



# Short introduction to the Semantic Web



Ivan Herman, W3C  
Last update: 2010-06-04

# WARNING TO THE READER!

---

- ▶ **This is an evolving slide set. This means:**
  - ▶ it changes frequently
  - ▶ there may be bugs, inconsistencies
  - ▶ it may try to reflect the latest view of technology evolution but that is often a moving target
- ▶ **“Frozen” versions are instantiated for a specific presentation, and those become stable**

# The Music site of the BBC

The screenshot shows a web browser window with the URL <http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294>. The page title is "BBC - Music - Eric Clapton". The navigation bar includes "MUSIC BETA", "GENRES", "ARTISTS", "REVIEWS", "NEWS", and "BLOG". A search bar labeled "QUICK FIND" is present with the placeholder text "Enter an artist name ...". The breadcrumb trail reads "BBC Music > Artists > Eric Clapton".

The main content area features the artist's name "Eric Clapton" and his birth date "Born 30 March 1945.". A large photograph of Clapton playing a light blue electric guitar is shown, with the caption "David Redfern/Redferns" at the bottom right. To the right of the photo is a "MOST PLAYED ON BBC RADIO 2" badge.

Below the photo is a "Biography" section with the following text: "Eric Patrick Clapton, CBE (born 30 March 1945) is an English blues-rock guitarist, singer, songwriter and composer. Clapton has been inducted into the Rock and Roll Hall of Fame as a solo performer, as a member of rock bands; the Yardbirds and Cream. Clapton is the only person ever to be inducted three times. Often viewed by critics and fans alike as one of the most important and influential guitarists of all time, Clapton was ranked fourth in Rolling Stone magazine's list of the '100 Greatest Guitarists of All Time' and #53 on their list of the Immortals: 100 Greatest Artists of All Time".

On the right side of the page, there are two sections: "Latest Tracks Played On The BBC" and "Audio Previews From Latest Album Review".

**Latest Tracks Played On The BBC**

- Promises**  
BBC Radio 2 | [Ken Bruce 22/02/2010](#)
- Bad Love**  
BBC Radio 2 | [Alex Lester 22/02/2010](#)
- Lay Down Sally**  
BBC Radio 2 | [Chris Evans Breakfast 18/02/2010](#)
- I Ain't Gonna Stand For It**  
BBC Radio 2 | [Alex Lester 15/02/2010](#)
- Wonderful Tonight**  
BBC Radio 2 | [Ken Bruce 10/02/2010](#)

**Audio Previews From Latest Album Review**

**Me And Mr Johnson**

- 8 Milkcow's Calf Blues
- 10 Come on in My Kitchen

# The Music site of the BBC

BBC - Music - Eric Clapton

http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

Netvibes Feedly Social Private Mailing lists SW Python RDFa it! Bookmarks Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

this focus, he is credited as an innovator in a wide variety of genres. These include blues-rock (with John Mayall & the Bluesbreakers and The Yardbirds) and psychedelic rock (with Cream). Clapton's chart success was not limited to the blues, with chart-toppers in Delta Blues (Me and Mr. Johnson), pop ("Change the World") and reggae (Bob Marley's "I Shot the Sheriff") (He is often credited for bringing reggae and Bob Marley to the mainstream.) Two of his most successful recordings were the hit love song "Layla", which he played with the band Derek and the Dominos, and Robert Johnson's "Crossroads", which has been his staple song since his days with Cream.

[Read more at Wikipedia...](#)

**WIKIPEDIA** This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can [edit this article at Wikipedia](#). Find out more about our use of this data .

**Links & Information**

**LINKS**

- [Official homepage at ericclapton.com](#)
- [Fanpage at whereseric.com](#)
- [Wikipedia article on Eric Clapton](#)
- [MySpace at mspace.com/ericclapton](#)
- [Last.fm page on Eric Clapton](#)
- [MusicBrainz entry on Eric Clapton](#)

**MEMBER OF** [Derek and the Dominos](#), [Blind Faith](#) (1968-1969), [Cream](#) (1966-1968), [John Mayall & The Bluesbreakers](#) (1965-1966), [The Yardbirds](#) (1963-1965)

**COLLABORATED ON** [J.J. Cale & Eric Clapton](#), [Eric Clapton & The Immediate All Stars](#), [Eric Clapton & The Impressions](#), [Eric Clapton & Jimmy Page](#), [Eric Clapton & David Sanborn](#), [Eric Clapton & Stan Webb's Chicken Shack](#), [Eric Clapton & The Powerhouse](#), [Eric Clapton & Stevie Ray Vaughan](#), [Eric Clapton & Marc Shaiman](#), [The Dirty Mac](#), [Bob Dylan](#), [Roger McGuinn](#), [Tom Petty](#), [Neil Young](#), [Eric Clapton & George Harrison](#), [Jimmie Vaughan](#), [Eric Clapton](#), [Bonnie Raitt](#), [Robert Cray](#), [B.B. King](#), [Buddy Guy](#), [Dr. John & Art Neville](#), [Elton John & Eric Clapton](#), [Michael Kamen](#), [Eric Clapton and David Sanborn](#), [B.B. King & Eric Clapton](#), [Mark Knopfler & Eric Clapton](#), [Paul McCartney & Eric Clapton](#), [Sting with Eric Clapton](#), [Steve Winwood & Eric Clapton](#)

Links and information come from MusicBrainz. You can add or edit information about [Eric Clapton at musicbrainz.org](#). Find out more about our use of this data. The BBC is not responsible for the content of external sites

**Latest News Stories**

**Played By**

Since December 2008

-  **Alex Lester**  
2 BBC Radio 2
-  **Steve Wright in the Afternoon**  
2 BBC Radio 2
-  **Sarah Kennedy**  
2 BBC Radio 2
-  **Wake Up to Wogan**  
2 BBC Radio 2
-  **Ken Bruce**  
2 BBC Radio 2
-  **Steve Wright's Sunday Love Songs**  
2 BBC Radio 2
-  **Jeremy Vine**  
2 BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete out more about this artist play count information.

# How to build such a site 1.

---

- ▶ Site editors roam the Web for new facts
  - ▶ may discover further links while roaming
- ▶ They update the site manually
- ▶ And the site gets soon out-of-date 😞

## How to build such a site 2.

---

- ▶ Editors roam the Web for new data published on Web sites
- ▶ “Scrape” the sites with a program to extract the information
  - ▶ I.e., write some code to incorporate the new data
- ▶ Easily get out of date again... 😞

## How to build such a site 3.

---

- ▶ Editors roam the Web for new data via API-s
- ▶ Understand those...
  - ▶ input, output arguments, datatypes used, etc
- ▶ Write some code to incorporate the new data
- ▶ Easily get out of date again... 😞

# The choice of the BBC

---

- ▶ Use external, public datasets
  - ▶ Wikipedia, MusicBrainz, ...
- ▶ They are available as data
  - ▶ not API-s or hidden on a Web site
  - ▶ data can be extracted using, eg, HTTP requests or standard queries

## In short...

---

- ▶ Use the Web of Data as a Content Management System
- ▶ Use the community at large as content editors

# And this is no secret...

BBC - Music - Eric Clapton

http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

Netvibes Feedly Social Private Mailing lists SW Python RDFa it! Bookmarklets Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MusicBrainz entry on [Eric Clapton](#)

---

**MEMBER OF** [Derek and the Dominos](#), [Blind Faith](#) (1968-1969), [Cream](#) (1966-1968), [John Mayall & The Bluesbreakers](#) (1966), [The Yardbirds](#) (1963-1965)

---

**COLLABORATED ON** [J.J. Cale & Eric Clapton](#), [Eric Clapton & The Immediate All Stars](#), [Eric Clapton & The Impressions](#), [Eric Clapton & Jimmy Page](#), [Eric Clapton & David Sanborn](#), [Eric Clapton & Stan Webb's Chicken Shack](#), [Eric Clapton & The Powerhouse](#), [Eric Clapton & Stevie Ray Vaughan](#), [Eric Clapton & Marc Shaiman](#), [The Dirty Mac](#), [Bob Dylan](#), [Roger McGuinn](#), [Tom Petty](#), [Neil Young](#), [Eric Clapton & George Harrison](#), [Jimmie Vaughan](#), [Eric Clapton](#), [Bonnie Raitt](#), [Robert Cray](#), [B.B. King](#), [Buddy Guy](#), [Dr. John & Art Neville](#), [Elton John & Eric Clapton](#), [Michael Kamen](#), [Eric Clapton and David Sanborn](#), [B.B. King & Eric Clapton](#), [Mark Knopfler & Eric Clapton](#), [Paul McCartney & Eric Clapton](#), [Sting with Eric Clapton](#), [Steve Winwood & Eric Clapton](#)

---

Links & information come from [MusicBrainz](#). You can add or edit information about [Eric Clapton at musicbrainz.org](#). [Find out more about this data](#). The BBC is not responsible for the content of external sites

---

**Latest News Stories**

NEWS FROM THE BBC

[Clapton recovering after surgery](#)  
Tue 27 Oct 2009 17:46 Musician Eric Clapton is recuperating after having an operation to remove gallstones,...

[Clapton to use a slowhand to take Wyman's wicket](#)

# Data on the Web

---

- ▶ **There are more and more data on the Web**
  - ▶ government data, health related data, general knowledge, company information, flight information, restaurants,...
- ▶ **More and more applications rely on the availability of that data**

But... data are often in isolation, “silos”



# Imagine...

---

- ▶ A “Web” where
  - ▶ documents are available for download on the Internet
  - ▶ but there would be no hyperlinks among them

# And the problem is real...

The image shows three overlapping web browser windows. The top-left window is titled "CoCoDat - Collation of Cortical Data - Mozilla Firefox" and displays the CoCoDat homepage. The top-right window is titled "Cell Centered Database - Mozilla Firefox" and shows the "Cell Centered Database Gallery" with a navigation menu and image thumbnails. The bottom window is titled "NeuronDB = Thalamic relay neuron - Overview (A) () - Mozilla Firefox" and displays a detailed overview of a thalamic relay neuron, including a diagram and a list of components.

**CoCoDat: Collation of Cortical [microcircuitry] Data**

CoCoDat is a microcircuitry database that publishes experimental reports. The data (morphology and cellular compartment), as well as the

- Morphology
- Firing properties
- Ionic currents
- Ionic conductances
- Synaptic currents
- Connectivity

The database is available for download of data tables but also a Search Board with manual or automatic relaxation of the search

- Brain region
- Layer
- Neuron type

<http://www.cocomac.org/cocodat/catalyzer/inc>

**Cell Centered Database Gallery**

National Center for Microscopy and Imaging Research

Data | Search | Gallery | Dictionary | Publications | MyCCDB | Data Download | Contact us | Help

2D image | Reconstruction | Segmentation | Animation

**NeuronDB = Thalamic relay neuron - Overview (A) ()**

Thalamic relay neuron

Mode: **Overview** Data/Search plus Connectivity plus Classical References/Notes Models

Region: Distal equivalent dendrite | Middle equivalent dendrite | Proximal equivalent dendrite | Soma | Axon hillock | Axon fiber | Axon terminal | All Compartments

Properties: **Receptors** Channels Transmitters **All Properties**

Interoperation: Gene and Chromosome Experimental Data (neurodatabase.org) Microscopy Data (CCDB)

Neuron type: principal  
Organism: Vertebrates

1. Equivalent dendrite Show other  
2. Distal equivalent dendrite Show other  
3. Middle equivalent dendrite Show other  
4. Proximal equivalent dendrite Show other  
5. Soma Show other

Done

# Data on the Web is not enough...

---

- ▶ We need a proper infrastructure for a real Web of Data
  - ▶ data is available on the Web
    - ▶ accessible via standard Web technologies
  - ▶ data are interlinked over the Web
  - ▶ ie, data can be integrated over the Web
- ▶ This is where Semantic Web technologies come in

# In what follows...

---

- ▶ We will use a simplistic example to introduce the main Semantic Web concepts

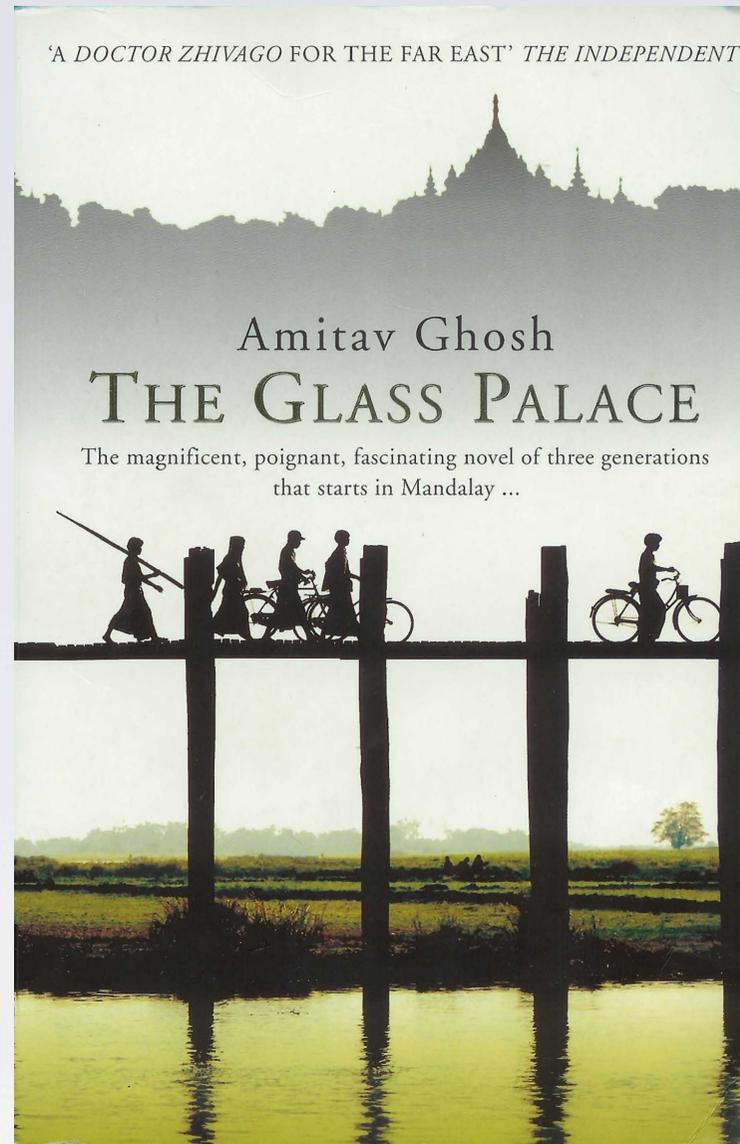
# The rough structure of data integration

---

- ▶ **Map the various data onto an abstract data representation**
  - ▶ make the data independent of its internal representation...
- ▶ **Merge the resulting representations**
- ▶ **Start making queries on the whole!**
  - ▶ queries not possible on the individual data sets

# We start with a book...

---



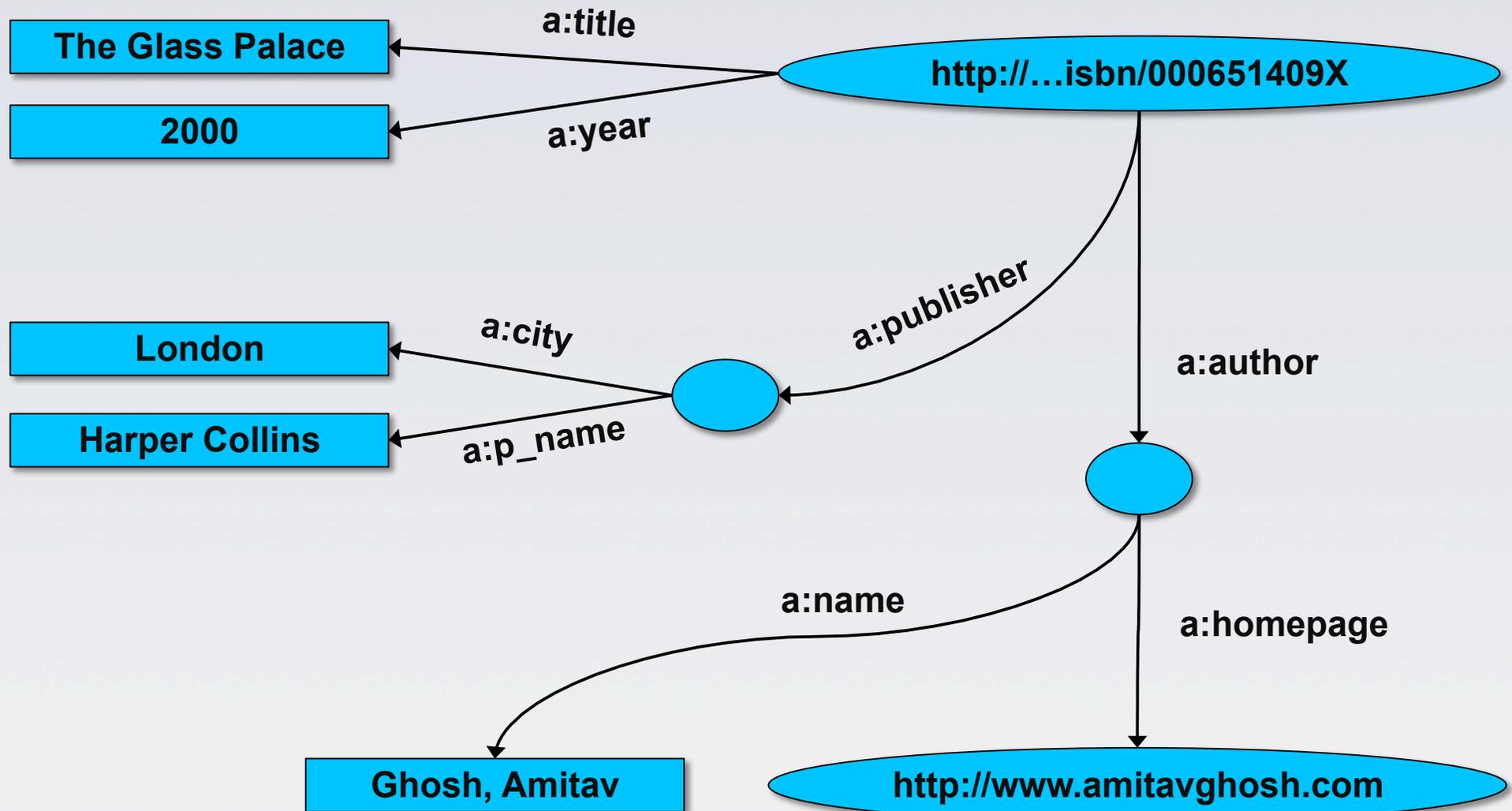
# A simplified bookstore data (dataset “A”)

<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 your 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 machine is immaterial for now

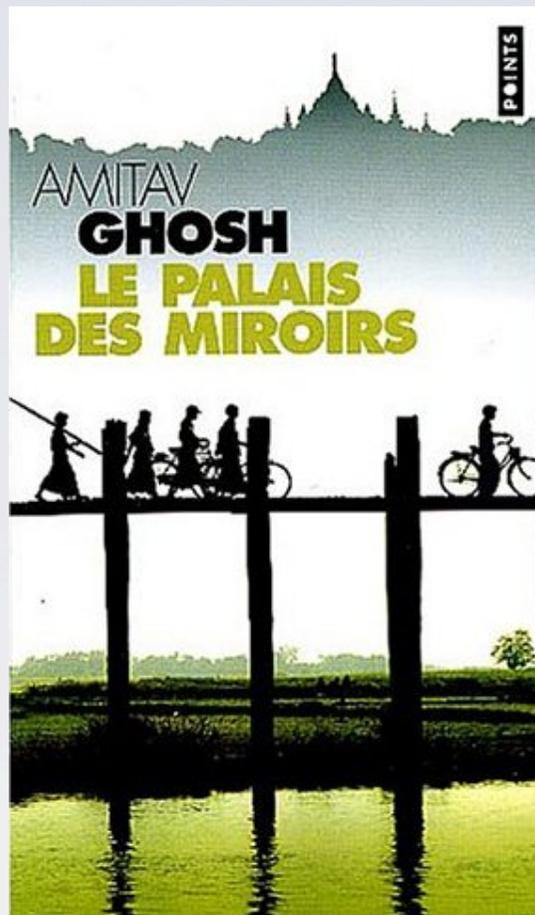
## Some notes on the exporting the data

---

- ▶ **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
    - ▶ etc.
- ▶ **One can export part of the data**

# Same book in French...

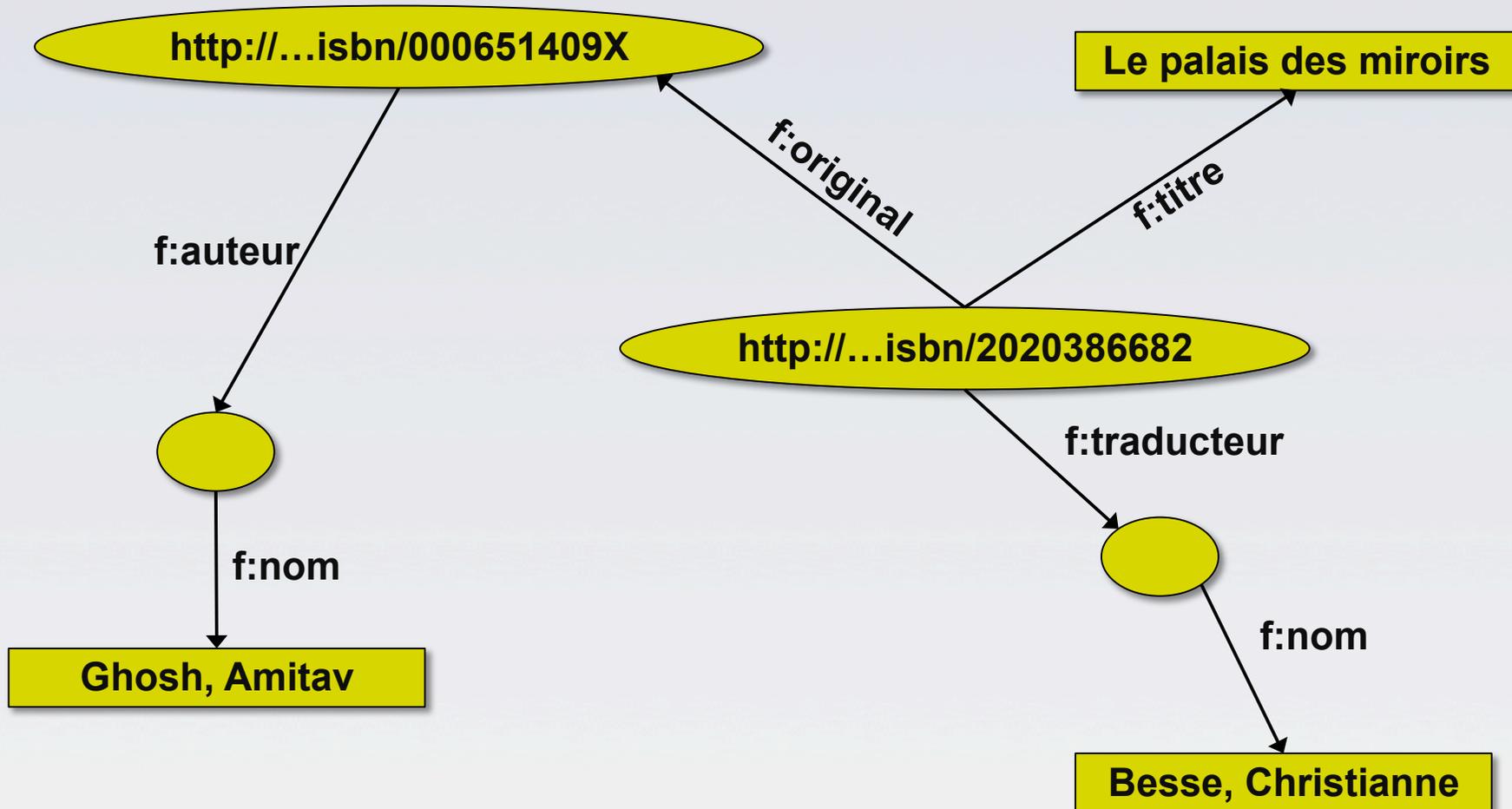
---



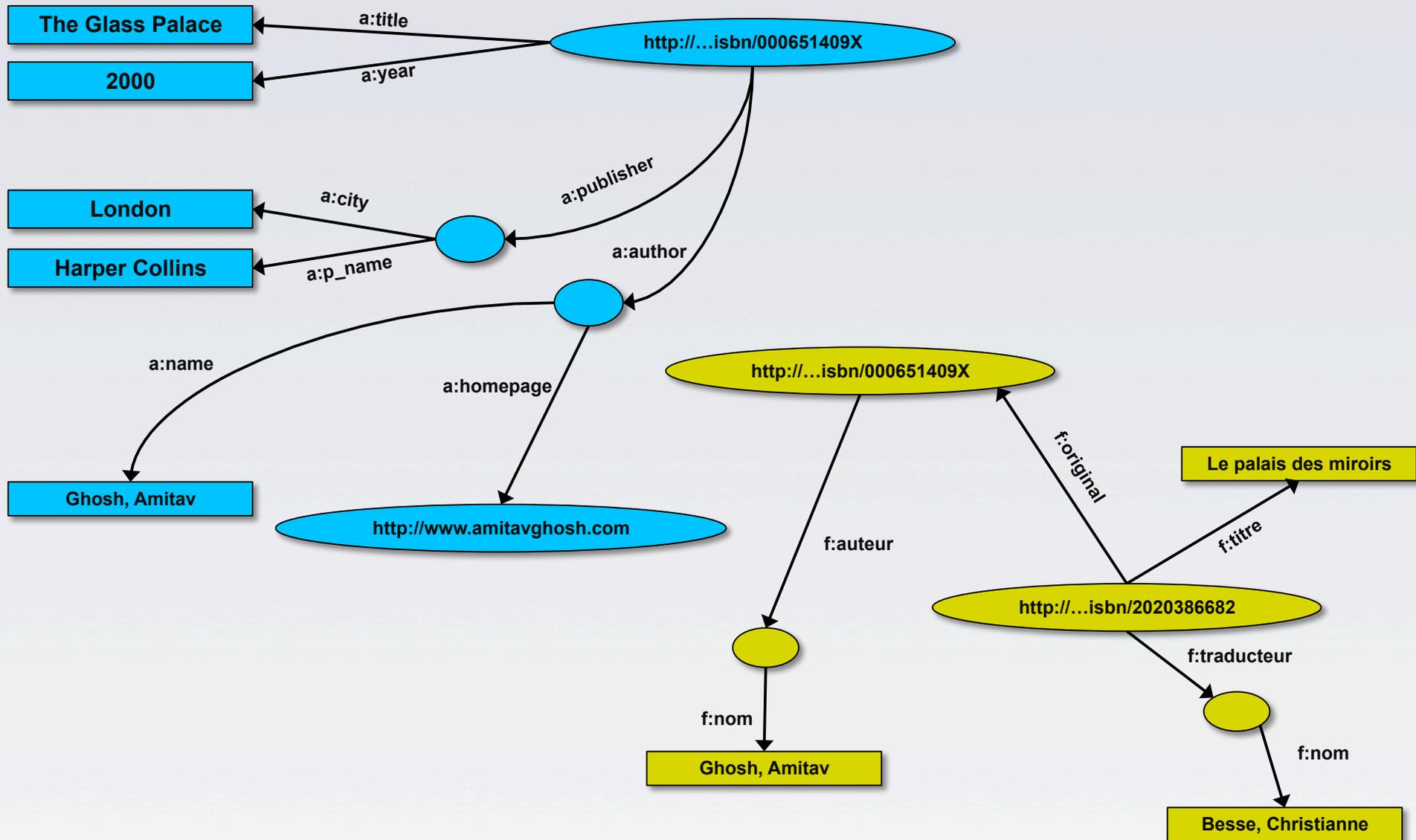
# Another bookstore data (dataset “F”)

A	B	C	D	
1	<b>ID</b>	<b>Titre</b>	<b>Traducteur</b>	<b>Original</b>
2	ISBN 2020286682	Le Palais des Miroirs	\$A12\$	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			

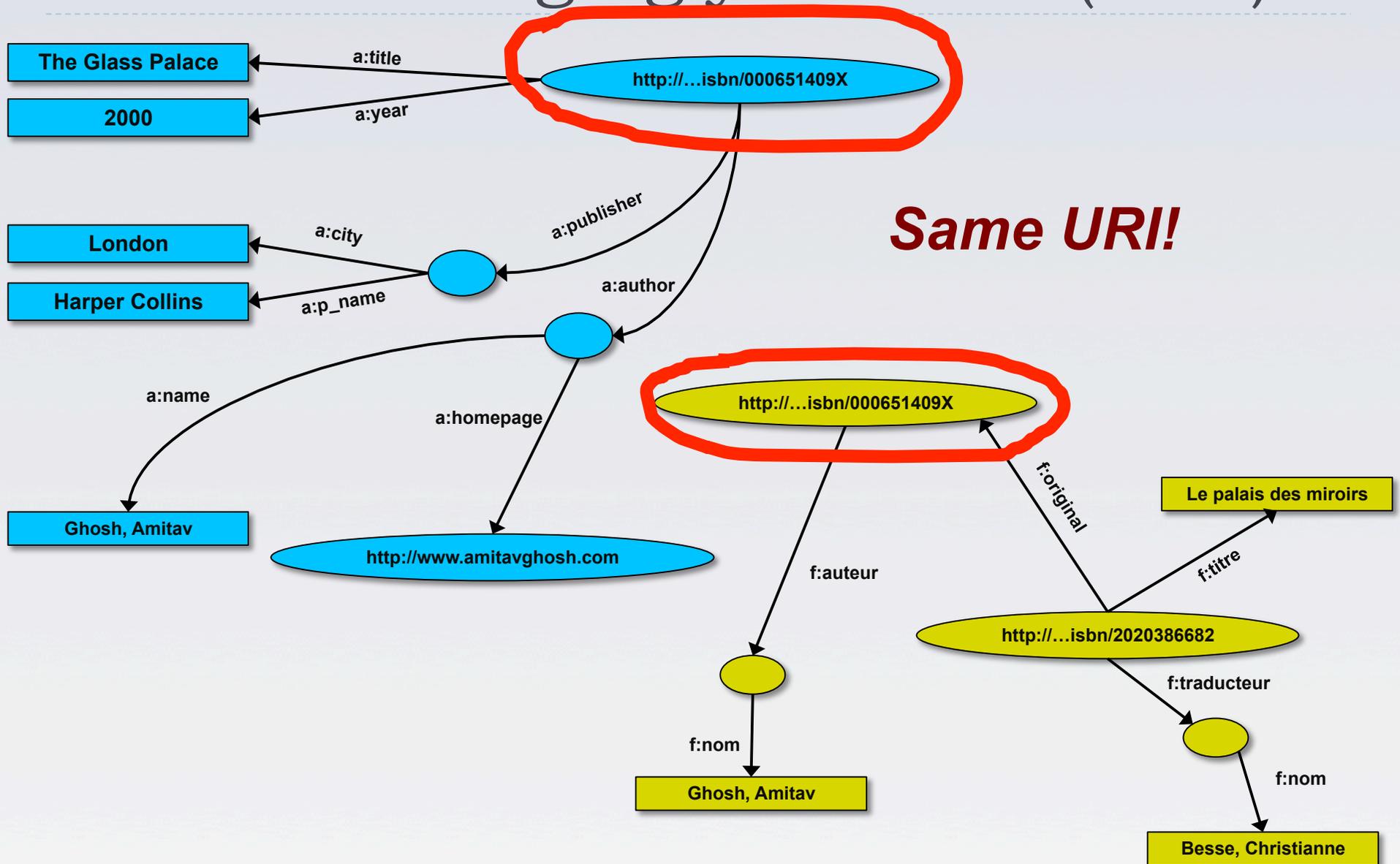
## 2<sup>nd</sup>: export your second set of data



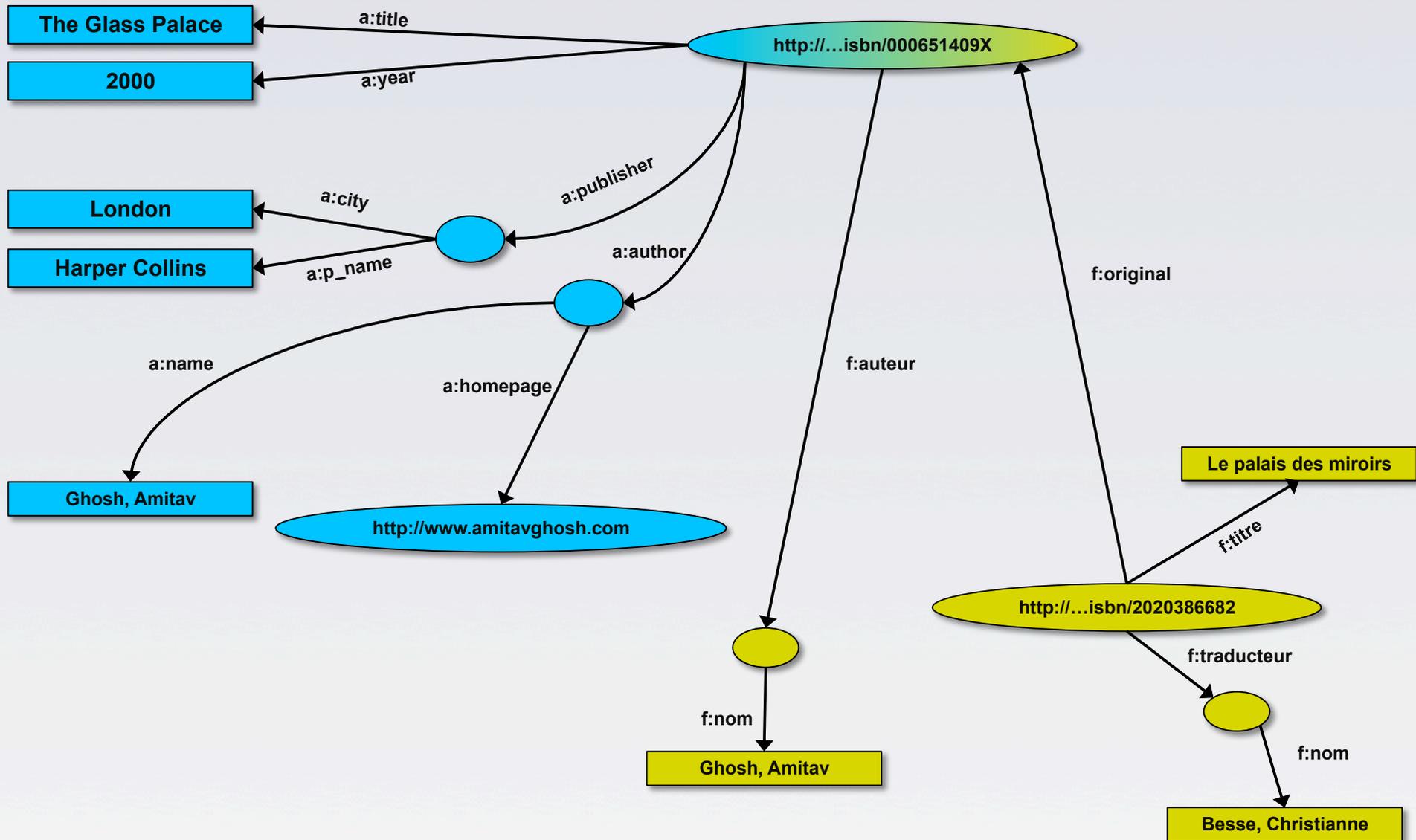
# 3<sup>rd</sup>: start merging your data



# 3<sup>rd</sup>: start merging your data (cont)

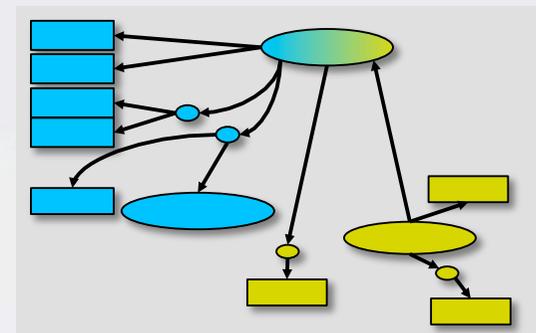


# 3<sup>rd</sup>: start merging your data



# 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”!

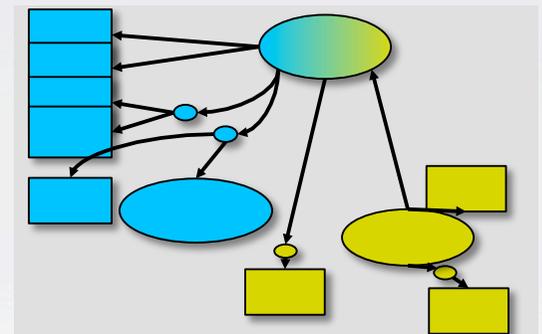
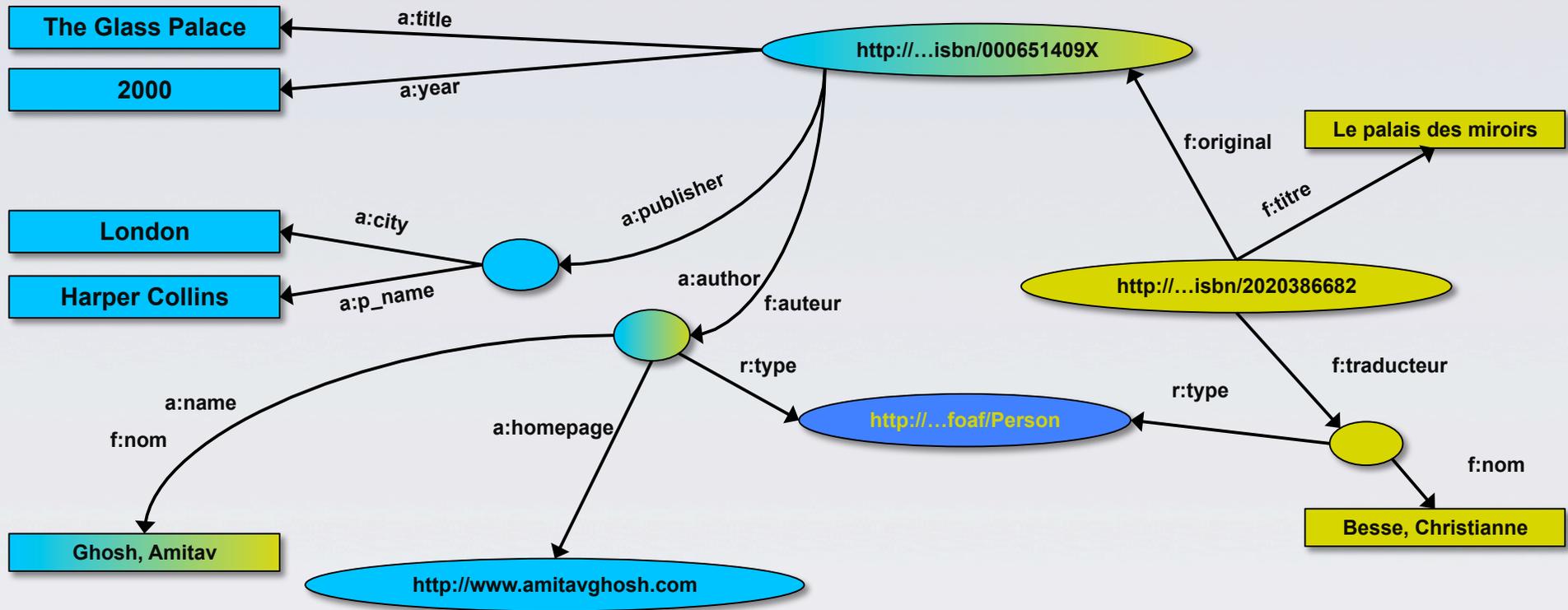


## However, more can be achieved...

---

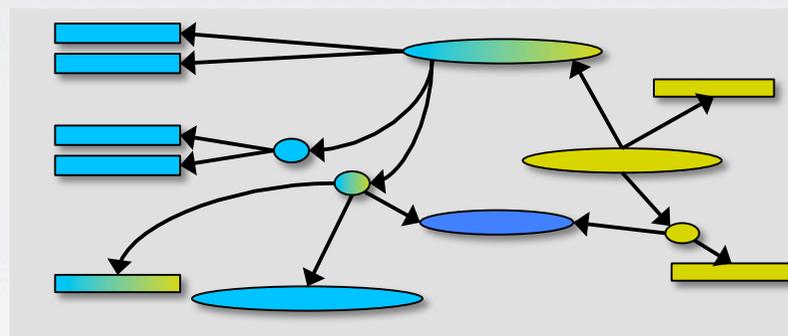
- ▶ We “feel” that a:author and f:auteur should be the same
- ▶ But an automatic merge does not know that!
- ▶ Let us add some extra information to the merged data:
  - ▶ a:author same as f:auteur
  - ▶ both identify a “Person”
  - ▶ a term that a community may have already defined:
    - ▶ a “Person”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

# 3<sup>rd</sup> 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 datasets “A” and datasets “F”
  - ▶ adding three simple extra statements as an extra “glue”

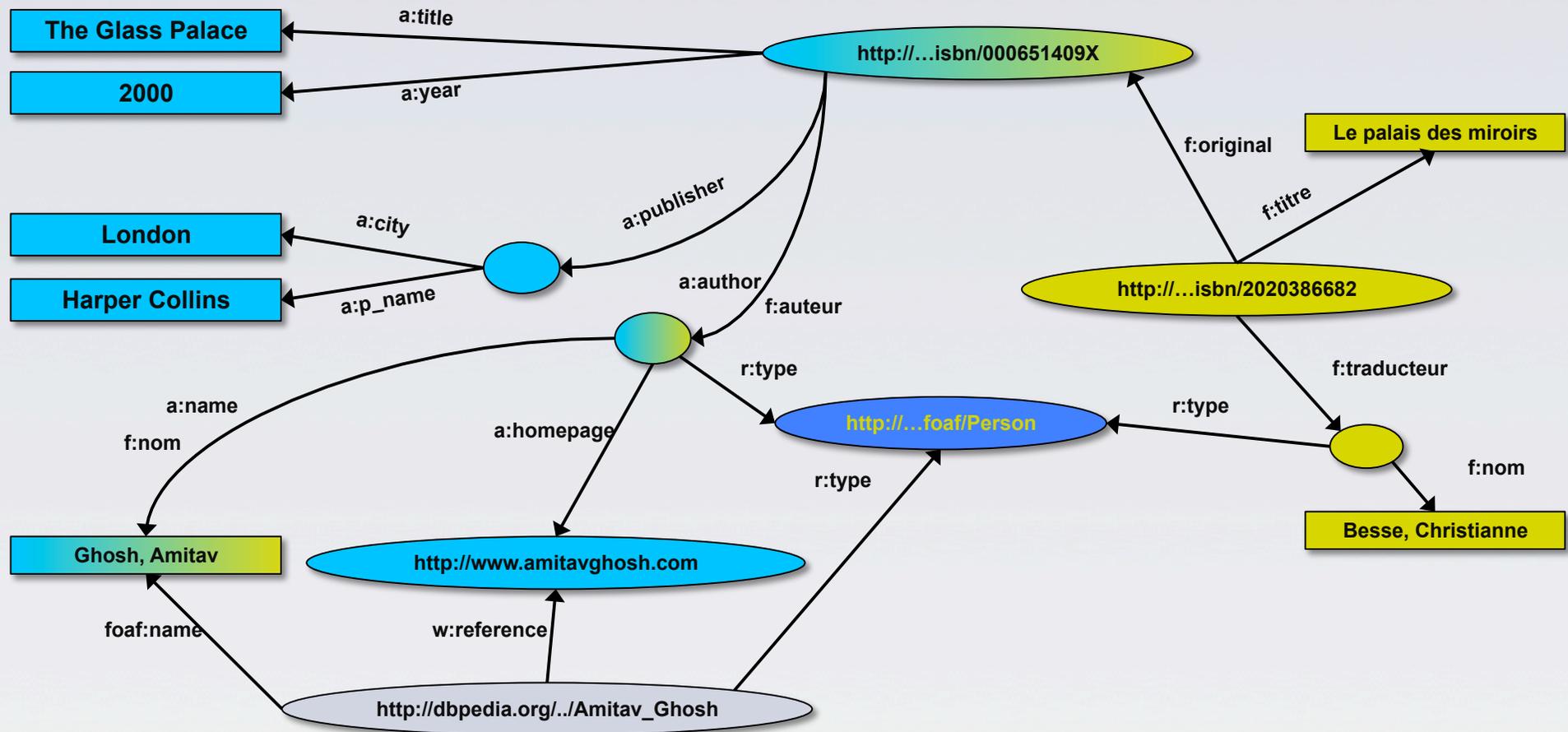


# Combine with different datasets

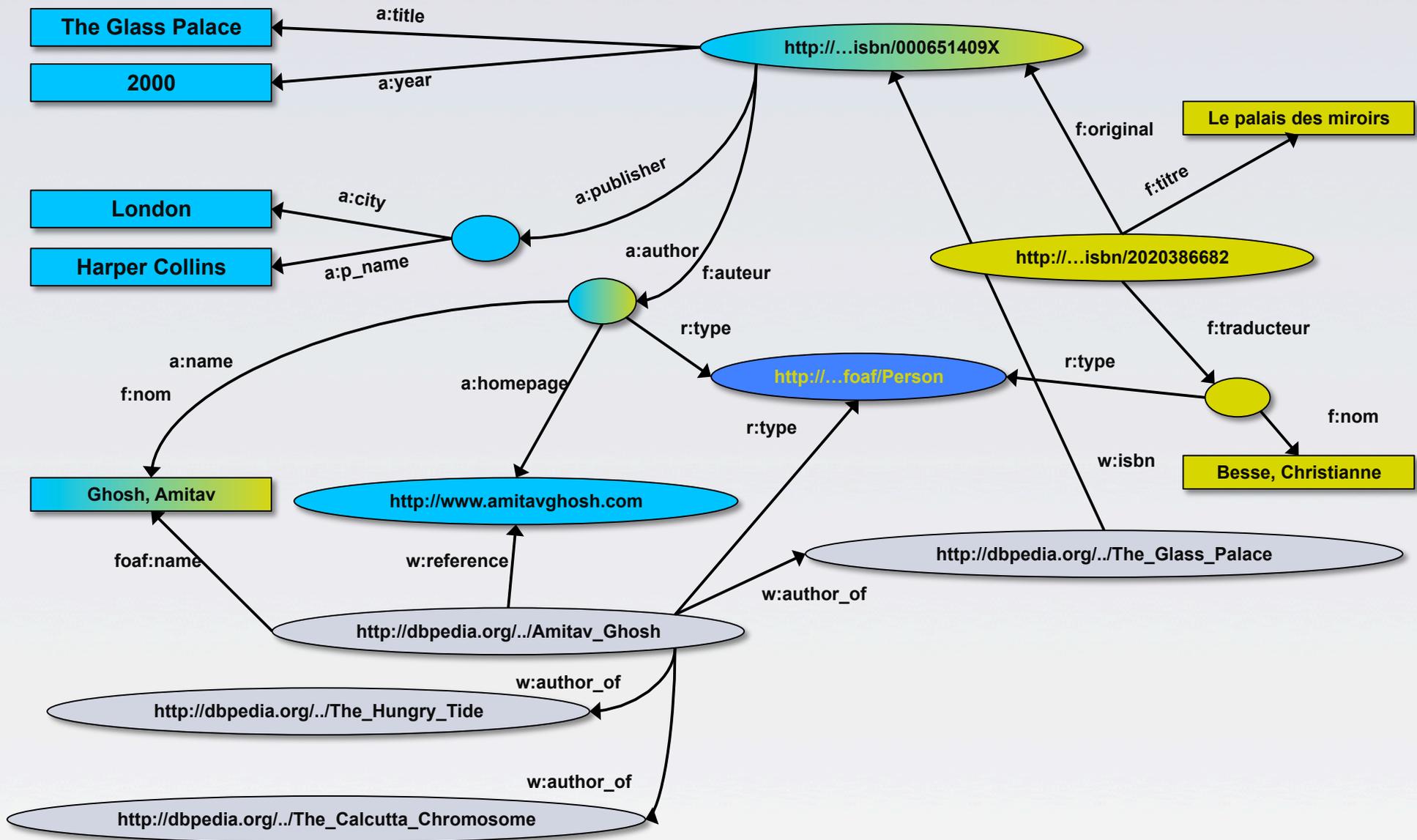
---

- ▶ 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...

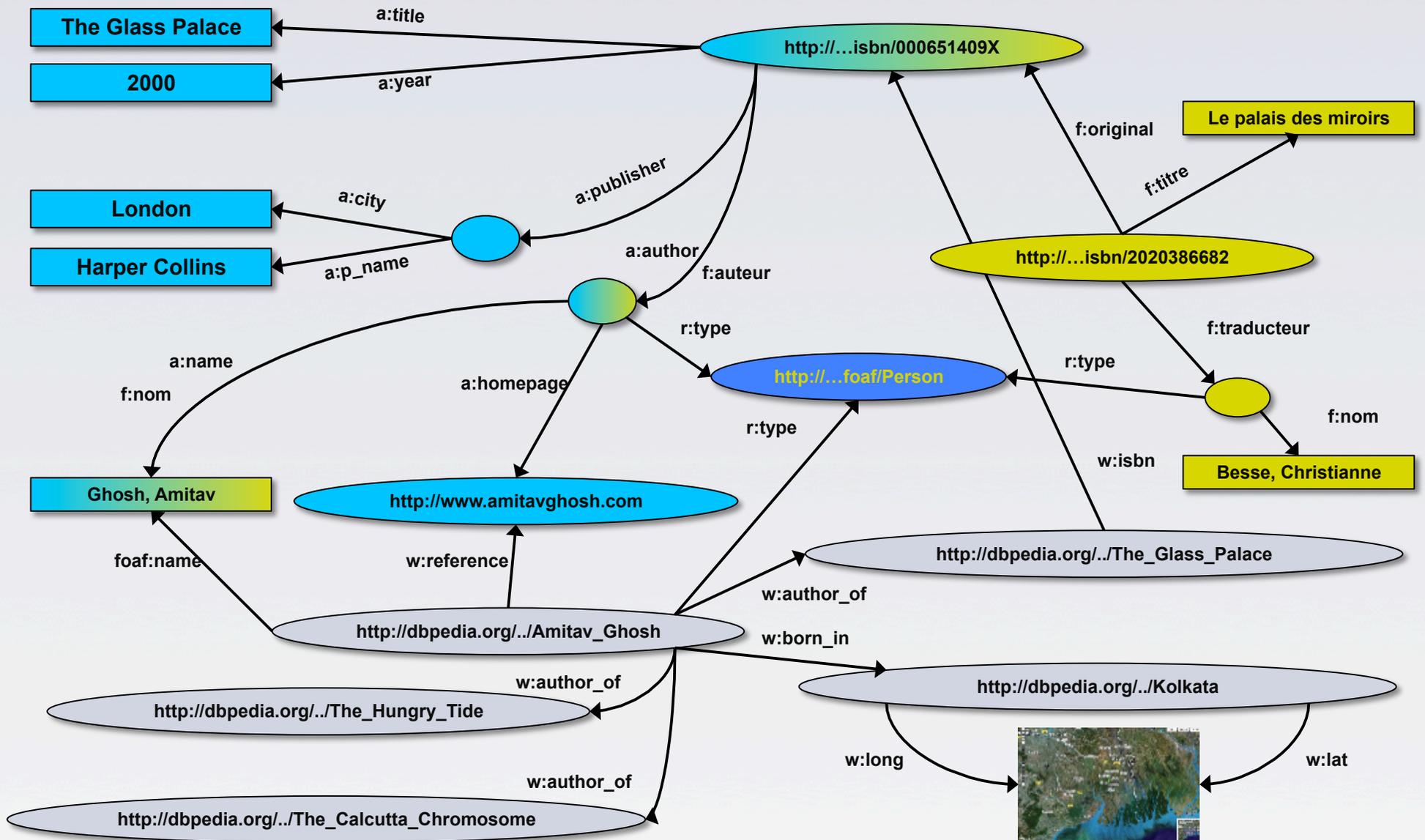
# Merge with Wikipedia data



# Merge with Wikipedia data



# Merge with Wikipedia data



# 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 Web users!
- ▶ The difference: a bit of extra rigour so that machines could do this, too

# What did we do?

---

- ▶ We combined different datasets that
  - ▶ are somewhere on the web
  - ▶ are of different formats (mysql, excel sheet, etc)
  - ▶ have different names for relations
- ▶ We could combine the data because some URI-s were identical (the ISBN-s in this case)

# What did we do?

---

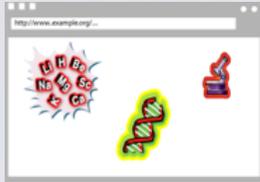
- ▶ We could add some simple additional information (the “glue”), also using common terminologies that a community has produced
- ▶ As a result, new relations could be found and retrieved

# It could become even more powerful

---

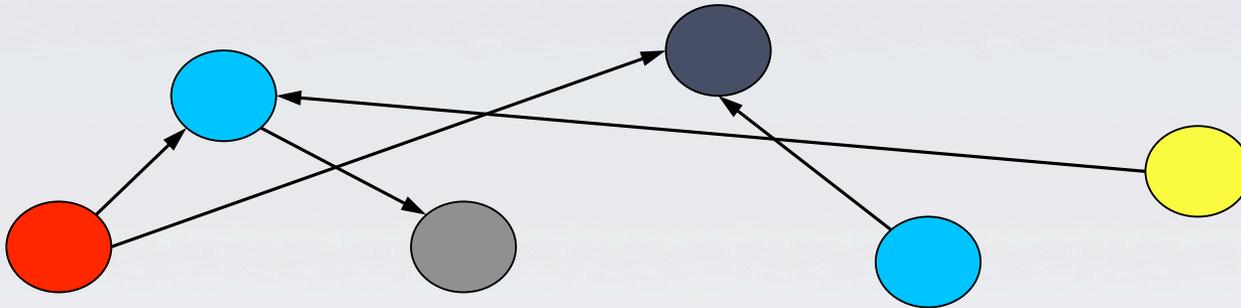
- ▶ We could add extra knowledge to the merged datasets
  - ▶ e.g., a full classification of various types of library data
  - ▶ geographical information
  - ▶ etc.
- ▶ This is where ontologies, extra rules, etc, come in
  - ▶ ontologies/rule sets can be relatively simple and small, or huge, or anything in between...
- ▶ Even more powerful queries can be asked as a result

# What did we do? (cont)



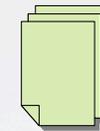
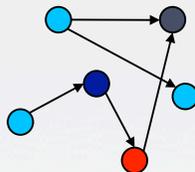
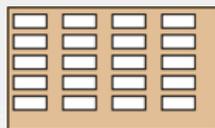
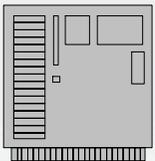
Applications

Manipulate  
Query  
...



Data represented in abstract format

Map,  
Expose,  
...



Data in various formats

# The abstraction pays off because...

---

- ▶ ... the graph representation is independent of the exact structures
- ▶ ... a change in local database schema's, XHTML structures, etc, do not affect the whole
  - ▶ “schema independence”
- ▶ ... new data, new connections can be added seamlessly

# The network effect

---

- ▶ Through URI-s we can link any data to any data
- ▶ The “network effect” is extended to the (Web) data
- ▶ “Mashup on steroids” become possible

# So where is the Semantic Web?

---

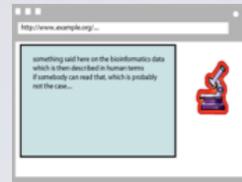
- ▶ The Semantic Web provides technologies to make such integration possible!

## Details: many different technologies

---

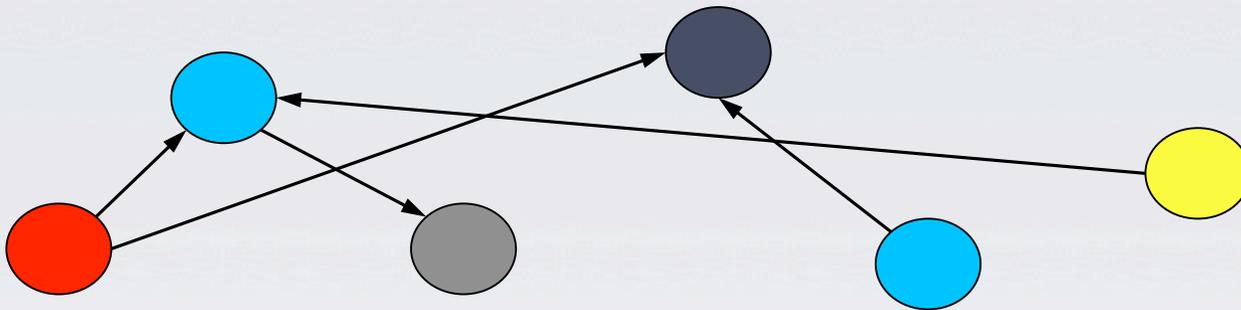
- ▶ an abstract model for the relational graphs: **RDF**
- ▶ add/extract RDF information to/from XML, (X)HTML: **GRDDL, RDFa**
- ▶ a query language adapted for graphs: **SPARQL**
- ▶ characterize the relationships and resources: **RDFS, OWL, SKOS, Rules**
  - ▶ applications may choose among the different technologies
- ▶ reuse of existing “ontologies” that others have produced (FOAF in our case)

# Using these technologies...



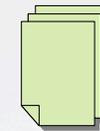
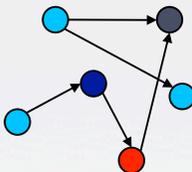
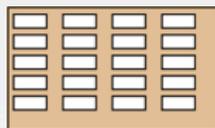
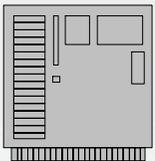
Applications

SPARQL,  
Inferences  
...



Data represented in RDF with extra knowledge (RDFS, SKOS, RIF, OWL,...)

RDB ⇔ RDF,  
GRDL, RDFa,  
...



Data in various formats

# Where are we today (in a nutshell)?

---

- ▶ The technologies are in place, lots of tools around
  - ▶ there is always room for improvement, of course
- ▶ Large datasets are “published” on the Web, ie, ready for integration with others
- ▶ Large number of vocabularies, ontologies, etc, are available in various areas

# Everything is not rosy, of course...

---

- ▶ **Tools have to improve**
  - ▶ scaling for very large datasets
  - ▶ quality check for data
  - ▶ etc
- ▶ **There is a lack of knowledgeable experts**
  - ▶ this makes the initial “step” tedious
  - ▶ leads to a lack of understanding of the technology

# CEO guide for SW: the “DO-s”

---

- ▶ **Start small:** Test the Semantic Web waters with a pilot project [...] before investing large sums of time and money.
- ▶ **Check credentials:** A lot of systems integrators don't really have the skills to deal with Semantic Web technologies. Get someone who's savvy in semantics.
- ▶ **Expect training challenges:** It often takes people a while to understand the technology. [...]
- ▶ **Find an ally:** It can be hard to articulate the potential benefits, so find someone with a problem that can be solved with the Semantic Web and make that person a partner.

# CEO guide for SW: the “DON’T-s”

---

- ▶ **Go it alone:** The Semantic Web is complex, and it's best to get help. [...]
- ▶ **Forget privacy:** Just because you can gather and correlate data about employees doesn't mean you should. Set usage guidelines to safeguard employee privacy.
- ▶ **Expect perfection:** While these technologies will help you find and correlate information more quickly, they're far from perfect. Nothing can help if data are unreliable in the first place.
- ▶ **Be impatient:** One early adopter at NASA says that the potential benefits can justify the investments in time, money, and resources, but there must be a multi-year commitment to have any hope of success

# Thank you for your attention!

These slides are also available on the Web:



<http://www.w3.org/People/Ivan/CorePresentations/IntroThroughExample/>

---