

SEMANTIC WEB DATA MANAGEMENT

from Web 1.0 to Web 3.0



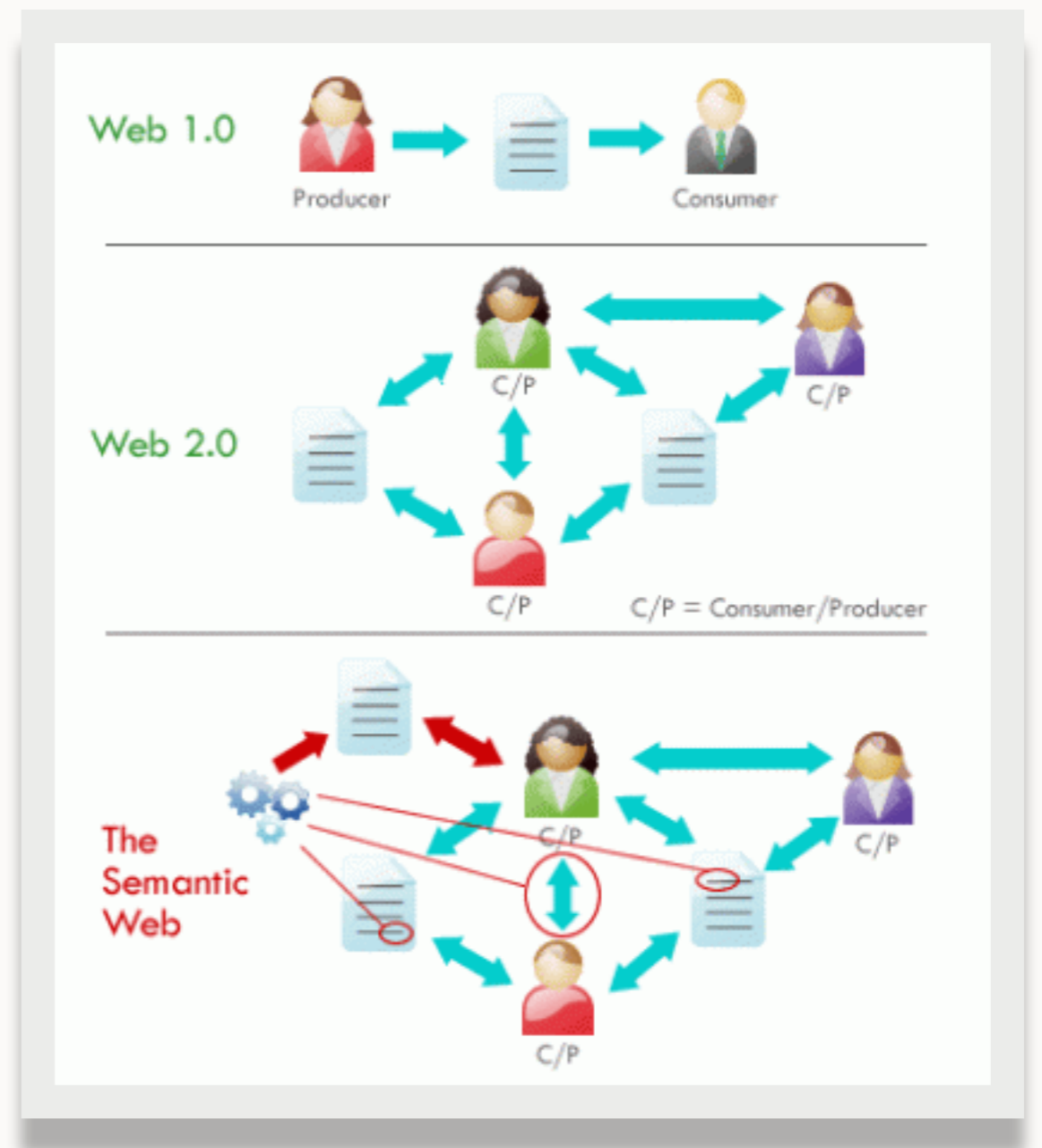
MOTIVATIONS

- Web evolution
- Self-describing Data
- XML, DTD, XSD
- RDF, RDFS, OWL



WEB 1.0, WEB 2.0, WEB 3.0

- ☑ Web 1.0 is a *one-way* platform
- ☑ Web 2.0 is a *two-way* platform where participation is a key-word.
- ☑ Web 3.0 shows *more intelligence*: the "web machine" *learns*, suggests and anticipates what people like and would like to get.



WEB 1.0 : *RECORD STRUCTURES*

- ☑ A flat file is a collection of **records**.
- ☑ A record consists of **fields**.
- ☑ Each record in a flat file has the same number and kinds of fields as any other record in the same file.
- ☑ The **schema** of a flat file describes the structure (i.e., the kinds of fields) of each record.
- ☑ A schema is an example of an **ontology**.

WEB 1.0 : *RECORD STRUCTURES*

Consider the following records in flat file:

```
011500 18.66 0 0 62 46.271020111 25.220010  
011500 26.93 0 1 63 68.951521001 32.651010  
020100 33.95 1 0 65 92.532041101 18.930110  
020100 17.38 0 0 67 50.351111100 42.160001
```

What do they mean?

METADATA: *DATA ABOUT DATA*

The explanation of what data means is called *metadata* or “*data about data*”

For a flat file or database the metadata is called the *schema*

NAME	LENGTH	FORMAT	LABEL
instudy	6	MMDDYY	Date of randomization into study
bmi	8	Num	Body Mass Index.
obesity	3	0=No 1=Yes	Obesity (30.0 <= BMI)
ovrwt	8	0=No 1=Yes	Overweight (25 <= BMI < 30)
Height	3	Num	Height (inches)
Wtkgs	8	Num	Weight (kilograms)
Weight	3	Num	Weight (pounds)
метарц	3	ИЛШ	метарц (болуцс)
мфкдс	8	ИЛШ	метарц (кѣјодѣшс)
неѣарц	3	ИЛШ	неѣарц (ѣуорес)

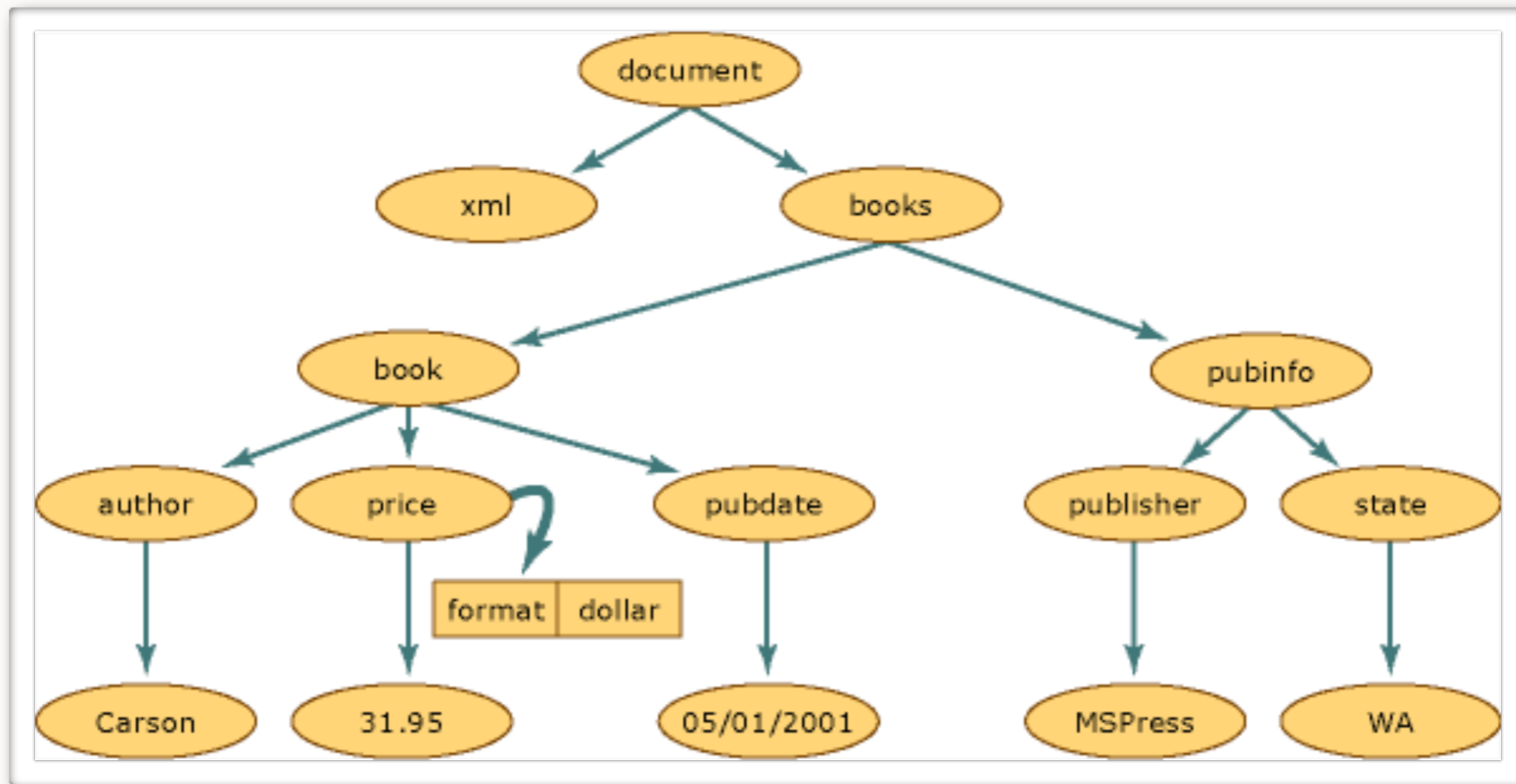
WEB 2.0: *SELF-DESCRIBING DATA*

The eXtensible Markup Language (XML)

- XML is a format for representing data.
- XML goes beyond flat files by allowing elements to contain other elements, forming a hierarchy.

XML	FLAT Files
Element	Record
Attribute	Field
DTD	Schema

HIERARCHICAL ORGANIZATION



THE MEANING OF A HIERARCHY

Hierarchies can be based on many principles:

subclass (subset)

instance (member)

more complex relationships

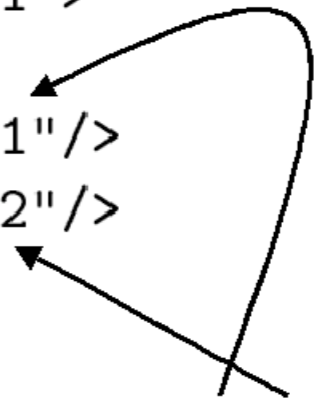
Hierarchies to be based on several principles at the same time.

XML hierarchies cannot represent these more general forms of hierarchy.

NON-HIERARCHICAL RELATIONSHIPS

- ✓ Hierarchical relationships are represented by one element contained inside another one.
- ✓ Non-hierarchical relationships are represented using reference attributes, such as the two arrows in the diagram.
- ✓ Containment and reference are very different in XML.

```
<molecule id="m1">  
  <atomArray>  
    <atom id="a1"/>  
    <atom id="a2"/>  
  </atomArray>  
  <bondArray>  
    <bond atomRefs2="a1 a2"/>  
  </bondArray>  
</molecule>
```



XML SEMANTICS

The infoset contains two kinds of relationship:

Unlabeled hierarchical relationship link

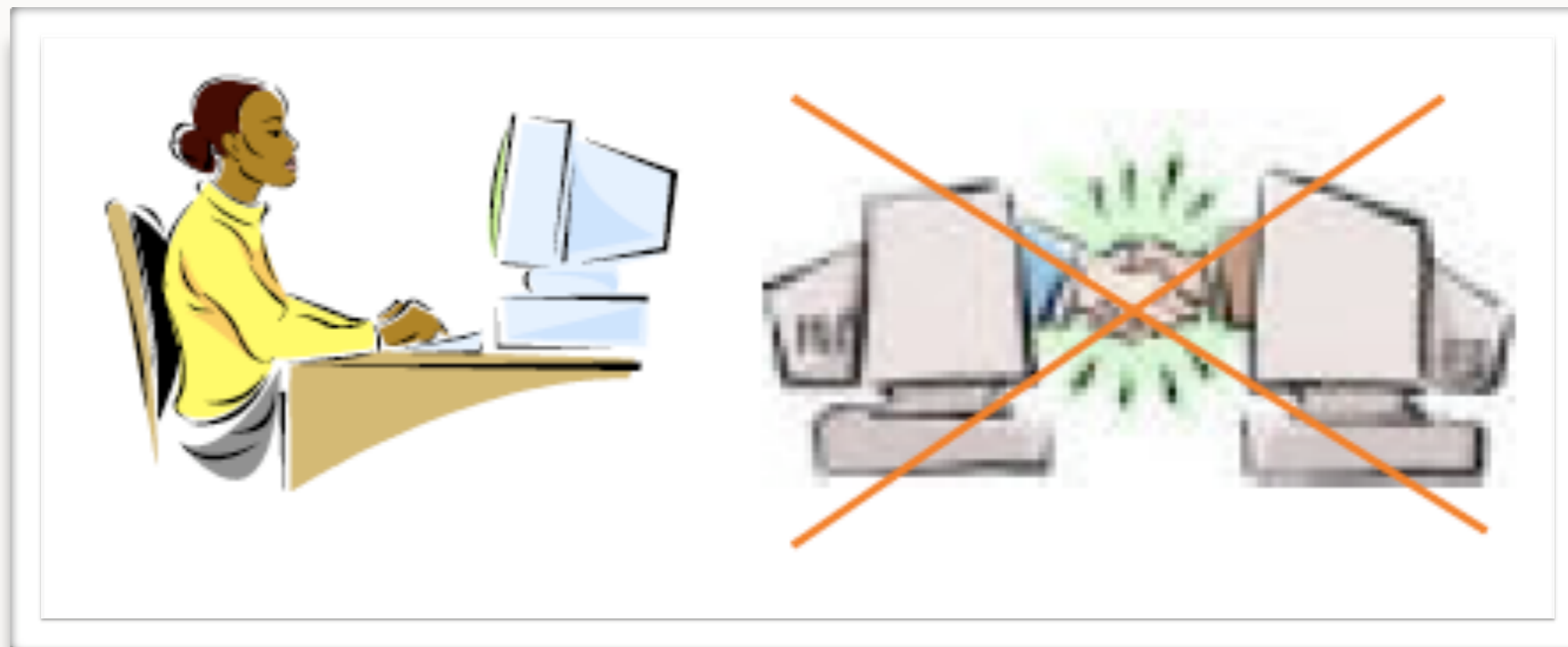
Labeled attribute link

The order of attributes does not matter. The infoset is the same no matter how they are arranged.

The order of hierarchical links does matter. The infoset is different if the elements are in a different order.

...LIMITATIONS OF THE WEB TODAY

The Web activities are mostly focus on **Machine-to-Human**, and **Machine-to-Machine** activities are **not** particularly **well supported** by software tools.



WEB 3.0: SEMANTIC WEB



THE SEMANTIC WEB

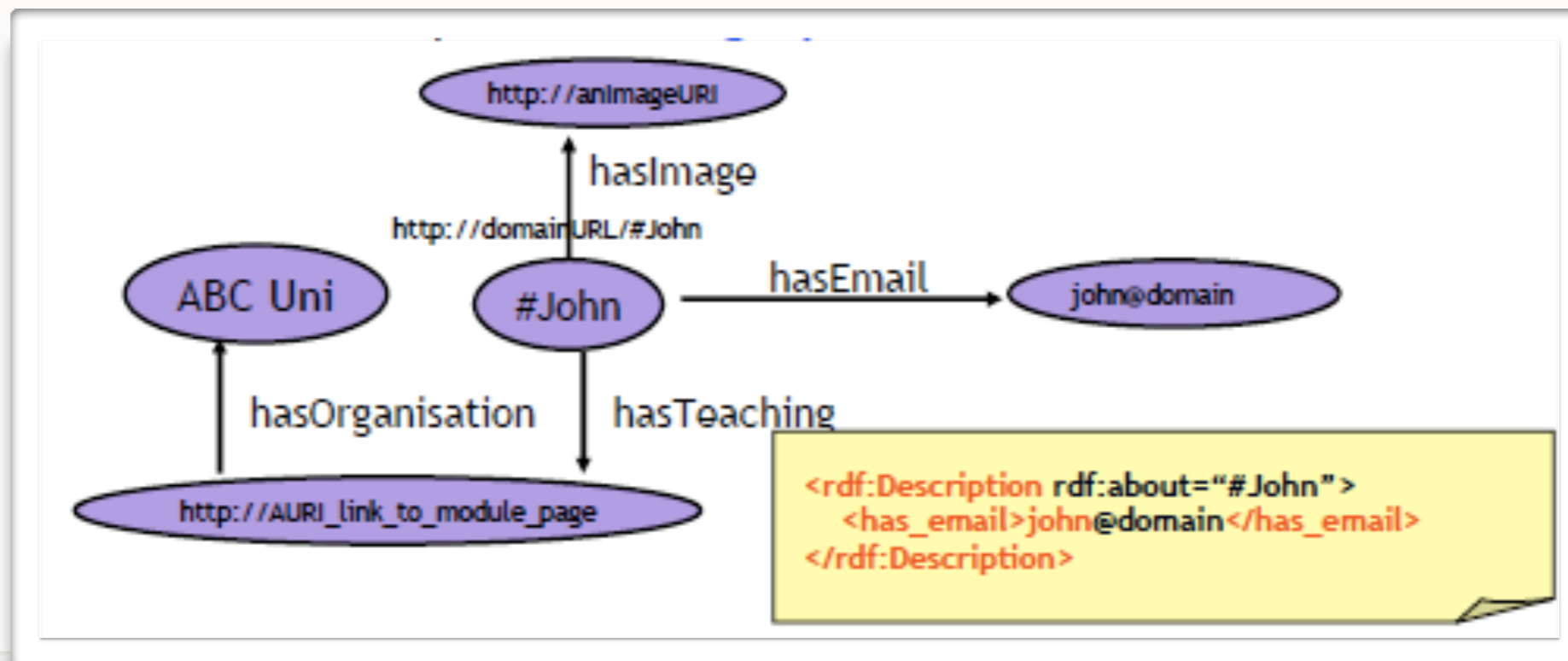
A new form of Web content
that is meaningful to computers
will unleash a revolution of new abilities

by
TIM BERNERS-LEE,
JAMES HENDLER and
ORA LASSILA

PHOTO CREDIT HERE

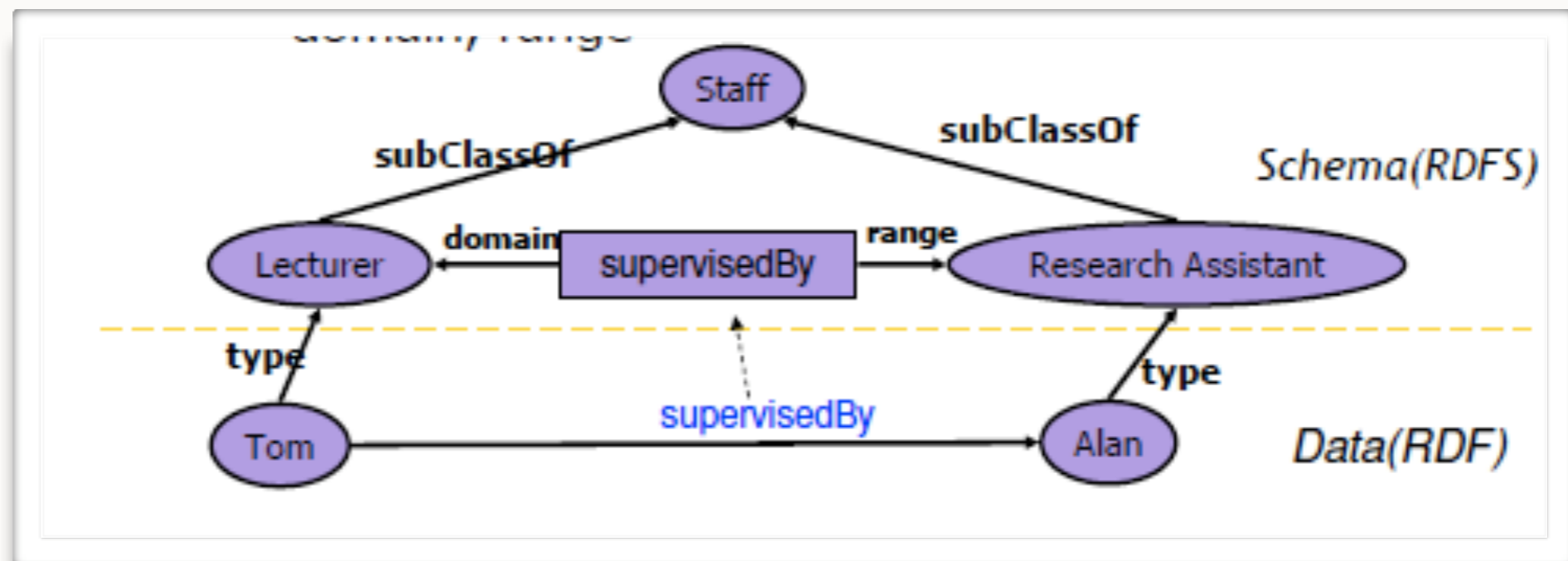
RDF FOR SEMANTIC ANNOTATION

- ✓ RDF provides metadata about Web resources
- ✓ **<subject, predicate, object>** (i.e Object -> Attribute-> Value triples)
- ✓ It has an **XML** syntax
- ✓ Chained triples form a **graph**

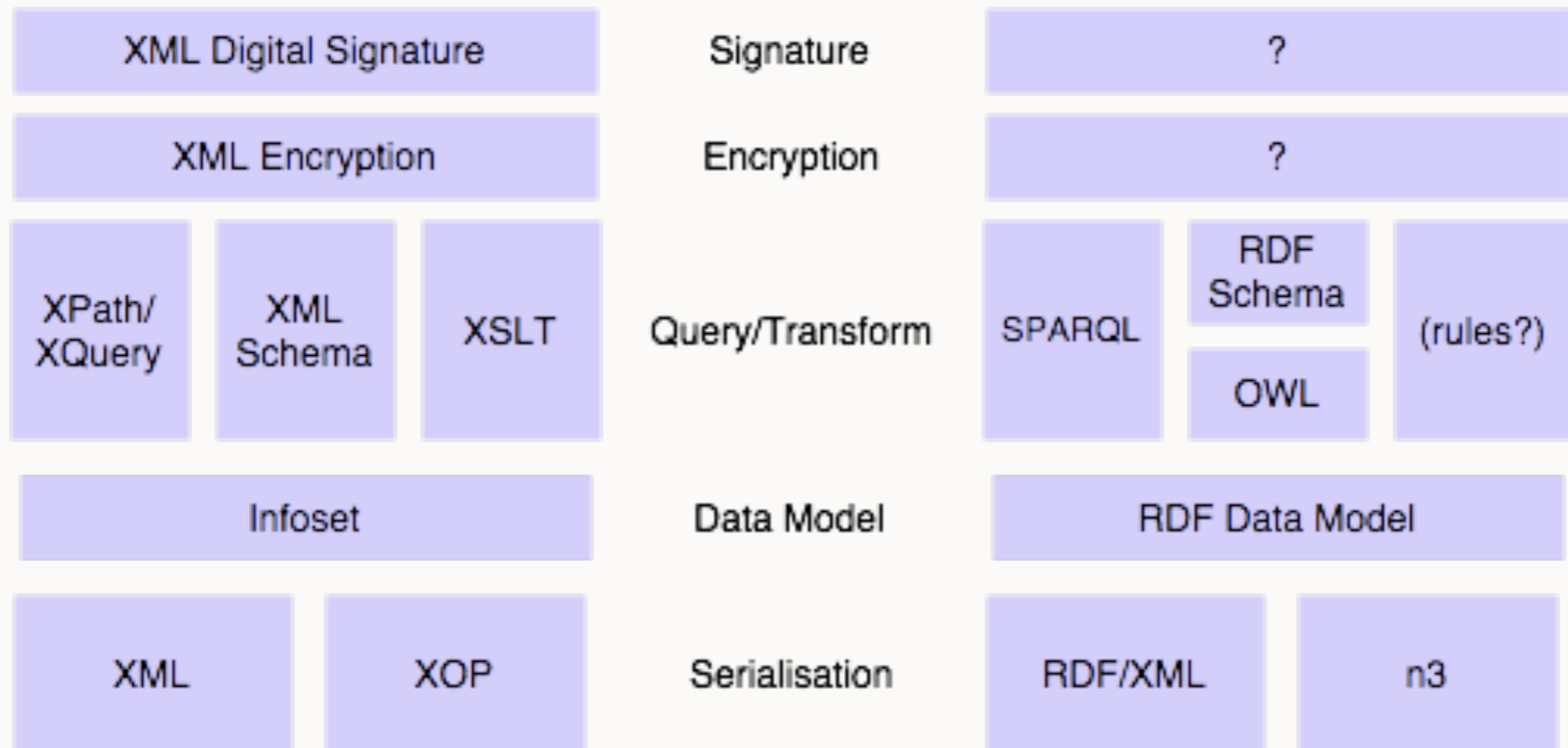


RDFS AND OWL

- ☑ Defines vocabulary for RDF
- ☑ Organizes this vocabulary in a typed hierarchy
 - Class, subClassOf, type
 - Property, subPropertyOf
 - domain, range



XML VS RDF

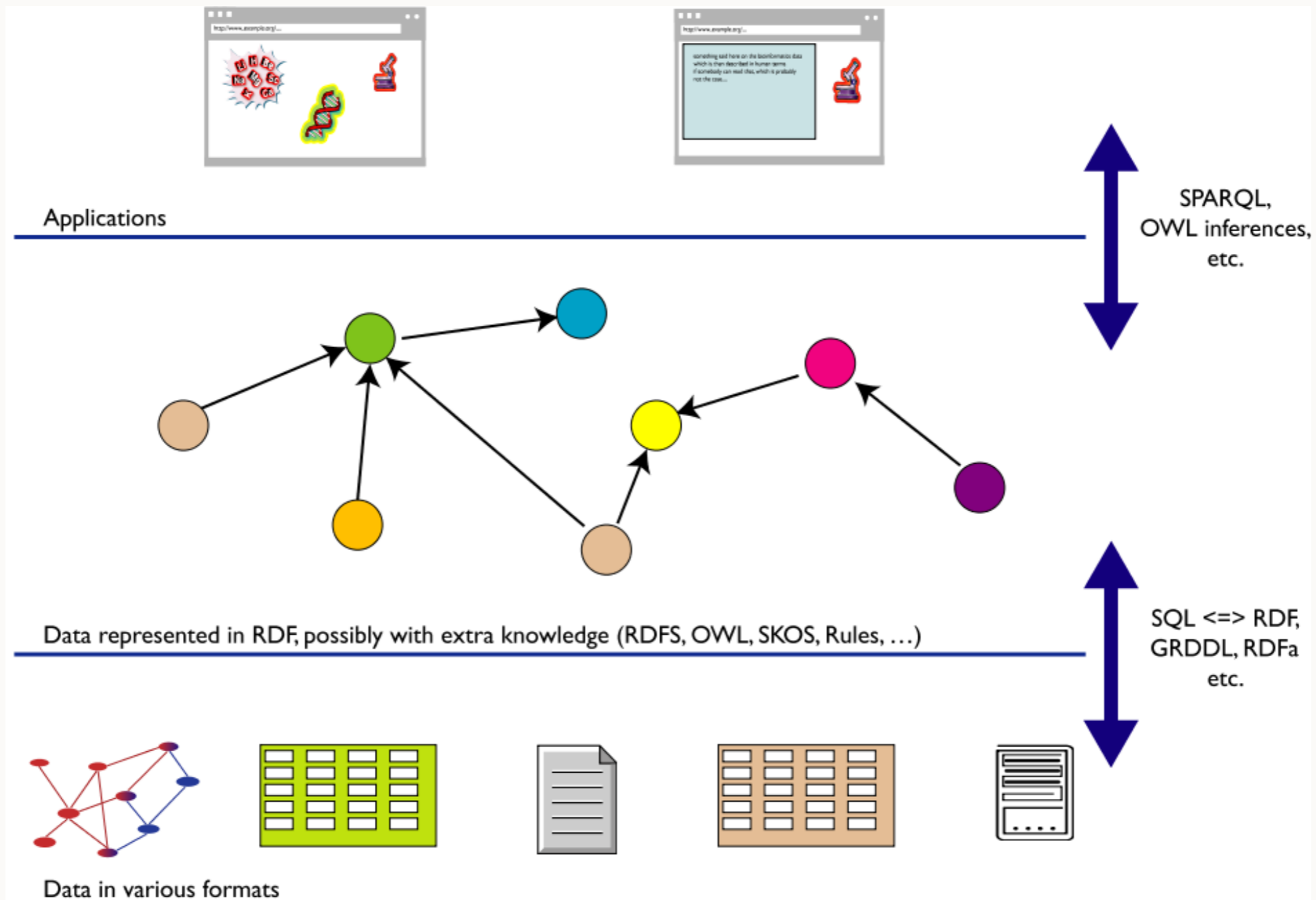


OPEN PROBLEMS

- Data Storage
- Data retrieval
- Data Visualization



RELEVANT AMOUNT OF SEMANTIC DATA



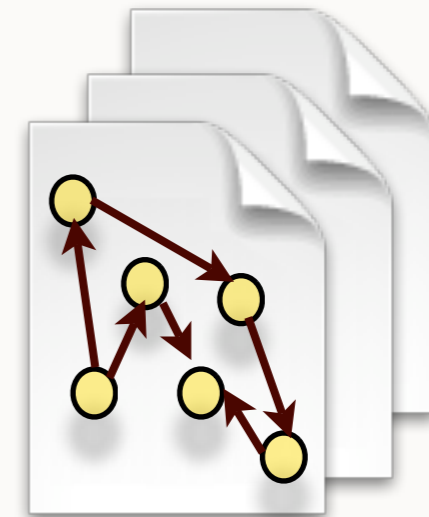
DATA STORAGE



WEB



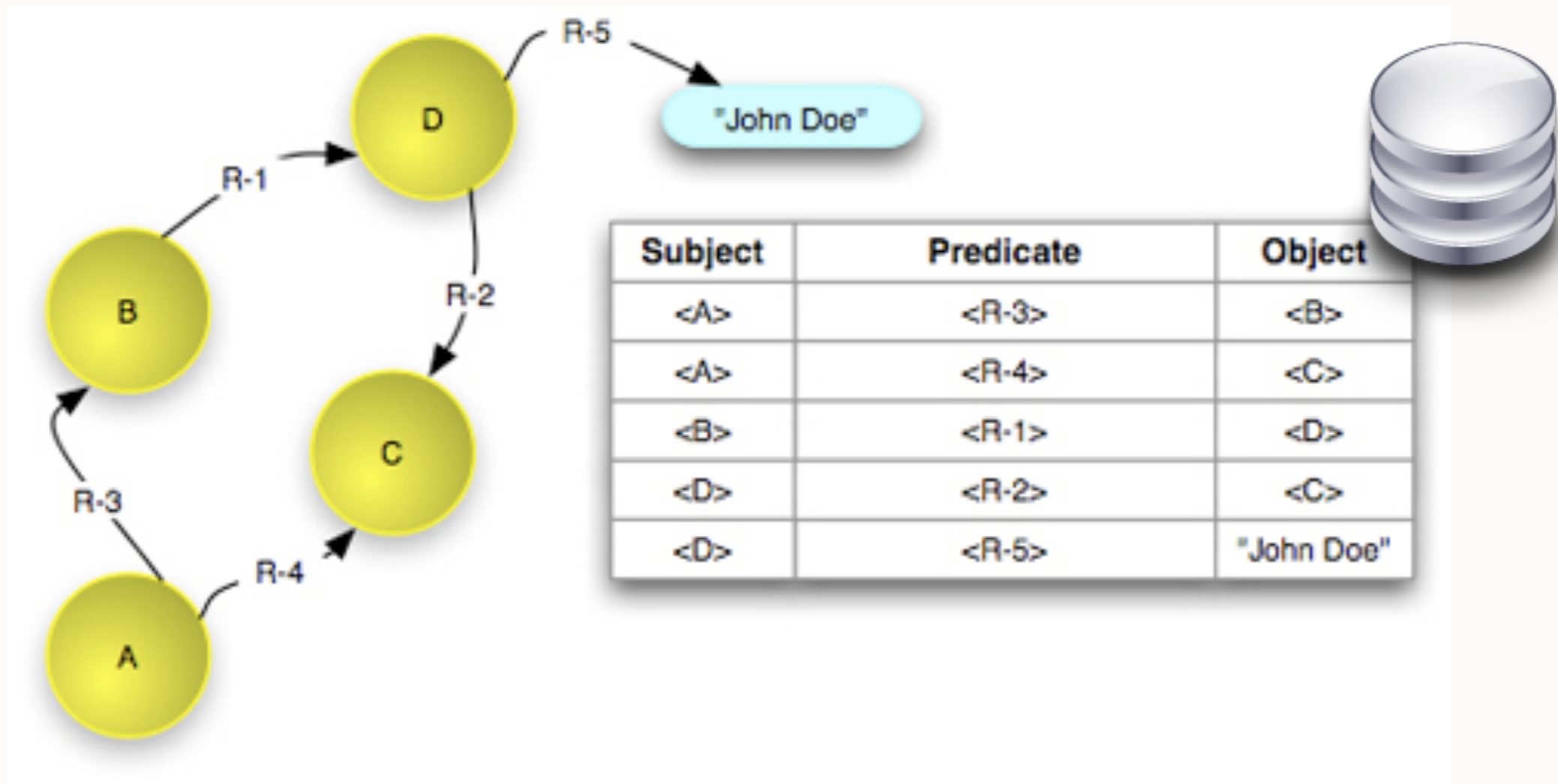
RDF



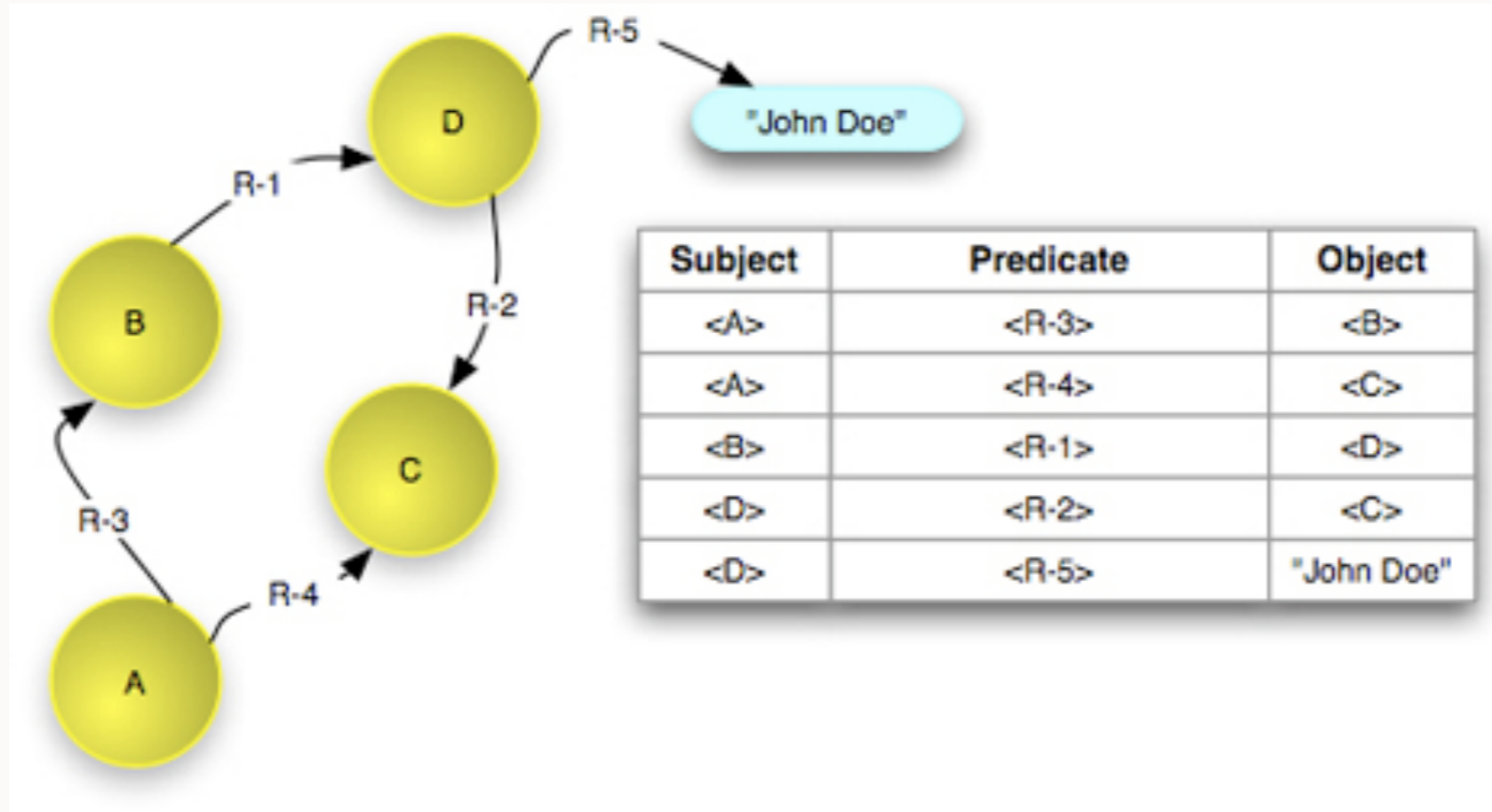
DBMS



DATA STORAGE

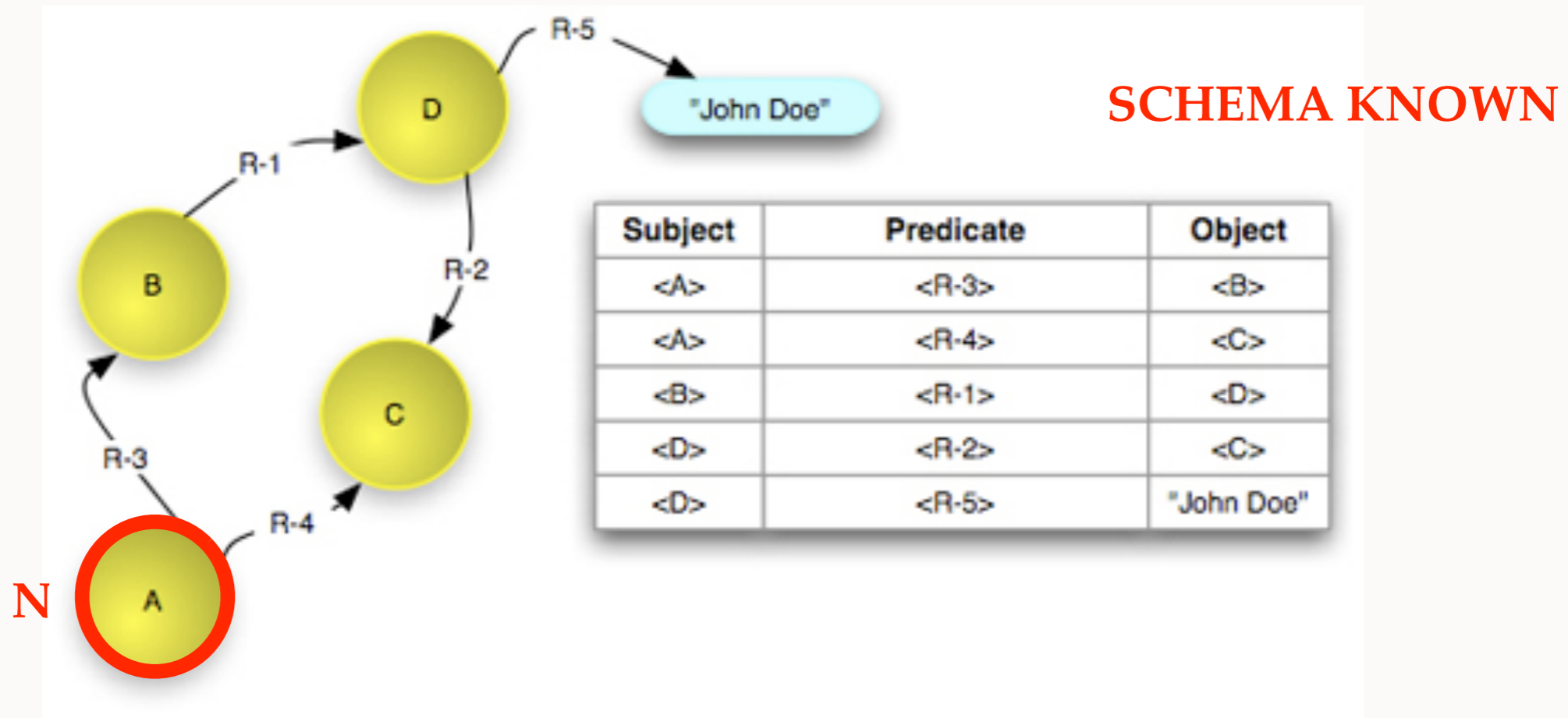


DATA (INFORMATION) RETRIEVAL



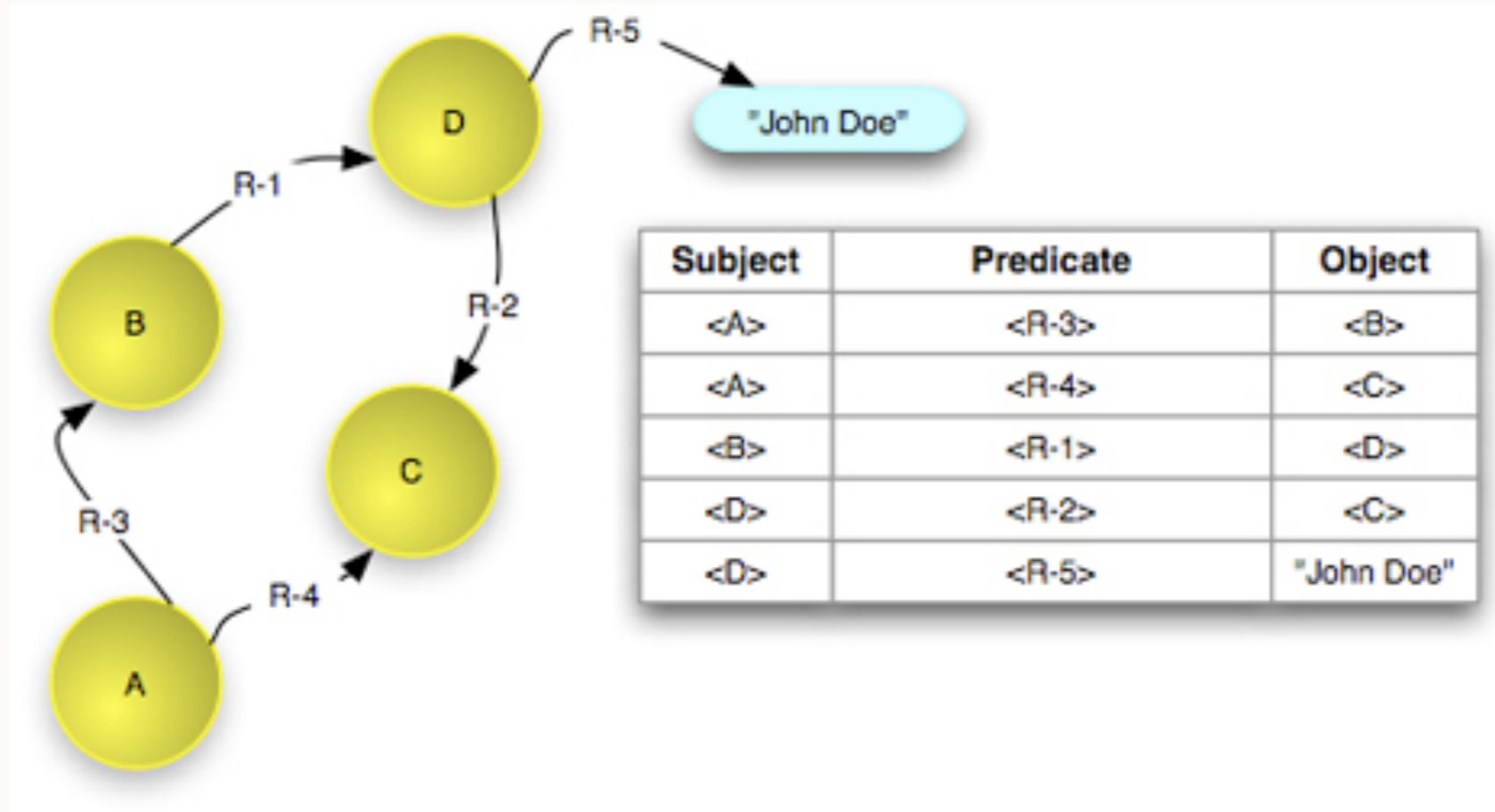
Query 1: "All nodes **N** having out-coming predicates into **B** and **C** at least"

DATA (INFORMATION) RETRIEVAL



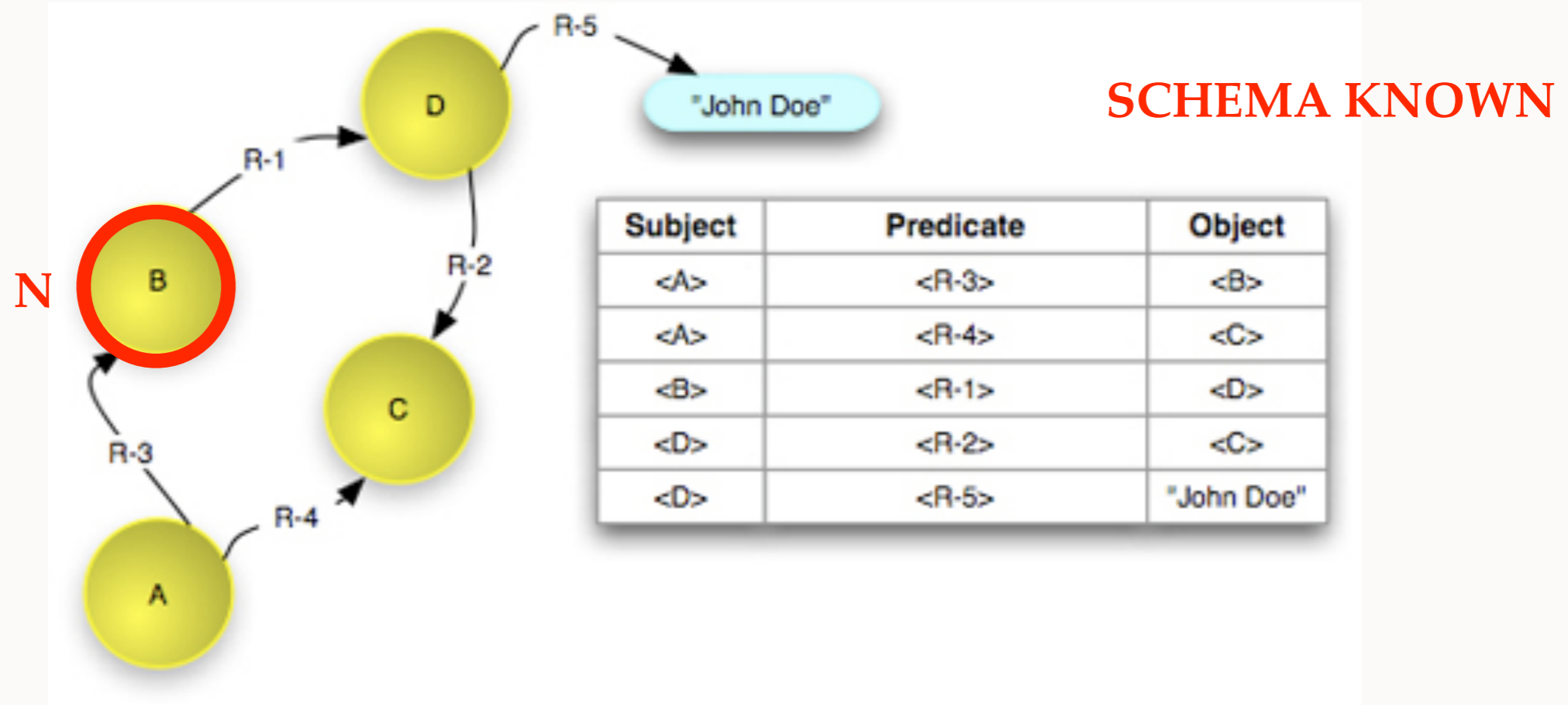
```
SELECT T1.subject As N
FROM triples T1, triples T2
WHERE T1.object = 'B' AND T2.object = 'C' AND
      T1.subject = T2.subject
```

DATA (INFORMATION) RETRIEVAL



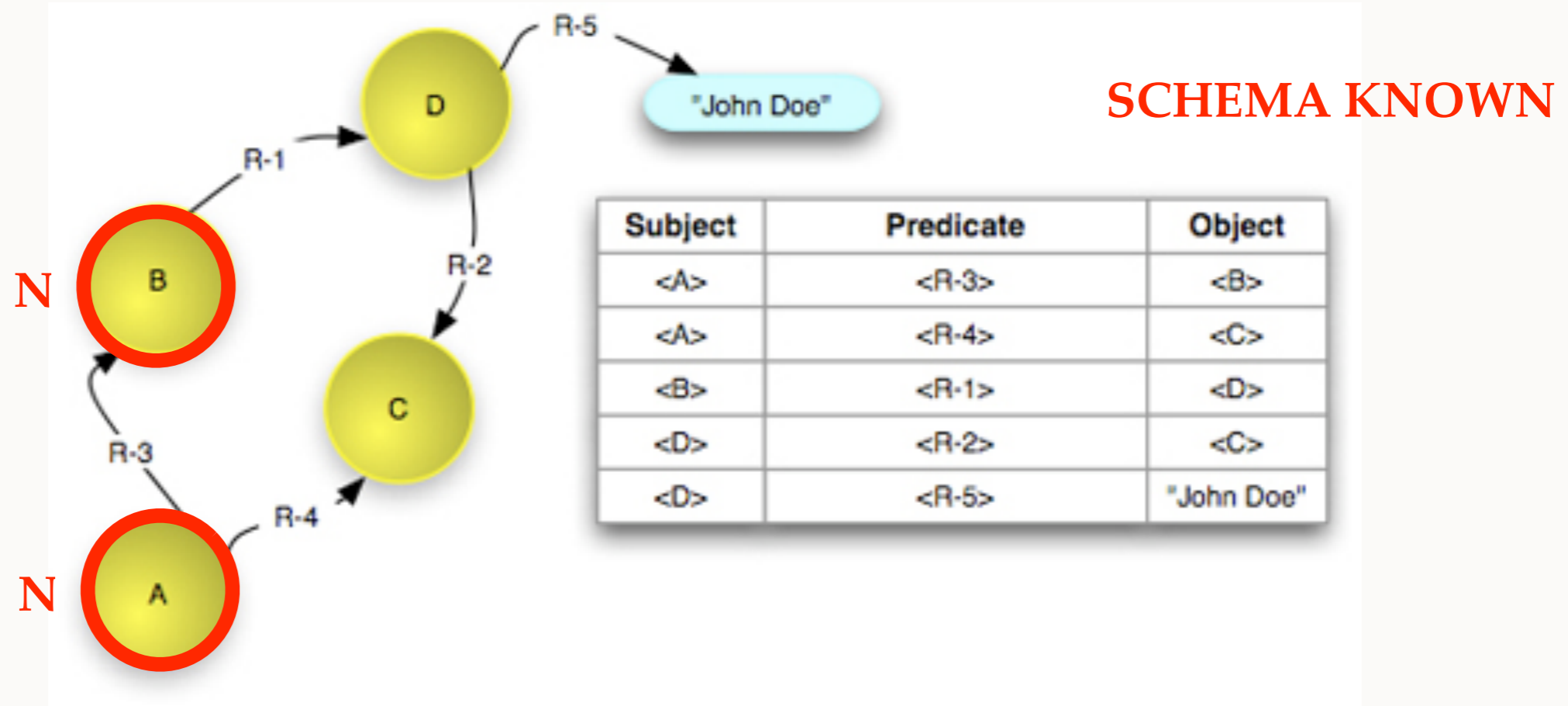
Query 2: "All nodes **N** having a relation into **D**"

DATA (INFORMATION) RETRIEVAL



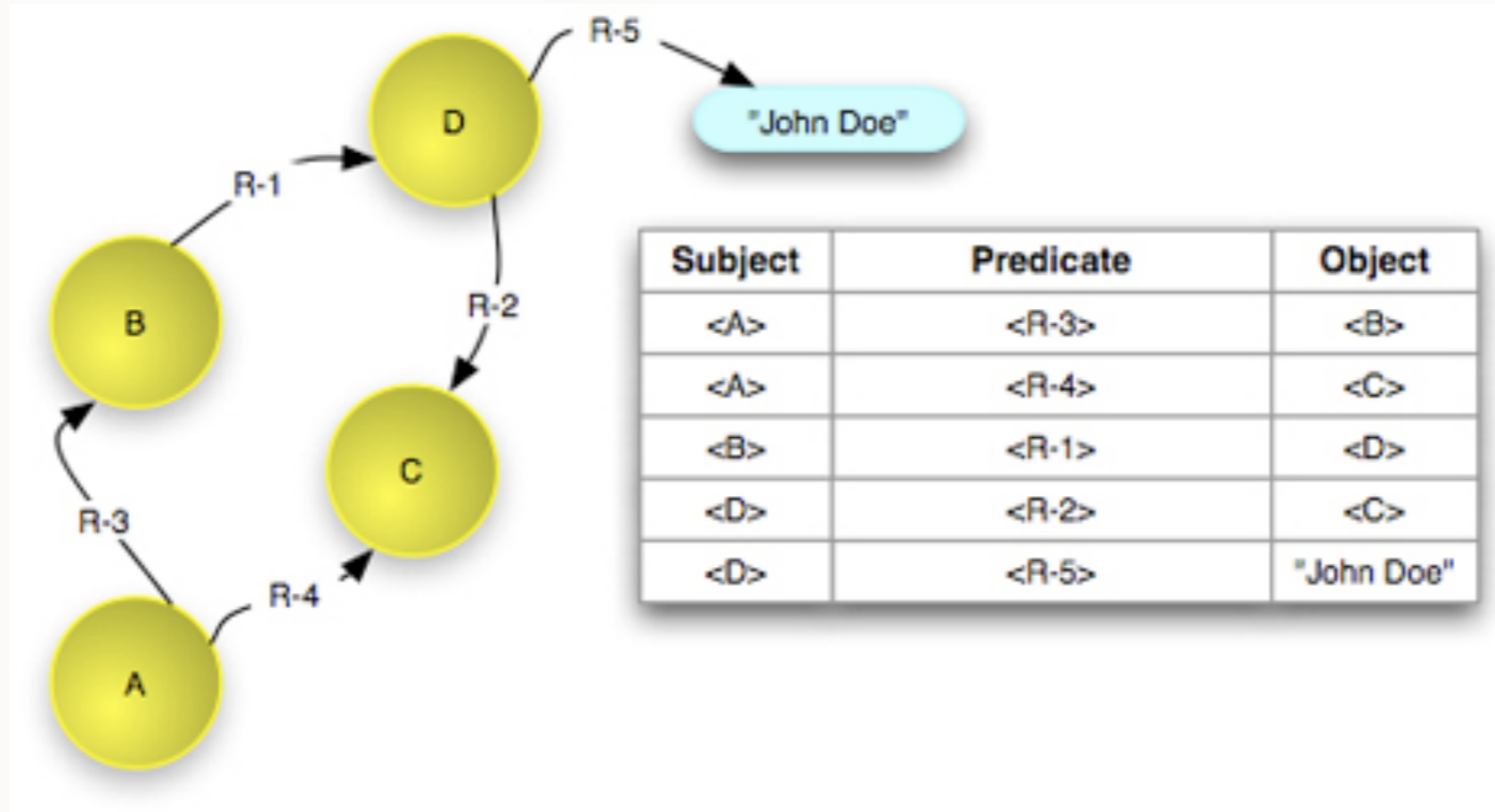
```
SELECT T.subject As N
FROM triples T
WHERE T.object = 'D'
```

DATA (INFORMATION) RETRIEVAL



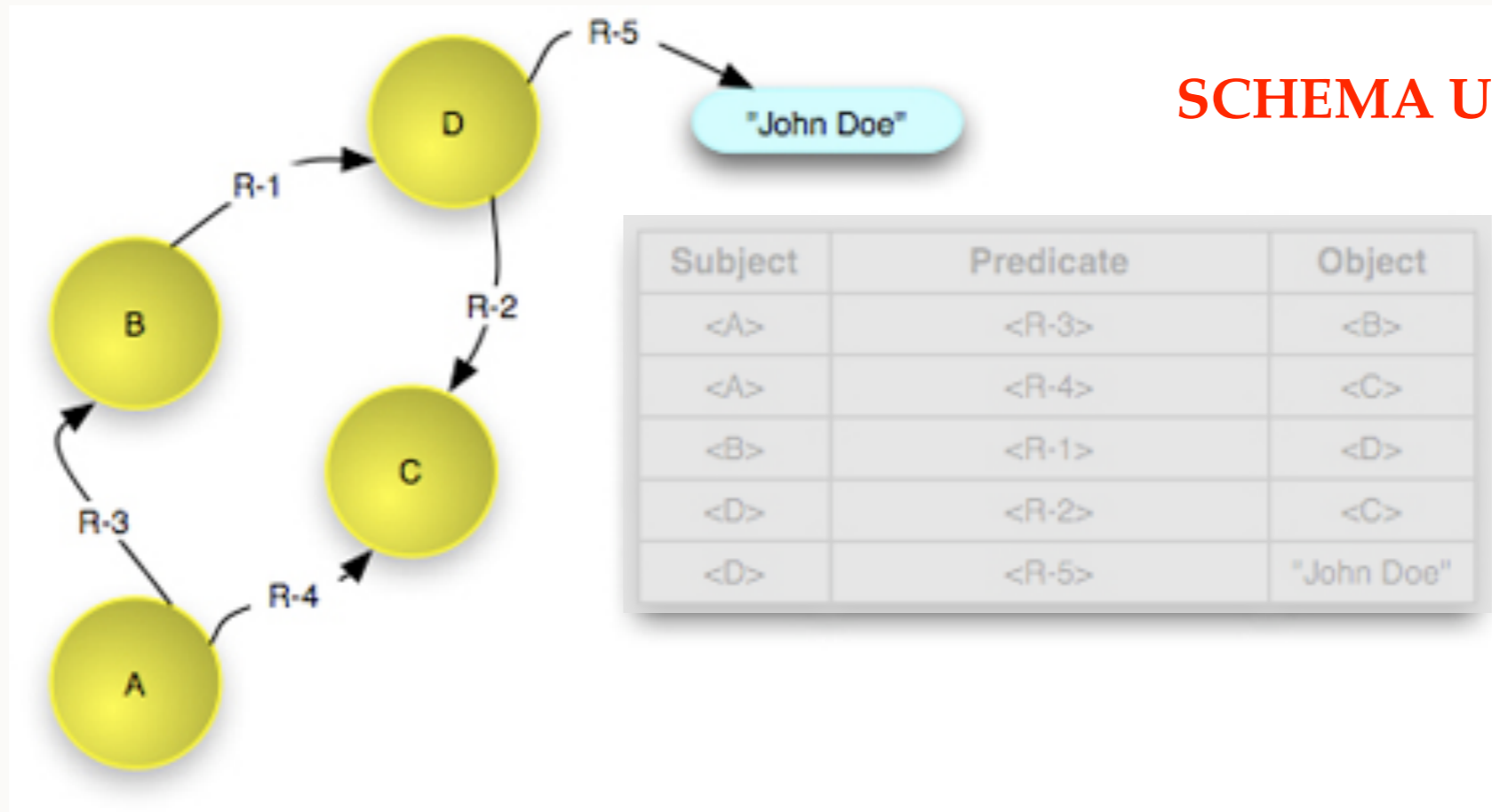
```
SELECT T.subject As N
FROM triples T
WHERE T.object = 'D'
```


KEYWORD SEARCH



Query 3: "D R-1 John Doe"

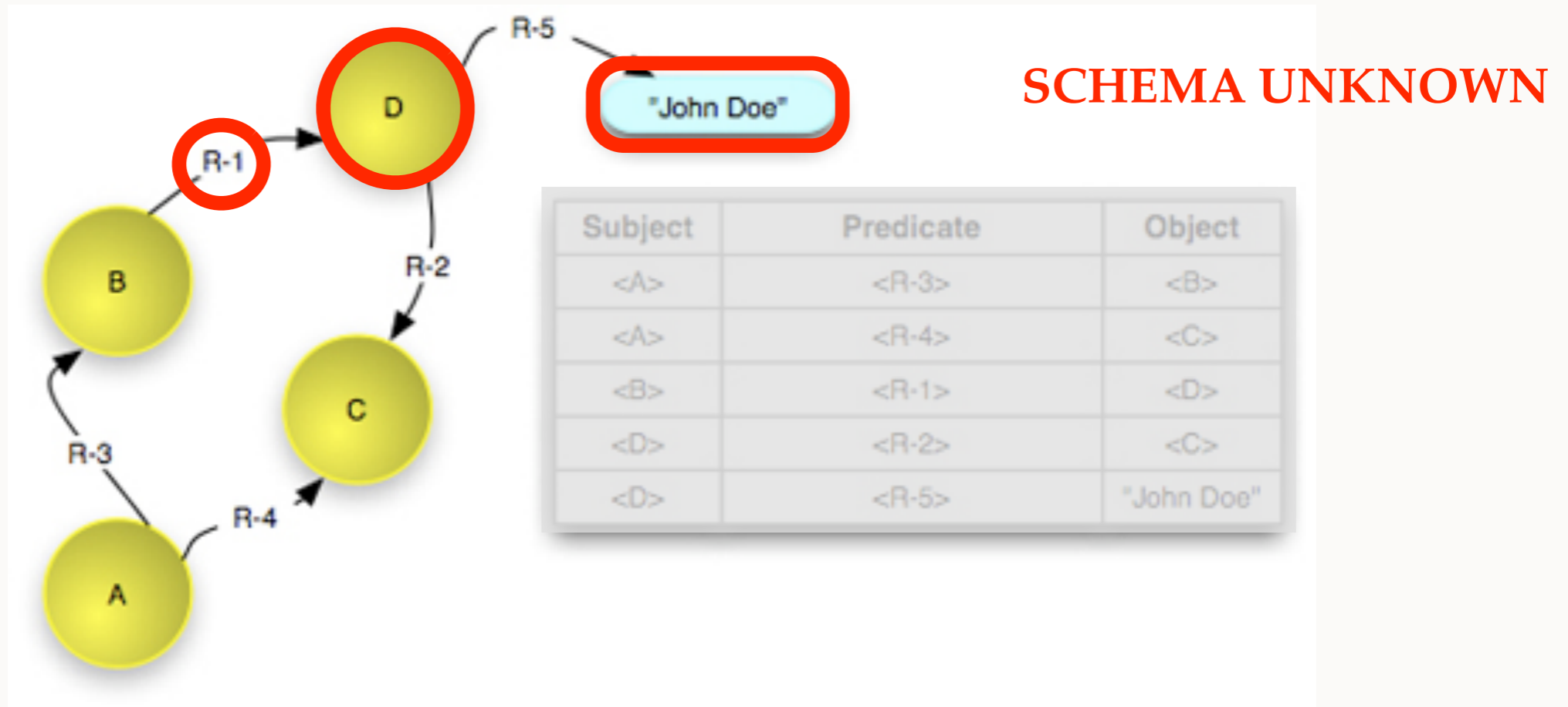
KEYWORD SEARCH



SCHEMA UNKNOWN

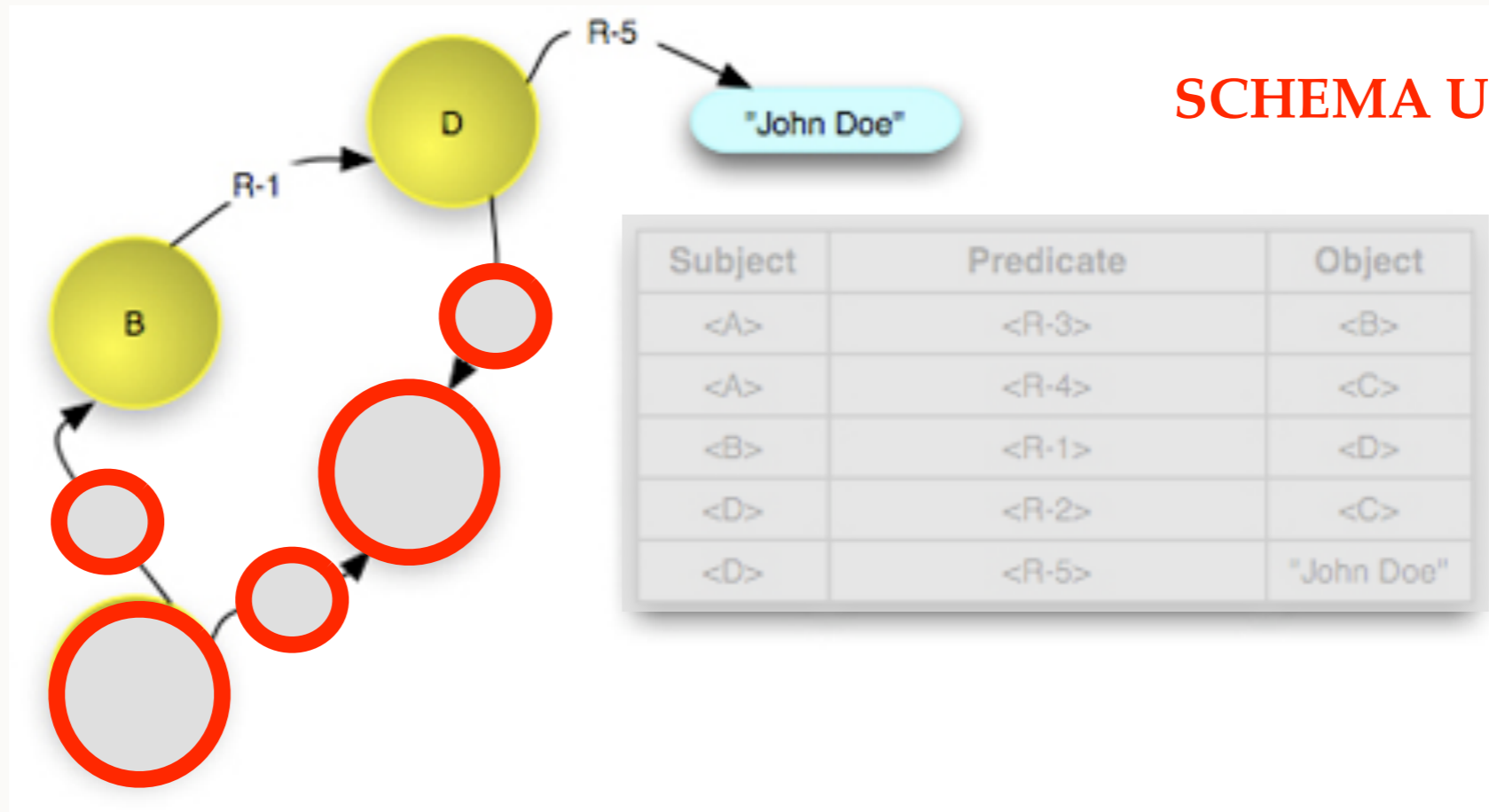
Query 3: "D R-1 John Doe"

KEYWORD SEARCH



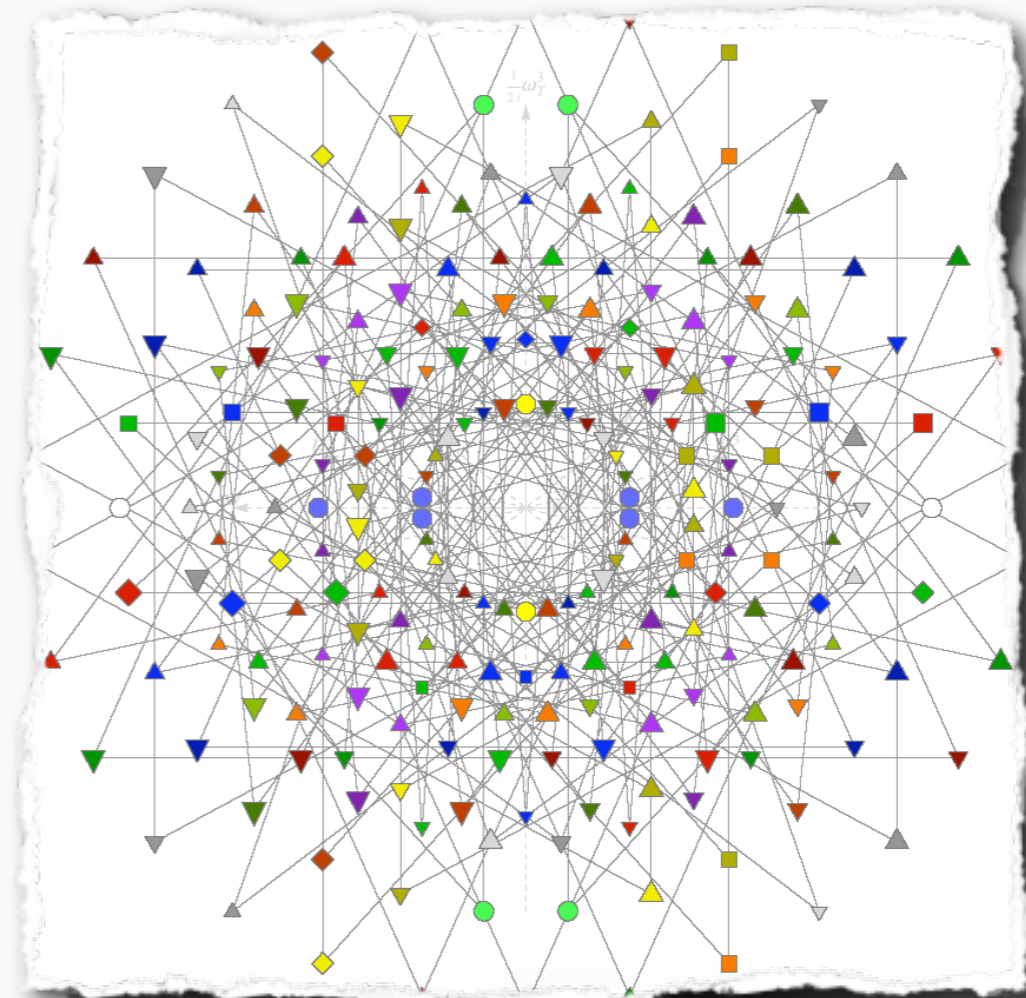
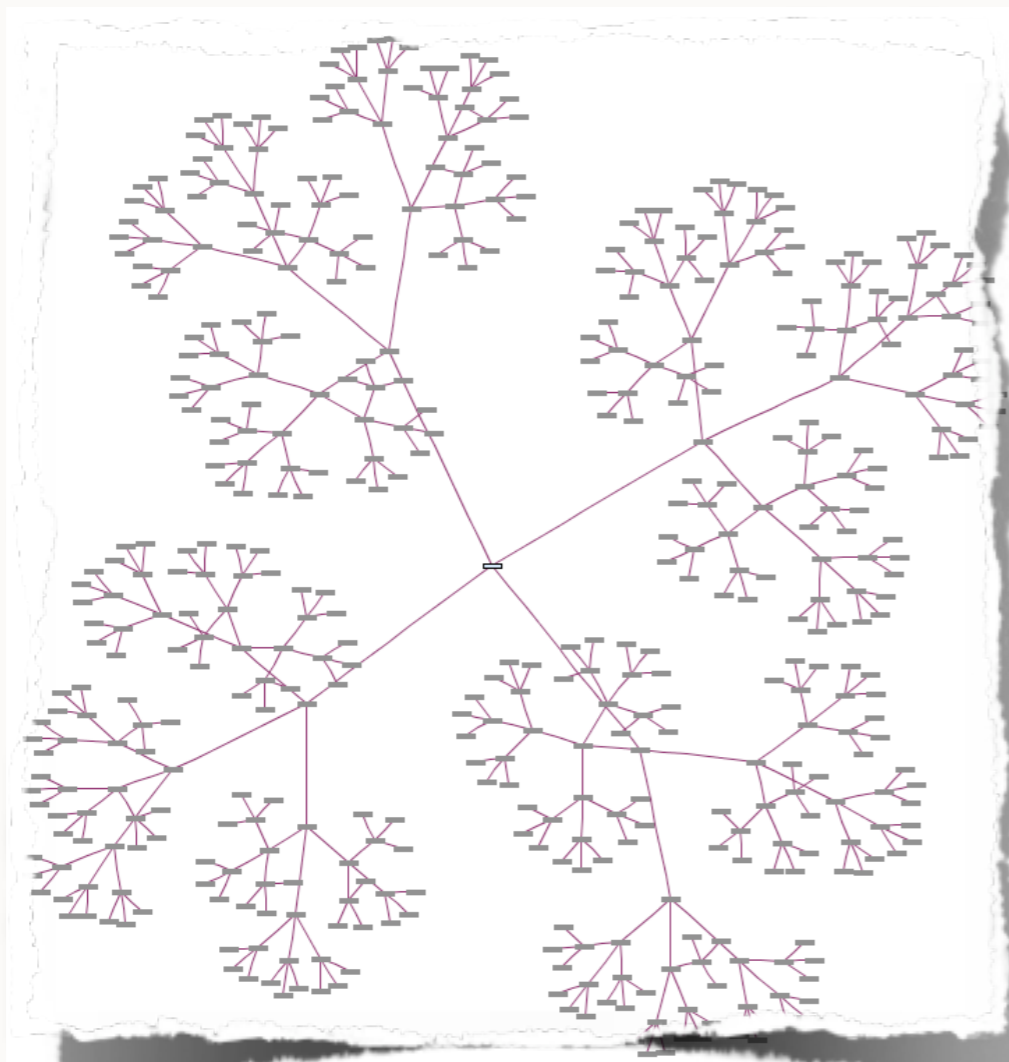
Query 3: "D R-1 John Doe"

KEYWORD SEARCH



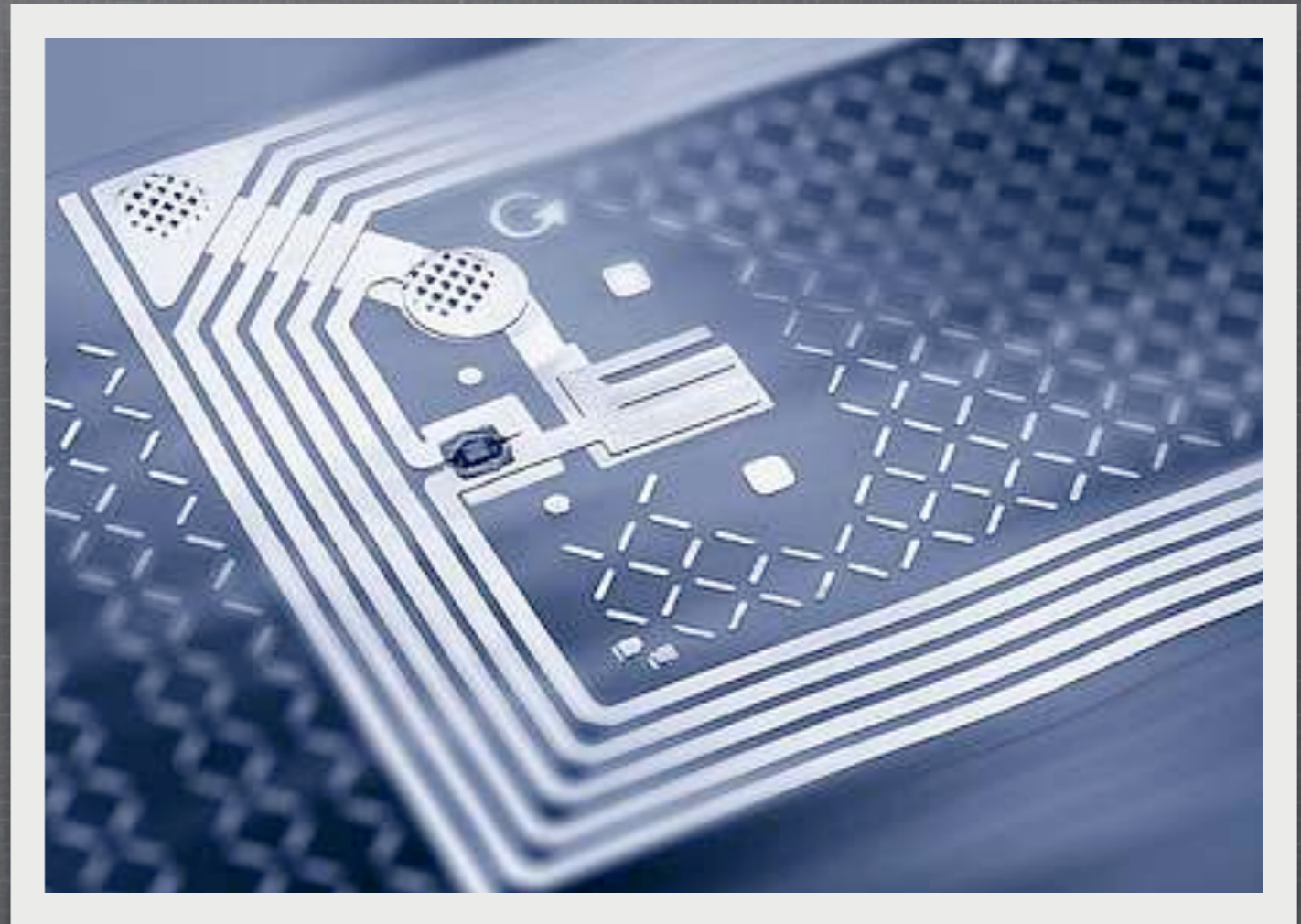
Query 3: "D R-1 John Doe"

DATA VISUALIZATION



APPLICATION SCENARIOS

- Data Extraction
- Semantic RFID
- Semantic Web Services



WEB DATA EXTRACTION BY SEMANTIC ANNOTATION

Biblio description

Title

RDF Semantics - W3C Recommendation 10 February 2004

Author

Patrick Hayes see [homepage](#)

WEB DATA EXTRACTION BY SEMANTIC ANNOTATION

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml"
  xml:base="http://www.dc4plus.com/references/rdf_sem.html"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" >

<head profile="http://ns.inria.fr/grddl/rdfa/">
  <title>Biblio description</title>
</head>

<body>
  <h1>Biblio description</h1>
  <dl about="http://www.w3.org/TR/2004/REC-rdf-mt-20040210/">
    <dt>Title</dt>
    <dd property="dc:title">RDF Semantics - W3C Recommendation 10 February 2004</dd>
    <dt>Author</dt>
    <dd rel="dc:creator" href="#a1">
      <span id="a1">
        <link rel="rdf:type" href="[foaf:Person]" />
        <span property="foaf:name">Patrick Hayes</span>
        see <a rel="foaf:homepage" href="http://www.ihmc.us/users/user.php?UserID=42">homepage</a>
      </span>
    </dd>
  </dl>
</body>

</html>
```

he is a Person

it is a title

it is a homepage

WEB DATA EXTRACTION BY SEMANTIC ANNOTATION

Biblio description

Title

RDF Semantics - W3C Recommendation 10 February 2004

Author

Patrick Hayes see [homepage](#)

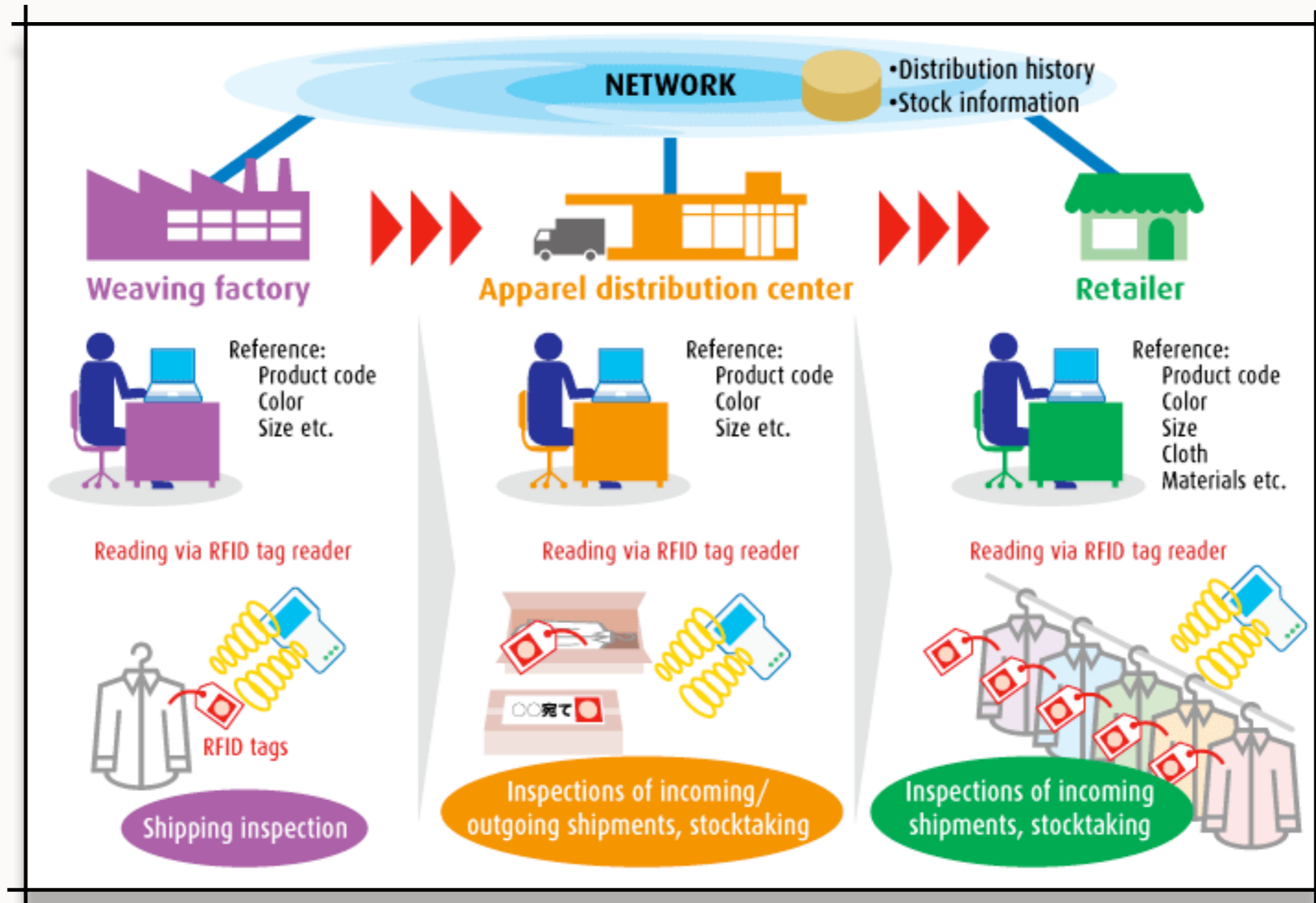


name	homepage
<i>Patrick Hayes</i>	<i>http://...</i>

title	creator
<i>RDF Semantics - W3C ...</i>	<i>Patrick Hayes</i>



RFID: RADIO FREQUENCY IDENTIFICATION



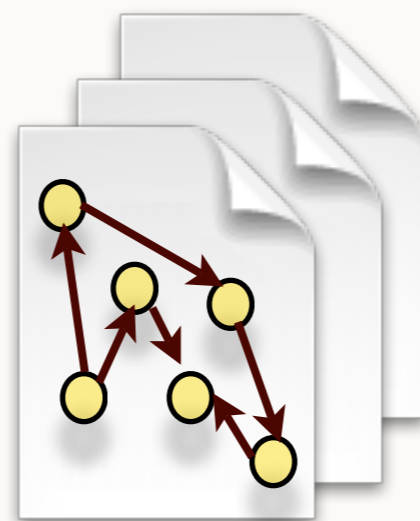
SEMANTIC RFID

FLAT REPRESENTATION

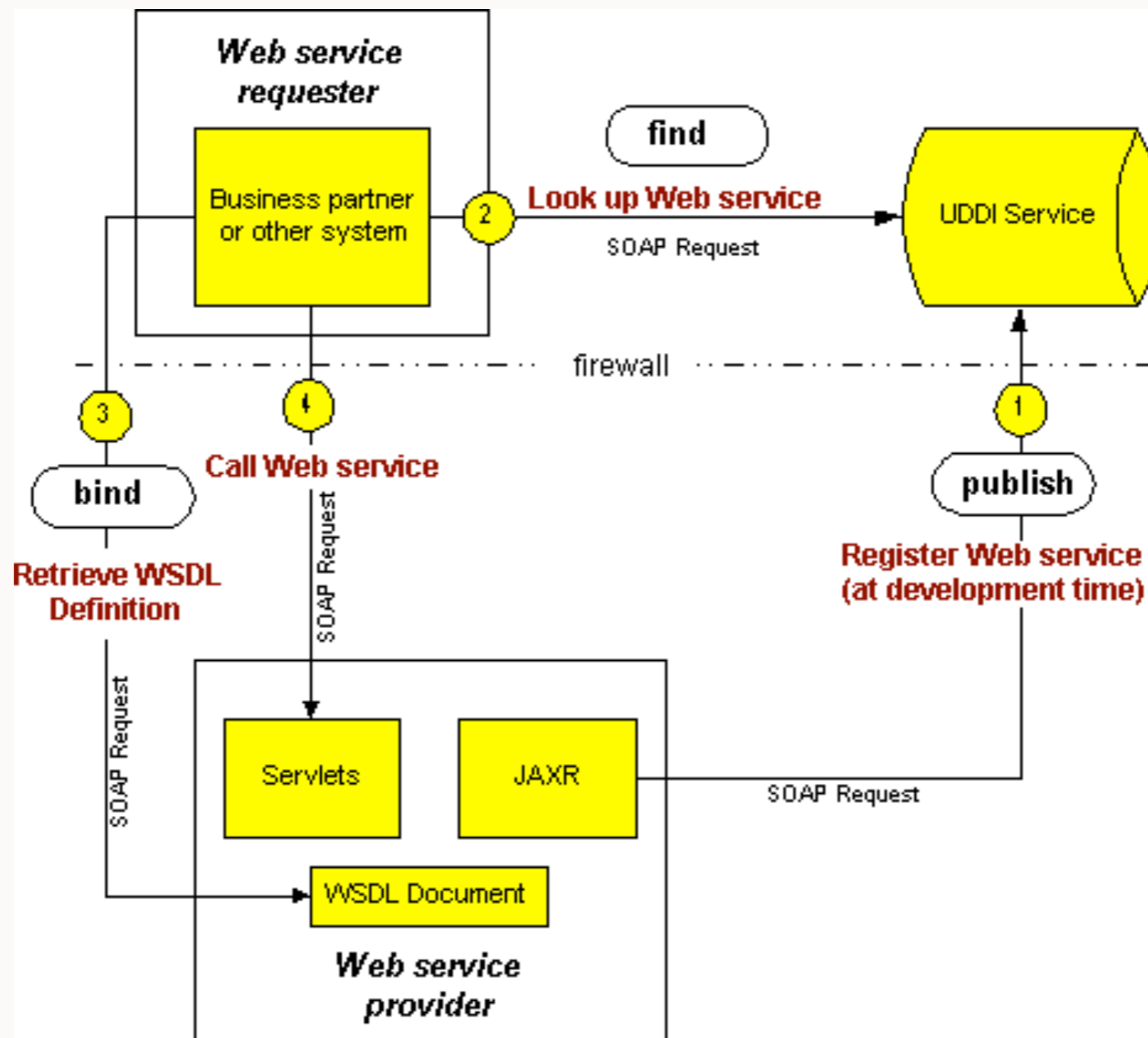


EPC	Location	time
<i>ID1</i>	<i>STORE1</i>	<i>2005-10-30 T 10:45 UTC</i>
<i>ID2</i>	<i>STORE2</i>	<i>2005-10-30 T 11:55 UTC</i>
<i>ID3</i>	<i>STORE3</i>	<i>2005-10-30 T 12:45 UTC</i>
---	---	---

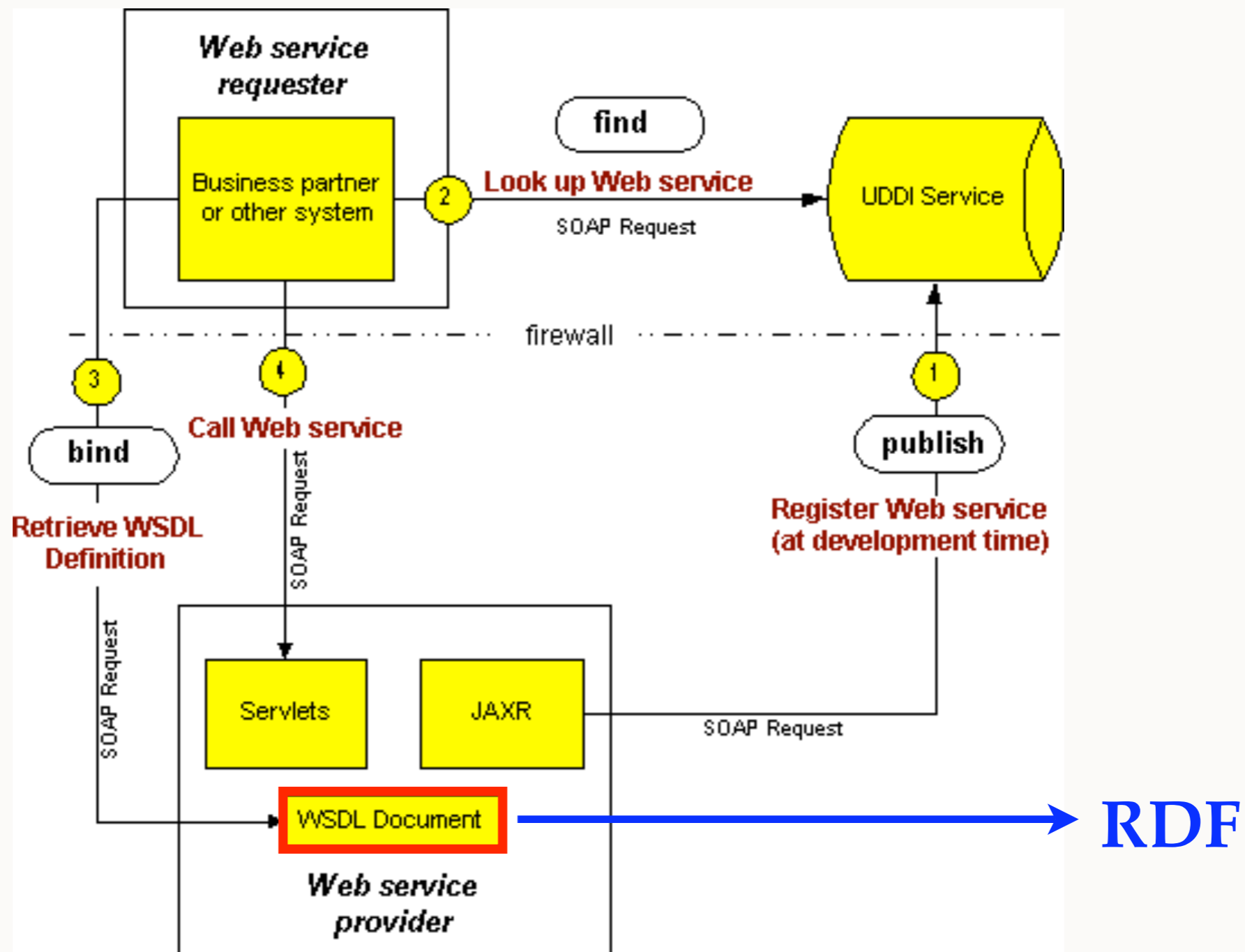
RDF



SEMANTIC WEB SERVICES



SEMANTIC WEB SERVICES



CONTACTS

Roberto De Virgilio

Dipartimento di Informatica e Automazione
Laboratorio Basi di Dati - Room 219

Tel: +39-06-57333229

Fax: +39-557-3030

Email rde79@yahoo.com

... THANKS

