



시맨틱 웹과 링크의 진화

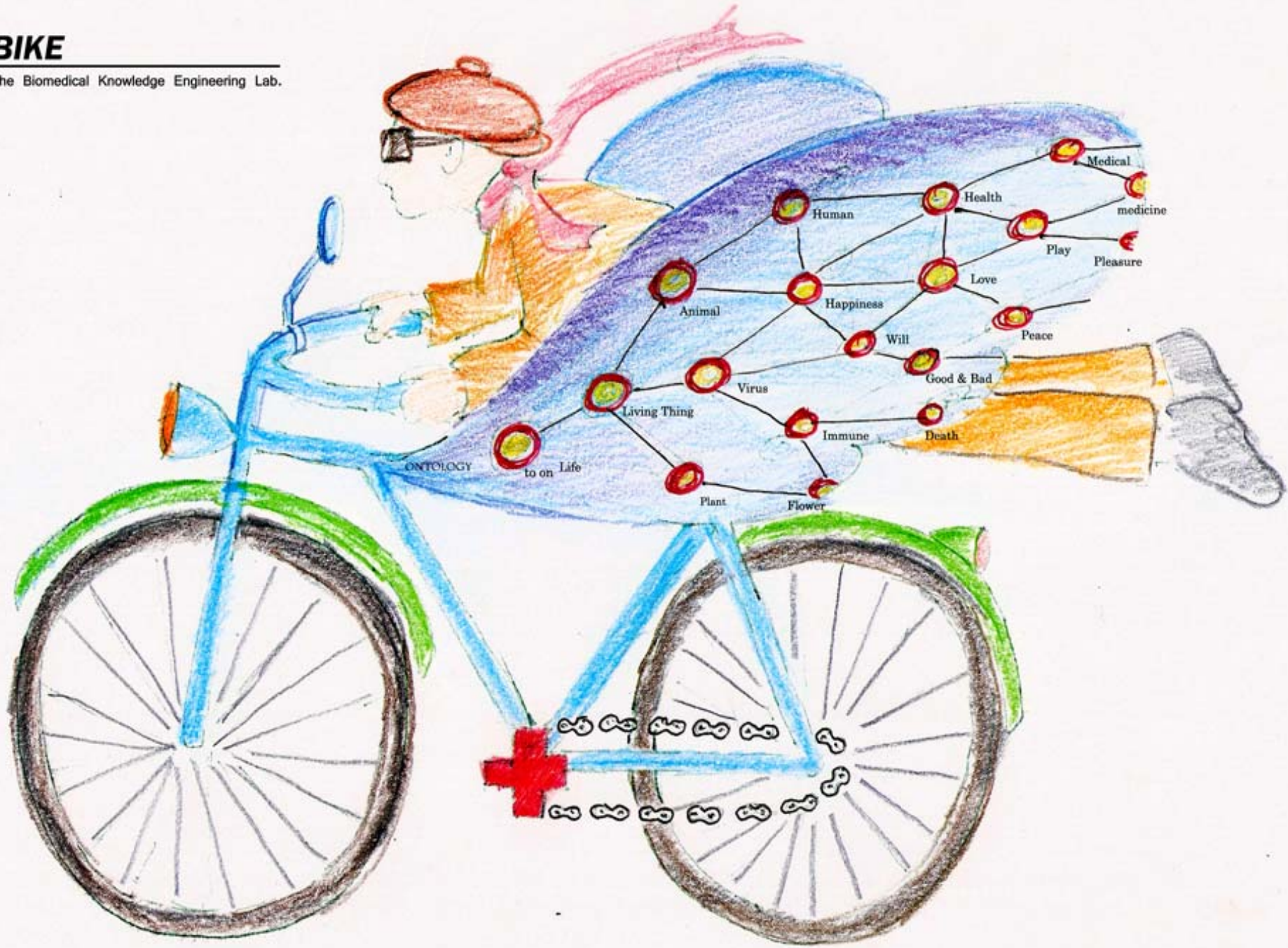
김 홍 기

서울대학교
의생명지식공학연구실



BIKE

The Biomedical Knowledge Engineering Lab.



웹의 진화



Social web

- 메쉬업
- 커뮤니티 포털
- **RSS, Weblogs, Wikis, Folksonomy**
- Semi-structured data
- **URI** 중심
- **Social networks**
- 2000 ~ 2010

The web

- 웹 검색 엔진
- 엔터프라이즈 포털
- **HTML, 웹 사이트**
- Unstructured data
- Document 중심
- “Push” Publish & Subscribe
- 1990 ~ 2000

Web 1.0

Semantic web

- 의미적 질의
- 개인화 서비스
- **RDF, OWL, Semantic Desktop**
- **온톨로지, Structured data**
- **Data** 중심
- **Data & Social Links**
- 2005 ~ 2020

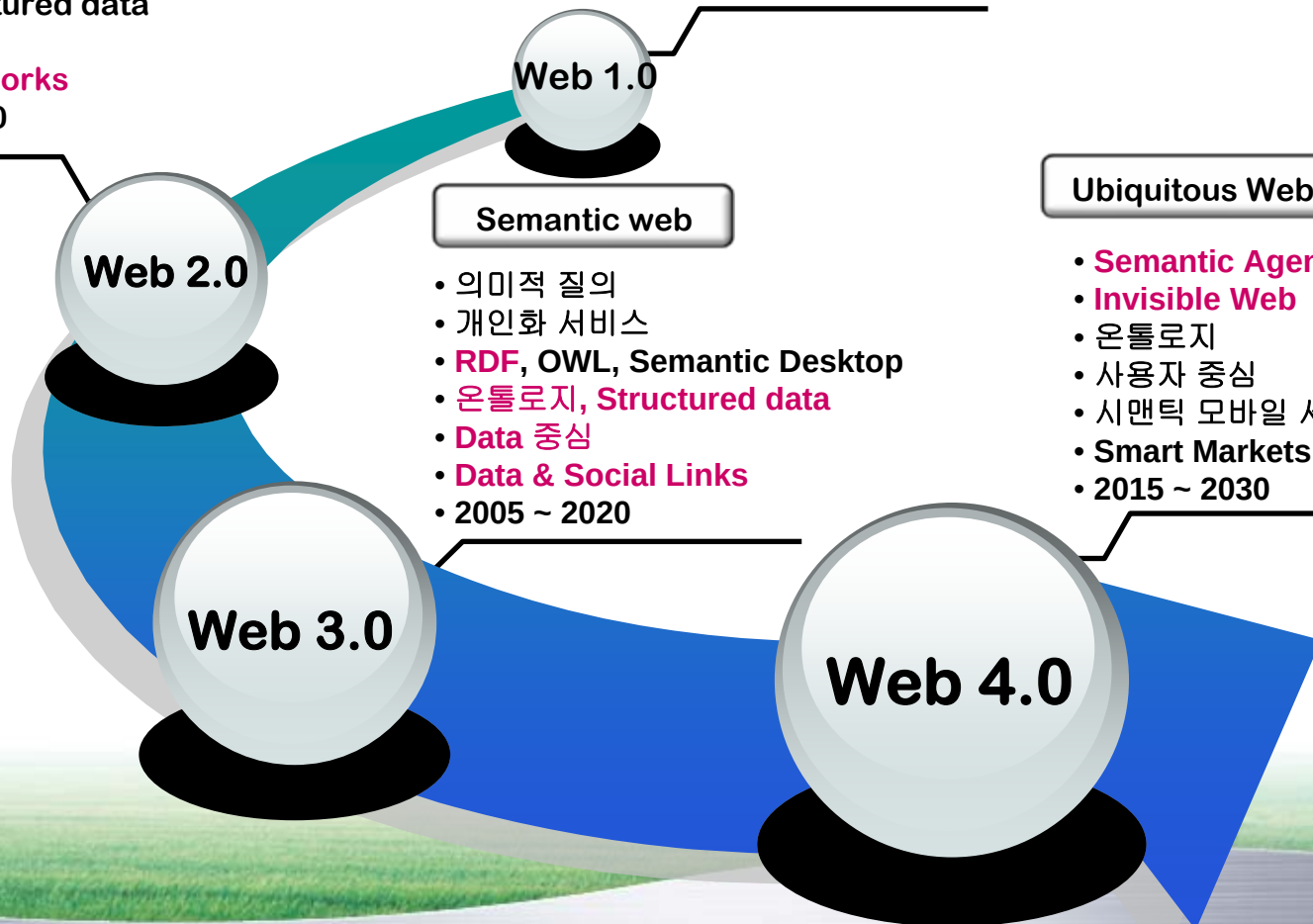
Ubiquitous Web

- **Semantic Agent Ecosystems**
- **Invisible Web**
- 온톨로지
- 사용자 중심
- 시맨틱 모바일 서비스
- Smart Markets
- 2015 ~ 2030

Web 2.0

Web 3.0

Web 4.0



검색: Document centered vs. Data centered?



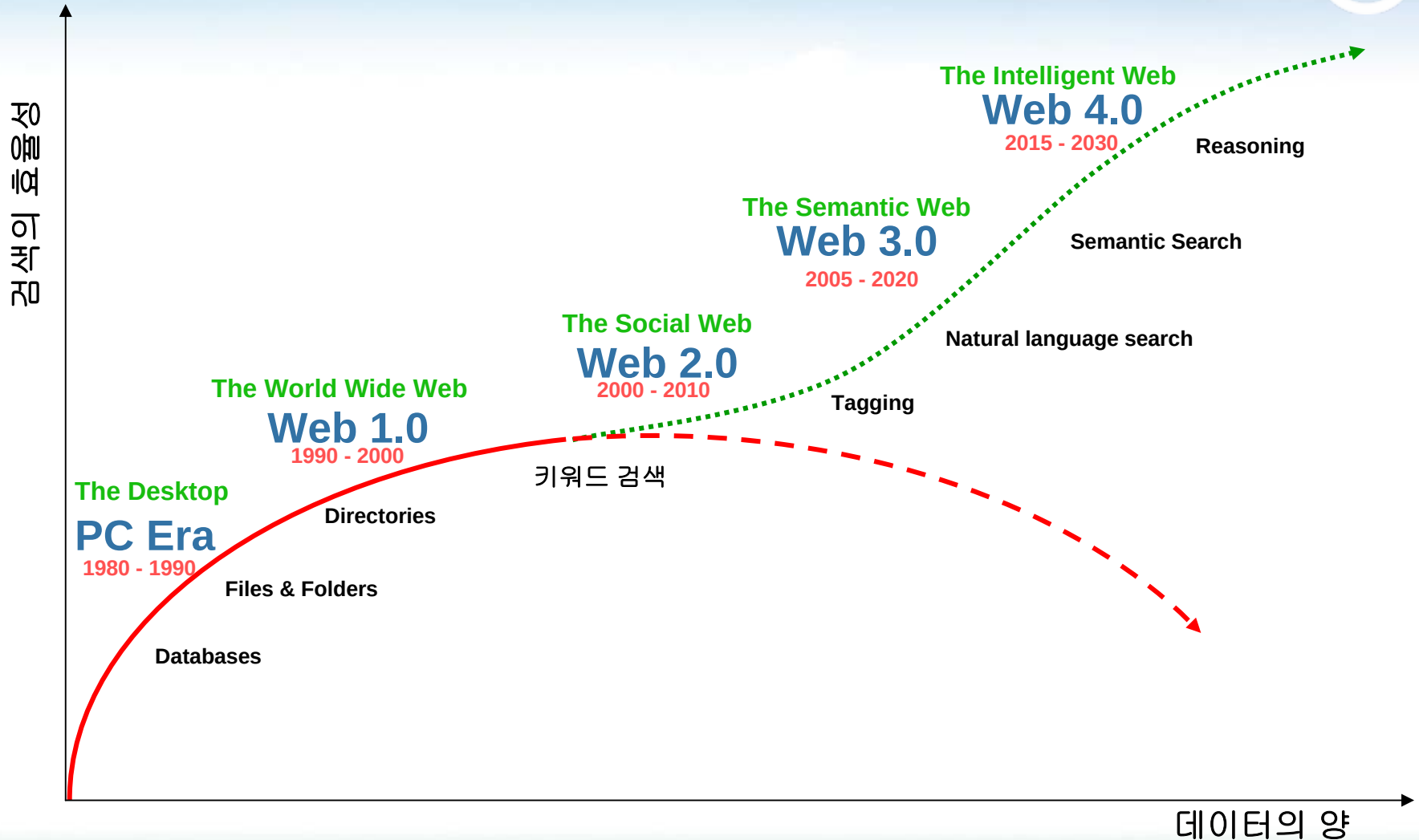
Unclassified &
Unstructured
Documents (웹1.0)



Semantically
classified and
related data (웹3.0)



검색의 진화: 키워드 검색을 넘어서



출처: <http://www.mindingtheplanet.net>

링크의 진화



Document Web

- Document-centric
- Document resources
- Unstructured data
- HTML, CSS
- URL-centric
- Web 1.0



Semi-Structured Web

- Data-centric
- Web database
- semi-structured data
- XML, JSON
- URI-centric
- Web 1.5



Social Web

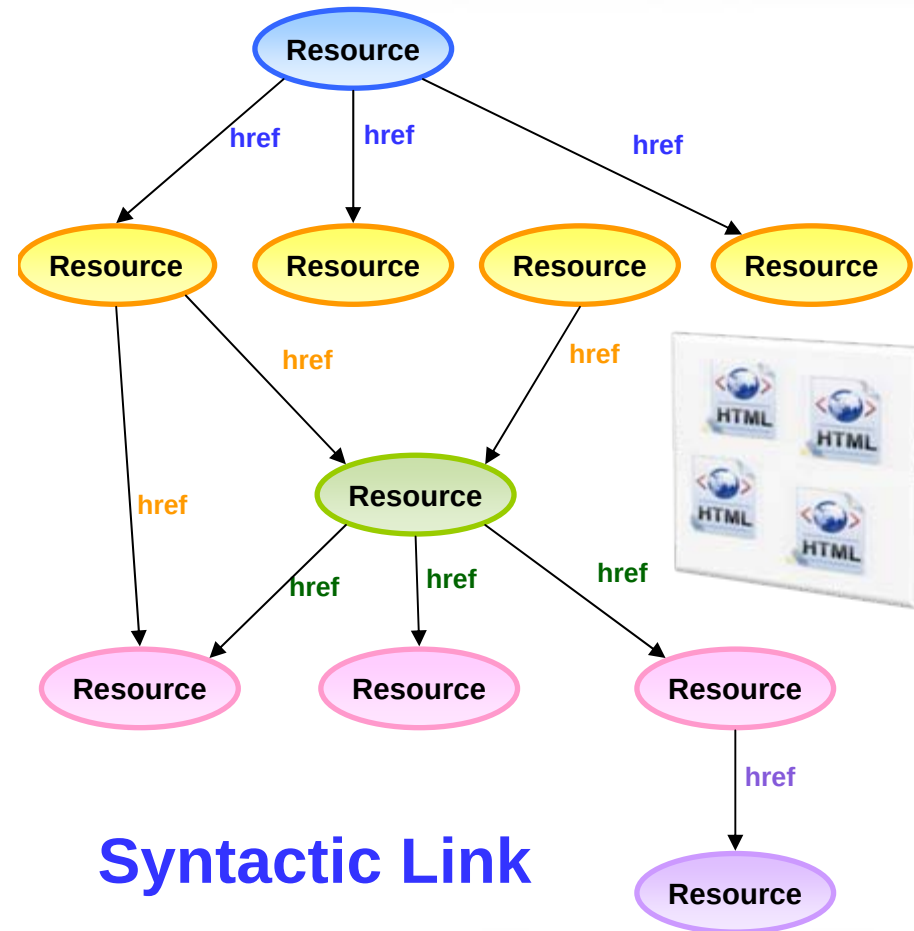
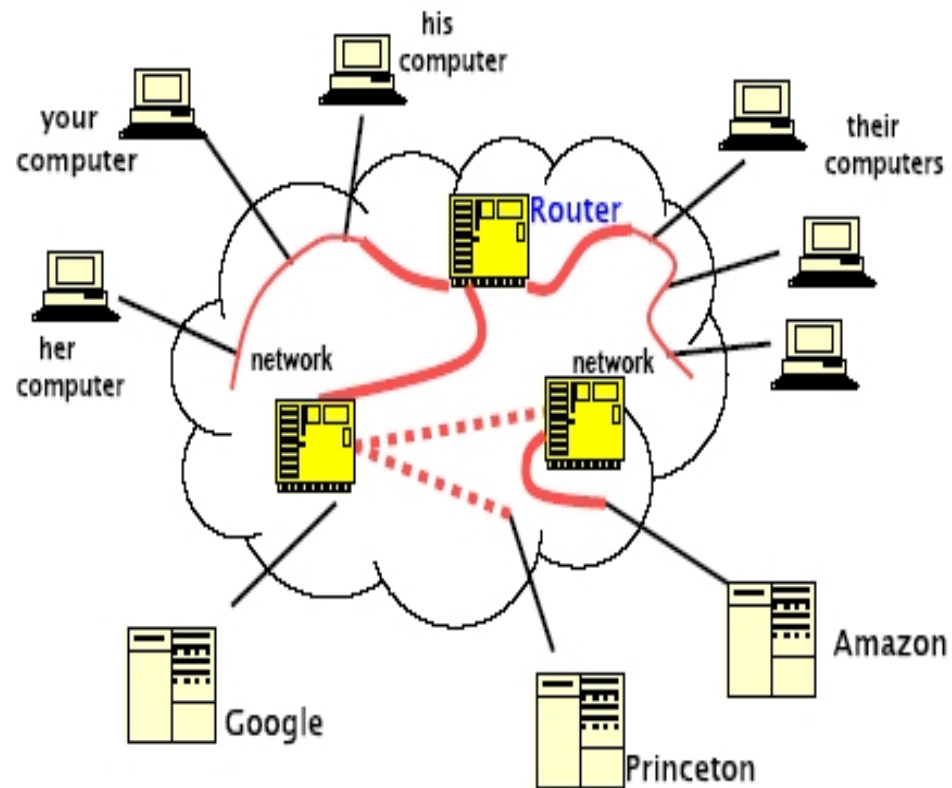
- Data-centric
- Relational resources
- Semi-structured data
- RSS, AJAX
- URI-centric
- Web 2.0



Semantic Web

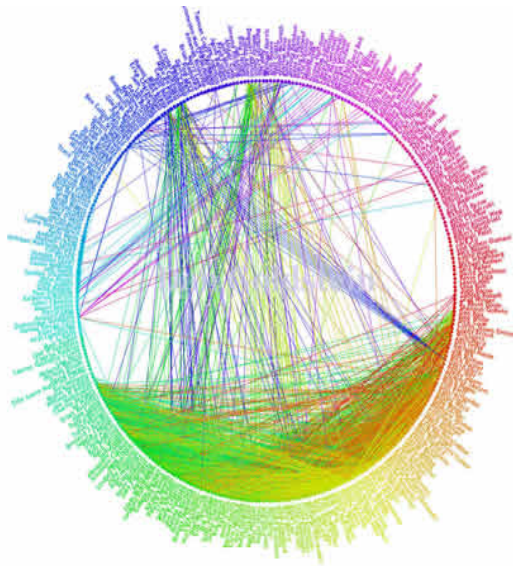
- Data-centric
- Linked Data
- Structured data
- RDF, RDF-S, OWL
- URI-centric
- Web 3.0

Internet Cloud & Hyperlink on WWW

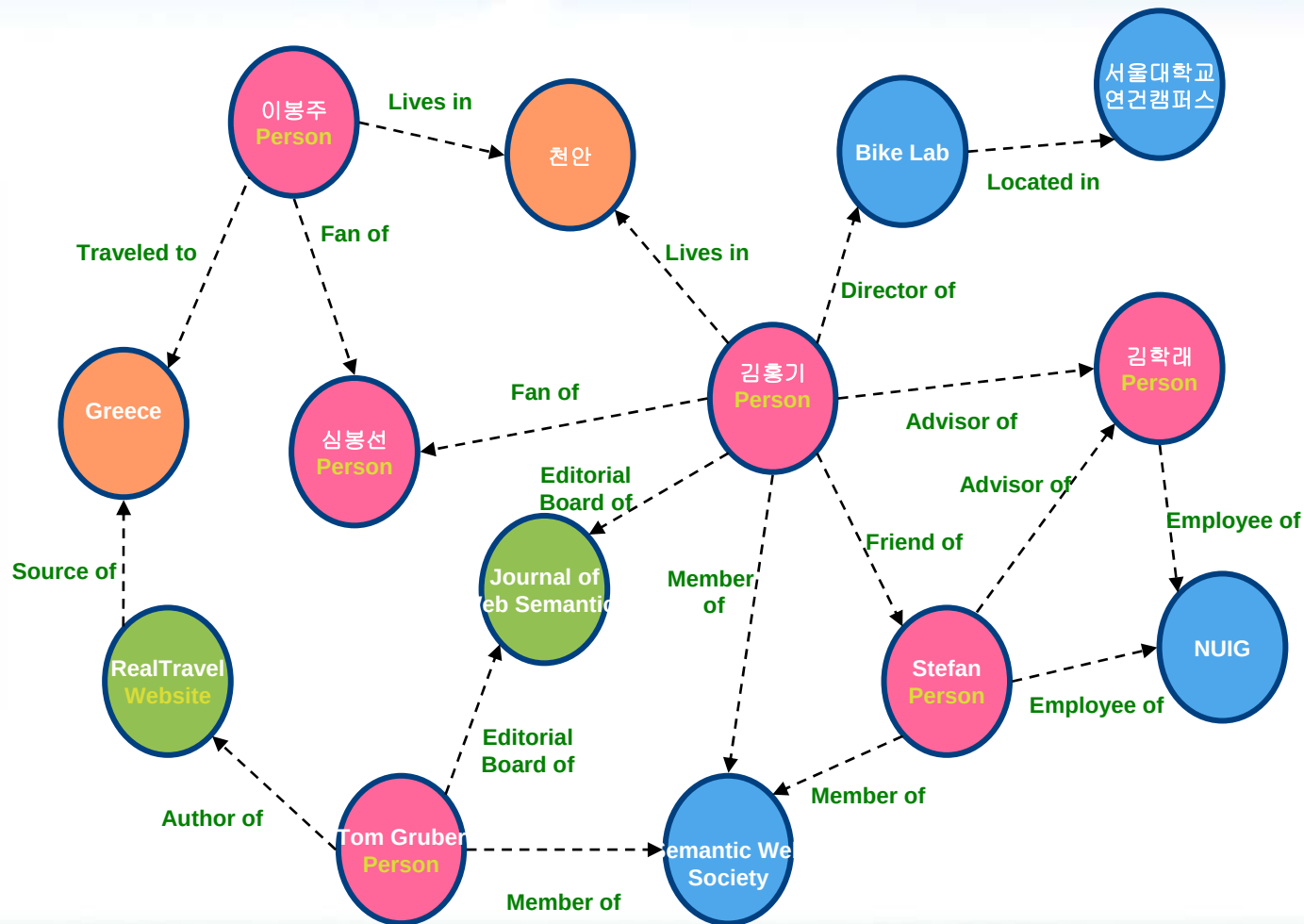


Syntactic Link

Graph based Semantic Link



Semantic Web
Cloud





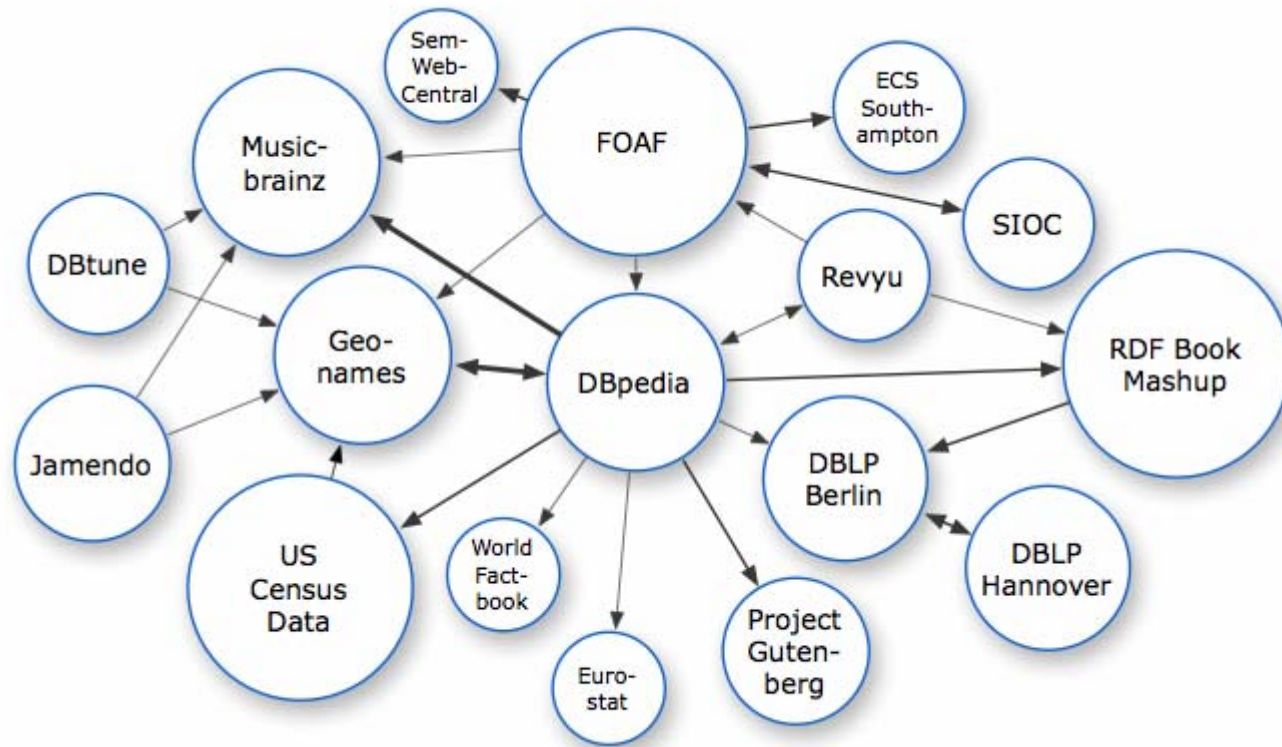
LINK

VS



CLOUD

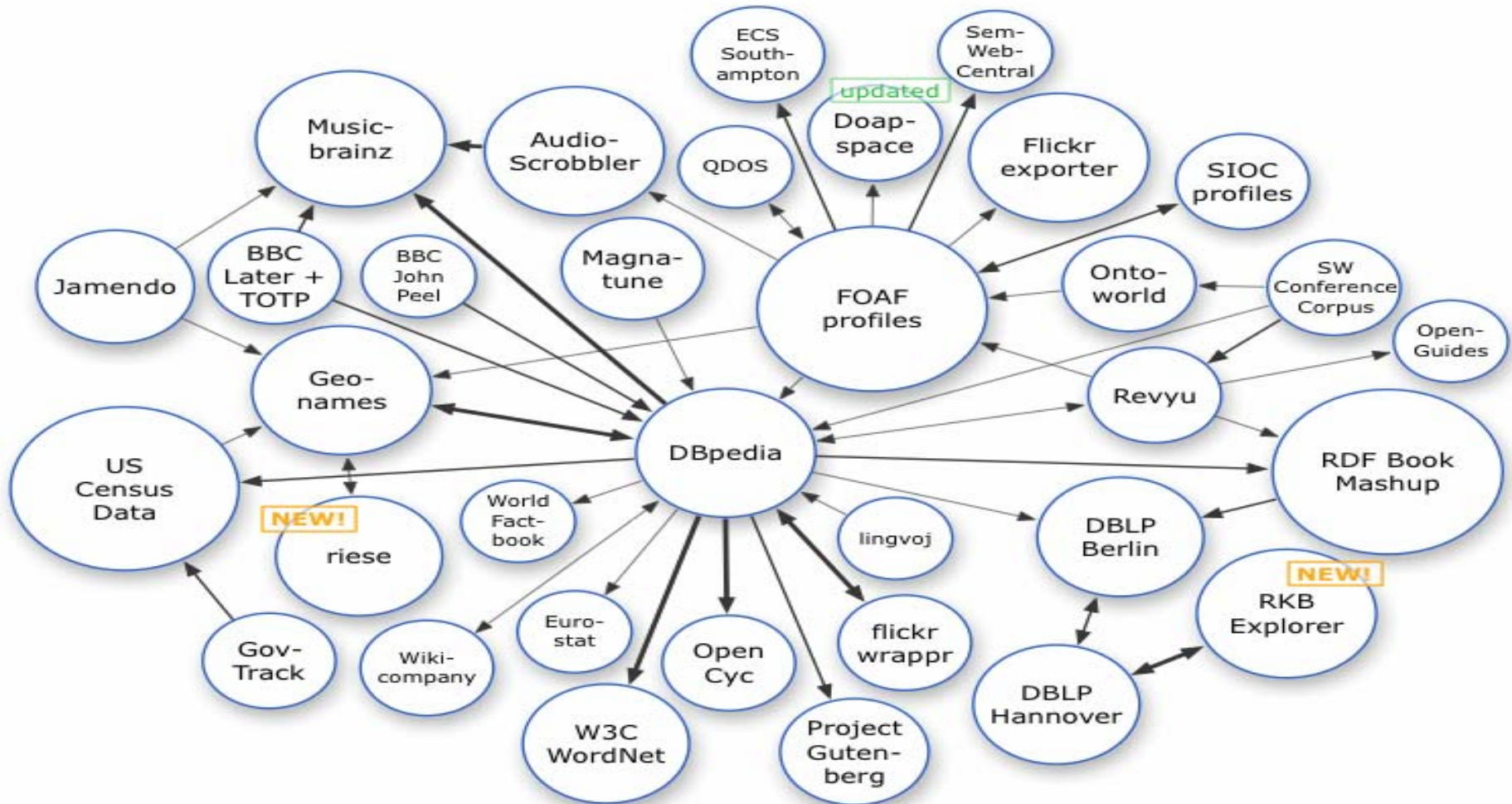
Linked Open Data Cloud (2006)



Christian Bizer, Tom Heath, Tim Berners-Lee: Linking Open Data, 17th W3C conference, 2008.



Linked Open Data Cloud (present)



Christian Bizer, Tom Heath, Tim Berners-Lee: Linking Open Data, 17th W3C conference, 2008.

Tabulator: Linked Data Browser



URI: <http://www3.wiwiss.fu-berlin.de:2020/all>
This is release 0.7 of the Tabulator Project. The [live development trunk](#) is also available.

[Add to outline](#) [Help](#) [About](#)

- ▶ The Tabulator Project
- ▶ Decentralised Information Group
- ▶ W3C (large: includes all specs and groups)
- ▶ W3C groups (Member-only data)
- ▶ Life Sciences demo data: GSK3-Beta Drug Target

▼ **D2R Server contents**

label D2R Server contents
seeAlso

▼ **List of all instances: Conferences**

mentions ▶ Conference
label List of all instances: Conferences
seeAlso ▶ D2R Server contents

▼ **International Semantic Web Conference 2002**

location Sardinia
date June 9-12, 2002
type ▶ Conference
label International Semantic Web Conference 2002
seeAlso ▶ List of all instances: Conferences
startDate 2002-10-09T00:00:00.0
is conference of ▶ Trusting Information Sources One Citizen at a Time
▶ Automatic Generation of Java/SQL based Inference Engines from RDF Schema and RuleML

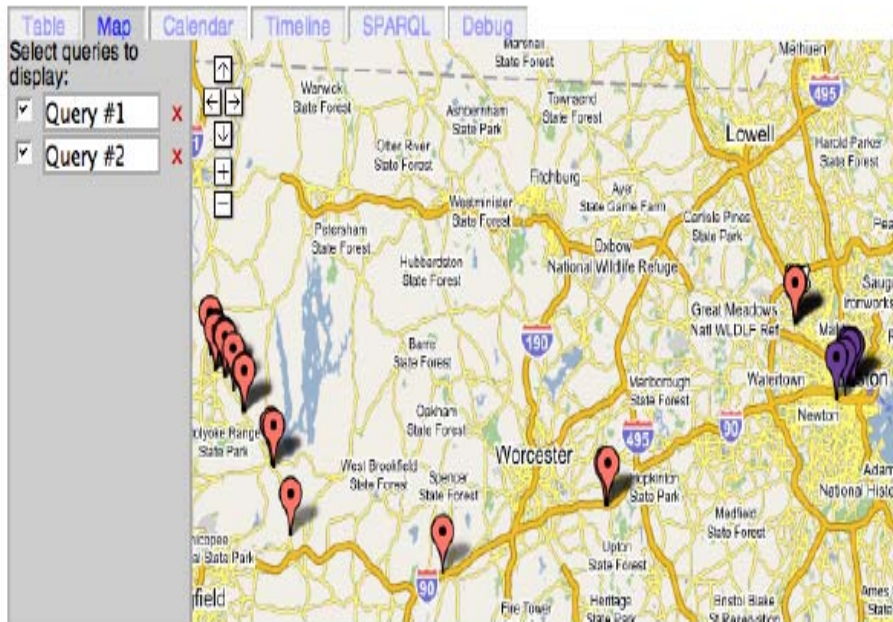
▼ **Three Implementations of SquishQL, a Simple RDF Query Language**

conference ▶ International Semantic Web Conference 2002
creator

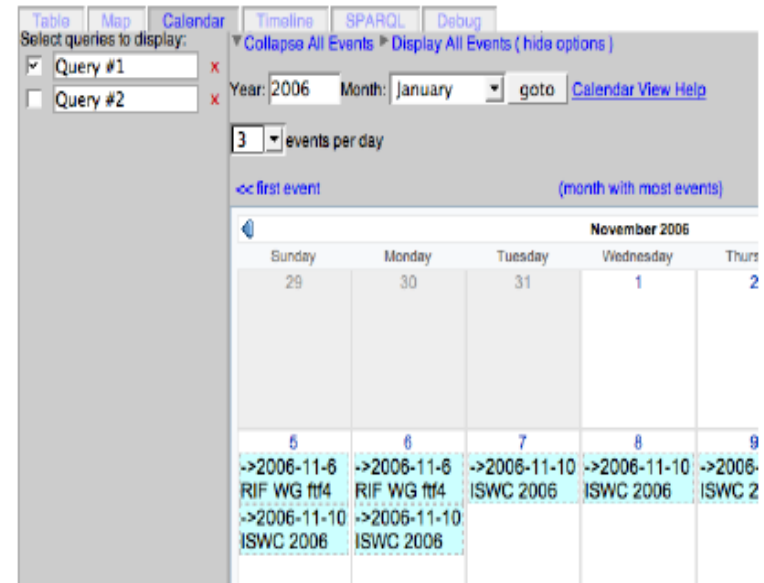
▼ **Andy Seaborne**

address Hewlett-Packard Laboratories, Bristol, BS34 8QZ, UK
has_affiliation ▶ Hewlett-Packard Laboratories, Bristol
research_interests ▶ Semantic Web
type ▶ Researcher
▶ Person
label Andy Seaborne
seeAlso ▶ List of all instances: Persons
homepage ▶ <http://www-uk.hpl.hp.com/people/afs/>

Views of RDF Data in Tabulator

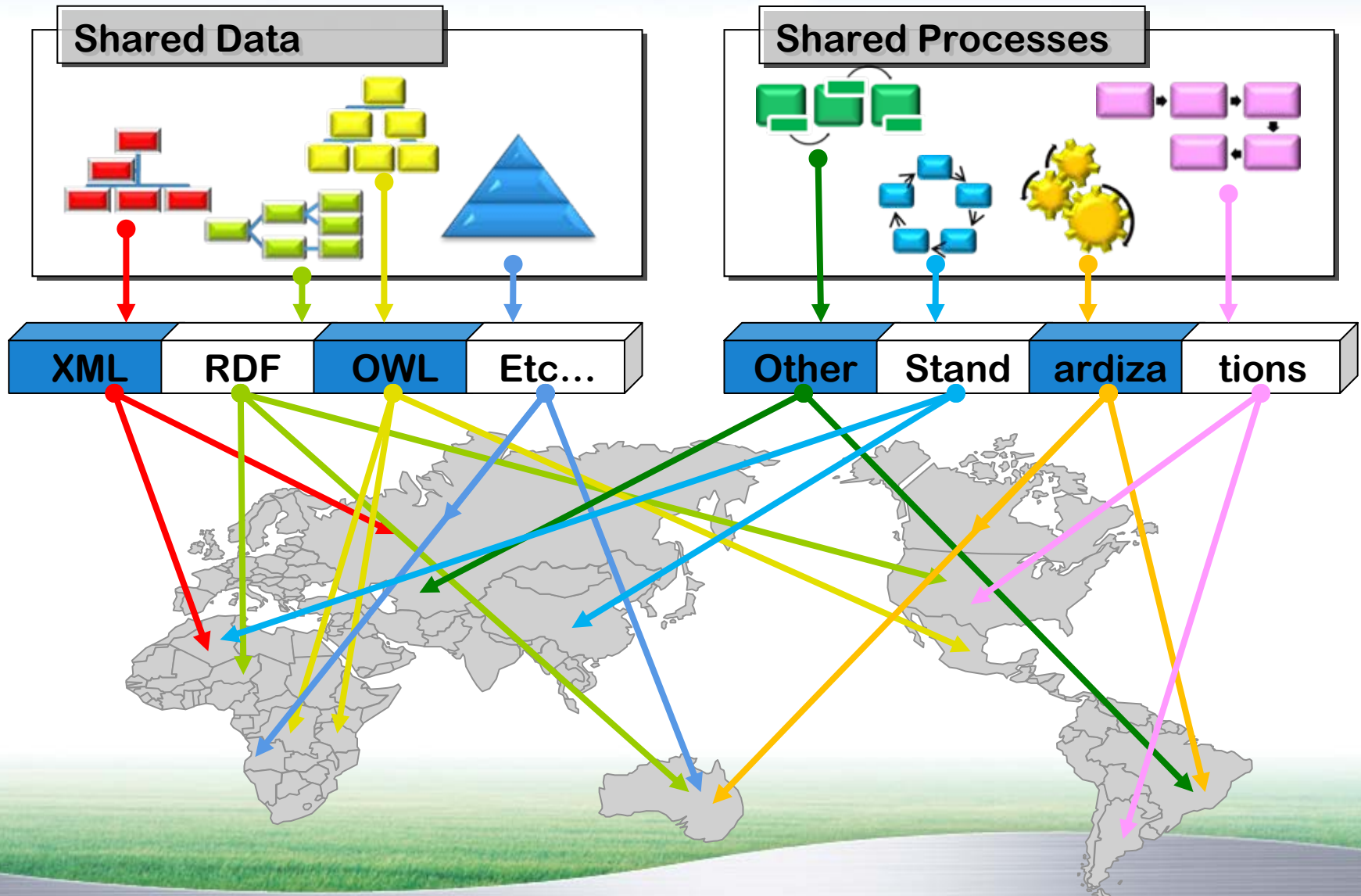


Map Viewer



Calendar Viewer

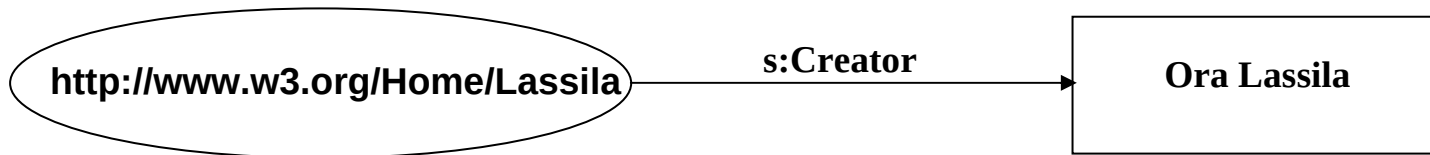
Data & Process sharing, Reusing



Two key ingredients of the SW



❖ RDF – 웹 메타데이터 표준 (W3C)



❖ Ontologies – “추론 가능한 데이터 모델”

- provide **shared understanding** of a domain
- organize knowledge in a **machine-comprehensible** way
- give an exploitable **meaning** to the **data**

Ontology = Conceptualization



뽕뽕



꼬까

다 ~ 내 꺼

맘마



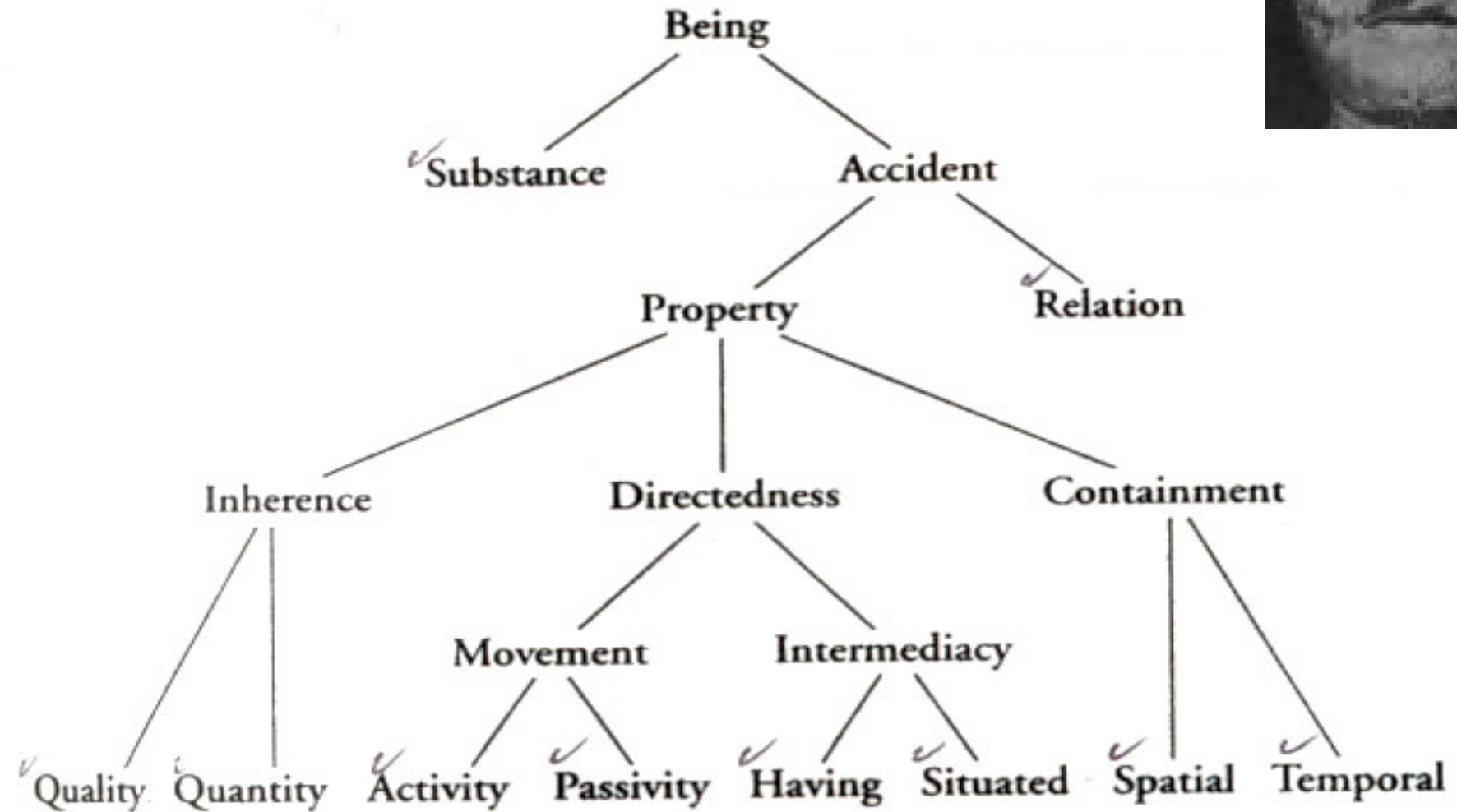
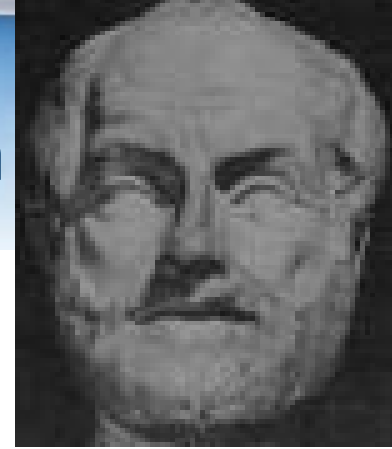
짱까



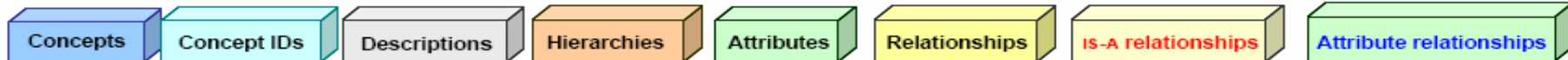
Ontology = Conceptualization



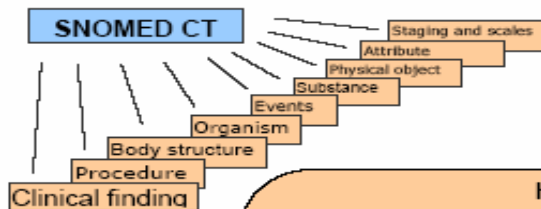
Ontology = Conceptualization



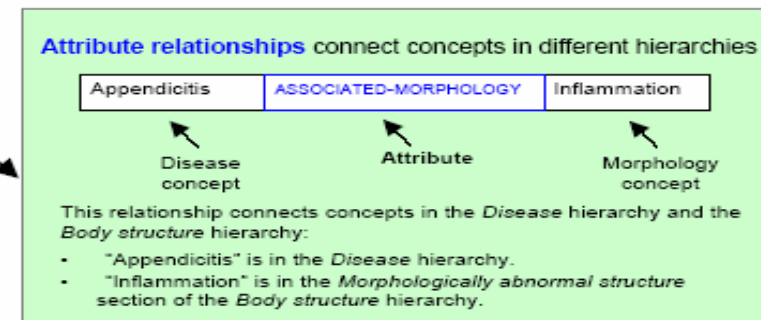
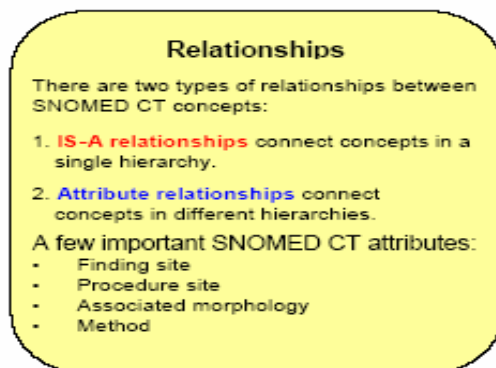
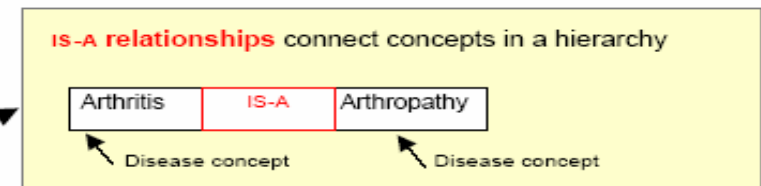
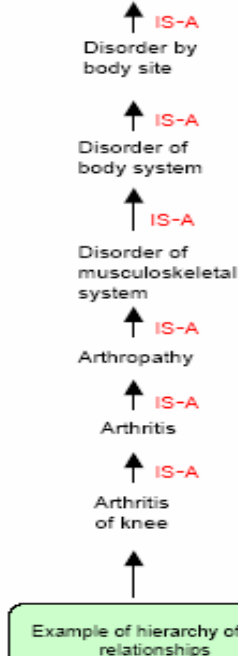
Ontology = Shared Conceptualization



SNOMED CT[®] Components



396285007	← Concept ID	← Fully specified name
1764202015	Bronchopneumonia (disorder)	
1776258010	Bronchopneumonia	
1785531017	Bronchial pneumonia	
IS-A		Pneumonia
ASSOCIATED-MORPHOLOGY		Inflammation
FINDING-SITE		Lung structure
FINDING-SITE		Bronchial structure



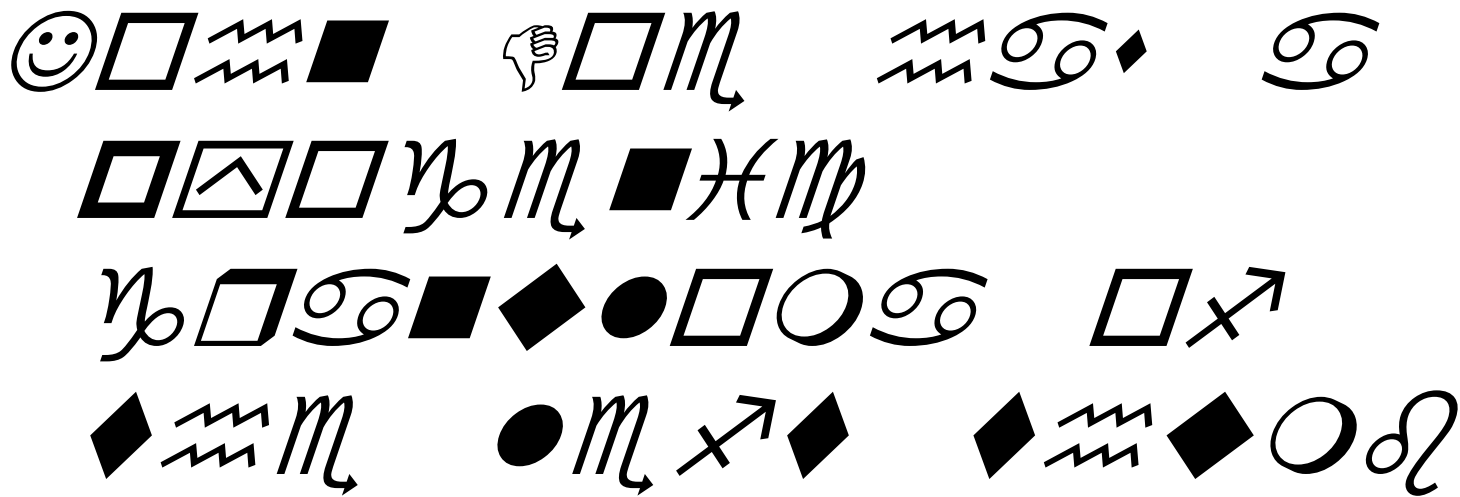
Unstructured data (only human readable)



사람의 관점

"John Doe has a pyogenic granuloma of the left thumb"

기계 또는 에이전트의 해석



Semi-structured data (not machine readable)



사람의 관점

<record>

<patient>John Doe</patient>

<diagnosis>pyogenic granuloma of the left thumb</diagnosis>

</record>

기계/에이전트의 관점

<□ㄴ ㄴ□□으>

<◆◆ㄱㄴㄴ◆> ☺□ㄴ■ ☞□ㄴ </◆◆ㄱㄴㄴ◆>

<으ㄴㄴㄴ■□◆ㄴ> □□□ㄴㄴ■ㄴㄴㄴㄴㄴㄴㄴㄴㄴㄴㄴ
□ㄴ☞ ◆ㄴㄴㄴ ●ㄴㄴ☞ ◆ㄴㄴ◆○ㄱ </으ㄴㄴㄴ■□◆ㄴ>

</□ㄴ ㄴ□□으>

XML을 데이터의 구조를 정의할 수 있지만, 의미를 정의할 수 없음

Ontology is Machine Processable



사람의 관점

Adult = Person $\cap$ $\exists$ age.over17

기계/에이전트의 관점

```
<owl:Class rdf:ID="Adult">
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Person"/>
    <owl:restriction>
      <owl:onProperty rdf:resource="#age"/>
      <owl:someValuesFrom
        df:resource="http://www.w3.org/TR/@@/owl-ex-dt#over17"/>
      </owl:Restriction>
    </owl:intersectionOf>
  </owl:Class>
```

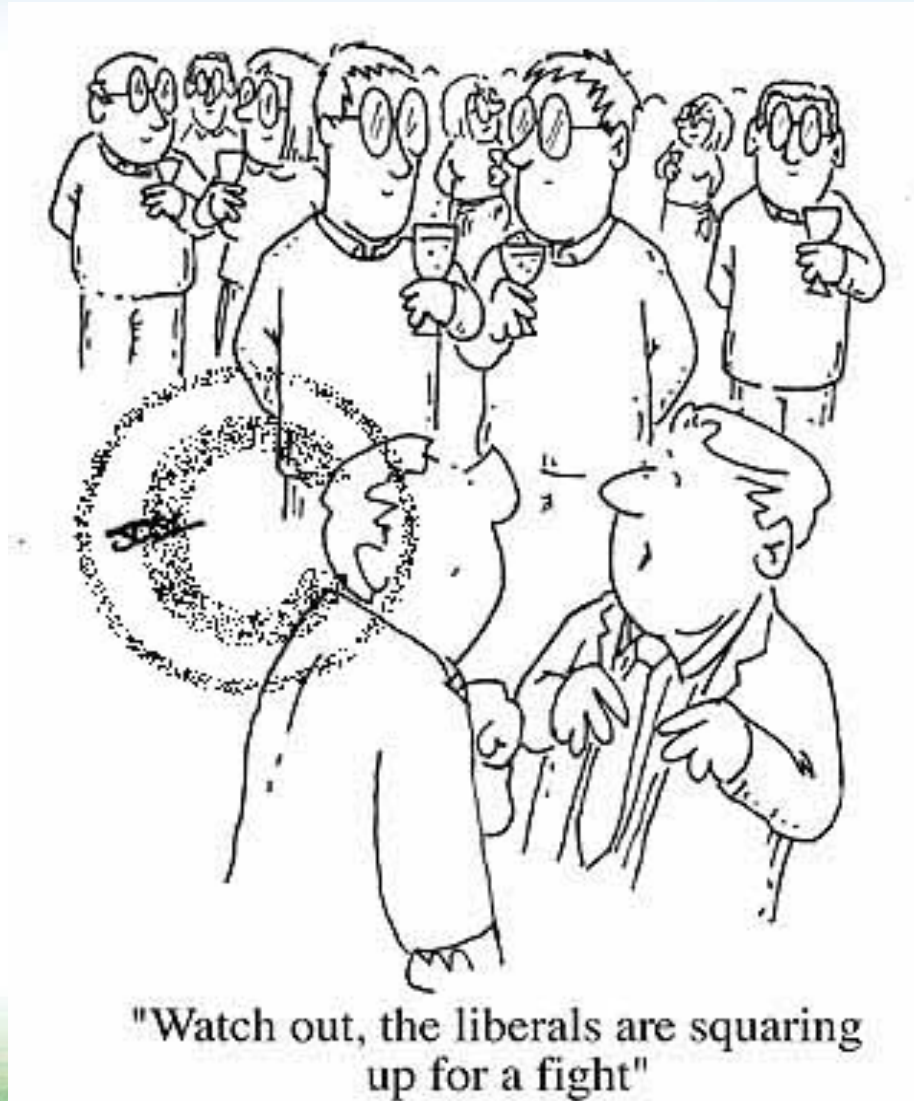

Debate between Naïve Realism and Representationalism



Communication and Semantic Modeling Problem



Making Semantics: Right Wing vs. Left Wing



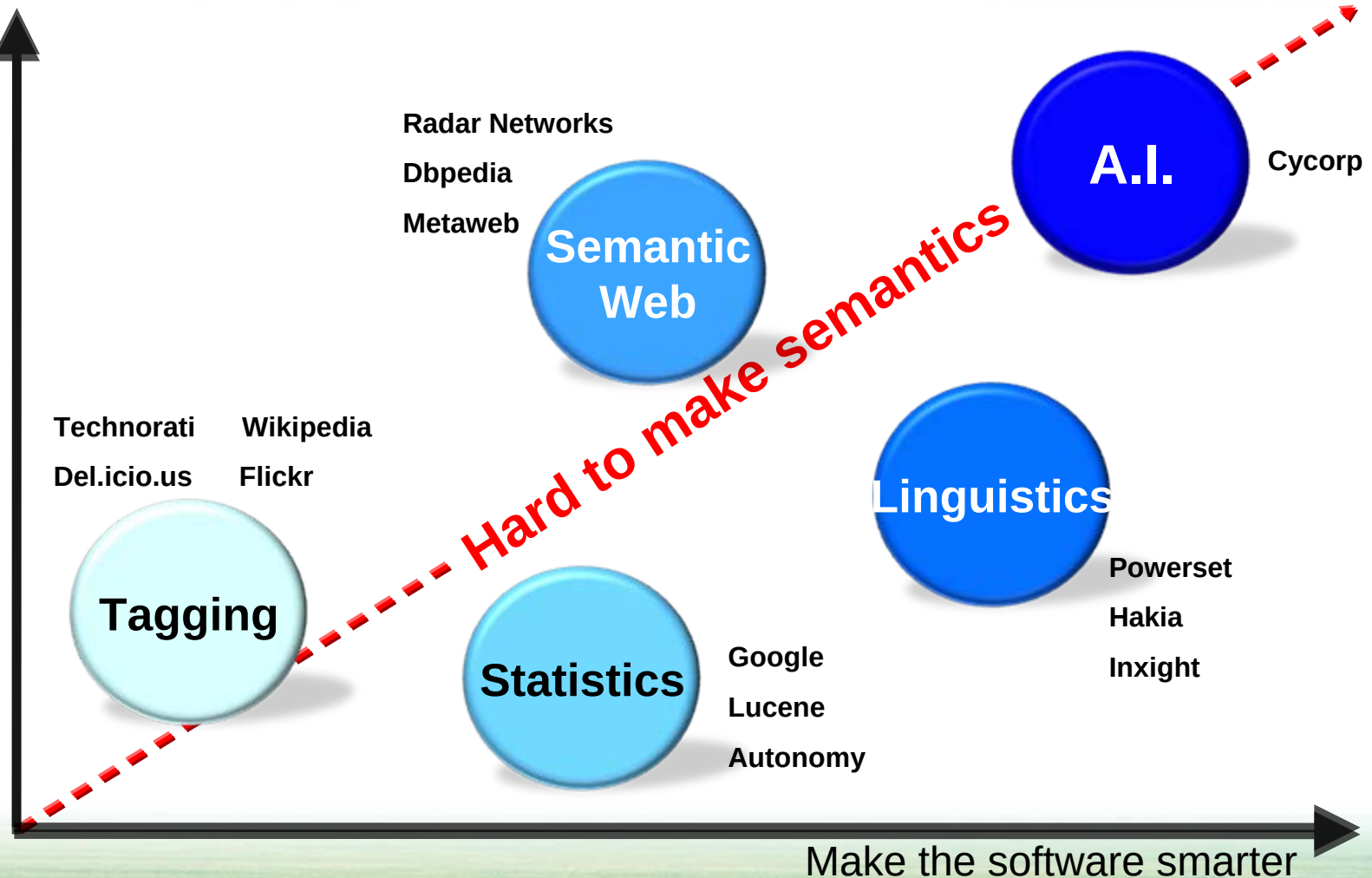
Taggers

Ontologists

Five Approaches to Semantics



Make the Data Smarter



RDF: Semantic Web over Web Resources



- ❖ RDF (Resource Description Framework):
resources = RDF로 표현되어지는 모든 것

Example of RDF Statement

Ora Lassila is the creator of the resource <http://www.w3.org/Home/Lassila>.

Subject (resource)	http://www.w3.org/Home/Lassila
Predicate (property)	Creator
Object (literal)	“Ora Lassila”

RDF Example



❖ Graphical View



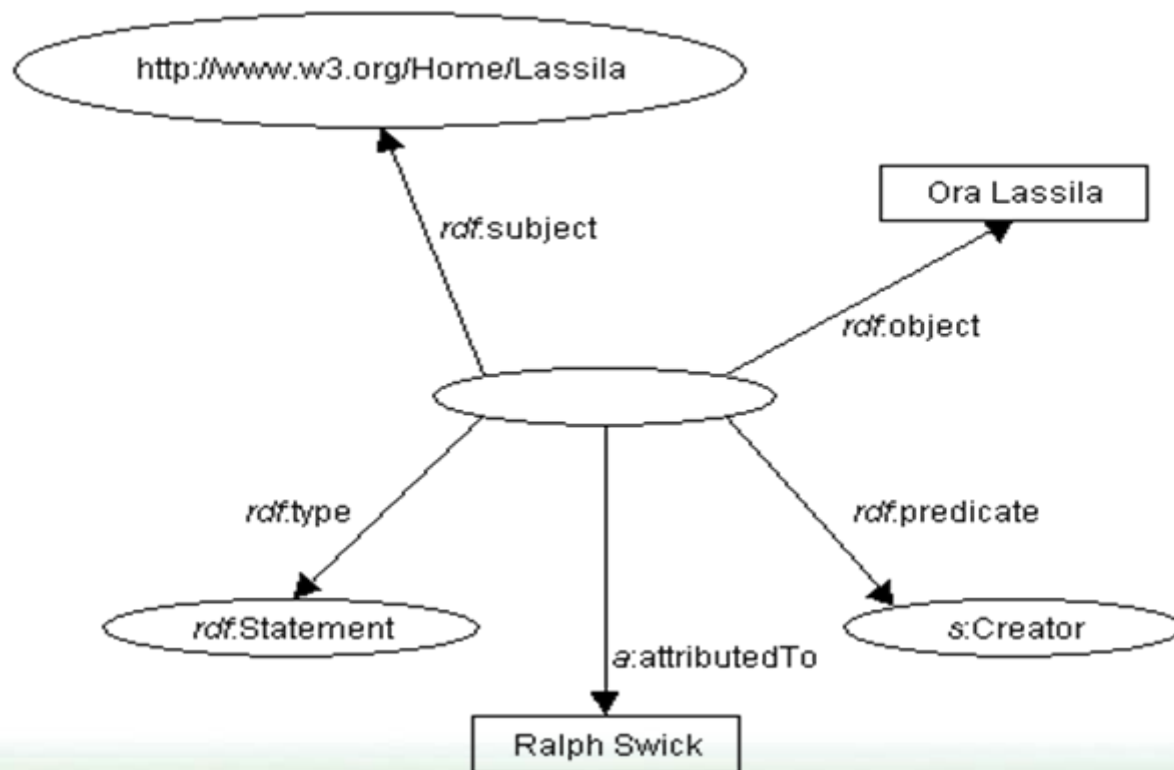
❖ Serialization Syntax

```
<rdf:RDF>  
  <rdf:Description about=  
    "http://www.w3.org/Home/Lassila">  
    <s:Creator>Ora Lassila</s:Creator>  
  </rdf:Description>  
</rdf:RDF>
```

Statements about Statements



“Ralph Swick says that Ora Lassila is the creator of the resource <http://www.w3.org/Home/Lassila>”





❖ RDF Schema

- Defines vocabulary for RDF
- Organizes this vocabulary in a typed hierarchy (Class, subClassOf, type, Property, subPropertyOf)
- Explicitly declares semantic relations between vocabulary terms

RDF+RDFS Example



RDF Document

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE rdf:RDF [
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax-ns#'>
  <!ENTITY mv 'http://protege.stanford.edu/mv#'>
]>
<rdf:RDF xmlns:rdf="&rdf;"
  xmlns:mv="&mv;"
  <mv:Truck rdf:about="&mv;test3_03">
    <mv:registeredTo rdf:resource="&mv;test3_04"/>
  </mv:Truck>
  <mv:Person rdf:about="&mv;test3_04"
    mv:name="Ora Lassila"/>
</rdf:RDF>
```

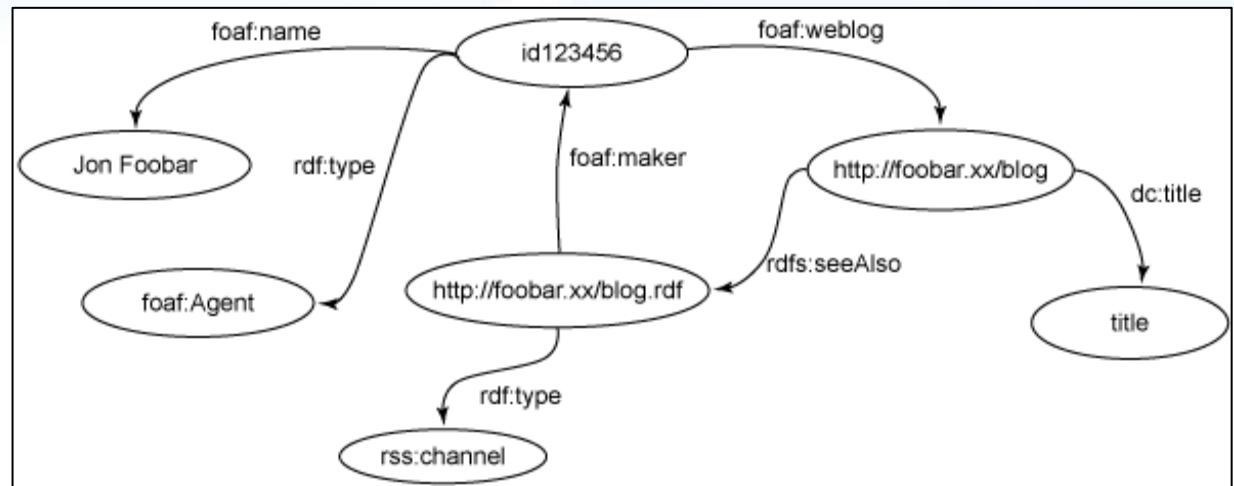
RDFS

```
<rdf:Class rdf:about="&mv;MotorVehicle">
  <rdf:subClassOf rdf:resource="&rdfs;Resource"/>
</rdf:Class>
<rdf:Class rdf:about="&mv;Person">
  <rdf:subClassOf rdf:resource="&rdfs;Resource"/>
</rdf:Class>
```

```
<rdf:Class rdf:about="&mv;Truck">
  <rdf:subClassOf rdf:resource="&mv;MotorVehicle"/>
</rdf:Class>
<rdf:Property rdf:about="&mv;registeredTo"
  a:maxCardinality="1">
  <rdf:domain rdf:resource="&mv;MotorVehicle"/>
  <rdf:range rdf:resource="&mv;Person"/>
</rdf:Property>
```

- ❖ SPARQL Protocol And RDF Query Language
- ❖ Query Language + Protocol + XML Results Format
- ❖ A query language for access and query RDF graphs
- ❖ Mashing up disparate data sources represented in RDF

SPARQL Query Example



Jon Foobar란 이름을
가진사람의 블로그의
URL을 찾기



Result
"http://foobar.xx/blog"

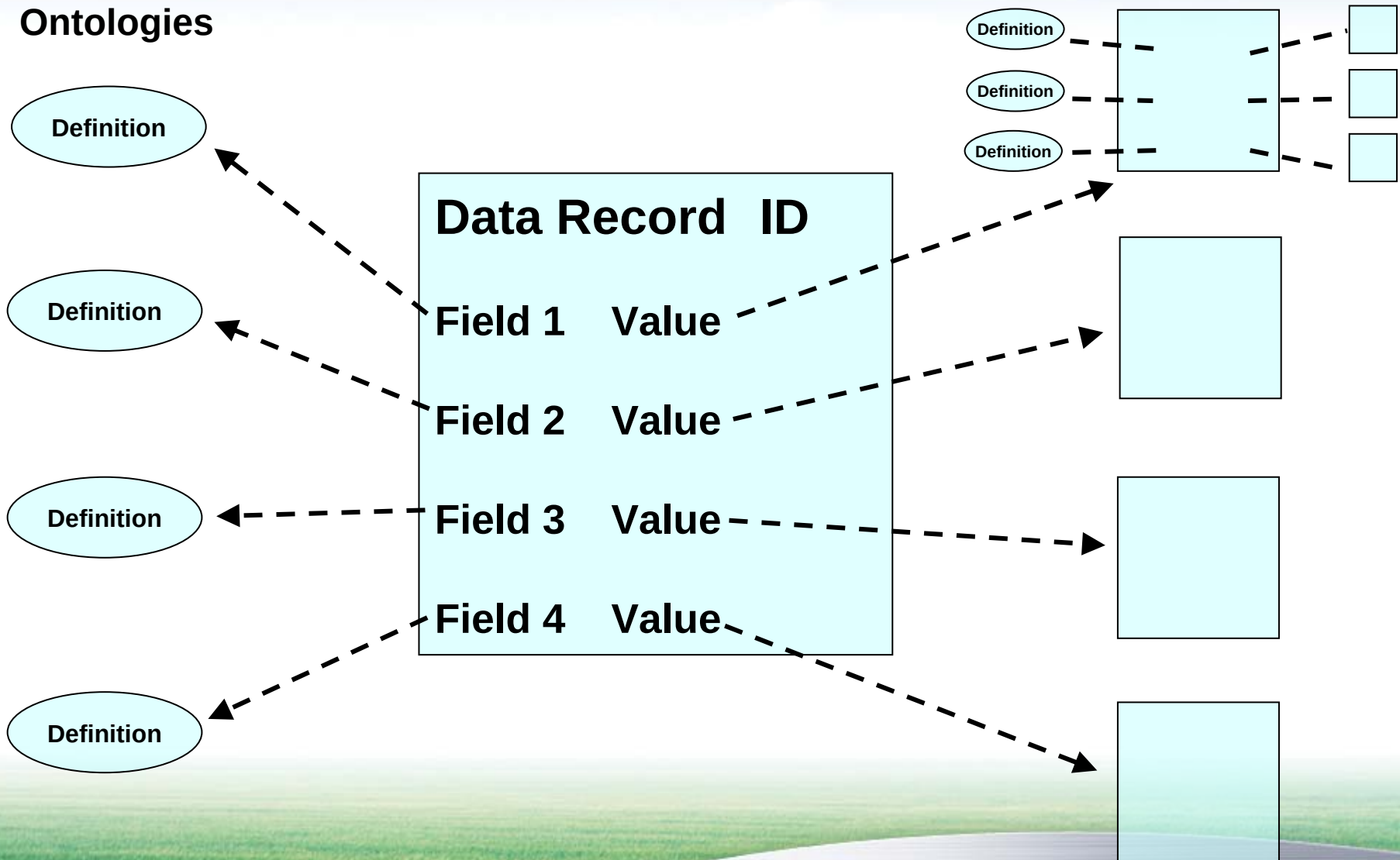
SPARQL query to find the URL of a contributor's blog

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?url
FROM <bloggers.rdf>
WHERE {
    ?contributor foaf:name "Jon Foobar" .
    ?contributor foaf:weblog ?url .
}
```

Semantic Web Data is Self-Describing Linked Data



Ontologies



RDBMS vs RDF Triplestore

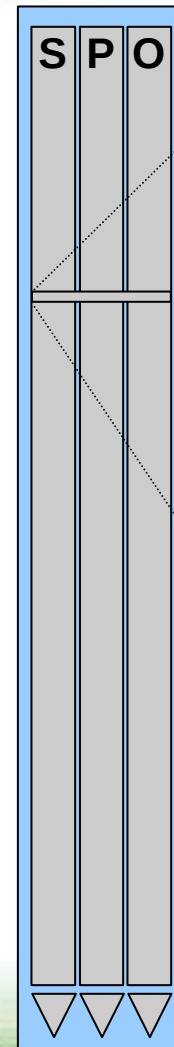


Person Table

<u>ID</u>	<u>f_name</u>	<u>l_name</u>
001	jim	wissner
002	nova	spivack
003	chris	jones
004	lew	tucker

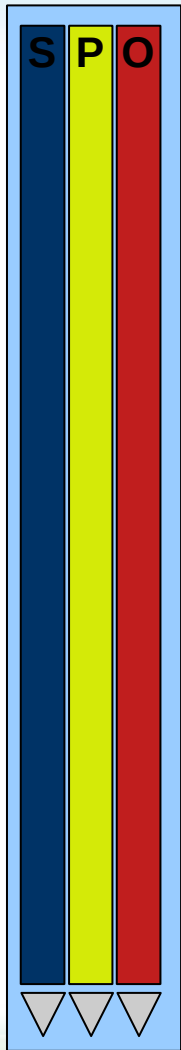
Colleagues Table

<u>SRC-ID</u>	<u>TGT-ID</u>
001	001
001	002
001	003
001	004
002	001
002	002
002	003
002	004
003	001
003	002
003	003
003	004
004	001
004	002
004	003
004	004

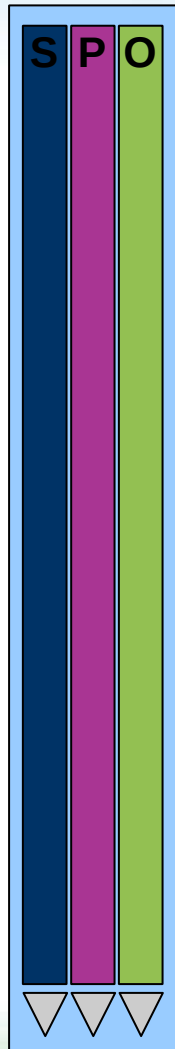


Subject	Predicate	Object
001	isA	Person
001	firstName	Jim
001	lastName	Wissner
001	hasColleague	002
002	isA	Person
002	firstName	Nova
002	lastName	Spivack
002	hasColleague	003
003	isA	Person
003	firstName	Chris
003	lastName	Jones
003	hasColleague	004
004	isA	Person
004	firstName	Lew
004	lastName	Tucker

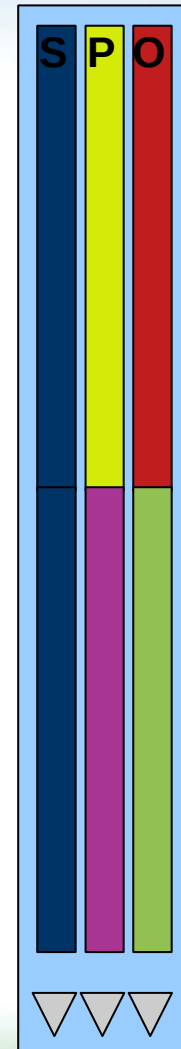
Merging Databases in RDF is Easy



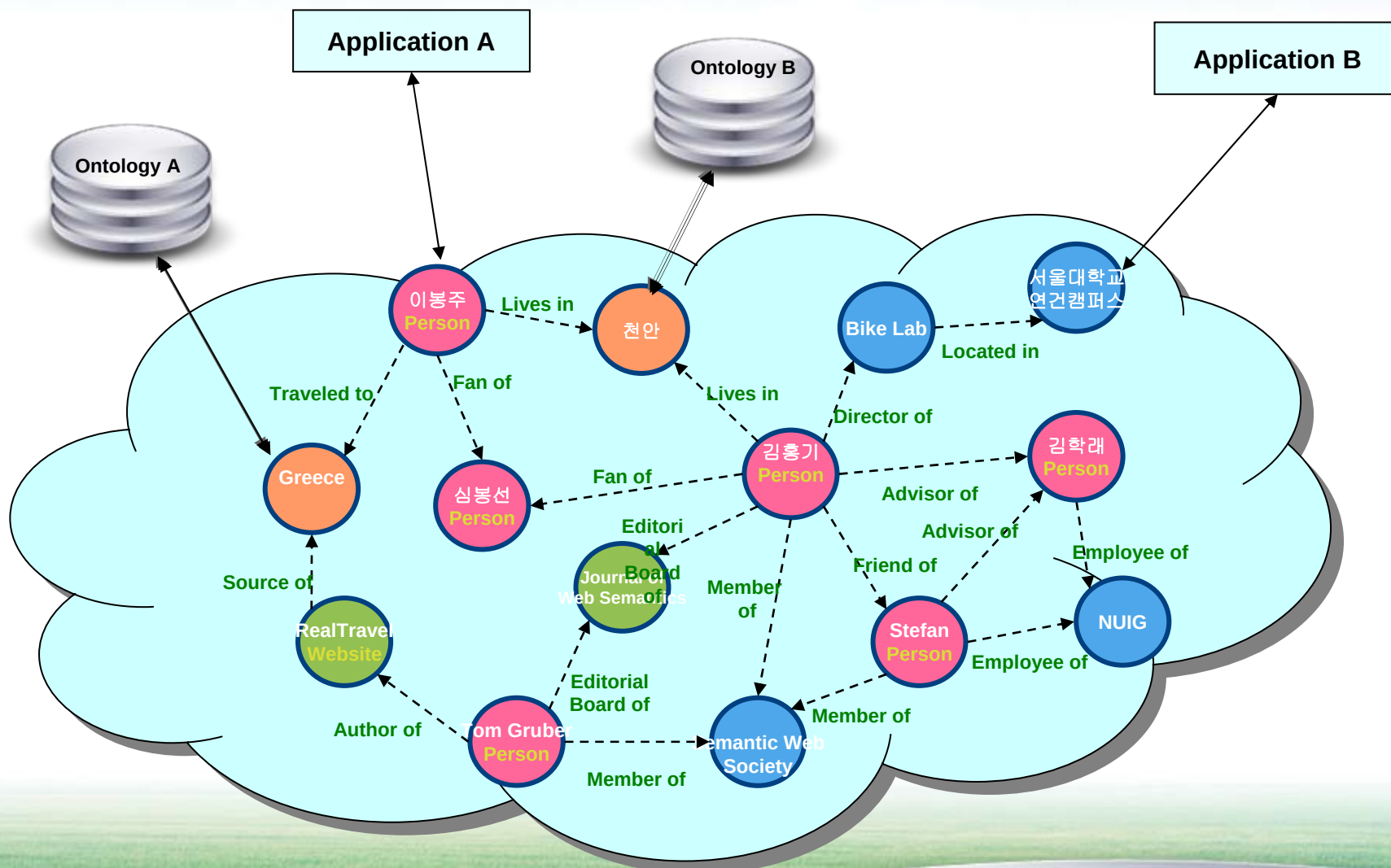
+



=



The Global Database of the Web!



Unlinked Medical Records



Structured Data Entry for Semantic Links



❖ Knowledge base

- 비구조적 텍스트를 구조적 형태로 표현하여 저장하는 방법은?

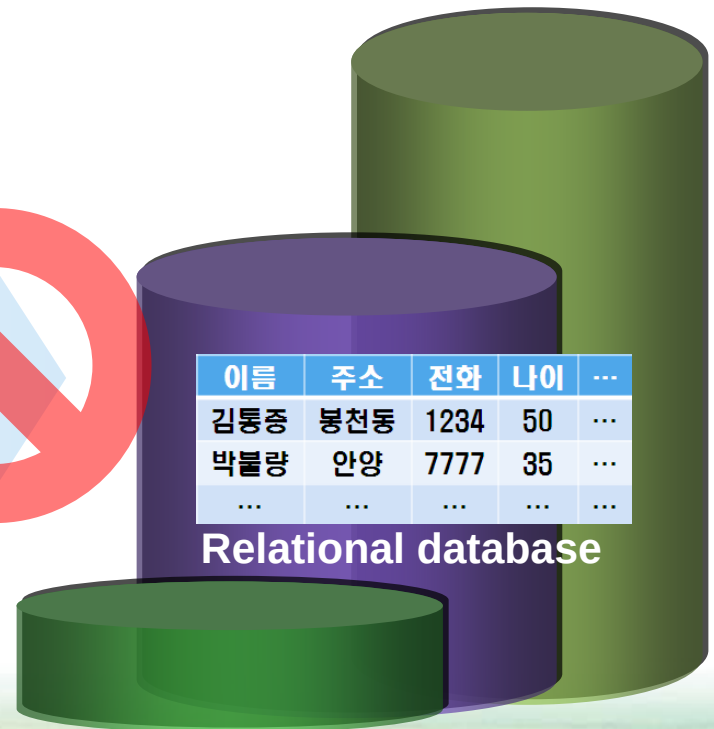
❖ Relational DB vs. RDF DB

받은 조직은 포르말린에 고정된
부분절제된 하악골 덩어리로 부패되어있는 상태임.
35, 35, 37번 치아를 포함하고 있으며
크기는 6.5 x 4.5 x 1.5 cm임.
병소는 37번 치아후방 부위를 천공시키고 있는
낭성병소로 내부에 치아를 포함하고 있음.
또 하악골의 협측과 설측 피지리골판이 얇아지고
부분적인 천공을 보임.
단면을 내어보았을 때 병소는 하악체내부에
매복된 치아의 치경부위를 감싸는 낭구조이며
내부는 비어 있음. 대표부위를 포매함



이름	주소	전화	나이	...
김통중	봉천동	1234	50	...
박불량	안양	7777	35	...
...	...	...	...	...

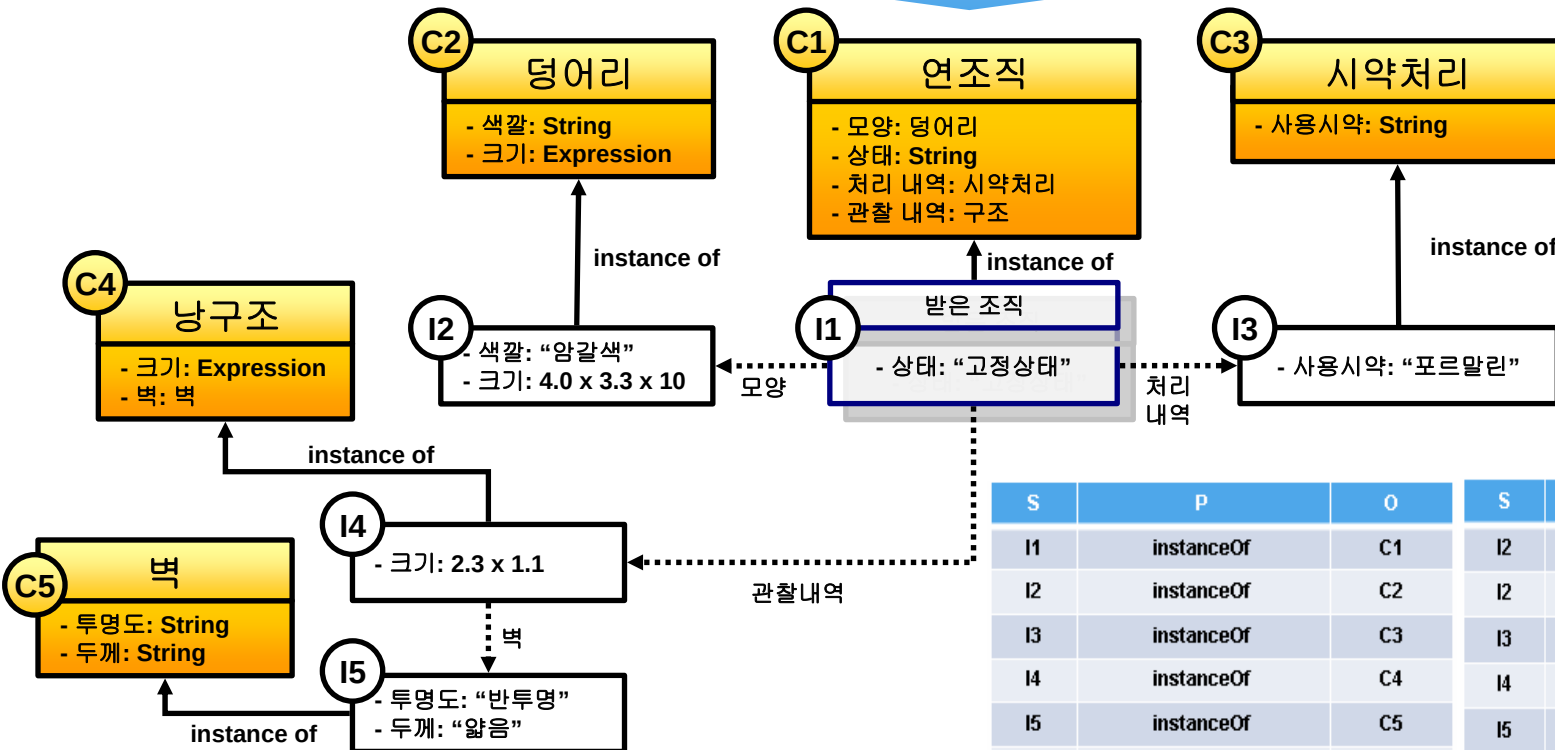
Relational database



RDF for Structuring Data



받은 조직은 포르말린에 고정된 상태이며,
크기 40x3.3x1.0cm의 암갈색 연조직 덩어리로,
한 부위에 크기 2.3x1.1cm크기의 낭구조가 관찰됨



S	P	O	S	P	O
I1	instanceOf	C1	I2	hasDataValue:색깔	"암갈색"
I2	instanceOf	C2	I2	hasDataValue:크기	4.0x3.3x1.0
I3	instanceOf	C3	I3	hasDataValue:사용시약	"포르말린"
I4	instanceOf	C4	I4	hasDataValue:크기	2.3 x 1.1
I5	instanceOf	C5	I5	hasDataValue: 투명도	"반투명"
I1	hasDataValue:상태	"고정상태"	I5	hasDataValue:두께	"얇음"
I1	hasObjectValue:모양	I2			
I1	hasObjectValue:처리내역	I3			
I1	hasObjectValue:관찰내역	I4			
I4	hasObjectValue:벽	I5			

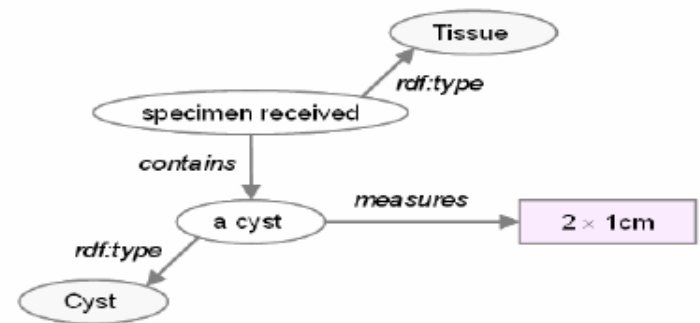
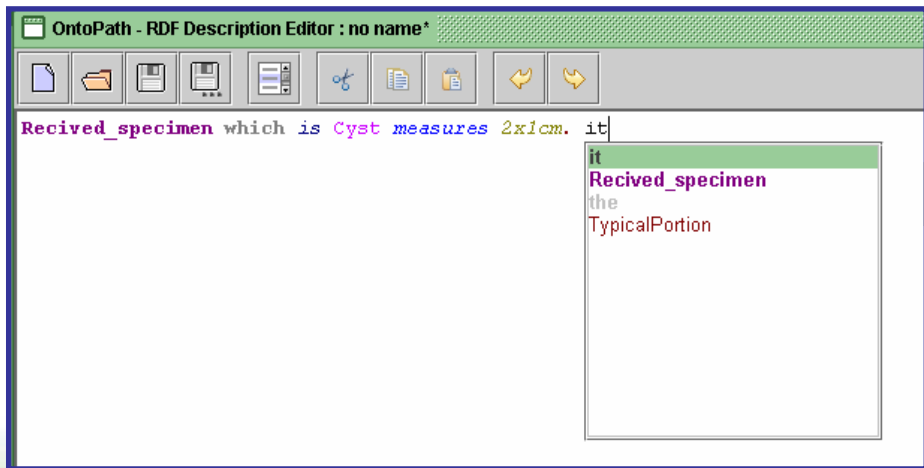
Narratives to RDF

❖ OntoPath

- translation from a sentence to RDF triples

❖ Ontology

- Vocabulary for description
- Semantic relation



Description Editor for Structured Data Entry



http://oogl.snu.ac.kr - Pathology Report - Microsoft Internet Explorer

Request	
No.	S91-644
Date	07/08/1991
Patient	Seung Han (M/62)
Patient No.	192974
Client	OS (outpatient/S)
In Charge	Dr. Nam
History	Perceive objectOf Mass and atSiteOf mouseFloor and atTimeBeforeNow 8month. the Mass hasProcess Grow which hasState slow.
Anatomic Site	anterior mouth floor
Type	Incision
Clinical Diagnosis	SquamousCellCarcinoma

Report	
Gross Description	
specimen_received which is SoftTissue isFixedIn formalin and hasColor darkBrown and measures 4x3x1cm. it contains Cyst which measures 2x1cm. the Cyst contains CystWall which hasThickness thin and hasTransparency transparent. TypicalPortion isProcessed embedded.	
Slides	IH&E
Diagnosis	SquamousCellCarcinoma, WellDifferentiated
Images	
Capture	Remove
image1	
Reset Done	

OntoPath - RDF Description Editor : gross36.rdf+

specimen_received which is SoftTissue isFixedIn formalin and hasColor darkBrown and measures 4x3x1cm. it contains Cyst which measures 2x1cm. the Cyst contains CystWall which hasThickness thin and hasTr

hasFeature
hasShape
hasSizeAbout
hasState
hasThickness
hasTransparency
isFixedIn
isProcessed
it
measures
SoftTissue
specimen_received
Substance

Problems

Select an appropriate term

pathont.rdfs

insert

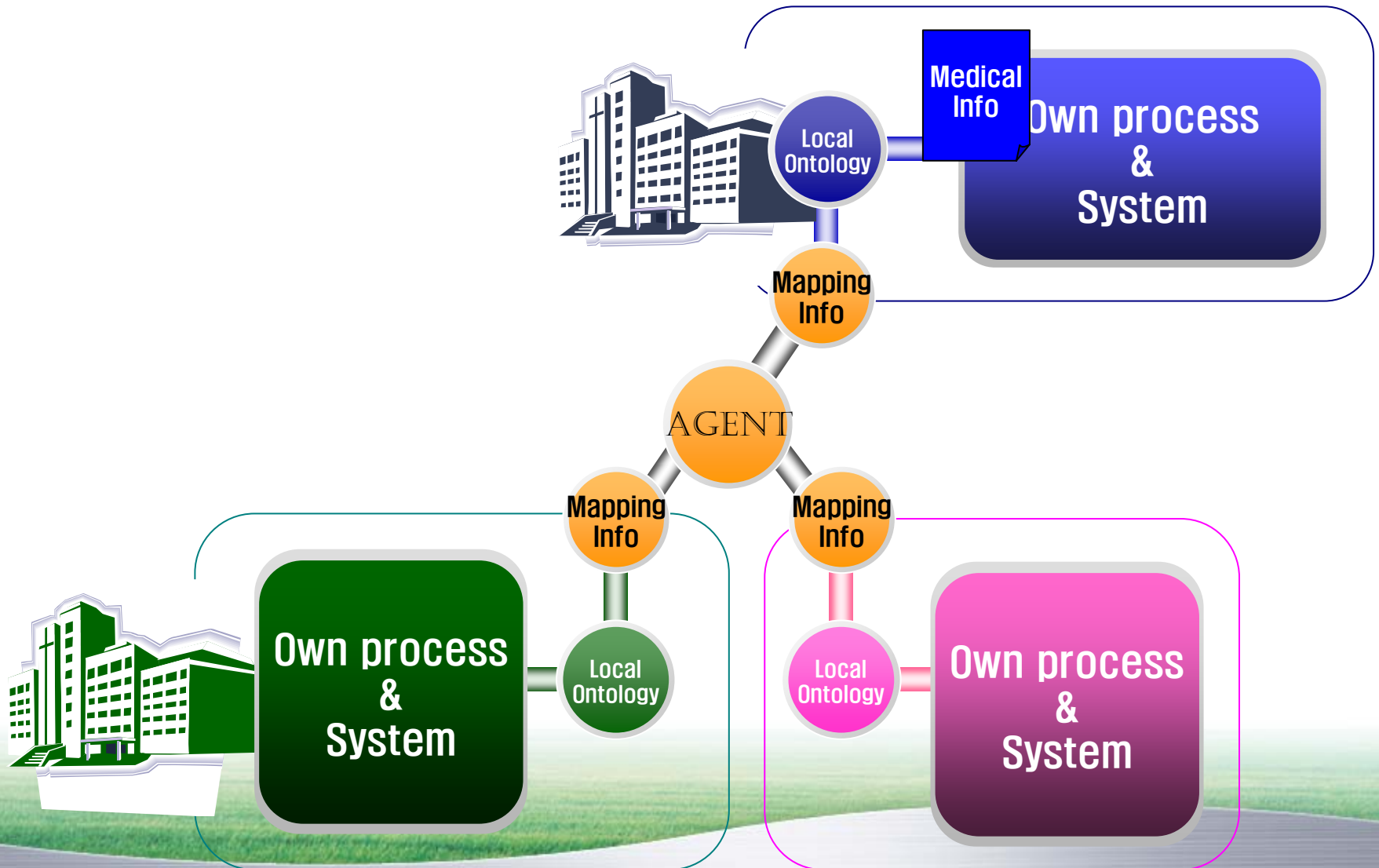
4 : 28

Semantic Annotation for PDF & Other Documents (manual & automatic)

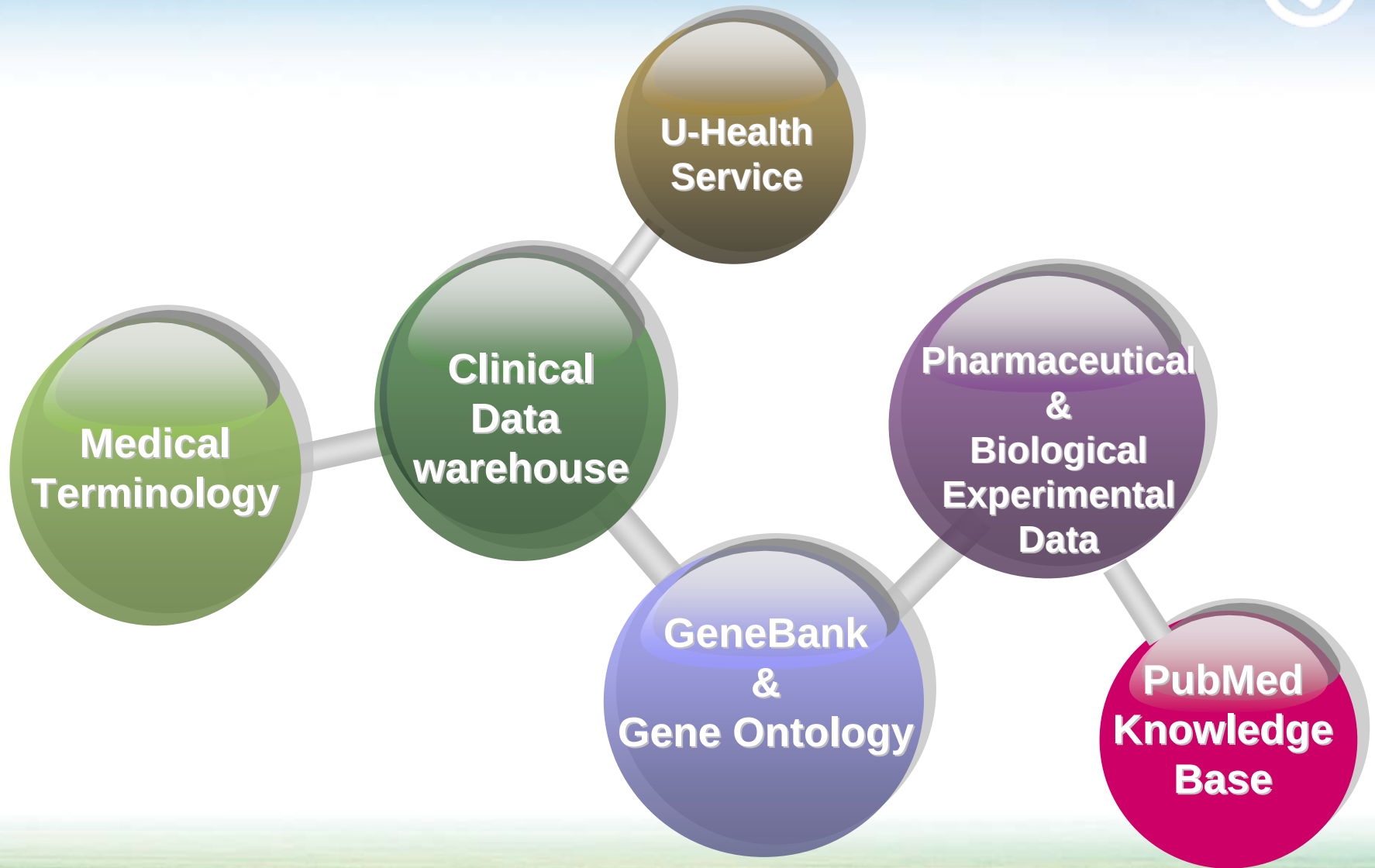


The screenshot displays the Fox Studio application window. The top menu bar includes 'File', 'Search', 'Tool', 'Window', and 'Help'. Below the menu is a toolbar with icons for 'Automatic Annotation' and other functions. The left sidebar shows a hierarchical tree of ontology classes under 'owl:Thing', including 'Event', 'Organization', 'Person', 'Product', 'Project', 'Publication' (with sub-items 'Article', 'Book', 'Manual', 'Misc', 'Proceedings'), 'Report', and 'Thesis'. The main window displays a PDF document titled 'FOIS98.pdf'. The document content includes a title 'Formal Ontology and Information Systems' in a blue box, the author 'Nicola Guarino', and his affiliation 'National Research Council, LADSEB-CNR, Corso Stati Uniti 4, I-35127 Padova, Italy' with email 'guarino@ladseb.pd.cnr.it'. Below the main text, there is a section for 'Abstract. Research on ontology is becoming increasingly widespread in the com-'. At the bottom, a panel titled 'Thesis_1' shows 'Annotations' with 'rdfs:label' set to 'Formal Ontology and Information Systems' and 'Properties' including 'abstract', 'address', 'author', and 'keywords'.

Semantic web + EHR (clinical information sharing)



Data Links for Future Healthcare



BIKE

Thank You !

