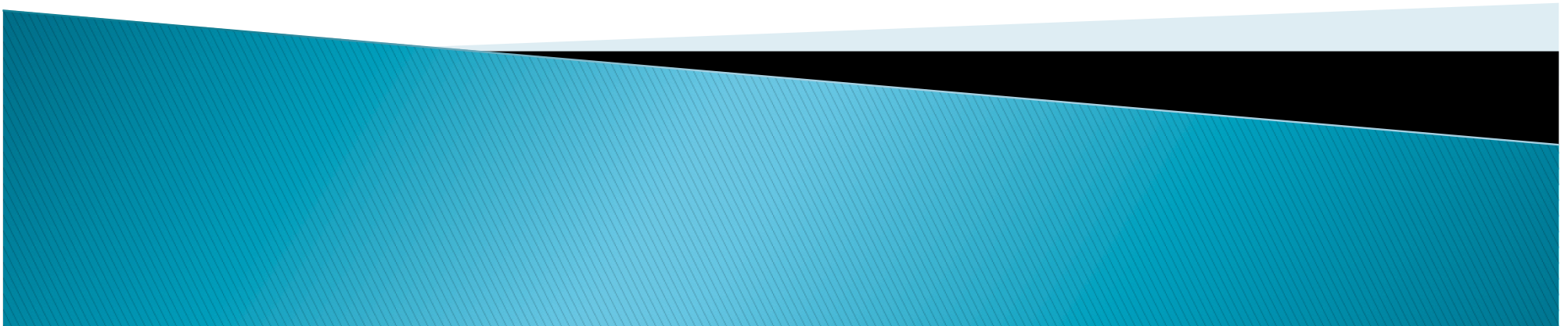


# How does the Semantic Web Work?

Ivan Herman, W3C,  
“Semantic Café”, organized by the W3C Brazil Office  
São Paulo, Brazil, 2010-10-15



# 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! Bookmarklets Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG QUICK FIND Enter an artist name ...

BBC Music > Artists > Eric Clapton

## Eric Clapton

Born 30 March 1945.

MOST PLAYED ON BBC RADIO 2



David Redfern/Redferns

### 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

### Biography

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.

# The Music site of the BBC

The screenshot shows the BBC Music website for Eric Clapton. The browser address bar displays the URL: <http://www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294>. The page features a Wikipedia article snippet on the left, a 'Links & Information' section in the middle, and a 'Played By' list on the right. Red circles highlight specific elements: the 'LINKS' section, the 'MEMBER OF' section, the 'COLLABORATED ON' section, and the 'MySpace' link.

**Wikipedia Article Snippet:**

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](http://ericclapton.com)
- Fanpage at [whereseric.com](http://whereseric.com)
- Wikipedia article on [Eric Clapton](#)
- MySpace at [myspace.com/ericclapton](http://myspace.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 & information come from MusicBrainz. You can add or edit information about [Eric Clapton](#) at [musicbrainz.org](http://musicbrainz.org). Find out more about our use of this data. The BBC is not responsible for the content of external sites

**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...😞

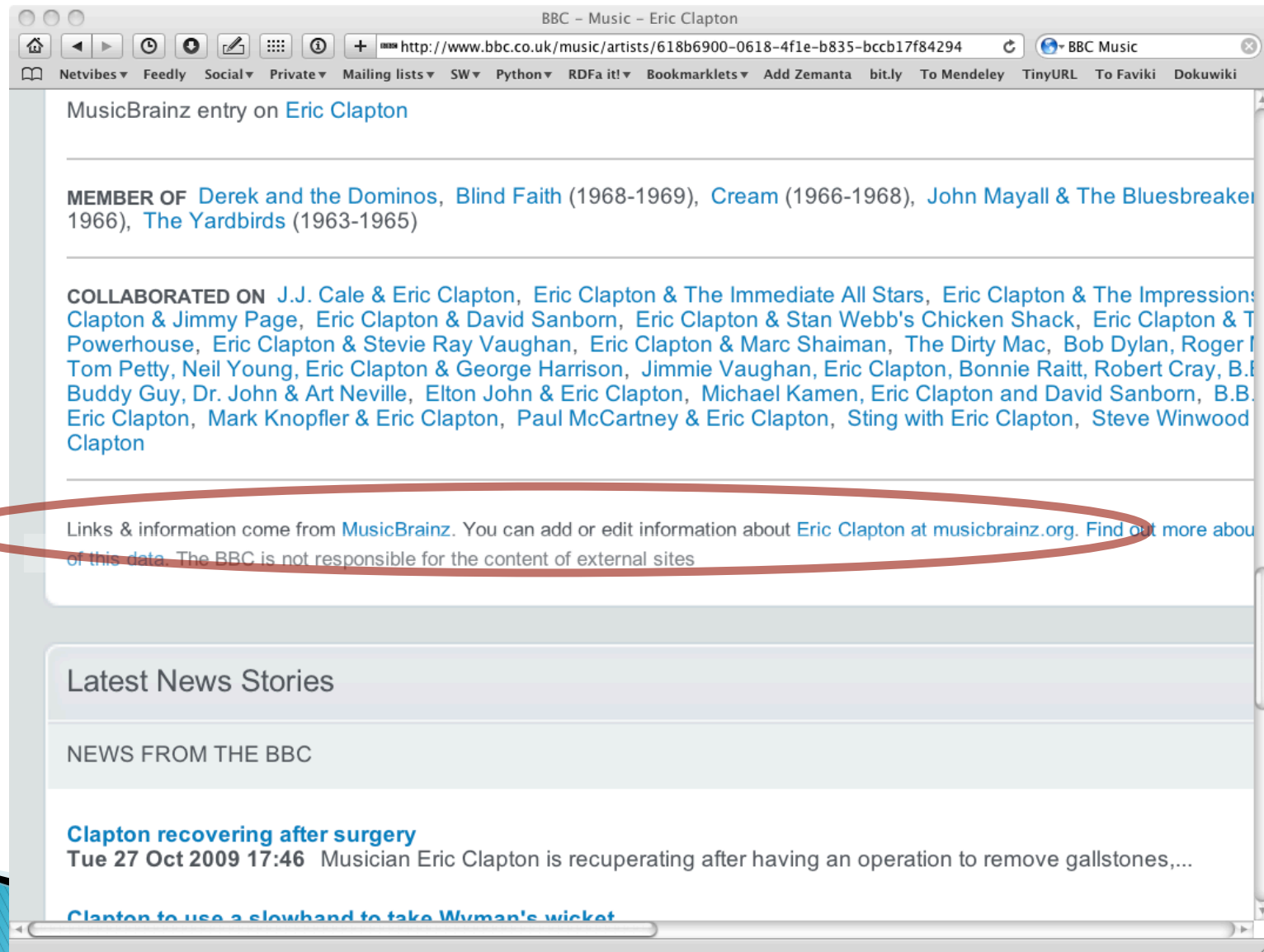
# The choice of the BBC

- ▶ 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...





# 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 displays three overlapping web browser windows, illustrating a complex data integration problem. The windows are:

- CoCoDat - Collation of Cortical Data - Mozilla Firefox**: Shows the CoCoDat homepage with a navigation bar (CoCoMac, DATABASES, ORT, EXAMPLES) and a list of properties: Morphology, Firing properties, Ionic currents, Ionic conductances, Synaptic currents, and Connectivity. It also mentions the database is available for download and has a Search Board.
- Cell Centered Database - Mozilla Firefox**: Shows the Cell Centered Database Gallery homepage with a navigation bar (Data, Search, Gallery, Dictionary, Publications, MyCCDB, Data Download, Contact us, Help) and a list of image types: 2D image, Reconstruction, Segmentation, and Animation.
- NeuronDB = Thalamic relay neuron - Overview (A) () - Mozilla Firefox**: Shows the NeuronDB overview page for a Thalamic relay neuron. It includes a navigation bar (Overview, Data/Search, plus Connectivity, plus Classical References/Notes, Models) and a list of regions: Distal equivalent dendrite, Middle equivalent dendrite, Proximal equivalent dendrite, Soma, Axon hillock, Axon fiber, Axon terminal, and All Compartments. It also includes a list of properties: Receptors, Channels, Transmitters, and All Properties. The page features a diagram of a neuron and a list of properties with corresponding 'Show' and 'other' buttons.

# Data on the Web is not enough...

- ▶ We need a proper infrastructure for a real Web of Data
  - data is available on the Web
  - data are interlinked over the Web (“Linked Data”)
- ▶ I.e., data can be integrated over the Web



I.e.,... we need to connect the silos



# In what follows...

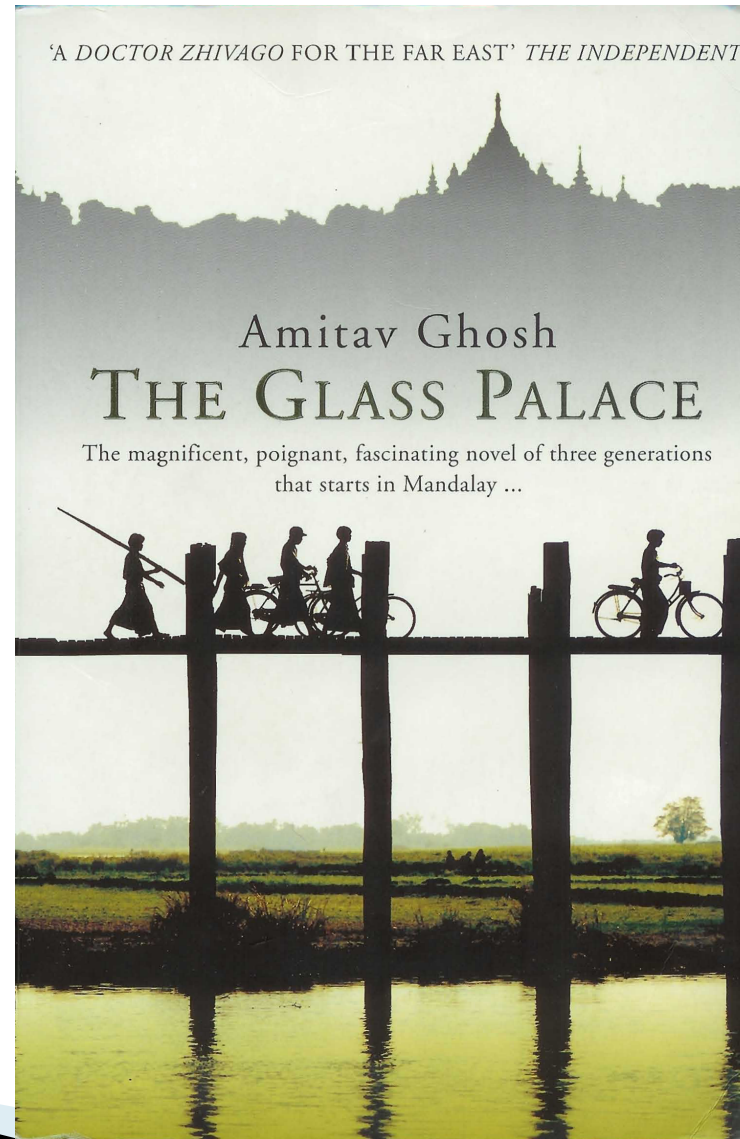
- ▶ We will use a simplistic example to introduce the main Semantic Web concepts



# 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...



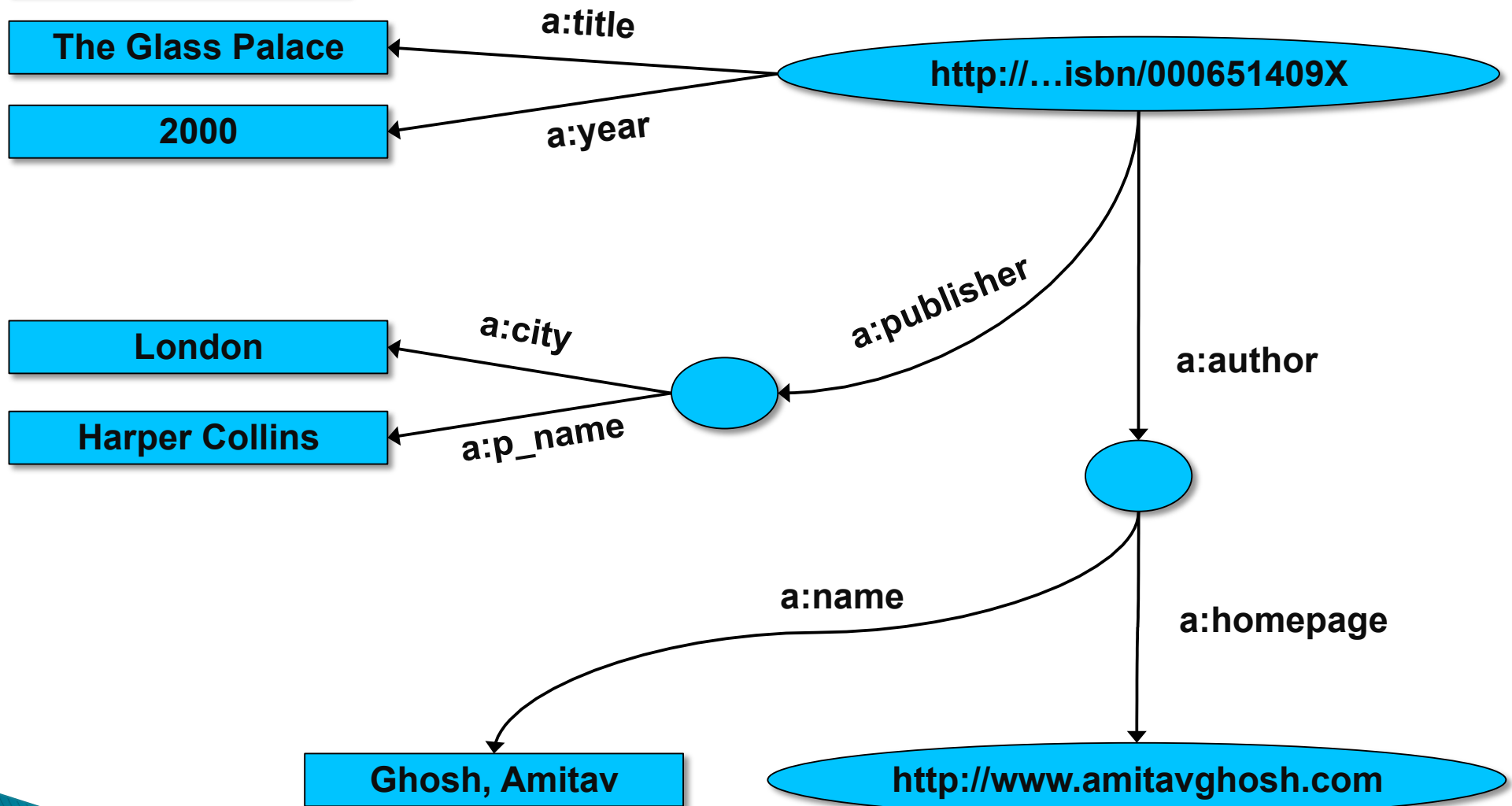
# A simplified bookstore data (dataset “A”)

| ID                  | Author | Title            | Publisher | Year |
|---------------------|--------|------------------|-----------|------|
| ISBN 0-00-6511409-X | id_xyz | The Glass Palace | id_qpr    | 2000 |

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

| ID     | Publisher's name | City   |
|--------|------------------|--------|
| id_qpr | Harper Collins   | London |

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

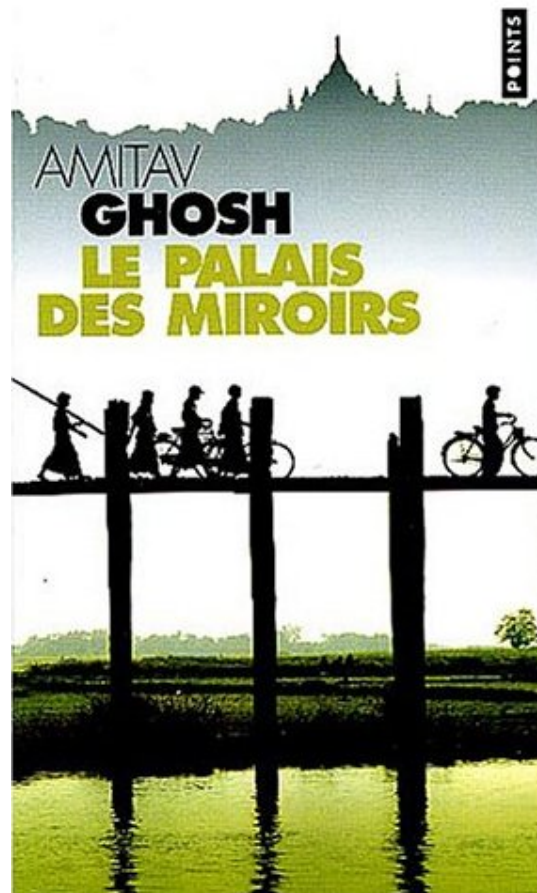


# Some notes on the exporting the data

- ▶ 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



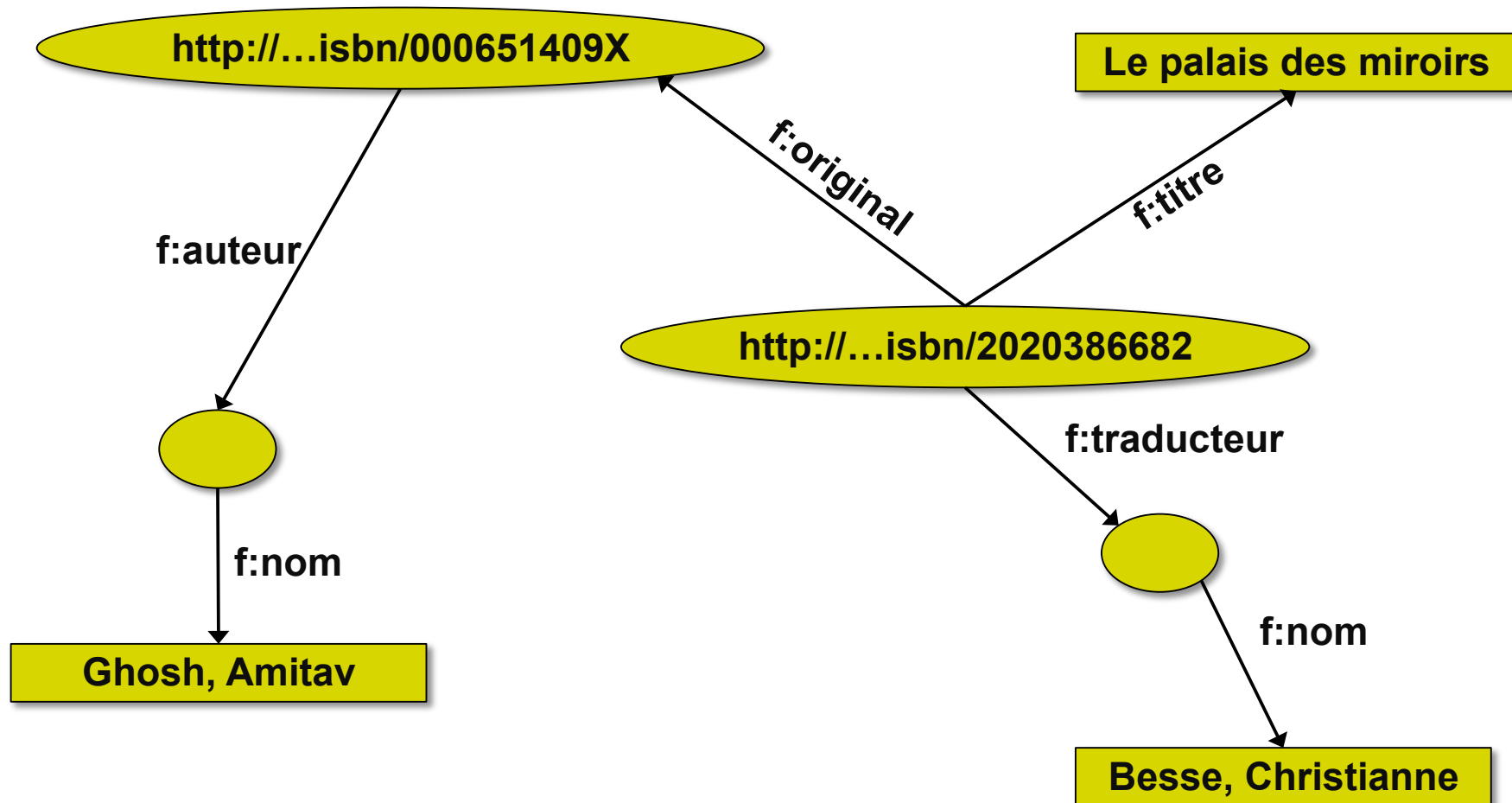
# Same book in French...



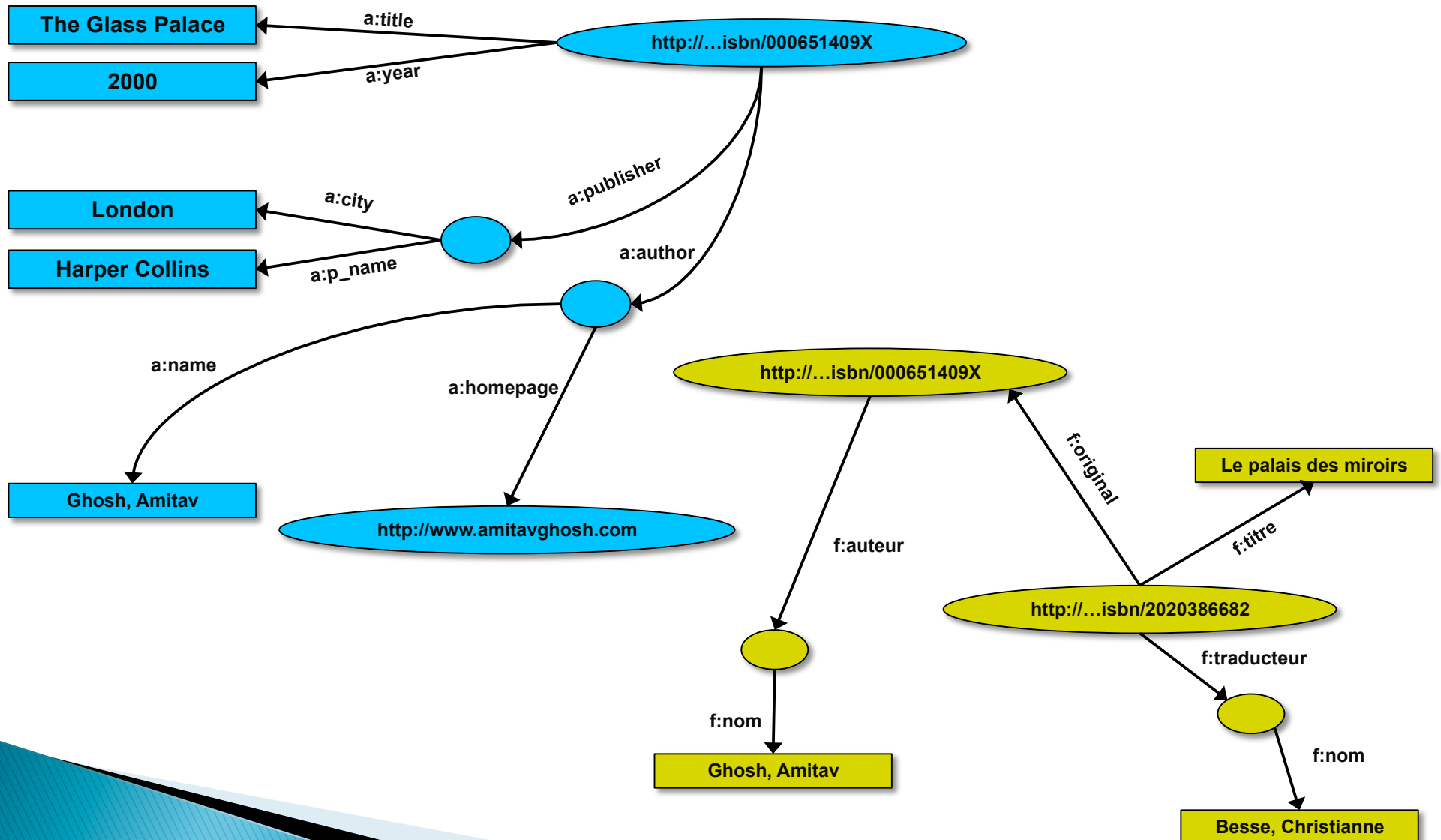
# Another bookstore data (dataset “F”)

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

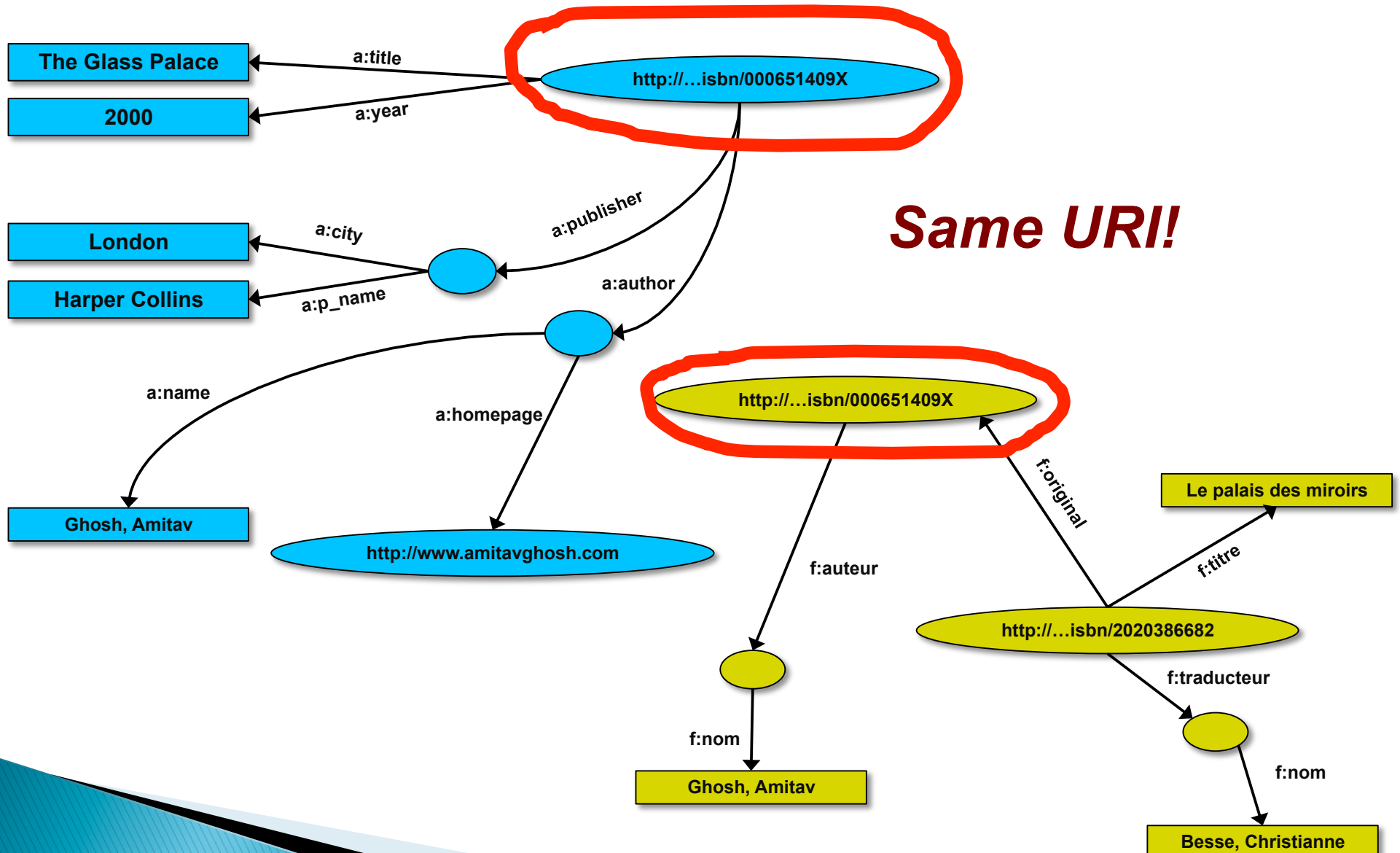
## 2<sup>nd</sup>: export your second set of data



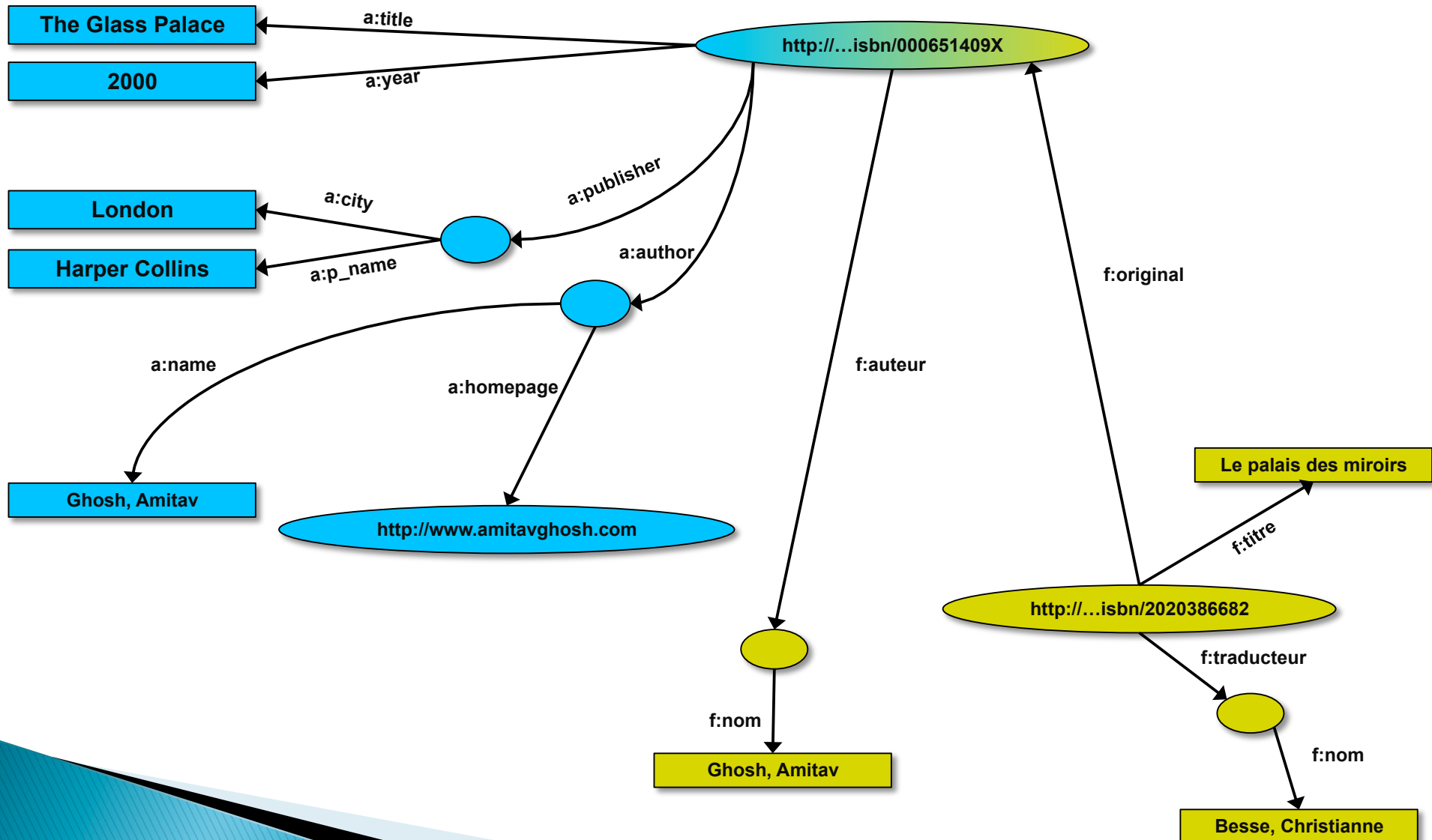
# 3<sup>rd</sup>: start merging your data



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



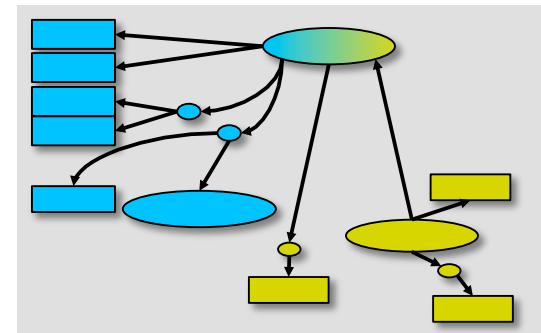
# 3<sup>rd</sup>: start merging your data





# Start making queries...

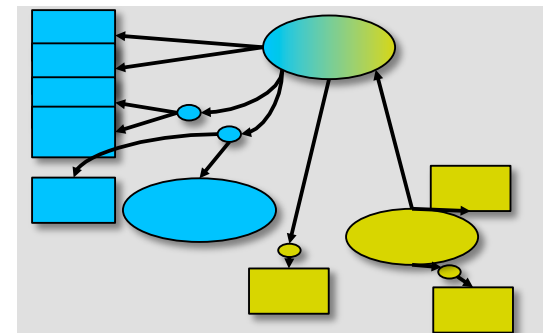
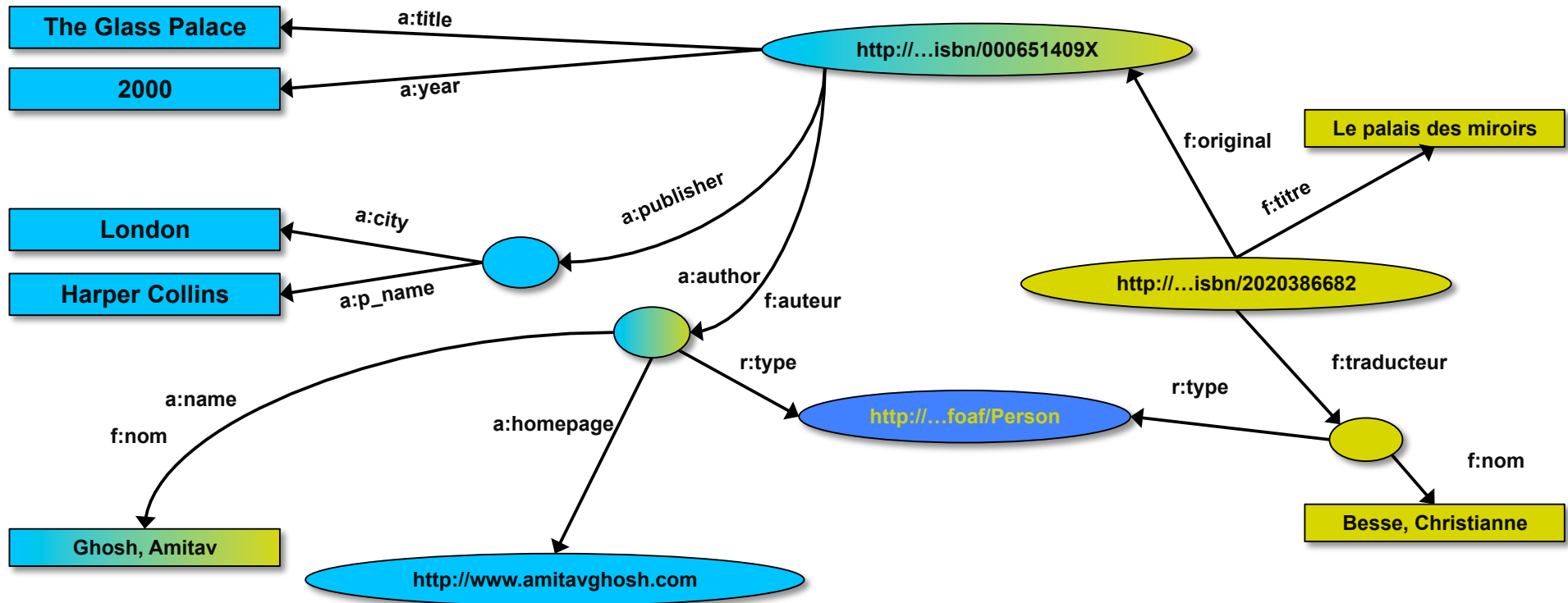
- ▶ User of data “F” can now ask queries like:
  - “give me the title of the original”
    - well, ... « donne-moi le titre de l’original »
- ▶ This information is not in the dataset “F”...
- ▶ ...but can be retrieved by merging with dataset “A”!



# However, more can be achieved...

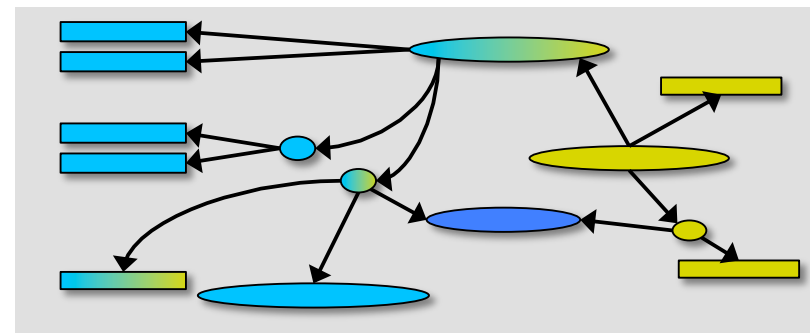
- ▶ We “feel” that a:author and f:auteur should be the same
- ▶ But an automatic merge does not know that!
- ▶ Let us add some extra information to the merged data:
  - a:author same as f:auteur
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely 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!

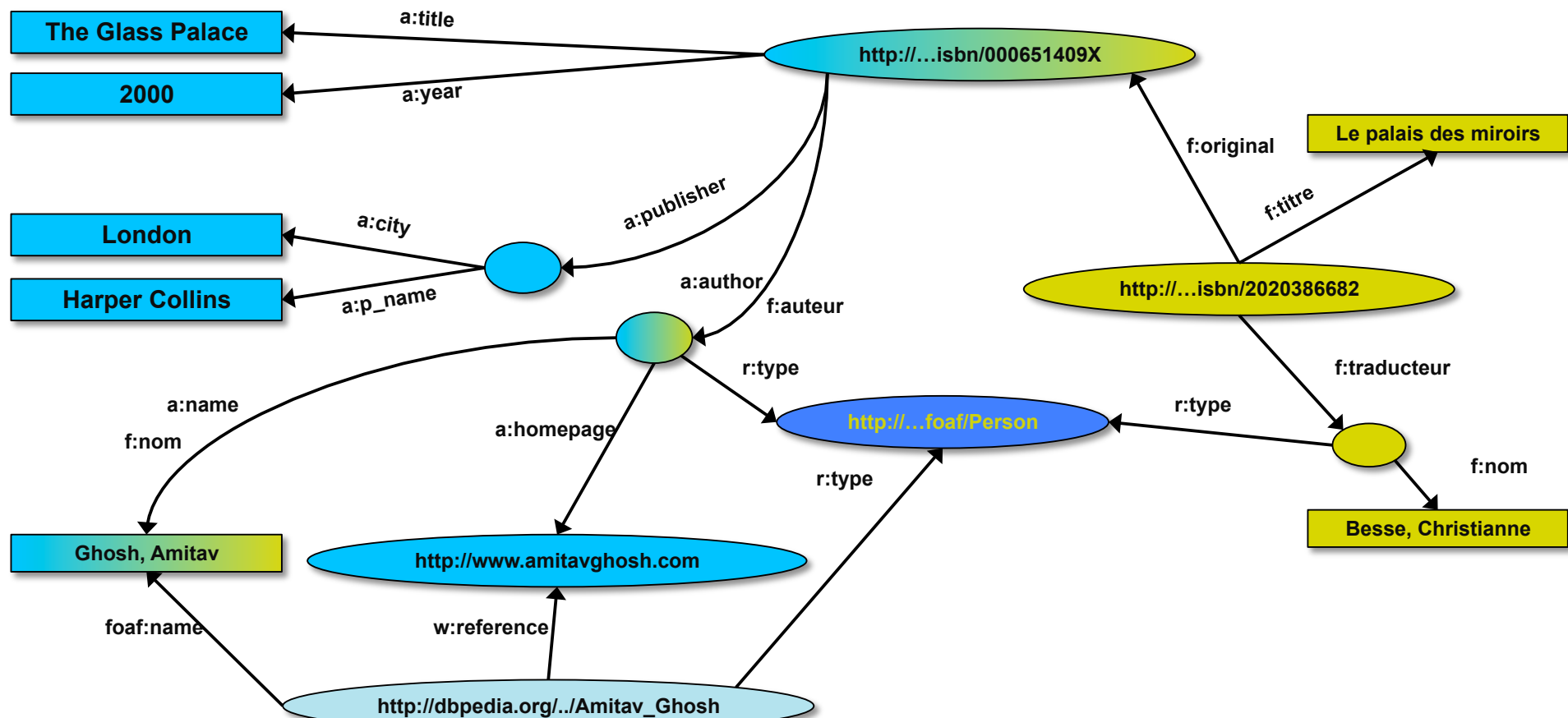
- ▶ User of dataset “F” can now query:
  - “donnes-moi la page d’accueil de l’auteur de l’original”
    - well... “give me the home page of the original’s ‘auteur’”
- ▶ The information is not in datasets “F” or “A”...
- ▶ ...but was made available by:
  - merging datasets “A” and datasets “F”
  - adding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



# Combine with different datasets

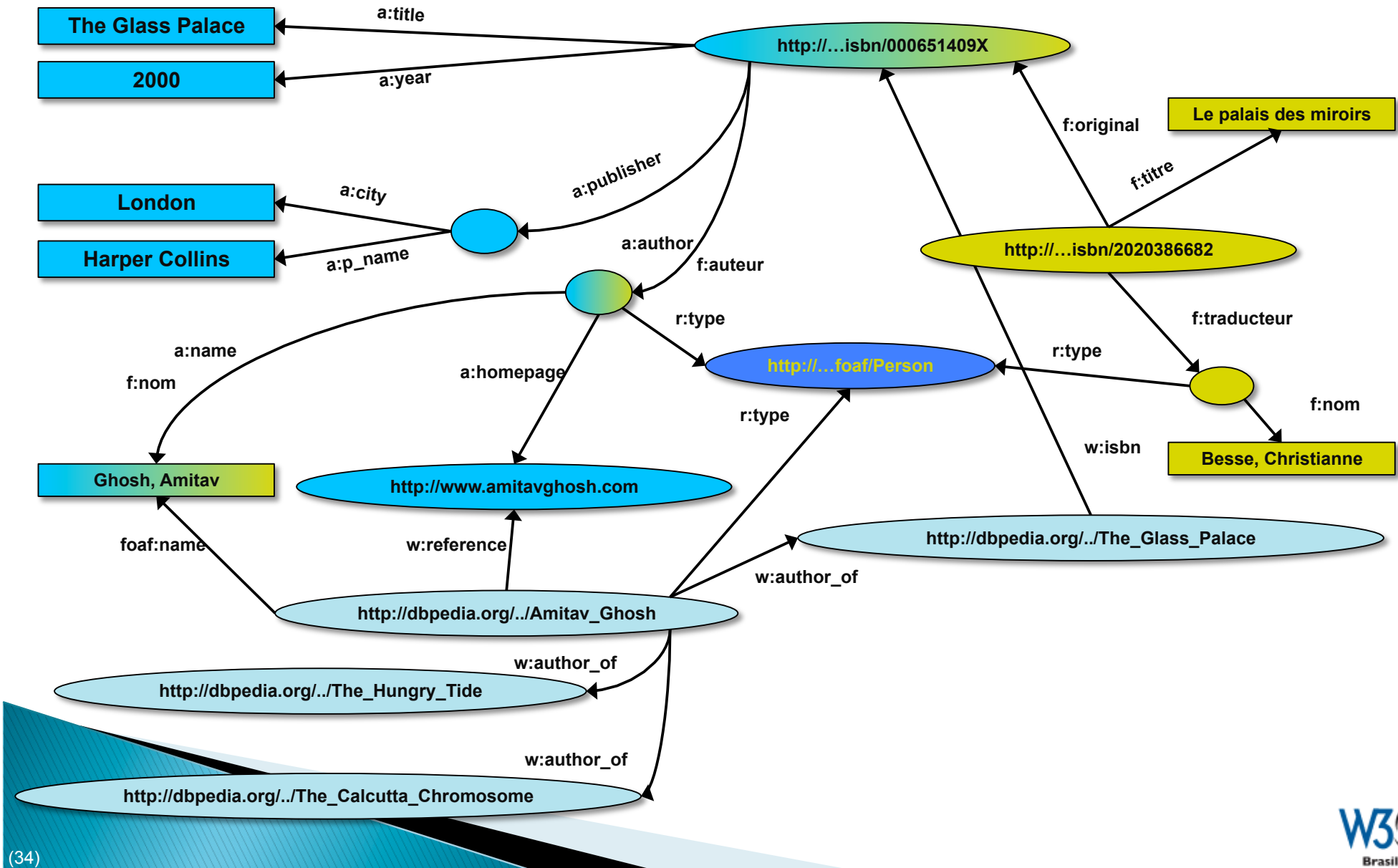
- ▶ Using, e.g., the “Person”, the dataset can be combined with other sources
- ▶ For example, data in Wikipedia can be extracted using dedicated tools
  - e.g., the “[dbpedia](#)” project can extract the “infobox” information from Wikipedia already...

# Merge with Wikipedia data

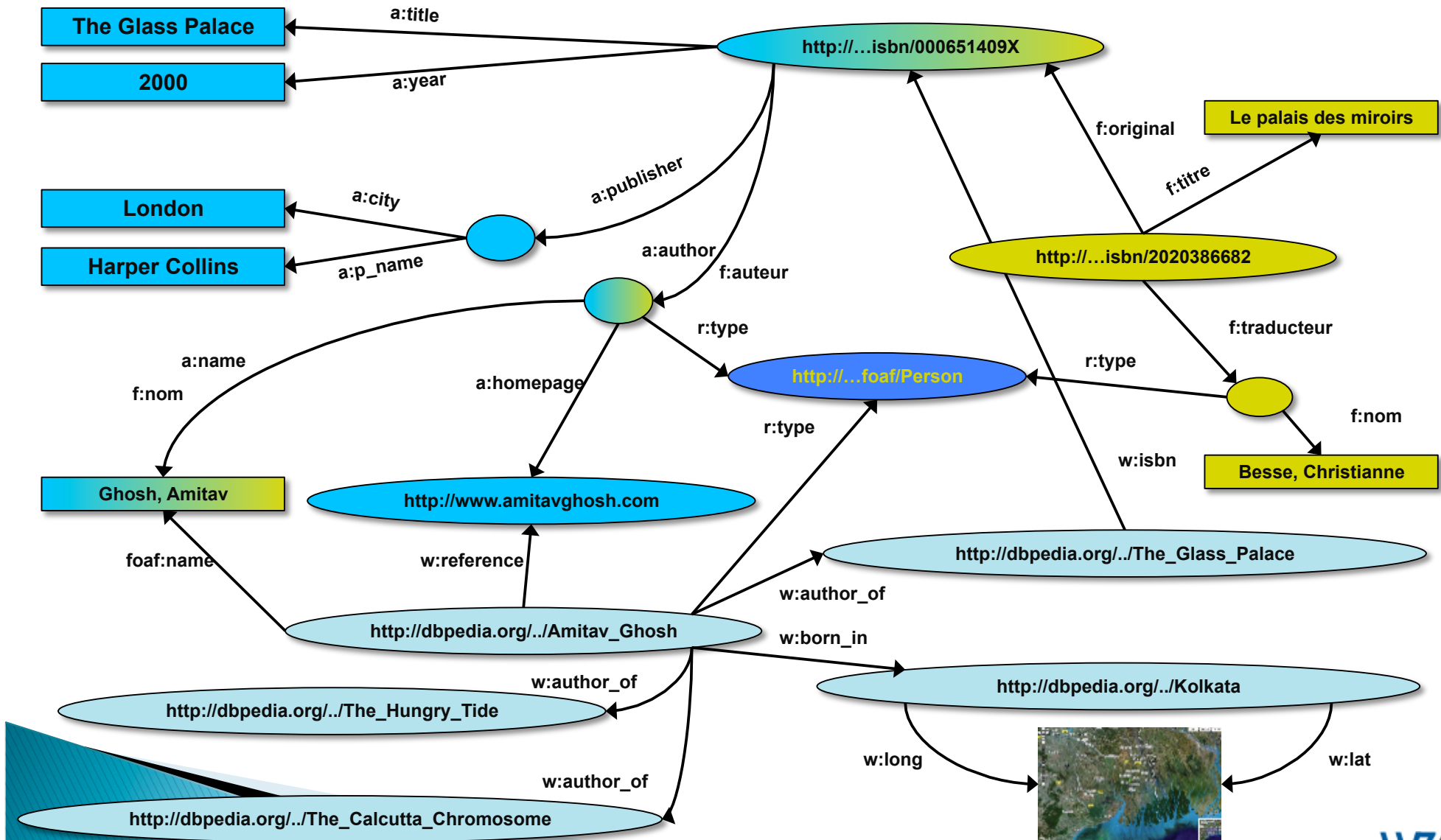




# Merge with Wikipedia data



# 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?

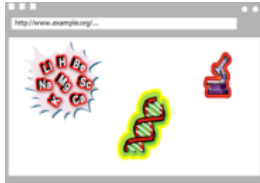
- ▶ We could add some simple additional information (the “glue”), also using common terminologies that a community has produced
- ▶ As a result, new relations could be found and retrieved

# 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

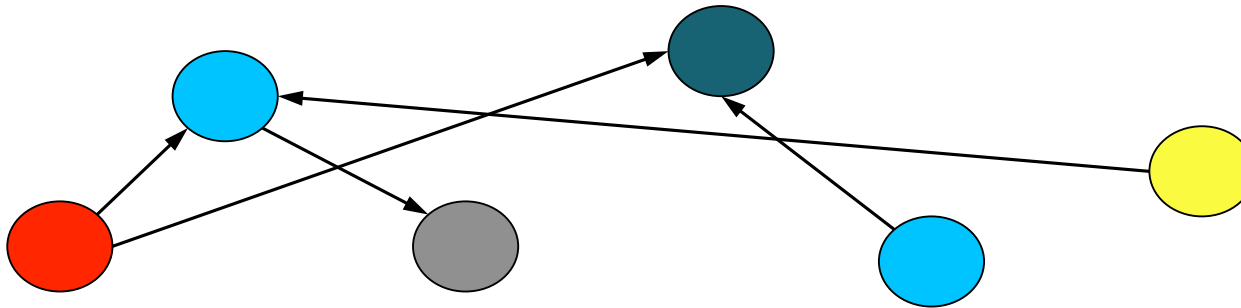


# What did we do? (cont)



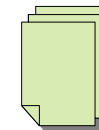
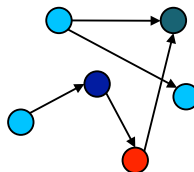
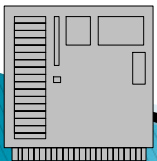
Applications

Manipulate  
Query  
...



Data represented in abstract format

Map,  
Expose,  
...



Data in various formats

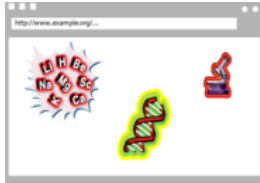
# So what is the Semantic Web?

- ▶ The Semantic Web is a collection of technologies to make such integration of Linked Data possible!

# Details: many different technologies

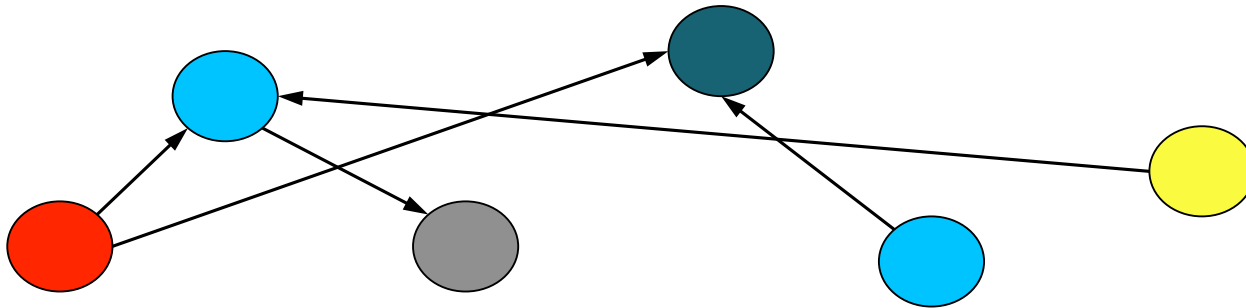
- ▶ an abstract model for the relational graphs: **RDF**
- ▶ add/extract RDF information to/from XML, (X)HTML: **GRDDL, RDFa**
- ▶ a query language adapted for graphs: **SPARQL**
- ▶ characterize the relationships and resources: **RDFS, OWL, SKOS, Rules**
  - applications may choose among the different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

# Using these technologies...



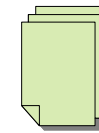
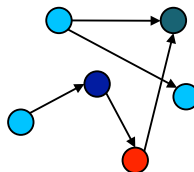
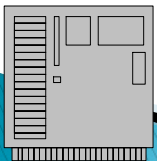
Applications

SPARQL,  
Inferences  
...



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

RDB  $\Leftrightarrow$  RDF,  
GRDDL, RDFa,  
...



Data in various formats

# Remember 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! Bookmarklets Add Zemanta bit.ly To Mendeley TinyURL To Faviki Dokuwiki

MUSIC BETA GENRES ARTISTS REVIEWS NEWS BLOG QUICK FIND Enter an artist name ...

BBC Music > Artists > Eric Clapton

## Eric Clapton

Born 30 March 1945.

MOST PLAYED ON BBC RADIO 2



David Redfern/Redferns

### Biography

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.

### 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

# Remember 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! Bookmarkslets 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](http://ericclapton.com)
- Fanpage at [whereseric.com](http://whereseric.com)
- Wikipedia article on [Eric Clapton](#)
- MySpace at [myspace.com/ericclapton](http://myspace.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 & information come from MusicBrainz. You can add or edit information about [Eric Clapton](#) at [musicbrainz.org](http://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.](#)

# What happens is...

- ▶ Datasets (e.g., MusicBrainz) are published in RDF
- ▶ Some simple vocabularies are involved
- ▶ Those datasets can be queried together via SPARQL
- ▶ The result can be displayed following the BBC style



(47)



# What has been achieved so far?

- ▶ A set of core technologies are in place
- ▶ *Lots* of data (billions of relationships) are available in standard format
  - see the Linked Open Data Cloud

# What has been achieved so far?

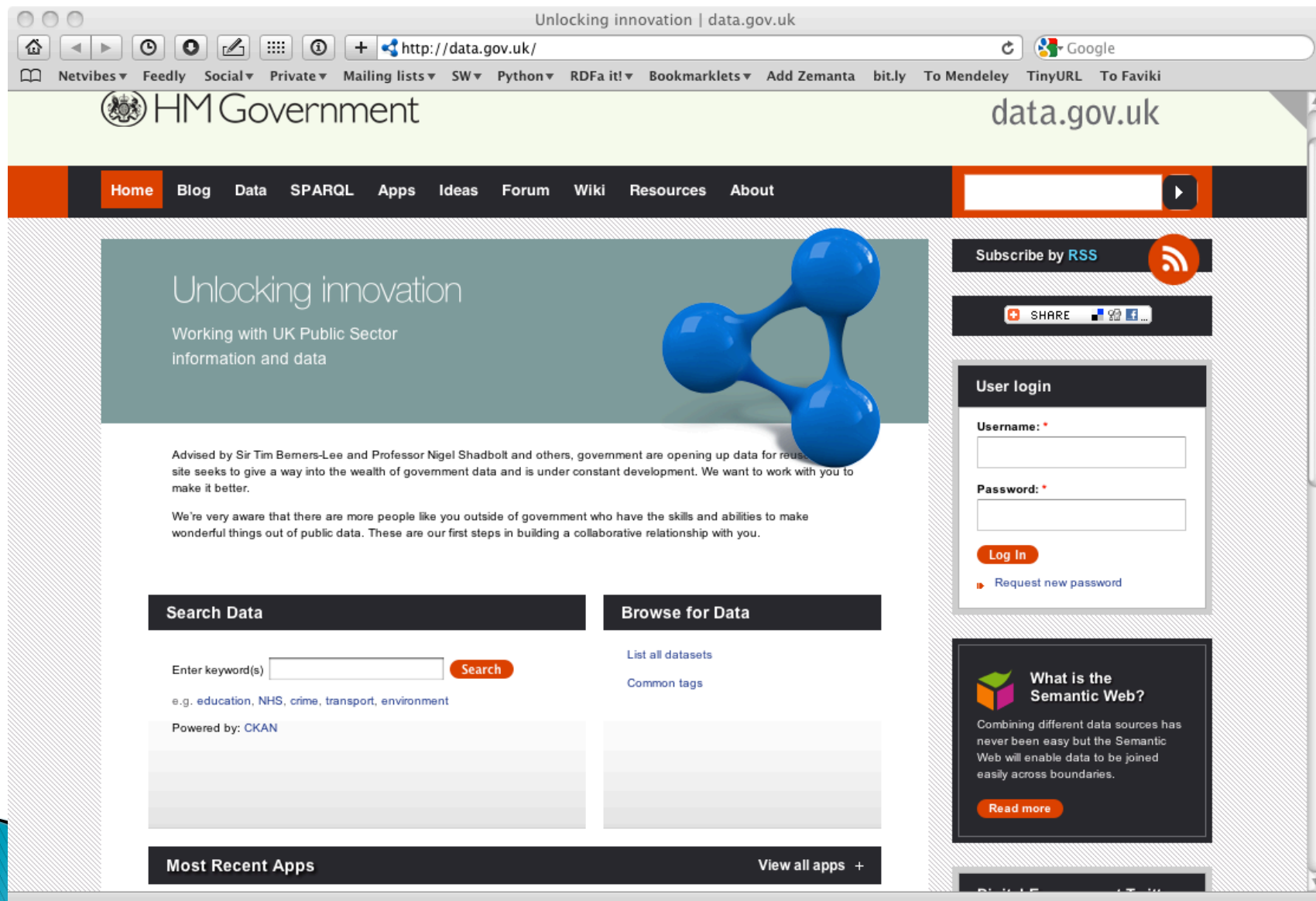
- ▶ There is a vibrant community of
  - academics: universities of Southampton, Oxford, Stanford, PUC
  - small startups: Garlik, Talis, C&P, TopQuadrant, Cambridge Semantics, OpenLink, ...
  - major companies: Oracle, IBM, SAP, ...
  - users of Semantic Web data: Google, Facebook, Yahoo!
  - publishers of Semantic Web data: New York Times, US Library of Congress, open governmental data (US, UK, France,...)

# And, of course, applications emerge

- ▶ Companies, institutions begin to use the technology:
  - BBC, Vodafone, Siemens, NASA, BestBuy, Tesco, Korean National Archives, Pfizer, Chevron, ...
    - see <http://www.w3.org/2001/sw/UseCases>
- ▶ Truth must be said: we still have a way to go
  - deployment may still be experimental, or on some specific places only



# An example for unexpected reuse...

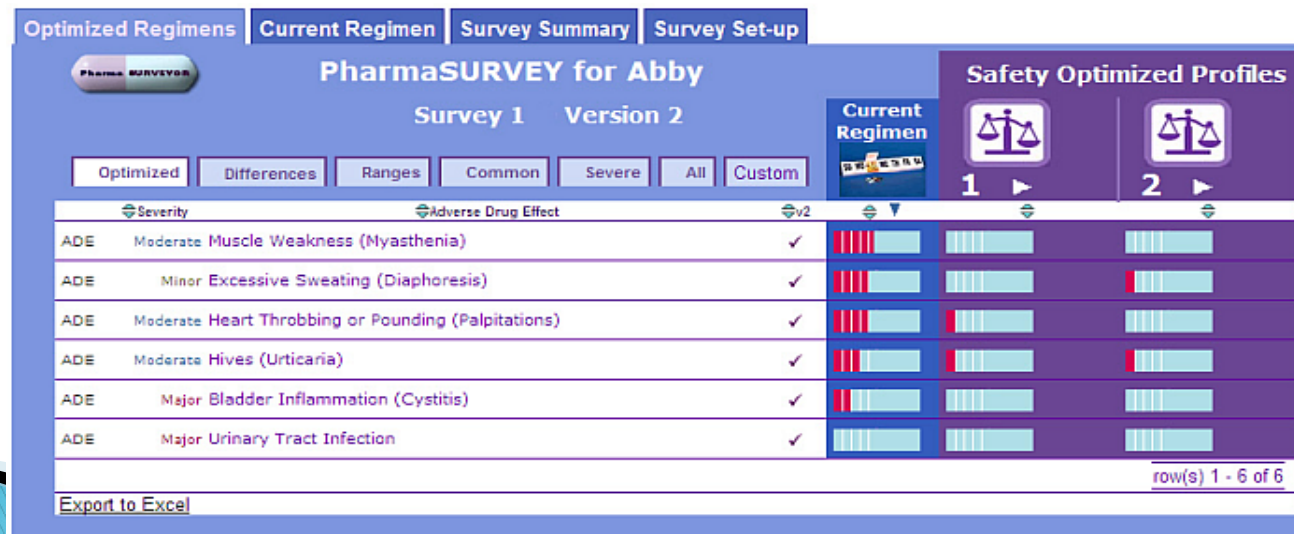


(52)



# Help in choosing the right drug regimen

- ▶ Help in finding the best drug regimen for a specific case, per patient
- ▶ Integrate data from various sources (patients, physicians, Pharma, researchers, ontologies, etc)
- ▶ Data (eg, regulation, drugs) change often, but the tool is much more resistant against change



The screenshot displays the PharmaSURVEYOR software interface. At the top, there are tabs for 'Optimized Regimens', 'Current Regimen', 'Survey Summary', and 'Survey Set-up'. Below these, the title 'PharmaSURVEY for Abby' is shown, along with 'Survey 1' and 'Version 2'. A 'Pharma SURVEYOR' logo is on the left. A navigation bar includes buttons for 'Optimized', 'Differences', 'Ranges', 'Common', 'Severe', 'All', and 'Custom'. The main table compares two regimens across six adverse drug effects (ADEs). The columns are 'Severity', 'Adverse Drug Effect', 'v2', and two 'Current Regimen' columns (1 and 2). Each ADE row shows a severity level, the ADE name, a checkmark in the 'v2' column, and bar charts in the 'Current Regimen' columns. The ADEs listed are: Moderate Muscle Weakness (Myasthenia), Minor Excessive Sweating (Diaphoresis), Moderate Heart Throbbing or Pounding (Palpitations), Moderate Hives (Urticaria), Major Bladder Inflammation (Cystitis), and Major Urinary Tract Infection. At the bottom right, it says 'row(s) 1 - 6 of 6' and there is an 'Export to Excel' link.

| Severity     | Adverse Drug Effect                        | v2 | Current Regimen 1 | Current Regimen 2 |
|--------------|--------------------------------------------|----|-------------------|-------------------|
| ADE Moderate | Muscle Weakness (Myasthenia)               | ✓  | [Bar Chart]       | [Bar Chart]       |
| ADE Minor    | Excessive Sweating (Diaphoresis)           | ✓  | [Bar Chart]       | [Bar Chart]       |
| ADE Moderate | Heart Throbbing or Pounding (Palpitations) | ✓  | [Bar Chart]       | [Bar Chart]       |
| ADE Moderate | Hives (Urticaria)                          | ✓  | [Bar Chart]       | [Bar Chart]       |
| ADE Major    | Bladder Inflammation (Cystitis)            | ✓  | [Bar Chart]       | [Bar Chart]       |
| ADE Major    | Urinary Tract Infection                    | ✓  | [Bar Chart]       | [Bar Chart]       |

row(s) 1 - 6 of 6

[Export to Excel](#)



# eTourism: provide personalized itinerary

The screenshot displays the Zaragoza Turismo website interface. At the top, there's a navigation bar with 'TOURISM' and 'THE CITY COUNCIL' tabs. A sidebar on the left lists various services like 'Tourist Services', 'What to visit', and 'Guided Visit'. The main content area is titled 'ITINERARY FOR 17/06/08' and includes a 'Proposed routes' section with 'Morning' and 'Afternoon' itineraries. The 'Morning' itinerary lists sites like the Basilica of the Pilar and the Ibercaja Camón Aznar Museum. The 'Afternoon' itinerary lists sites like the Church of la Mantería and the Church of San Ildefonso. A 'Zaragoza street plan' section shows a map of the city with a highlighted route. The website also features a 'Back to Create' button and a 'Quality' link in the sidebar.

- ▶ Integration of relevant data in Zaragoza
- ▶ Use rules to provide a proper itinerary

# 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
- ▶ But we are getting there!

# Why is all this good?

- ▶ A huge amount of data (“information”) is available on the Web
- ▶ Sites struggle with the dual task of:
  - providing quality data
  - providing usable and attractive interfaces to access that data

# Why is all this good?

- ▶ Semantic Web technologies allow a separation of tasks:
  1. publish quality, interlinked datasets
  2. “mash-up” datasets for a better user experience

*“Raw Data Now!”*

*Tim Berners-Lee, TED Talk, 2009*

<http://bit.ly/dq7H7Z>

# Why is all this good?

- ▶ The “network effect” is also valid for data
- ▶ There are unexpected usages of data that authors may not even have thought of
- ▶ “Curating”, using, exploiting the data requires a different expertise

# Thank you for your attention!

These slides are also available on the Web:

<http://www.w3.org/2010/Talks/1015-SauPaulo-SemCafe-IH/>

