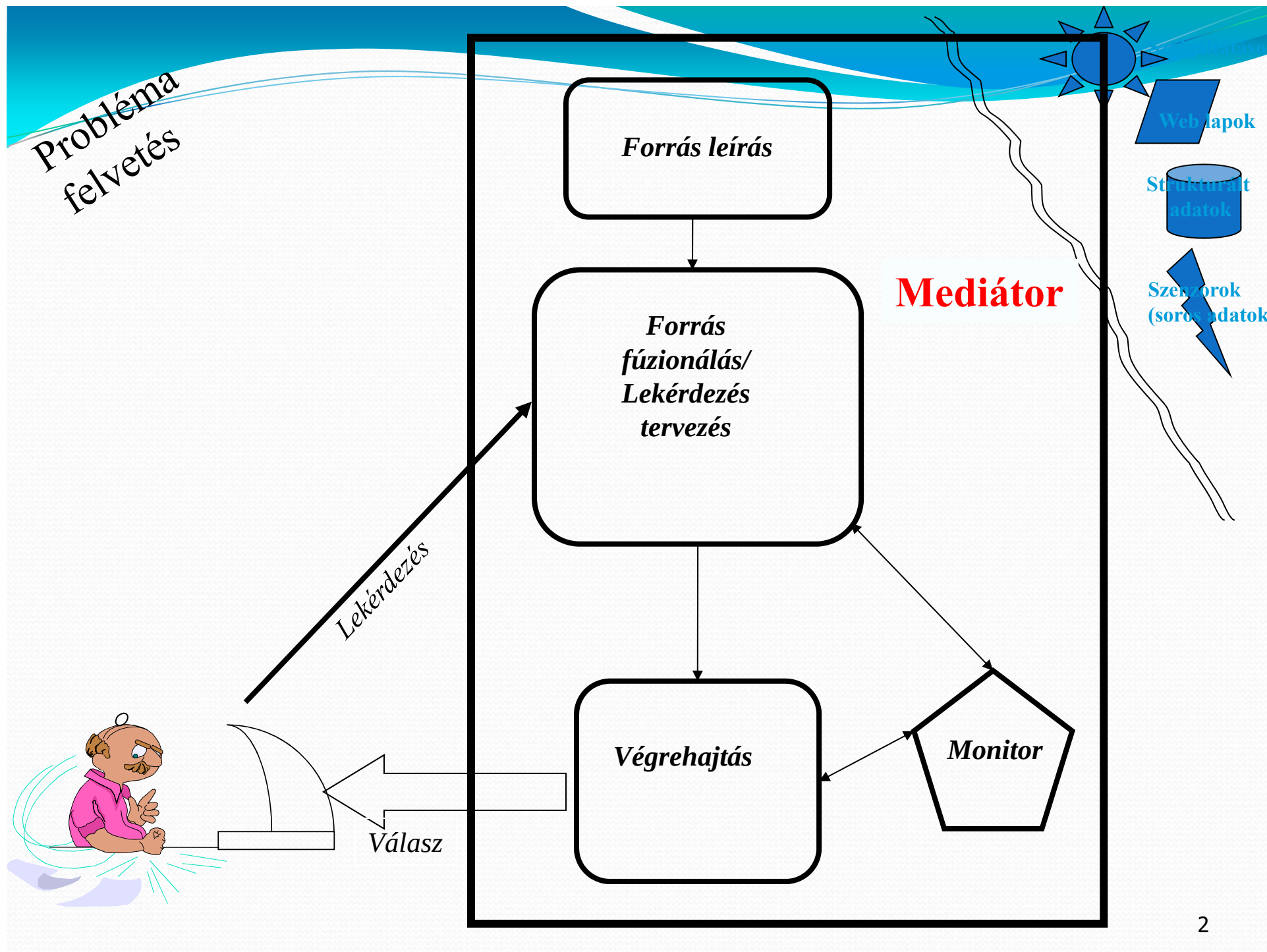


Integrációs és ellenőrzési technikák
VIMIAC04, 2019. tavasz

SZEMANTIKUS WEB

Méréstechnika és Információs Rendszerek Tanszék
<https://www.mit.bme.hu/oktatas/targyak/vimiaco4>



A web napjainkban

- Résztvevők dokumentumokat, adatokat publikálnak, URL címeket adva elérhetővé teszik az információkat
 - Kapcsolódás, hivatkozás linkekkel
- Integrációs próbálkozások: mashup oldalak (ad hoc)
 - Webszolgáltatások integrációja (eltérő API, logika, struktúra)
 - Adatgyűjtés kereső robotokkal (crawler programokkal)
 - Újra és újra felfedezzük a „kereket”
- Egészítsük ki a webet standard adatelérési módokkal: “Adatweb”

Az adatok webje

- Amire szükségünk van:
 - Publikáljuk adatainkat úgy, hogy felfedezhetőek legyenek a weben:
 - Standardizáljuk az adatok leírását, elérését
 - Dokumentumok eléréhez hasonló, de általános címezés: URI
 - Az URI-kal elérhető forrásokat kapcsoljuk
 - És engedjük a hálózat hatását érvényesülni, ahogy mi is böngészünk a weben...
- Példák: 2009 Semantic Technology Conference, San Jose, California, USA, June 15, 2009, Ivan Herman, W3C, ivan@w3.org

Weblapok értelmezése

- A felhasználók megértik, hogy a link egy munkahely honlapjára mutat
- Tudjuk értelmezni, hogy ez egy kutatóhely leírása
- *Ami hiányzik az „adatok webjének” építéséhez:*
 - *jelentéssel kiegészített, értelmezhető linkek*

Tehát bővítsünk:

- Adjunk a linkekhez kiegészítő információt, címkézzük fel őket
- A címkék legyenek géppel értelmezhetőek
 - Kategorizálás
 - Esetleg következtetés

Adatok hálój

- Az adatok webje:
 - Használjunk URI-kat adatok és (nemcsak) dokumentumok publikálására
 - Kapcsoljuk össze az adatokat
 - Jellemezzük/osztályozzuk a linkeket információk hozzáadásával
 - Használjunk standard technológiákat
- *Ez a szemantikus web alapja*



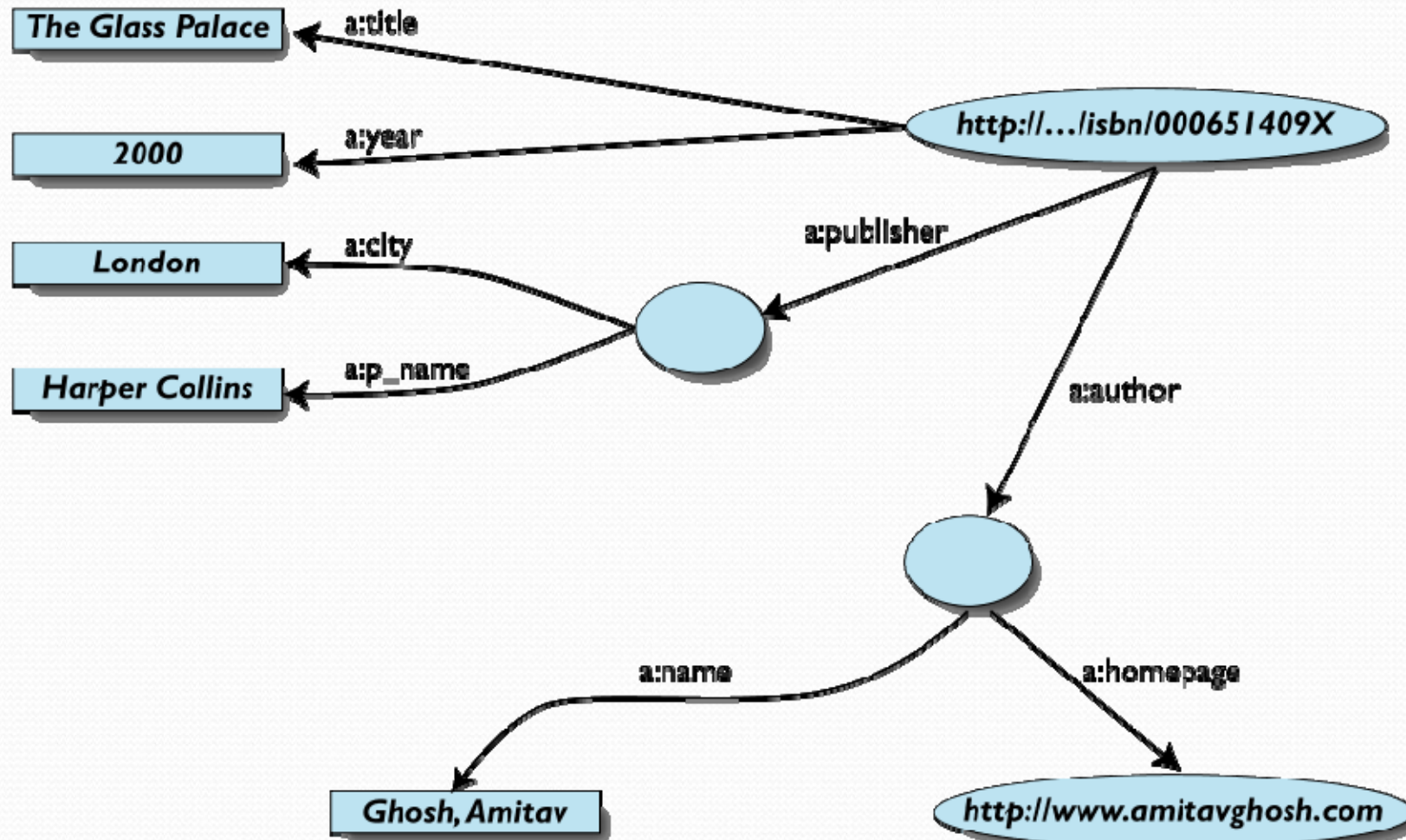
Példa: könyvesbolt adatai

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

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

ID	Publ. Name	City
id_qpr	Harper Collins	London

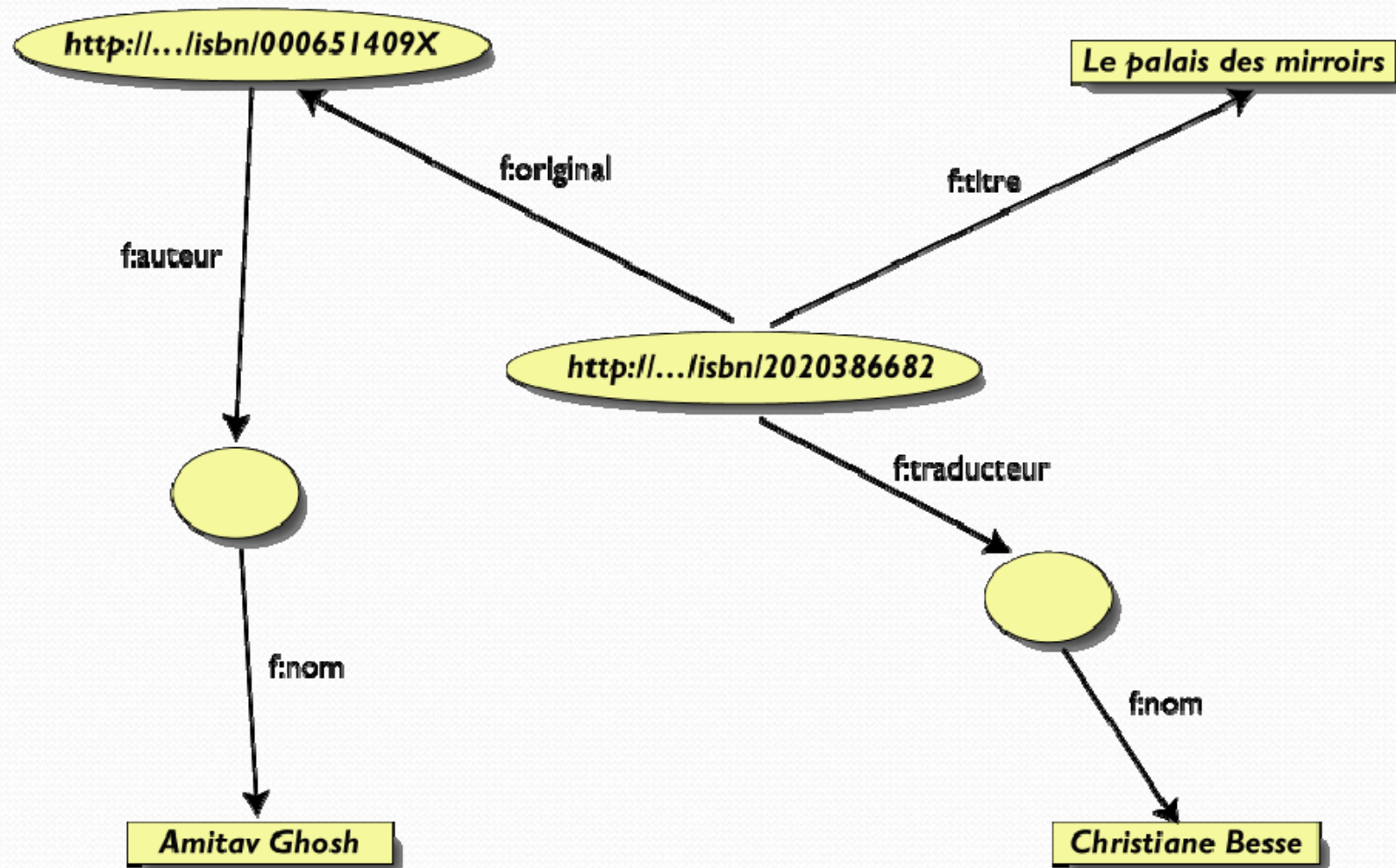
Adatok exportálása relációkként



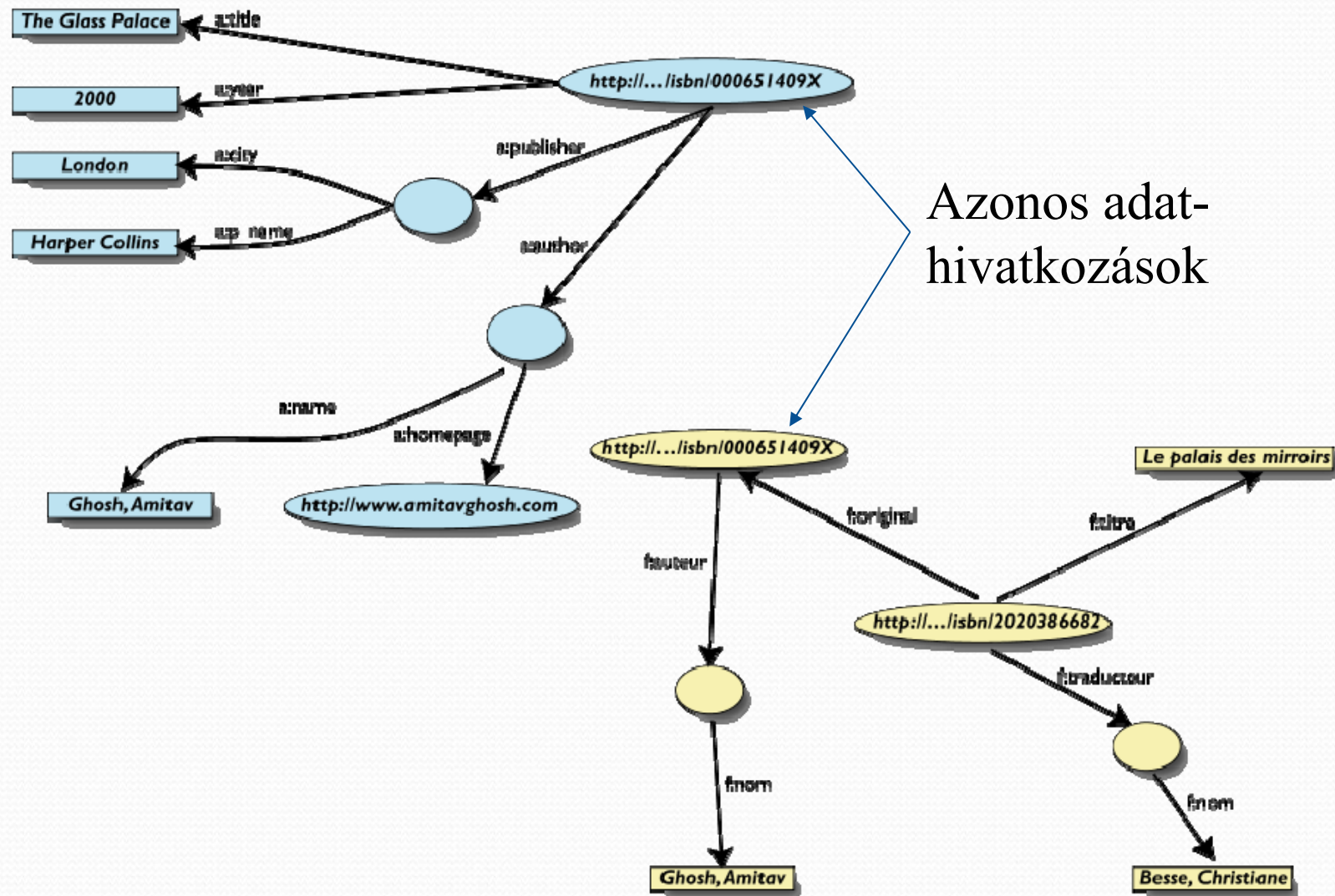
Egy másik könyvesbolt adatai

	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			

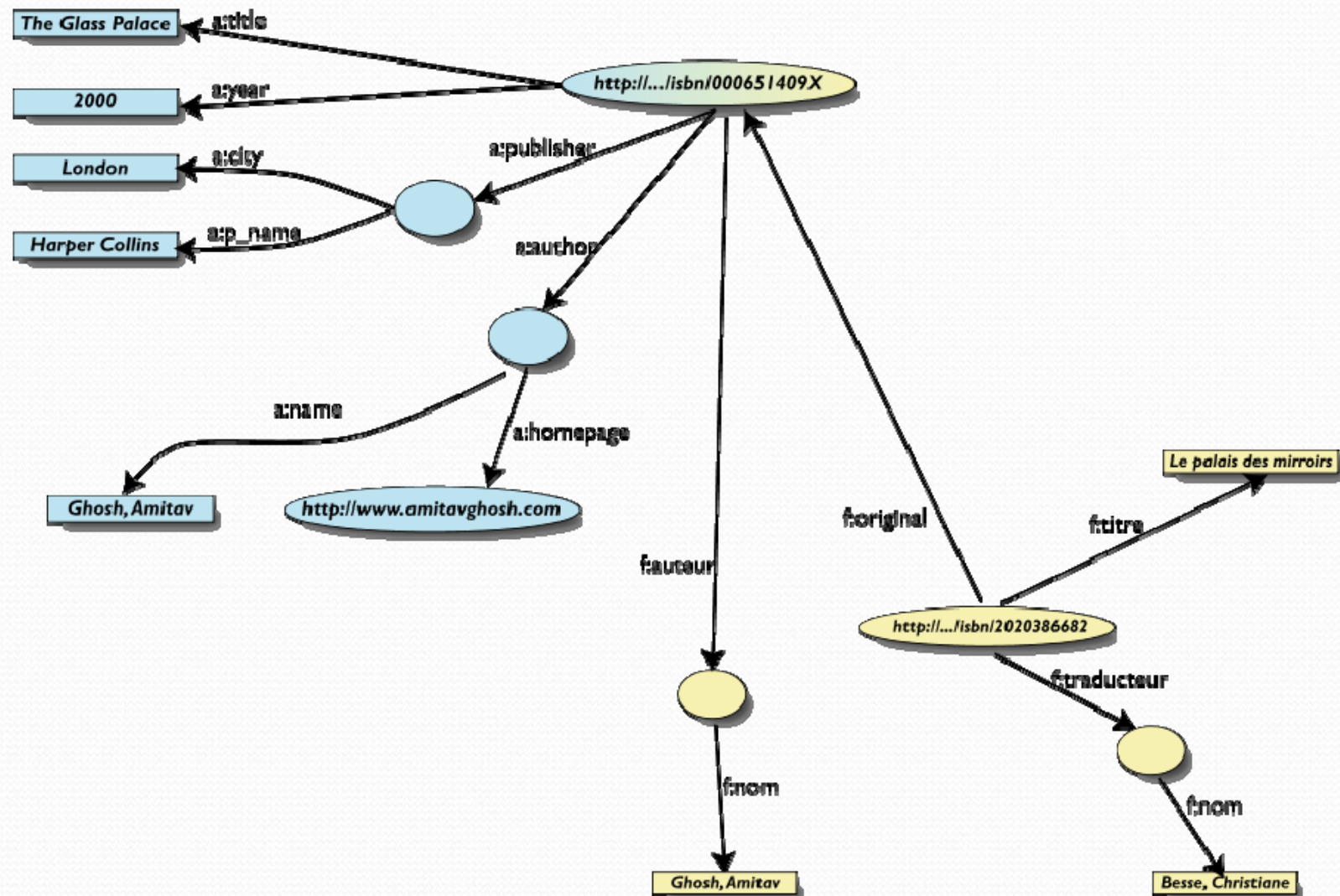
A második könyvesbolt adatainak exportja



Kapcsoljuk össze az adatokat

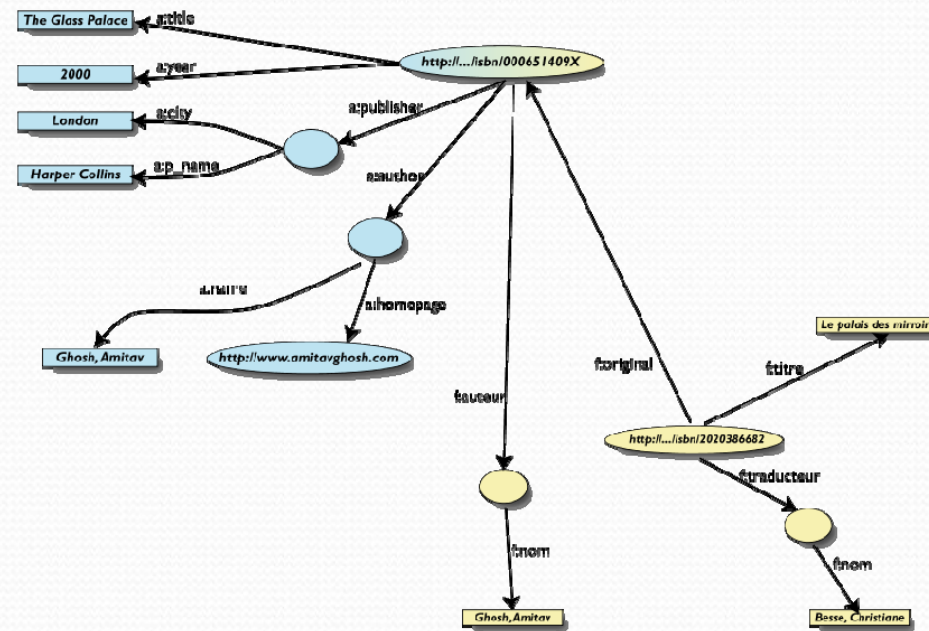


Amennyiben identikusak az elemek:



Írjunk lekérdezéseket a kapcsolatokon keresztül:

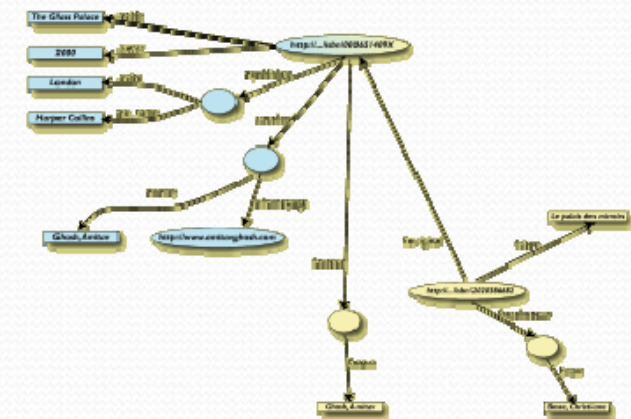
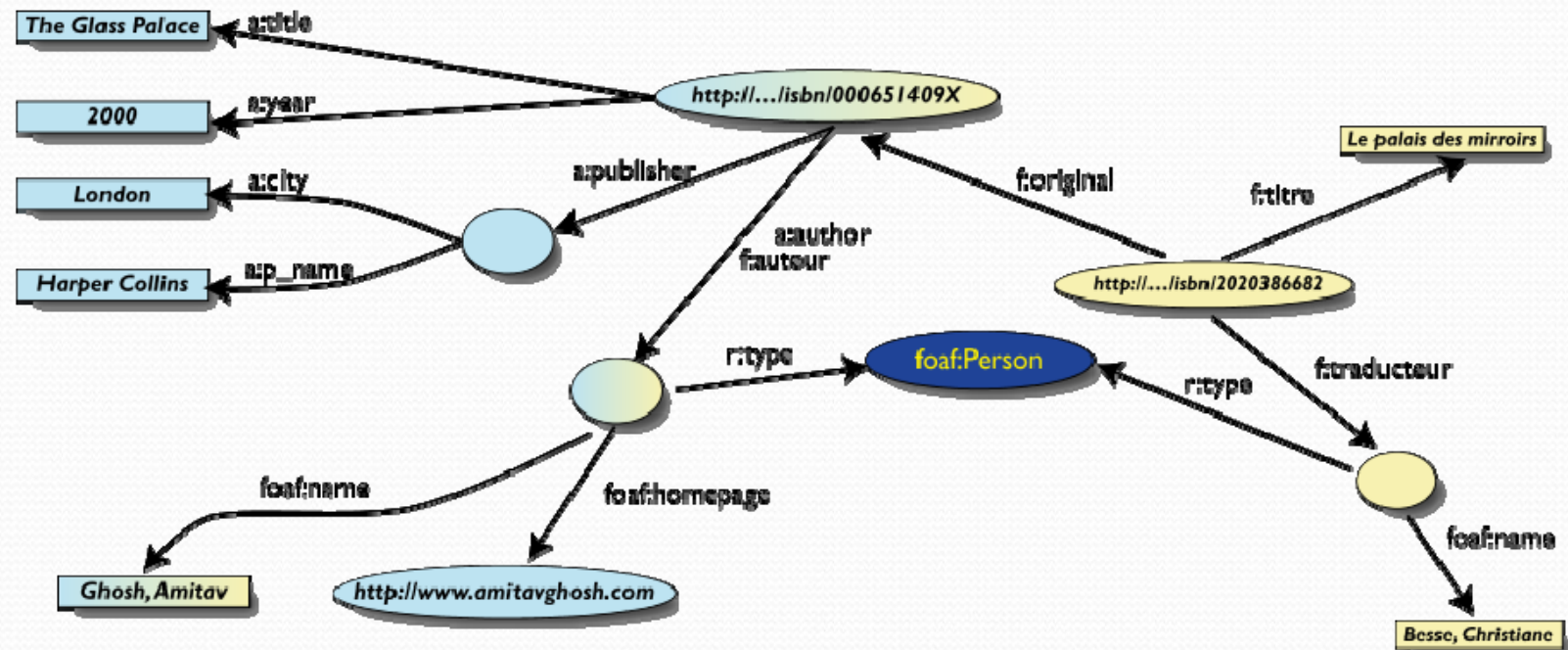
- Az első könyvesbolt adatai kiegészíthető például az eredeti könyvre vonatkozó információkkal



További kapcsolatok is felfedezhetőek...

- Vélhetően az **a:author** és az **f:auteur** azonos elemre mutat
- Automatikus összekapcsoláshoz: adjunk további információt a leíráshoz
 - **a:author** legyen azonos **f:auteur** erőforrással
 - Mindkettő személyt azonosít
 - Ilyen fogalmakat már a webes közösség definiált:
 - egy “Person” elem azonosítható a nevével és a honlapjával
 - Használjuk ezt kategóriaként

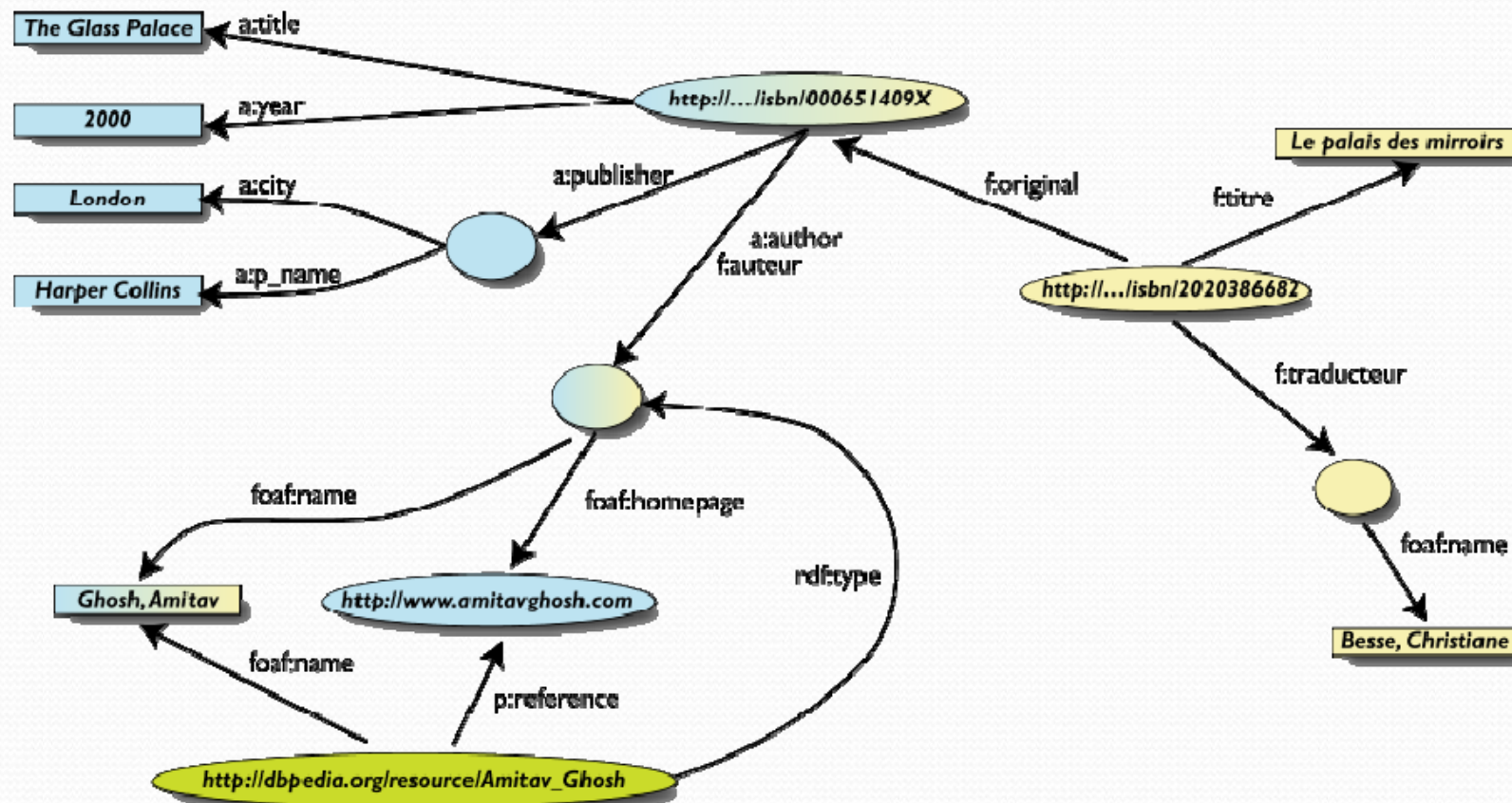
Adatháló kiegészíthető, lekérdezhető így:



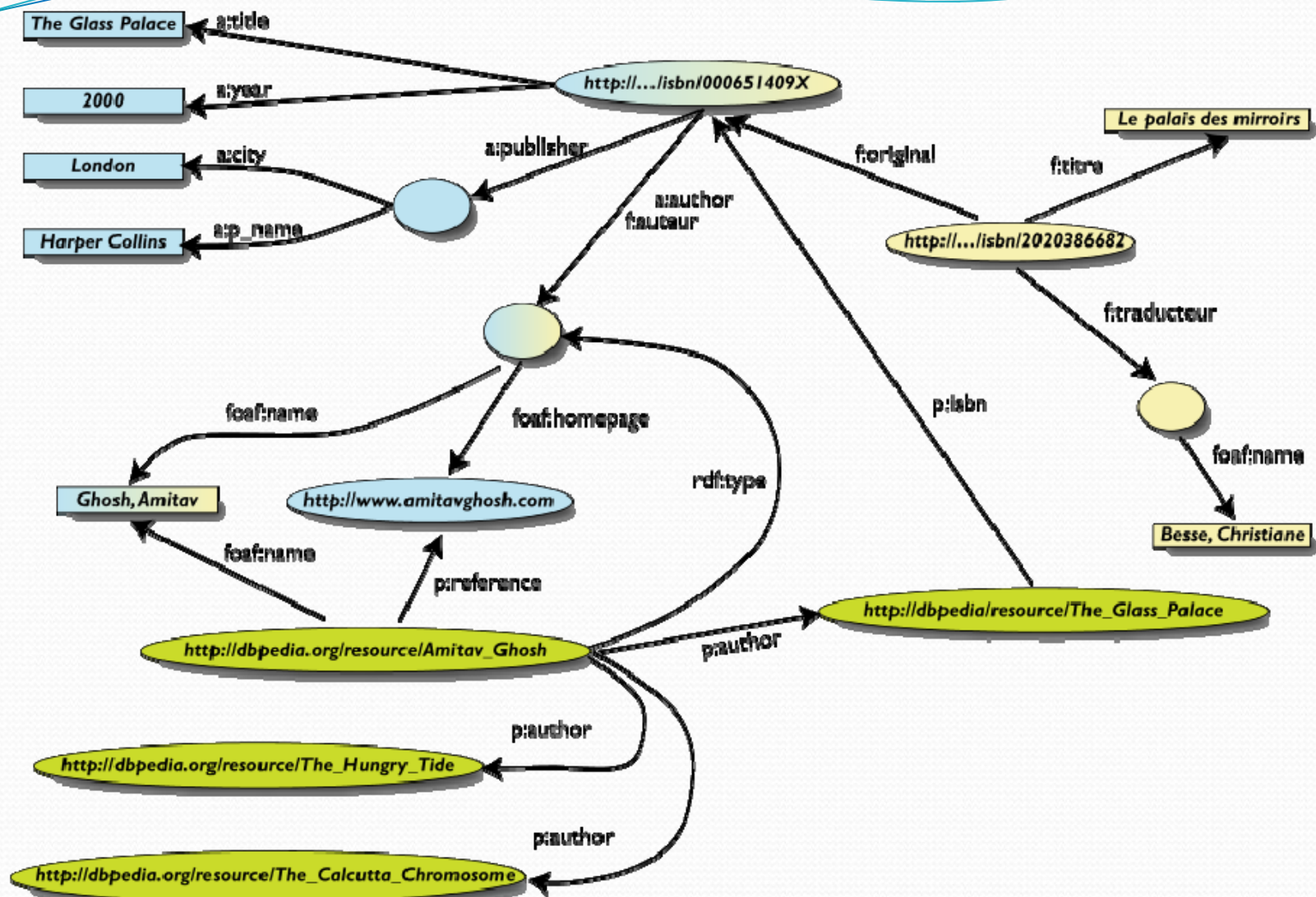
Bővítés újabb adathalmazokkal

- Például a “Person” típus esetén, használhatóak a Wikipédia adatai:
 - pl., a “[dbpedia](#)” projekt már feldolgozta a Wikipédia adatait

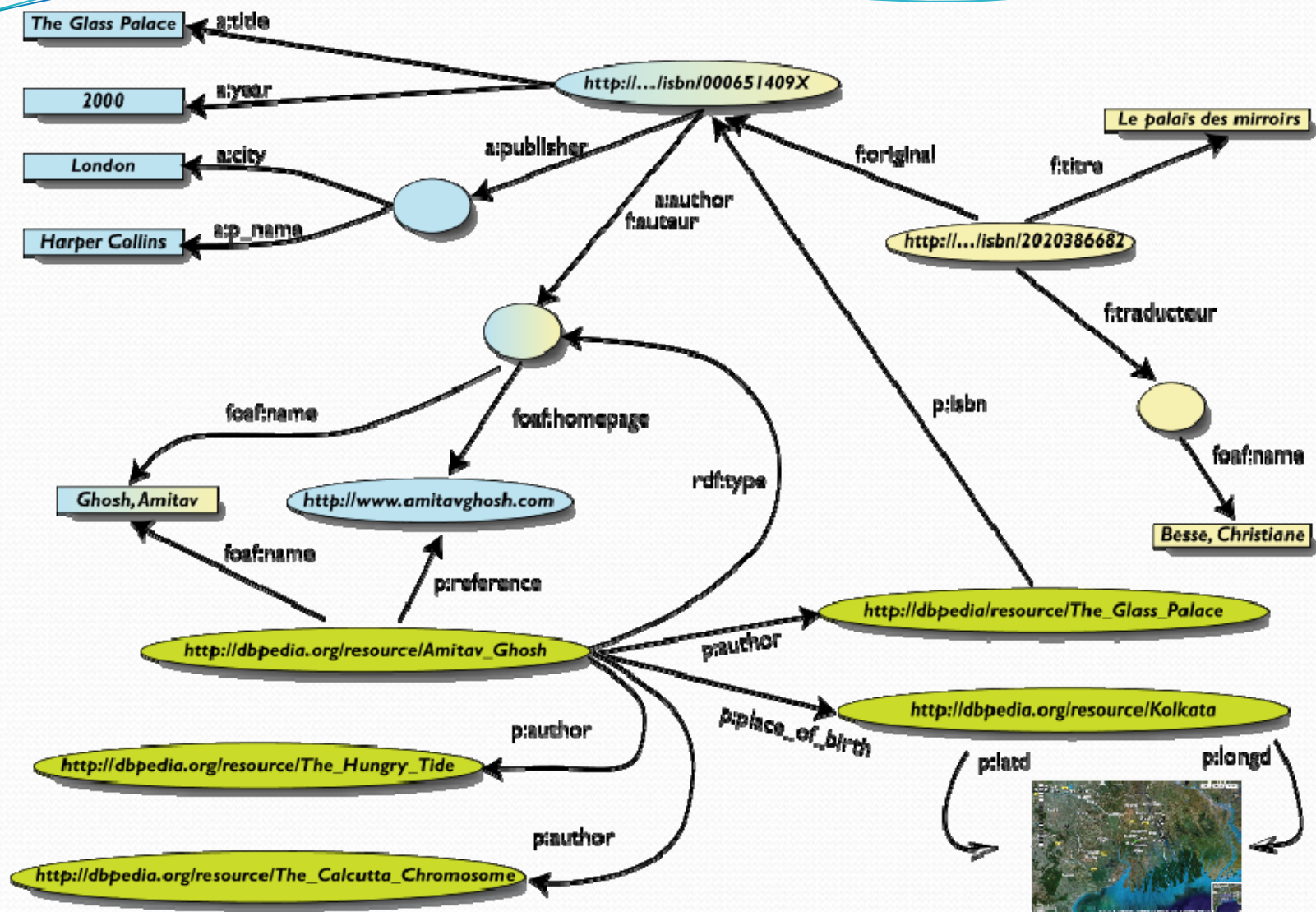
Összekapcsolás Wikipedia adatokkal



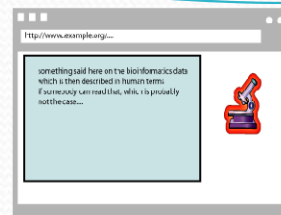
Összekapcsolás Wikipedia adatokkal



És így tovább...

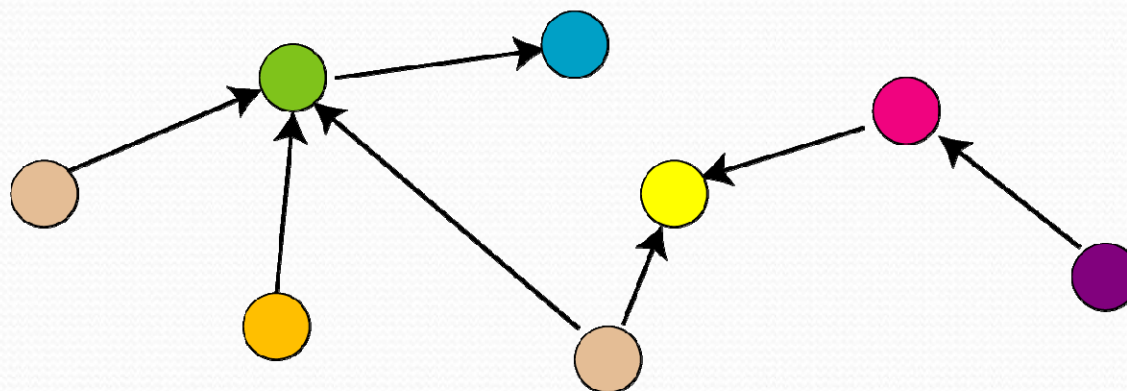


Az adatok webje kialakulófélben...



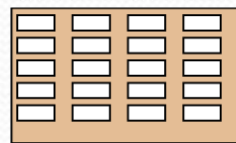
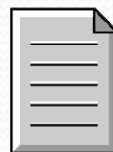
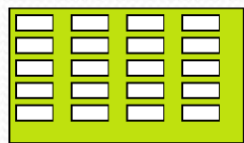
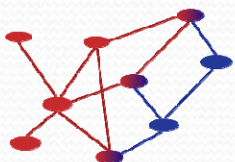
Alkalmazások

Lekérdezés,
adatmódosítás



Absztrakt adatstruktúra

Leképezés,



Forrás adatok különböző formátumokban

The background is a solid blue gradient, transitioning from a darker blue at the top to a lighter blue at the bottom. A series of white, wavy lines are positioned near the top edge, creating a sense of movement and depth.

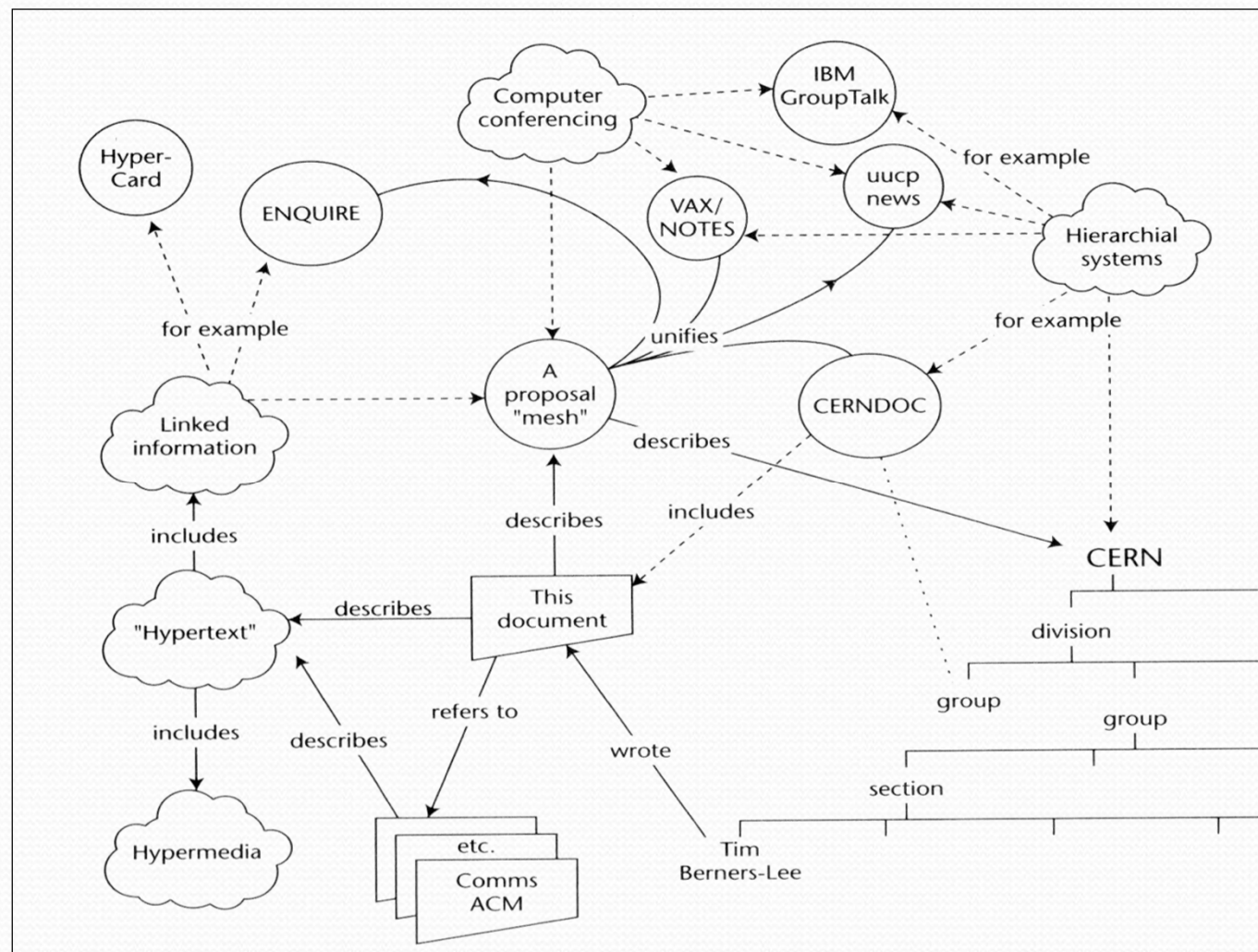
Szemantikus web

Szemantikus Web

- A szemantikus web alkalmas megközelítés, illetve megfelelő nyelvekkel, eszközökkel támogatja az intelligens információs rendszerek fejlesztését az elosztott információs környezetben.
- A SzW alapja a hagyományos web hálózat, így egyáltalán nem nyilvánvaló, hogy alkalmas a feladatra.
- A SzW technológia lehetőséget teremt az ágens alapú intelligens megoldások felhasználására a web területen.

A Szemantikus Web eredete

- Tim Berners-Lee eredeti 1989-es WWW javaslata a Web-et információ menedzselő funkciókkal ellátott objektumok kapcsolataiként jellemzi.



<http://www.w3.org/History/1989/proposal.html>

W3C szervezet célkitűzései

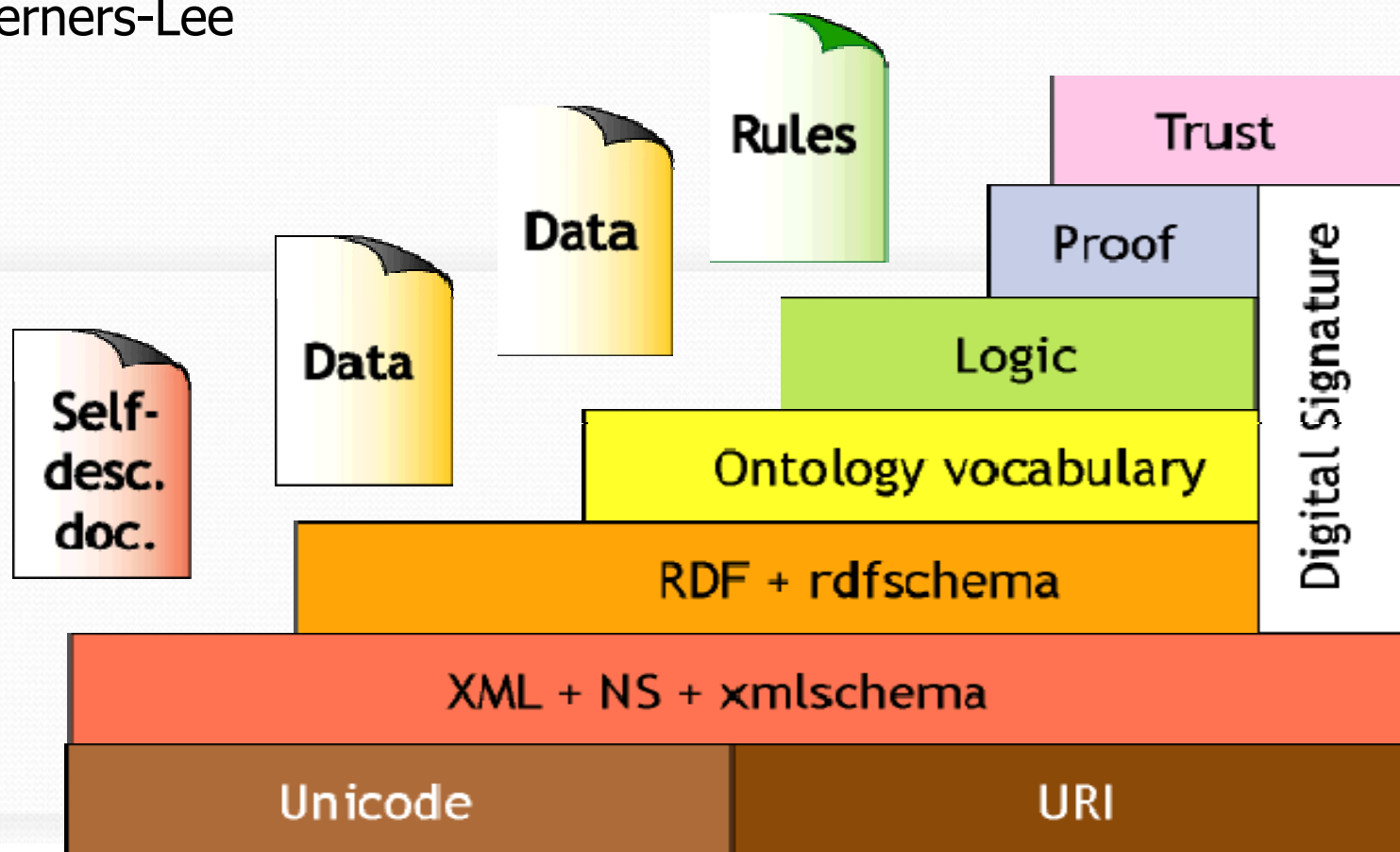
- Megközelítés – számítógépek jobb kihasználtságának biztosítása:

„A szematikus web egy kiterjesztése a jelenlegi web-nek, amelyben az információknak jól definiált jelentést adhatunk, lehetővé téve a gépek és felhasználók jobb együttműködését..” -- Berners-Lee, Hendler and Lassila, The Semantic Web, Scientific American, 2001

- A jelenlegi web *tárol* dolgokat, míg a szemantikus web képes *működtetni* dolgokat.

TBL szemantikus web felépítése

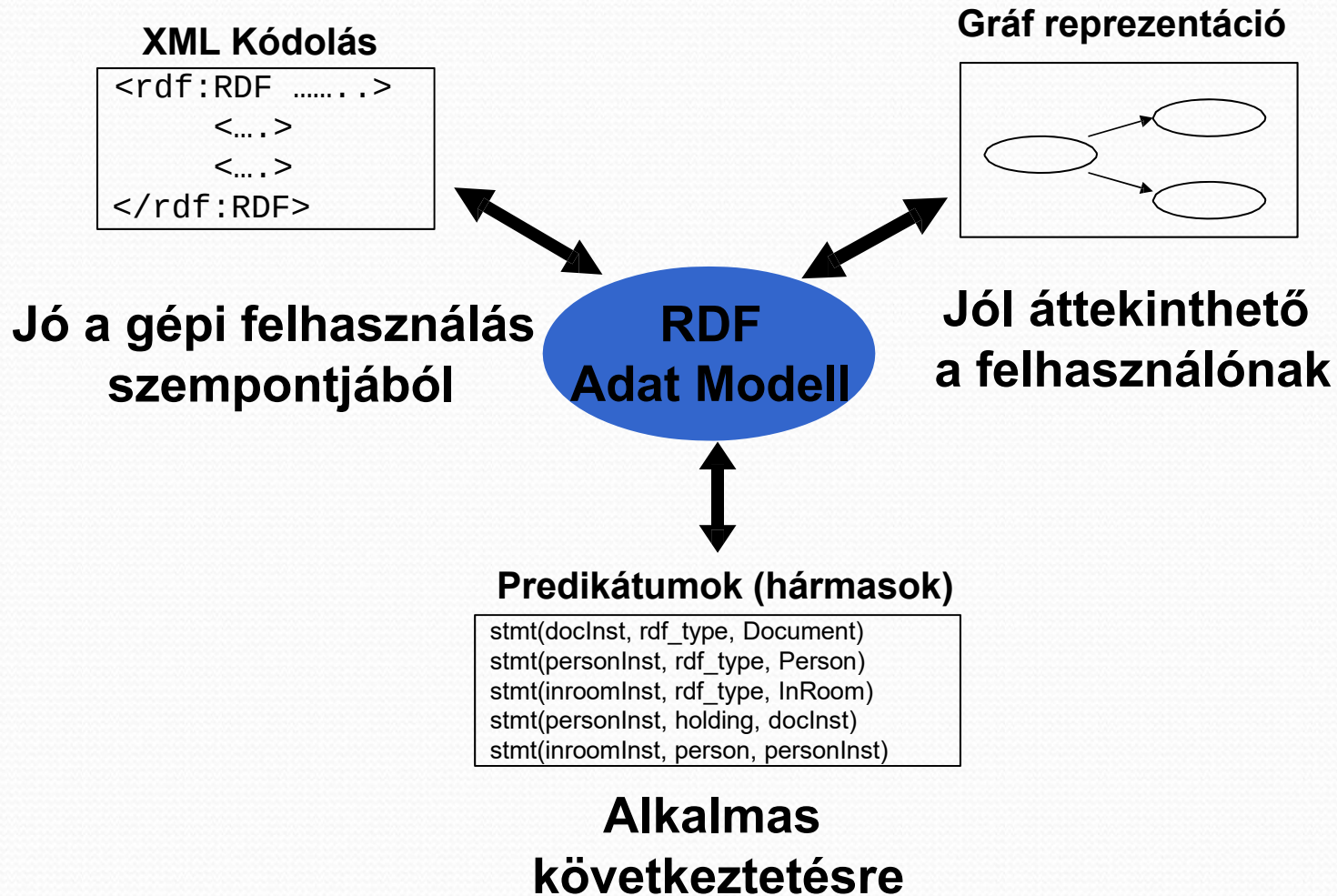
"A szemantikus web
elérhetővé teszi a tudást,
mint a web a hipertext-et --
Tim Berners-Lee



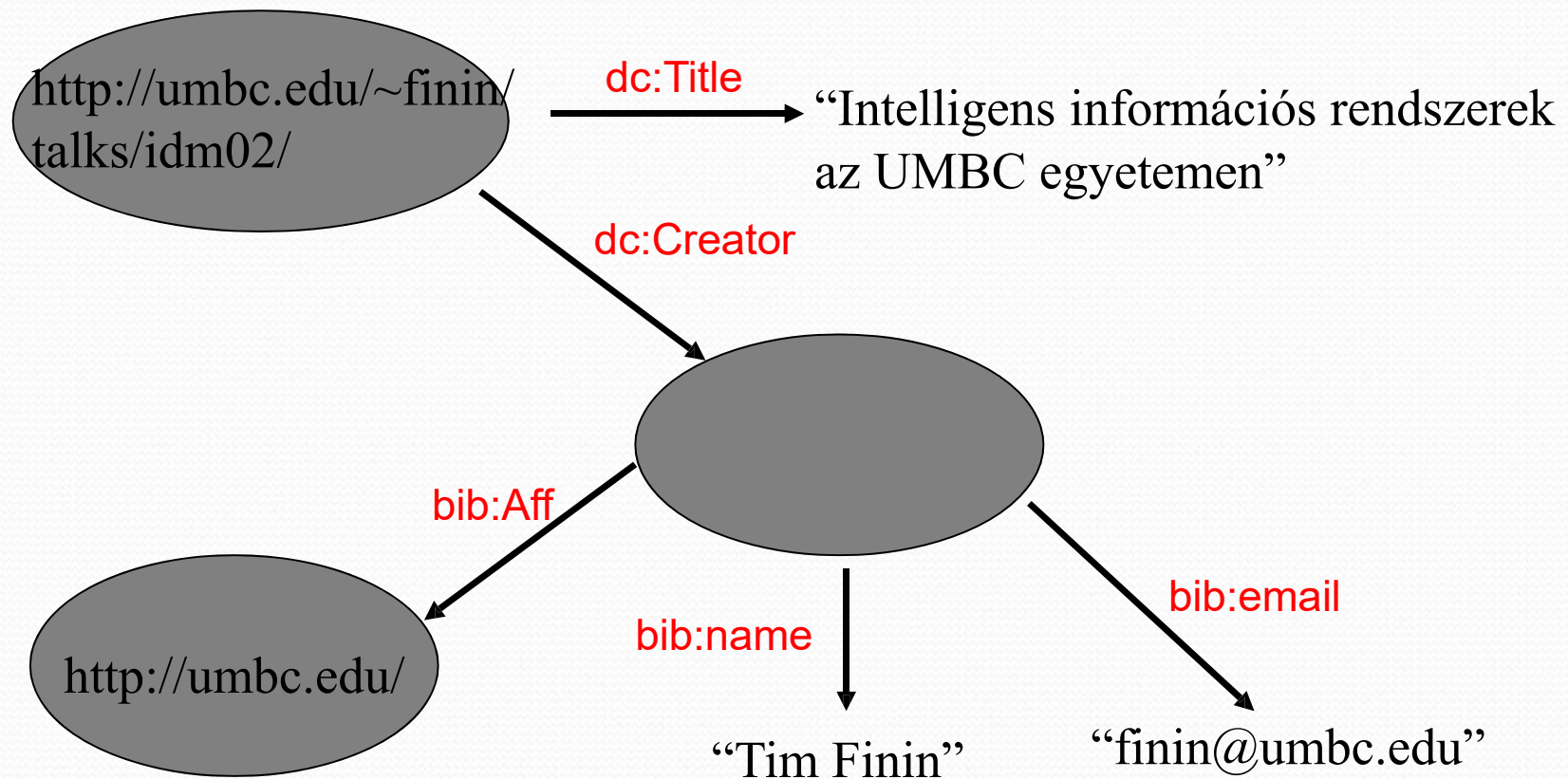
Dokumentumok

- **RDF Primer**
 - URI: <http://www.w3.org/TR/rdf-primer>
- **OWL Guide**
 - URI: <http://www.w3.org/TR/owl-guide/>
- **RDF Test Cases**
 - URI: <http://www.w3.org/TR/rdf-testcases/>
- **RDF: Concepts and Abstract Syntax**
 - URI: <http://www.w3.org/TR/rdf-concepts/>
- **RDF szemantika**
 - URI: <http://www.w3.org/TR/rdf-mt/>
 - Precíz, gráfokon alapuló szemantika
- **RDF/XML szintaxis**
 - URI: <http://www.w3.org/TR/rdf-syntax-grammar/>
- **RDF Vocabulary Description Language (RDF Schema)**
 - URI: <http://www.w3.org/TR/rdf-schema/>
- **Semantic Web/RDF Interest Group**
 - Vitafórum, alkalmazások
 - URI: <http://www.w3.org/RDF/Interest>
- **RDF Logic**
 - Nyilvános levelezési lista részletesebb szakmai vitákhoz
 - URI: <http://lists.w3.org/Archives/Public/www-rdf-logic/>
- **Annotation and Collaboration**
 - Nyilvános levelezési lista RDF-alapú annotációs rendszerekről
 - URI: <http://lists.w3.org/Archives/Public/www-annotation/>
- **W3C Semantic Web Home page**
 - URI: <http://www.w3.org/2001/sw/>

RDF - az első SzW nyelv



Egyszerű RDF példa



A példa XML szintaxissal

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bib="http://daml.umbc.edu/ontologies/bib/">
<description about="http://umbc.edu/~finin/talks/idm02/">
  <dc:title>Intelligent Information Systems on the Web and in the
Aether</dc:Title>
  <dc:creator>
    <description>
      <bib:Name>Tim Finin</bib:Name>
      <bib:Email>finin@umbc.edu</bib:Email>
      <bib:Aff resource=http://umbc.edu/>
    </description>
  </dc:Creator>
</description>
</rdf:RDF>
```


Hármassokat alkalmazó reprezentáció

- RDF kifejezések leírhatóak hármassokkal:
- <alany> <állítmány> <tárgy>

Megengedett szintaxis:
<URI><URI><URI>
<URI><URI><string>

RDF tervezési szempontok

- Egyszerű adatmodell
- Formális szemantika és egyszerű következtetési lehetőségek
- Bővíthető URI
- XML alapú szintaktika (is)
 - XML séma adattípusok
- Bárki megfogalmazhat állításokat az erőforrásokról

Alapelvek (1/2)

- Külön értelmezhetően definiálva
 - Modell struktúra (RDF gráf)
 - Interpretációs szemantika (vonzatok)
 - Szintaktikák (XML, TN, N3, ...)
- Mindössze két alap adattípus
 - URI/URIref: minden URI-val azonosított
 - Literálisok
 - String vagy más XSD adattípus

Alapelvek (2/2)

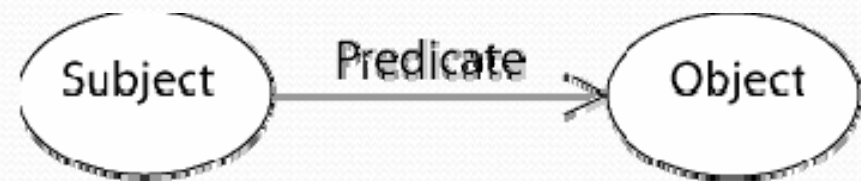
- Integrálható a webes információkkal
 - XML séma adattípusok
 - Referenciák http elérésű információkhoz
- Nyílt világ feltételezés
 - Bárki megfogalmazhat állításokat bármilyen erőforráshoz
 - Nem garantált a teljesség
 - Nem garantált a konzisztencia

Alapelemek

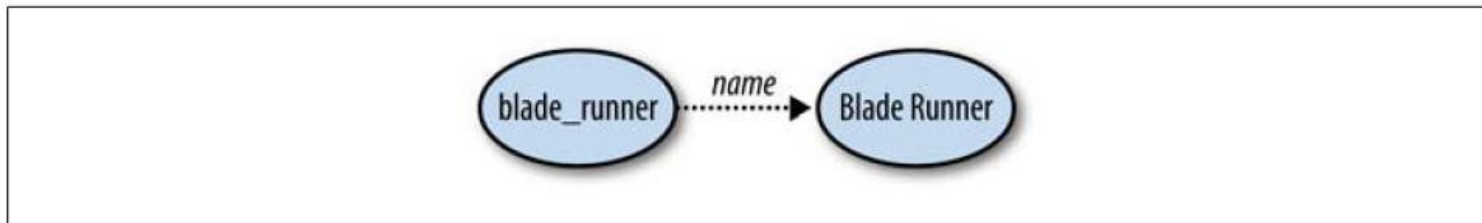
- Gráf adatmodell
- URI alapú szótárak
- Adattípusok
- Literálisok
- XML szerializációs szintaktika
- Egyszerű tények leírása
- Következtetés

Graph adatmodell

- Hármások: alany, állítmány, tárgy
- Kifejezések: hármások gyűjteménye



Példa



Terminológia és kényszerek

- Alany és tárgy a csomópontok
- Állítmány és tulajdonság szinonimák
- Különleges meg nem nevezett csomópontok: üres csomópontok
- Alany: URI referencia vagy üres csomópont
- Állítmány: URI referencia
- Tárgy: URI referencia, literális, üres csomópont



Információ a hármaskban



RDF

CompanyHomePage

PersonID	Homepage
FulvioCorno	http://www.polito.it/

Relációs adatbázis

**Első rendű logikai
predikátum**

```
HasCompanyHomePage(  
  'FulvioCorno',  
  'http://www.polito.it/');
```

Hármasok vs adatbázisok

FEATURE	RELATIONAL DATABASE	KNOWLEDGEBASE
Structure	Schema	Ontology statements
Data	Rows	Instance statements
Administration language	DDL	Ontology statements
Query language	SQL	SPARQL
Relationships	Foreign keys	Multidimensional
Logic	External of database/triggers	Formal logic statements
Uniqueness	Key for table	URI

RDF áttekintés

A formális modell alapelemei:

Két alaphalmaz: erőforrások (resources) és literálisok (literals)

Az erőforrások egy fontos részhalmaza: Tulajdonságok (properties).

Definiálunk egy hármassokból álló halmazt: Állítások (statements), amelyek formája: {alany, állítmány, tárgy}

, ahol az alany egy erőforrás,
az állítmány egy tulajdonság,
a tárgy vagy erőforrás vagy literális.

RDF adatmodell

- Erőforrások (Resources)
 - URI azonosítja
 - Kijelentés vonatkozik rá
- Tulajdonságok (Properties)
 - Erőforráshoz kapcsolt jellemző
 - A tulajdonság is erőforrás
- Literálok (Literals)
 - Karaktersorozatok
- Kijelentések (Statements)
 - Alany (subject); erőforrás
 - Állítmány (predicate); tulajdonság
 - Tárgy (object); erőforrás vagy literál

Ezek segítségével megfogalmazhatunk leírásokat:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description about="http://www.w3.org/Home/Lassila">
    <s:Creator>
      <rdf:Description about="http://www.w3.org/staffId/85740">
        <rdf:type resource="http://description.org/schema/Person"/>
        <v:Name>Ora Lassila</v:Name>
        <v:Email>lassila@w3.org</v:Email>
      </rdf:Description>
    </s:Creator>
  </rdf:Description>
</rdf:RDF>
```

Az RDF szerepe az SZV hierarchiában

- Technológiát és módszert ad ahhoz, hogy dokumentumainkhoz jelentést rendeljünk egy jól olvasható formában
- Jó lehetőség, de
 - a szemantikai információk nem túl hasznosak, amíg strukturálatlan és nem tudjuk konzisztens módon értelmezni.

(XML séma kevés: csak szintaktikáról szól, nem ad lehetőséget a dokumentumon kívüli dolgok leírására)

Séma hiányában ugyanannak a tartalomnak sokféle reprezentációja lehetséges:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
    syntax-ns#" />
  <rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <Creator>
      <rdf:Description
        about="http://www.w3.org/staffId/85740">
        <rdf:type
          resource="http://desc.org/schema/Person"/>
        <Name>Ora Lassila</Name>
        <Email>lassila@w3.org<Email>
      </rdf:Description>
    </Creator>
  </rdf:Description>
</rdf:RDF>
```

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
    syntax-ns#" />
  <rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <author>
      <rdf:Description
        about="http://www.w3.org/staffId/85740">
        <rdf:type
          resource="http://desc.org/schema/Person"/>
        <name>
          <surname>Lassila</surname>
          <given>Ora</given>
        </name>
        <email>lassila@w3.org</email>
      </rdf:Description>
    </author>
  </rdf:Description>
</rdf:RDF>
```

SPARQL

- Lekérdező nyelv RDF-hez
- SPARQL gráf illesztésen alapuló lekérdező nyelv
- Gráf minták

- példa:

```
<http://example.org/book/book1>  
<http://purl.org/dc/elements/1.1/title>  
  ?title .
```

- ?title - változó.

Egy egyszerű SPARQL Query

- **Adat:**

```
<http://example.org/book/book1>  
  <http://purl.org/dc/elements/1.1/title>  
  "SPARQL Tutorial".
```

- **Lekérdezés:**

```
SELECT ?title  
WHERE { <http://example.org/book/book1>  
        <http://purl.org/dc/elements/1.1/title>  
        ?title . }
```

- **Eredmény:**

title
"SPARQL Tutorial"

További példa

- **Adat:**

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
```

```
_:a foaf:name "Johnny Lee Outlaw" .
```

```
_:a foaf:mbox <mailto:jlow@example.com> .
```

```
_:b foaf:name "Peter Goodguy" .
```

```
_:b foaf:mbox <mailto:peter@example.org> .
```

```
_:c foaf:mbox <mailto:carol@example.org> .
```


- **Lekérdezés:**

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?name ?mbox
```

```
WHERE { ?x foaf:name ?name . ?x foaf:mbox ?mbox }
```

- **Eredmény:**

name	mbox
"Peter Goodguy"	<mailto:peter@example.org>
"Johnny Lee Outlaw"	<mailto:jlow@example.com>

Lekérdezésel RDF literálisokkal

- **Példa RDF adatokra**

```
@prefix dt: <http://example.org/datatype#> .  
@prefix ns: <http://example.org/ns#> .  
@prefix : <http://example.org/ns#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
  
:x ns:p "cat"@en .  
:y ns:p "42"^^xsd:integer .  
:z ns:p "abc"^^dt:specialDatatype .
```


RDF Literálisok illesztése

- Lekérdezés 1:

```
SELECT ?v WHERE { ?v ?p "cat" }
```

Lekérdezés 2:

```
SELECT ?v WHERE { ?v ?p "cat"@en }
```

eltérő eredményt ad.

- Csak a második találja meg az előző példában az eredményt:

v
<http://example.org/ns#x>

Üres csomópontok a lekérdezésekben

- **Adat:**

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
```

```
_:a foaf:name "Alice" .
```

```
_:b foaf:name "Bob" .
```

- **Lekérdezés:**

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?x ?name
```

```
WHERE { ?x foaf:name ?name . }
```

- **Eredmény:**

x	name
_:c	"Alice"
_:d	"Bob"

Üres csomópontok a lekérdezésekben (modell bővítés)

- **Adat:**

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
```

```
_:a foaf:name "Alice" .
```

```
_:b foaf:name "Bob" .
```

```
_:a foaf:knows _:b .
```

```
_:b foaf:knows _:a .
```

- **Lekérdezés:**

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
```

```
SELECT ?x ?name1 ?y ?name2
```

```
WHERE { ?x foaf:name ?name1 . ?y foaf:name ?name2 .  
        ?x foaf:knows ?y }
```

- **Eredmény:**

?x	name1	?y	name2
_:c	"Alice"	_:d	"Bob"
_:d	"Bob"	_:c	"Alice"

RDF esettanulmányok

- Dublin Core
 - Magas szintű szótár definiálása
 - Elektronikus dokumentumok megtalálása
- Open Directory Project (OPD)
 - Webes katalógus keresők számára
- MusicBrainz
 - Hanganyagok (cd, mp3 ...) metaadatainak lekérésére
- RSS: RDF Site Summary
 - Hírek, események közzététele
- Wordnet
 - Szabadon letölthető szótár
 - Nem csak címszavakat, hanem kapcsoltaokat is leír