

# Lecture #A

---

NEWM N510: Web-Database Concepts

## Semantic Web

kharrazi@iupui.edu  
<http://www.info510.com>

# Lecture in a Nutshell

1. Definition
2. Purpose
3. Relationship to The Hypertext Web
4. Skeptical Reactions
5. Components
6. Challenges
7. Projects
8. Example

# 1. Definition

- Semantic Web is a group of methods and technologies to **allow machines to understand the meaning** (*semantics*) of information on the World Wide Web
- According to the original vision, the availability of **machine-readable metadata** would enable automated agents and other software to access the Web more intelligently. The **agents would be able to perform tasks automatically** and locate related information on behalf of the user.
- While the term "Semantic Web" is not formally defined it is mainly used to describe the **model and technologies proposed by the W3C**. These technologies include the Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.
- Many of the technologies proposed by the W3C **already exist and are used in various projects**. The Semantic Web as a global vision, however, has remained largely unrealized and its critics have questioned the feasibility of the approach.

## 2. Purpose

- **Humans** are capable of using the Web to carry out tasks such as finding the French word for "directory," reserving a library book, and searching for a low price for a DVD. However, one **computer** cannot accomplish all of these tasks without human direction, because web pages are **designed to be read by people, not machines**.
- The semantic web is a vision of **information that is understandable by computers**, so computers can perform more of the tedious work involved in finding, combining, and acting upon information on the web.
- Tim Berners-Lee originally expressed the vision of the semantic web as:  
*"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."*

## ***Purpose cont.***

- Semantic Web application areas are **experiencing intensified interest** due to the rapid growth in the use of the Web, together with the innovation and renovation of information content technologies.
- The Semantic Web is regarded as an **integrator across different content and information applications and systems**, and provide mechanisms for the realization of Enterprise Information Systems.
- Often the terms '**Semantics**', '**metadata**', '**ontologies**' and '**Semantic Web**' are used inconsistently. In particular, these terms are used as everyday terminology by researchers and practitioners, spanning a vast landscape of different fields, technologies, concepts and application areas. Furthermore, there is confusion with regards to the current status of the enabling technologies envisioned to realize the Semantic Web.
- Tim Berners-Lee has described semantic web part of 'Web 3.0':  
*"People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty — on Web 2.0 and access to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource..."*

### 3. Relationship to The Hypertext Web

#### □ Limitations of HTML

- Many files on a typical computer can be loosely divided into: **Documents** like mail messages, reports, and brochures are read by humans. **Data**, like calendars, address-books, playlists, and spreadsheets are presented using an application program which lets them be viewed, searched and combined in many ways.
- Currently, the World Wide Web is based mainly on documents written in Hypertext Markup Language (HTML), a markup convention that is used for coding a body of text interspersed with multimedia objects such as images and interactive forms.
- **Metadata tags** provide a method by which computers can categorize the content of web pages:

```
<meta name="keywords" content="Hospital Information System, HIS, Electronic Health Record, EHR, Electronic Medical Record" >
```

```
<meta name="description" content="New Commercial EHR Systems" >
```

```
<meta name="author" content="John Doe – HIMMS" >
```

## *Relationship to The Hypertext Web cont.*

- With HTML and a tool to render it (perhaps web browser or a user agent), one can create and present a page that **lists admitted patients**.

The HTML of this list page can make simple, document-level assertions such as "this document's title is 'Patient List'", but there is **no capability within the HTML itself to assert unambiguously that**, for example, patient HK783545e is a Male with a chief complaint of Abdominal Pain, or that he is 45 yrs old.

Rather, HTML can only say that the span of text "HK783545e" is something that should be positioned near "Male" and "Abdominal Pain".

There is no way to say "This is a demographic list" or even to establish that "Male" is a kind of gender or that "Abdominal Pain" is a chief complaint. There is also **no way to express that these pieces of information are bound together** in describing a patient, distinct from other patients perhaps listed on the page.

- **Microformats** represent unofficial attempts to extend HTML syntax to create machine-readable semantic markup about objects such formatted name, telephone numbers and etc.

## *Relationship to The Hypertext Web cont.*

### □ Semantic Web Solutions

- The Semantic Web takes the solution further. It involves publishing in languages specifically designed for data: Resource Description Framework (**RDF**), Web Ontology Language (**OWL**), and Extensible Markup Language (**XML**). HTML describes documents and the links between them. **RDF, OWL, and XML, by contrast, can describe arbitrary things** such as people, meetings, or airplane parts → resulting network of Linked Data the **Giant Global Graph**, in contrast to the HTML-based World Wide Web.
- These technologies are combined in order to provide descriptions that **supplement or replace the content of Web documents**. Thus, content may manifest itself as descriptive data stored in Web-accessible databases, or as markup within documents (particularly XHTML interspersed with XML, or purely in XML with layout or rendering XSL stored separately).
- The machine-readable descriptions enable content managers to add meaning to the content → a machine can process knowledge itself:

```
<item>appendicitis</item>
```

```
<item rdf:about="http://en.wikipedia.org/wiki/Appendicitis">appendicitis</item>
```



## 4. Skeptical Reactions

### ❑ Practical Feasibility

- Critics (e.g. [Which Semantic Web?](#)) question the basic feasibility of a complete or **even partial fulfillment of the semantic web**.

**Metacrap** → human behavior and personal preferences. For example, people lie: they may include spurious metadata into Web pages in an attempt to **mislead Semantic Web engines** that naively assume the metadata's veracity. This phenomenon was well-known with metatags that fooled the search engine ranking algorithms into elevating the ranking of certain Web pages.

### ❑ Censorship and Privacy

- **Text-analyzing techniques** can now be easily bypassed by using other words, metaphors for instance, or by using images in place of words. An advanced implementation of the semantic web would make it much easier for **governments to control** the viewing and creation of online information, as this information would be much easier for an automated content-blocking machine to understand → **geo location meta-data**: very little anonymity associated with the authorship of articles.

## *Skeptical Reactions cont.*

### ❑ Doubling Output Formats

- It would be much more **time-consuming to create and publish content** because there would need to be two formats for one piece of data: one for human viewing and one for machines. However, many web applications in development are addressing this issue by creating a **machine-readable format upon the publishing of data** or the request of a machine for such data. The development of microformats has been one reaction to this kind of criticism.

### ❑ Need

- The idea of a semantic web, able to describe, and associate meaning with data, necessarily involves **more than simple XHTML mark-up code**. It is based on an assumption that, in order for it to be possible to endow machines with an ability to accurately interpret web content, far more than the mere ordered relationships involving letters and words is **necessary as underlying infrastructure** → Otherwise, most of the supportive functionality would have been available in **Web 2.0**, and it would have been possible to derive a semantically capable Web with minor additions.

## 5. Components

- The semantic web comprises the standards and tools of XML, XML Schema, RDF, RDF Schema and OWL that are organized in the **Semantic Web Stack**. The [OWL Web Ontology Language Overview](#) describes the function and relationship of each of these components of the semantic web:

**XML** provides an elemental syntax for content structure within documents, yet associates no semantics with the meaning of the content contained within.

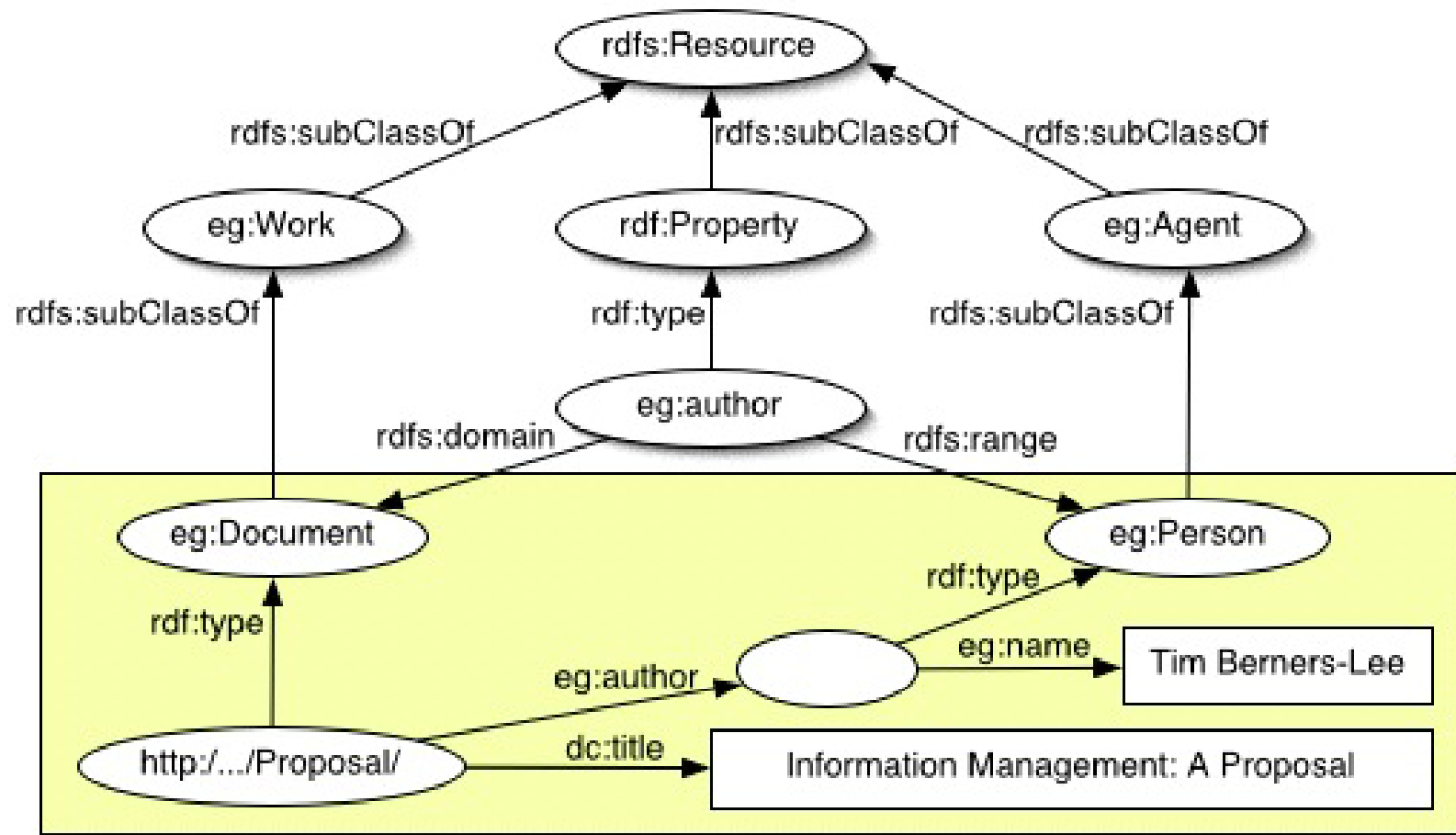
**XML Schema** is a language for providing and restricting the structure and content of elements contained within XML documents.

**RDF** is a simple language for expressing data models, which refer to objects ("resources") and their relationships. An RDF-based model can be represented in XML syntax.

**RDF Schema** extends RDF and is a vocabulary for describing properties and classes of RDF-based resources, with semantics for generalized-hierarchies of such properties and classes.

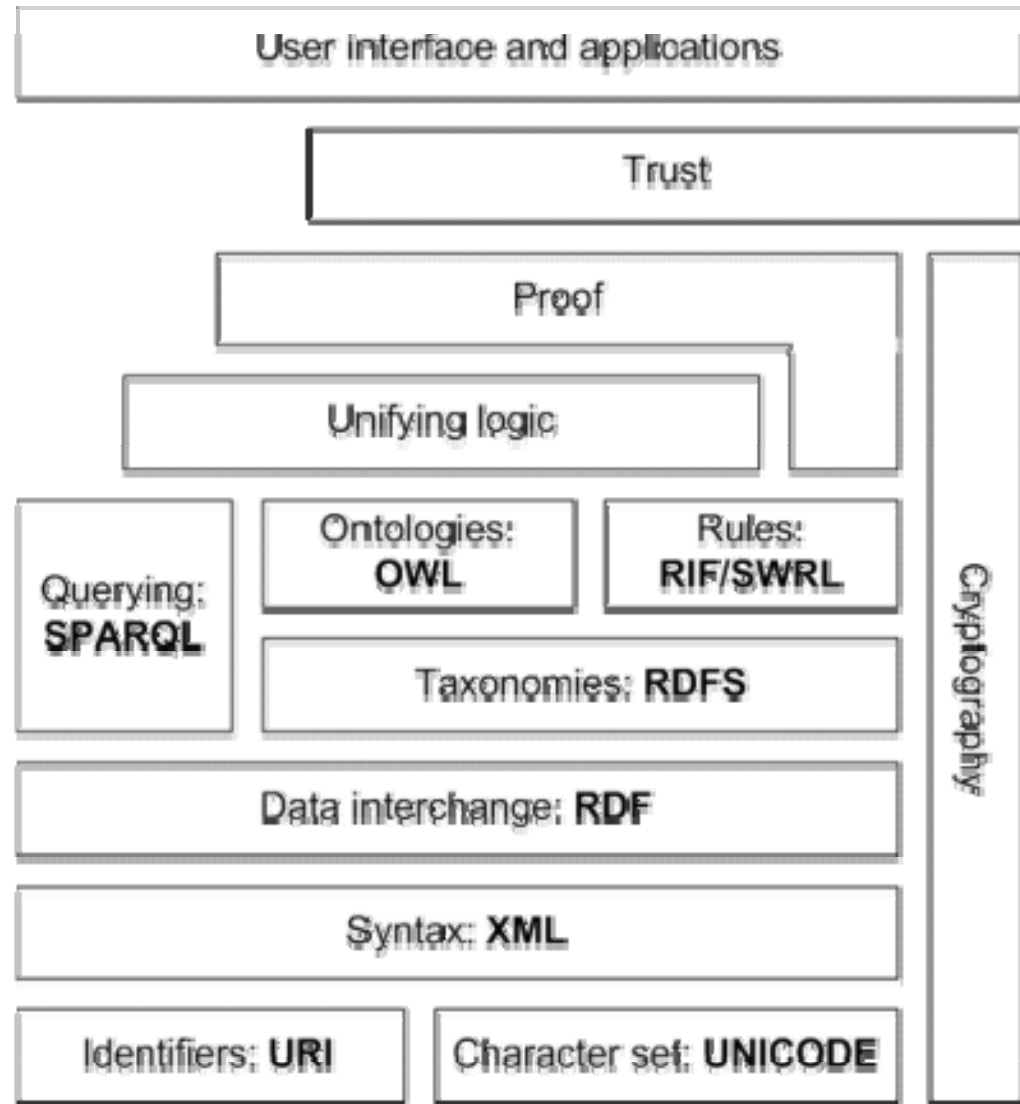
**OWL** adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes.

**SPARQL** is a protocol and query language for semantic web data sources.

*Components cont.*

RDF example (documents and their authors)

## Components cont.



Semantic Web Stack

## 6. Challenges

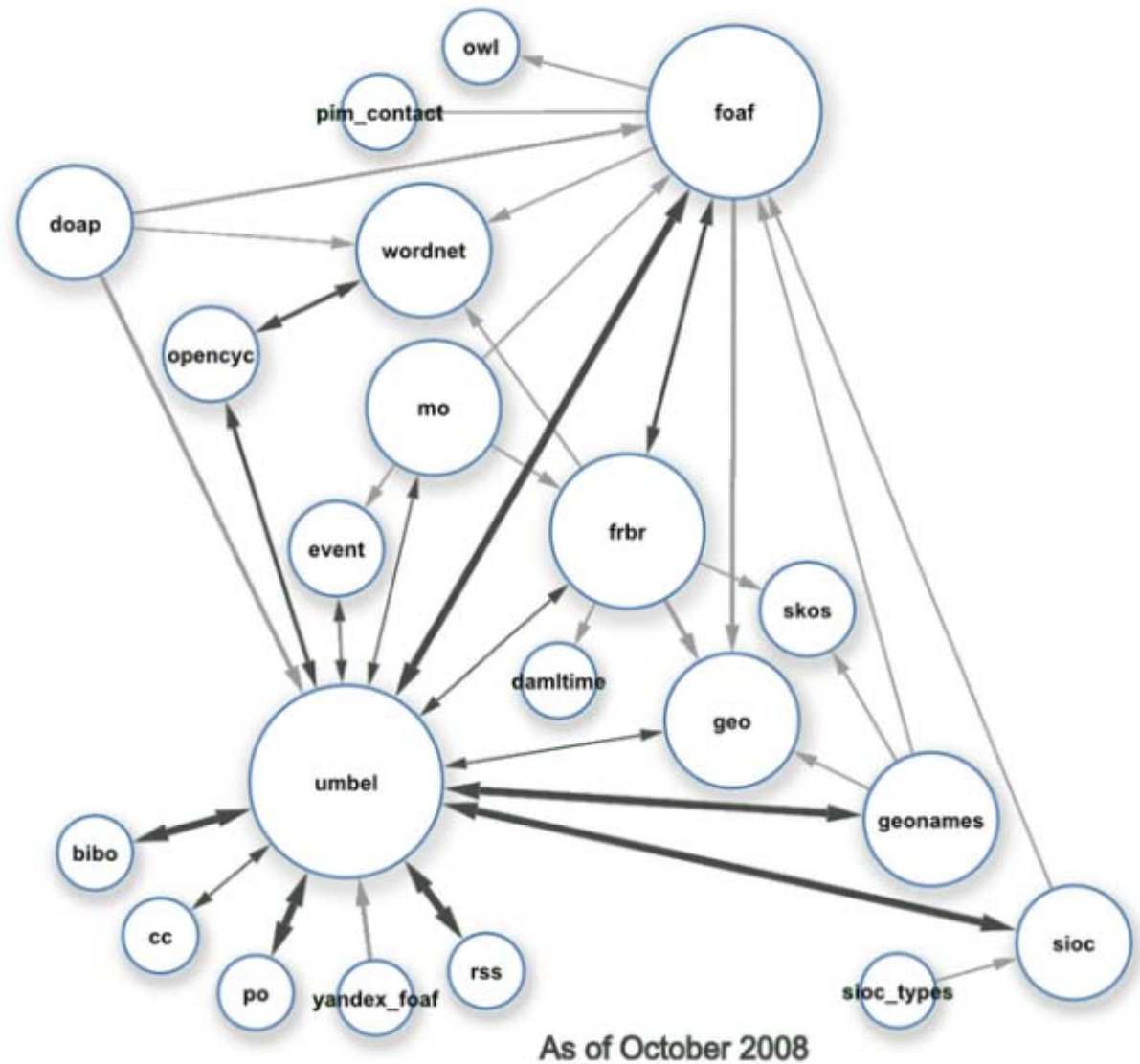
- Some of the challenges for the Semantic Web include vastness, vagueness, uncertainty, inconsistency and deceit. **Automated reasoning systems** will have to deal with all of these issues:
- **Vastness**: The World Wide Web contains at least 24 billion pages as of 2010. The SNOMED CT medical terminology ontology contains 370,000 class names, and existing technology has not yet been able to eliminate all semantically duplicated terms.
- **Vagueness**: These are imprecise concepts like "young" or "tall". This arises from the vagueness of user queries, of concepts represented by content providers, of matching query terms to provider terms and of trying to combine different knowledge bases with overlapping but subtly different concepts → Fuzzy logic
- **Uncertainty**: These are precise concepts with uncertain values. For example, a patient might present a set of symptoms which correspond to a number of different diagnoses each with a different probability → Probabilistic reasoning techniques
- **Inconsistency**: These are logical contradictions which will inevitably arise during the development of large ontologies. Deductive reasoning fails catastrophically when faced with inconsistency, because "anything follows from a contradiction" → Defeasible reasoning and paraconsistent reasoning
- **Deceit**: This is when the producer of the information is intentionally misleading the consumer of the information → Cryptography techniques

## 7. Projects

- **Dbpedia**: is an effort to publish structured data extracted from Wikipedia: the data is published in RDF and made available on the Web for use under the GNU Free Documentation License, thus allowing Semantic Web agents to provide inferencing and advanced querying over the Wikipedia-derived dataset and facilitating interlinking and extension in other data-sources.
- **FoaF**: A popular application of the semantic web is Friend of a Friend which uses RDF to describe the relationships people have to other people and the "things" around them. FOAF permits intelligent agents to make sense of the thousands of connections people have with each other, their jobs and the items important to their lives; connections that may or may not be enumerated in searches using traditional web search engines. Because the connections are so vast in number, human interpretation of the information may not be the best way of analyzing them. FOAF is an example of how the Semantic Web attempts to use the relationships within a social context.
- **GoodRelations**: is a popular vocabulary for expressing product information, prices, payment options, etc. GoodRelations has been adopted by BestBuy, Yahoo, OpenLink Software, O'Reilly Media, the Book Mashup, and many others.

## Projects cont.

- Linking Open Data:** is a W3C-led effort to create openly accessible, and interlinked, RDF Data on the Web. The data in question takes the form of RDF Data Sets drawn from a broad collection of data sources.





## 8. Example

- **Example (1):** Organizing a visit to a Quebec Hospital: Flight / Hotel / Pictures / Center
- Flight → High Cost Airline

The screenshot displays the United Airlines website interface for a flight search. The search parameters are: Depart Sun, Nov 07, 2010 from Indianapolis, IN (IND) to Quebec, QC (YQB); Return Sat, Nov 13, 2010 from Quebec, QC (YQB) to Indianapolis, IN (IND). The lowest fare available is \$859.30 USD per adult. The search results show several flight options, including a highlighted option with a total price of \$931.25 USD per adult. The highlighted option consists of a departure flight (United 6031, 1:18 PM IND - 1:34 PM ORD) connecting to a return flight (United 6903, 7:00 PM ORD - 10:20 PM YQB). Other flight options are also visible, such as United 7486 and United 6915.

**UNITED** My profile | Worldwide sites | Customer service

Planning & booking | Reservations & check-in | Mileage Plus® | Services & information | Search site

Looking for Continental Airlines service? Visit [con.continental.com](http://con.continental.com) | Shop for flights | Special deals | Travel Options by United

**Results based on schedule and price**  
Click the flight(s) you would like to select. \*Additional [baggage charges](#) may apply

Search → Review → Purchase

**Lowest fare available**  
**859.30**  
USD per adult

Additional taxes & fees apply. **LOW FARE** DETAILS

**Depart Sun, Nov 07, 2010**  
Indianapolis, IN (IND) to Quebec, QC (YQB)

**Return Sat, Nov 13, 2010**  
Quebec, QC (YQB) to Indianapolis, IN (IND)

**1:18 PM IND - 1:34 PM ORD**  
United 6031 [View seats](#)  
Operated by UNITED EXPRESS/EXPRESSJET AIRLINES [Flight details](#)

**Only 3 tickets left at this fare**

**931.25**  
USD per adult

Additional taxes & fees apply  
[Price breakdown](#)  
[Fare rules](#)  
**Continue**

**8:09 AM YQB - 10:00 AM ORD**  
United 6915 [View seats](#)  
Operated by UNITED EXPRESS/SKYWEST AIRLINES [Flight details](#)

connecting to

**7:00 PM ORD - 10:20 PM YQB**  
United 6903 [View seats](#)  
Operated by UNITED EXPRESS/SKYWEST AIRLINES [Flight details](#)

**3:05 PM IND - 3:20 PM ORD**  
United 7486 [View seats](#)  
Operated by UNITED EXPRESS/SHUTTLE AMERICA  
\* [Exclus Details](#)

**8:09 AM YQB - 10:00 AM ORD**  
United 6915 [View seats](#)  
Operated by UNITED EXPRESS/SKYWEST AIRLINES [Flight details](#)

**Modify search**

**Departing times**

- Anytime
- Before noon
- Noon - 6PM
- After 6PM

**Returning times**

- Anytime
- Before noon

## Example cont.

- Flight → Moderate Cost Airline

The screenshot shows the Continental Airlines website interface. At the top, there's a navigation bar with links for Home, Reservations, Travel Information, Deals & Offers, OnePass Frequent Flyer, Products & Services, and About Continental. A search progress bar indicates the current step is 'CHOOSE FLIGHTS'. The main heading is 'Select Departing Flight'. Below this, there's a promotional banner for a round trip from Indianapolis, IN (IND) to Québec City, QC Canada (YQB) on Sunday, Nov. 7, 2010, for 1 traveler in Economy class. The fare is \$710 with stops, or \$2,192 flexible. A secondary offer mentions 'Pay no annual fee for the first year. Get \$50 back and up to 30,000 bonus miles.' Below the promotional text is a table titled 'Select Your Departing Flight for Sun., Nov. 7, 2010:' with columns for Price, Departing, Arriving, Travel Time, and OnePass Miles. Two flight options are listed: one with stops from \$800 and one direct flight for \$1,096.

**OUR LOWEST FARES**

With Stops from	Flexible Fare from
<b>\$710</b>	<b>\$2,192</b>

Select Your Departing Flight for Sun., Nov. 7, 2010:

Price	Departing	Arriving	Travel Time	OnePass Miles
<b>RIGHTS WITH STOPS FROM \$710</b>				
From <b>\$800</b>	Depart: 8:20 p.m. Sun., Nov. 7, 2010 Indianapolis, IN (IND)	Arrive: 8:20 p.m. Sun., Nov. 7, 2010 New York/Newark, NJ (EWR - Liberty)	Flight Time: 2 hr 0 min	OnePass Miles/Elite Qualification*: 655 / 150%
Change Planes: Connective in New York/Newark, NJ (EWR - Liberty) to 00 minutes.				
	Depart: 9:00 p.m. Sun., Nov. 7, 2010 New York/Newark, NJ (EWR - Liberty)	Arrive: 10:42 p.m. Sun., Nov. 7, 2010 Québec City, QC Canada (YQB)	Flight Time: 1 hr 42 min Travel Time: 4 hr 22 min	OnePass Miles/Elite Qualification*: 441 / 100% Total Miles: 1,096

## Example cont.

- Flight → Country Specific (Local) Airline

**AIR CANADA** **GO FAR** Canadian Edition | [Contact Us](#)

Home Search **Select** Review Passengers Purchase Seats Itinerary

### Select Flights

Display only fares eligible for upgrade - [Learn more](#)

All fares displayed on this page are in **Canadian dollars**, per person for each way of travel, and do not include taxes, fees or some other charges. [Learn more...](#)

**Relax in our Maple Leaf Lounge**

[Learn more...](#)

**Video on demand\***

\*On most flights

[Learn more...](#)

**Select departing flight**

A charge of \$30 CAD may apply for a [second checked bag](#) in Economy Class (Air Canada and Jazz operated flights).

Please note that **United Airlines** and **Continental Airlines** charge a fee for all **checked baggage** for codeshare flights they operate between Canada and the U.S., or Canada and Mexico. [Learn more.](#)

Day's lowest fare→	Tue 02-Nov \$519	Wed 03-Nov \$516	Thu 04-Nov \$516	Fri 05-Nov \$519	Sat 06-Nov \$519	Sun 07-Nov \$434	Mon 08-Nov \$434	Tue 09-Nov \$434	Wed 10-Nov \$434	Thu 11-Nov \$434	Fri 12-Nov \$434
--------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------	------------------------

From: **Indianapolis, Indianapolis Int'l, IN (IND)**  
 To: **Quebec, QC (YQB)** [Compare our fare options](#)

Op.	Flights	Depart	Arrive	Aircraft	Connections	Tango	Tango Plus	Latitude	Executive Class Lowest	Executive Class Flexible
<b>Connecting Flights</b>										
1	AC4721	15:05	15:20	E70	Chicago (ORD)	-	\$910	\$1238	-	-
2	AC4546	19:00	22:20	CRJ						

## Example cont.

- Hotel → High Cost Hotel Chain

The screenshot shows the Hilton Hotels & Resorts website interface. At the top left is the Hilton logo. To its right are fields for 'Sign in' (Username or HHonors #, Password or PIN) and a 'Search by keyword' field. Below these are navigation tabs: 'Specials & Packages', 'Reservations', 'Meetings', 'Social Gatherings', 'Hilton HHonors', 'Travel Guides', 'Destinations', and 'My Favorite Hotels'. The main content area is titled 'Search Results' and includes a 'Change Your Search' sidebar on the left. The sidebar contains filters for Location (City: quebec, Search Within: 25 mi/km, State/Province: Quebec, Country: Canada), Brand (Hilton Hotels, All Hotels), and Dates (Check-in: 7 Nov 2010, Check-out: 12 Nov 2010). The search results show one hotel: 'Hilton Quebec' at 1100 Rene Levesque East, Quebec, Canada, G1R 4P3, 1-418-647-2411, 0.2 miles from city center. The hotel is marked as 'Available' and has a rate of 'From 119.20 CAD / night'. There are buttons for 'Compare Selected Hotels', 'View Rates', and 'Add to Favorite Hotels'. A 'Clear Selected Hotels' link is at the bottom left.

## Example cont.

- Hotel → Moderate Cost Hotel Chain

**Marriott** Find & reserve EXPLORE & PLAN MEETINGS & EVENTS MARRIOTT REWARDS MY ACCOUNT Sign in

Courtyard Quebec City Home

**Courtyard Quebec City**

Photo Tour  
 Guest Rooms in Detail  
 Deals at This Hotel  
 About This Hotel  
 Restaurants & Lounges  
 Visitor's Guide: Quebec City  
 Maps & Transportation  
 Plan Events & Meetings  
 Use Marriott Rewards Points  
 Printable Hotel Fact Sheet  
 Find other Courtyard Hotels

850 Place D'Youville  
 Quebec City, Quebec G1R 2P6 Canada  
 Phone: 1 418 694 4004  
 Fax: 1 418 694 4007  
 Sales: 1 418 694 4004  
 Sales fax: 1 418 694 4478  
 Toll-free: 1 866 694 4004

photo tour

The one and only, Courtyard by Marriott Quebec City offers a boutique style ambiance just within steps of Old Quebec, one of the few remaining walled city in North America. The hotel is walking distance to most of the summer events like the inescapable Cirque du Soleil or

[Read More](#)

**Hotel Highlights**

- Awarded #1 in Staff Service by Courtyard America with over 780+ hotels for a 6th

**Last-Minute Weekends**

**Check Rates & Availability**

Check-in date (mm/dd/yy)  
 11/7/10  
 Check-out date (mm/dd/yy)  
 11/13/10  
 No. of rooms: 1 Guests / room: 1  
 Check nearby locations

**Optional Information**

## Example cont.

- Hotel → Low Cost Hostel

The screenshot shows the Hostel Traveler.com website interface. At the top, there is a navigation bar with the logo, a globe, and buttons for "Book now", "travelers", "work with us", "reviews", and "my bookings". A "User ID" field and a "Login" button are also present. Below the navigation bar, there is a search bar with "English" selected. The main content area features a "Get Rates and Availability" section with a "Book Online! Compare lodgings, check rates and availability, and securely book reservations." message. A "Booking Request" form is displayed, showing a search for "1 property" in "Quebec City". The selected property is "Auberge Aux Deux Lions" with a check-in date of Nov 07, 2010, and a check-out date of Nov 14, 2010. The form also includes fields for "Rooms" (1), "Adults" (2), "Children" (0), and "Maximum desired rate (optional)" (US Dollar). A sidebar on the left contains navigation links like "Homepage", "My Reservations", "Popular Hostels", "Best Hostels", "Cheapest Hostels", and "Link to Us". A "Find the best places to stay" section on the right provides recommendations for Quebec City and Quebec.

## Example cont.

- Hotel → Country Specific (Local) Hotel

All in Canadian-French!

The screenshot displays the website for Le Vincent Auberge urbaine. The header features the hotel name, a 4-star rating, and the slogan "À deux pas de tout...". A navigation menu on the left includes links for "BIENVENUE L'AUBERGE", "LES HÔTES", "LES CHAMBRES", "LES FORFAITS", "LOCALISATION", and "CONTACT". The main content area is titled "Les forfaits" and lists several packages with their respective prices and descriptions. A "Réservation en ligne" button is located in the top right corner.

**Le Vincent**  
Auberge urbaine ★★★★★

+ Réservation en ligne

À deux pas de tout ...

**Les forfaits**

L'Auberge Le Vincent vous offres ses forfaits.

Forfait Italien	Forfait Détente	Forfait Amoureux
		
Souper au restaurant le Cicco Café. Repas pour 2 personnes induant 5 services et 1 bouteille de vin sélectionnée.	Forfait Détente avec Massothérapie La Quintessens Spa Urbain pour 1 personne. 1 heure de massage.	Vous attendent à la chambre pour ce forfait romantique : 1 bouteille de vin mousseux, 2 coupes et des fleurs.
Tarif : 75\$ + taxes	Tarif : 55\$ + taxes	Tarif : 80 \$ + taxes
<b>Forfait Détente Max</b>	<b>Forfait 3 Musées</b>	
		

QUARTIER SAINT-ROCH, 295, RUE SAINT-VALLIER EST, QUÉBEC (QUÉBEC) Q1K 3P5 FAX: (418) 523-5999 RÉSERVATIONS (418) 523-3000

## Example cont.

- Flight + Hotel → Specialized Travel Website

The screenshot displays the Expedia website interface for a travel search. At the top, there is a navigation bar with the Expedia logo and a search bar. Below the navigation bar, a yellow banner contains the text "Create your trip from Indianapolis to Quebec (and vicinity)".

On the left side, there is a sidebar with search filters:
 

- Change your search:** Departing (11/7/2010), Returning (11/13/2010), Star Rating (Show all), Lodging Type (Show All), Name contains (Go).
- Change Travelers:** 2 Adults, 1 Room (Change travelers).

The main content area shows:
 

- Maps:** Area map view, Hotel map view.
- Hotel amenities:** Narrow your search.
- Show hotels in this area:** Quebec (and vicinity) (all areas) (Go).
- View trips:** 1 - 25, 26 - 50, 51 - 73.
- Sort by:** Expedia Picks (selected), Price, Hotel Name, City, Star Rating, Traveler Opinion.
- Sponsored Listing:** Auberge du Littoral, Quebec, QC. 5-star rating. Description: "Skin from Old Quebec: Rooms, Suites & Apartment-Hotel located between Montmorency Falls and Old Quebec. Free internet access and free parking lot." (SEE DETAILS ▶)
- Flight + Hotel Trip:** Hotel Clarendon. Description: "Built in 1870, this historic Art Nouveau and Art Deco hotel is located in Old Québec, 1 block from Château Frontenac." (More lodging info). Promotion: "Hotel promotion - 48 Hour Sale: save up to 25%". Book online or call 1-800-222-0892.
- Price breakdown:**

Booked separately:	\$2588
Trip Savings:	- \$1137
<b>Total Price:</b>	<b>\$1451</b>
<b>\$726 per person</b>	
Includes: Flight + Hotel, Taxes & Fees**	

On the right side, there are two promotional banners:
 

- Get the best price with Expedia—guaranteed!** (Learn more)
- NEW! Save up to 50% with Unpublished Rate Hotels** (FIND OUT HOW)



## Example cont.

- Flight + Hotel → Specialized Travel Website

The screenshot shows the Orbitz website interface for a search from Indianapolis to Quebec City. The search parameters are: 11/07/2010 - 11/13/2010 for 2 travelers. The package includes 2 round trip tickets and 6 nights in 1 room. The results are displayed in an ORBITZ MATRIX™ DISPLAY table.

Sort packages by:	Air Canada	United Airlines	Delta Air Lines	Continental Airlines	Multiple Carriers
Best value	2+ stops	2+ stops	1 stop	1 stop	2+ stops
Lowest price		4 Seats left at this price!			3 Seats left at this price!
Distance					
Star rating					
Howard Johnson Quebec City   ★★★★★ Quebec City 7.4 Miles West of Quebec City User score 3.9 out of 5	\$769 USD per person 2 tickets 6 nights Save \$902	\$835 USD per person 2 tickets 6 nights Save \$675	\$1,072 USD per person 2 tickets 6 nights	\$1,074 USD per person 2 tickets 6 nights	\$1,100 USD per person 2 tickets 6 nights
ORBITZ SPECIAL OFFER Stay 3, Save 30%					
Hotel Clarendon   Quebec ★★★★★	\$855	\$924	\$1,457	\$1,460	\$1,496

## Example cont.

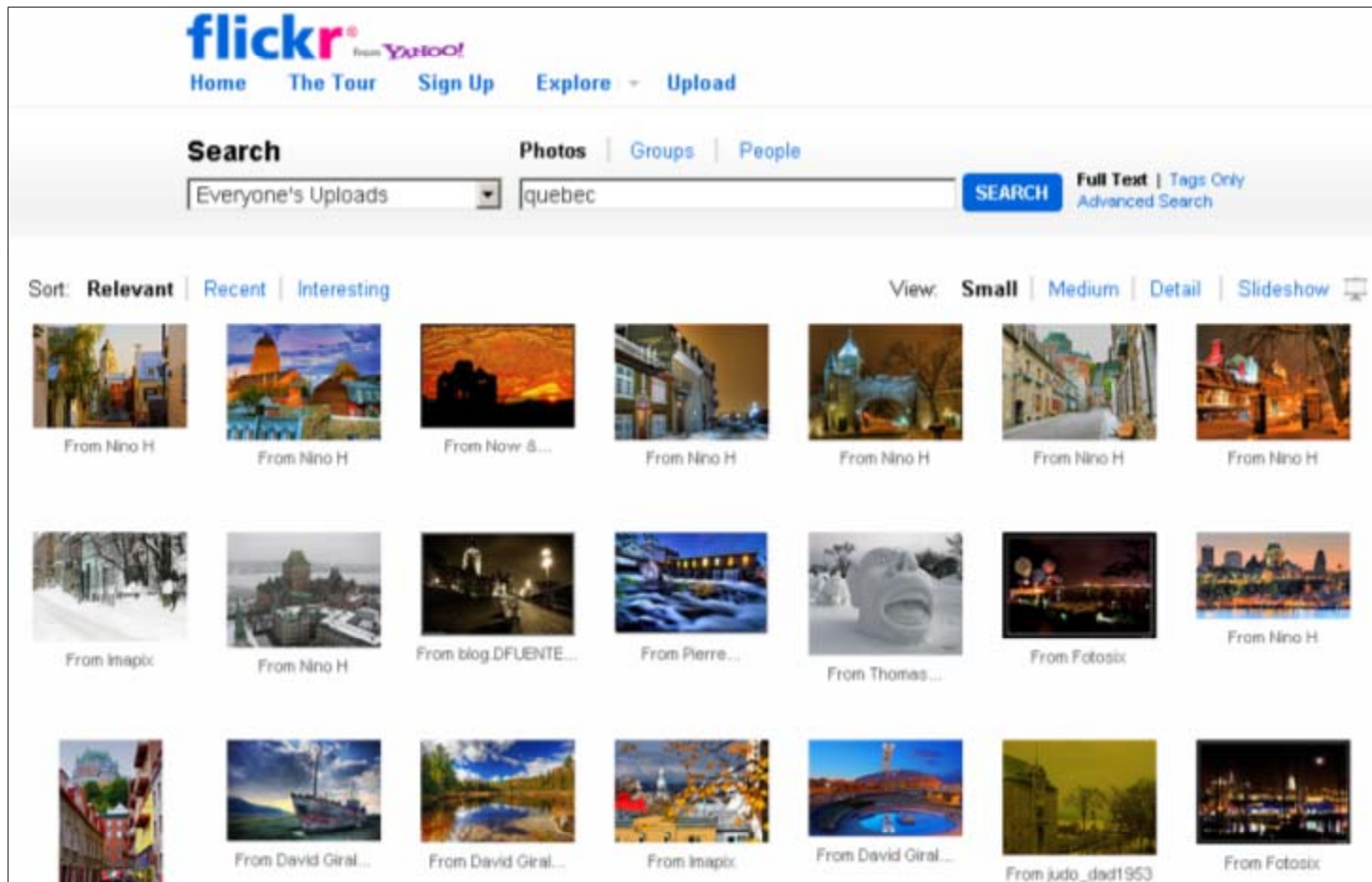
- Photo → Photo Search Engine (Storage) Website

The screenshot shows a Google search for "quebec" with the following elements:

- Search Bar:** Contains the text "quebec" and a "Search" button. To the right, it says "SafeSearch moderate" with a dropdown arrow.
- Results Summary:** "About 19,200,000 results (0.39 seconds)" and a link for "Advanced search".
- Related searches:** [quebec map](#), [quebec flag](#), [quebec city](#), [quebec montreal](#), [quebec winter](#)
- Image Grid:** A collection of images including:
  - A map of Quebec province.
  - A sunset view of a city on a hill.
  - A waterfall in a lush green setting.
  - Four Quebec flags (fleur-de-lis) arranged in a 2x2 grid.
  - A large, ornate building, likely a cathedral or government building.
  - Aerial views of the city of Quebec.
  - Views of the city at night with lights reflecting on the water.
  - Views of the city during the day from different angles.
  - Another map of Quebec.
- Left Sidebar (Filters):**
  - Everything** (selected)
  - Images** (selected)
  - More** (dropdown arrow)
  - Any size:** Large, Medium, Icon, Larger than..., Exactly...
  - Any type:** Face, Photo, Clip art, Line drawing
  - Any color:** Full color, Black and white, and a color palette.
  - Standard view** (selected)
  - Show sizes** (link)

## Example cont.

- Photo → Photo Sharing Website



## ***Example cont.***

- Because these website are mainly human readable and not computer readable → you had to search and try a large number of websites with different languages, styles and etc → then **you had to combine and integrate all those websites** to find your desired information → this is a very long and exhausting process
- All of these websites have a large amount of information hidden in their backend → DBs (e.g., MySQL) and XML data islets → however **you can only see the frontend** (e.g., united.com) or a **predefined combination of them** (e.g., expedia.com)
- Goal → We like to be able to link to data independently of their presentation and use it **the way we want** not the way that the designer wanted to present → agents (e.g., web programs) should be able to interpret data from various website the way we want → **Web of Data**
- **Mashup Websites** → collect data from other websites (that are not necessarily web semantic compatible but provide their information via web services) and offer the information in one spot. Indeed, Mashup websites give us a sneak peak of what the true power of Web of Data can be.

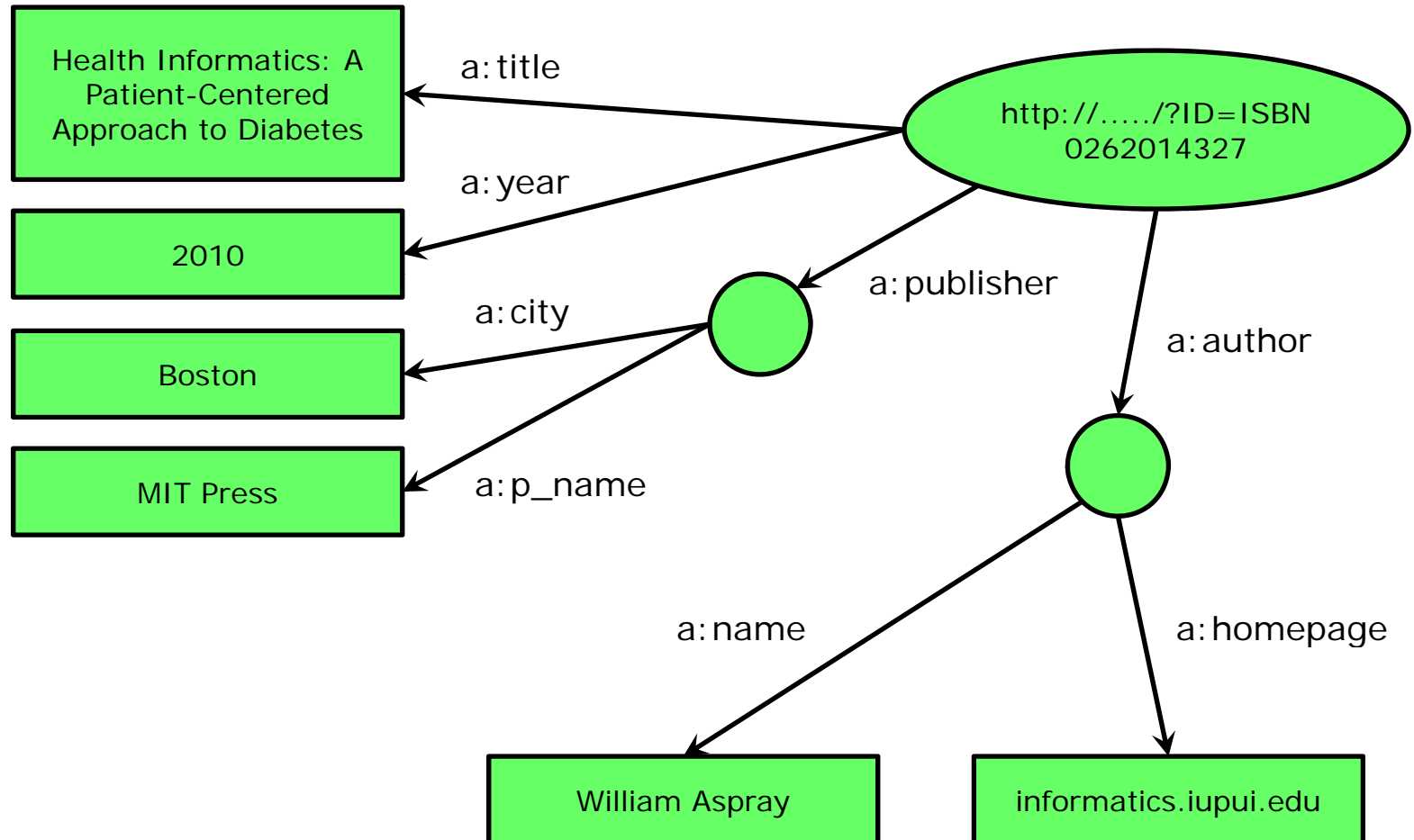
### Example cont.

- **Example (2):** Bookstore Database En, Fr, and Wikipedia Integration → (a) **Map** the various data onto an abstract data representation (b) **Merge** the resulting representations (c) Start making **queries** on the whole.
- Sample En Bookstore Dataset (e.g., MySQL)

ID	Author	Title	Publisher	Year
ISBN0262014327	id_xyz	Health Informatics: A Patient-Centered Approach to Diabetes	id_qpr	2010

ID	Name	Home Page
id_xyz	Aspray, William	<a href="http://informatics.iupui.edu">http://informatics.iupui.edu</a>

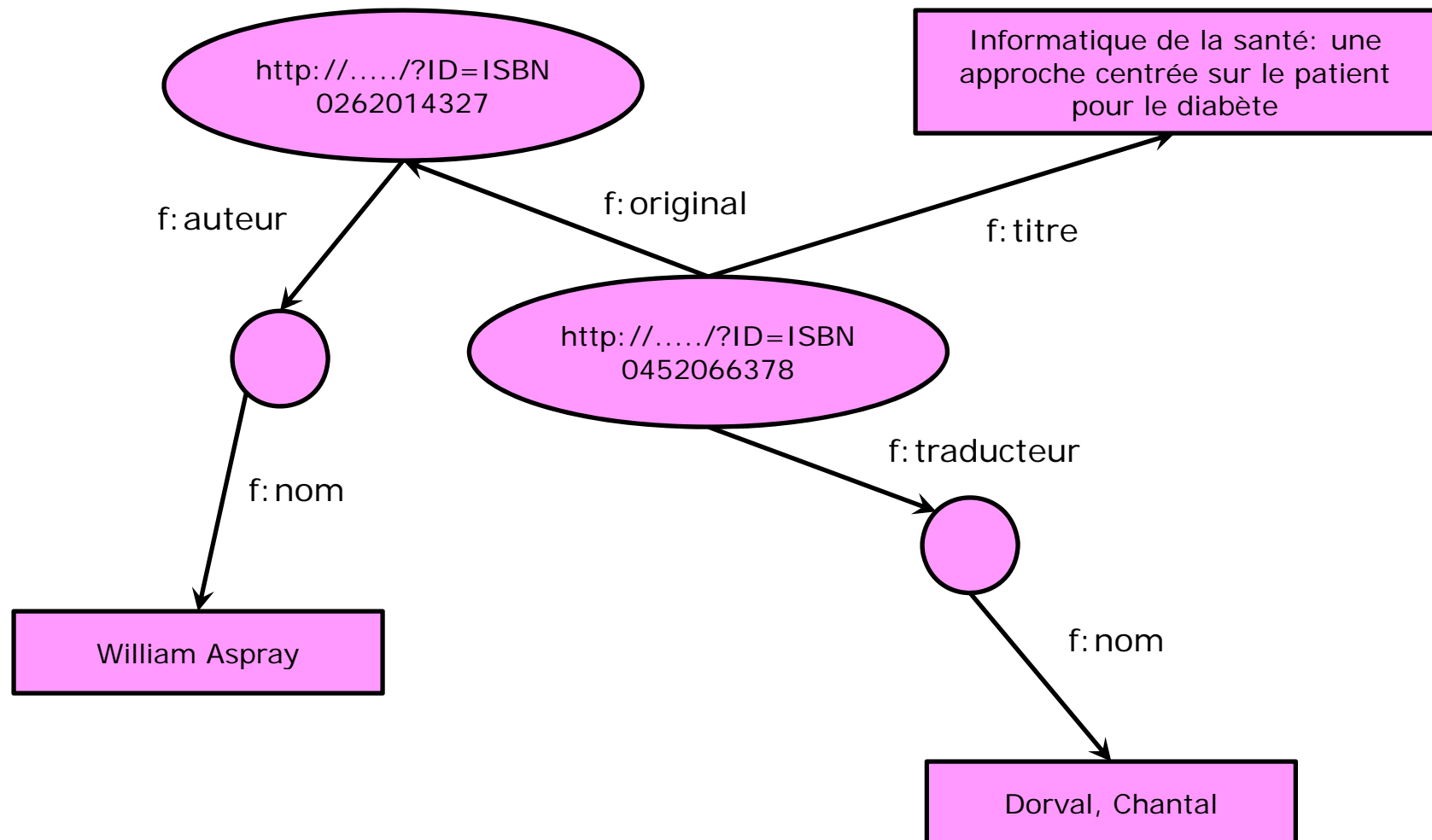
ID	Publ. Name	City
id_qpr	The MIT Press	Boston

*Example cont.*

**Example cont.**

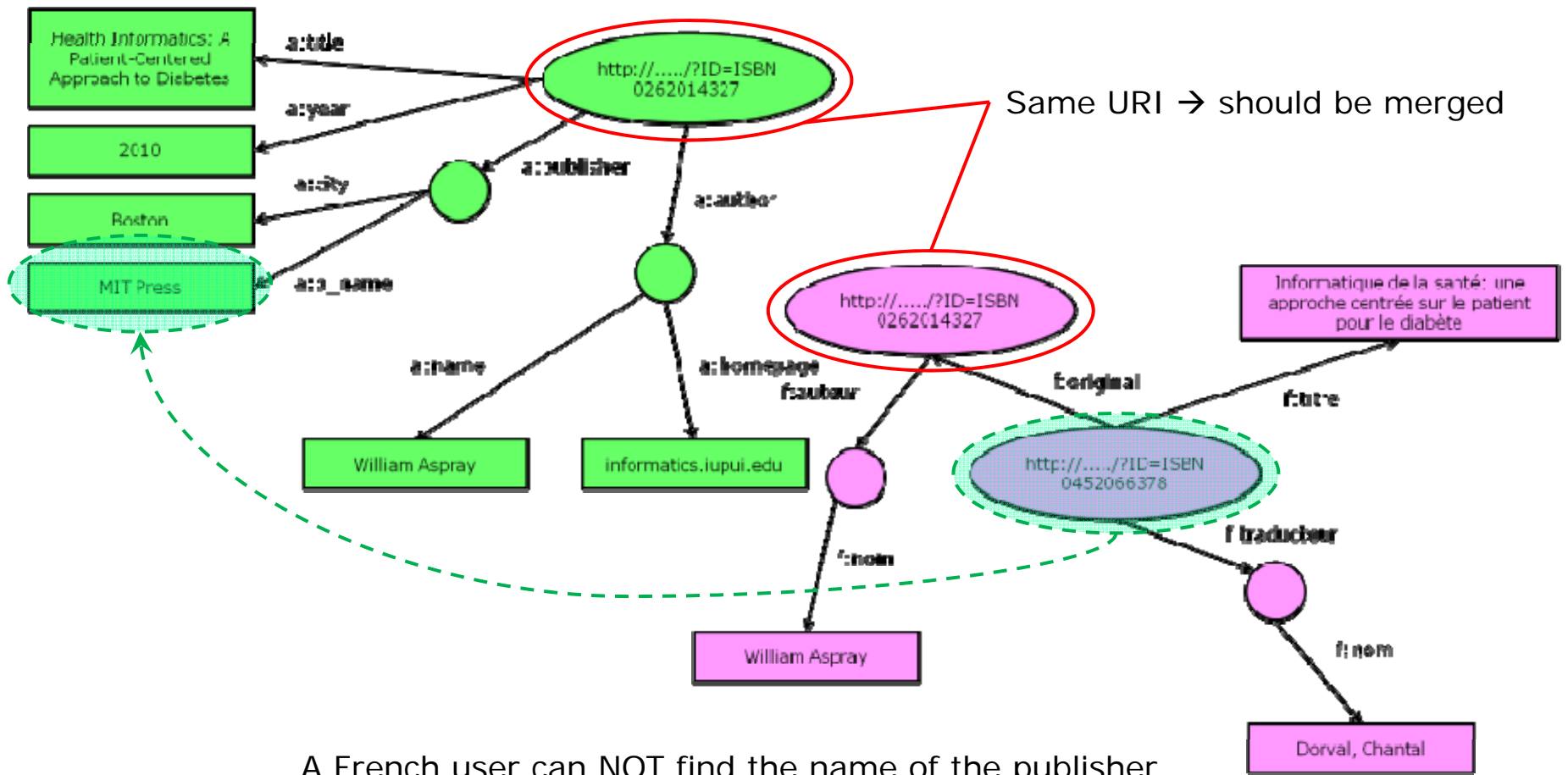
- Sample Fr Bookstore Dataset (e.g., Excel)

	A	B	D	E
1	<b>ID</b>	<b>Titre</b>	<b>Traducteur</b>	<b>Original</b>
2	ISBN0452066378	Informatique de la santé: une approche centrée sur le patient pour le diabète	\$A\$13	ISBN0262014327
3				
4				
5				
6	<b>ID</b>	<b>Auteur</b>		
7	ISBN0262014327	\$A\$12		
8				
9				
10				
11	<b>Nom</b>			
12	Aspray, William			
13	Dorval, Chantal			

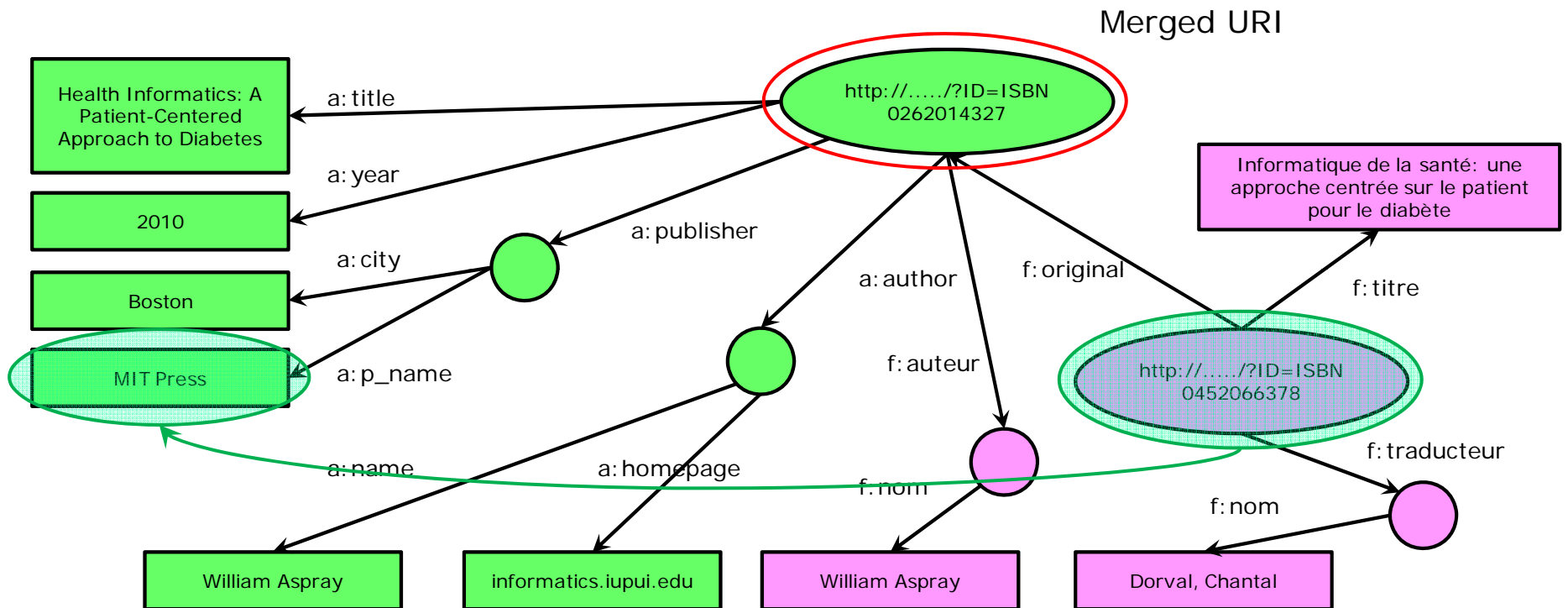
**Example cont.**



### Example cont.

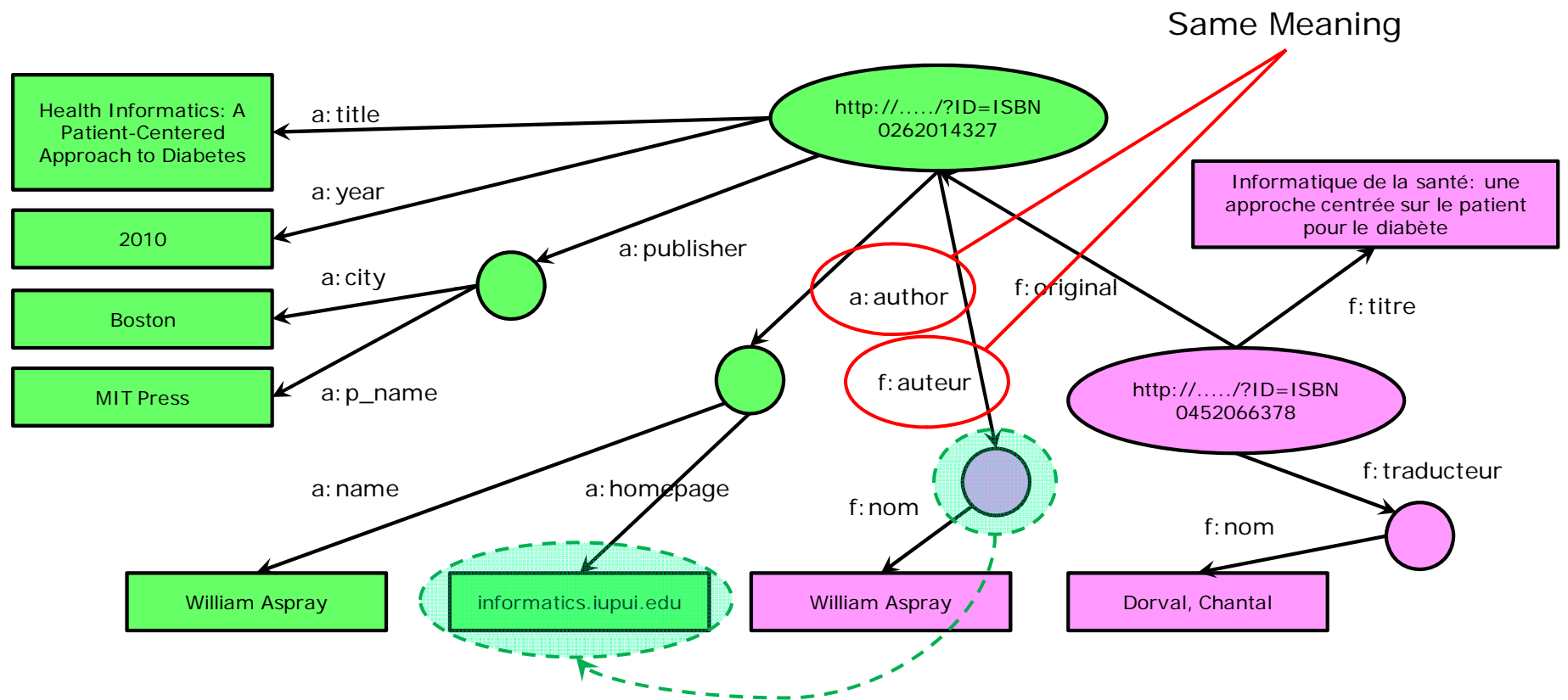


### Example cont.



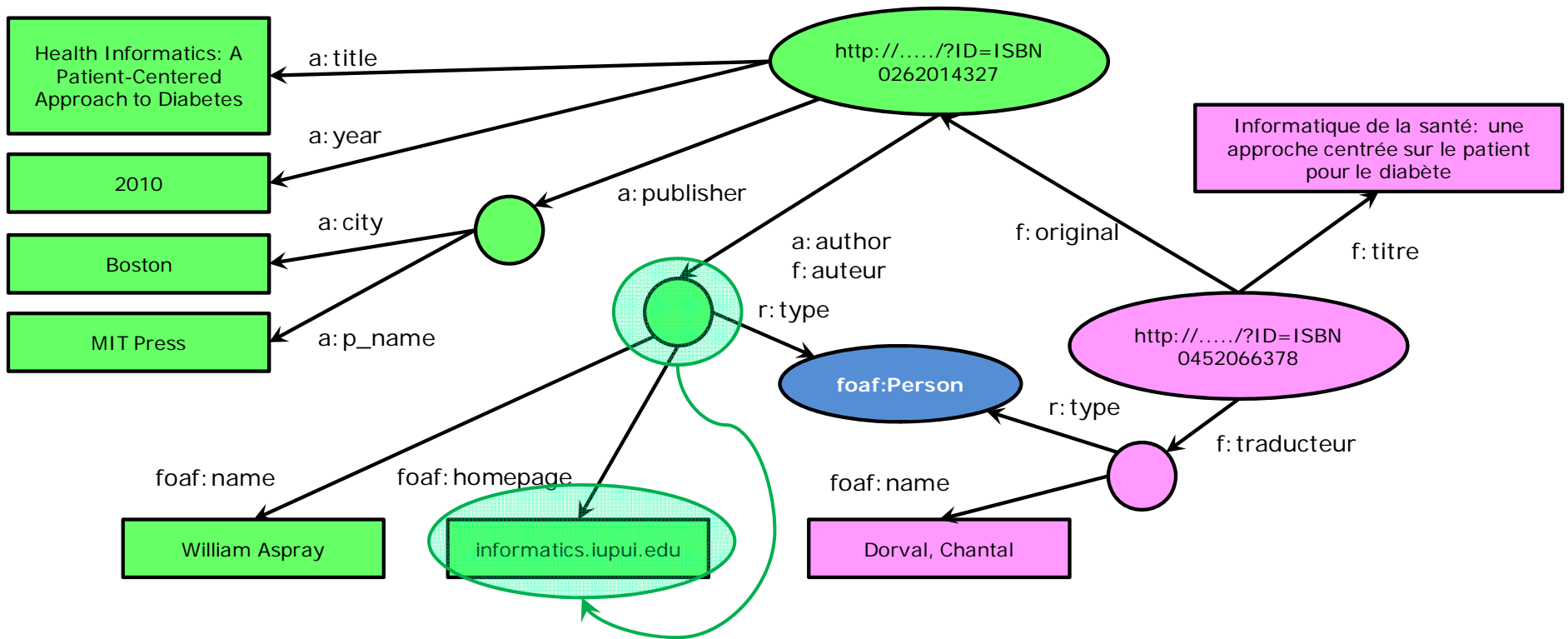
A French user can find the name of the publisher

### Example cont.



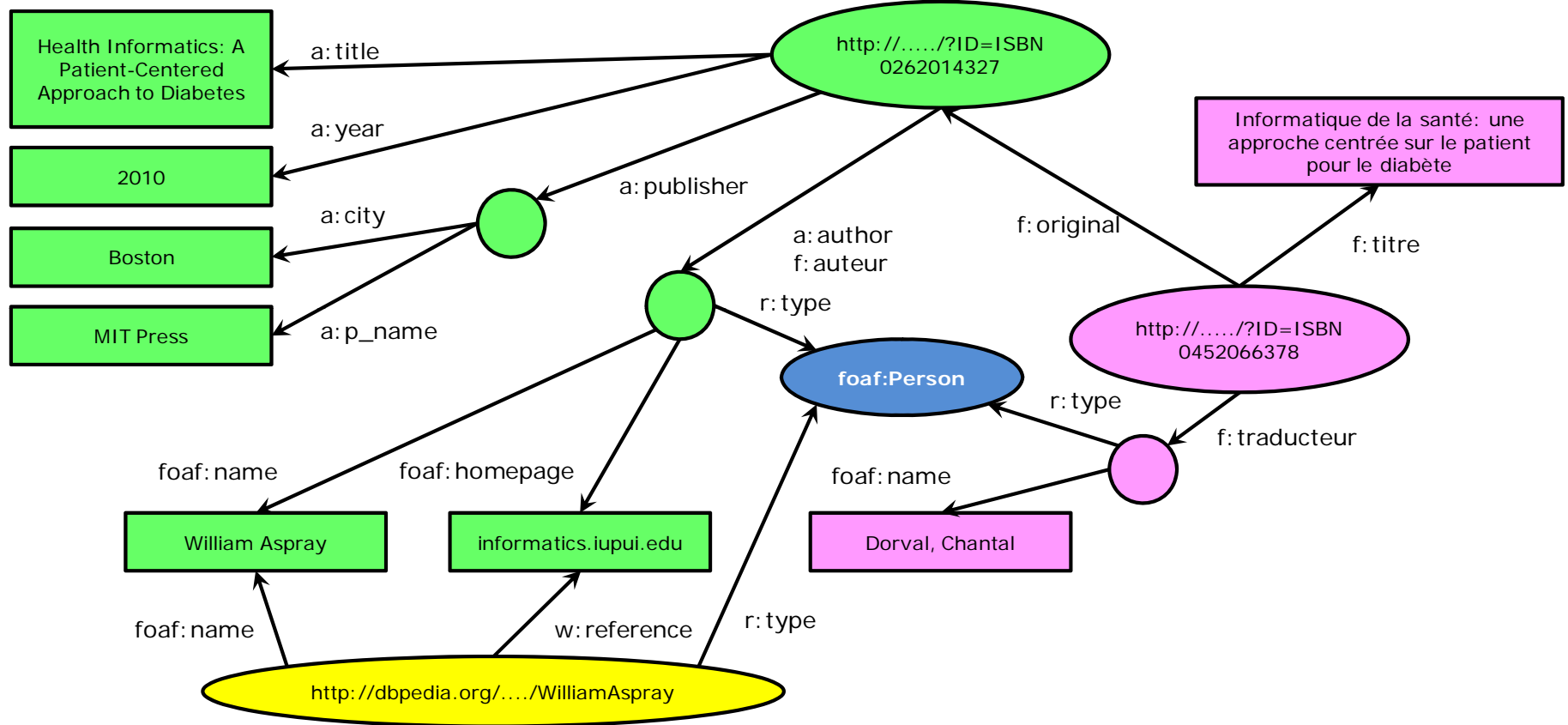
A French user can NOT find the author's website

### Example cont.



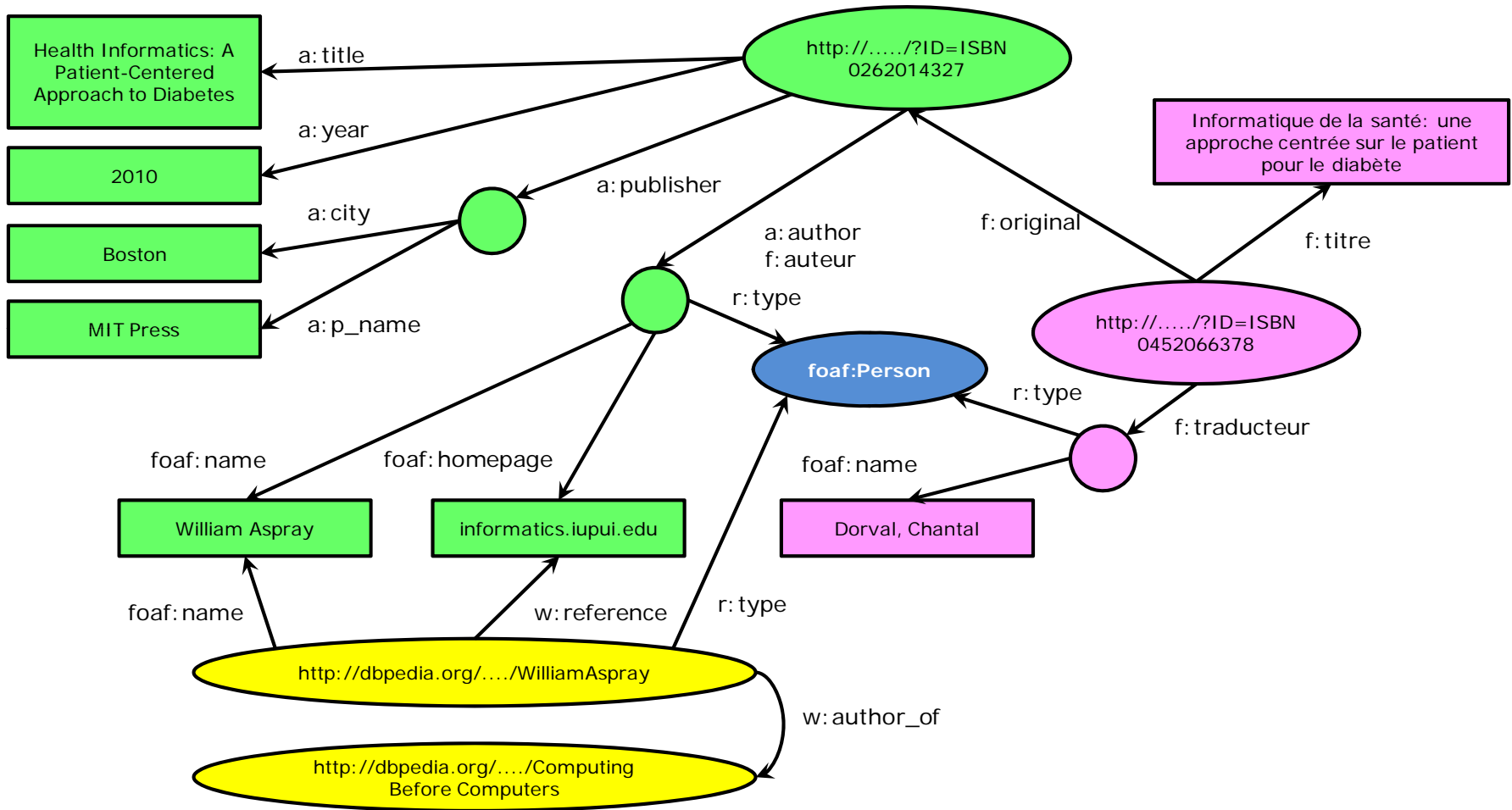
A French user can find the author's website

**Example cont.**



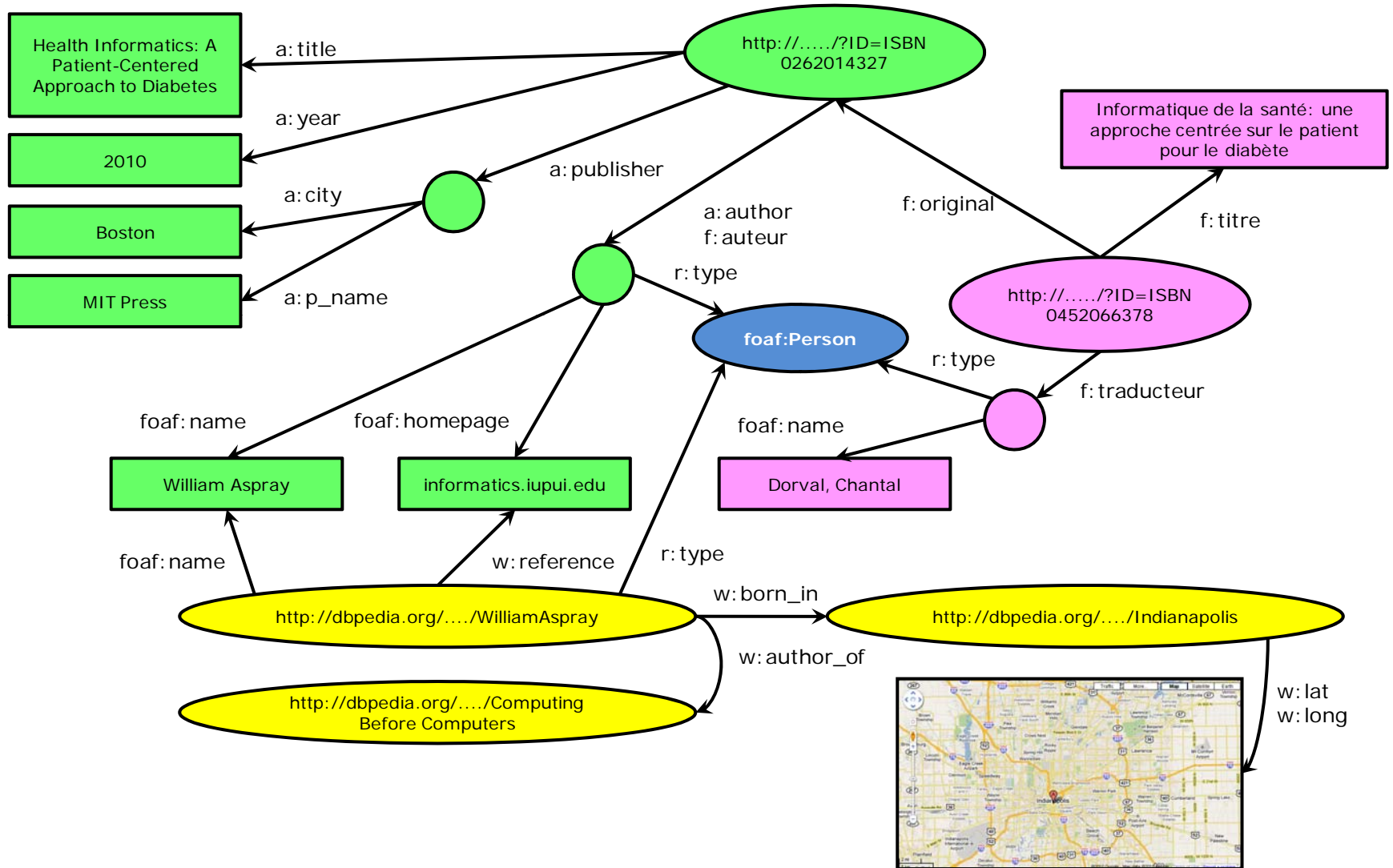
Integration into Wikipedia via DBpedia

**Example cont.**



Further integration into Wikipedia via DBpedia

**Example cont.**



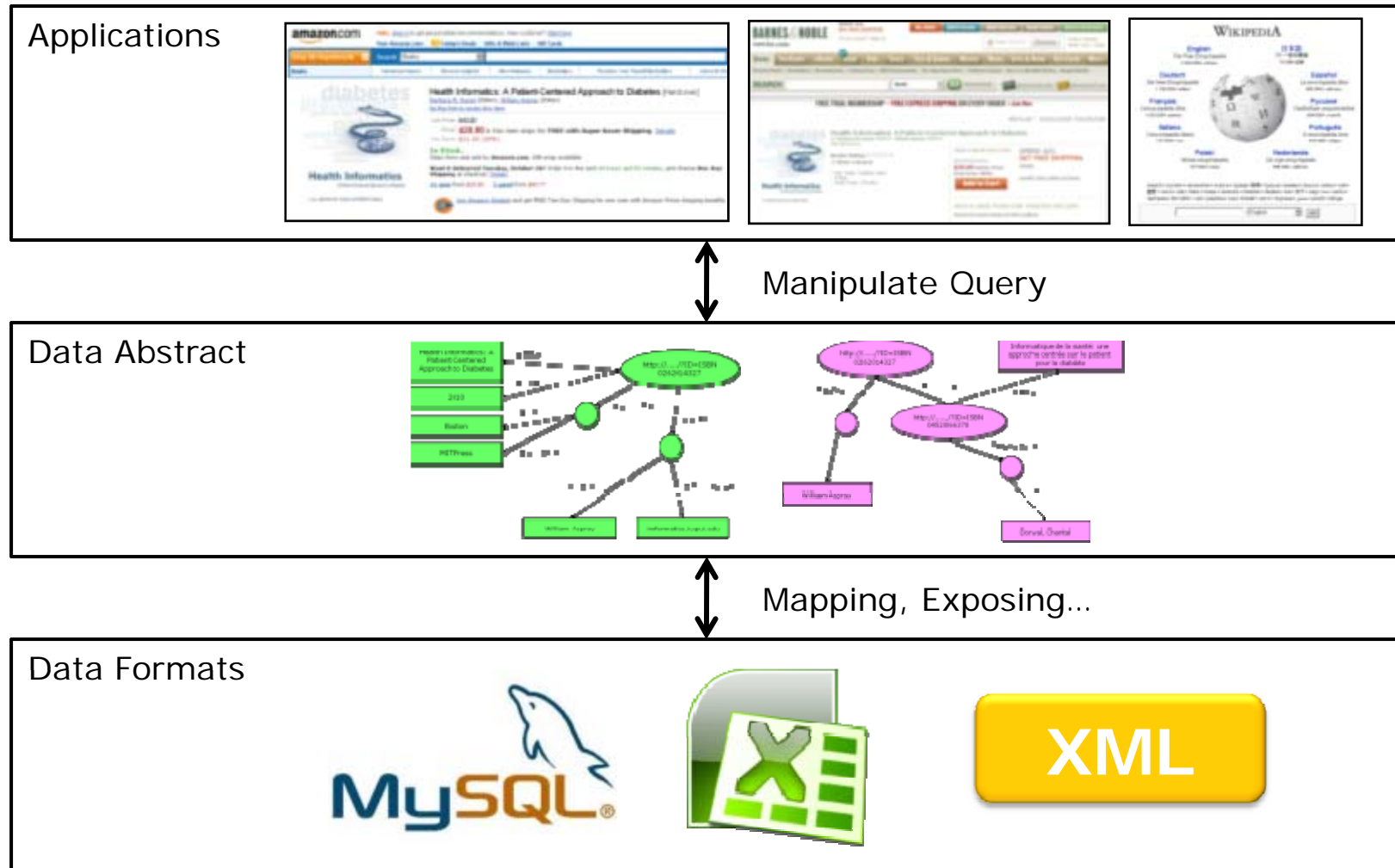
Further integration into Wikipedia via Dbpedia + Google Map

## *Example cont.*

- ❑ What did we do?
  - We combined **different datasets** that
    - are somewhere on the web
    - are of different formats (mysql, excel sheet, XHTML, etc)
    - have different names for relations
  - We could combine the data because some **URI-s were identical** (the ISBN-s in this case)
  - We could add some simple additional information, **using common terminologies** that a community has produced
  - As a result, **new relations could be found** and retrieved



## Example cont.



## Example cont.

Applications



↕ SPARQL

Data Abstract  
(OWL, RDF)



↕ RDB → RDFa

Data Formats



# Summary

1. Definition
2. Purpose
3. Relationship to The Hypertext Web
4. Skeptical Reactions
5. Components
6. Challenges
7. Projects
8. Example