



# Semantic Web and Semantic Search

Vasant Honavar

Artificial Intelligence Research Laboratory

Department of Computer Science

Center for Computational Intelligence, Learning, & Discovery

Iowa State University

[honavar@cs.iastate.edu](mailto:honavar@cs.iastate.edu)

[www.cs.iastate.edu/~honavar/](http://www.cs.iastate.edu/~honavar/)

[www.cs.iastate.edu/~honavar/aigroup.html](http://www.cs.iastate.edu/~honavar/aigroup.html)

# Disclaimer

- Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not reflect the views of the National Science Foundation.

# Syntactic web vs Semantic web

- Syntactic web
  - Computers do the presentation (easy) and
  - People do the linking and interpreting (hard)
- Semantic web
  - Why not get computers to do more of the hard work?

[Goble 03]

# Working with the Syntactic Web

- Semantic search
- Complex queries involving **background knowledge**
  - Find information about “animals that use sonar but are not either bats, dolphins or whales”
- Integrating information from multiple sources
  - Book me a holiday next weekend somewhere warm, not too far away, and where they speak French or English

Hopeless for machines and tedious for people

# Looking for a "Blue Car with Red Doors"



Red Car with  
Blue Doors



Navy Sedan with  
Crimson Hatches



Blue Car and  
a Red Door



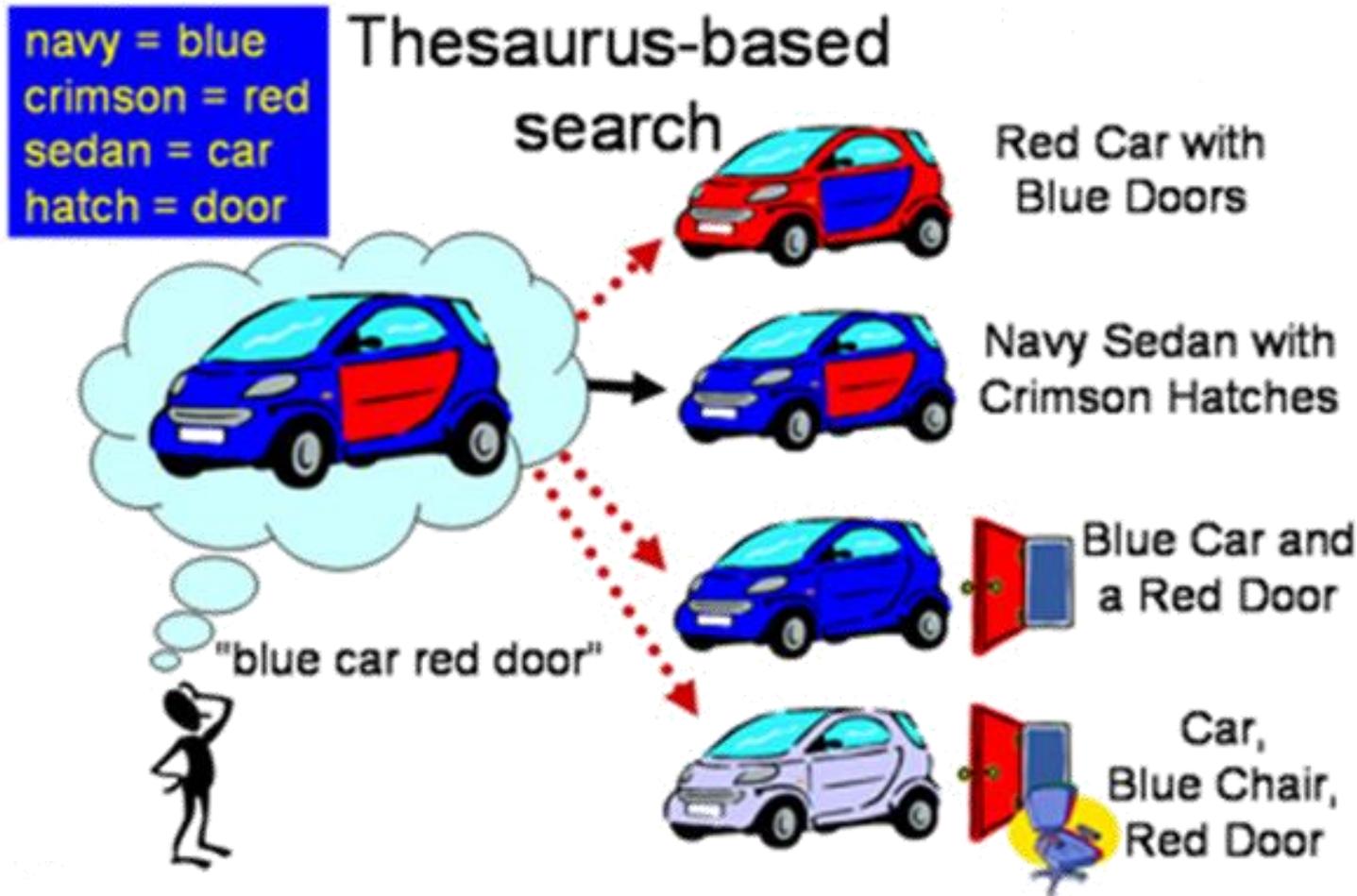
Car,  
Blue Chair,  
Red Door

Steven Kraines, Science Integration Project – Human, The University of Tokyo

# Simple word-matching



Steven Kraines, Science Integration Project – Human, The University of Tokyo



Steven Kraines, Science Integration Project – Human, The University of Tokyo

navy = blue  
crimson = red  
sedan = car  
hatch = door

# Semantic Matching



Red Car with Blue Doors



?



Navy Sedan with Crimson Hatches

"blue car red door"



Blue Car and a Red Door



hasColor(car, blue)  
hasColor(door, red)  
hasPart(car, door)



Car, Blue Chair, Red Door

Steven Kraines, Science Integration Project – Human, The University of Tokyo

navy = blue  
crimson = red  
sedan = car  
hatch = door

# Semantic Matching



"blue car red door"



hasColor(car, blue)  
hasColor(door, red)  
hasPart(car, door)



Red Car with Blue Doors

hasColor(car, red)  
hasColor(door, blue)  
hasPart(car, door)



Navy Sedan with Crimson Hatches

hasColor(sedan, navy)



Blue Car and a Red Door

hasColor(car, red)  
hasColor(door, blue)



Car, Blue Chair, Red Door

hasColor(chair, blue)

Steven Kraines, Science Integration Project – Human, The University of Tokyo

# Semantic Web

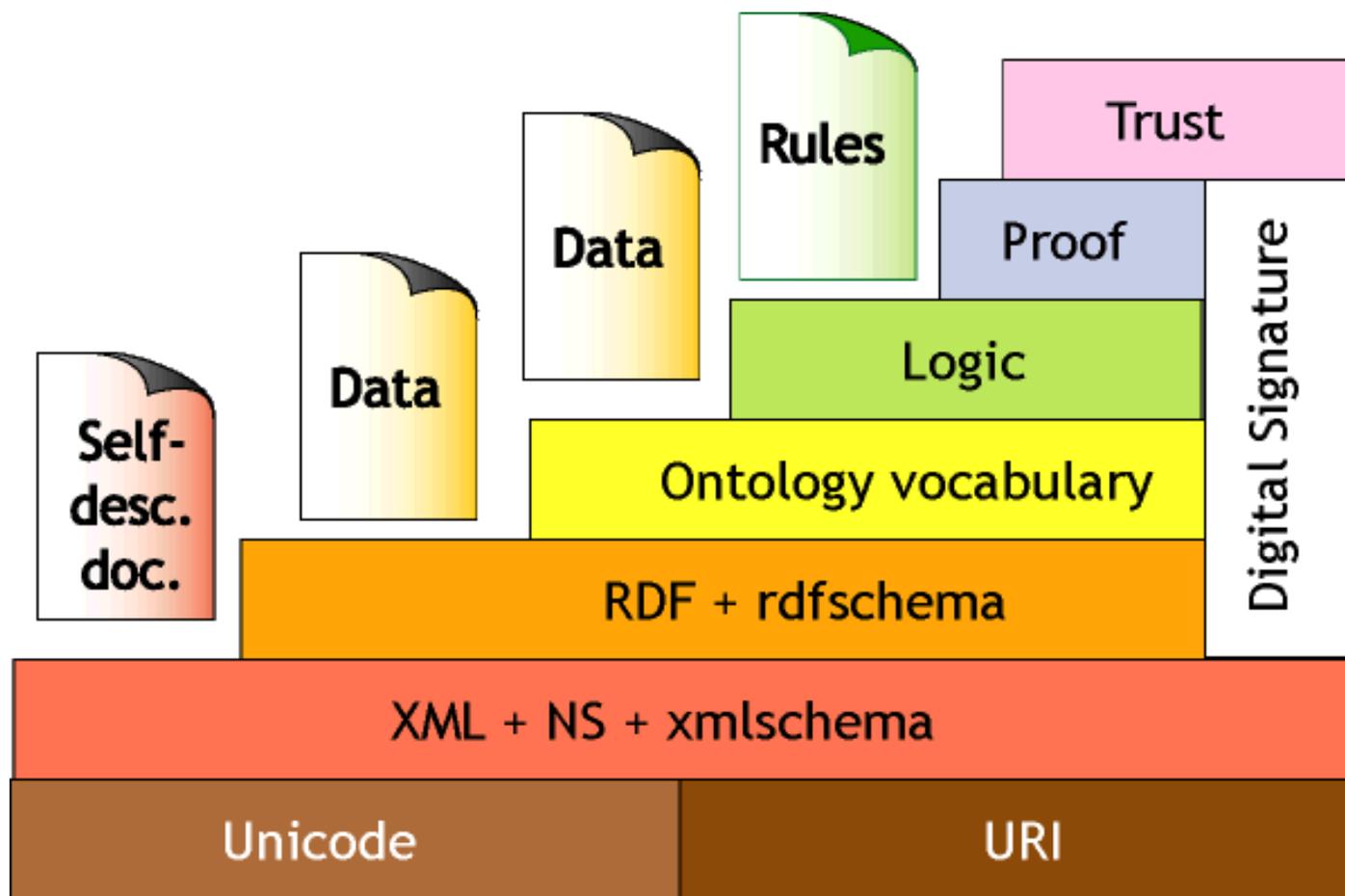
“The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications”

<http://www.w3.org/sw/>



[Berners-Lee, Lassila, Hendler, 2001]

## Tim Berners-Lee's Vision of Semantic Web (IJCAI-01)



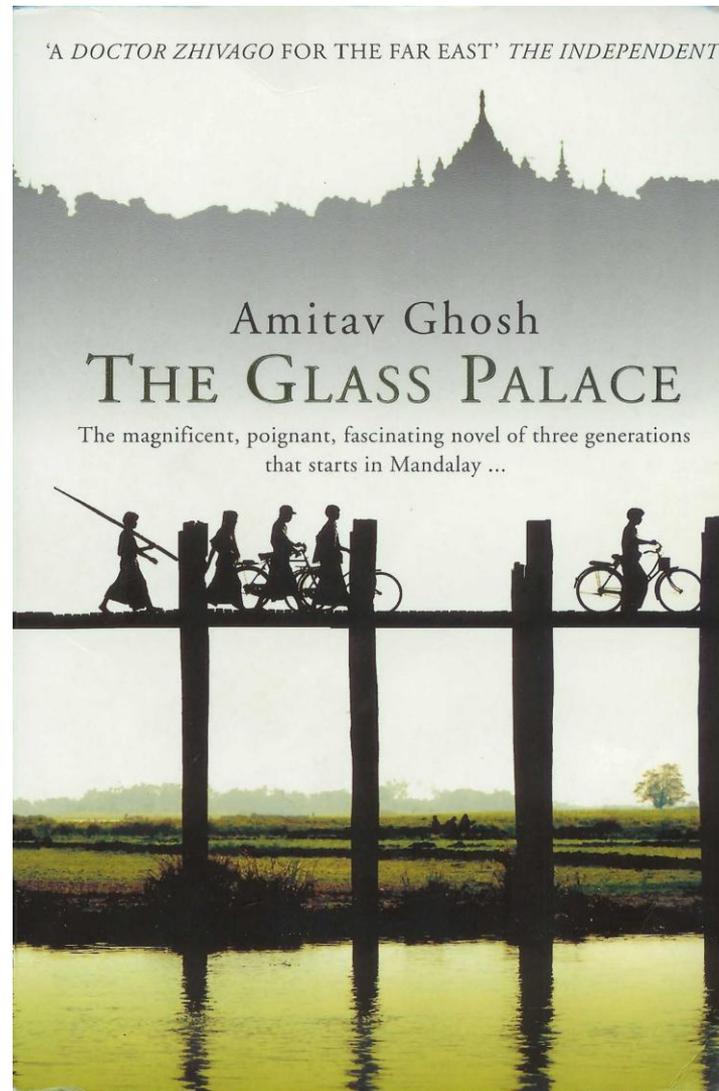
# Semantic Web

- Expose data on the web in an interoperable form (RDF)
- Expose knowledge on the web with interoperable semantics (ontologies: RDF Schema, OWL)
- Apply (lightweight) inference for
  - Searching for information
  - Answering complex queries
  - Integrating multiple sources of knowledge and data
  - Composing composite services from component services
  - Learning predictive models from disparate data sources
  - Unexpected reuse

# Data integration

- Represent data in RDF stores
- Combine data from RDF stores
- Query the integrated RDF data

# A book in English



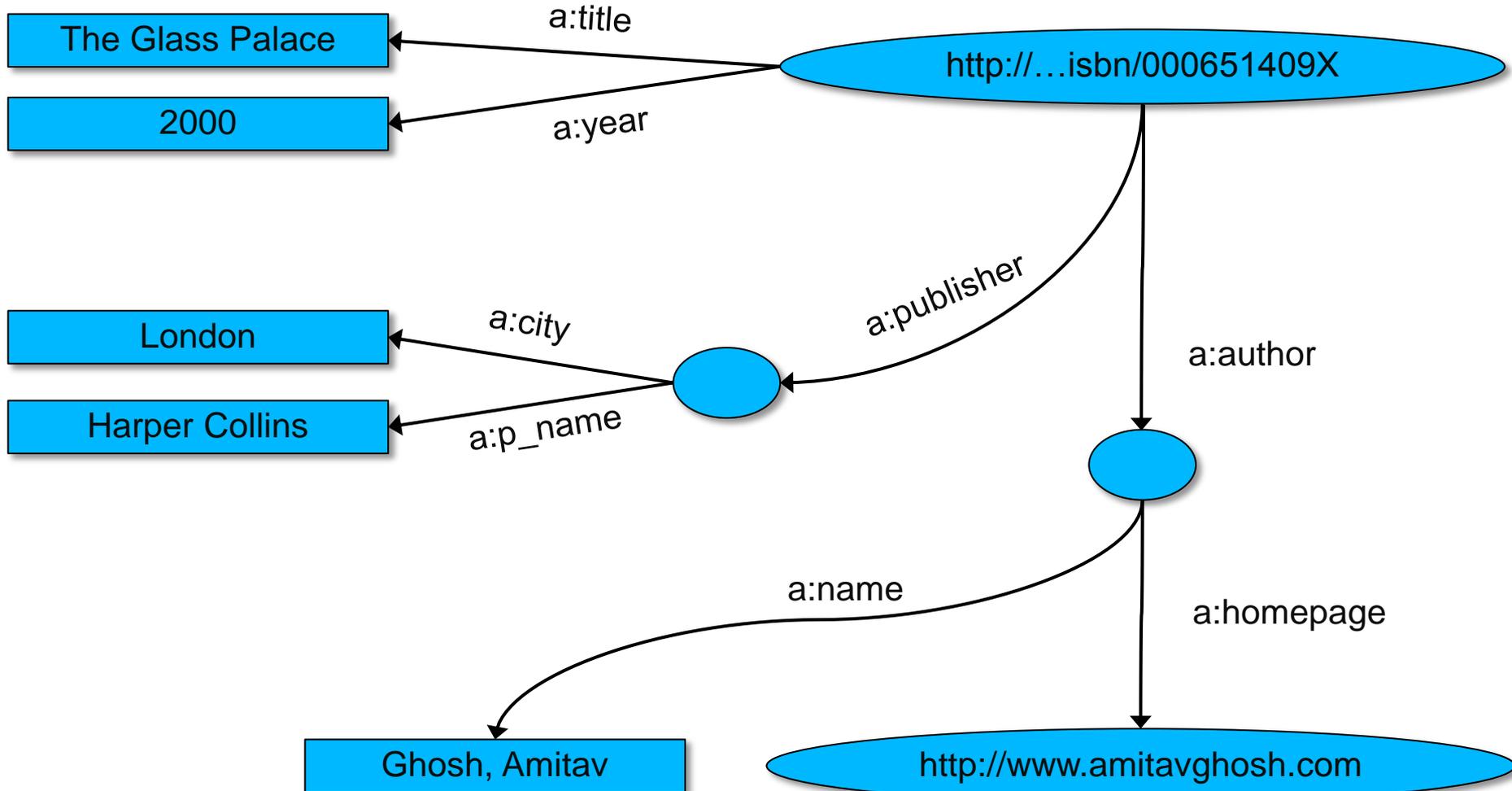
[Ivan Herman, W3C]

## A simplified bookstore data (dataset “A”)

<b>ID</b>	<b>Author</b>	<b>Title</b>	<b>Publisher</b>	<b>Year</b>
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

<b>ID</b>	<b>Name</b>	<b>Homepage</b>
id_xyz	Ghosh, Amitav	<a href="http://www.amitavghosh.com">http://www.amitavghosh.com</a>

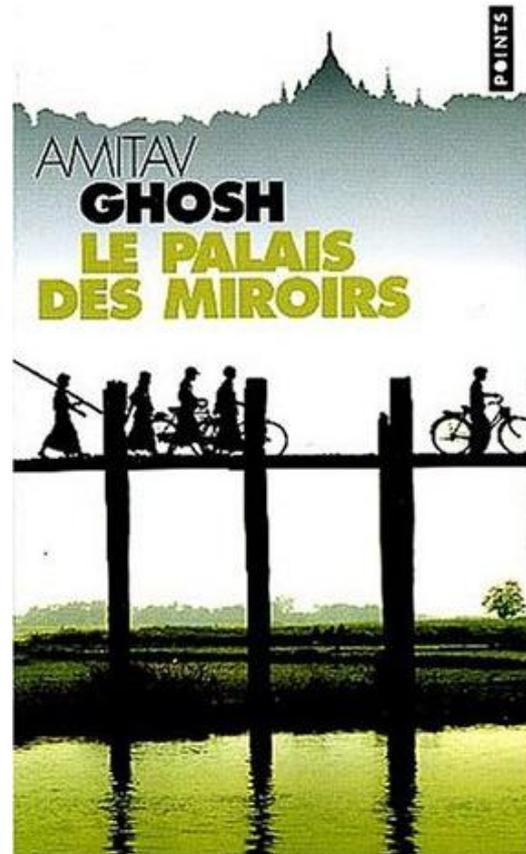
<b>ID</b>	<b>Publisher's name</b>	<b>City</b>
id_qpr	Harper Collins	London

1<sup>st</sup>: export data as a set of relations

## Some notes on the exporting the data

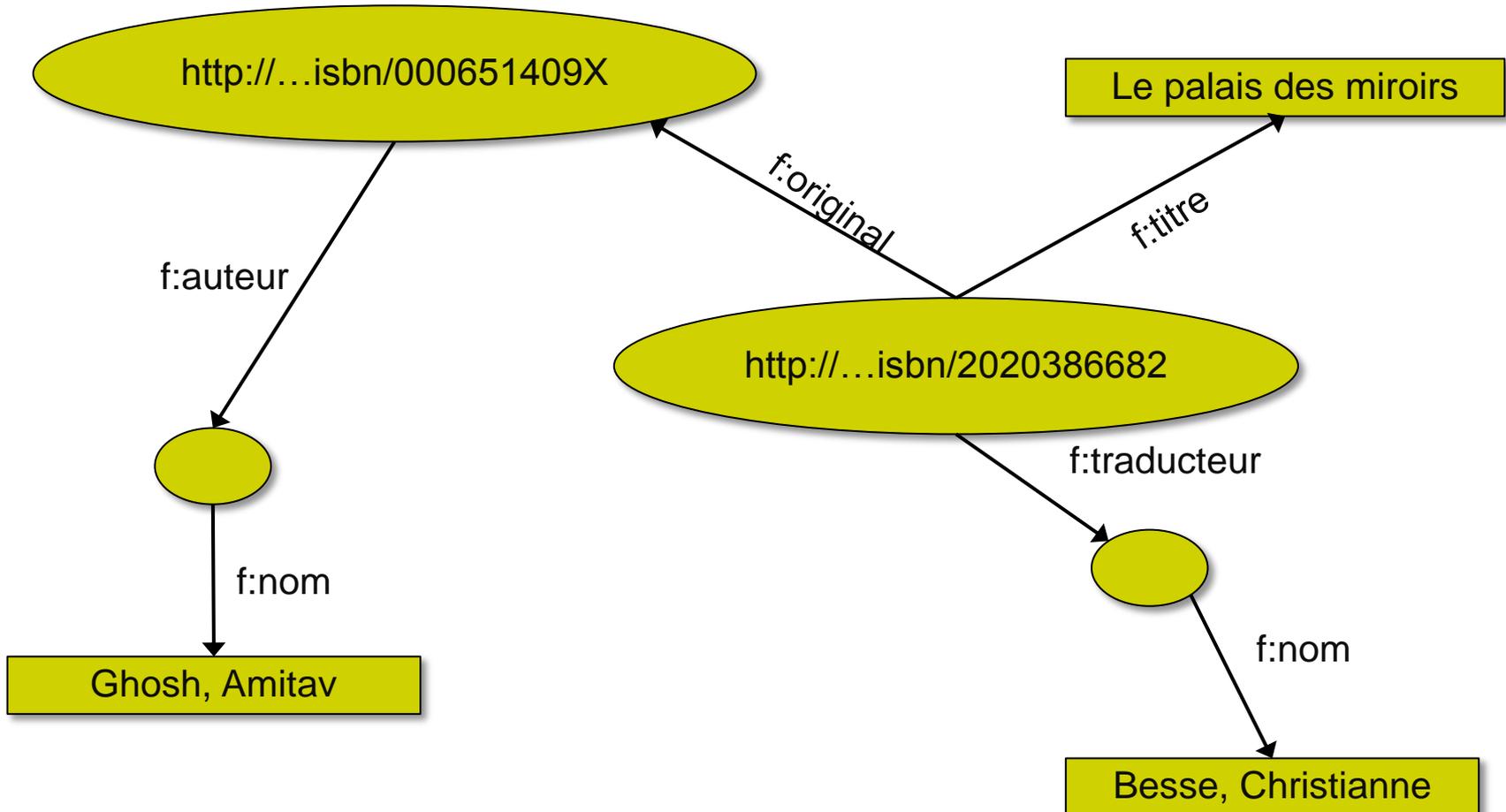
- Relations form a graph
  - the nodes refer to the “real” data or contain some literal
  - how the graph is represented in the machine is immaterial for now
- Data export does not necessarily mean physical conversion of the data
  - Relations can be generated on-the-fly at query time
    - via SQL “bridges”
    - scraping HTML pages
    - extracting data from Excel sheets
    - Extracting data from text
- One can export part of the data

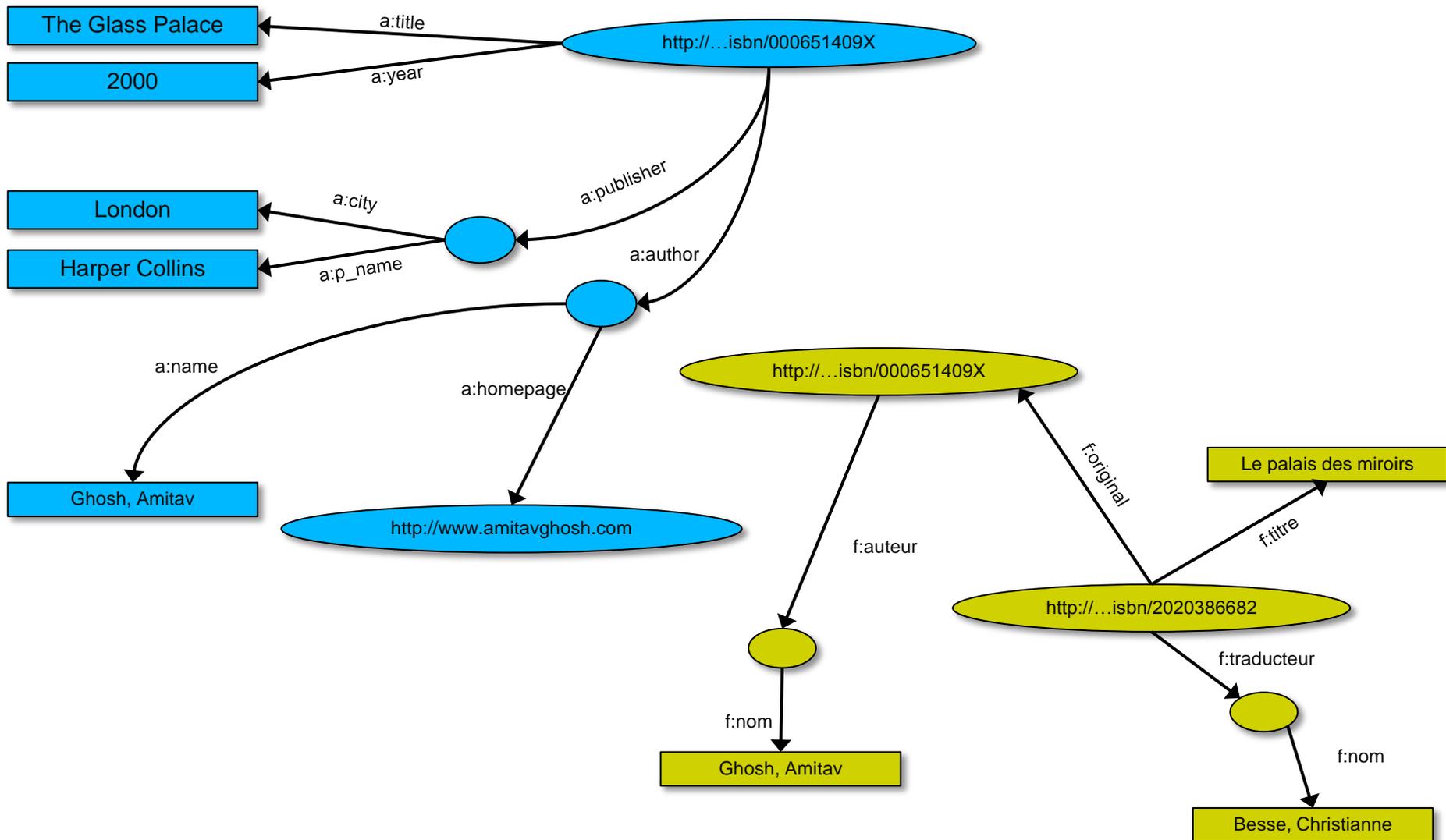
The same book in French...

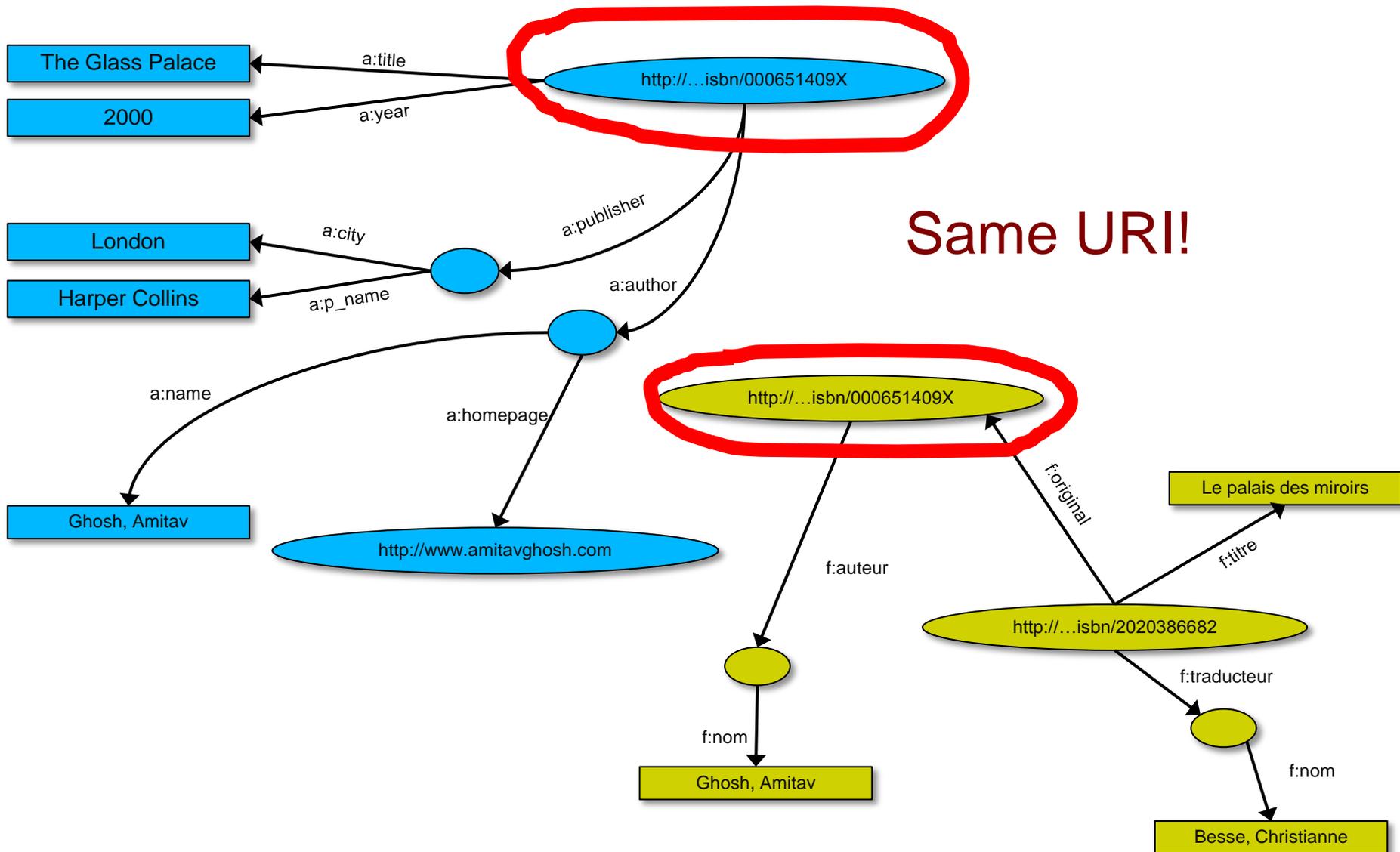


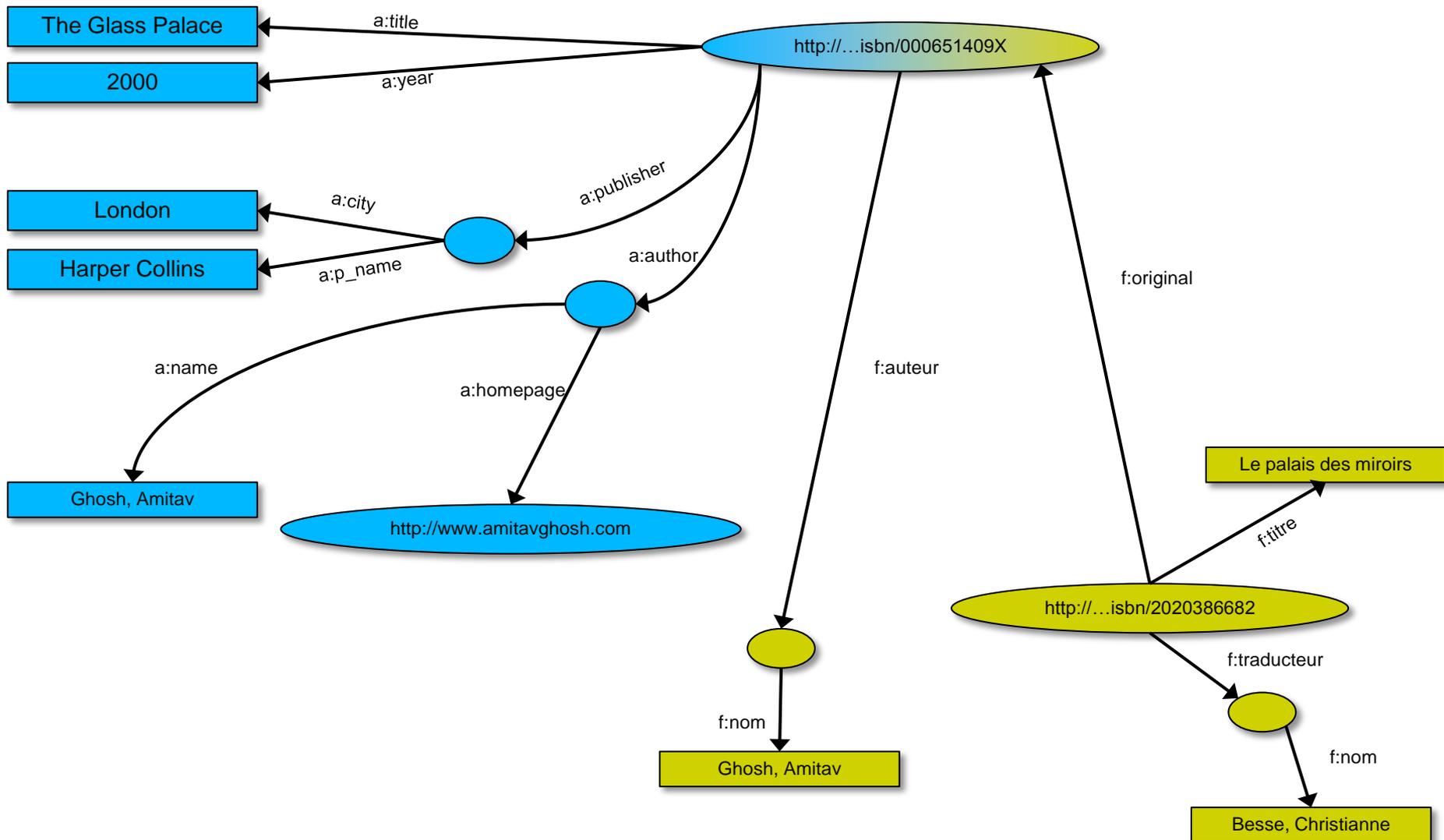
## A second bookstore data (dataset "F")

A	B	C	D
1	<b>ID</b>	<b>Titre</b>	<b>Traducteur</b>
2	ISBN 2020286682	Le Palais des Miroirs	ISBN 0-00-6511409-X
3			
4			
5			
6	<b>ID</b>	<b>Auteur</b>	
7	ISBN 0-00-6511409-X	\$A11\$	
8			
9			
10	<b>Nom</b>		
11	Ghosh, Amitav		
12	Besse, Christianne		

2<sup>nd</sup>: export the second set of data

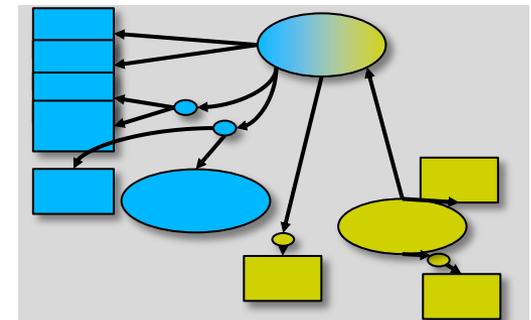
3<sup>rd</sup>: merge your data

3<sup>rd</sup>: merge your data (cont)

3<sup>rd</sup>: merge your data

# Start making queries...

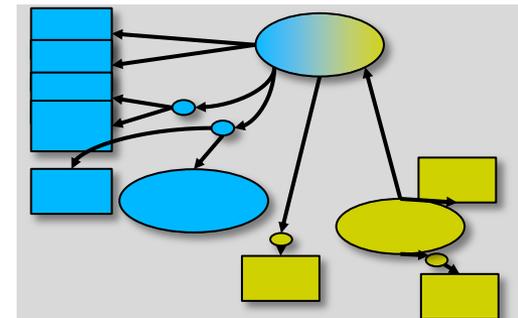
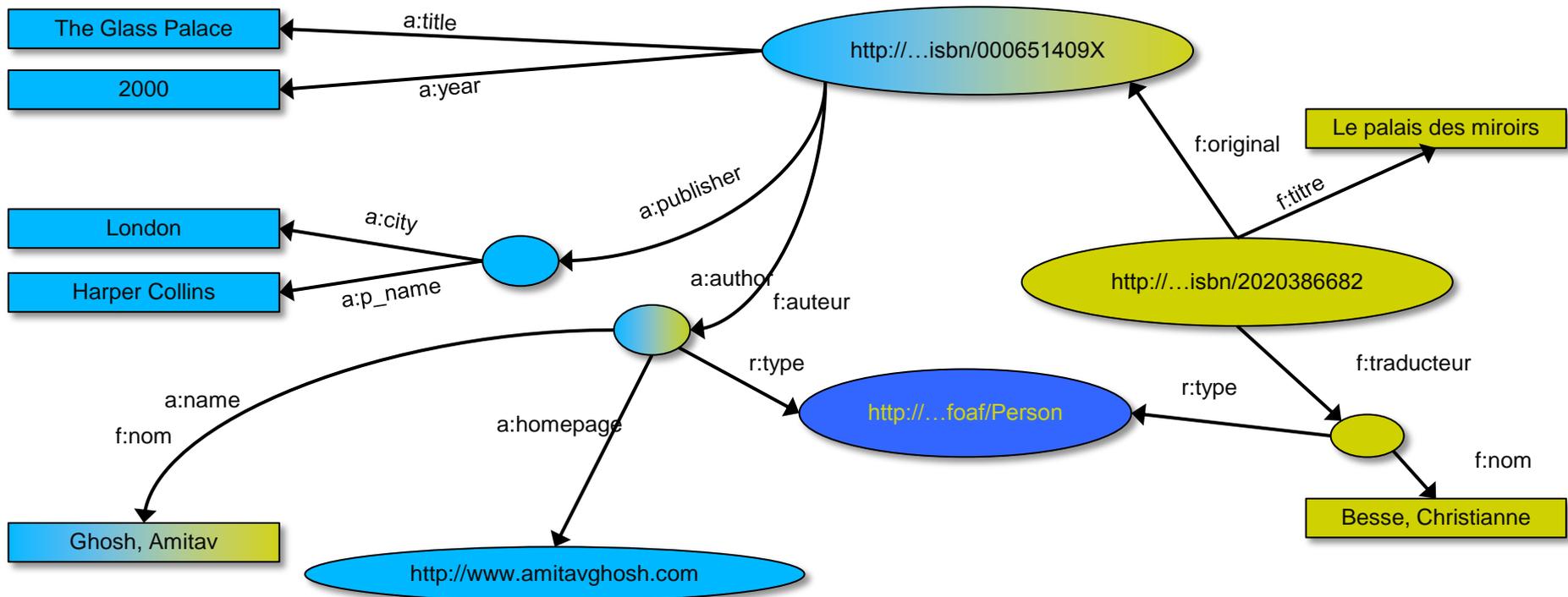
- User of data “F” can now ask queries like:
  - “give me the title of the original”
    - well, ... « donnes-moi le titre de l’original »
- This information is not in the dataset “F” ...
- ...but can be retrieved by merging with dataset “A”!



## There is more we can do!

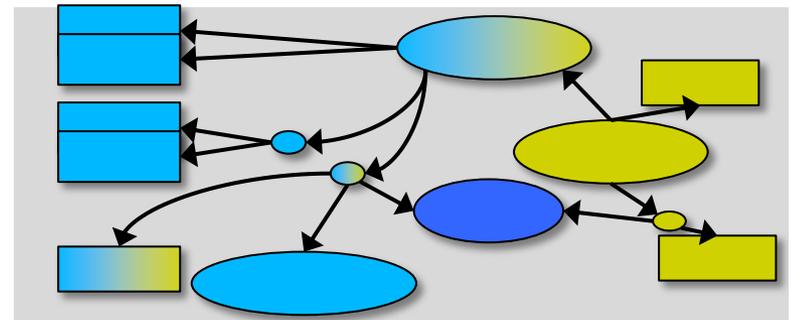
- We “think” that a:author and f:auteur should be the same
- But an automatic merge does not know that!
- Let us add some extra information to the merged data:
  - a:author same as f:auteur
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources

# Querying revisited: use the extra knowledge



## Start making richer queries!

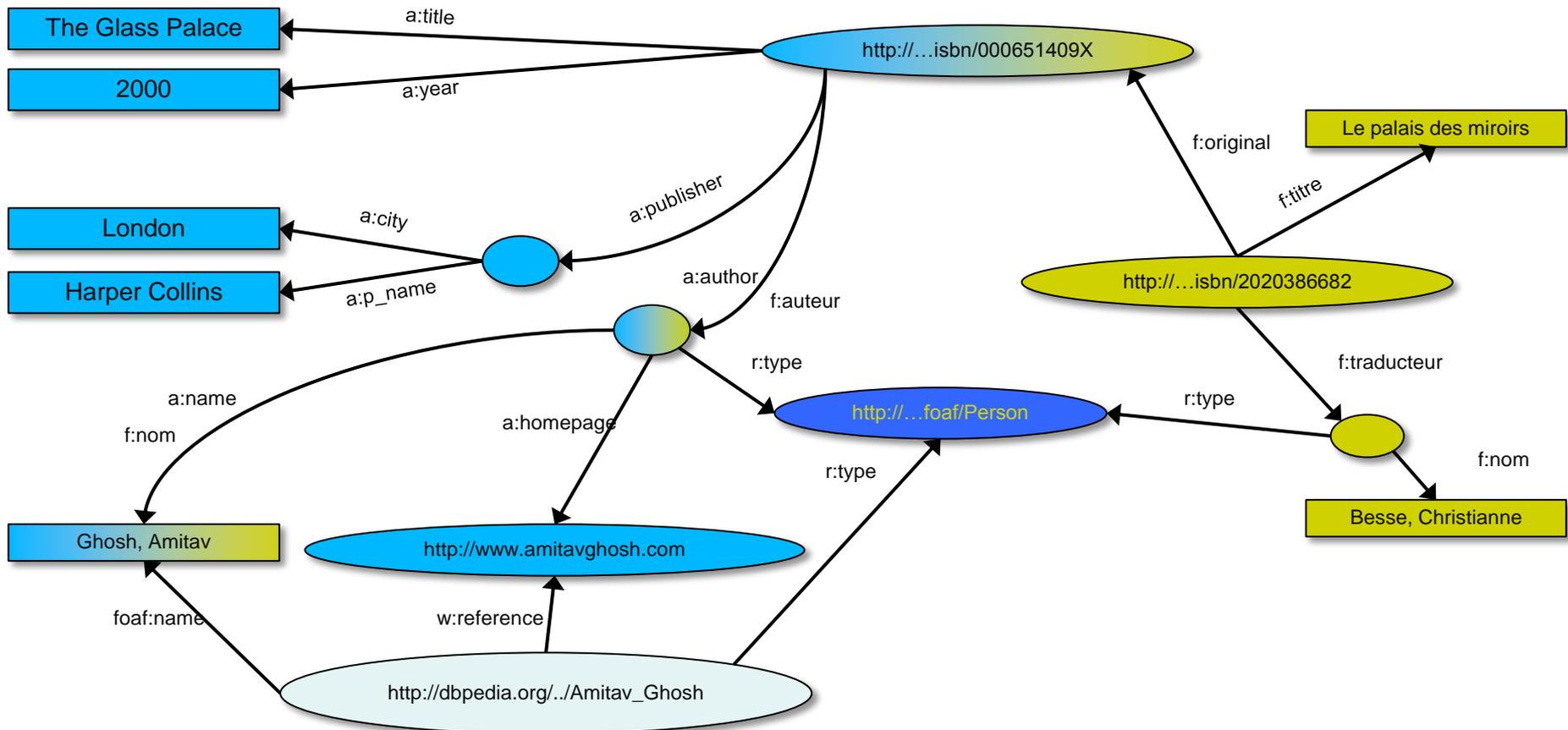
- User of dataset “F” can now query:
  - “donnes-moi la page d’accueil de l’auteur de l’original”
    - well... “give me the home page of the original’s ‘auteur’”
- The information is not in datasets “F” or “A” but was made available by:
  - merging dataset “A” and dataset “F”
  - adding three simple extra statements as an extra “glue”



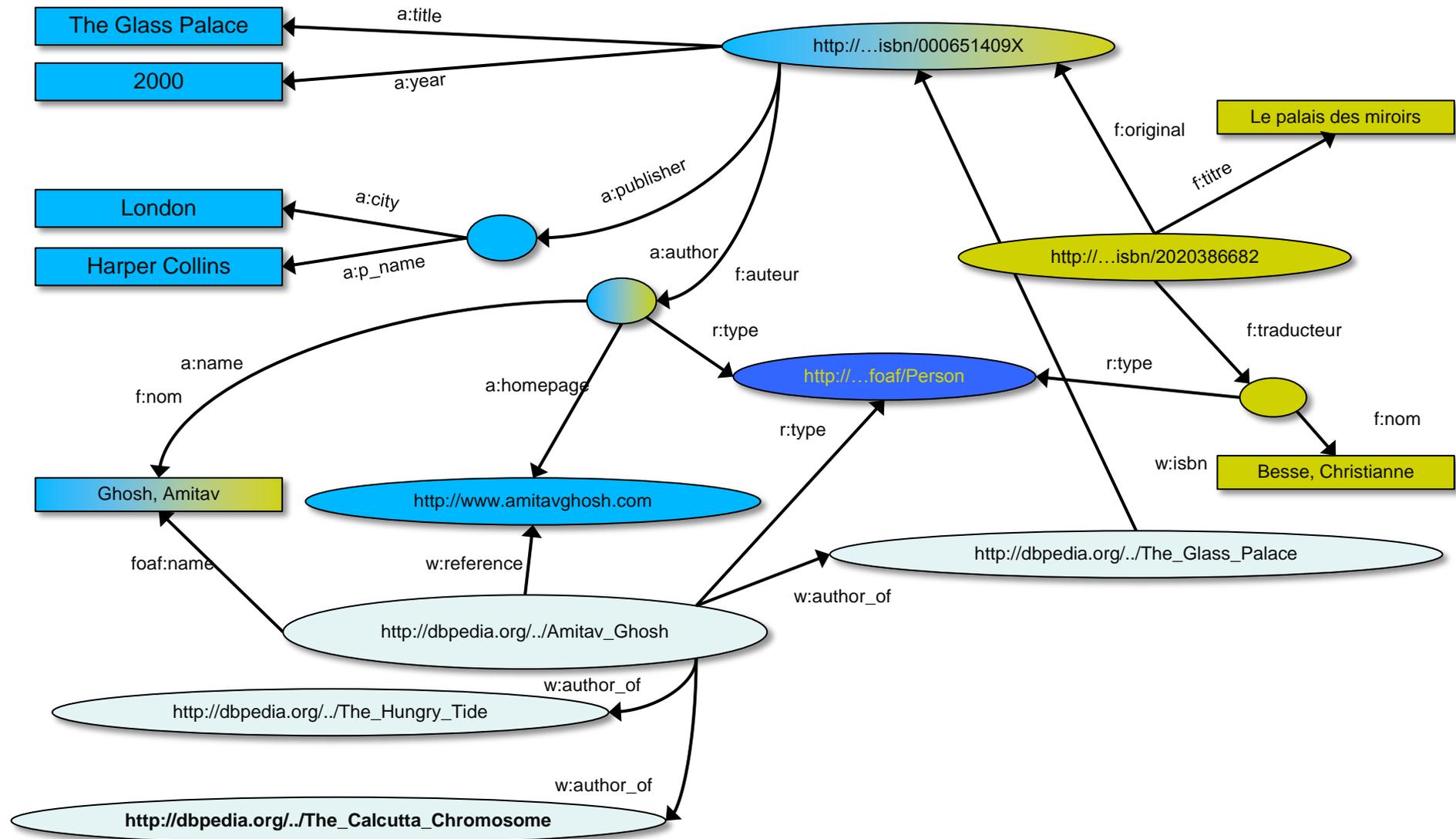
## Combine with different datasets

- 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...

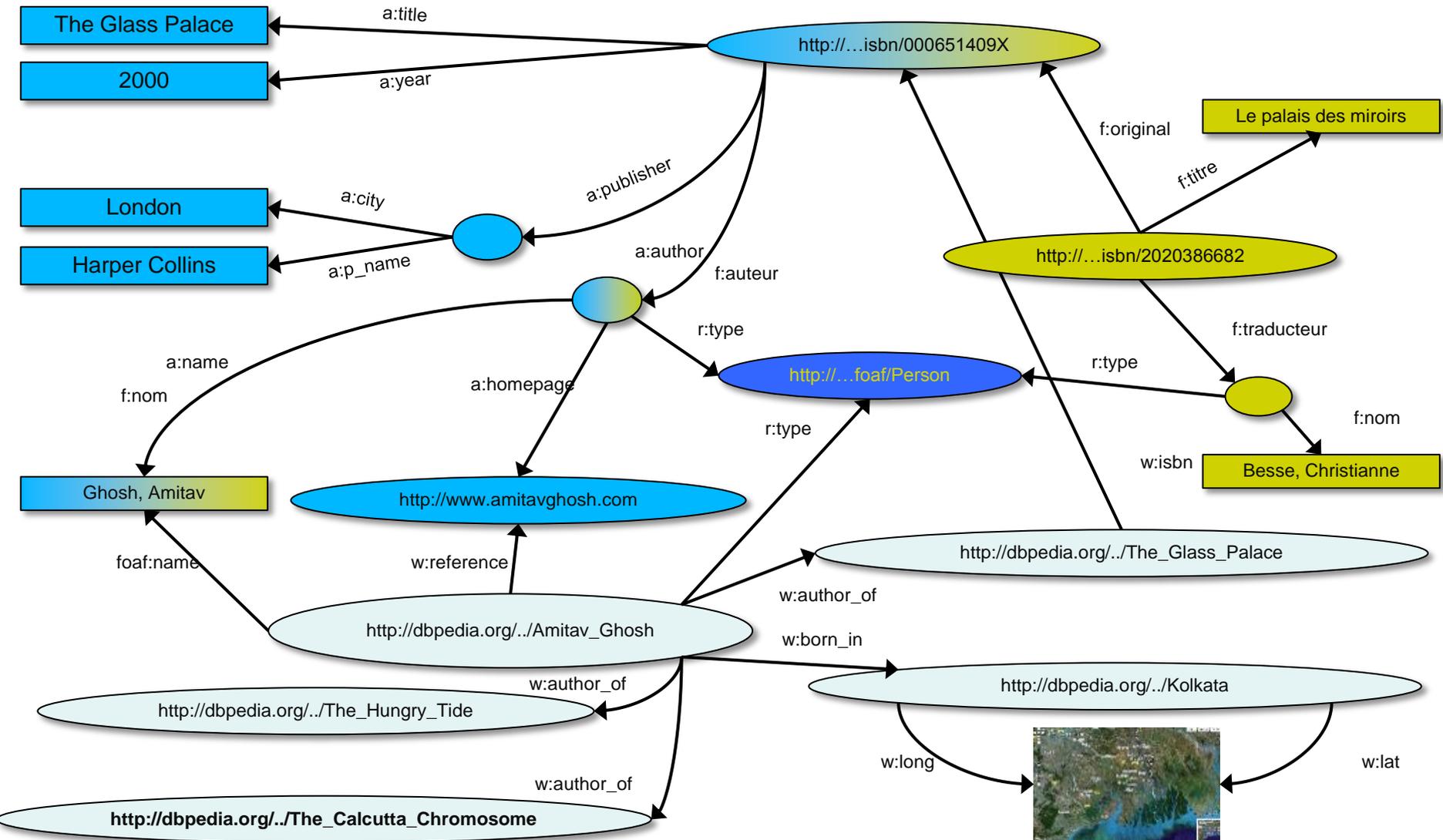
## Merge with Wikipedia data



## Merge with Wikipedia data



# Merge with Wikipedia data



## We can do more

- We could 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, extra rules, etc. come in
  - Ontologies/rule sets can be relatively simple and small, or very complex, or anything in between...
- Even more powerful queries can be asked as a result

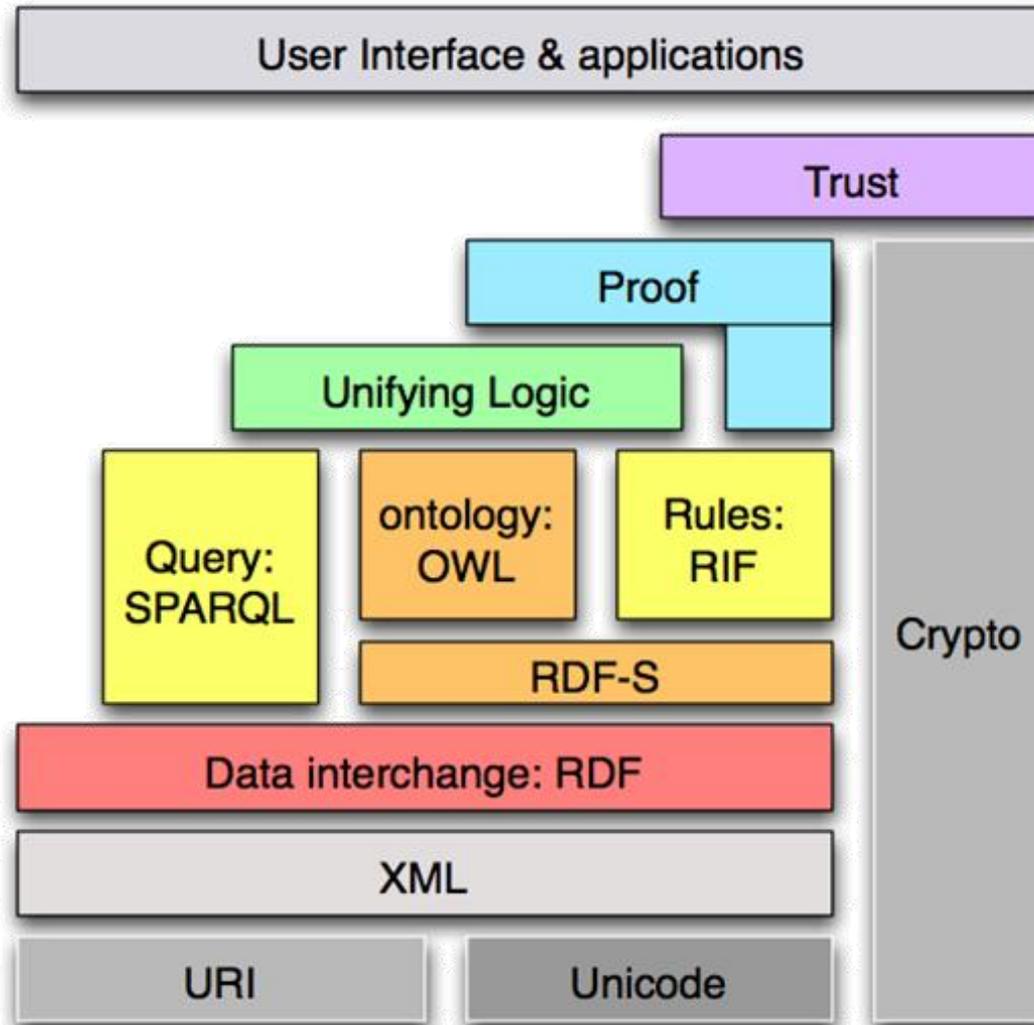
# What did we do?

- Expose data on the web in an interoperable form (RDF)
- Expose knowledge on the web with interoperable semantics (ontologies: RDF Schema, OWL)
- Apply (lightweight) inference for
  - Answering complex queries
  - Integrating multiple sources of knowledge and data

# Semantic Web

- Semantic web utilizes
  - Low expressivity logic (RDF) that allows for some inference
    - Property inheritance
    - Domain/range inference
  - Medium expressivity logic (OWL) that supports additional inference
    - In(equality)
    - Number restrictions
    - Data types ...

## Semantic Web Stack (updated, W3C, 2006)



# Alternative perspectives on the semantic web

- Semantic web
  - annotated web
  - web of data (RDF triple stores)
- Semantic web
  - In the large
  - In the small



# Example data source: DBpedia

- DBpedia is a community effort to
  - extract structured (“infobox”) information from Wikipedia
  - provide a query endpoint to the dataset
  - interlink the DBpedia dataset with other datasets on the Web



# Linked data enable new applications

- Build your own NYT linked data application

The screenshot shows a web browser window titled "Who Went Where" displaying a linked data application for The New York Times. The application is titled "Linked Open Data BETA" and features a search interface for "Alumni In The News" at "Harvard University". The search results show profiles for Henry Kissinger and Jacques Chirac, along with a list of news articles related to them. A sidebar on the left shows a network diagram of various data sources like LIBRIS, DBpedia, and KEGG.

**The New York Times** **Linked Open Data BETA** [View Application Source](#)

**Alumni In The News** **Harvard University**

Enter a school name below and see our coverage of that school's alumni.

Harvard University

**Henry Kissinger**  
Born: May 27, 1923

Henry Kissinger Is Released From Hospital in South Korea - March 15, 2010  
Got Your Back - July 19, 2009  
SPORTS OF THE TIMES; Kissinger's Soccer Diplomacy - March 31, 2009  
THE TV WATCH; A Question Revisited, but the Words Come None Too Easily for Palin - September 26, 2008  
OP-ED COLUMNIST; Park Avenue Diplomacy - September 24, 2008  
Palin Will Meet With Kissinger and Foreign Leaders - September 22, 2008  
Using Star Power to Repair Nigeria's Image - July 10, 2008  
OP-ED CONTRIBUTOR; Listening to Compromise - July 08, 2008  
EDITORIAL OBSERVER; Thinking the Unthinkable: A World Without Nuclear Weapons - June 30, 2008  
ON THE WHITE HOUSE; One Trip, Dual Purposes - April 27, 2008

**Jacques Chirac**  
Born: November 29, 1932

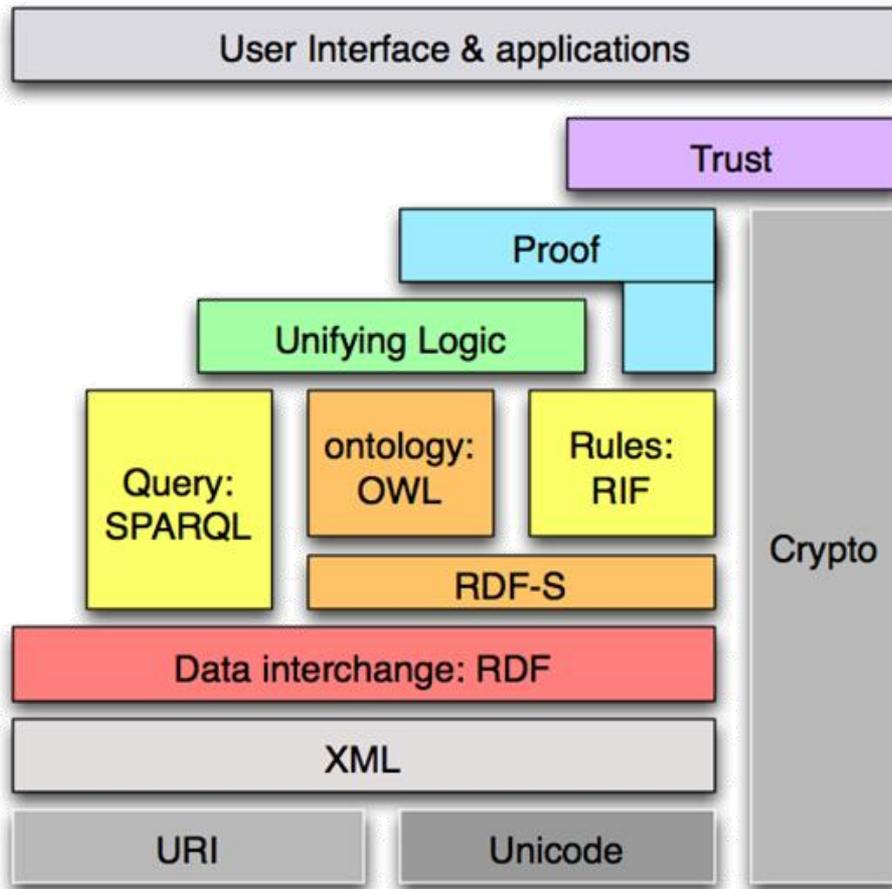
Chirac Faces New Inquiry On Charges Of Corruption - December 19, 2009  
Ex-Leader Of France Faces Trial - October 31, 2009  
French Power Elite Face a Fall From Grace - October 28, 2009  
French President Raises Eyebrows With Remark on a Contentious Political Trial - September 25, 2009  
INSIDE EUROPE; Order to Pay Back Farm Subsidies Comes at a Bad Time for France - August 11, 2009  
WORLD BRIEFING | EUROPE; France: Chirac Creates Foundation - June 10, 2008  
Putin Maintains Presidential Air in Paris Trip - May 31, 2008  
A Statesman Without Borders - February 03, 2008  
Chirac Under Investigation for Misuse of Funds as Paris Mayor - November 22, 2007  
Correction: For the Record - November 05, 2007

**Kennedy, John Fitzgerald**  
Born: May 29, 1917

ESSAY; The Making of the President, Then and Now - March 21, 2010  
Lost in Time, Letters Capture American Grief For a President - March 09, 2010

<http://open.blogs.nytimes.com/2010/03/30/build-your-own-nyt-linked-data-application/>

# We have barely touched on the bottom few layers of the semantic web cake



## Relevant technologies

- Data: RDF, SPARQL
- Vocabularies, Thesauri, Concepts: RDF-S, SKOS, OWL
- Domain Vocabularies and Ontologies e.g., GO (gene ontology)
- Reasoners e.g., Pellet
- Text processors e.g., OpenCalais

# Semantic Web

- Expose data on the web in an interoperable form (RDF)
- Expose knowledge on the web with interoperable semantics (ontologies: RDF Schema, OWL)
- Apply (lightweight) inference for
  - Searching for information
  - Answering complex queries
  - Integrating multiple sources of knowledge and data
  - Composing composite services from component services
  - Learning predictive models from disparate data sources
  - Unexpected reuse

# We can do a lot already, but many problems remain

Practically useful yet computationally tractable approaches to

- Representing and using context
- Selective knowledge sharing across ontologies
- Privacy preserving query answering
- Incorporating uncertainty, imprecision
- Representing and reasoning about time and space
- Representing and reasoning about preferences
- Capturing, representing, and reasoning about provenance
- Learning from data and knowledge
- Learning semantics-preserving mappings across ontologies

# Thank You!