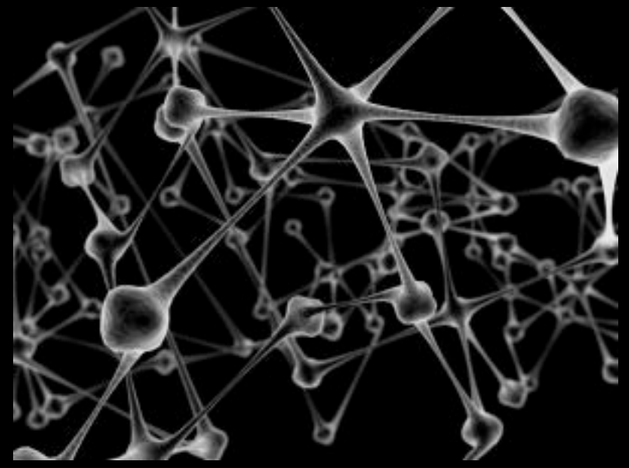


# Semantic Technologies

Graham McLeod
Inspired & Promis Ltd

Semantics refers to the representation of meaning with words and symbols. The principles have long been understood and used in the human realm for the organization and enrichment of knowledge. With the increasing power and connectedness of computers and new standards for interoperation, semantics are enabling machines to approach human intelligence and to provide much more powerful tools in the management, integration, enhancement, analysis and interpretation of information.

The marriage of these technologies and the World Wide Web, foretold years ago by the inventor of the web, Tim Berners Lee, brings about the Semantic Web. This promises to be at least as fundamental a change as the www was... Welcome to Web 3.0

# Semantic Technology Coverage

What is it?

What can it do?

Why is it important?

Concepts, Definitions

RDF

OWL

SparQL

Content Markup

Open Linked Data

Applications and Demo's

Discussion

Sources

# Semantic

Refers to the representation of meaning with symbols

Symbols can be words, or graphics (e.g. nodes in a graph)

As distinct from representation/syntax

E.g. a diagram or sentence or matrix is a representation of an underlying meaning

The same meaning can be represented in different ways: e.g. English or French

# Concept

Any thing can be pointed at by a URL (URI)

Now add a relationship, which is typed

And structure to the description of things

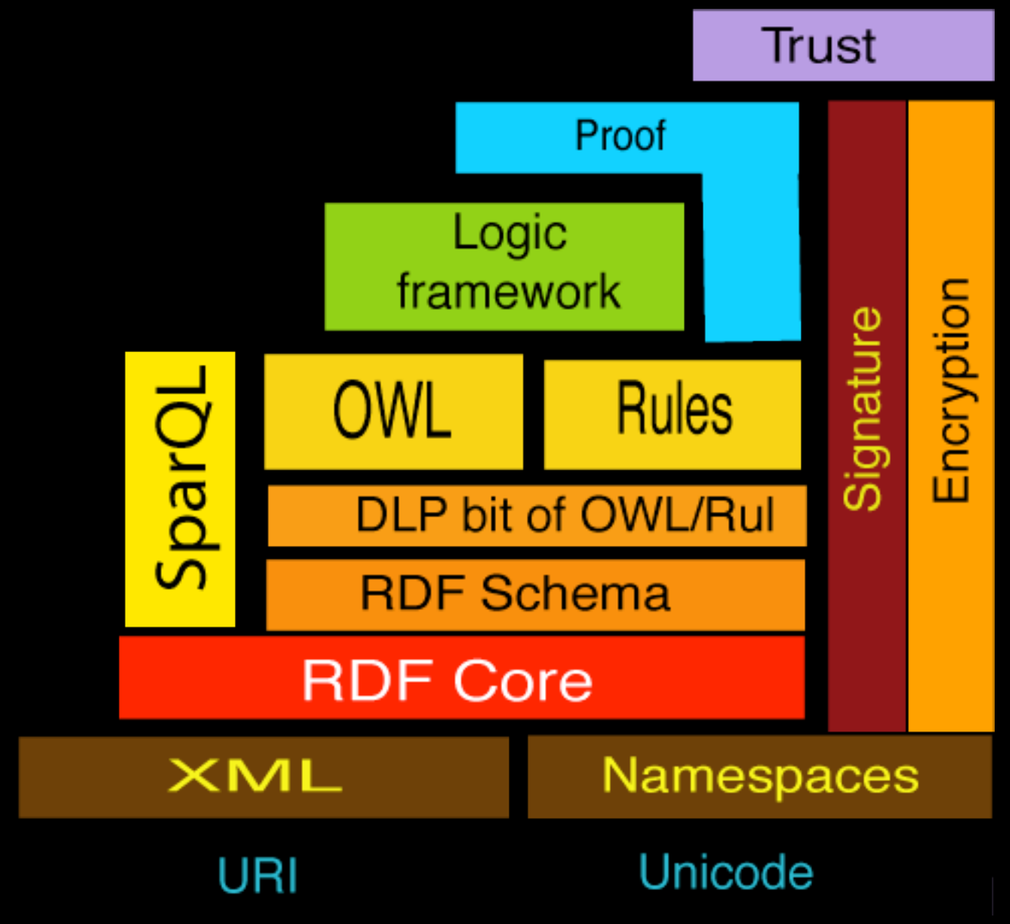
and...

You can run queries over the world wide web as a database!

We just need some standards....

# Once more with Meaning

- Everything has a URL (actually a URI) - we can find them (or information about them..)
- Things are typed and we know what they are
- Links are typed and we know what they mean
- Queries can span multiple sites as one conceptual database
- Automatically merge and infer new information



# Semantic Web

A term coined to describe the result of deploying semantic concepts on the World Wide Web

Raft of standards defined by W3C

- modeling data e.g. RDF, RDFS
- ontology (organising knowledge), rules, inferencing e.g. OWL
- querying data e.g. SPARQL
- analysis and inferencing

http://www.ludssurfingpage.bravepages.com/gallery.htm

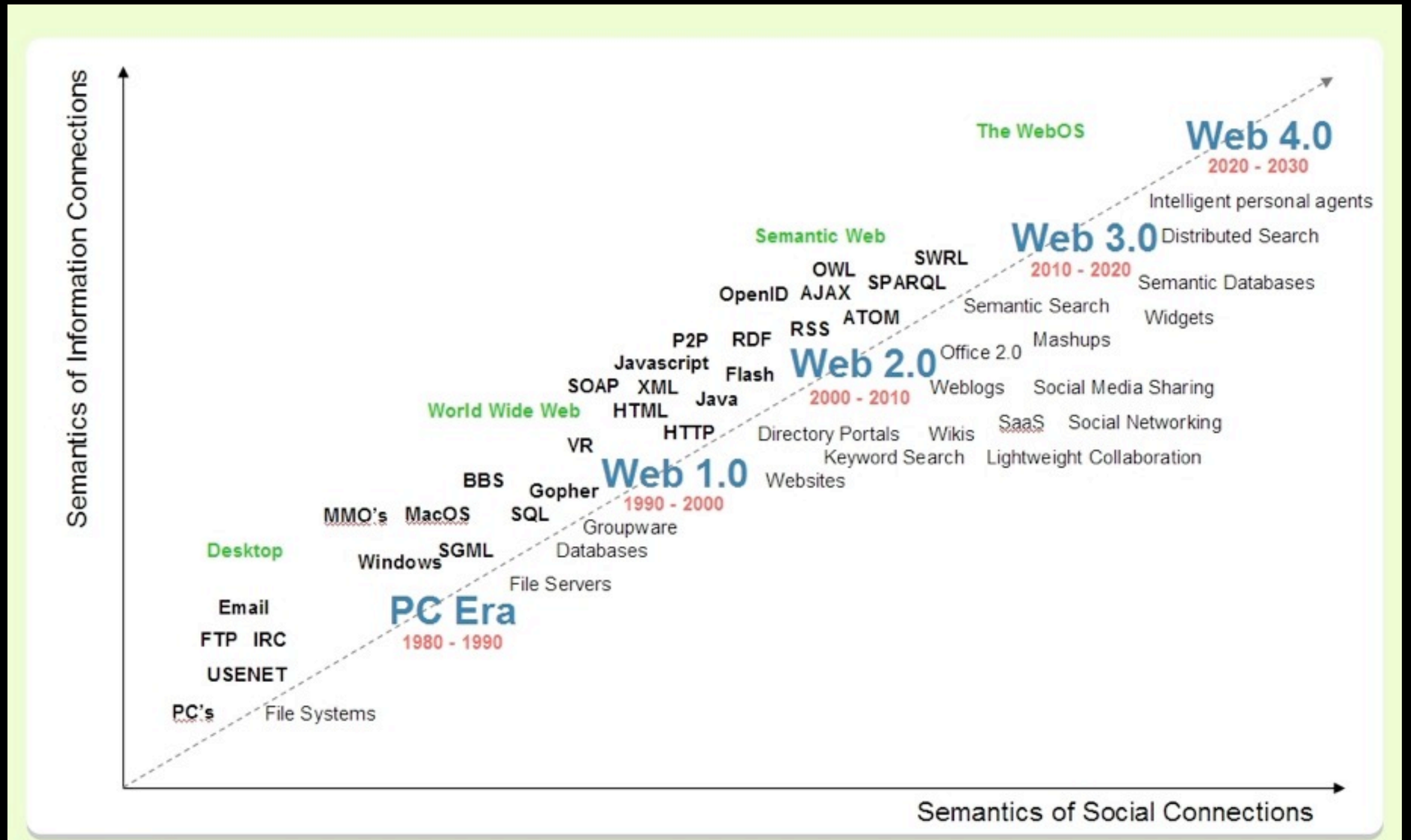
Can Robots Surf?

Web 1.0 - Content
Human Navigation
Search

Web 2.0 Transactional/Services
Social
Fine Grained Interactivity
Mobile
Visual / Video

Web 3.0 Semantic
Machine Navigation
Query web as a massive interlinked data source

# Generations of Web




SOME RIGHTS RESERVED

Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

# What and Why

Natural language - quick easy knowledge sharing, a natural human capability

Communicating enough meaning to result in an action..

Represent, combine, share knowledge

Between people, people and programs and programs and programs

Enable systems that operate on knowledge, rather than data

Distribution on web scale

Across communities

Flexibly enough to integrate in real time with other data we have not seen before

# What and Why (2)

Tabular Data, Relational Data, Object Data => Semantic Data

Relationships at least as important as data

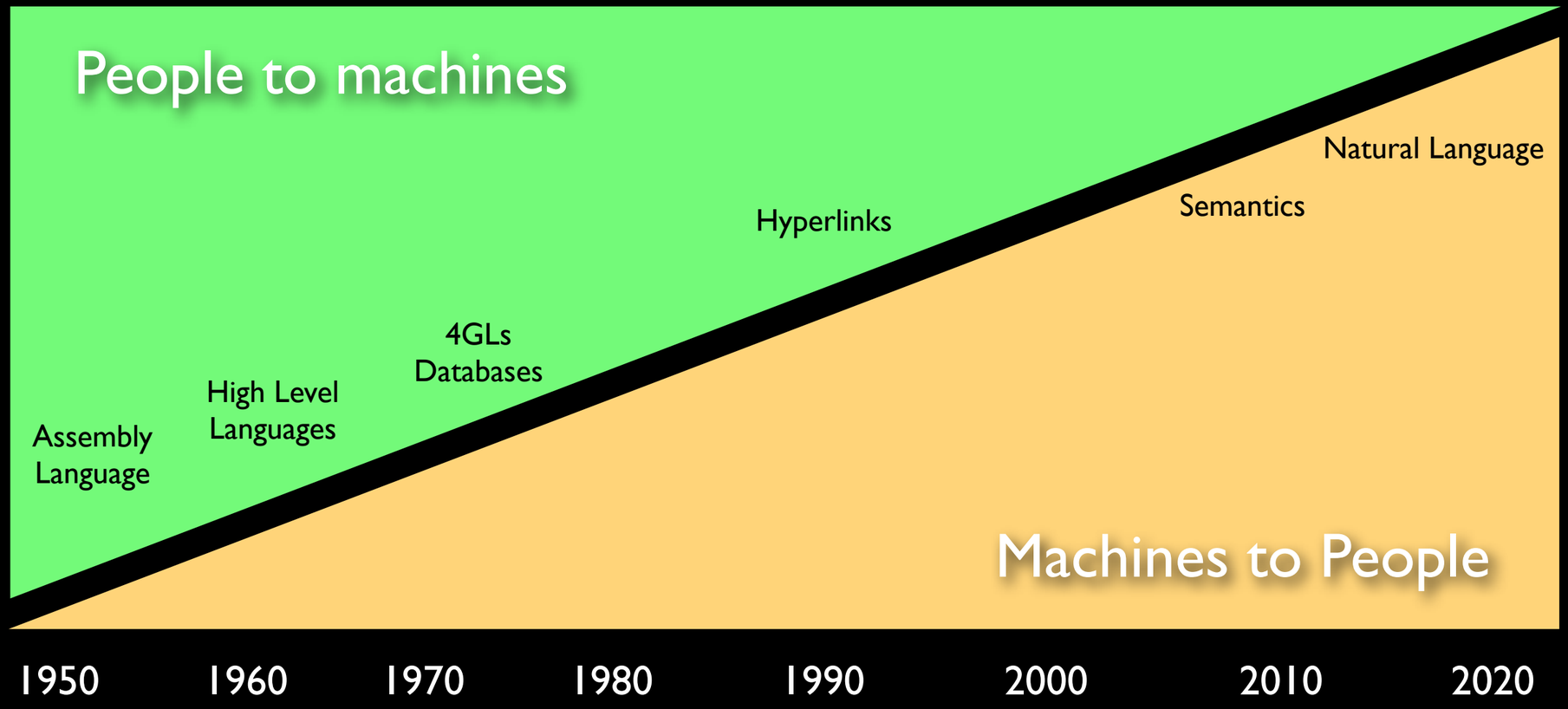
Meta data is data

Schema is inherent in the data + relationships

Easier design, schema creation, integration, query...

Flexible, perpetual “beta”

# Who's Adapting?



# Meaning

Sentences: Subject [predicate / property] object

John knows Graham

John salary 30 000

Entities/Domain Concepts

Triple Store

# RDF Basics

What and where from - W3C

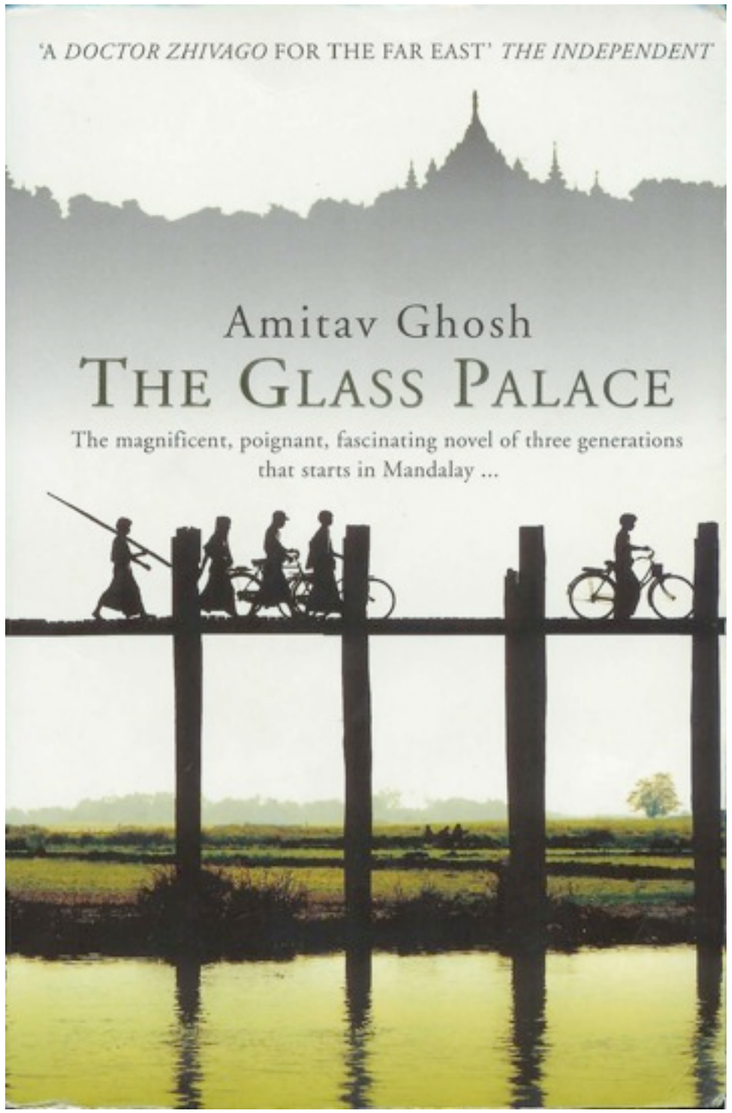
RDF Data Model - Sees world as a conceptual connected graph

URIs as strong keys - every fact uniquely identified and addressable

Blank nodes - where data does not need to be exposed outside

Literal Values - special provisions and typing

# We start with a book...




©Ivan Herman, W3C

# A simplified bookstore data (dataset “A”)

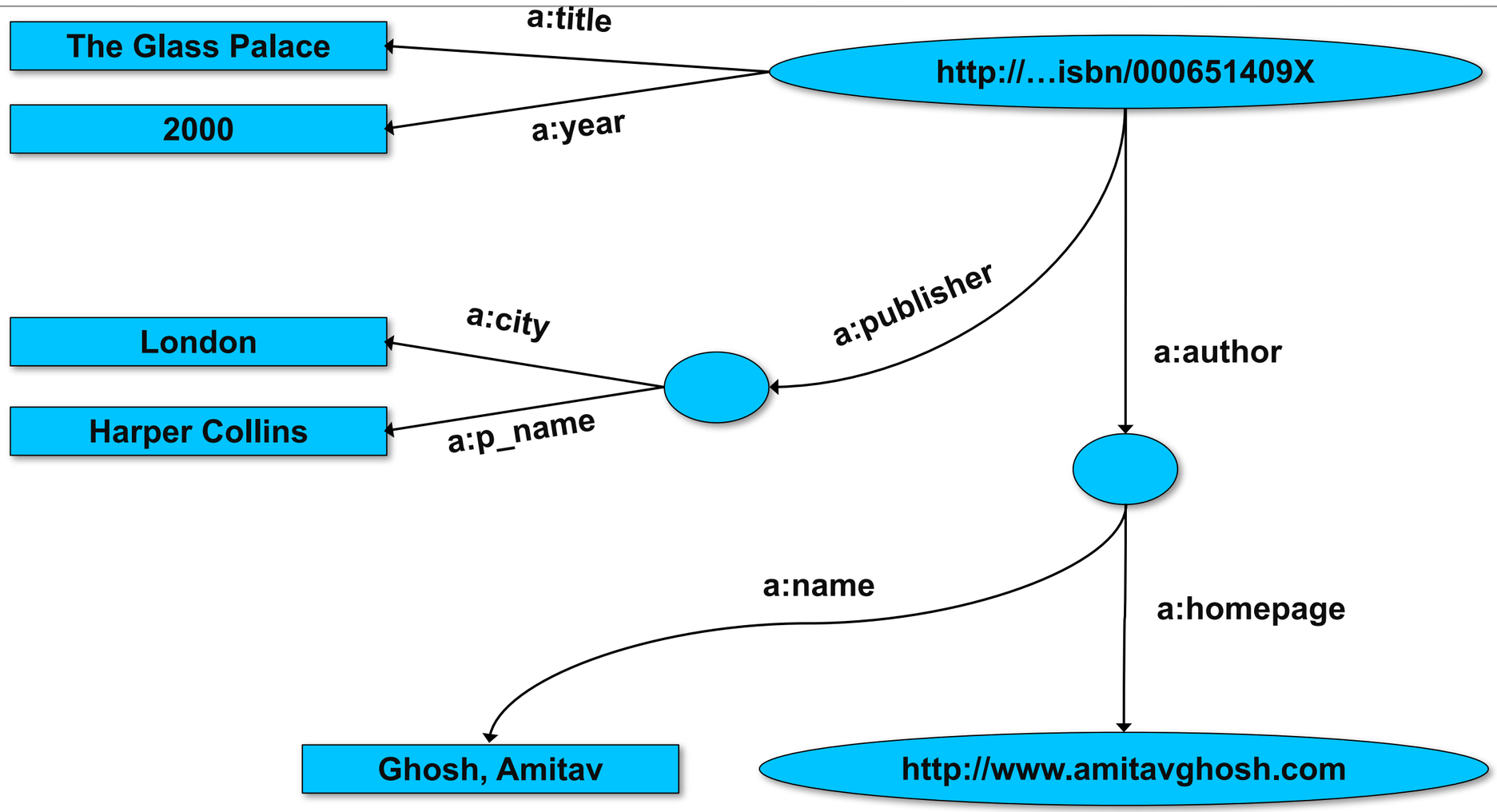
| ISBN | Author | Title | Publisher | Year |
|---|---|---|---|---|
| 000651140 9X | id_xyz | The Glass Palace | id_qpr | 2000 |

| ID | Name | Homepage |
|---|---|---|
| id_xyz | Ghosh, Amitav | http://www.amitavghosh.com |

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

©Ivan Herman, W3C

# 1st: export your data as a set of *relations*



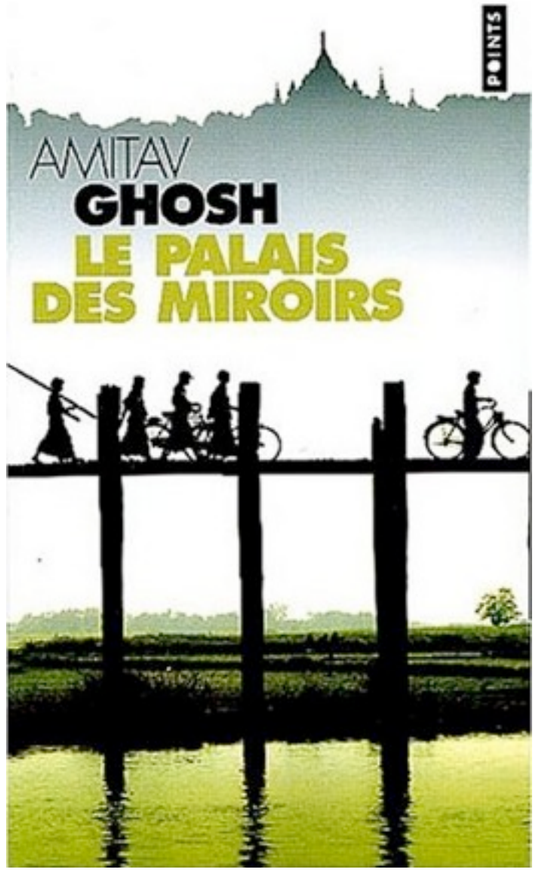

©Ivan Herman, W3C

# 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

©Ivan Herman, W3C

# Same book in French…



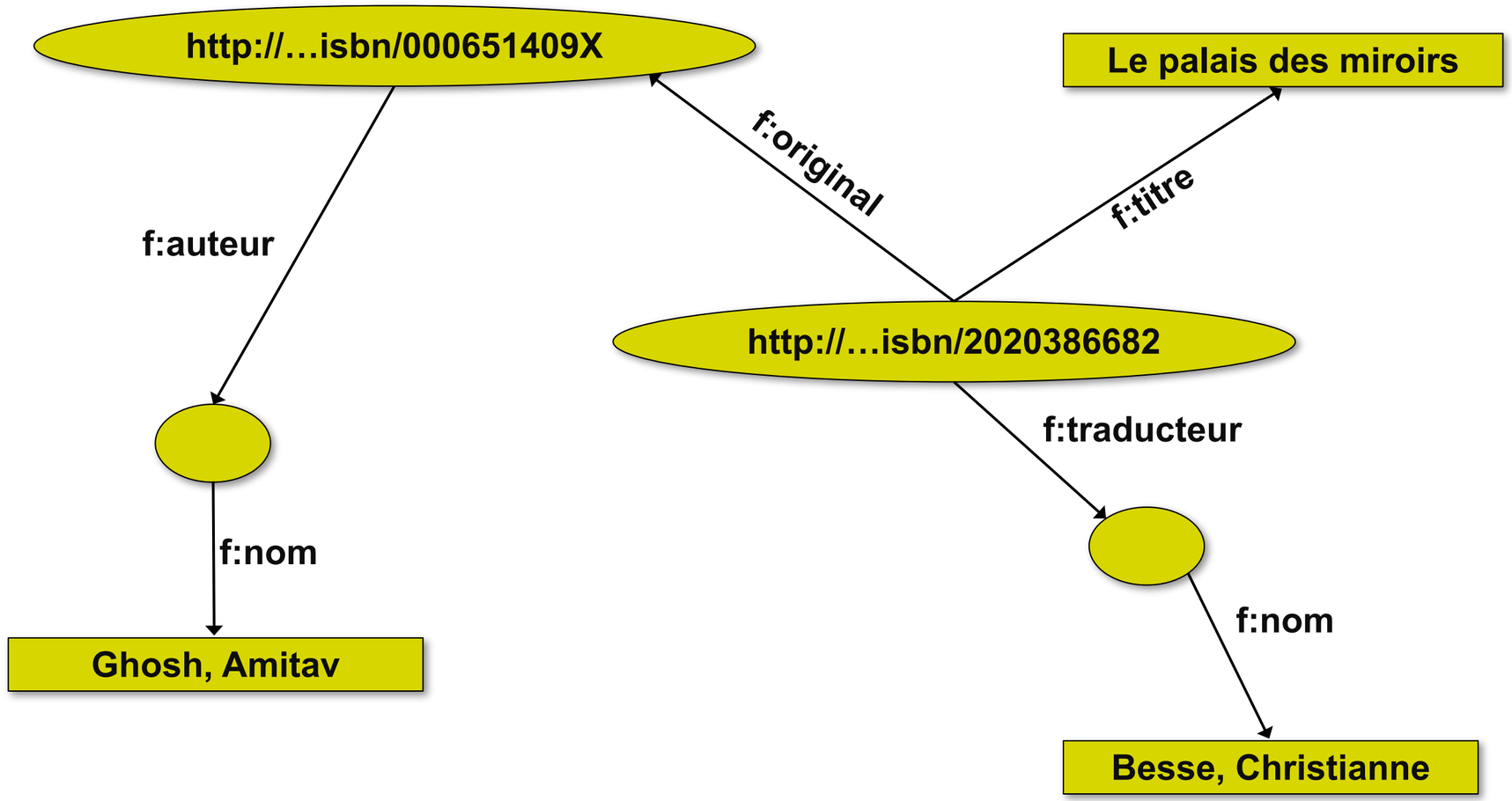

©Ivan Herman, W3C

# Another bookstore data (dataset “F”)

| | A | B | C | D |
|---|---|---|---|---|
| 1 | ID | Titre | Traducteur | Original |
| 2 | ISBN 2020286682 | Le Palais des Miroirs | $A12$ | ISBN 0-00-6511409-X |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | ID | Auteur | | |
| 7 | ISBN 0-00-6511409-X | $A11$ | | |
| 8 | | | | |
| 9 | | | | |
| 10 | Nom | | | |
| 11 | Ghosh, Amitav | | | |
| 12 | Besse, Christianne | | | |

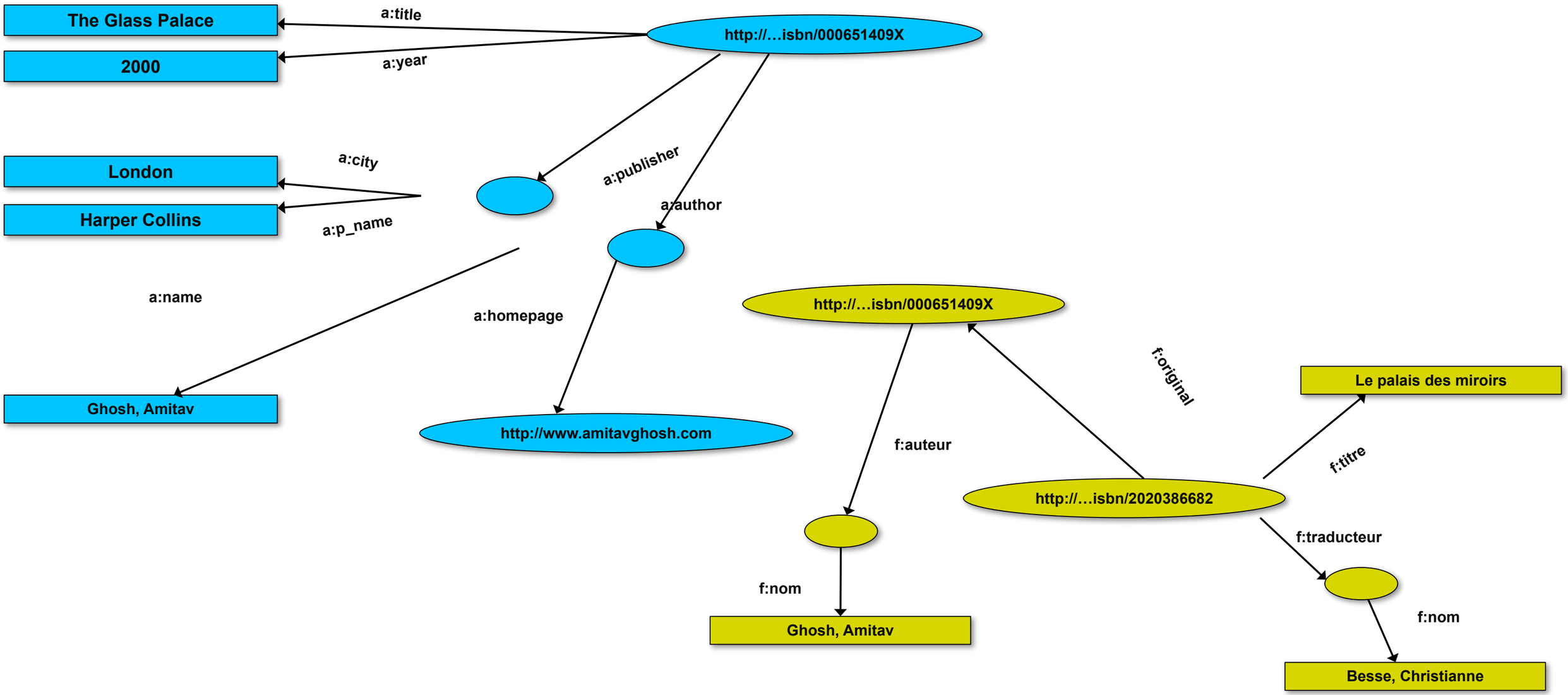
©Ivan Herman, W3C

# 2nd: export your second set of data



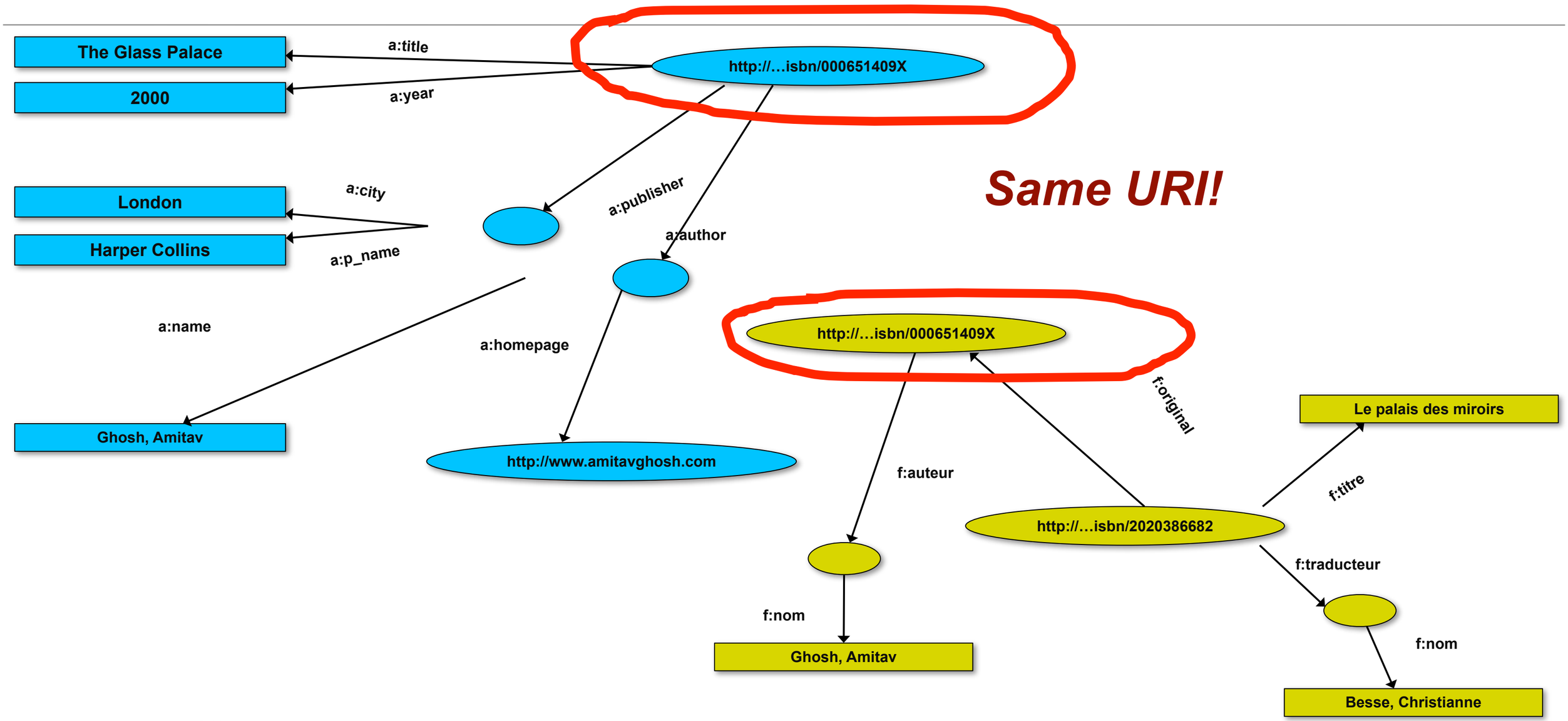

©Ivan Herman, W3C

# 3rd: start merging your data



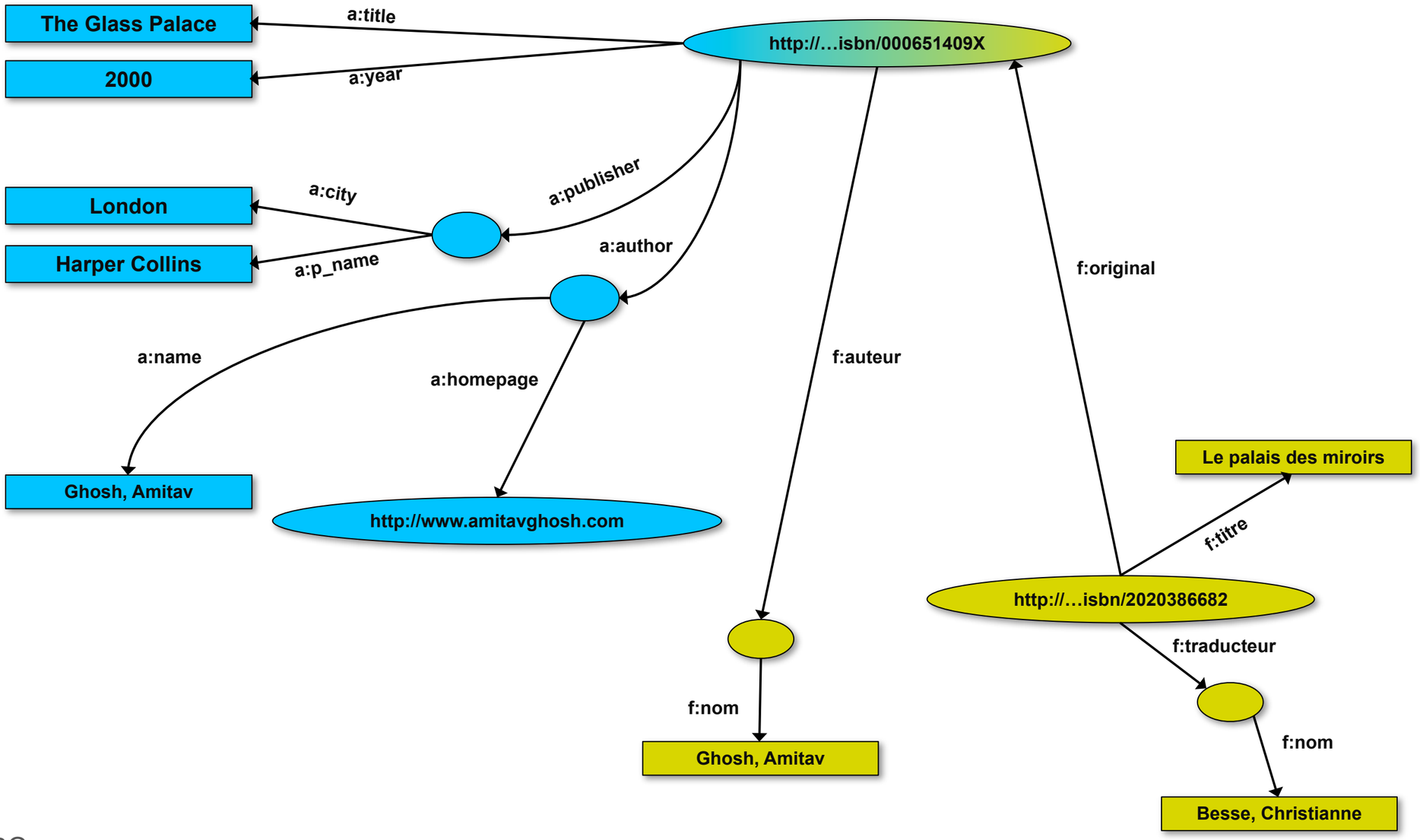

©Ivan Herman, W3C

# 3rd: start merging your data (cont)




©Ivan Herman, W3C

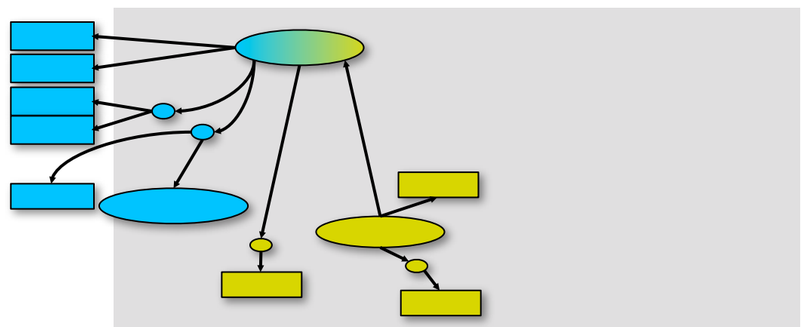
# 3rd: start merging your data




©Ivan Herman, W3C

# 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”!



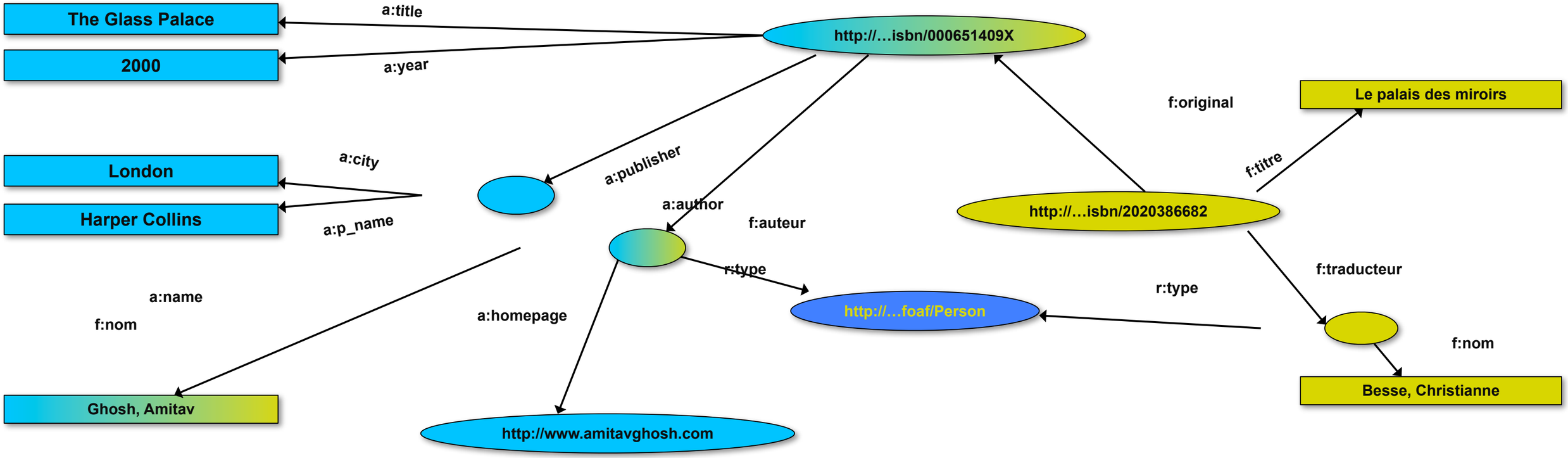
©Ivan Herman, W3C

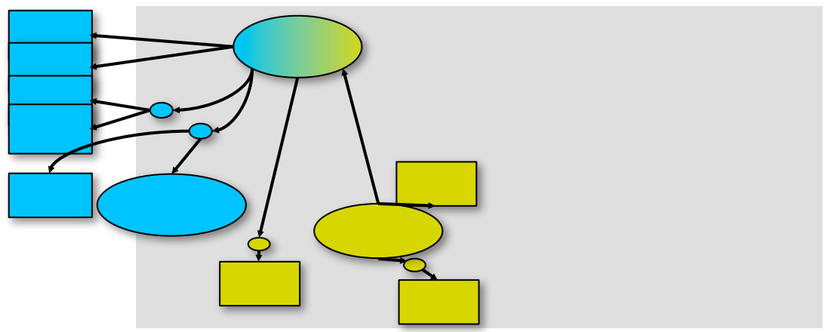
# However, more can be achieved…

- We “feel” that a:author and f:auteur should be the same
- But an automatic merge doest 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

©Ivan Herman, W3C

# 3rd revisited: use the extra knowledge

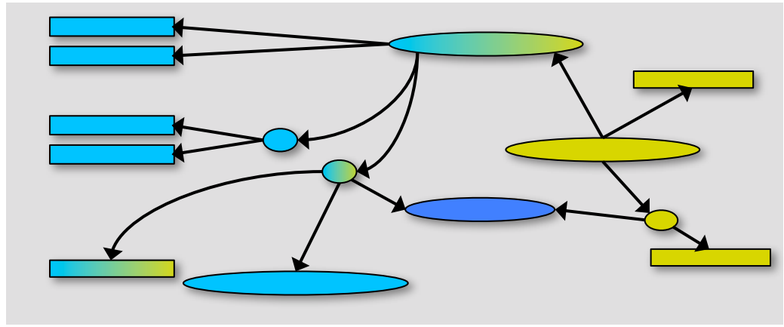


©Ivan Herman, W3C

# 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”



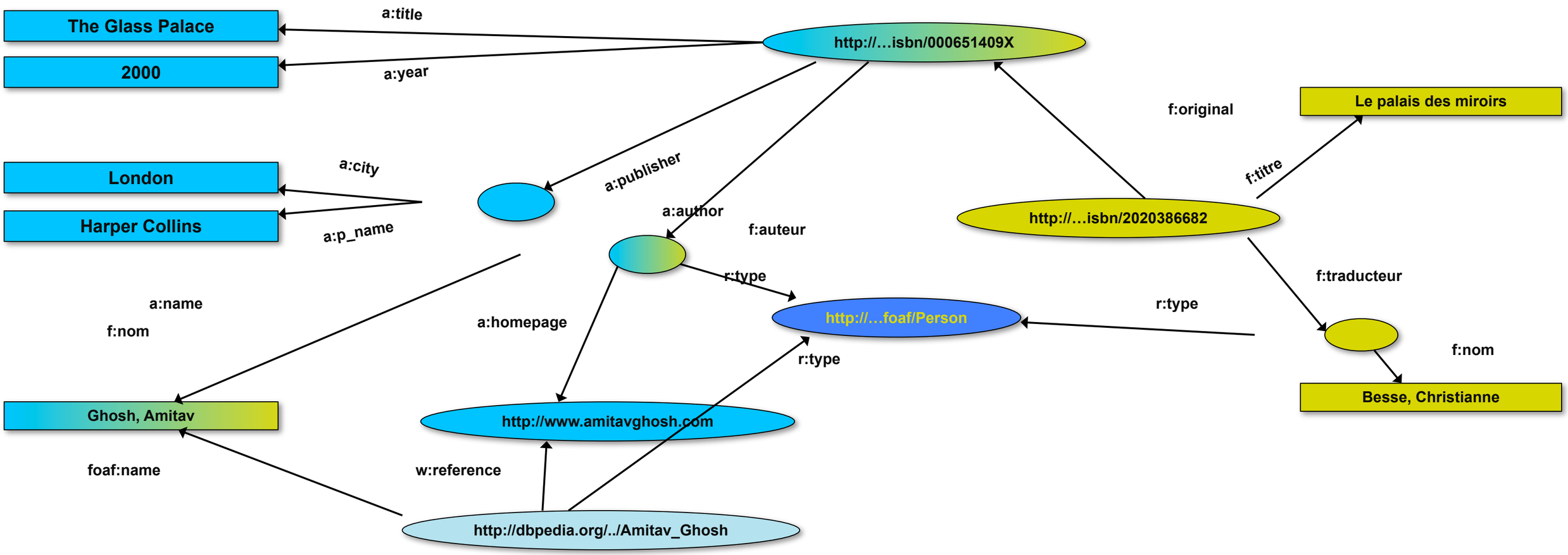
©Ivan Herman, W3C

# 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…

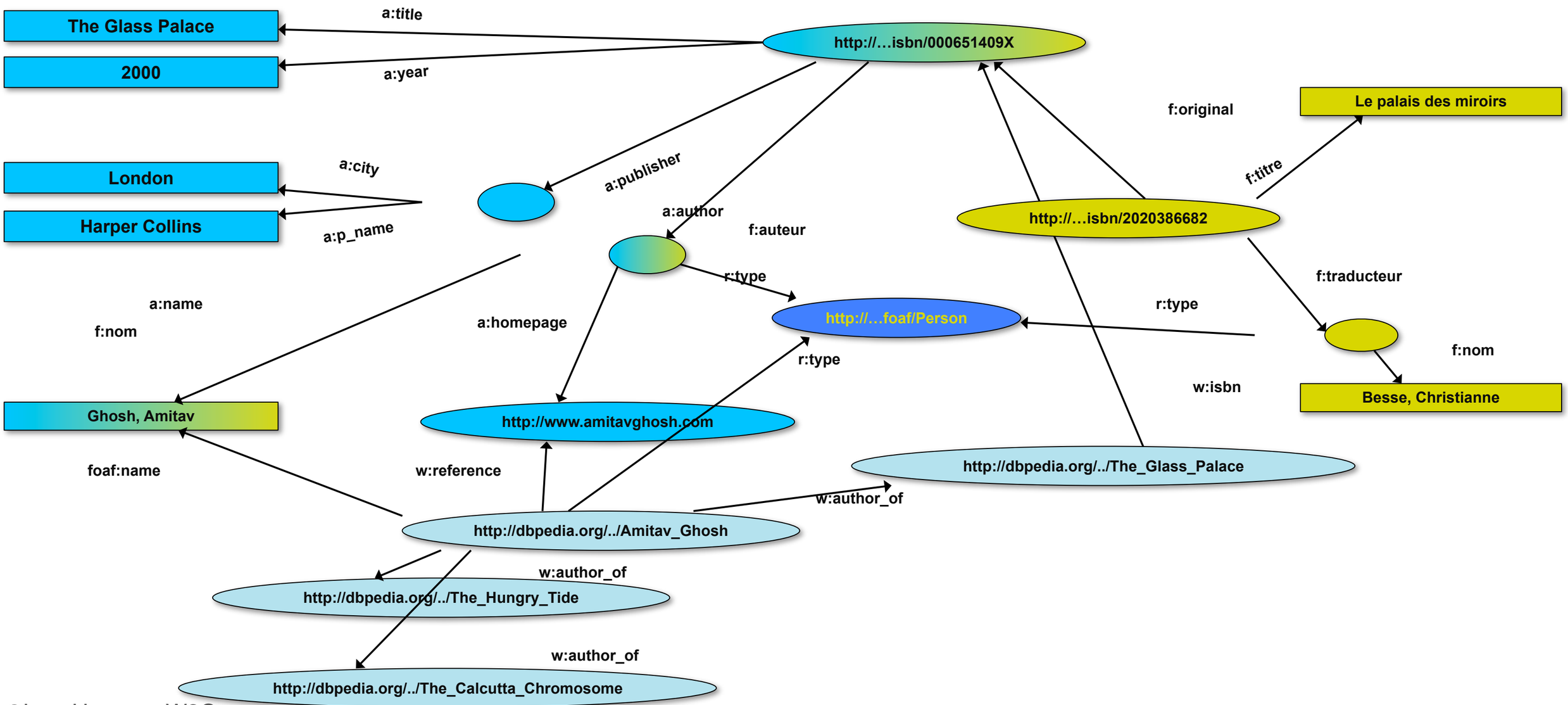
©Ivan Herman, W3C

# Merge with Wikipedia data



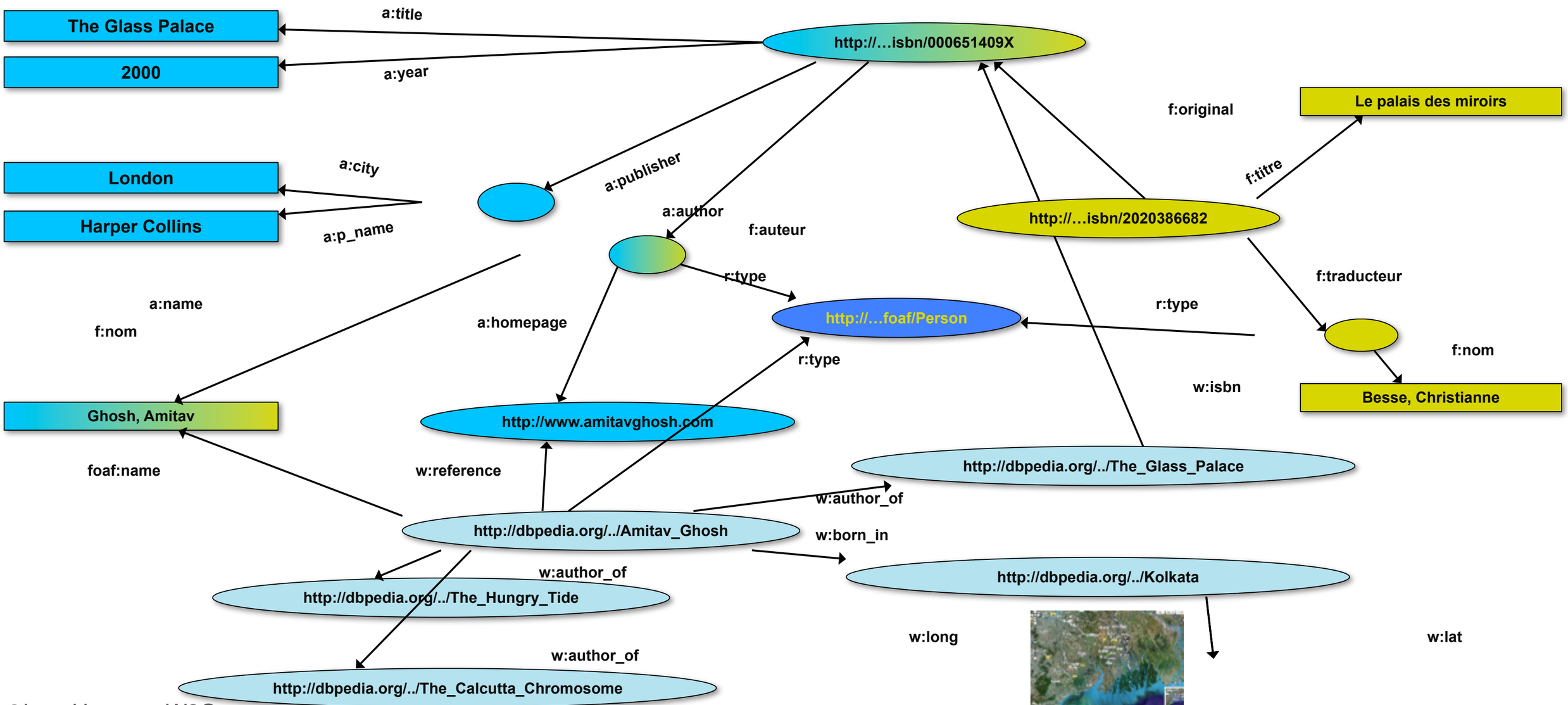

©Ivan Herman, W3C

# Merge with Wikipedia data




©Ivan Herman, W3C

# Merge with Wikipedia data




©Ivan Herman, W3C

# Triple Store

Structure

Indexes

Adding and Removing information

Queries

Merging graphs

Graph examples

# RDF Serialisation

N-Triples

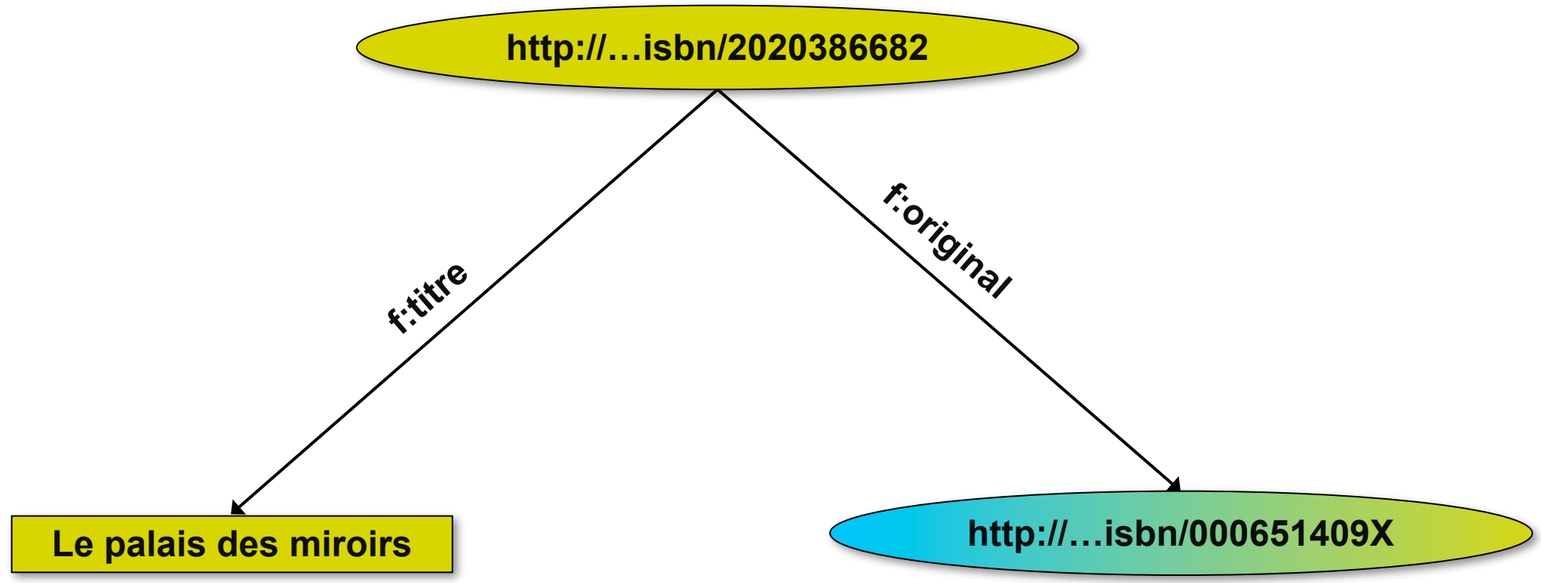
N3

RDF/XML

Turtle

RDFa

# A simple RDF example (in RDF/XML)

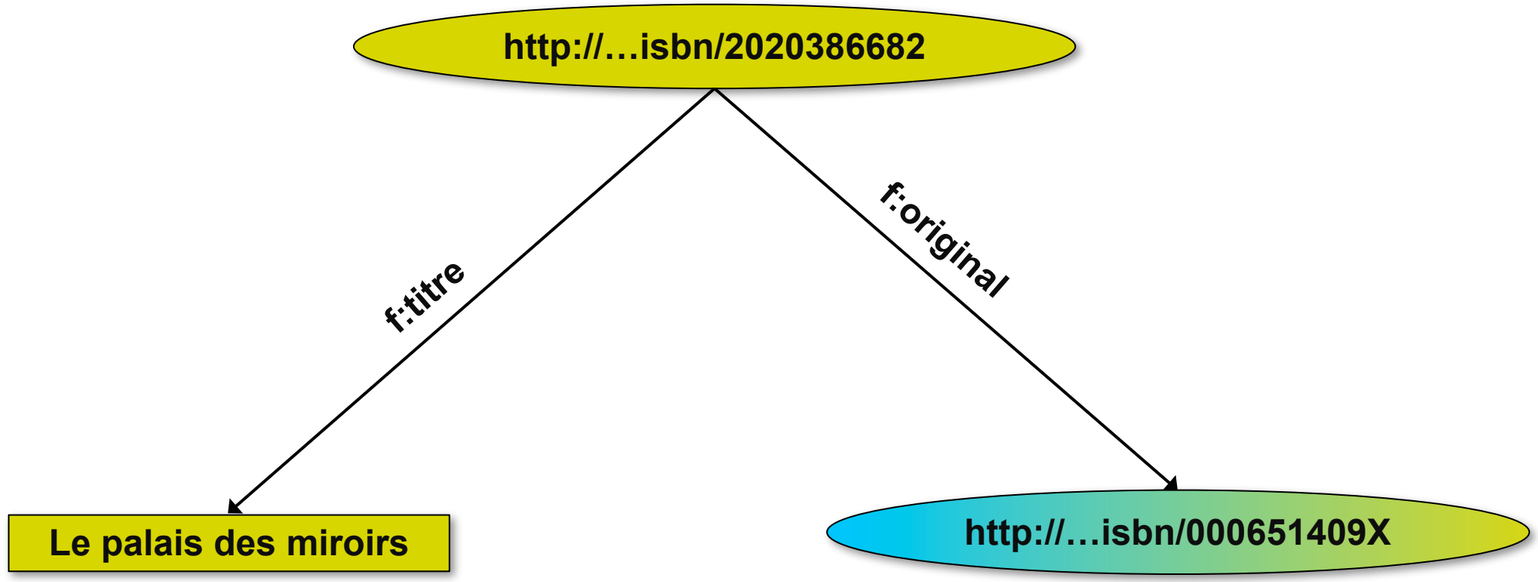



```
<rdf:Description rdf:about="http://…/isbn/2020386682">
    <f:titre xml:lang="fr">Le palais des mirroirs</f:titre>
    <f:original rdf:resource="http://…/isbn/000651409X"/>
</rdf:Description>
```

(Note: namespaces are used to simplify the URI-s)

©Ivan Herman, W3C

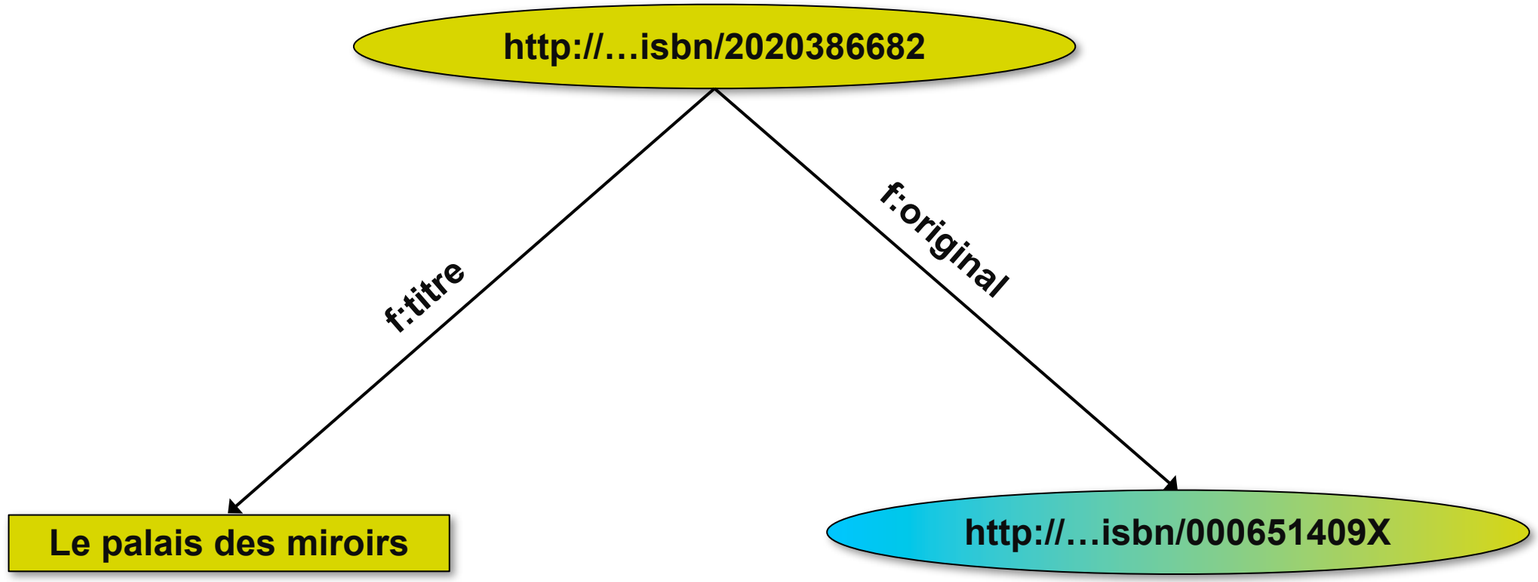
# A simple RDF example (in Turtle)




```
<http://…/isbn/2020386682>
    f:titre "Le palais des mirroirs"@fr ;
    f:original <http://…/isbn/000651409X> .
```

©Ivan Herman, W3C

# A simple RDF example (in RDFa)




```
<p about="http://…/isbn/2020386682">The book entitled
“<span property="f:title" lang="fr">Le palais des mirroirs</span>”
is the French translation of the
“<span rel="f:original" resource="http://…/isbn/000651409X">Glass
Palace</span>”</p> .
```

©Ivan Herman, W3C

# SPARQL

Query Language that operates on RDF data - simple to use

- Select
- Optional and Filter
- More complex queries
- Construct
- ASK and Describe
- Offset, limit and order by

# Example SparQL Query

Finding the links where you can pay to download an Album or a Track

```
1         PREFIX mo: <http://purl.org/ontology/mo/>
2         SELECT ?download_link_album ?download_link_track
3         WHERE
4         {
5             {
6                 <http://mm.musicbrainz.org/record/a89e1d92-5381-4dab-ba51-733137d0e431> a mo:Record.
7                 {
8                     <http://mm.musicbrainz.org/record/a89e1d92-5381-4dab-ba51-733137d0e431> mo:mailorder ?download_link_album
9                 }
10                UNION
11                {
12                    <http://mm.musicbrainz.org/record/a89e1d92-5381-4dab-ba51-733137d0e431> mo:paiddownload ?download_link_album
13                }
14            }
15            UNION
16            {
17                <http://mm.musicbrainz.org/record/a89e1d92-5381-4dab-ba51-733137d0e431> mo:has_track ?track.
18                ?track a mo:Track.
19                {
20                    ?track mo:mailorder ?download_link_track
21                }
22                UNION
23                {
24                    ?track mo:paiddownload ?download_link_track
```

# Ontology

What and why - adds classification, organization, meta data relationships

Classes and properties

Naming

Reifying relationships

OWL - Web Ontology Language

Protege - a free Ontology modeling tool from Stanford Univ.

# Semantic Resources

FOAF - Friend of a Friend - Social Graph ex Social Media (e.g. Facebook)

Freebase - Semantic resource indexing people, places, things, media...

- Community Built (like WikiPedia)

DBPedia - Mining WikiPedia to discover the semantics in the content

# OWL continued

Functional properties

Inverse properties

Disjoint classes

Examples

# Publishing

Embedding Semantics - Every organization/individual a publisher

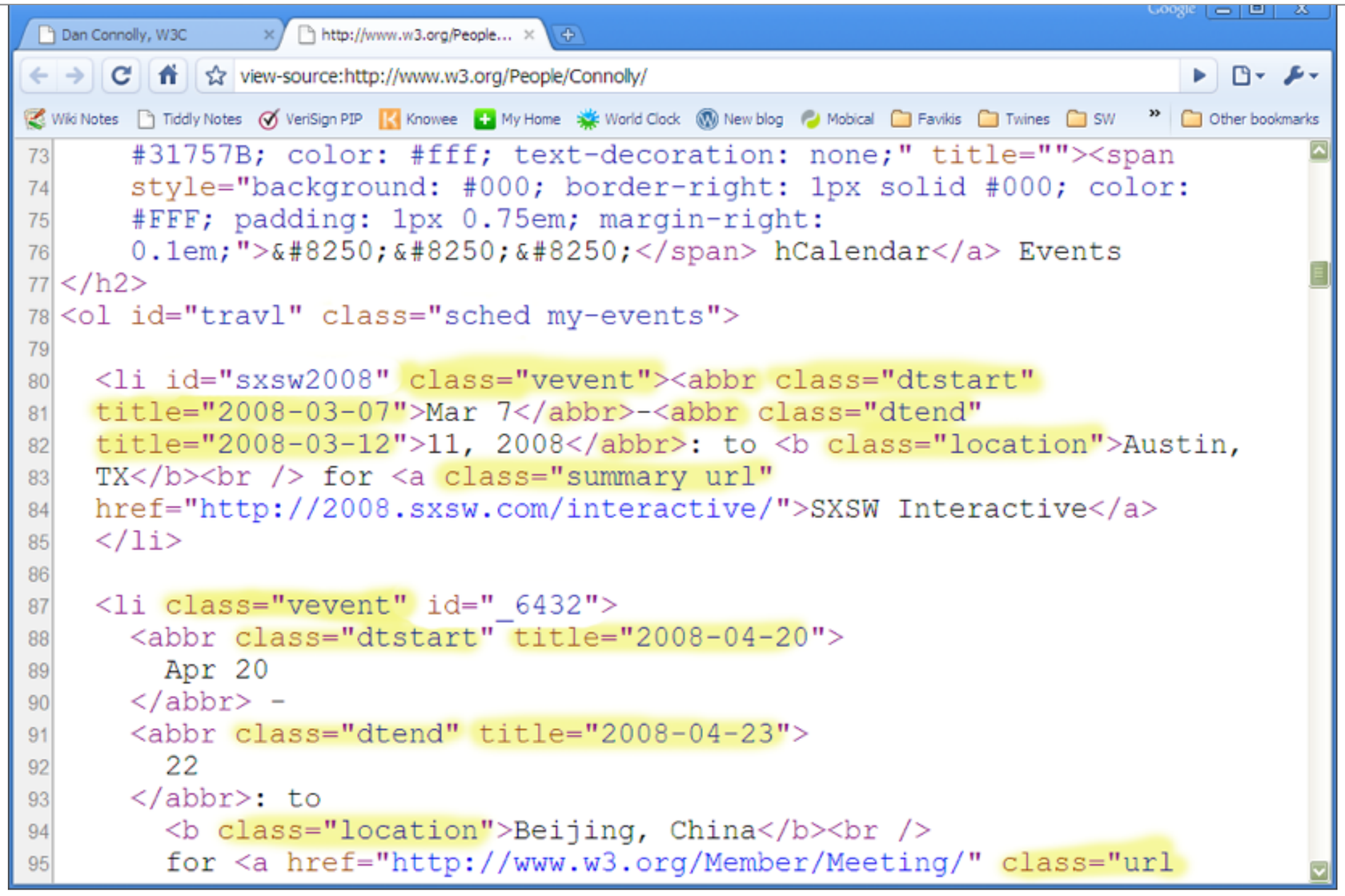
Microformats

- RDFa
- Yahoo SearchMonkey
- Google's Rich Snippets
- Semantic Media Wiki

# Microformat example: hCalendar

©Ivan Herman, W3C

# Behind the scenes…



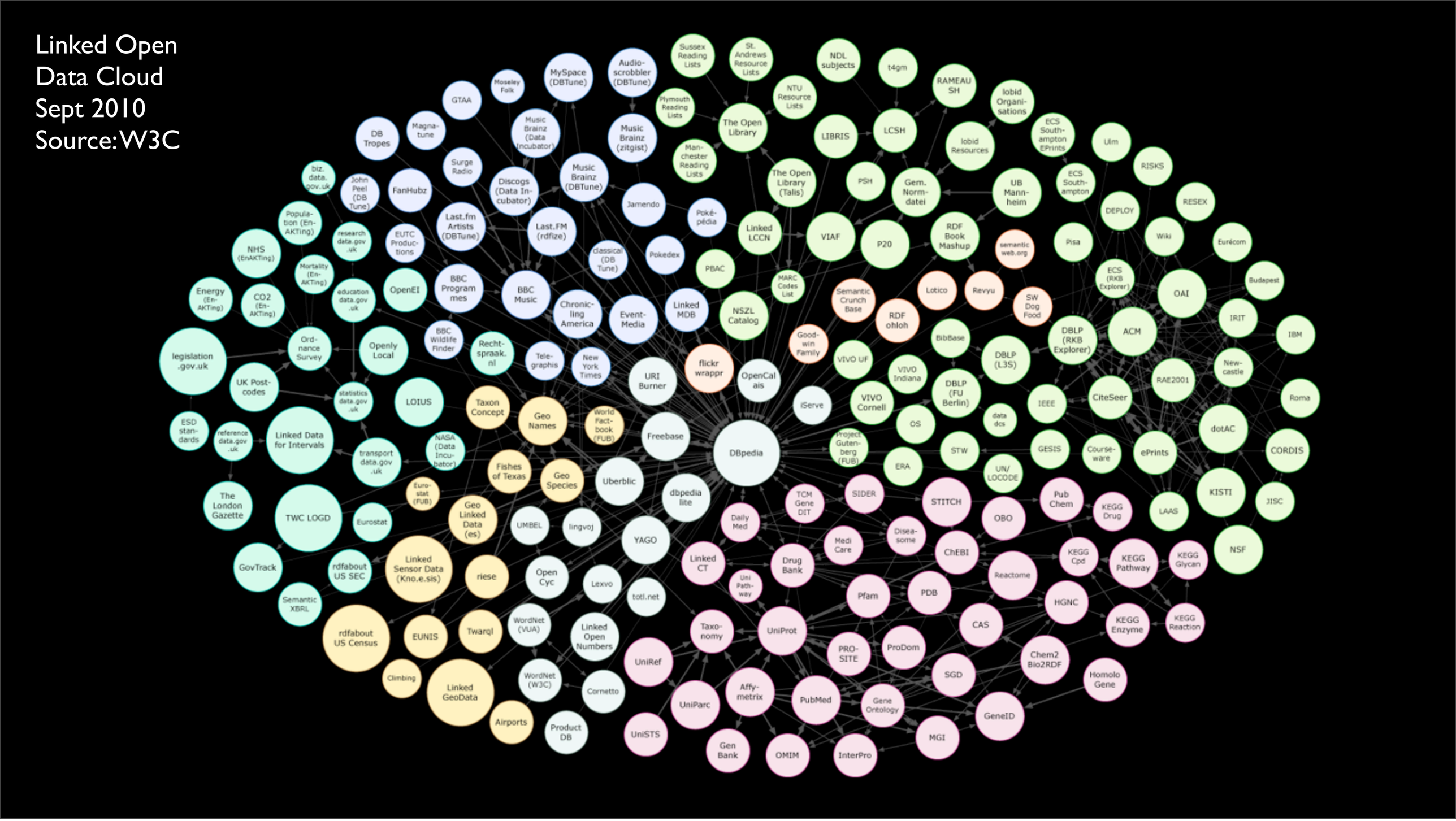

©Ivan Herman, W3C

# Tools

| | |
|---|---|
| Sesame | Parsing, storing, inferencing and querying over RDF data. API to various data stores. |
| Jena | Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine. |
| Redland | A set of free software C libraries that provide support for the Resource Description Framework (RDF). |
| Mulgara | An open source semantic data store written in Java. Provides SparQL and TQL support. |
| Virtuoso | A high performance object-relational SQL database which has embedded support for SparQL queries. |
| Simile project's Exhibit | An AJAX-based DHTML web program for displaying/sorting lightweight structured information. (e.g., information structured as JSON - JavaScript Simple Object Notation). |

Linked Open Data Cloud - March 2008
Source: W3C

Linked Open Data Cloud Sept 2010 Source: W3C



# Examples

Siri (now Google!)

Refiner (now Google!)

Invention Machine http://inventionmachine.com/Portals/56687/html/tour/tour.html?source=blog

MediaGlobe http://youtu.be/yTqSSkpzOO0

dbPedia

Parallax

SeMe4 http://apps.seme4.com/see-uk/#/bus-stops/by-population/ward/SO17%201BJ

VisualDataWeb http://www.visualdataweb.org/tools.php

http://www.youtube.com/watch?v=G0r1G-duPcQ&feature=related

WiGiPedia http://www.wigis.net/wigi/index.php/wigipedia

# WiGiPedia

- Editing the semantics in WikiPedia
- As a Graph!
- See at: http://www.youtube.com/watch?v=zevwssi6cLA

# Refiner

- Collect, merge, clean, refine and enhace
- Data from multiple sources
- Quickly and easily using desktop tool
- Similar interface to spreadsheet

# SeMe4's SeeUK

- Using open linked public data
- Integrating geo-spatial data with government information, statistics etc.
- To compare neighbourhoods and superimpose/integrate with maps
- Try it at: http://apps.seme4.com/see-uk/

# MediaGlobe

- Find semantic entities and relationships in live video
  - Including via optical character recognition
- Index video by person, place, topic
- Instantly jump to relevant video scenes

# Siri

A live application - location limited

Voice Recognition

Artificial Intelligence

Semantic Data Search, Integration and Inferencing

# References & Resources

Segaran, Evans & Taylor, Programming the Semantic Web, O'Reilly

WWW Consortium http://www.w3c.org especially a tutorial by Ivan Herman

Dean Allemang and Jim Hendler, Semantic Web for the Working Ontologist

Ray Kurzweil

HP Labs, Jena http://jena.sourceforge.net

Dave Beckett, Redland http://librdf.org

Mulgara, Semantic Store http://www.mulgara.org/

Semantic Web Company, Virtuoso http://www.semantic-web.at/

Simile/Exhibit http://www.simile-widgets.org/exhibit/

Google - Google Refine

Newsfeed: semanticweb.com

## Graham McLeod


- mcleod@iafrica.com
- www.inspired.org
- www.pro-mis.com
- http://grahammcleod.typepad.com/ (blog)