

Introduction

Werner Nutt

Acknowledgment

These slides are based on the slide sets

- RDF
by Mariano Rodriguez
- Overview and XML
for the course Semantic Web Technologies at TU Dresden
by Sebastian Rudolph

- The Semantic Web Idea
- Semantic Technologies

- The Semantic Web Idea
- Semantic Technologies

The Web

All aspects of daily **life** are being “**webized**”:

- everyday information (news, weather forecast)
- entertainment
- buying and selling
- administration (eGovernment)
- education (eLearning, eEducation)
- social contacts (social networking platforms, dating sites)

Why Semantic Web?

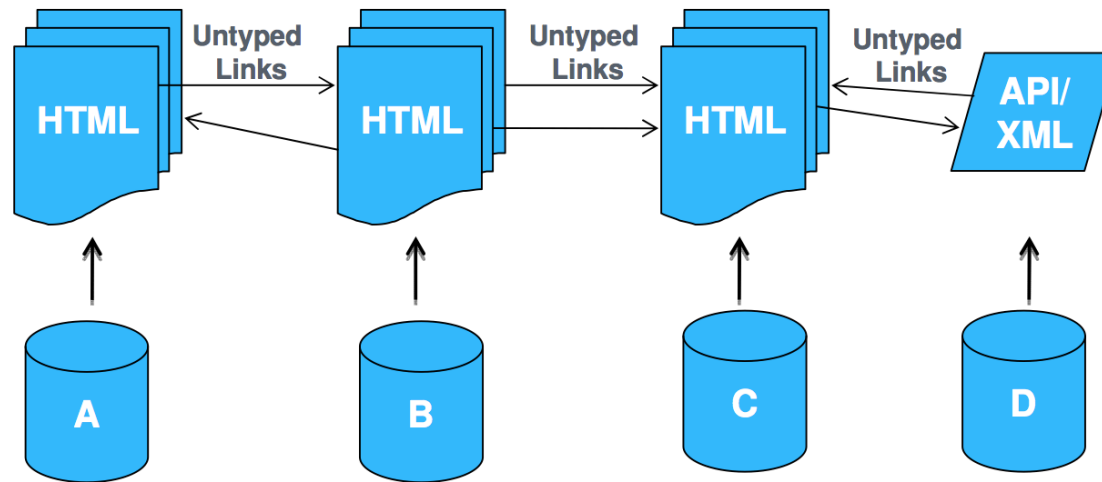
Syntax vs. Semantics

Syntax (from greek *συνταξις* composition, sentential structure) denotes the (normative) **structure of data**, i.e., it characterizes what makes data “well-formed”

Semantics (greek *σημαυτικός* belonging to the sign) denotes the **meaning of data**, i.e., it characterizes what conclusions can be drawn from it

- “4+) = (“ syntactically wrong, semantically ...
- “3 + 4 = 12” syntactically correct, semantically wrong
- “3 + 4 = 7” syntactically correct, semantically correct

Web of Documents



- Primary objects: **documents**
- Degree of structure in data: **low**
- Semantics of content: **implicit**
- Designed for: **human consumption**

Links between **documents**

Problems of the Web of Documents

- wealth of **information**
- ... targeted at **human users**

Ian Horrocks FRS

Professor of Computer Science and Fellow of [Oriol College](#)

INTERESTS

My research interests include knowledge representation, [ontologies](#) and [ontology languages](#), modal and [description logics](#), [automated reasoning](#), implementation and optimisation of [reasoning systems](#), and applications in areas such as [e-Science](#) and the [Semantic Web](#).

RECENT PUBLICATIONS

My [recent publications](#) can be found [here](#), and a [complete list](#) of my publications [here](#).



NAME

Ian Horrocks

POSTS

Professor of Computer Science, Fellow of Oriol College

PROJECTS

[ExODA](#), [Score!](#), [Optique](#)

CONTACT INFORMATION

+44 (0)1865 273939

Room 304
Wolfson Building
Parks Road
Oxford OX1 3QD

ian.horrocks@cs.ox.ac.uk

LINKS

[Projects](#)

[Tools](#)

[Ontologies](#)

Problems of the Web of Documents

- wealth of information
- ... targeted at human users

SEBASTIAN RUDOLPH

Position

Head of the Group

Email address

sebastian.rudolph@tu-dresden.de

Telephone

+49 (0)351 463 38516

Office

Room 2035, Nöthnitzer Straße 46,
Dresden-Räcknitz

External page

<http://www.sebastian-rudolph.de/>

**CURRICULUM VITAE**

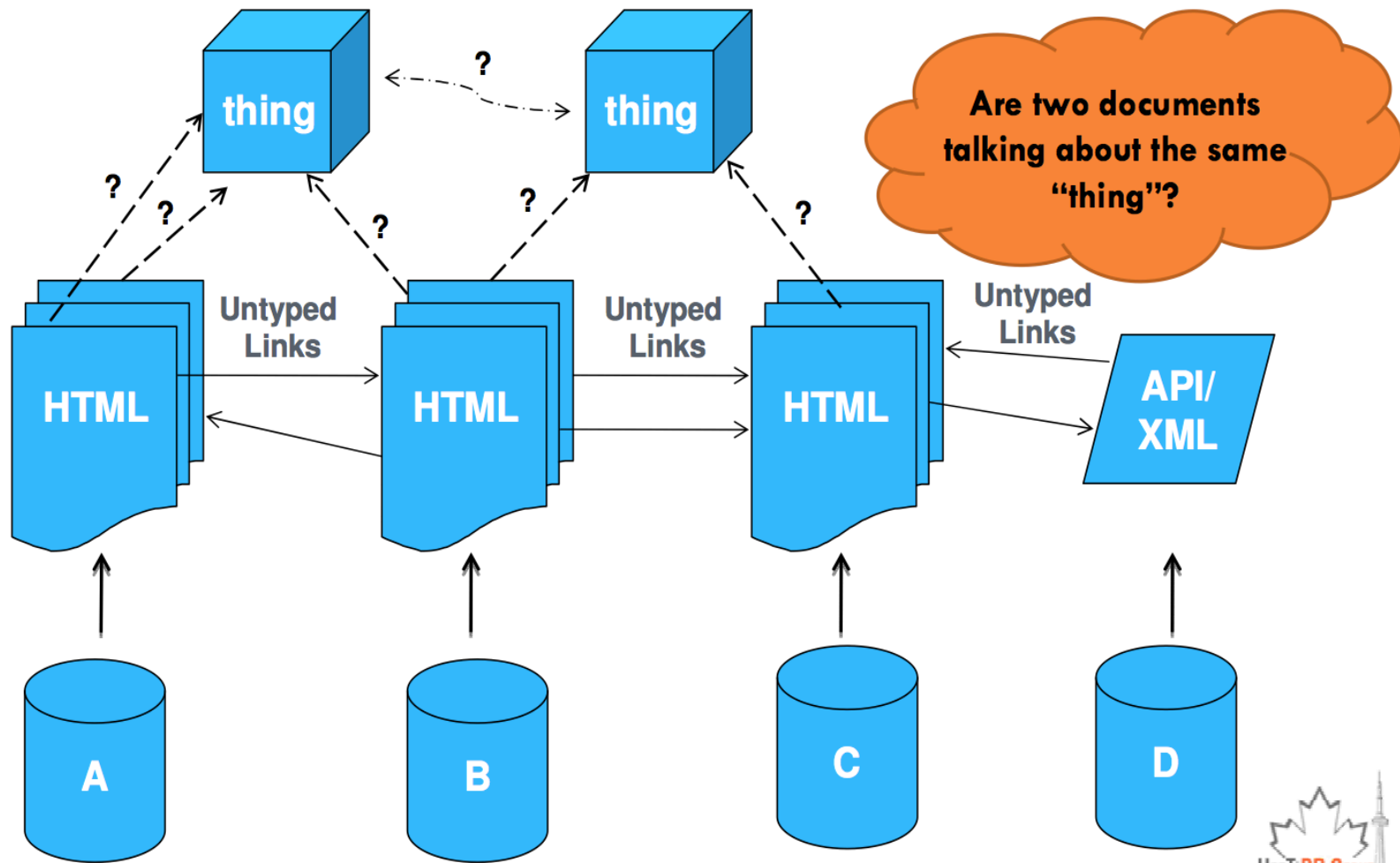
- 1995 - 2000 studies for high school teaching
(mathematics/physics/computer science), [Dresden University](#)
- 2000 - 2003 scholarship holder "Graduiertenkolleg 334", [Dresden University](#)
- 2003 - 2005 research assistant at the [chair for Psychology of Teaching and Learning](#), [Dresden University](#)
- 2006 dissertation in Mathematics (Dr.rer.nat.) at the [Institute for Algebra](#), [Dresden University](#)

Pages with Computer-readable Layout Information

```
<h1 class="documentFirstHeading">Sebastian Rudolph</h1>
<div id="bodyContent" class="plain">

<dl>
<dt>Position</dt>
<dd>Head of the Group</dd>
<dt>Email address</dt>
<dd>sebastian.rudolph@tu-dresden.de</dd>
<dt>Telephone</dt>
<dd>+49 (0)351 463 38516</dd>
<dt>Office</dt>
<dd>Room 2035, Nöthnitzer Straße 46, Dresden-Räcknitz</dd>
<dt>External page</dt>
</dl>
<dl>
<dd><a href="http://www.sebastian-rudolph.de/">http://www.sebastian-
rudolph.de/</a></dd>
</dl>
<h2>Curriculum vitae</h2>
<ul>
```

Web of Documents



Problems of the Web of Documents

- **localizing information** problematic
- today's search engines are good,
but still **keyword-based**

The screenshot shows a Google search interface with the query 'kohl' entered in the search bar. The search results are displayed in a grid format, with the 'Images' tab selected. The results are categorized into five groups, each with a representative image and a label below it:

- Eyeliner**: A collage of images showing various eyeliners and close-ups of eyes.
- Gemüse**: A collage of images showing various types of green vegetables, including cabbages and leafy greens.
- Cosmetic**: A collage of images showing a young girl, a woman's face, and various cosmetic products.
- Egyptian**: A collage of images showing an Egyptian mortar and pestle, a woman's face with blue eye makeup, and various cosmetic products.
- Eyes**: A collage of images showing various eye makeup products and close-ups of eyes.

Below the main grid, there are five larger images that are not part of the search results grid:

- A portrait of a man with glasses.
- A portrait of a man in a suit and tie, with a European Union flag in the background.
- A collage of images showing various eye makeup products and close-ups of eyes.
- A collage of images showing various types of green vegetables, including cabbages and leafy greens.

Problems of the Web of Documents

- How about this **query**:
 - How many romantic comedy Hollywood movies are directed by a person who is born in a city that has average temperature above 15 degrees!?
- You need to:
 - Find **reliable sources** containing facts about movies (genre & director), birthplaces of famous artists/directors, average temperature of cities across the world, etc.
 - ➔ The result: several lists of thousands of facts
 - **Integrate all the data**, join the facts that come from heterogeneous sources

Even if possible, it may take hours to answer just a single query!

Problems of the Web of Documents

Heterogeneity of present **information** on diverse levels:

- character encoding (e.g. ASCII vs. Unicode)
- used natural languages
- positioning of information on web pages
- desirable: information integration across the Web

Problems of the Web of Documents

- **implicit knowledge**, i.e. many pieces of information are not provided explicitly, but follow from the combination of the given data
- requires “logical thinking”
- ➔ automated deduction



Problems of the Web of Documents

Approaches toward a solution:

- ① Ad hoc: **Deployment of AI methods**
(most notably NLP techniques) to evaluate
existing unstructured information on the Web

- ② A priori: **Structure information on the Web**
at authoring time
in a way facilitating later automated deployment

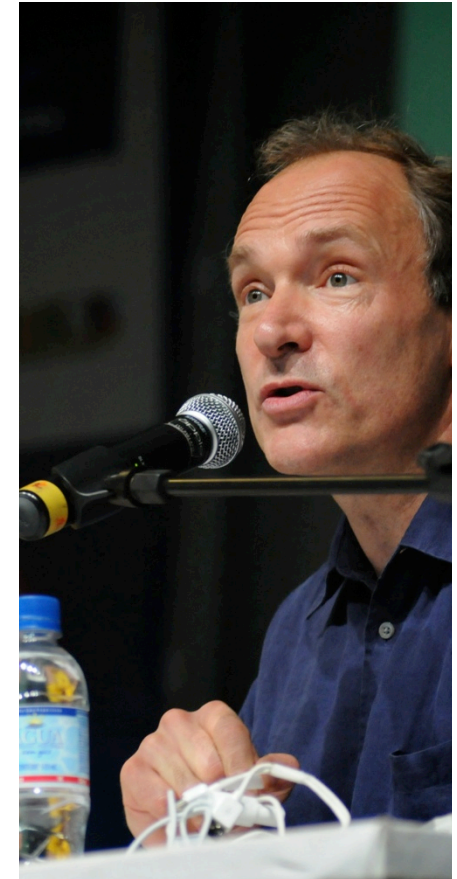
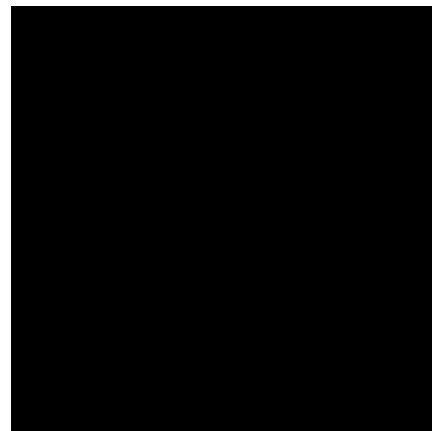
➔ **Semantic Web**



The Vision

I have a dream for the Web in which computers become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The intelligent agents people have touted for ages will finally materialize.

Berners-Lee, 1999



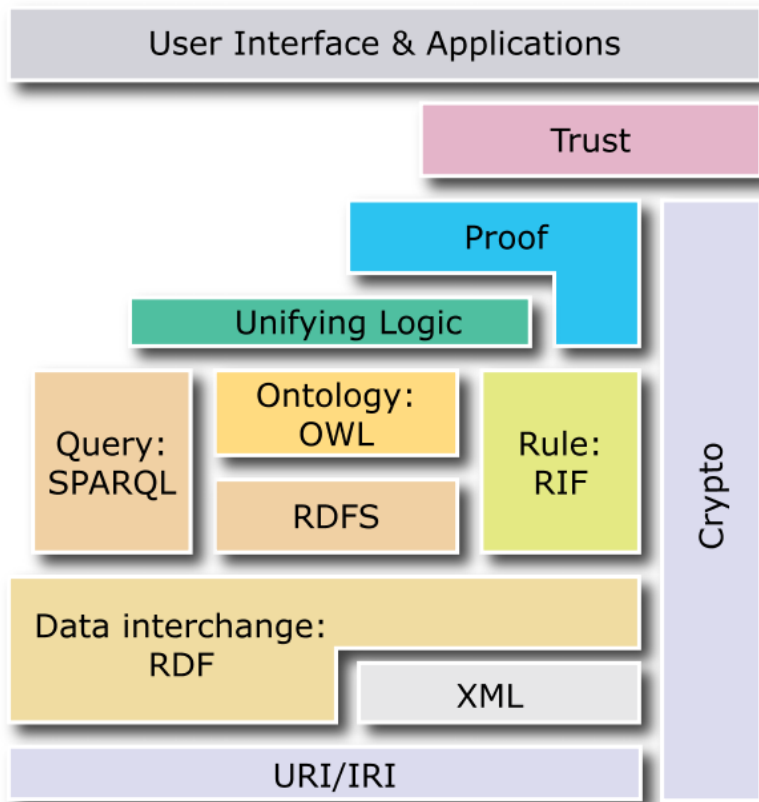
Semantic Web

Two essential prerequisites for the implementation:

- ① **open standards** for describing information
 - clearly defined
 - flexible
 - extendable

- ② **methods for retrieving information**
from such descriptions

Semantic Web: Standards



- 1994 First public presentation of the Semantic Web idea
- 1998 Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
- 2000 Start of large research projects about ontologies in the US and Europe
- 2002 Start of standardization of a new ontology language (OWL) based on research results
- 2004 Finalization of the standard for data (RDF) and ontology (OWL)
- 2008 Standardization of a query language (SPARQL)
- 2009 Extension of OWL to OWL 2.0
- 2010 Standard Rule Interchange Format (RIF)
- 2013 SPARQL 1.1

- The Semantic Web Idea
- Semantic Technologies

Not Only About the Web

- The Semantic Web vision has generated **technologies** that are applied **outside the web context** including:
 - Retailing, supermarkets (see web pages of Best Buy)
 - Health care, medicine (see SNOMED ontology)
 - Enterprise intelligence (see solutions by Ontotext)
 - Research: Bio, Geo, Cultural heritage, etc.
(see this week's KRDB seminar by Alasdair Gray)
 - Government (“smart cities”)
 - Software development
(“ontology-driven software development”)
 - ...

Introduction to the Semantic Web Approach

How does the Semantic Web approach help us

- *merge data sets*
- *infer new relations between data items, and*
- *integrate new data sources?*

Sketch of Data Integration with Semantic Technologies

1. Map the various data onto an *abstract data representation*
 - ➔ Make the data independent of its internal representation ...
2. *Merge* the resulting representations
3. Start *querying* the whole merged set
 - ➔ We can answer queries that were impossible over the individual data sets

Data Set “A”: A *Simple Book Store*

Books

ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

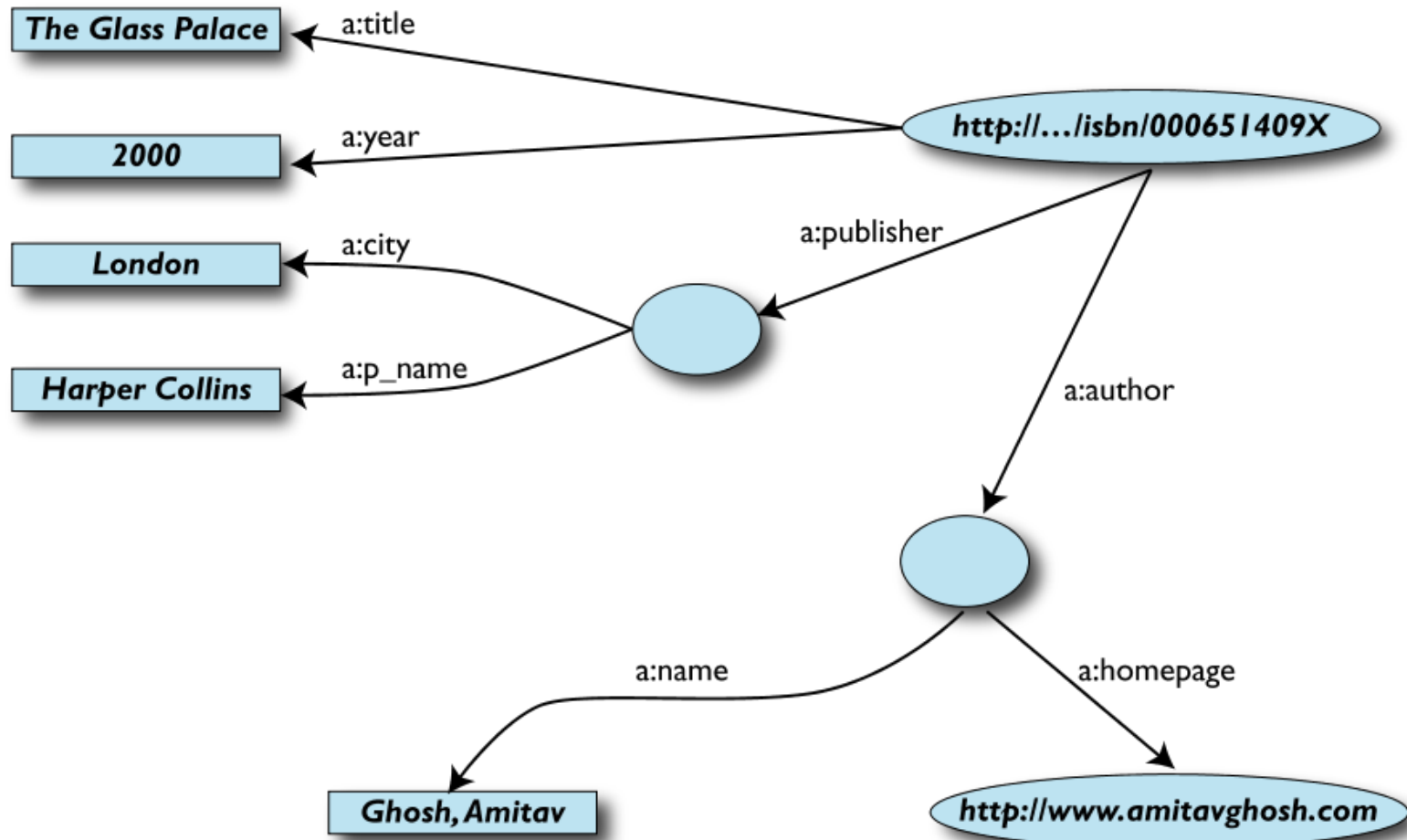
Authors

ID	Name	Home page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

Publishers

ID	Publisher Name	City
id_qpr	Harper Collins	London

1st: Export Your Data as a set of Binary Relations (= Labeled Graph)



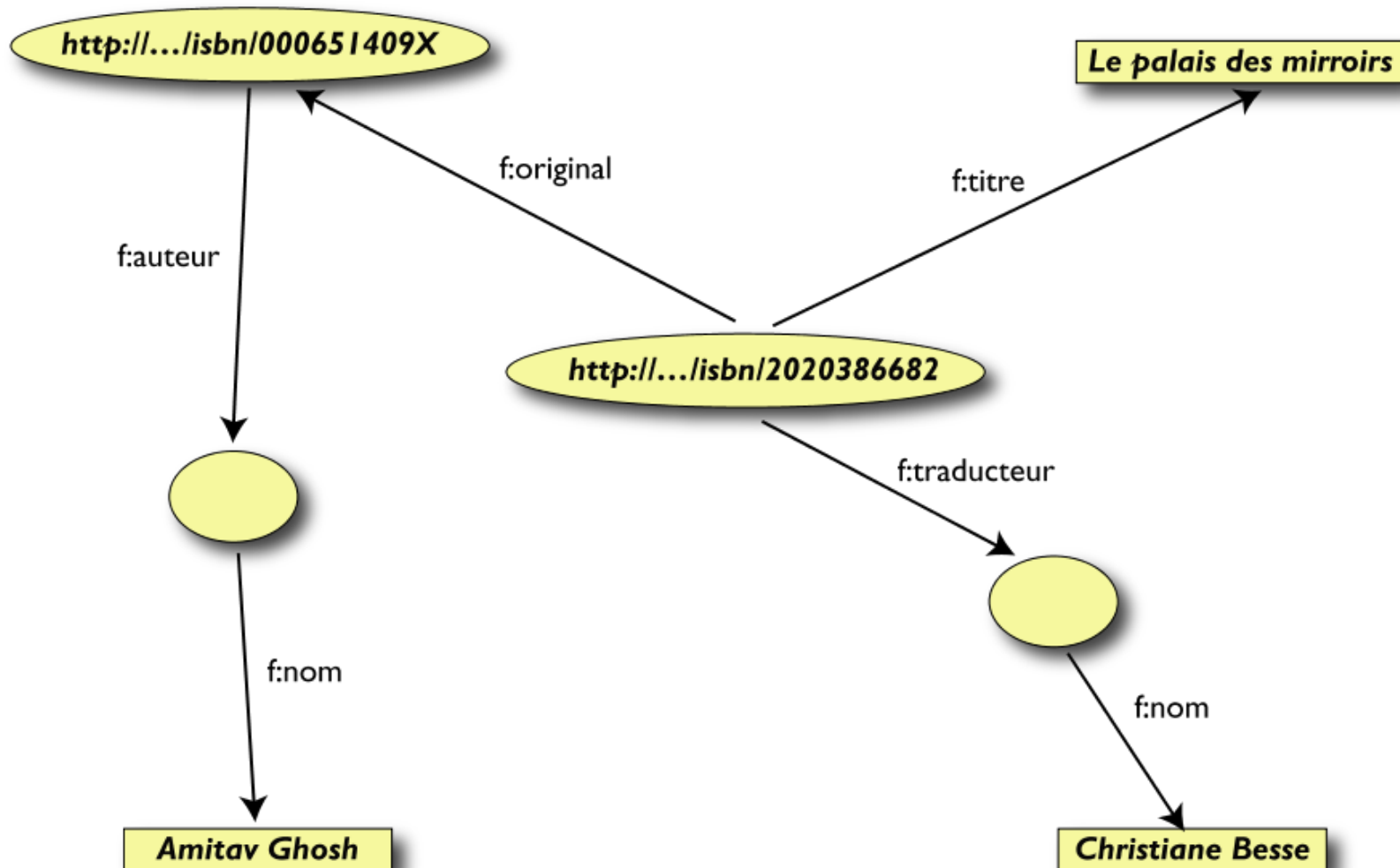
Remarks on the Data Export

- Data export does **not** necessarily mean **physical conversion** of (all) the data
 - Relations can be virtual, 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

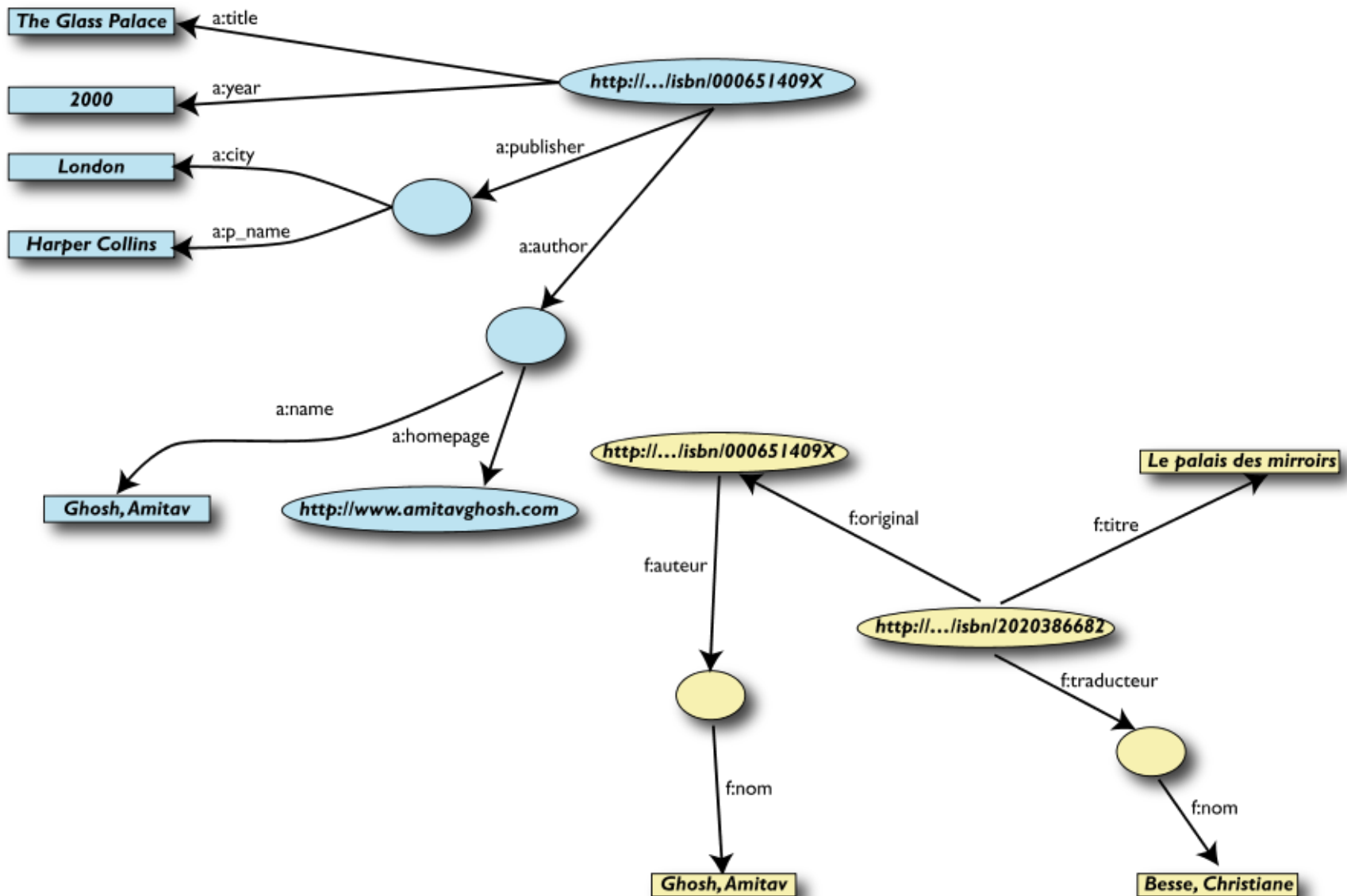
Data set “F”: Another Book Store’s Data

	A	B	D	E
1	ID	Titre	Traducteur	Original
2	ISBN0 2020386682	Le Palais des miroirs	A13	ISBN-0-00-651409-X
3				
6	ID	Auteur		
7	ISBN-0-00-651409-X	A12		
11	Nom			
12	Ghosh, Amitav			
13	Besse, Christianne			

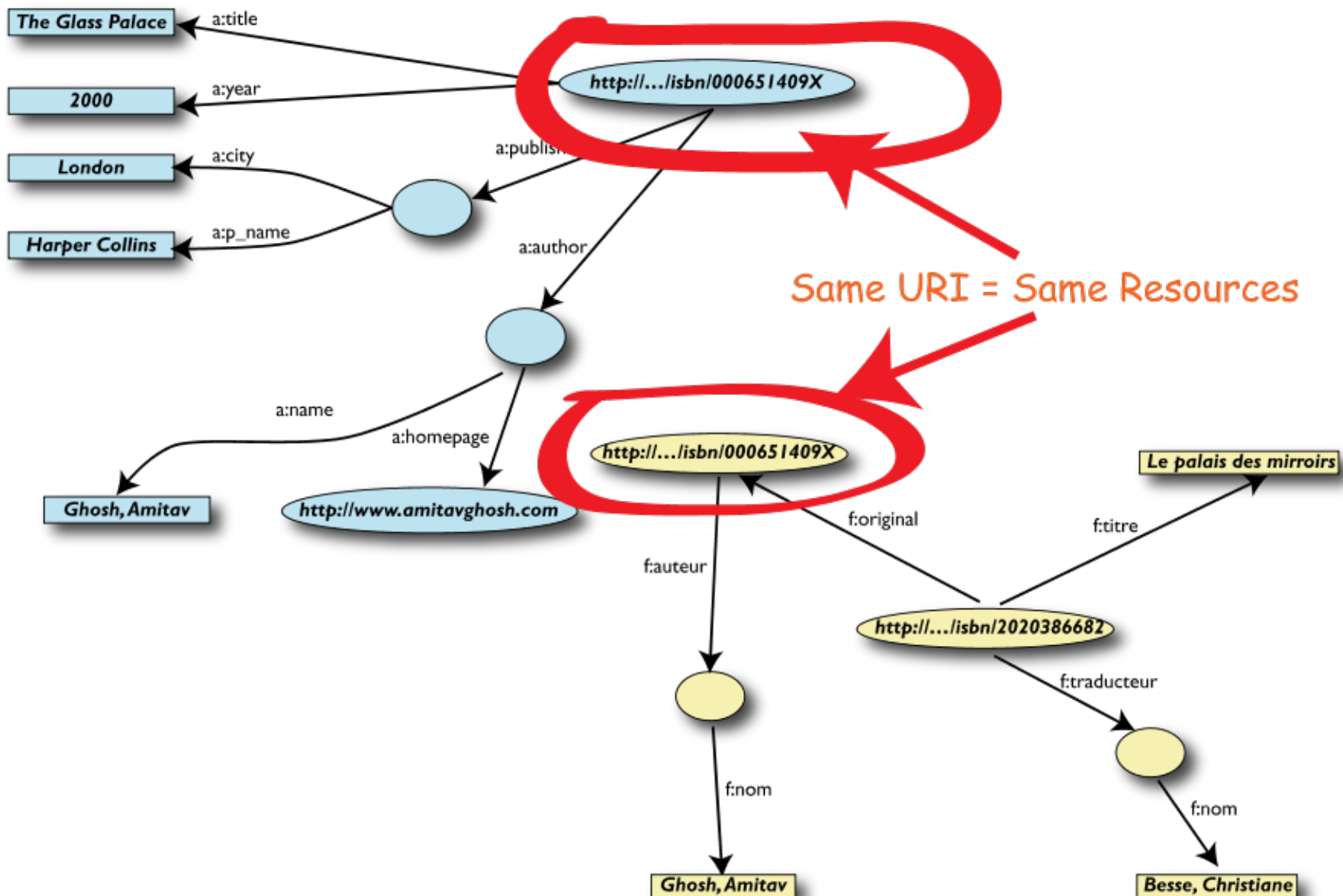
2nd: Export Your Second Set of Data



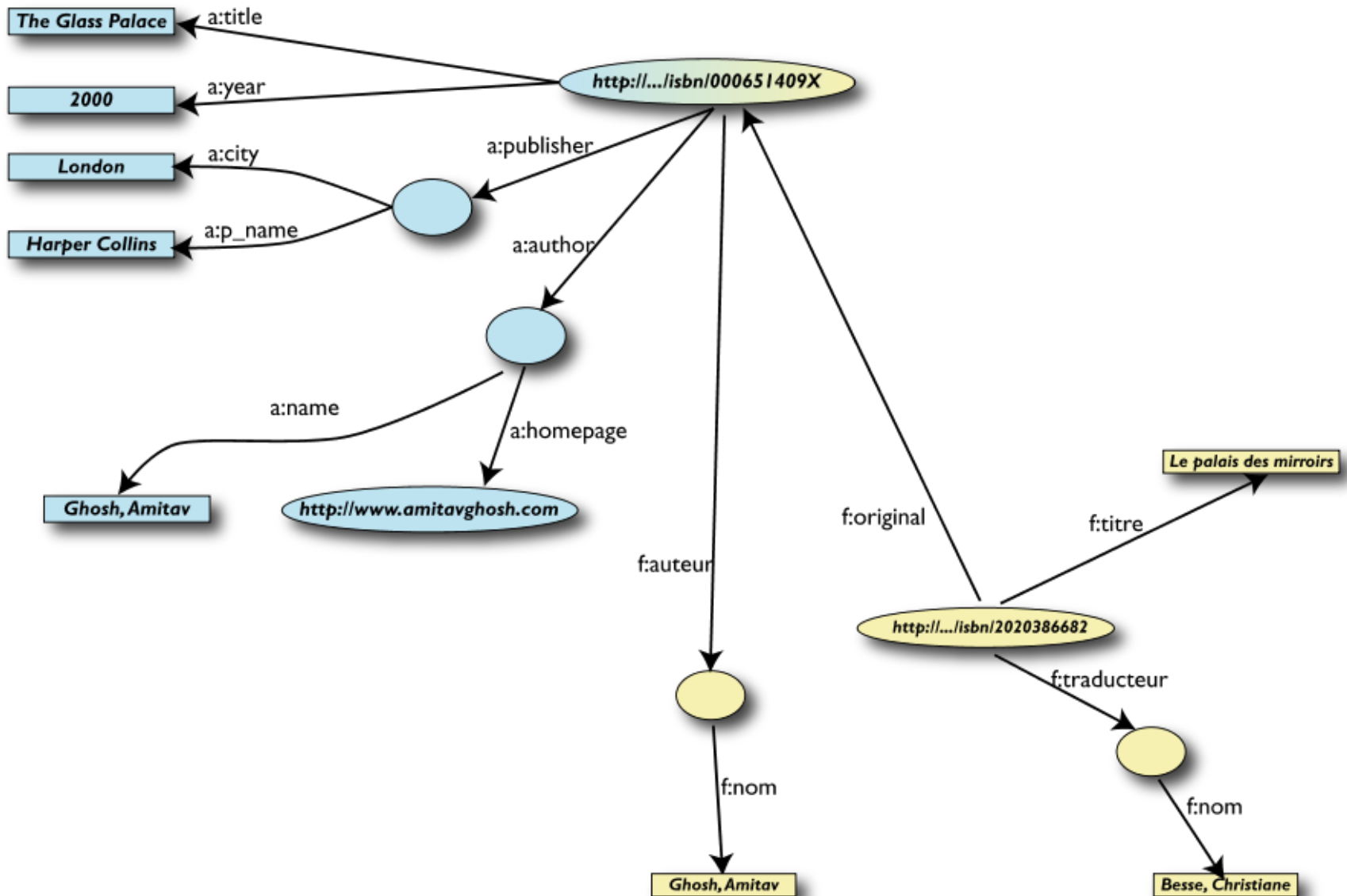
3rd: Start Merging Your Data



3rd: Start Merging Your Data (cont'd)



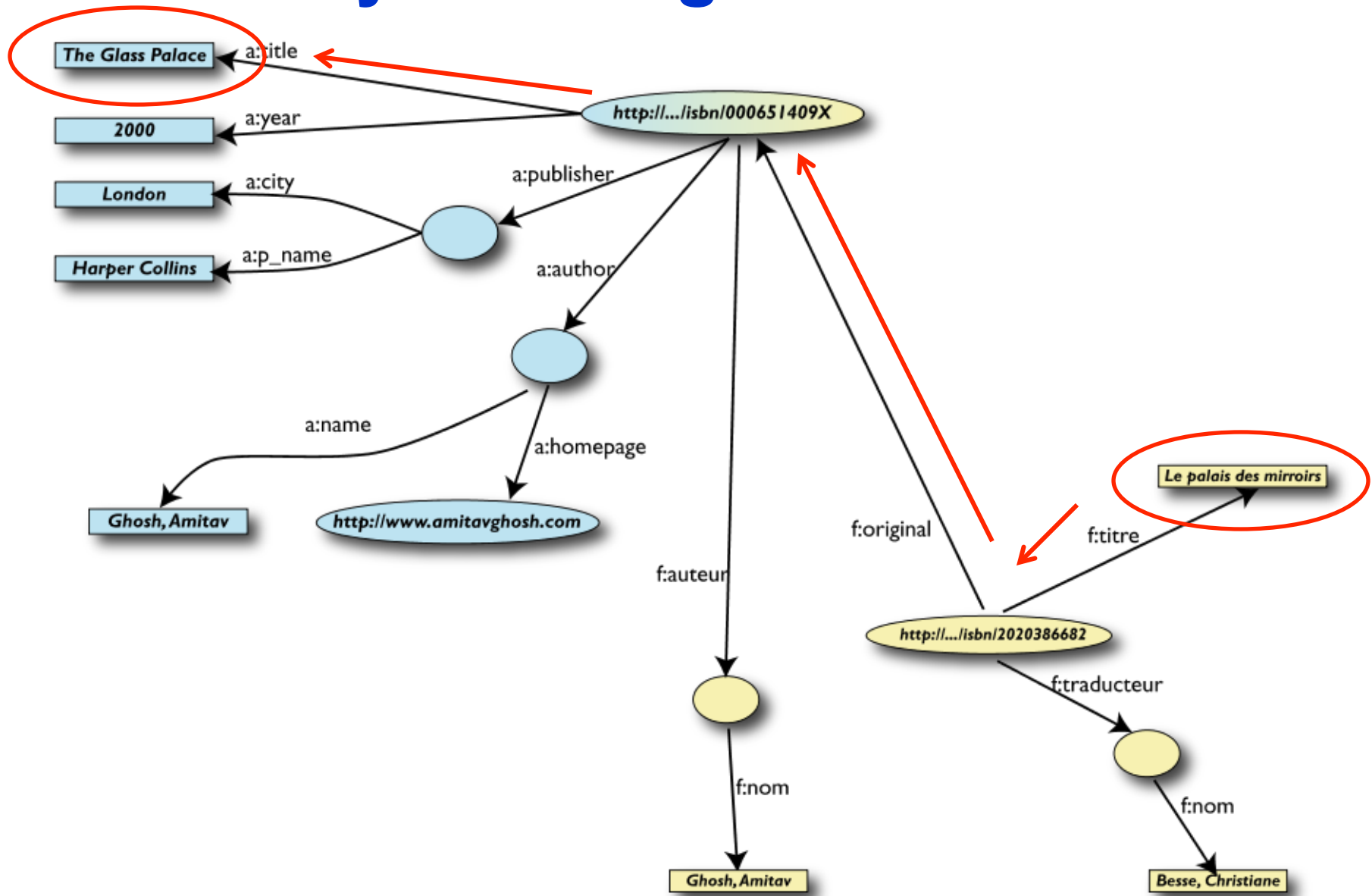
4th: Merge Identical Resources



Start Posing Queries...

- User of data set “F” can now ask queries like:
 - “What is the title of the original version of *Le Palais des miroirs*?”
- This information is not in the data set “F”...
- ...but can be retrieved after merging it with data set “A”!

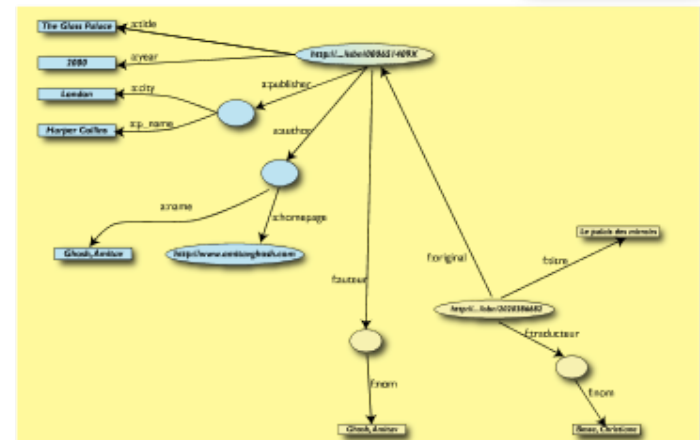
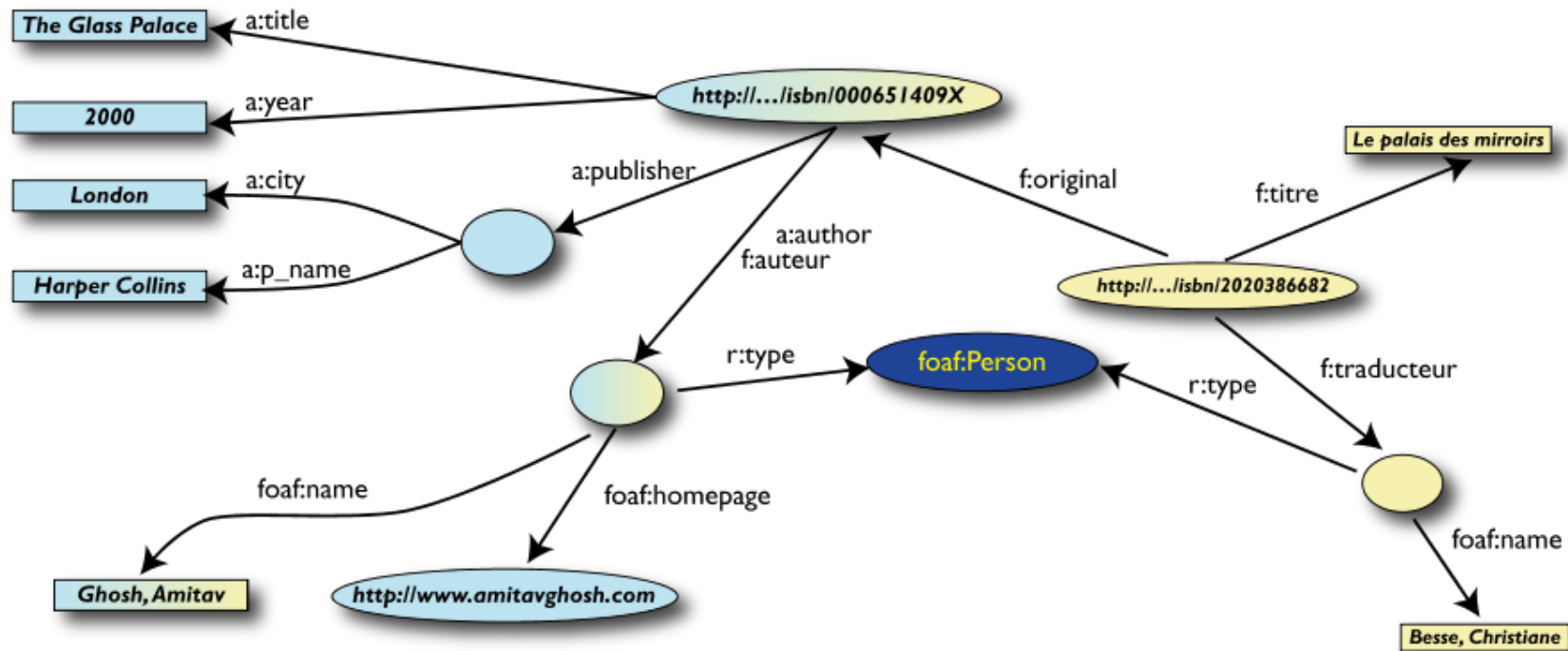
5th: Query the Merged Data Set



However, We Can Achieve More ...

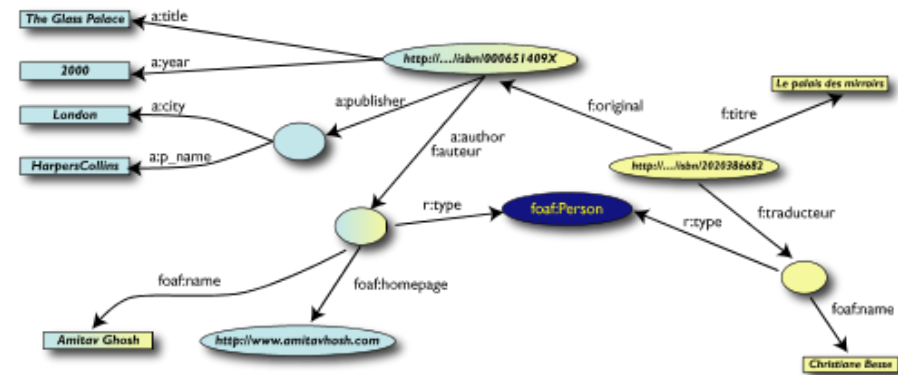
- We “know” that *a:author* and *f:auteur* are really the same
- But our automatic merge does not know that!
- Let us add some extra information to the merged data:
 - *a:author* is **equivalent to** *f:auteur*
both denote a *Person*,
a category (type) for certain resources
 - *a:name* and *f:nom* are equivalent to *foaf:name*
 - *a person is uniquely identified by his/her name*
(= *foaf:name*)

3rd Revisited: Use the Extra Knowledge

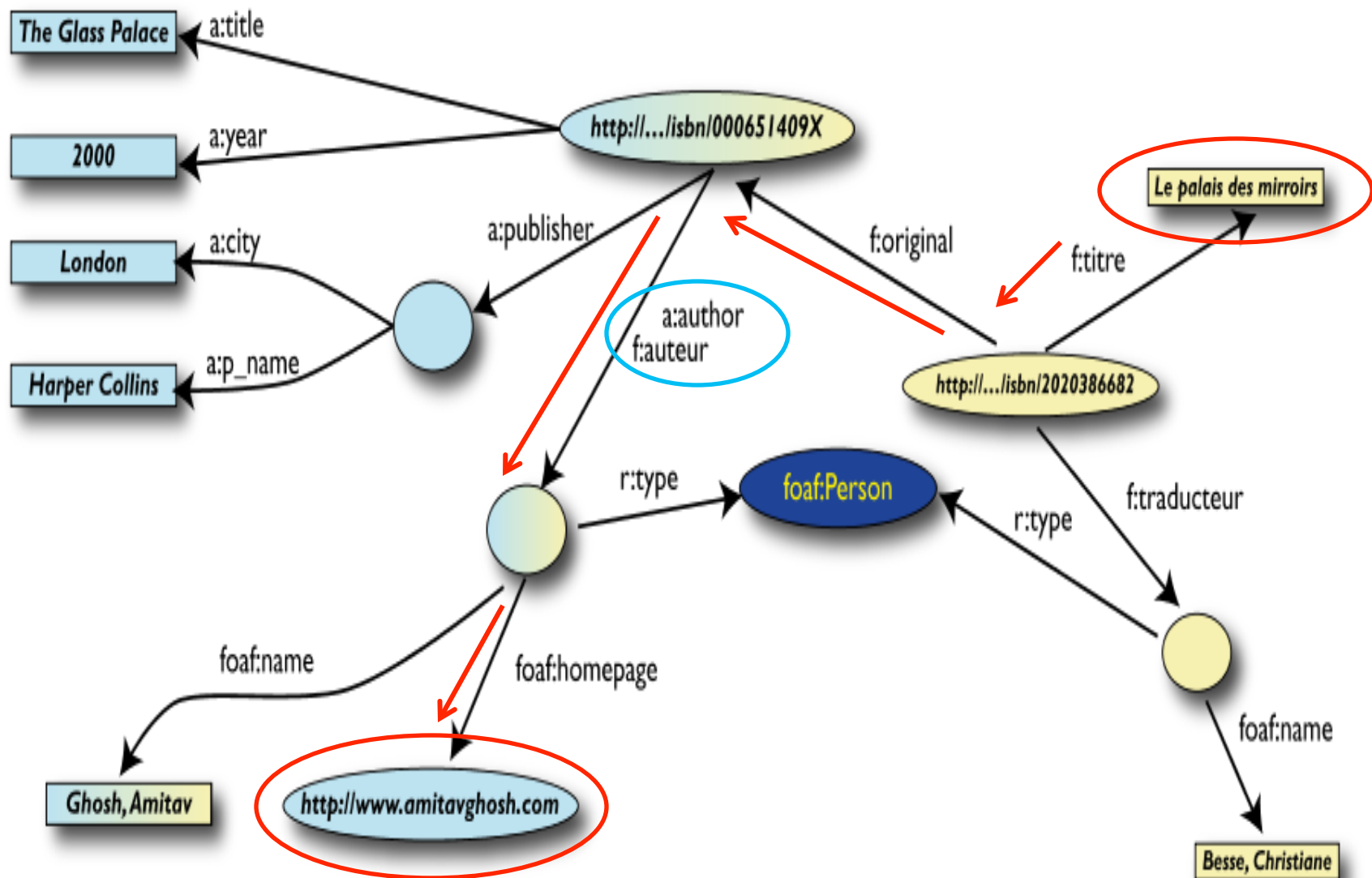


Start Making Richer Queries!

- User of data set “F” can now query:
 - “What is the home page of the ‘auteur’ of the original version of *Le Palais des miroirs* ?”
- The information is not in data set “F” *nor is it in* “A”...
- ...but was made available by:
 - Merging data sets “A” and “F”
 - Adding three “glue” statements



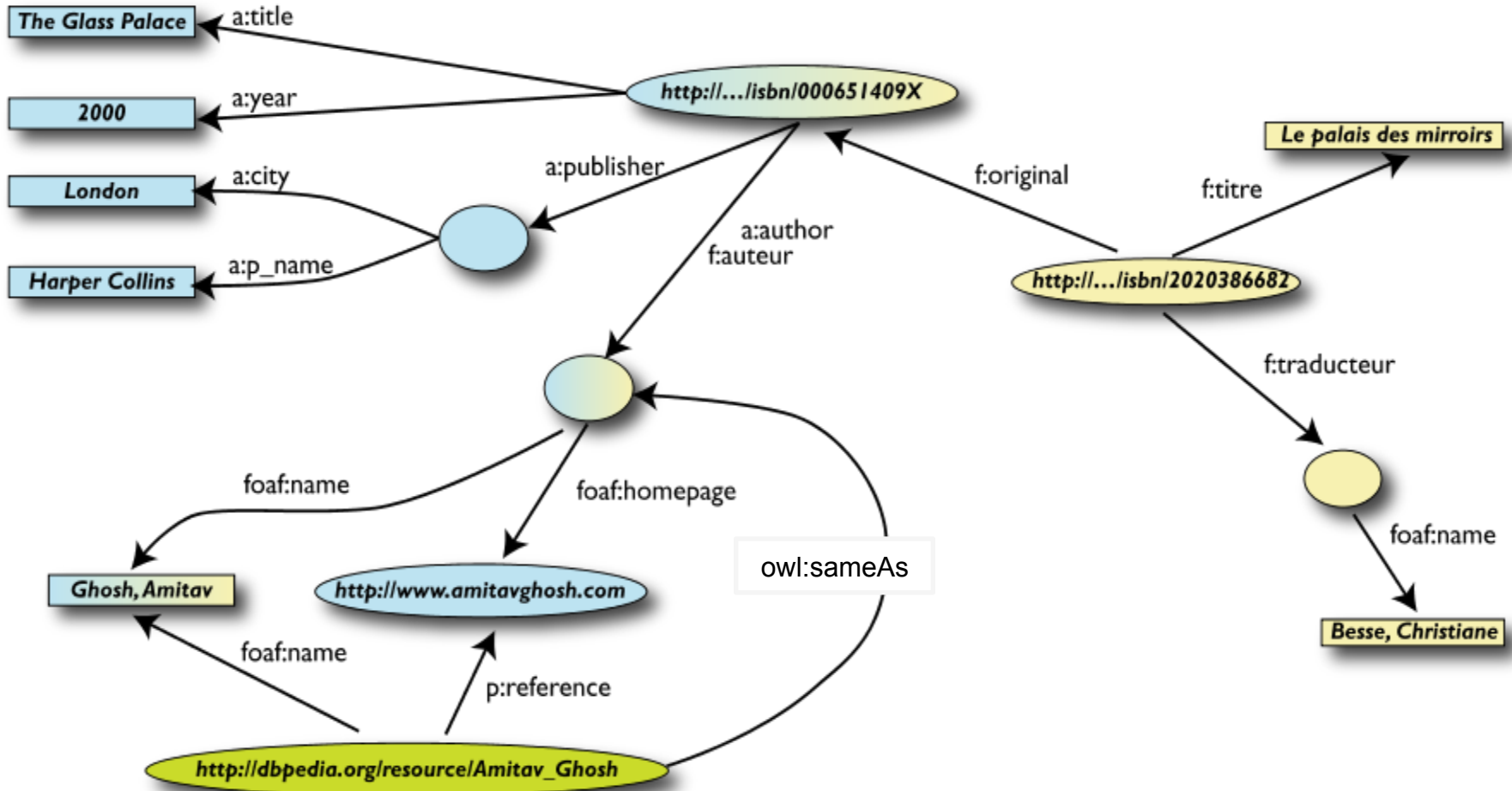
6th: Richer Queries



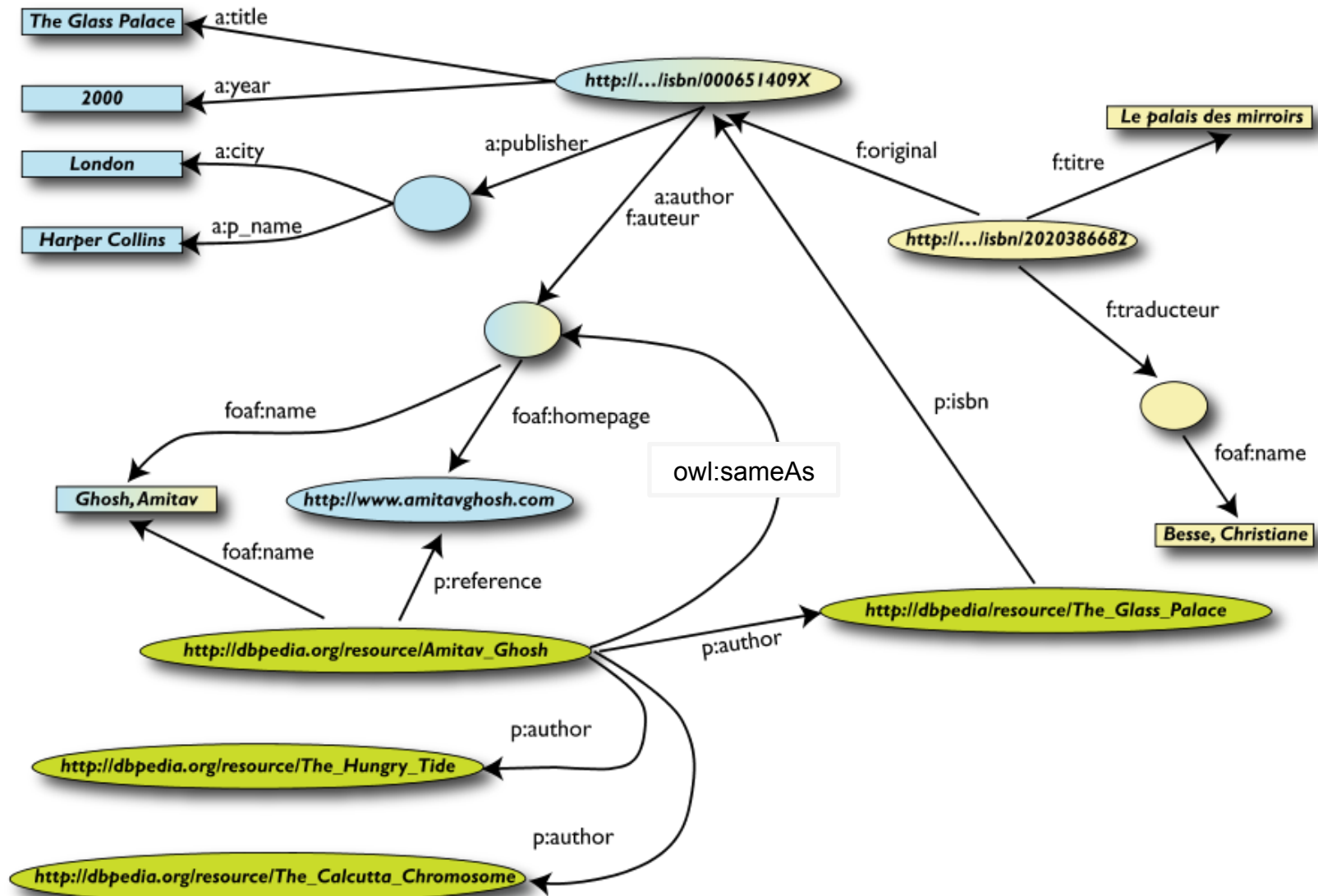
Bring in Other Data Sources

- Using the fact that Amitav Ghosh is of class Person, we can combine our data with Person data from additional sources
- Factual information from Wikipedia has been extracted with dedicated tools in the DBpedia project
(FU Berlin, U Leipzig)
 - let's combine our data with DBpedia

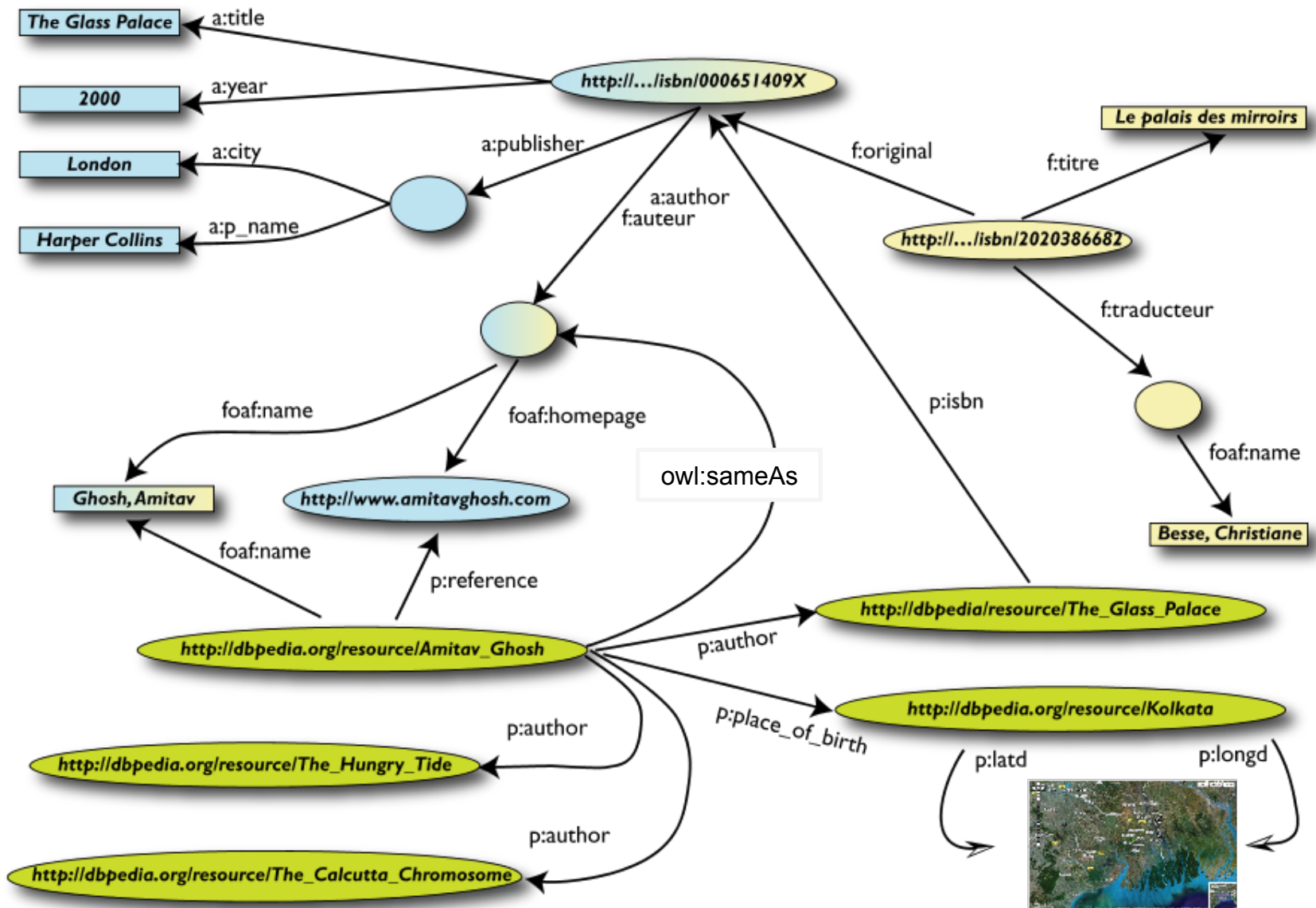
7th: Merge with DBpedia Data



7th (cont'd): Merge with DBpedia Data



7th (cont'd): Merge with DBpedia Data



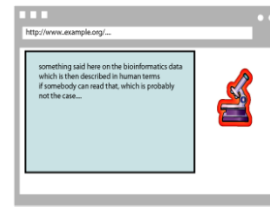
Rigour Makes the Difference

- What happened via automatic means is done every day by Web users!
- The difference: a bit of extra rigour so that machines can do this, too

What Did We Do?

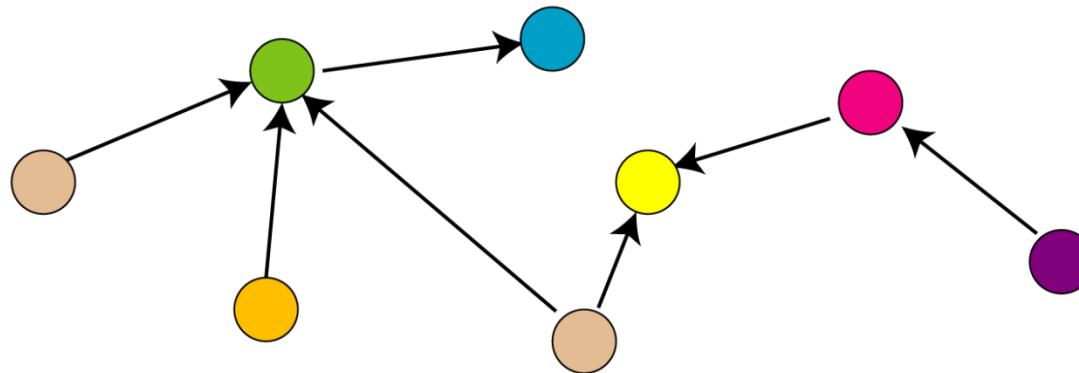
- We **combined** different data sets that
 - may be internal or somewhere on the Web
 - are of different formats
(RDBMS, Excel spreadsheet, (X)HTML, etc)
 - have different names for the same relations
 - We could *combine the data* because some **URIs were identical**
 - i.e. the ISBNs in this case
 - We could add some simple additional information (the “**glue**”) to help **further merge data** sets by applying some simple **reasoning**
- ➔ We could find new relationships between data

What Did We Do? (cont'd)



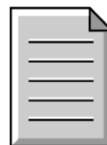
Applications

Query,
Manipulate,
etc.



Data represented in abstract format

Map,
Expose,
etc.



Data in various formats

The Abstraction Pays Off Because...

- ...the graph representation is independent of the details of the native structures
- ...a change in local database schemas, HTML structures, etc. does not affect the abstract graph
 - “schema independence”
- ...new data, new connections can be added using the same data model
- ... the approach can be applied both at the Web level and the enterprise level

So where is the Semantic Web?

Semantic Web technologies make such integration possible

Semantic Technologies in Business

- SNOMED
- Oracle (Server)
- IBM (DB2, Watson)
- Schema.org
- Good-relations
- SAP
- BBC
- Best Buy
- Startups



Semantic Technologies: Overview

- A set of technologies and frameworks that enable semantic **data management, data integration and the web of data**
 - Resource Description Framework (RDF)
 - A variety of data interchange formats (e.g., RDF/XML, N3, Turtle, N-Triples)
 - Semantic languages such as RDF Schema (RDFS), the Web Ontology Language (OWL) and the Query language (SPARQL)
 - Software infrastructure (RDF/SPARQL frameworks, Triple stores, Data integrators, Query engines, Reasoners)
 - Publicly available **connected** dataset and open data initiatives (LOD)

Reading Material

- FSWT Section 1.4
- PSW Chapter 1
- SWP Part I, Chapter 1