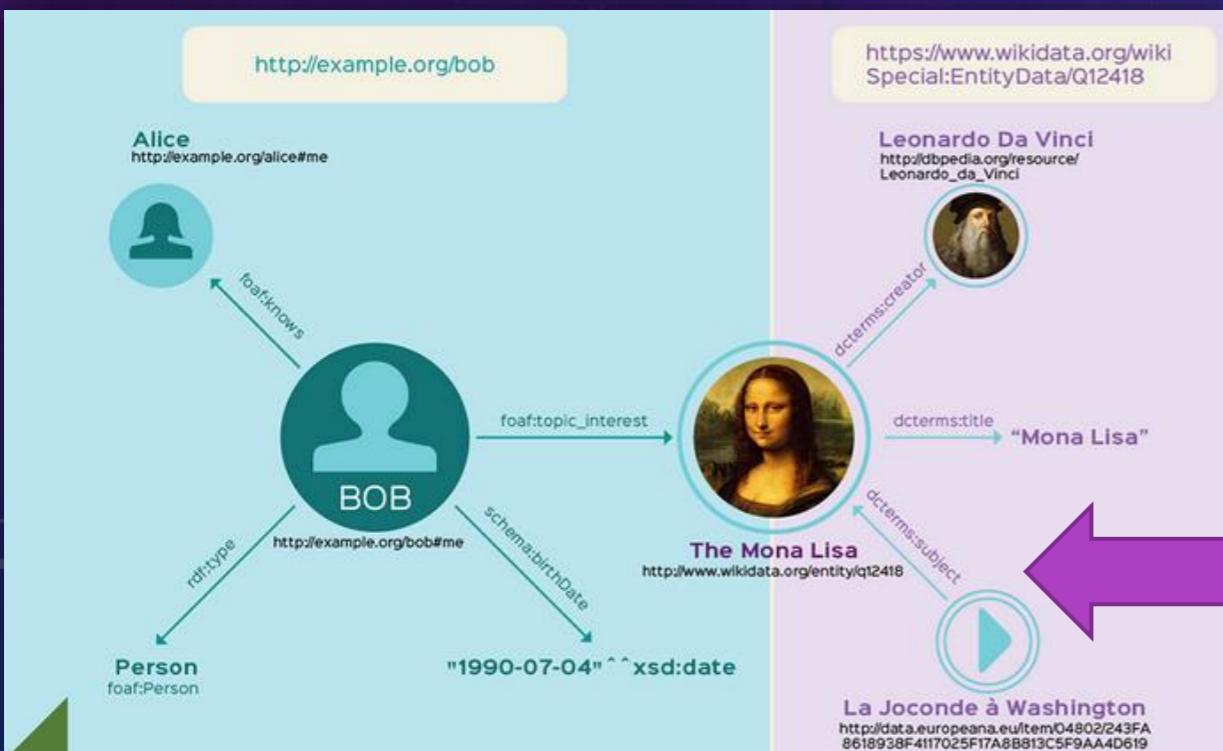
```
01  {
02    "@context": "example-context.json",
03    "@id": "http://example.org/bob#me",
04    "@type": "Person",
05    "birthdate": "1990-07-04",
06    "knows": "http://example.org/alice#me",
07    "interest": {
08      "@id": "http://www.wikidata.org/entity/Q12418",
09      "title": "Mona Lisa",
10      "subject_of": "http://data.europeana/item/04802/243FA8618938F41177025F17A8B813C5F9AA4D619",
11      "creator": "http://dbpedia.org/resource/Leonardo_da_Vinci"
12    }
13 }
```

SERIALIZAÇÃO EM JSON-LD

- o código descreve o recurso Bob, identificado pelo IRI `http://example.org/bob#me`, através da chave `@id`, (linha 03) e o tipo de recurso descrito (linha 04).
- As linhas 05 a 07 definem objetos que o sujeito Bob se relaciona. As triplas no JSON-LD são:
 - `http://example.org/bob#me birthDate 1990-07-04`
 - `http://example.org/bob#me knows http://example.org/alice#me`
 - `http://example.org/bob#me interest http://www.wikidata.org/entity/Q12418`
- Dois pontos que devemos destacar no código JSON-LD
 - i) A linha 07 identifica outro recurso no qual o sujeito Bob se relaciona, onde este recurso equivale a outro grafo, contendo uma estrutura própria e sendo identificado pelo IRI descrito na linha 08;
 - ii) **não é possível identificar o esquema e a semântica dos recursos.** Isto acontece por que o contexto no qual os recursos estão inseridos ficam em outro documento. Este contexto é identificado pelo objeto `@context`, descrito na linha 02.

SERIALIZAÇÃO EM JSON-LD

- o código que descreve o contexto é...
- Através do **contexto** é possível fazer o mapeamento do código JSON-LD do grafo na Monalisa



```
01  {
02    "@context": {
03      "foaf": "http://xmlns.com/foaf/0.1/",
04      "Person": "foaf:Person"
05      "interest": "foaf:topic_interest",
06      "knows": {
07        "@id": "foaf:knows",
08        "@type": "@id"
09      },
10      "birthdate": {
11        "@id": "http://schema.org/birthDate",
12        "@type": "http://www.w3.org/2001/XMLSchema#date"
13      },
14      "dcterms": "http://purl.org/dc/terms/",
15      "title": "dcterms:title",
16      "creator": {
17        "@id": "dcterms:creator",
18        "@type": "@id"
19      },
20      "subject_of": {
21        "@reverse": "dcterms:subject",
22        "@type": "@id"
23      }
24    }
25  }
```

TURTLE

- Surgiu para ampliar as possibilidades de descrição de um documento N-Triples, o formato descreve prefixos e IRIs relativos na estrutura do documento (Prud'hommeaux, et al., 2014). **Turtle é simples e fácil de ler.**

```
01  BASE      <http://example.org/>
02  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
03  PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
04  PREFIX schema: <http://schema.org/>
05  PREFIX dcterms: <http://purl.org/dc/terms/>
06  PREFIX wd: <http://www.wikidata.org/entity/>
07
08  <bob#me>
09    a foaf:Person ;
10    foaf:knows <alice#me> ;
11    schema:birthDate "1990-07-04"^^xsd:date ;
12    foaf:topic_interest wd:Q12418
13
14  wd:Q12418
15    dcterms:title "Mona Lisa" ;
16    dcterms:creator <http://dbpedia.org/resource/Leonardo_da_Vinci> .
17
18  <http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4d619>
19    dcterms:subject wd:Q12418 .
```

TURTLE

- Usa @PREFIX para encurtar IRI (URL) = Namespace
 - No nosso exemplo
 - @PREFIX foaf: <http://xmlsns.com/foaf/0.1/>
 - Permite que se escreva
 - foaf:knows <alice#me>
 - foaf:topic_interest wd:Q12418

SERIALIZAÇÃO EM TURTLE

- As 6 primeiras linhas mostram os IRIs que podem ser definidos como prefixos e IRI base do documento, (características não permitidas no N-Triples)
- As linhas 08, 14 e 18 apresentam sujeitos com seus predicados e objetos logo abaixo deles. Este tipo de organização e endentação torna bastante intuitiva a leitura do documento, facilitando assim a identificação das triplas RDF.

```
01      BASE    <http://example.org/>
02      PREFIX  foaf: <http://xmlns.com/foaf/0.1/>
03      PREFIX  xsd:  <http://www.w3.org/2001/XMLSchema#>
04      PREFIX  schema: <http://schema.org/>
05      PREFIX  dcterms: <http://purl.org/dc/terms/>
06      PREFIX  wd:   <http://www.wikidata.org/entity/>
07
```

```
08      <bob#me>
09          a foaf:Person ;
10          foaf:knows <alice#me> ;
11          schema:birthDate "1990-07-04"^^xsd:date ;
12          foaf:topic_interest wd:Q12418
13
14      wd:Q12418
15          dcterms:title "Mona Lisa" ;
16          dcterms:creator <http://dbpedia.org/resource/Leonardo_da_Vinci> .
17
18      <http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4d619>
19          dcterms:subject wd:Q12418 .
```

SERIALIZAÇÃO EM TURTLE

```
07  
08 <bob#me>  
09     a foaf:Person ;  
10     foaf:knows <alice#me> ;  
11     schema:birthDate "1990-07-04"^^xsd:date ;  
12     foaf:topic_interest wd:Q12418  
13
```

- A linha 09 apresenta mais uma característica do formato Turtle.
 - A primeira tripla descrita para Bob é *bob#me a foaf:Person*. O elemento que é responsável por relacionar o sujeito ao predicado deste exemplo é o token “*a*”.
 - O token “*a*” possui a mesma semântica da propriedade *rdf:type* e é usada para dizer que *bob#me* é do tipo *foaf:Person*.
- **Por que o “a”?** A letra “a” representa um artigo da língua inglesa, sendo traduzido para português como “um” ou “uma”. Ou seja, ao ler a tripla descrita na linha inglesa seria *Bob is a person* (no português: *Bob é uma pessoa*). Logo, o token “*a*” foi adicionado para tornar o formato Turtle ainda mais intuitivo para o ser humano.

RESUMO DE SERIALIZAÇÃO E UTILIZAÇÃO

Serialização RDF	Tipo de Código	Quando usar
RDFa	Código RDF embutido em HTML	SEO (<i>Search Engine Optimization.</i>)
RDF/XML	Código RDF com estrutura em XML	Aplicações que usam estruturas em XML
JSON-LD	Código RDF com estrutura JSON	Aplicações que usam JSON
N-Triples	Código RDF com estrutura de Triplas	Processamento e intercâmbio de Big Data em RDF.
Turtle	Código RDF para facilitar a leitura humana	Processamento e intercâmbio de Big Data em RDF
TriG	Código com estrutura Turtle	Representação de Múltiplos grafos
N-Quads	Código RDF com estrutura de Triplas	Processamento e Intercâmbio de grandes catálogos de dados.

REFERÊNCIAS

- **Data Scientist: The Sexiest Job of the 21st Century**
 - <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>
- **The Semantic Web** - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities
 - <http://www.scientificamerican.com/article/the-semantic-web/>
- Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)