Semantic Web
Motivating Example
A Motivating example

• Here’s a motivating example, adapted from a presentation by Ivan Herman
• It introduces semantic web concepts
• And illustrates the benefits of representing your data using the semantic web techniques
• And motivates some of the semantic web technologies
We start with a book...
A simplified bookstore data

<table>
<thead>
<tr>
<th>ID</th>
<th>Author</th>
<th>Title</th>
<th>Publisher</th>
<th>Year</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Homepage</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_xyz</td>
<td>Ghosh, Amitav</td>
<td><a href="http://www.amitavghosh.com">http://www.amitavghosh.com</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Publisher’s name</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_qpr</td>
<td>Harper Collins</td>
<td>London</td>
</tr>
</tbody>
</table>
Export data as a set of relations

The Glass Palace

2000

London

Harper Collins

http://...isbn/000651409X

a:title

a:year

a:city

a:p_name

a:publisher

a:author

Ghosh, Amitav

http://www.amitavghosh.com

a:name

a:homepage
Notes on exporting the data

• Relations form a graph
  – Nodes refer to “real” data or some literal
  – We’ll defer dealing with the graph representation

• Data export doesn’t necessarily mean physical conversion of the data
  – relations can be generated on-the-fly at query time

• All of the data need not be exported
Same book in French...
# Bookstore data (dataset “F”)

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>Titre</th>
<th>Traducteur</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ID</td>
<td>Auteur</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ISBN 0-00-6511409-X</td>
<td>$A11$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Nom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>Ghosh, Amitav</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Besse, Christianne</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Export data as a set of relations

http://...isbn/000651409X

f:original

Le palais des miroirs

http://...isbn/2020386682

f:titre

Ghosh, Amitav

f:nom

Besse, Christianne

f:nom
Start merging your data
Merging your data

The Glass Palace
2000
London
Harper Collins
Ghosh, Amitav
http://www.amitavghosh.com
http://...isbn/00651409X

Le palais des miroirs
Besse, Christianne
http://...isbn/20386682

Same URI!
Start making queries...

• User of data “F” can now ask about the title of the original
• This information is not in the dataset “F”...
• ...but can be retrieved by merging with dataset “A”!
However, more can be achieved...

• Maybe *a:author* & *f:auteur* should be the same

• But an automatic merge doesn’t know that!

• Add extra information to the merged data:
  – *a:author same as f:auteur*
  – both identify a “Person”
  – Where *Person* is a term that may have already been defined, e.g.:
    • A “Person” is uniquely identified by a full name, homepage, Facebook page, Google+ page or email address
    • It can be used as a “category” for certain type of resources
Use this extra knowledge
This enables richer queries

• User of dataset “F” can now query:
  – “donnes-moi la page d’accueil de l’auteur de l’original”
    • well... “give me the home page of the original’s ‘auteur’”
• The information is not in datasets “F” or “A”...
• ...but was made available by:
  – Merging datasets “A” and datasets “F”
  – Adding three simple extra statements
  – Inferring the consequences
Combine with different datasets

• Using, e.g., the “Person”, the dataset can be combined with other sources
• For example, data in Wikipedia can be extracted using dedicated tools
  – e.g., the “DBpedia” project can extract the “infobox” information from Wikipedia already...
Merge with Wikipedia data

The Glass Palace
2000
London
Harper Collins
Ghosh, Amitav
http://www.amitavghosh.com
http://dbpedia.org/..../Amitav_Ghosh

http://...isbn/000651409X

Le palais des miroirs
http://...isbn/2020386682

Besse, Christianne
http://...foaf/Person

The Glass Palace
2000
London
Harper Collins
Ghosh, Amitav
http://www.amitavghosh.com
http://dbpedia.org/..../Amitav_Ghosh

http://...isbn/000651409X

Le palais des miroirs
http://...isbn/2020386682

Besse, Christianne
Is that surprising?

- It may look like it but, in fact, it should not be...
- What happened via automatic means is done every day by human Web users!
- What is needed is a way to let machines decide when classes, properties and individuals are the same or different
This can be even more powerful

• Add extra knowledge to the merged datasets
  – e.g., a full classification of various types of library data
  – geographical information
  – etc.

• This is where ontologies, rules, etc., come in
  – ontologies/rule sets can be relatively simple and small, or huge, or anything in between...

• Even more powerful queries can be asked as a result
So where is the Semantic Web?

The Semantic Web provides technologies to make such integration possible!

Key integration datasets, like DBpedia, have emerged