The thirty minute guide to RDF and Linked Data

Originally created by Ian Davis and Tom Heath, but please share and reuse

What is RDF?

RDF is the data format for linked data

Resource Description Framework

In essence a simple data model

It's about writing down relations between things

See http://www.w3.org/RDF/

What is RDF for?

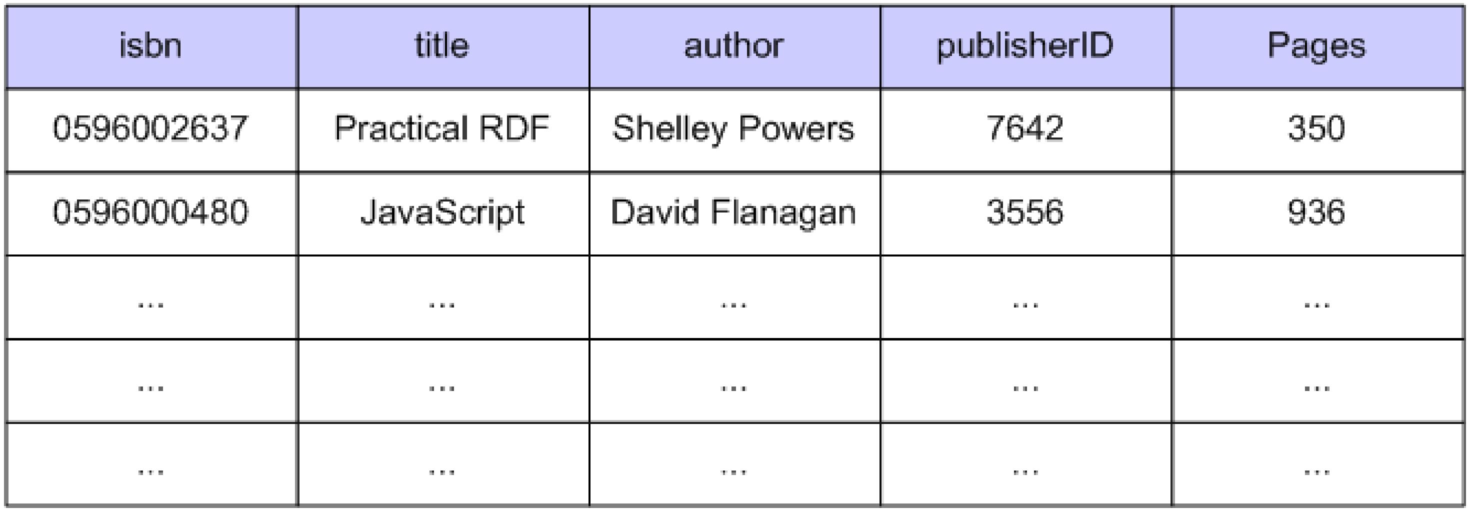
The web is a global, universal information space for documents

Can we do same for data?

Make the web into a database?

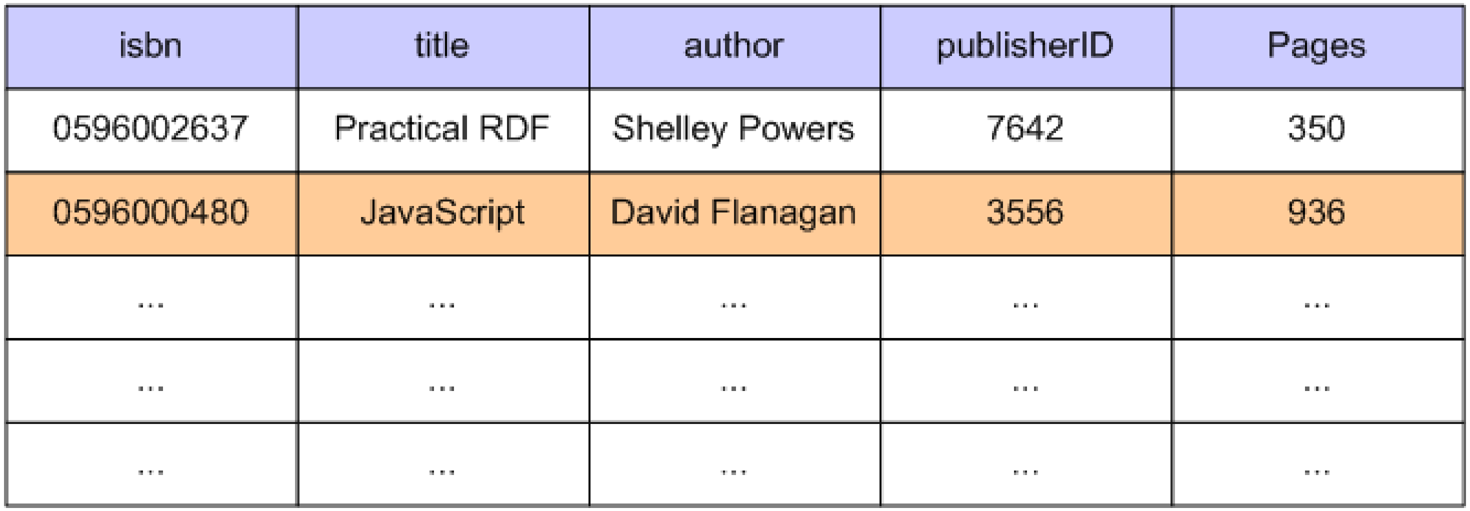
RDF is the data format for that database

Imagine a typical database table



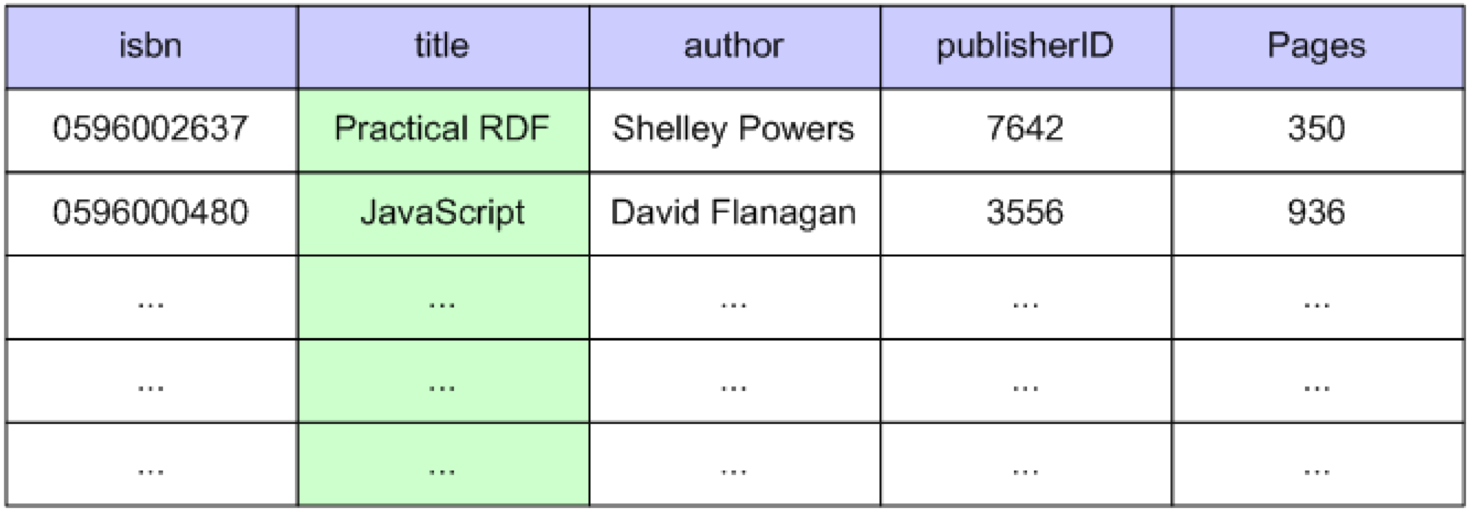
isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

The rows represent the “things”



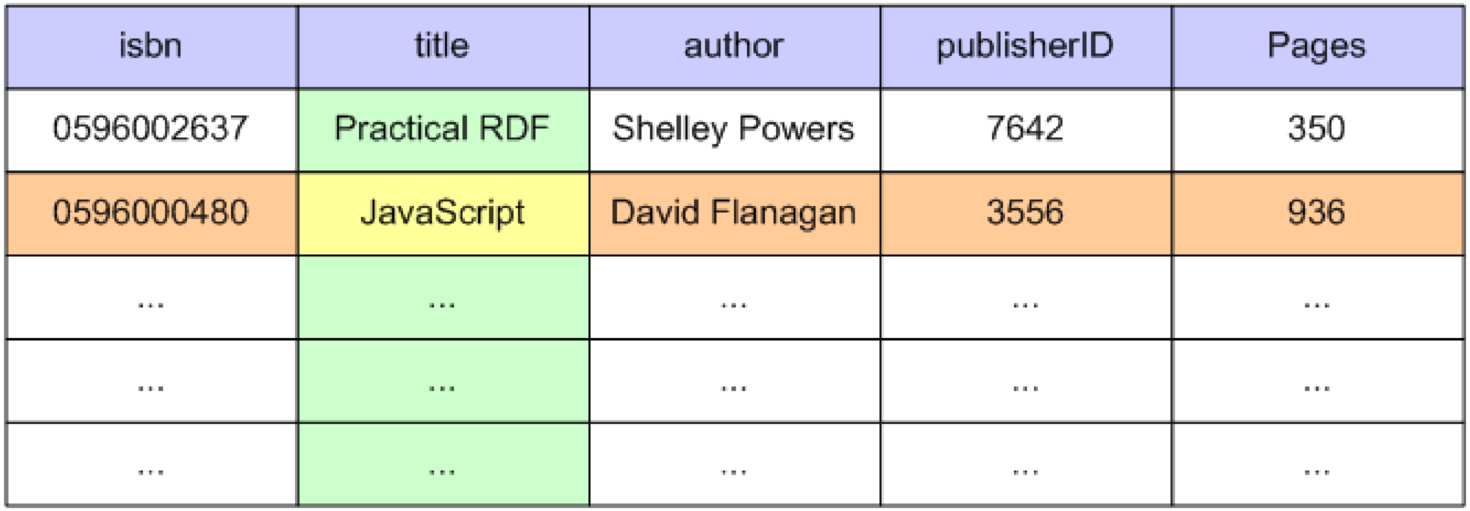
isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Columns represent the “properties”



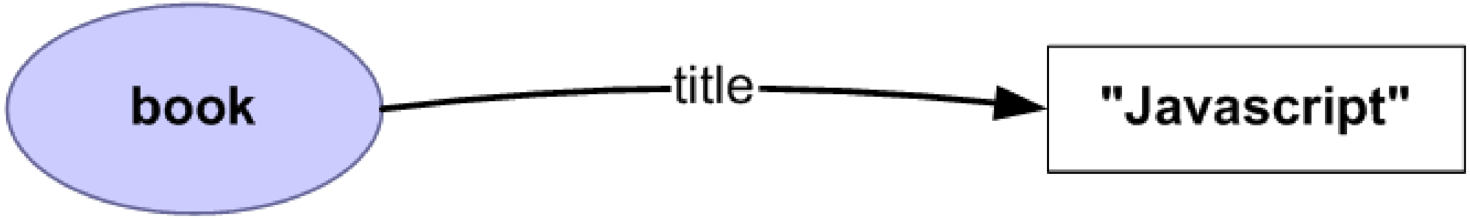
isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Intersections represent properties of things

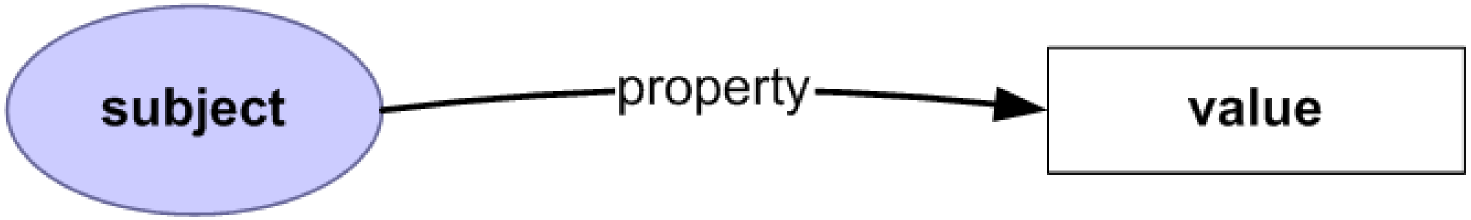


isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

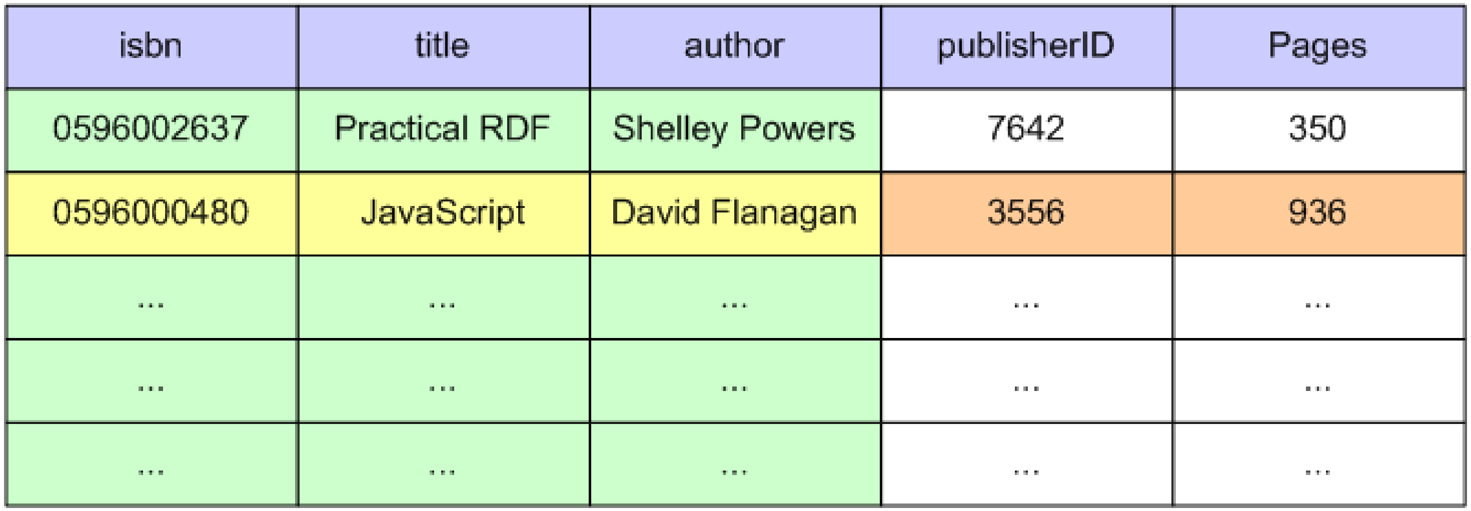
We can draw a picture of that information



The Essence of RDF: the “triple”

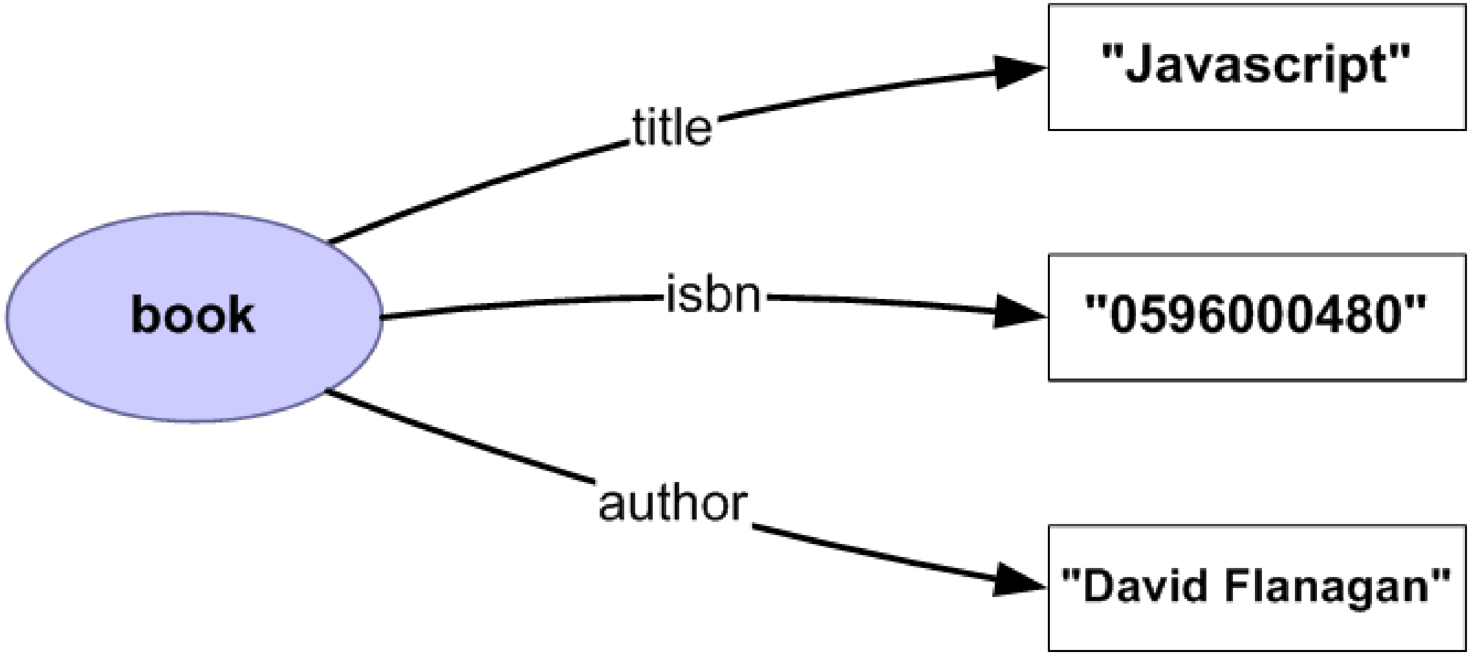


Selecting multiple properties

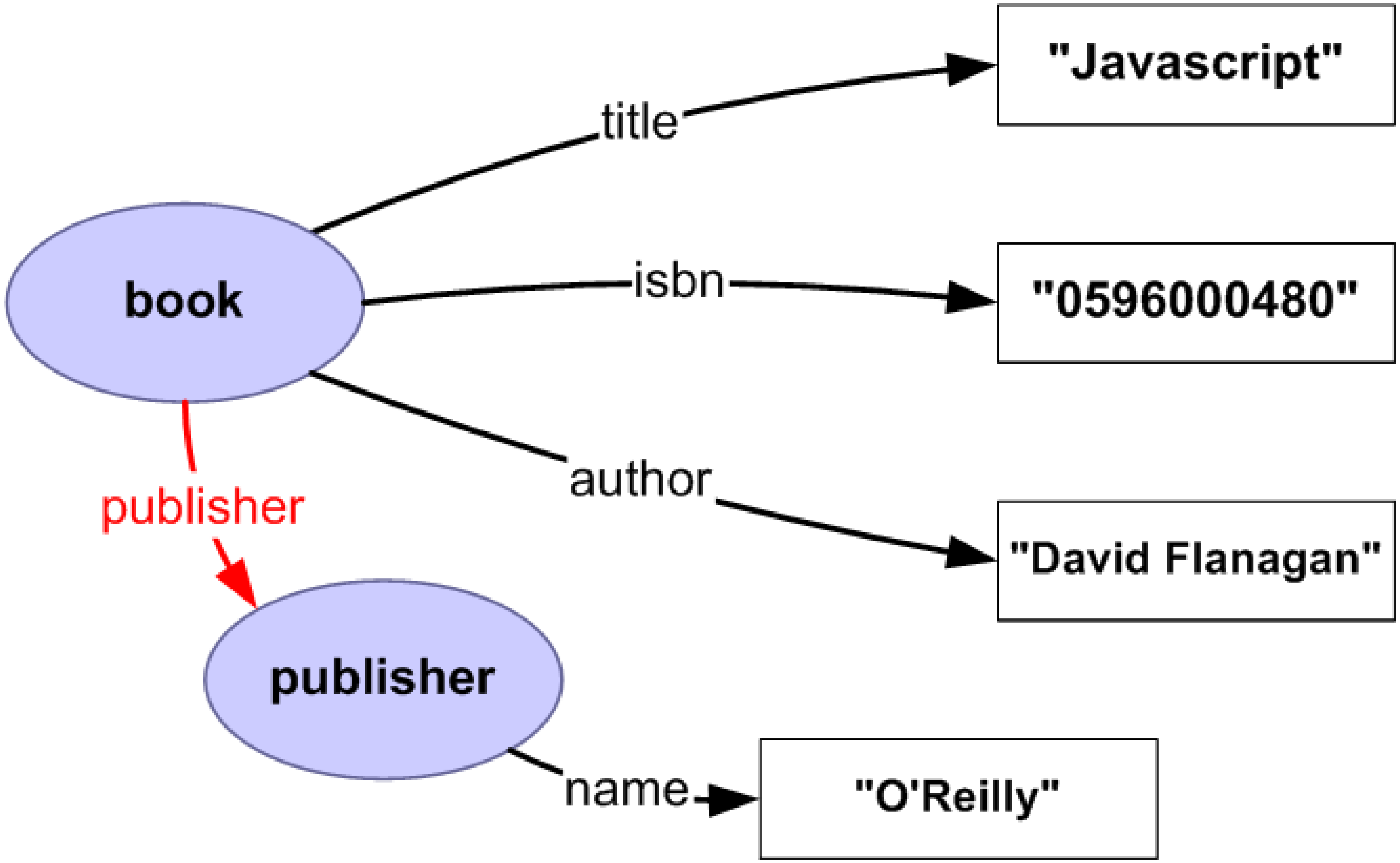


isbn	title	author	publisherID	Pages
0596002637	Practical RDF	Shelley Powers	7642	350
0596000480	JavaScript	David Flanagan	3556	936
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

multiple properties graphically



relations between “things”



graphs

Think in terms of graphs, not XML or documents

Nodes in graph are things, arcs are relationship between things

Using The Web's Infrastructure

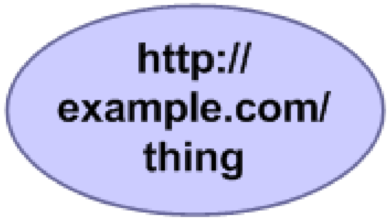
For a web-scale database we need to be able to identify things globally and uniquely

URIs (URLs) already provide those capabilities

Name things with URIs, specifically http://

This is the key to linked data

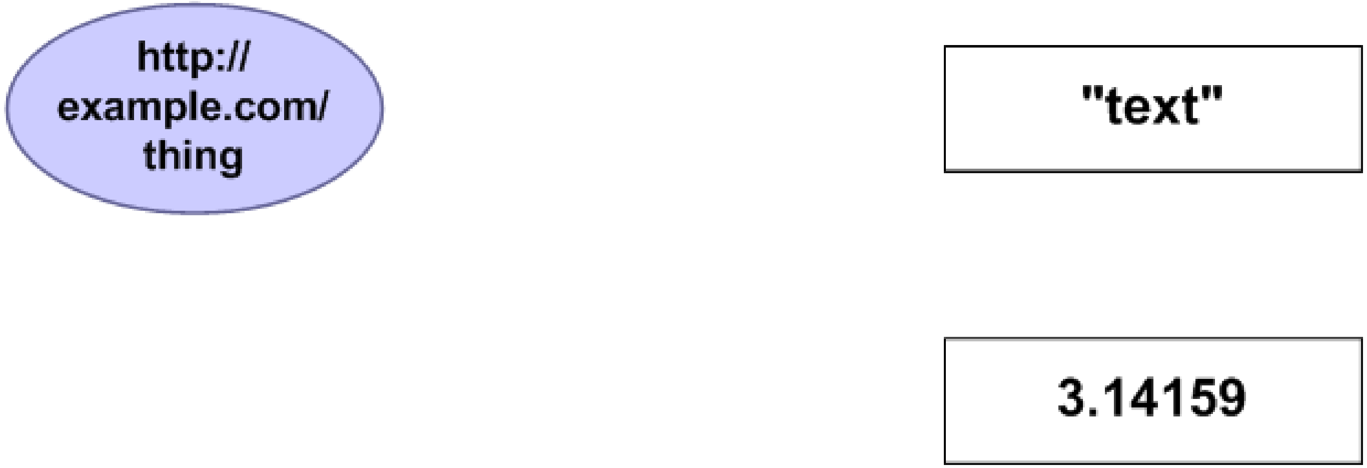
graphs can have named things (resources)



And text values (literals)

http://
example.com/
thing
"text"

and numeric values (also literals)

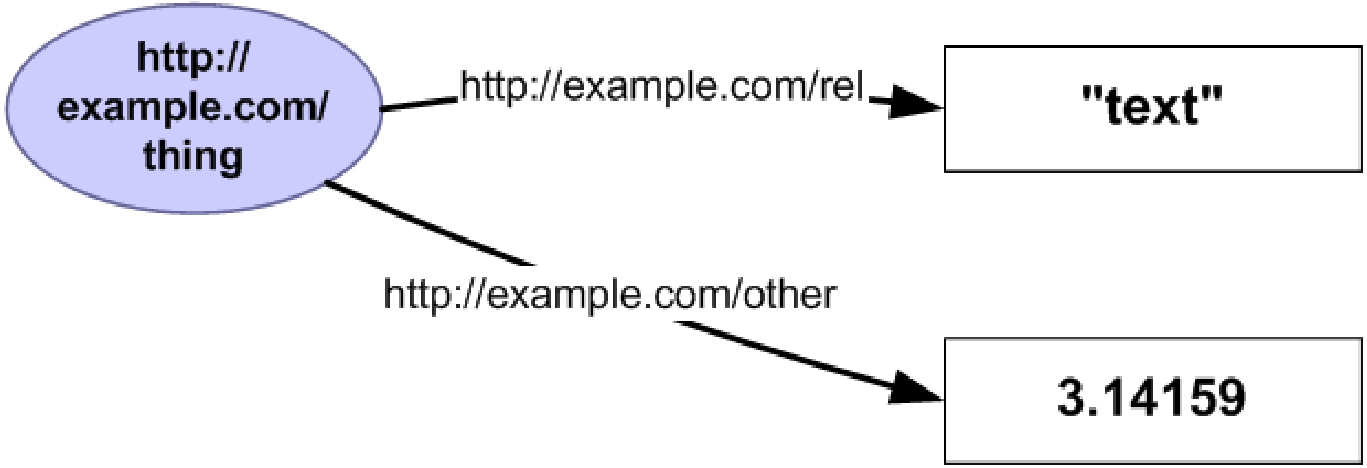


http://example.com/thing

"text"

3.14159

and named relations



Simple Rules

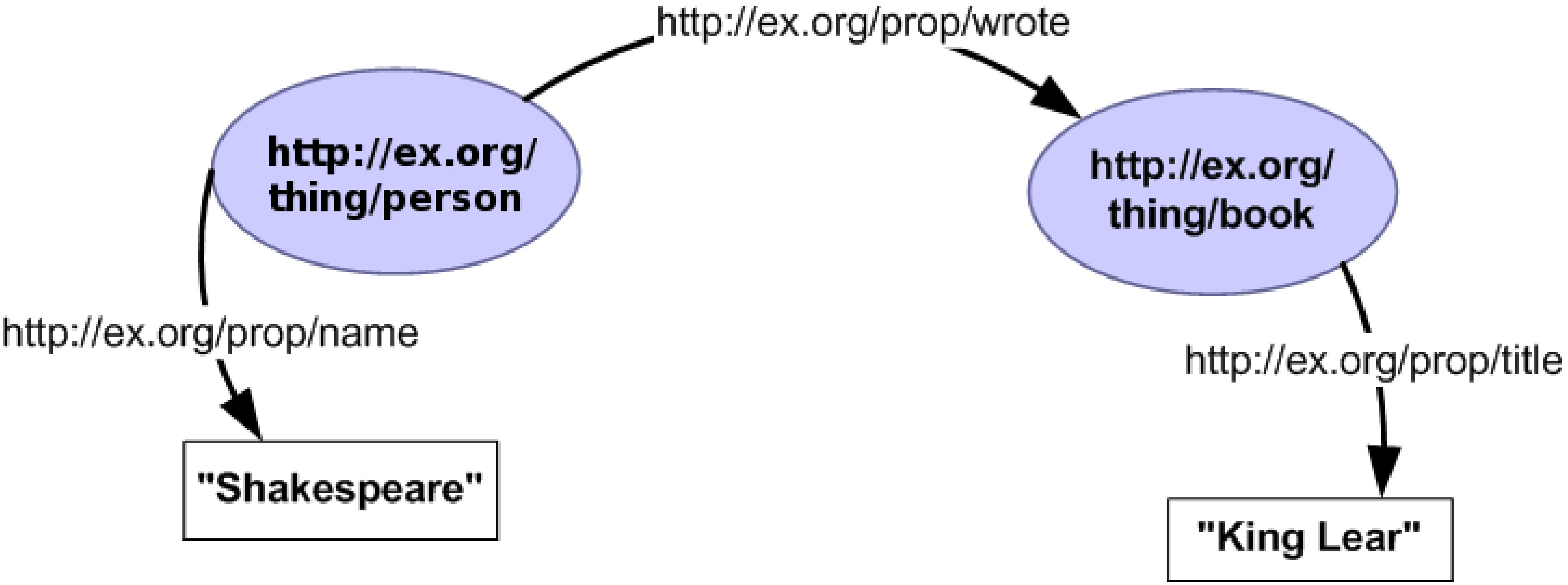
The URI identifies the thing you are describing

If two people create data using the same URI then they are describing the same thing

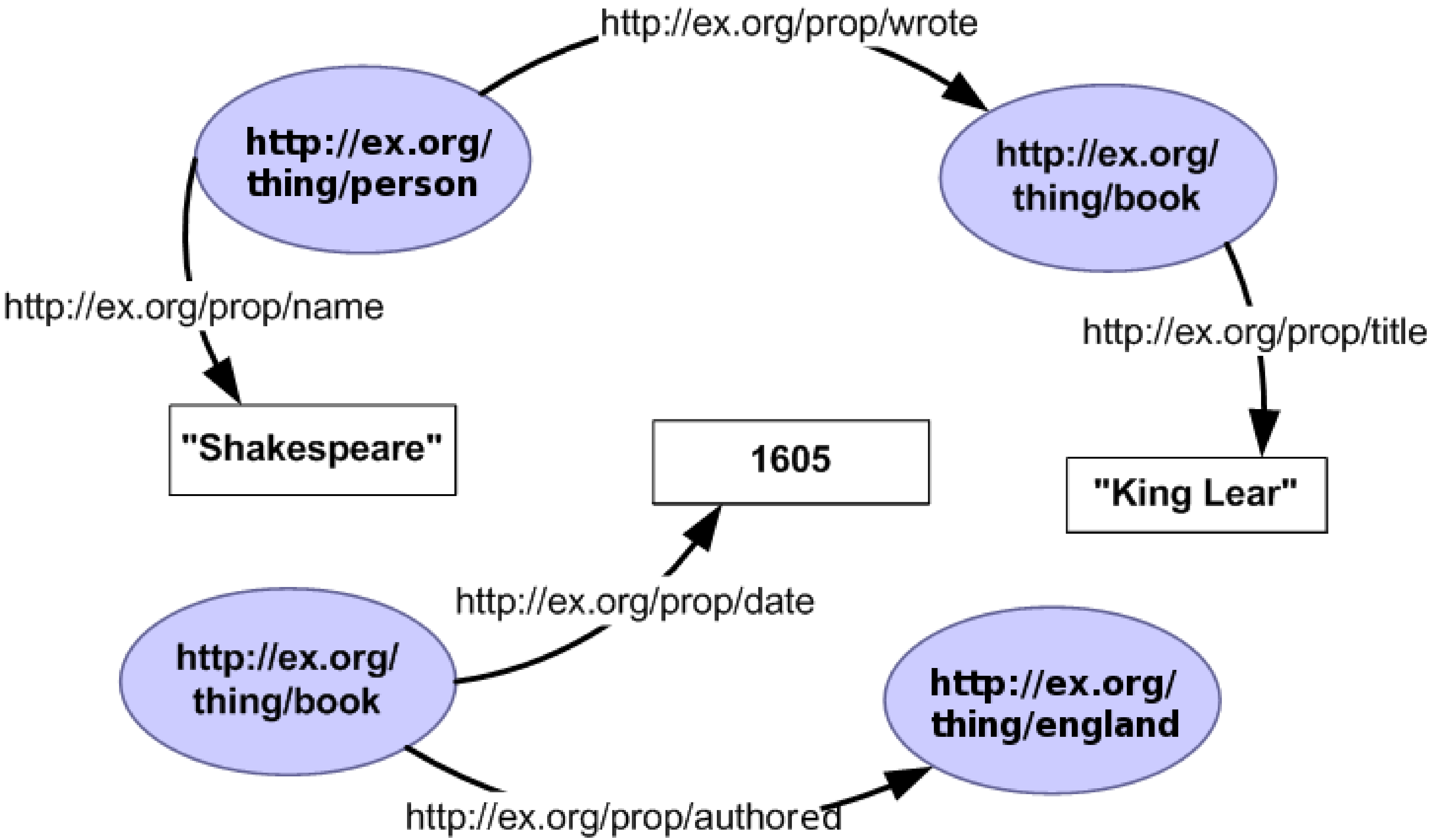
That makes it easy to merge data from different sources together

RDF data can use URIs from many different websites

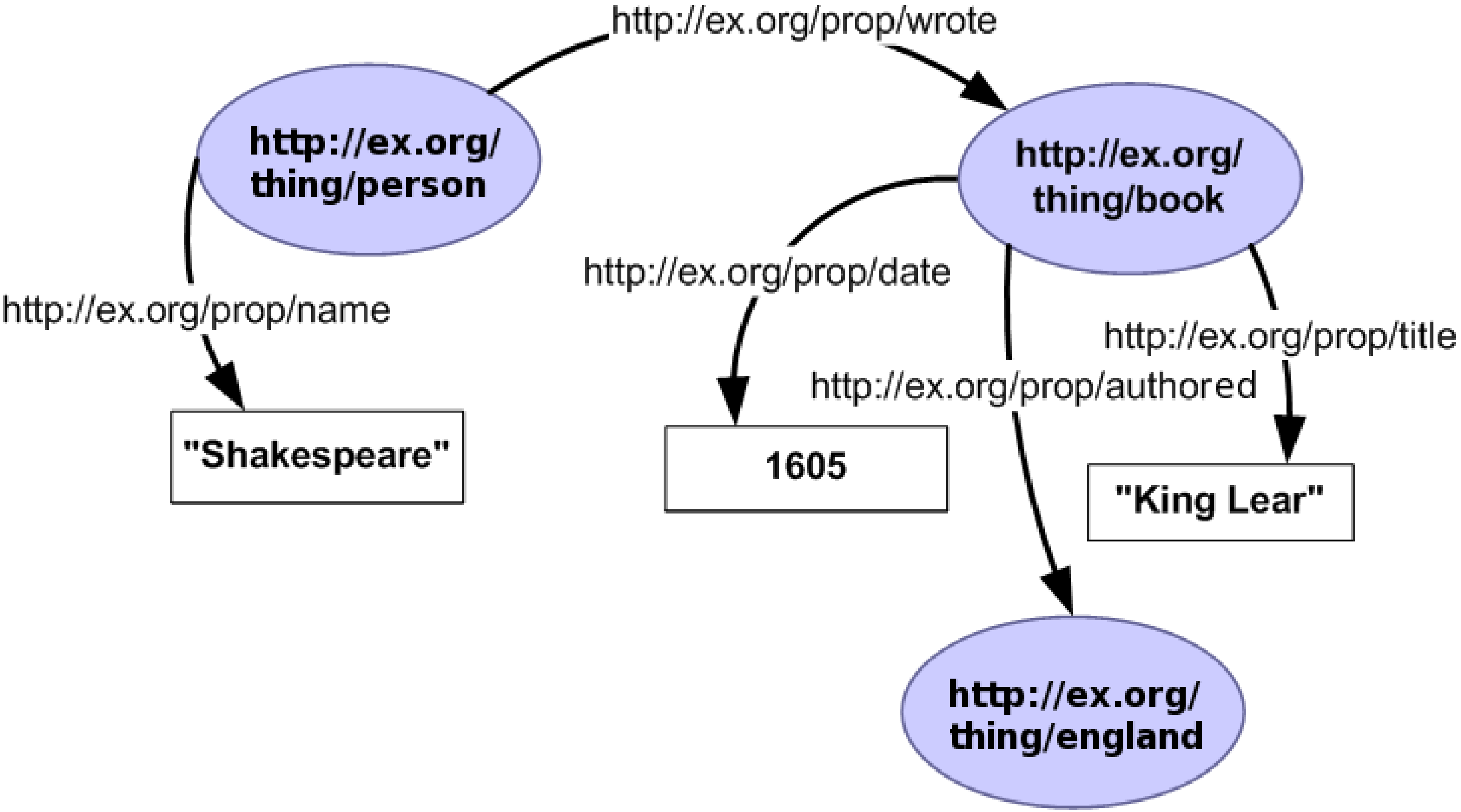
RDF MERGING 1 (GRAPH)



RDF MERGING 2 (GRAPH)



RDF MERGING 3 (GRAPH)



RDF MERGING 1 (XML)

```
<rdf:Description
rdf:about="http://ex.org/thing/person">
  <prop:name>Shakespeare</prop:name>
  <prop:wrote
rdf:resource="http://ex.org/thing/book"/>
</rdf:Description>

<rdf:Description
rdf:about="http://ex.org/thing/book">
  <prop:title>King Lear</prop:title>
</rdf:Description>
```

RDF MERGING 2 (XML)

```
<rdf:Description
rdf:about="http://ex.org/thing/person">
  <prop:name>Shakespeare</prop:name>
  <prop:wrote
rdf:resource="http://ex.org/thing/book"/>
</rdf:Description>

<rdf:Description
rdf:about="http://ex.org/thing/book">
  <prop:title>King Lear</prop:title>
</rdf:Description>

<rdf:Description
rdf:about="http://ex.org/thing/book">
```

RDF MERGING 3 (XML)

```
<rdf:Description
rdf:about="http://ex.org/thing/person">
  <prop:name>Shakespeare</prop:name>
  <prop:wrote
rdf:resource="http://ex.org/thing/book"/>
</rdf:Description>

<rdf:Description
rdf:about="http://ex.org/thing/book">
  <prop:title>King Lear</prop:title>
  <prop:date>1605</prop:date>
  <prop:authored
rdf:resource="http://ex.org/thing/england"/
>
```

URIs Are Active

This is the web - URIs can be retrieved

Fetch any URI from a graph to lookup more RDF about it

- ...which can be safely merged
- ...and may have other URIs which can be looked up

The result is a way to amass more information about things simply by following links

SPARQL

Once you have new data you can query it

SPARQL is the standard query language for linked data

It works by matching patterns in the graph

Getting Involved

As a data publisher you need to...

- convert some data to RDF
- publish it on your website like you do with HTML

As a data consumer you need to...

- learn one of the open source or commercial RDF toolkits
- find and retrieve the data you need
- use sparql to extract the useful bits for the task at hand

Linked Data Principles

Use URIs as names for things

Use HTTP URIs so people can look up those names

When someone looks up a URI, provide useful RDF information

Include RDF statements that link to other URIs so they can discover related things

Five Steps to Publishing Linked Data

1. Understand the Principles
2. Understand your Data
3. Choose URIs for Things in your Data
4. Set up Your Infrastructure
5. Link to other Data Sets

Step 1: understand the Principles

Use URIs as names for things

- anything, not just documents
- you are not your homepage
- information resources and non-information resources

Use HTTP URIs

- globally unique names, distributed ownership
- allows people to look up those names

Step 1: Understand the principles

Provide useful information in RDF

- when someone looks up a URI

Include RDF links to other URIs

- to enable discovery of related information

Step 2: Understand Your Data

What are the key things present in your data?

People

Places

Events

Books

Films

Musicians

Concepts

Photos

Comments

Reviews

Step 2: Understand Your Data

What vocabularies can be used to describe these?

FOAF, SIOC, Dublin Core, geo, SKOS, Review

Principles

- Reuse, don't reinvent
- Mix liberally

PART 3: Choosing URIs

Use HTTP URIs

Keep out of other peoples' namespaces

http://www.imdb.com/title/tt0441773/

    http://myfilms.com/tt0441773

Abstract away from implementation details

http://dbpedia.org/resource/Berlin

    http://www4.wiwiss.fu-berlin.de:2020/demos/dbpedia/cgi-bin/resources.php?id=Berlin

Hash or Slash?

http://mydomain.com/foaf.rdf#me

    http://mydomain.com/id/me

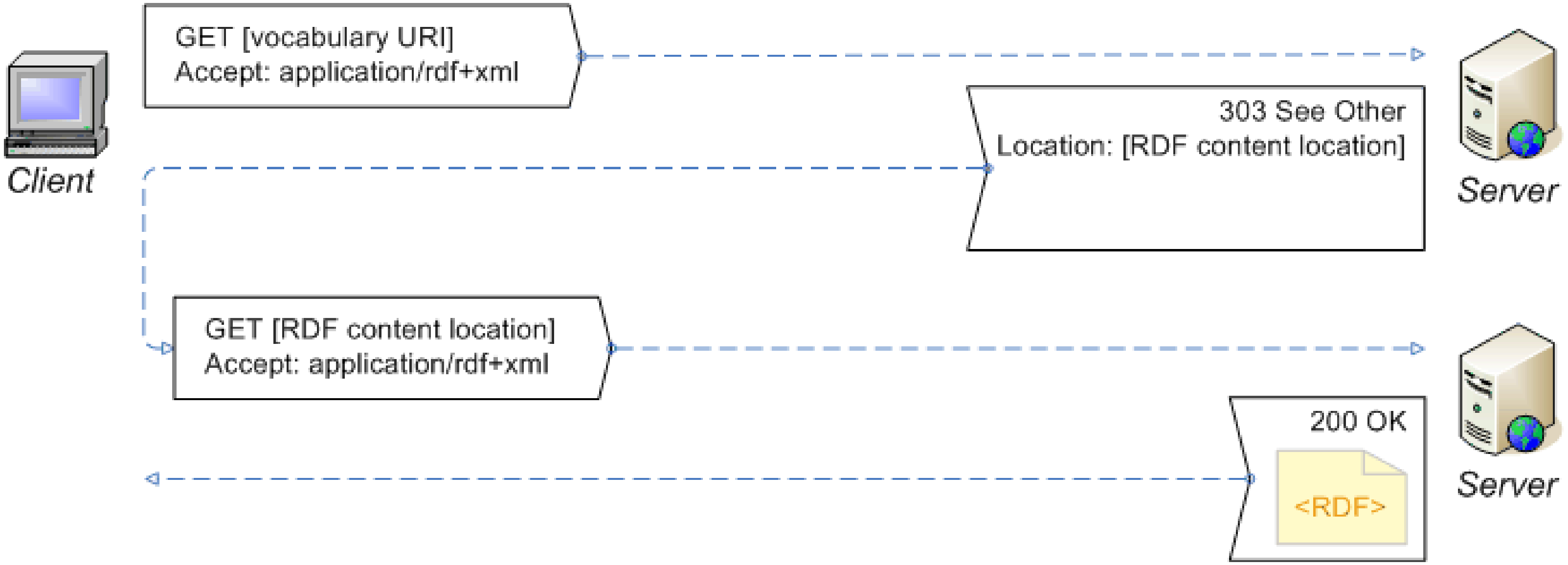
Part 3: Common Patterns FOR URIs

http://dbpedia.org/resource/New_York_City	← Thing
http://dbpedia.org/data/New_York_City	← RDF data
http://dbpedia.org/page/New_York_City	← HTML page

http://example.com/thing	← Thing
http://example.com/thing/rdf	← RDF data
http://example.com/thing/html	← HTML page

http://example.com/thing	← Thing
http://example.com/thing.rdf	← RDF data
http://example.com/thing.html	← HTML page

part 4: Setup infrastructire for Content Negotiation



Client

GET [vocabulary URI]
Accept: application/rdf+xml

303 See Other
Location: [RDF content location]

Server

GET [RDF content location]
Accept: application/rdf+xml

200 OK

<RDF>

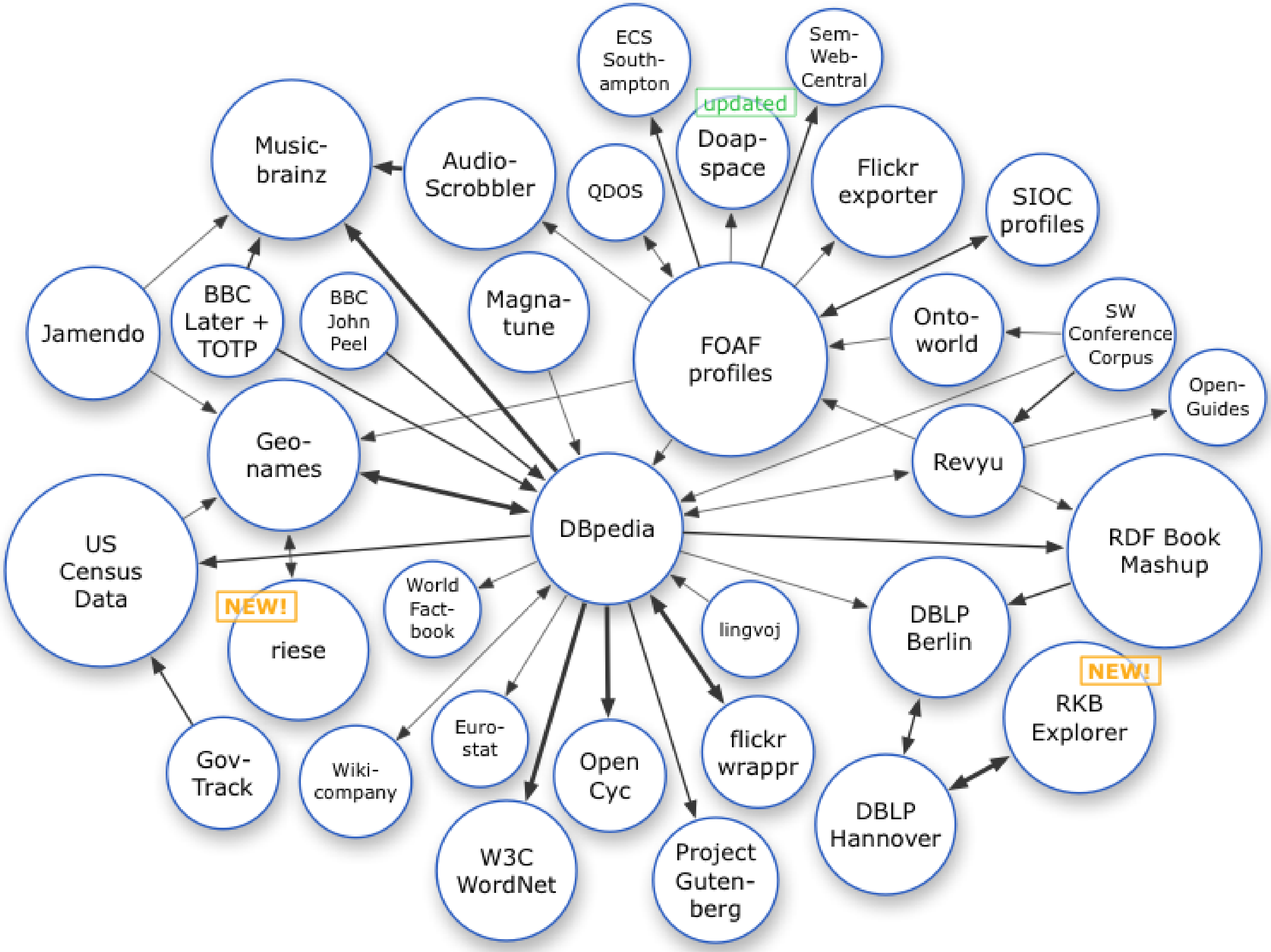
Server

PART 4: Setup Your Infrastructure

Testing your content negotiation

- Install the LiveHTTPHeaders and Modify Headers extensions for Firefox
- Try LiveHTTPHeaders against a URI
  - http://dbpedia.org/resource/Linked_Data
- Modify your headers to ask for application/rdf+xml
- What do you get back?
- Do the same with cURL
- http://dowhatimean.net/2007/02/debugging-semantic-web-sites-with-curl

part 5: Link to Other Available Data Sets



PART 5: Link to other Data Sets

Popular Predicates for Linking

owl:sameAs

foaf:homepage

foaf:topic

foaf:based_near

foaf:maker/foaf:made

foaf:depiction

foaf:page

foaf:primaryTopic

rdfs:seeAlso

part 5: Link to other Data Sets

Linking Algorithms

String Matching

- e.g. Lexical Distance between labels

Common Key Matching

- e.g. ISBN, Musicbrainz Ids

Property-based Matching

- Do these two things have the same label, type and coordinates?

Aim for reciprocal links

Summary

RDF is the data format for Linked Data

RDF uses URIs to name things, preferably HTTP

When a URI is fetched send a useful RDF description of the thing the URI names

RDF describes relationships between things

Link to other Data Sets