

DR. MIGUEL CERIANI

PROF. ALEJANDRO VAISMAN

---

# ONTOLOGÍAS EN LA WEB

## 1. INTRO A LA WEB SEMÁNTICA

## WEB SEMÁNTICA

- ▶ Web de los Datos (Modelo Genérico)
- ▶ Ontologías (Modelos para estos Datos)
- ▶ Lenguaje de Interrogación para la Web
- ▶ Arquitectura Aplicaciones para la Web Semántica

### EN ESTE CURSO

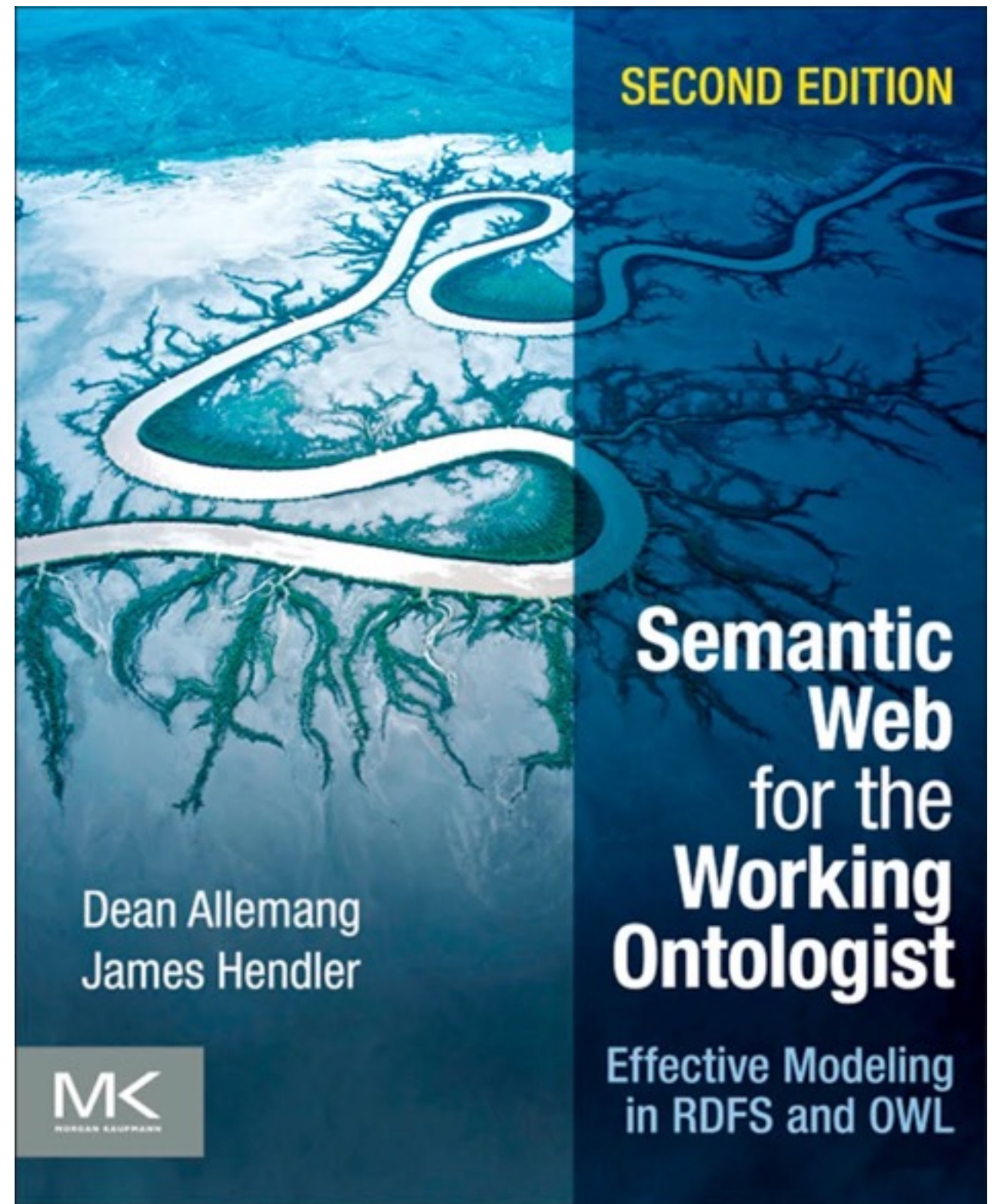
- ▶ Web de los Datos (Modelo Genérico): ~ 2 clases
- ▶ Ontologías (Modelos de Datos): ~ 10 clases
- ▶ Lenguaje de Interrogación: ~ 1 clase
- ▶ Arquitectura Aplicaciones: ~ 1 clase

## TEXTO UTILIZADO

*D. Allemang, J. Hendler*  
**Semantic Web for the Working  
Ontologist**  
(Second Edition)

Morgan Kaufmann, 2011

- ▶ Springer Store
- ▶ Amazon Store
- ▶ ScienceDirect



## ALGUNOS RECURSOS ONLINE

- ▶ Semantic University  
<http://www.cambridgesemantics.com/semantic-university>
- ▶ EdUcational Curriculum for the usage of LInked Data  
<http://www.euclid-project.eu/>
- ▶ Estandares W3C (World Wide Web Consortium)  
<https://www.w3.org/standards/semanticweb/>

## REFERENCIAS PARA ESTA CLASE (INTRO A LA WEB SEMÁNTICA)

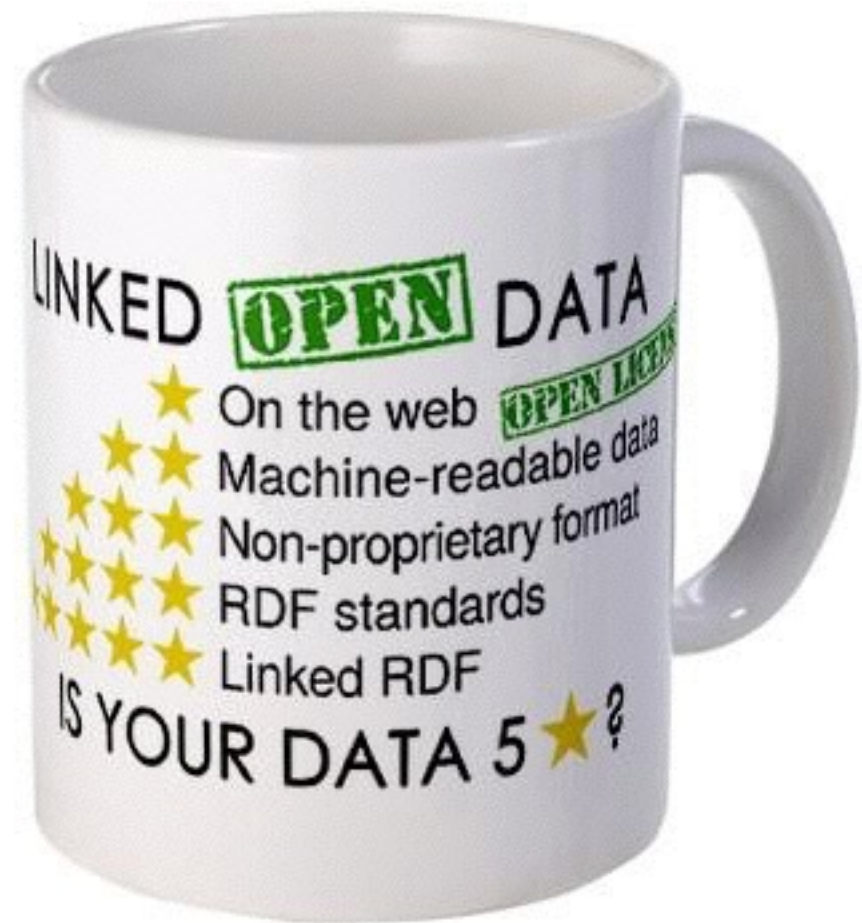
- ▶ Caps. 1 y 2 de "Semantic for the Working Ontologist"
- ▶ Tim Berners-Lee "Design Issues: Abstractions"  
<https://www.w3.org/DesignIssues/Abstractions>
- ▶ Semantic University:  
[Introduction to the Semantic Web](#), [RDF-101](#)
- ▶ Tim Berners-Lee "The Next Web", TED2009 (Video)  
[https://www.ted.com/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web](https://www.ted.com/talks/tim_berners_lee_on_the_next_web)



## TIM BERNERS-LEE: DE LA WEB A LA WEB SEMÁNTICA



"The Semantic Web", 2001



"Linked Data", 2006

## DE LA VOZ DEL IDEADOR

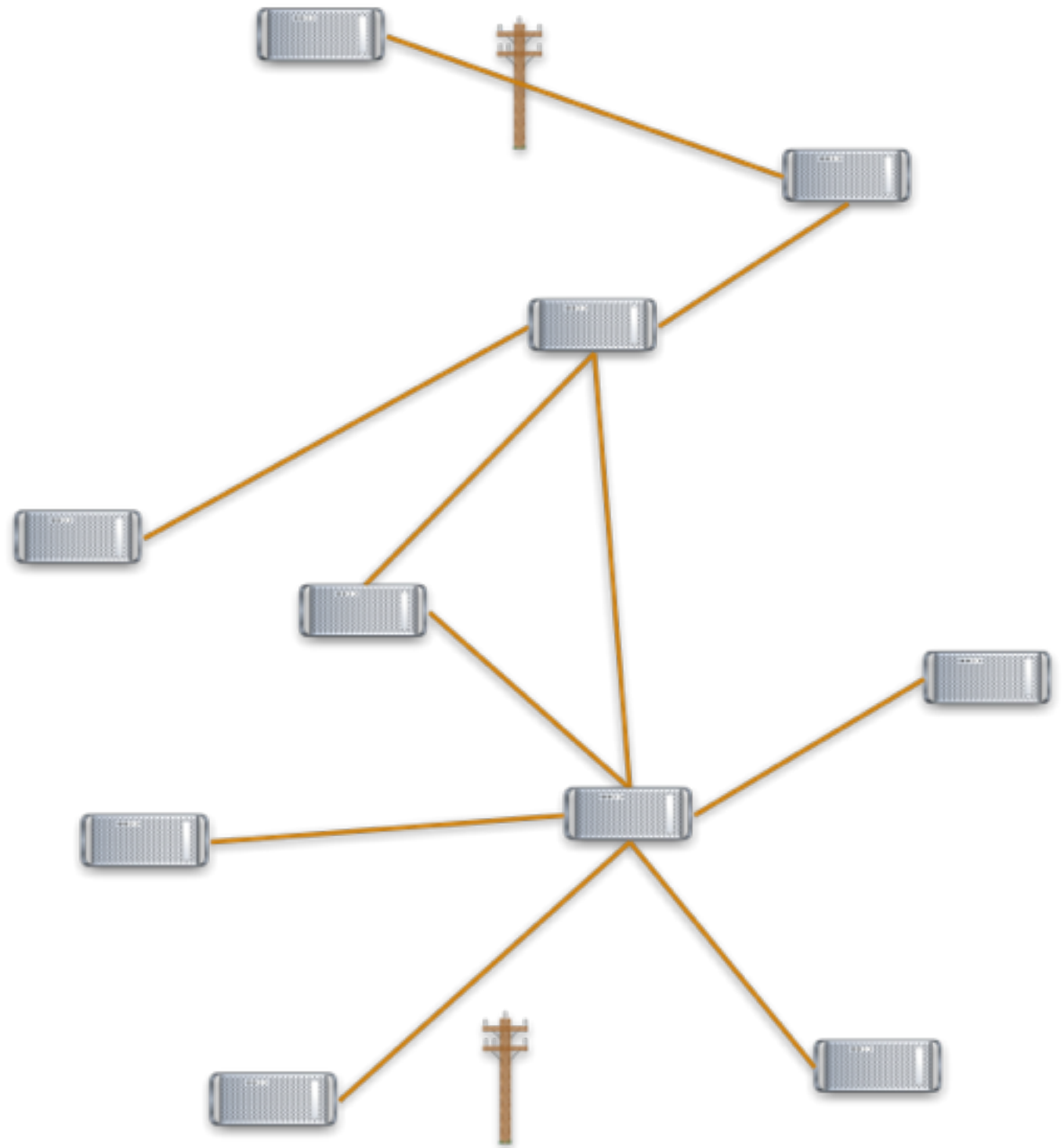
Time Berners-Lee:  
**The next web**  
TED2009

[https://www.ted.com/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web](https://www.ted.com/talks/tim_berners_lee_on_the_next_web)



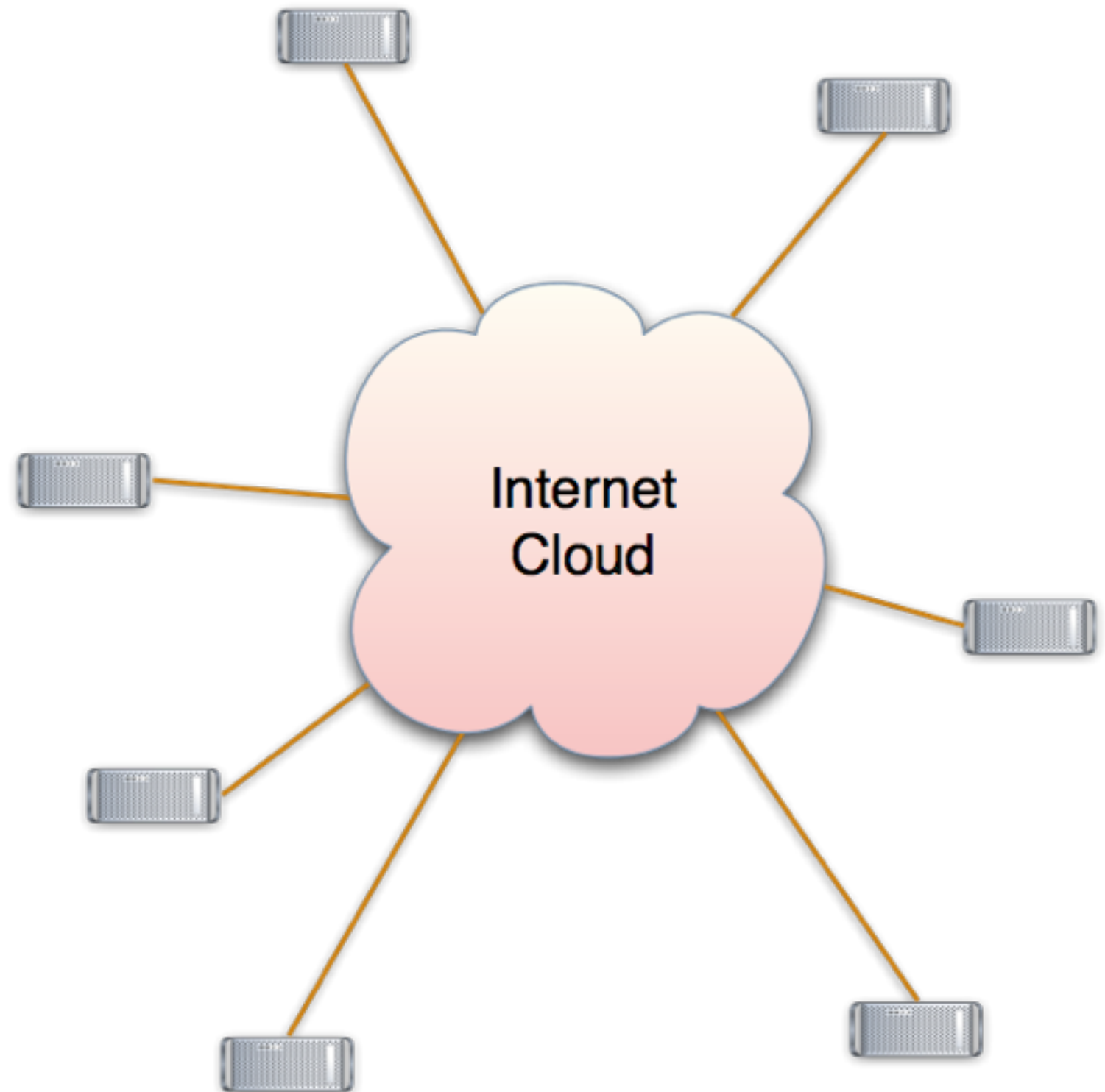
# RED DE COMPUTADORAS

- ▶ Cada computadora ve solo su entorno
- ▶ Para comunicar tengo que conocer las conexiones (topografía de la red)



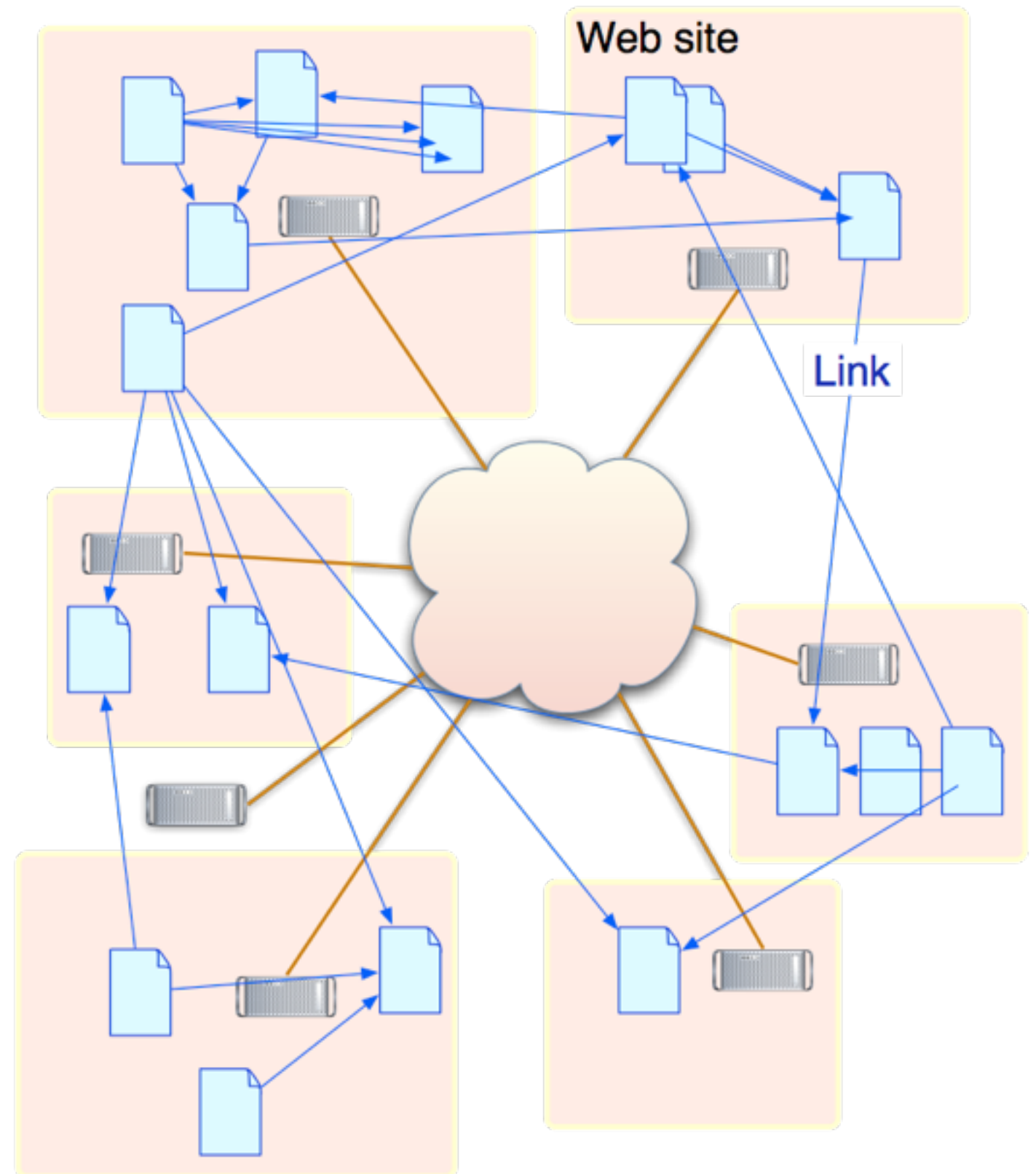
# I ABSTRACCIÓN: INTERNET

- ▶ Cada computadora ve toda la red (Internet)
- ▶ Para comunicar tengo que solo que conocer el "nombre" de la computadora (IP o DNS)



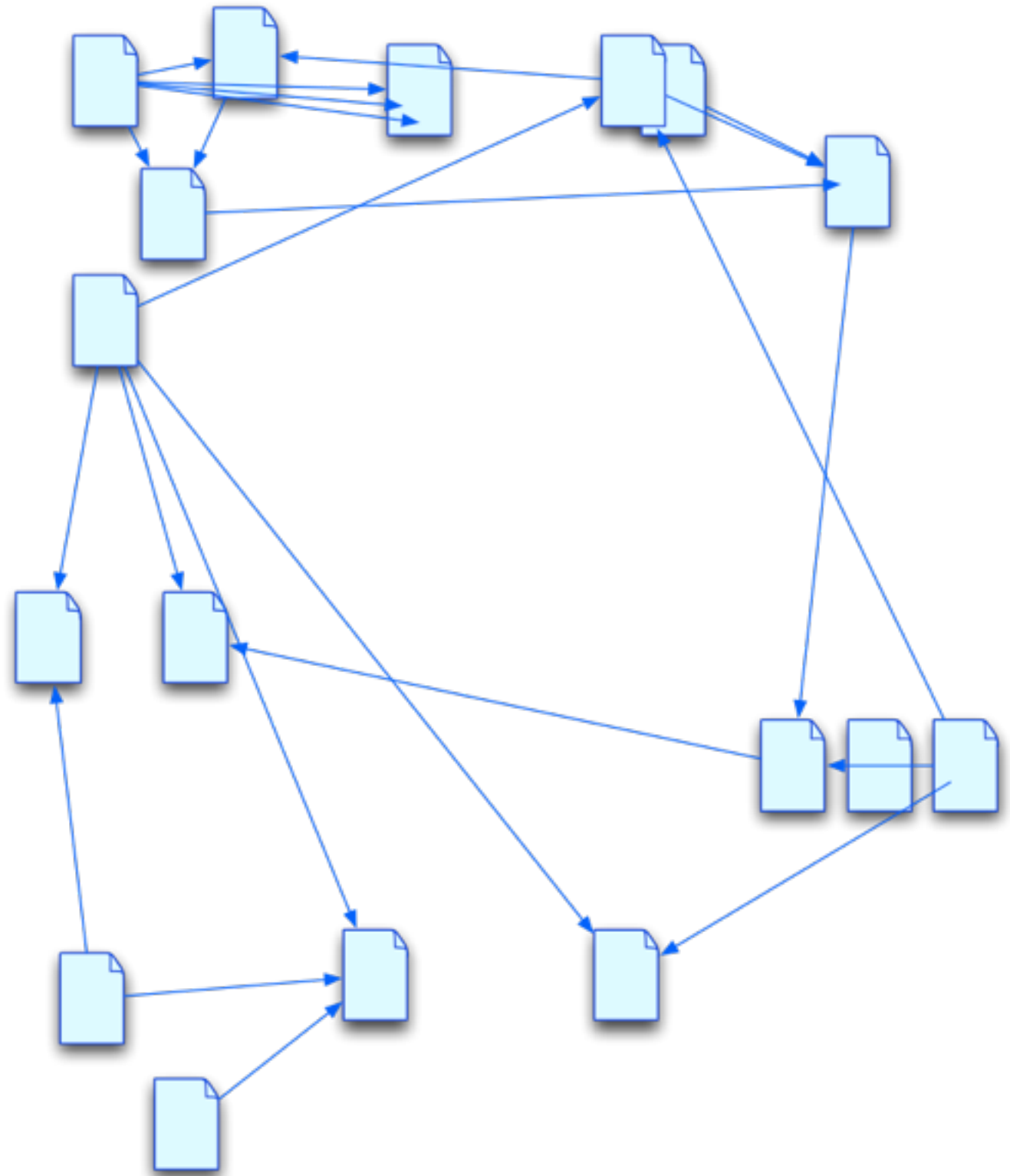
## II ABSTRACCIÓN: WEB

- ▶ cada servidor Web ofrece documentos puede tener enlaces a cualquier otro



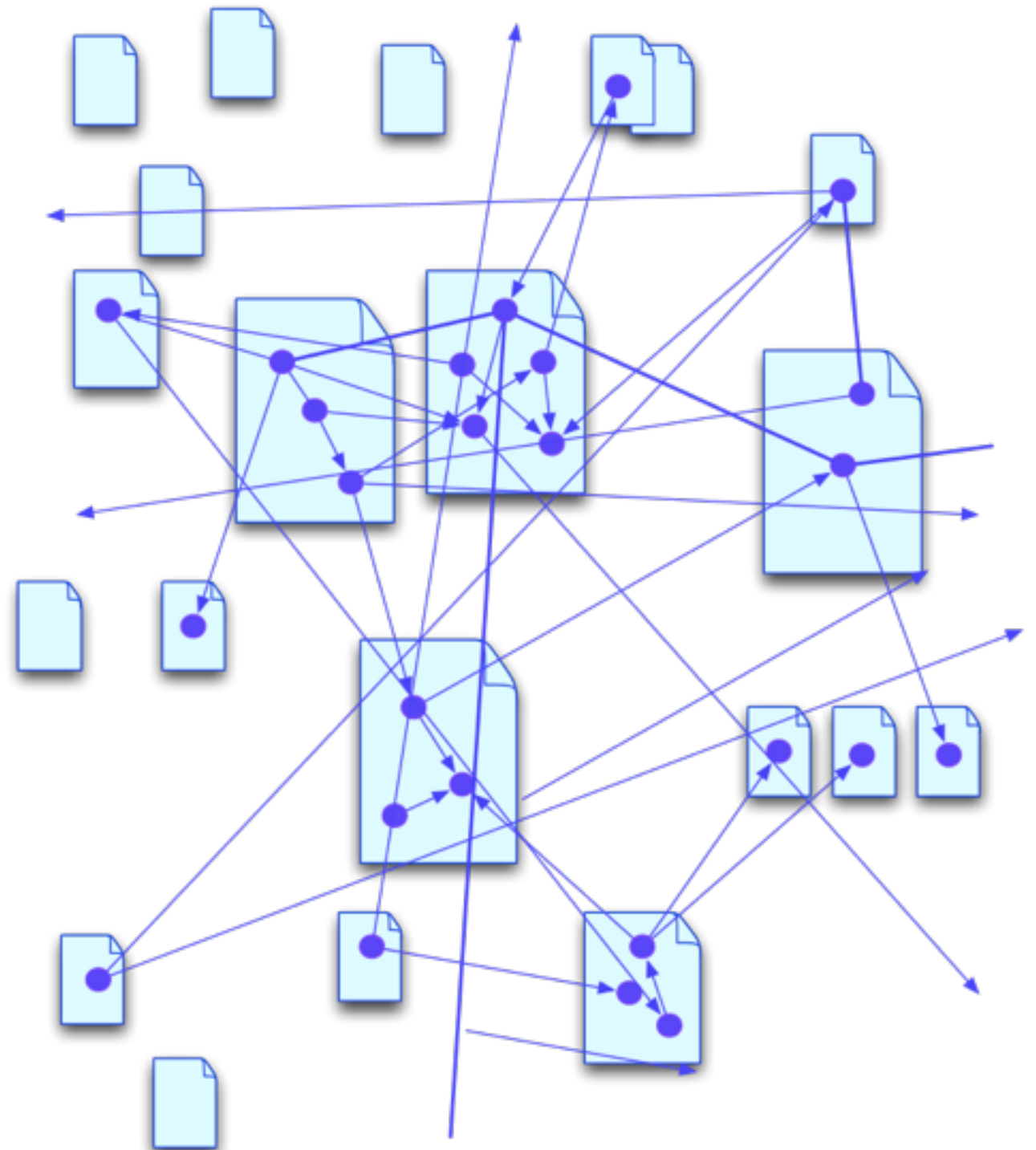
## II ABSTRACCIÓN: WEB

- ▶ cada servidor Web ofrece documentos puede tener enlaces a cualquier otro
- ▶ red de documentos



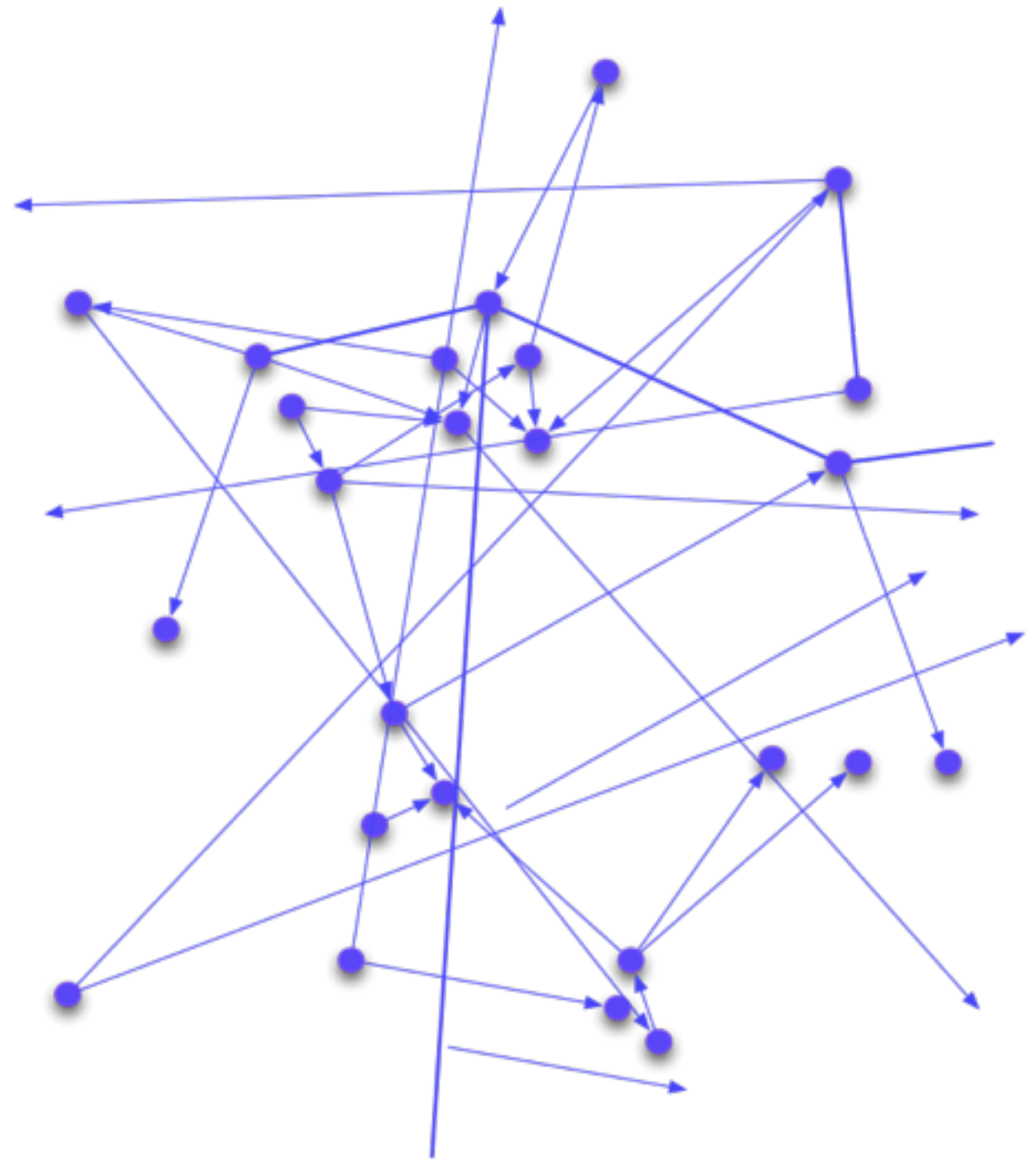
### III ABSTRACCIÓN: WEB SEMÁNTICO

- ▶ los documentos describen conceptos que están en relación entre ellos

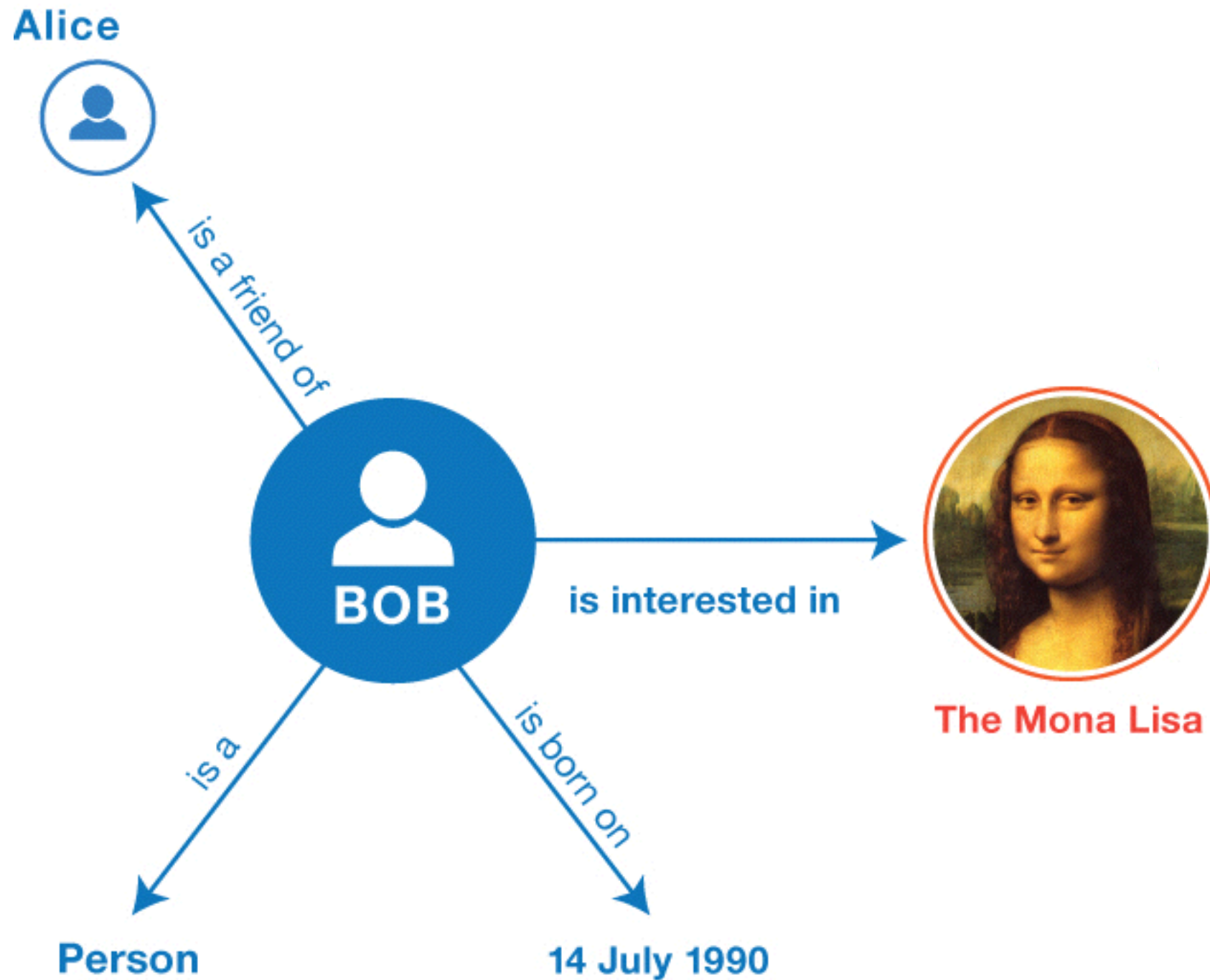


### III ABSTRACCIÓN: WEB SEMÁNTICO

- ▶ los documentos describen conceptos que están en relación entre ellos
- ▶ las relaciones entre los conceptos forman un grafo de informaciones

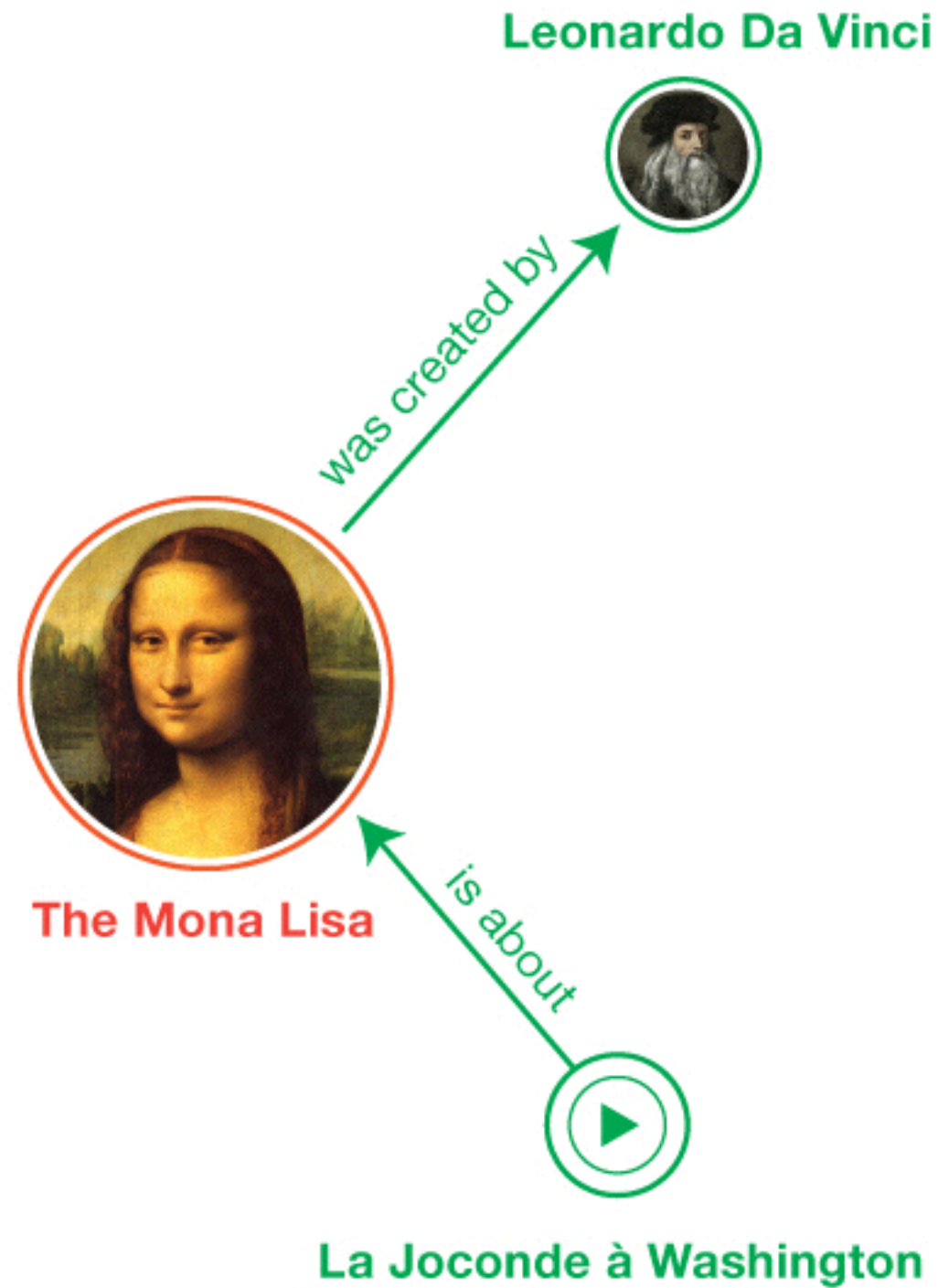


## GRAFO DE CONCEPTOS

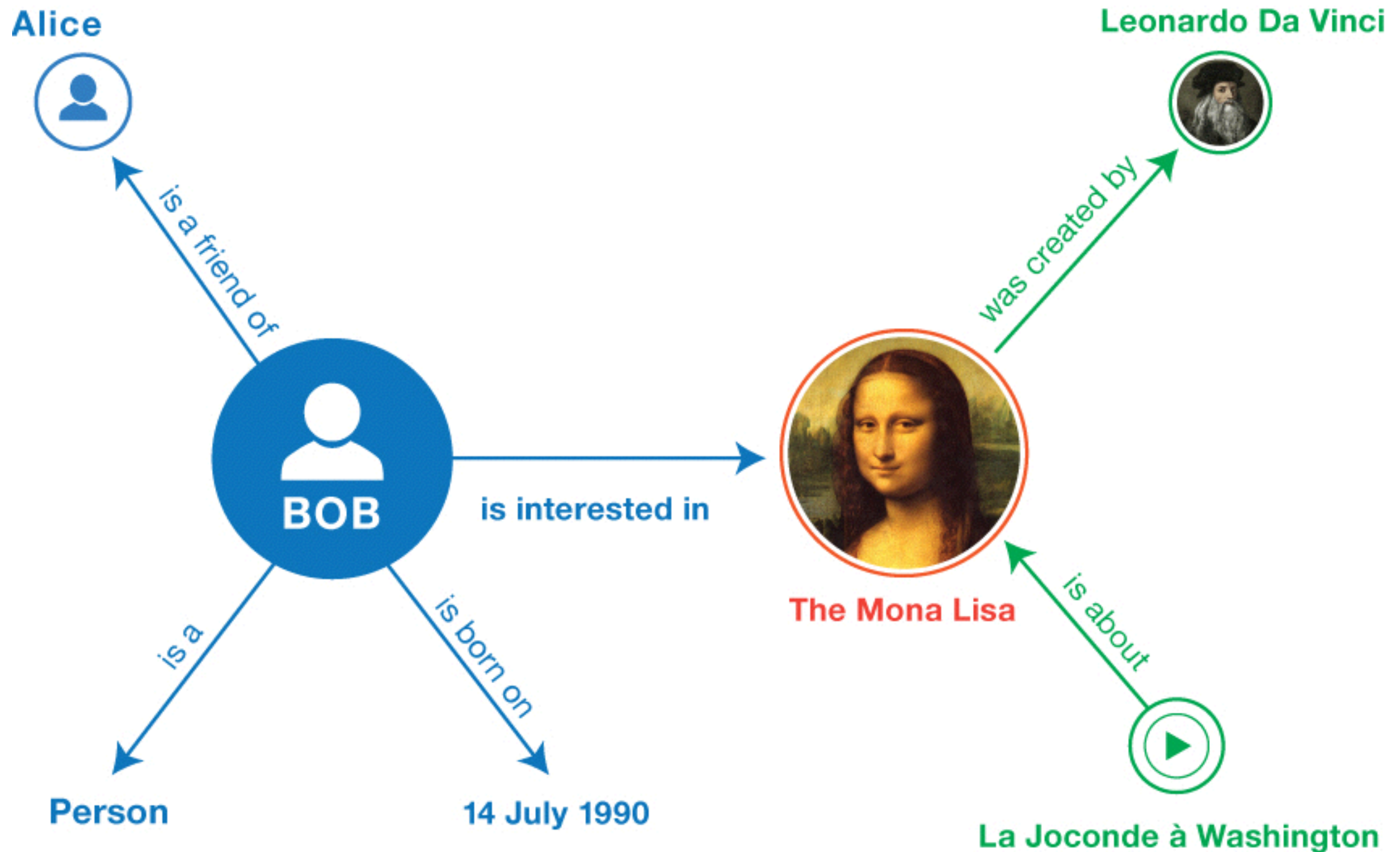




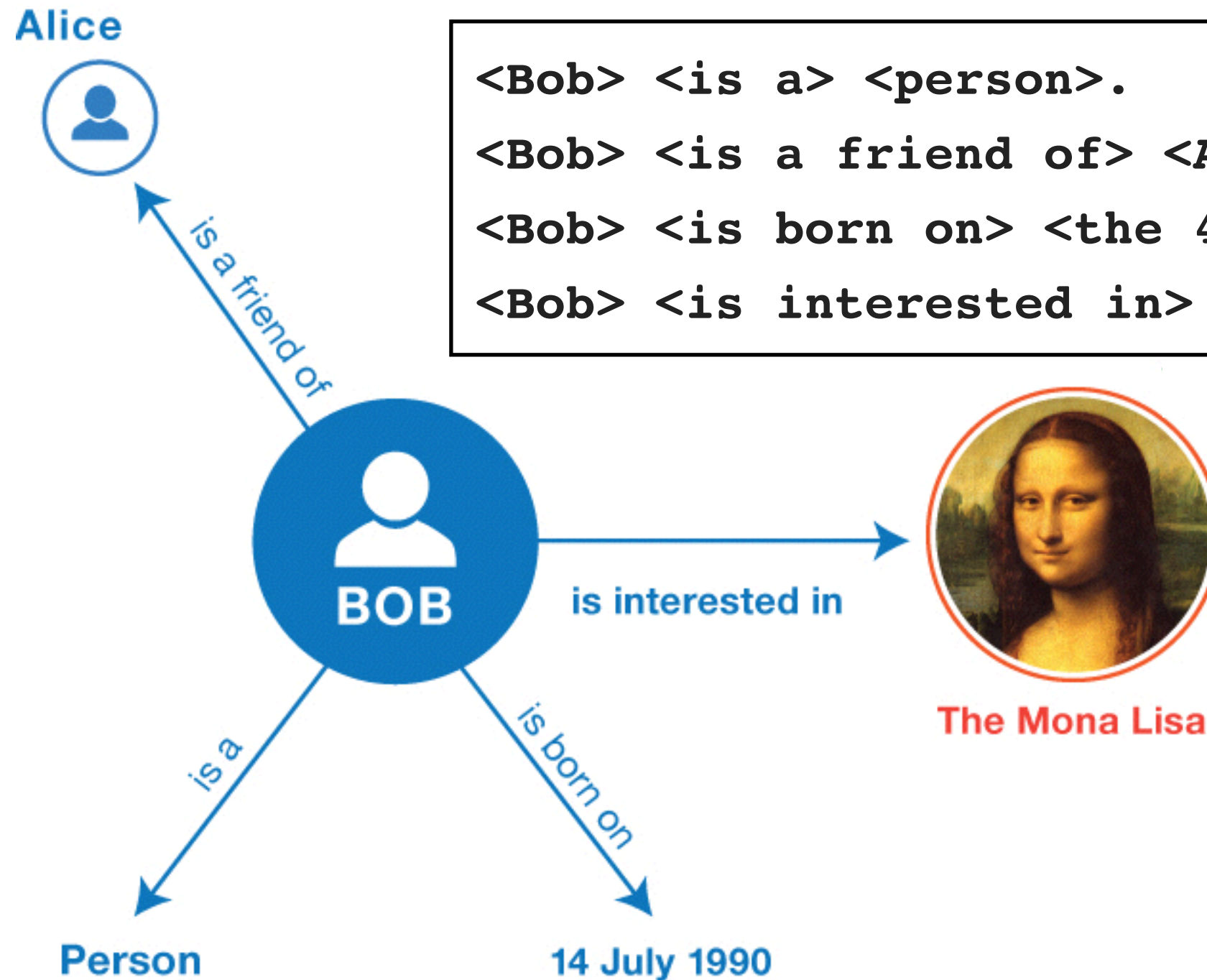
## GRAFO(S) DE CONCEPTOS



## GRAFO DE CONCEPTOS: UNIÓN



## GRAFOS DE CONCEPTOS: TRIPLAS (SUJETO PREDICADO OBJETO)



`<Bob> <is a> <person>.`

`<Bob> <is a friend of> <Alice>.`

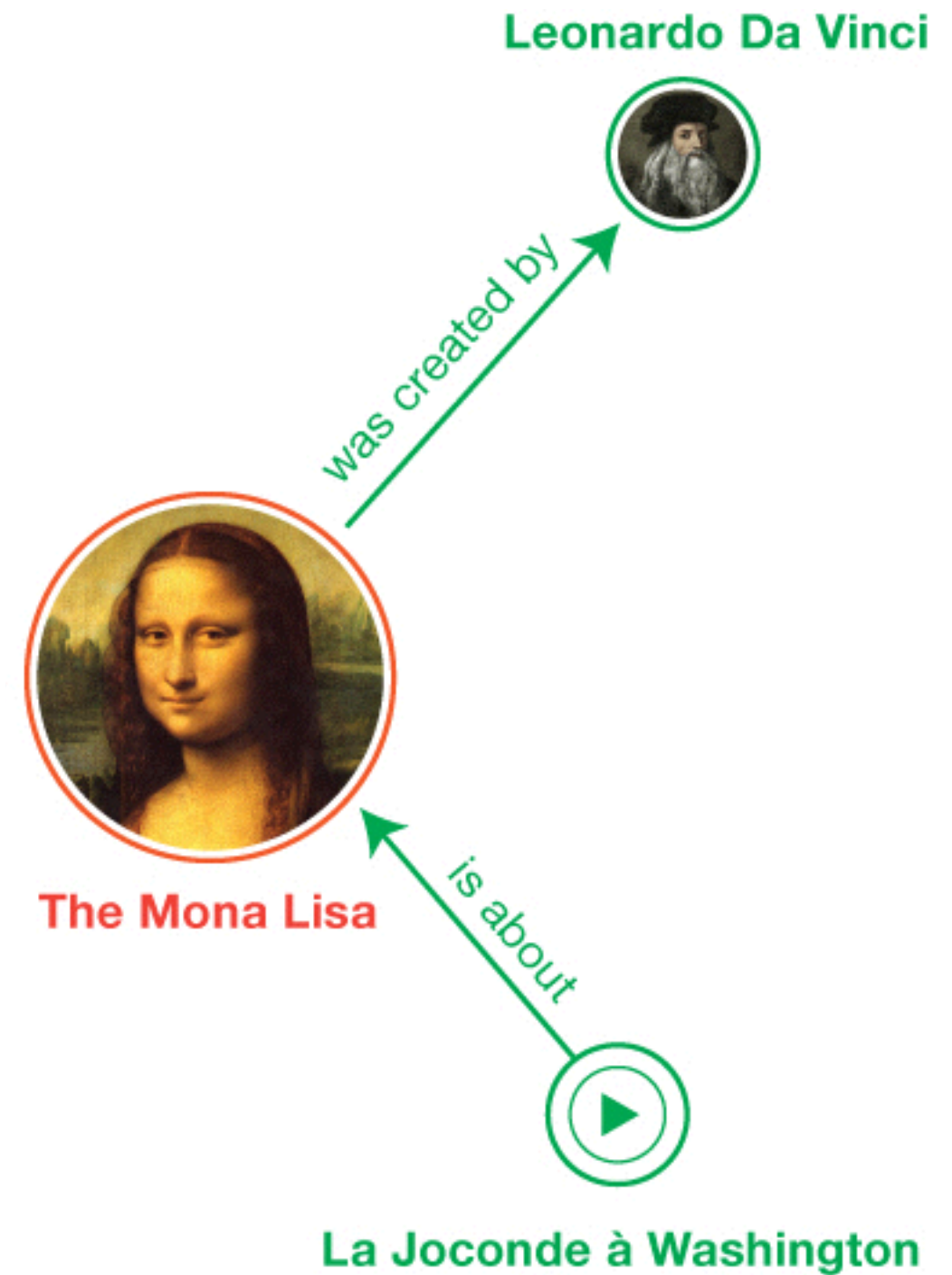
`<Bob> <is born on> <the 4th of July 1990>.`

`<Bob> <is interested in> <the Mona Lisa>.`



The Mona Lisa

## GRAFOS DE CONCEPTOS: TRIPLAS



<Leonardo da Vinci> <is the creator of> <the Mona Lisa>.  
<The video 'La Joconde...'> <is about> <the Mona Lisa>.

## GRAFO DE CONCEPTOS: UNIÓN DE TRIPLAS

<Bob> <is a> <person>.

<Bob> <is a friend of> <Alice>.

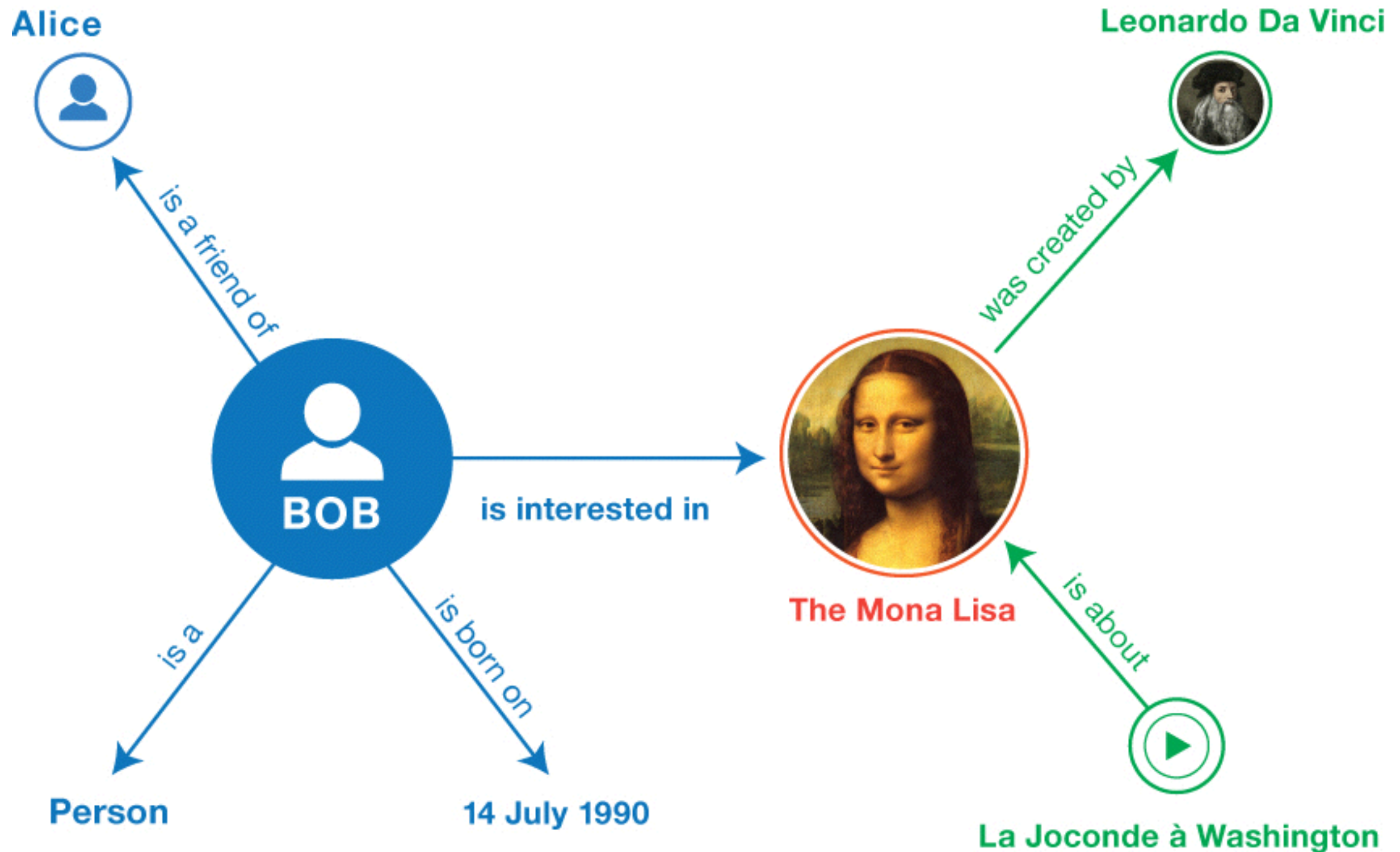
<Bob> <is born on> <the 4th of July 1990>.

<Bob> <is interested in> <the Mona Lisa>.

<Leonardo da Vinci> <is the creator of> <the Mona Lisa>.

<The video 'La Joconde...'> <is about> <the Mona Lisa>.

## GRAFO DE CONCEPTOS: UNIÓN



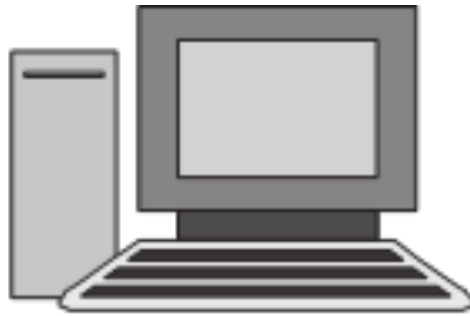


# DISTRIBUIR TABLA(S) DE DATOS

ID	Title	Author	Medium	Year
1	As You Like It	Shakespeare	Play	1599
2	Hamlet	Shakespeare	Play	1604
3	Othello	Shakespeare	Play	1603
4	"Sonnet 78"	Shakespeare	Poem	1609
5	Astrophil and Stella	Sir Phillip Sidney	Poem	1590
6	Edward II	Christopher Marlowe	Play	1592
7	Hero and Leander	Christopher Marlowe	Poem	1593
8	Greensleeves	Henry VIII Rex	Song	1525



## POR FILA?



1	<i>As You Like It</i>	Shakespeare	Play	1599
---	-----------------------	-------------	------	------



*Needs common schema—which column is which?*

4	Sonnet 78	Shakespeare	Poem	1609
6	Edward II	Christopher Marlowe	Play	1592



7	Hero and Leander	Christopher Marlowe	Poem	1593
3	Othello	Shakespeare	Play	1603

## POR COLUMNA?



Year	Medium
1599	Play
1604	Play
1603	Play
1609	Poem
1590	Poem
1592	Play
1593	Poem
1525	Song



*Needs to reference entities—which thing are we talking about?*



Author
Shakespeare
Shakespeare
Shakespeare
Shakespeare
Sir Phillip Sidney
Christopher Marlowe
Christopher Marlowe
Henry VIII Rex

Title
<i>As You Like It</i>
<i>Hamlet</i>
<i>Othello</i>
"Sonnet 78"
<i>Astrophil and Stella</i>
<i>Edward II</i>
<i>Hero and Leander</i>
<i>Greensleeves</i>

# POR FILA Y POR COLUMNA: TRIPLAS



	Medium
Row 7	Poem

	Title
Row 2	Hamlet

*Needs to reference both schema and entities*



	Author
Row 4	Shakespeare



	Year
Row 2	1604

	Medium
Row 6	Play

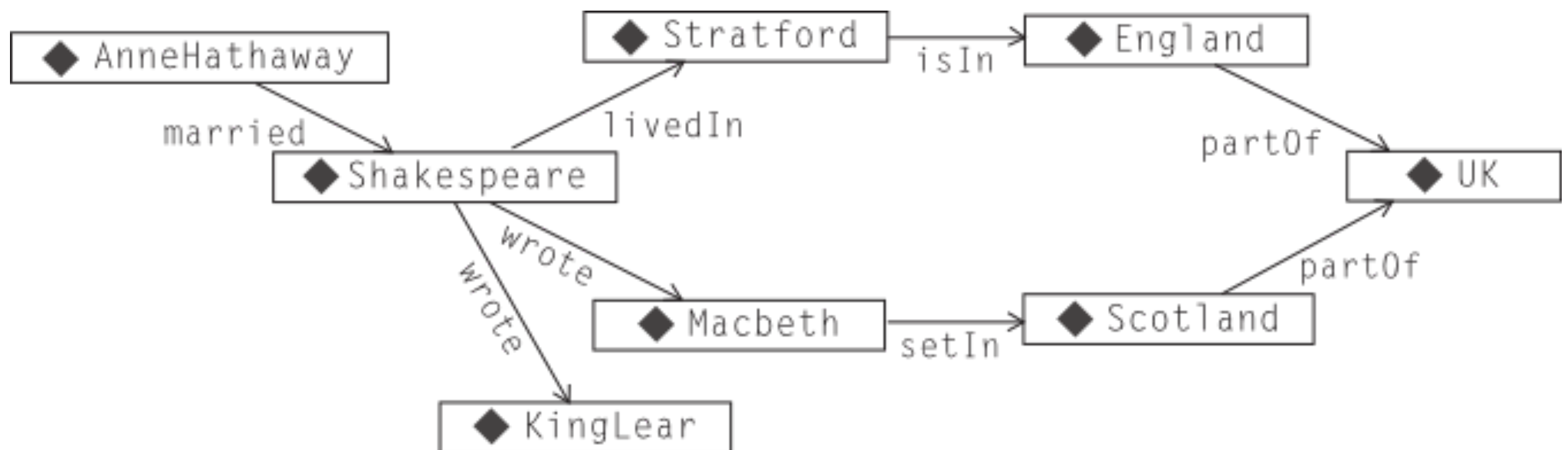
## POR FILA Y POR COLUMNA: TRIPLAS

Subject	Predicate	Object
Row 7	Medium	Poem
Row 2	Title	Hamlet
Row 2	Year	1604
Row 4	Author	Shakespeare
Row 6	Medium	Play

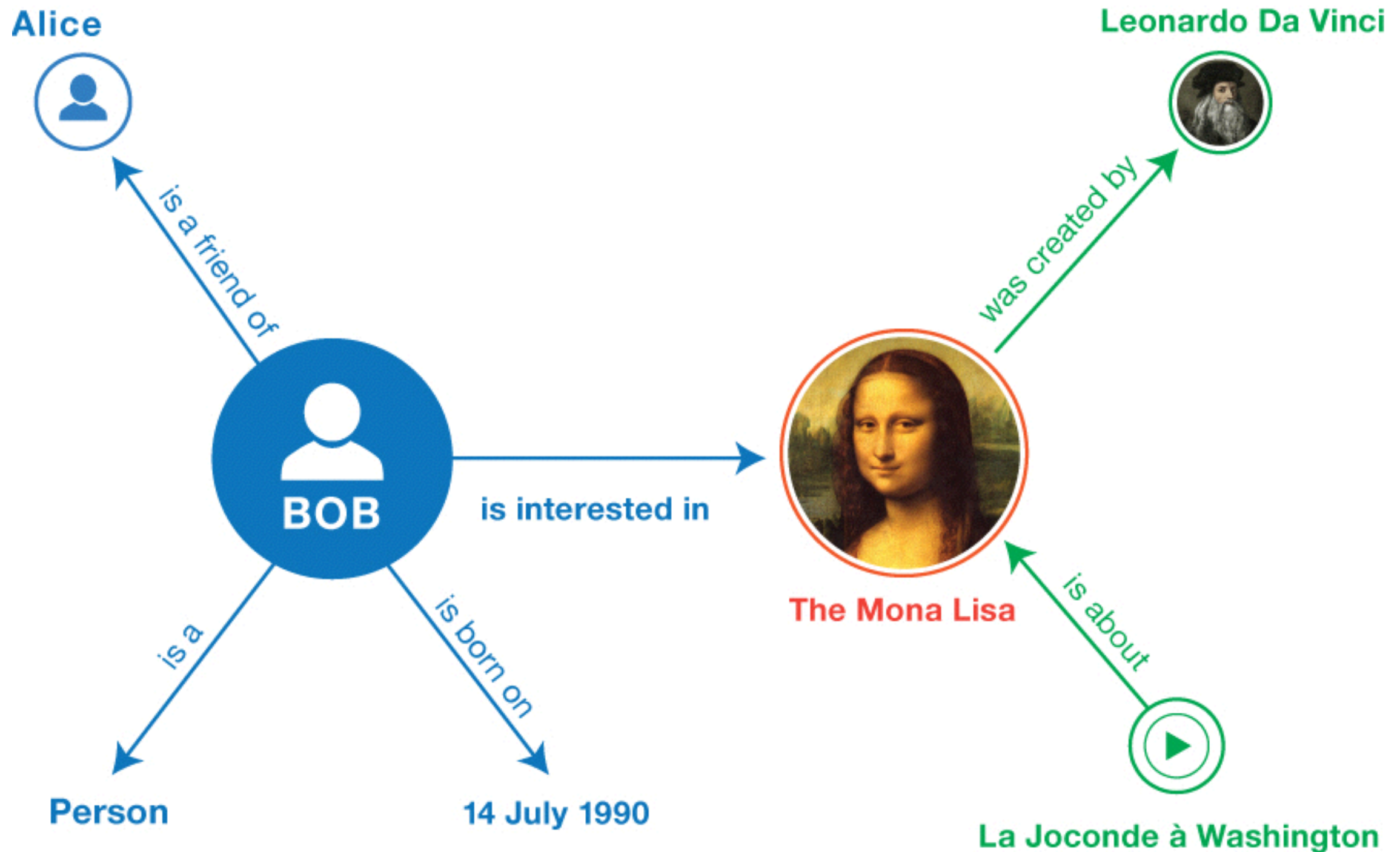
## MÁS TRIPLAS

Subject	Predicate	Object
Shakespeare	wrote	King Lear
Shakespeare	wrote	Macbeth
Anne Hathaway	married	Shakespeare
Shakespeare	livedIn	Stratford
Stratford	isIn	England
Macbeth	setIn	Scotland
England	partOf	UK
Scotland	partOf	UK

## DE LAS TRIPLAS AL GRAFO



## VOLVEMOS A LA UNIÓN DE GRAFOS





## GRAFO DE CONCEPTOS: PARA QUE FUNCIONE SIRVE...

- ▶ Adecuada Representación de Datos
- ▶ Arquitectura Aplicativa que lo permita la Integración
- ▶ Lenguaje de Interrogación para usarlo

# CÓMO REPRESENTAR LOS DATOS EN LA WEB?

- Necesitamos

- ▶ Identificar Universalmente los Conceptos
- ▶ Usar Vocabularios Establecidos

- Pero en la Web

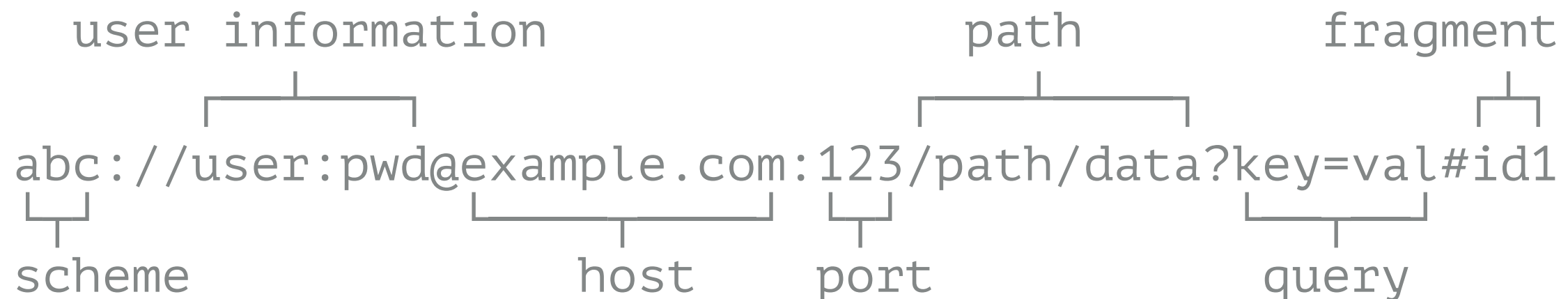
- ▶ No puedo Centralizar la Identificación de nuevos Conceptos
- ▶ No puedo Anticipar todos los Vocabularios
- ▶ Tengo que reconocer que Opiniones Discordantes son Inevitables (y Necesarias)

# SOLUCIÓN

- Identificación de Conceptos y Relaciones con URI  
([http://dbpedia.org/resource/Leonardo\\_da\\_Vinci](http://dbpedia.org/resource/Leonardo_da_Vinci),  
<http://purl.org/dc/terms/creator>, ...)
  - ▶ Posibilidad de usar Conceptos/Relaciones existentes o crear nuevos
- Modelo de Datos (Ontologías) también Distribuido:
  - ▶ Posibilidad de establecer posteriormente Identidad u otras Relaciones entre Conceptos/Relaciones

# URL, URI, IRI

- ▶ URI: Uniform Resource Identifier (identifica un recurso)



- ▶ URL: Uniform Resource Locator  
(un URI que localiza una representación del recurso)
- ▶ IRI: Internationalized Resource Identifier (un URI con caracteres UNICODE)

# URL, URI, IRI?

- ▶  $\text{URLs} \subset \text{URIs} \subset \text{IRIs}$
- ▶ Los estándares ahora hablan de IRIs
- ▶ Los textos y este curso hablan de URIs (viejos estándares) o IRIs (nuevos estándares)
- ▶ Prácticamente usamos siempre URLs

# URI Y PREFIJOS

- URIs

`http://dbpedia.org/resource/Leonardo_da_Vinci`

`http://purl.org/dc/terms/creator`

`http://purl.org/dc/terms/title`

- Asociando Prefijos

`dbpedia:        ↪     http://dbpedia.org/resource/`

`dcterms:       ↪     http://purl.org/dc/terms/`

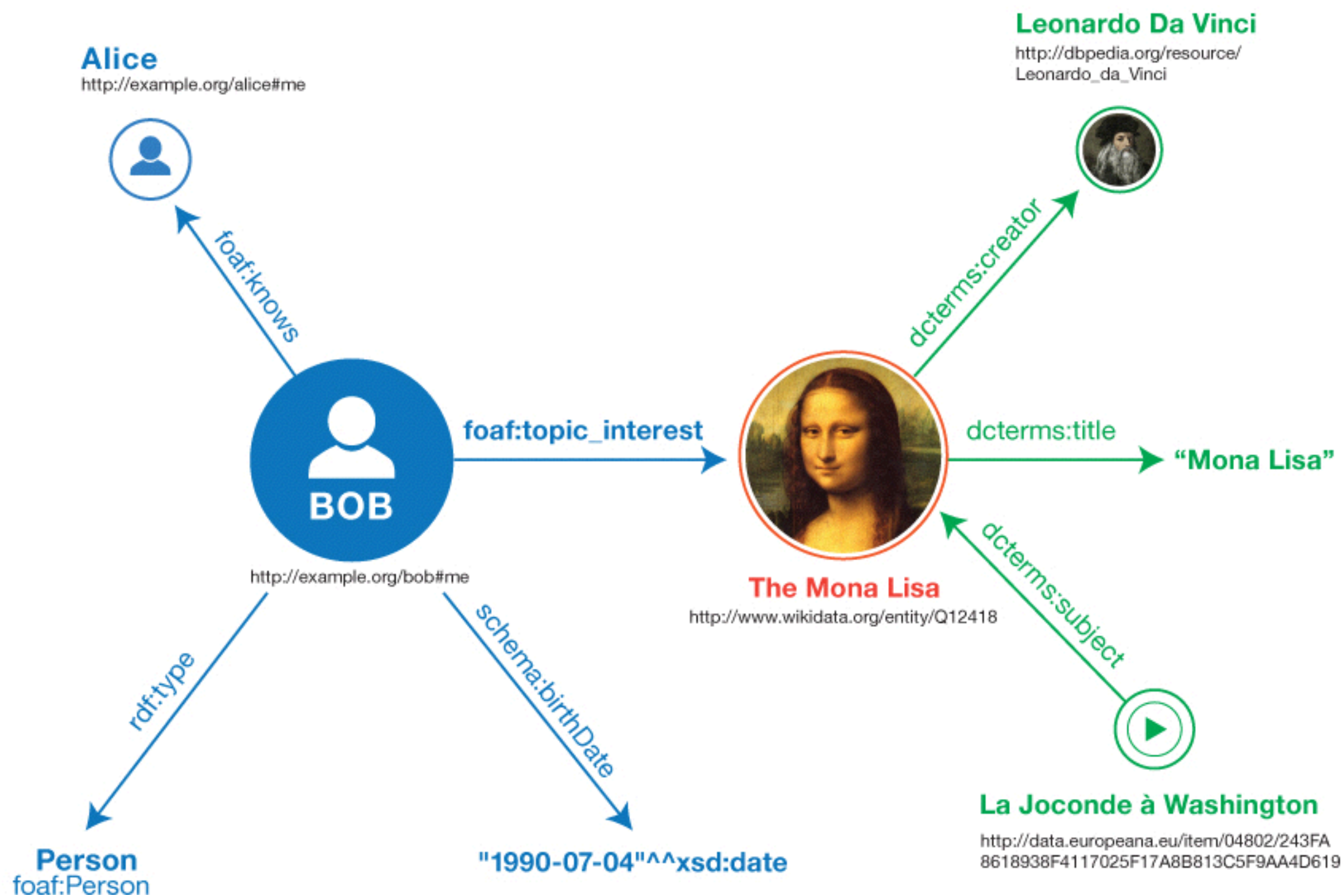
- Usando los Prefijos

`dbpedia:Leonardo_da_Vinci`

`dcterms:creator`

`dcterms:title`

## GRAFO DE CONCEPTOS: URI





# LA WEB SEMÁNTICA, 2001



# TECNOLOGÍAS

- ▶ Representación de Datos con Triplas de URIs:  
RDF (2002)
- ▶ Modelo de Datos Distribuido:  
RDFS, OWL (2002)
- ▶ Lenguaje de Interrogación:  
SPARQL (2004)
- ▶ Arquitectura:  
SPARQL Endpoint (2004), Principios Linked Data (2006),  
Graph Store (2013), Linked Data Platform (2015)

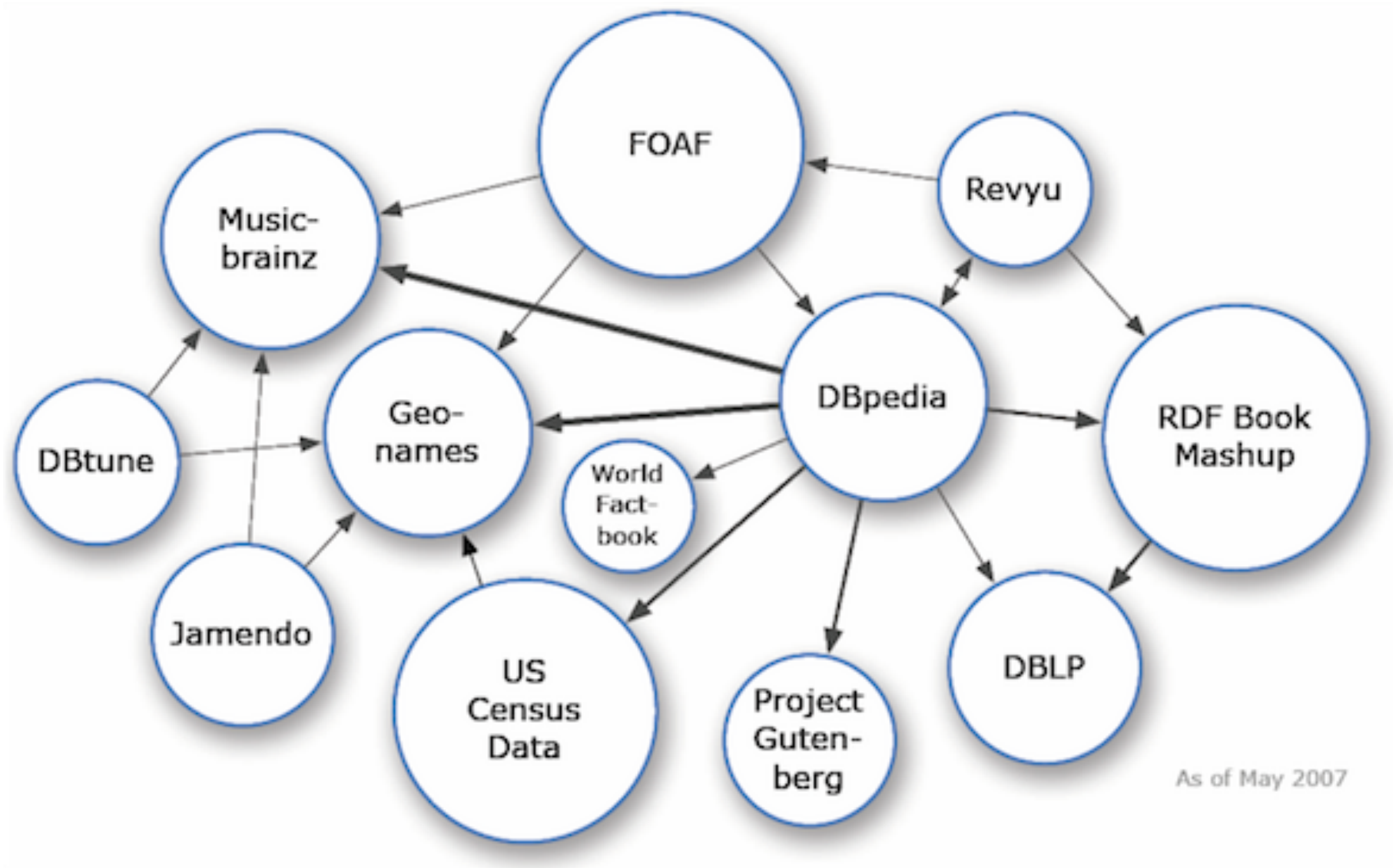
## BASE PARA LA WEB SEMÁNTICA: LINKED (OPEN) DATA, 2006

Calidad (estrellas) de un conjunto de datos

1. Datos (Públicos) en la Web
2. Elaborables por Maquinas
3. en Formatos Abiertos
4. en Formatos RDF
5. Enlazados (refieren a conceptos y vocabularios afuera del conjunto)

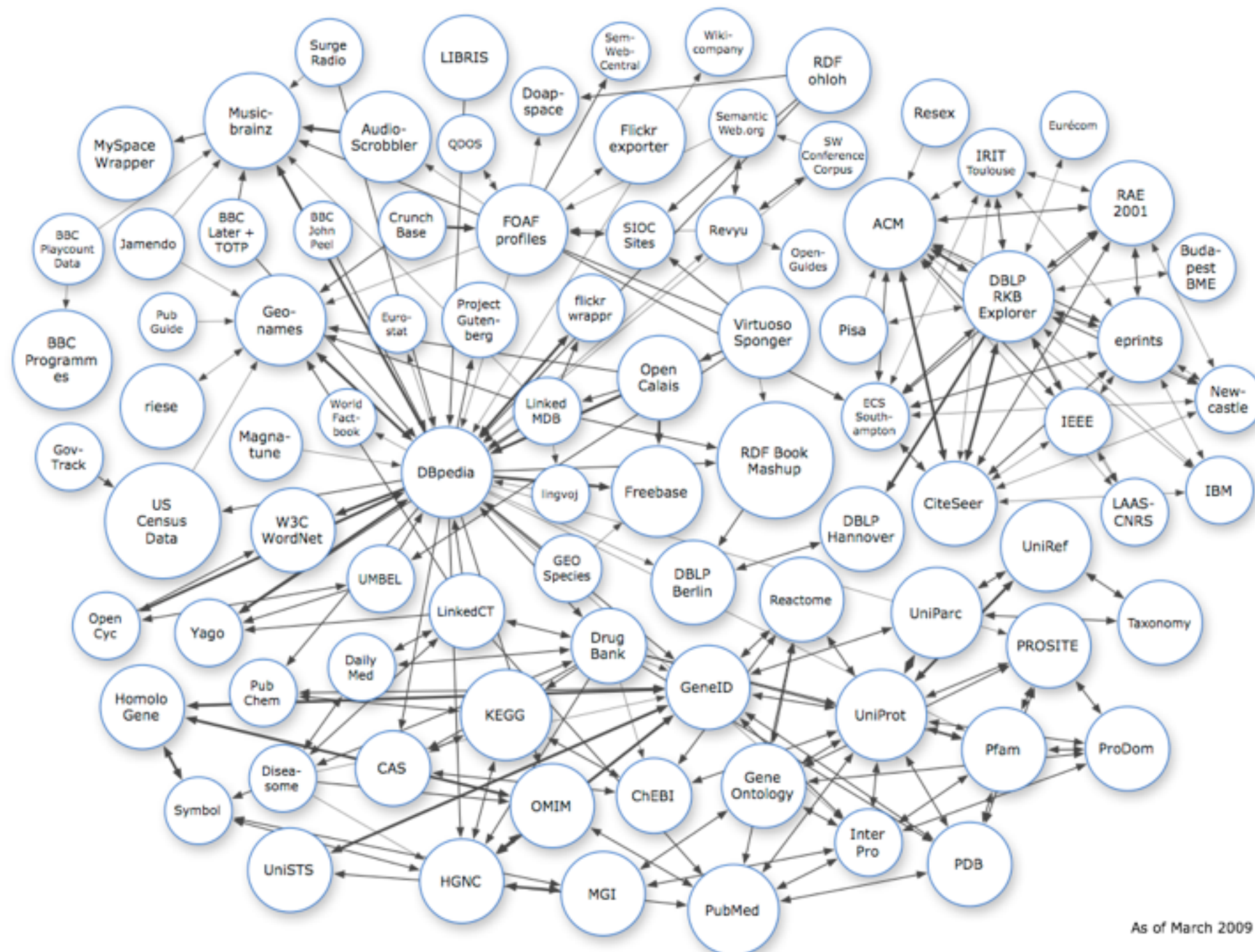


## LINKED OPEN DATA, 2007

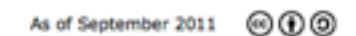




## LINKED OPEN DATA, 2009











# LINKED OPEN DATA: EJEMPLOS

- ▶ DBpedia  
Datos extraídos de InfoBox de Wikipedia
- ▶ WikiData, Freebase  
Datos insertados por los usuarios
- ▶ LinkedGeoData  
Datos extraídos de OpenStreetMaps



# EXPLORANDO DATOS: LINKED DATA BROWSER

- ▶ Un concepto a la vez
- ▶ Muestra las Relaciones (salientes y entrantes) con otros conceptos o datos
- ▶ Paradigma: textual, visual (grafo)
- ▶ Especifico para un Dataset o Genérico
- ▶ Puede tener un mecanismo de extensiones ("plug-in")

## BUSCANDO DATOS

Índice de Linked Open Data



<https://datahub.io/>

# ACTIVIDAD

- ▶ Explorar directamente de las respectivas paginas: DBpedia, WikiData, FreeBase
- ▶ Explorar visualmente con lod live ([en.lodlive.it](http://en.lodlive.it))
- ▶ Traten de entender el significado de las triplas
- ▶ Buscar datos sobre un tema o aplicación que les interese
- ▶ Ver si hay enlaces entre distintos datasets
- ▶ Pensar posibles interrogaciones, visualizaciones, aplicaciones
- ▶ Faltarían datos? Probar a buscar con datahub