An ontology of resources for Linked Data

Harry Halpin and Valentina Presutti

LDOW @ WWW2009 - Madrid, April 20th

Outline

- Premises and background
- Proposal overview
- Some details of IRW ontology
- Simple application in the LOD context
- Next steps

Premises

- General issue on identity and reference on the Web
  - We want to “handle” things that are not *web accessible*
- Practical issues in applications such as LOD
  - use of owl:sameAs, less contradictions, …
- We need a solution that goes beyond solving the current urgent issues, but it has to be easy
- Is there a low hanging fruit?
  - A lightweight ontology with Web and SW architecture core concepts and relations → IRW
  - Aligned modules for each specific application such as LOD → lod2irw
- Immediate benefit
  - We will argue by referring to one model i.e. IRW
    - means we have agreed on at least few things
  - Existing applications such as LOD can take immediate actions
    - E.g. self-describing resources in LOD

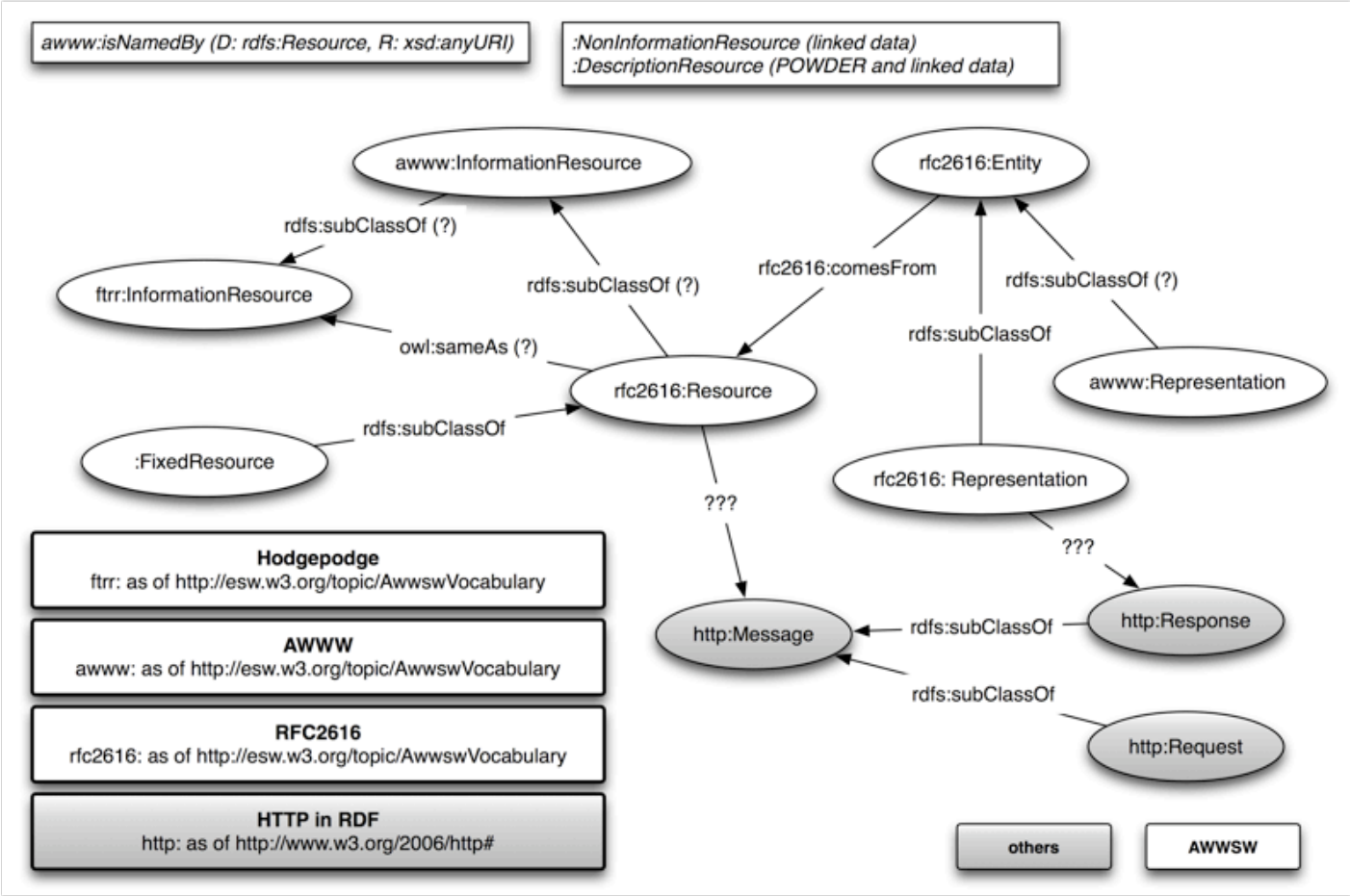
The background

- The problem of identity and reference on the web has been discussed for long in Web circles and outside them
  - E.g., TAG related awwsw is very active on this topic
- Standard documents
  - W3C TAG, RFC 2396, HTTP RFC, URI RFC 3986, IETF RFC
- Scientific literature
  - I3 workshop series, IJSWIS 4(2), etc..
- Notes and mailing list archives
  - E.g. TimBL's "Generic Resource", awwsw ml archive etc.
- Ontologies
  - RDFS, DUL + IOL, LMM, IRE, GEN…
- Projects and initiatives
  - OKKAM, LOD, etc..

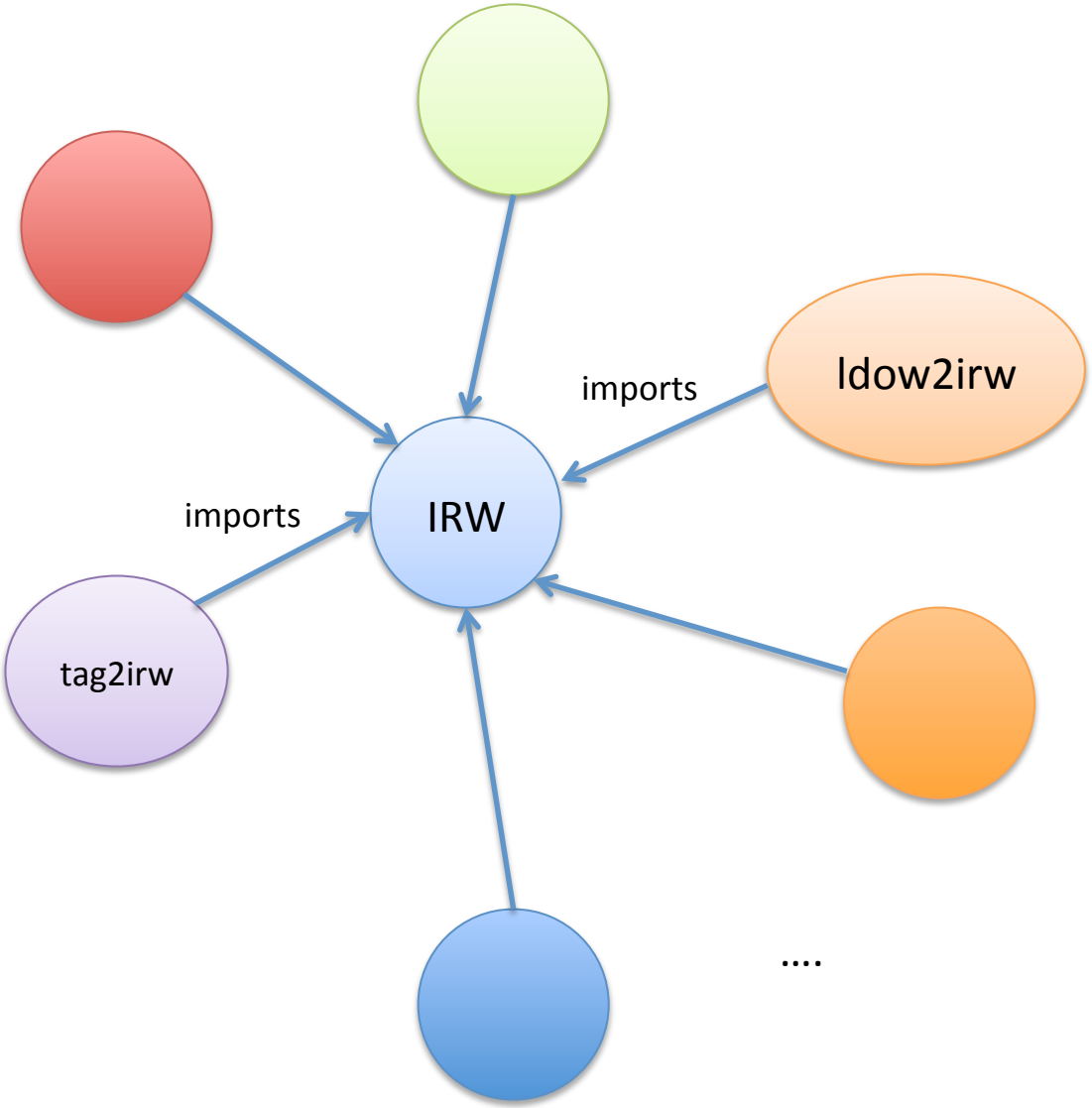
Towards a solution

- A simple *hub* OWL ontology
- Identity and Reference on the Web (IRW)
- Only few core concepts and relations
  - For the core concepts and relations we took practices from DUL+IOL, LMM and IRE
- Specific applications such as LOD have their small ontology that reuses IRW, if needed
  - "…many small ontological components largely created of pointers to each other…"
- Result
  - a network of ontologies for describing resources involved in applications on the web
- Why?
  - It can be used in applications such as LOD, over the Web
  - It helps the debate…

From recent awwsw discussion



IRW and specific modules



IRW Ingredients

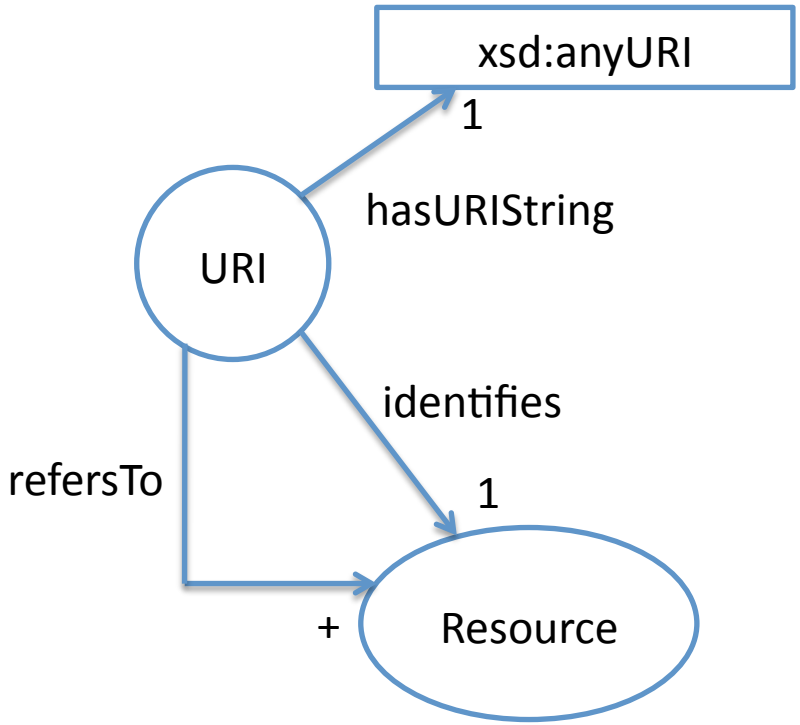
- Identification and Reference
- Resources and their types
- Access and Redirection
- “Hypertext” Web transaction

and

- LOD transaction

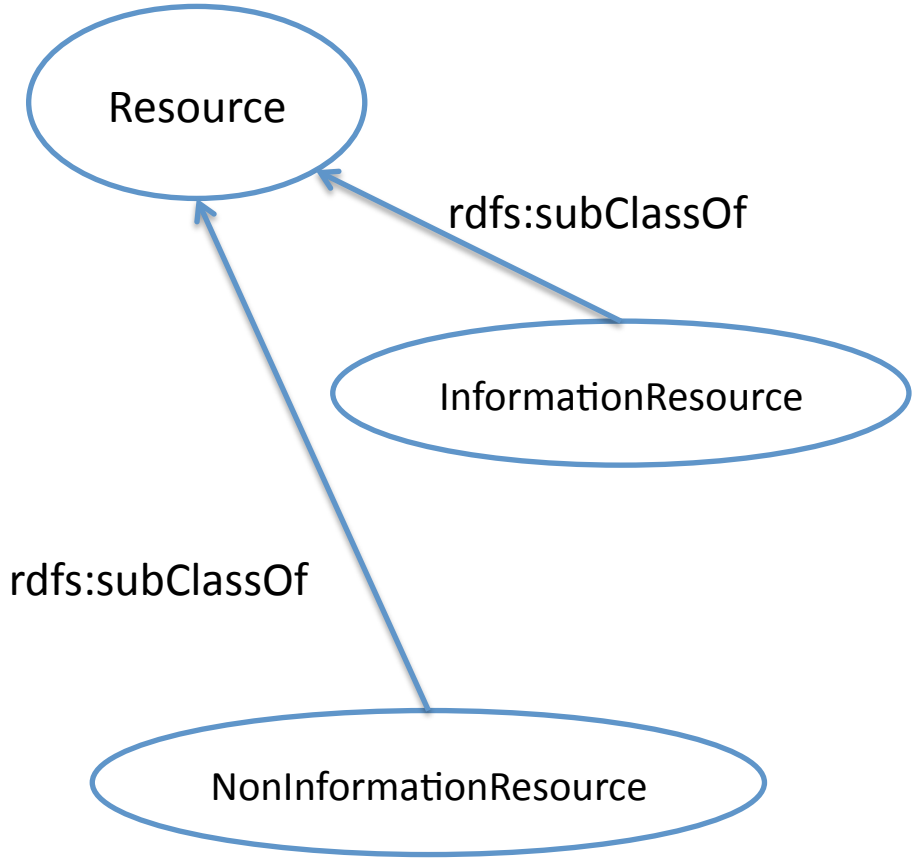
Identification and Reference

- A class for URIs allows us to talk about different types of URIs
- irw:identifies
  - Functional
  - The owner of the URI intends to identify a certain thing
- irw:refersTo
  - A URI refers to whatever results from its usage on the (Semantic) Web
  - Complies with *logicists' position*: the referent of a URI is ambiguous
- irw:identifies rdfs:subPropertyOf irw:refersTo



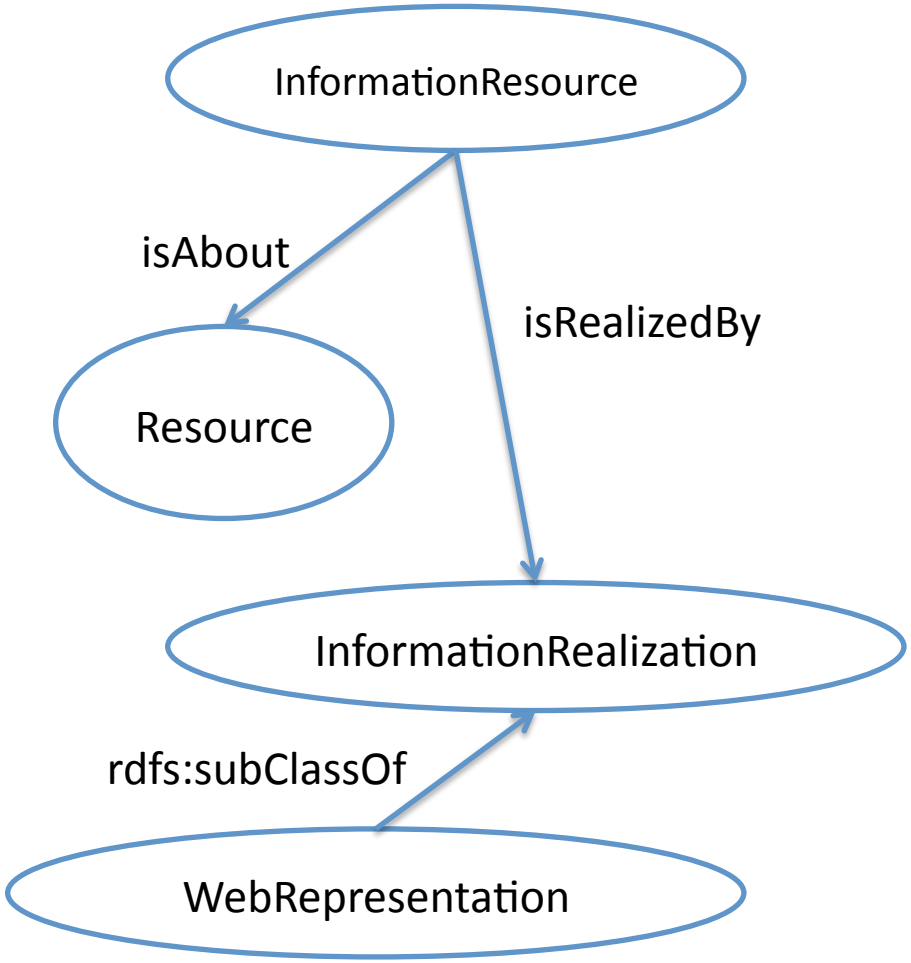
Resource types

- irw:Resource
  - is meant to express the same intuition as rdfs:Resource
  - In a possible OWL-full IRW it would be owl:equivalentTo rdfs:Resource
  - Two disjoint subclasses
    - IR
    - nIR



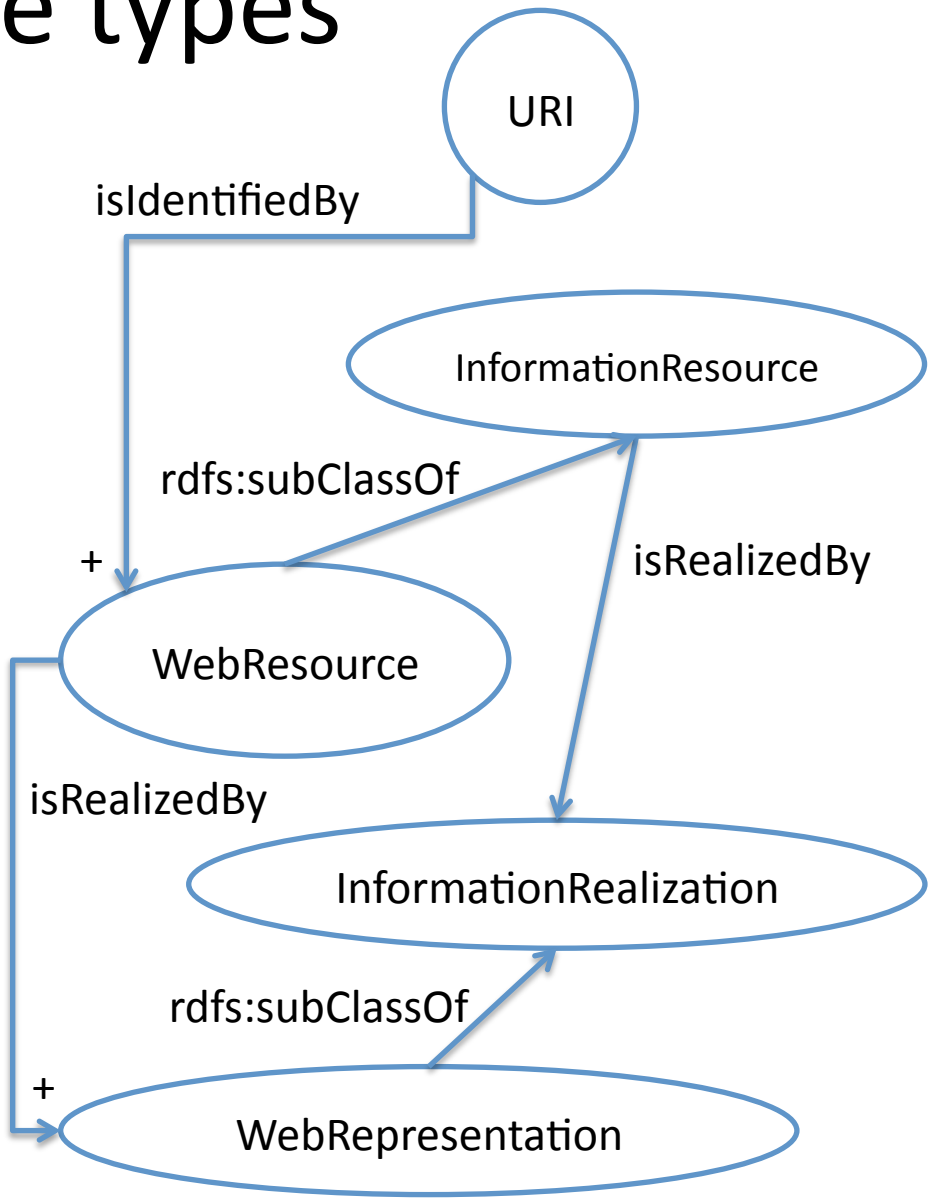
Resource types

- IR:
  - a resource which has the property that all of its essential characteristics can be conveyed in a message
    - i.e. TAG IR
  - E.g. the wikipedia text describing Rome
  - is about something
    - E.g. about Rome the place
  - is realized by some "information realization"
    - e.g., the web representation on the wikipedia server of the page describing Rome



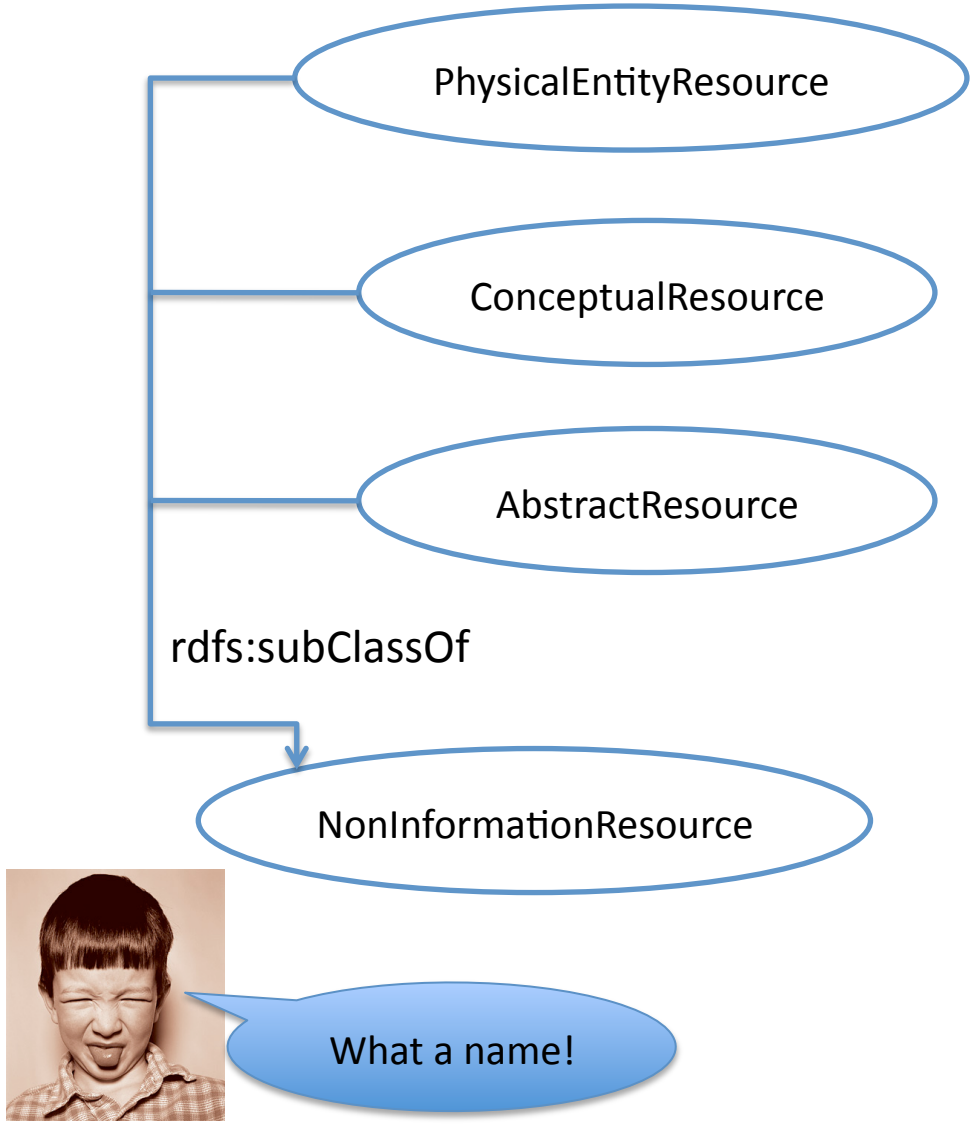
Resource types

- WebRepresentation
  - The realization of a message encoding that *goes on the wire*, according to an interaction protocol (e.g. http), in order to resolve a Web-accessible resource
- WebResource
  - A special kind of IR
  - It has at least a web representation
  - It is identified by at least one URI



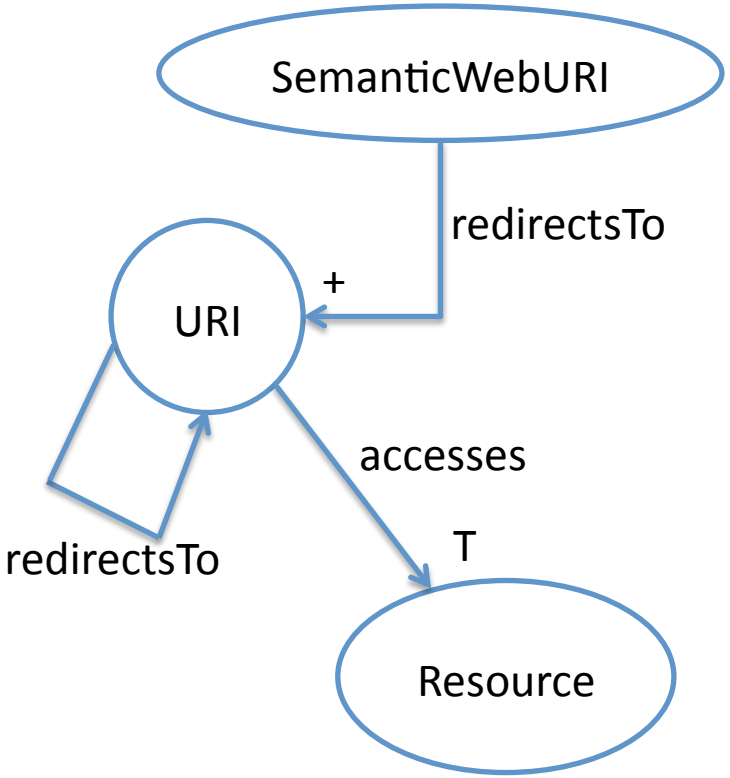
Resource types

- nIR: the complement of IR
  - Three subclasses help to give the intuition of what is an IR and what is a nIR
    - Non-normative modeling
    - There are many other plausible distinctions
  - Physical entities
    - A *touchable* resource
    - E.g. physical people, artifacts, places, etc.
  - Conceptual resources
    - They exist in the social communication process
    - E.g. legal entities, political entities, etc.
  - Abstract resources
    - Abstract things, combinatorial spaces. They cannot be located in space-time
    - E.g. functions, the infinite set of names that can be defined in a namespace, etc.



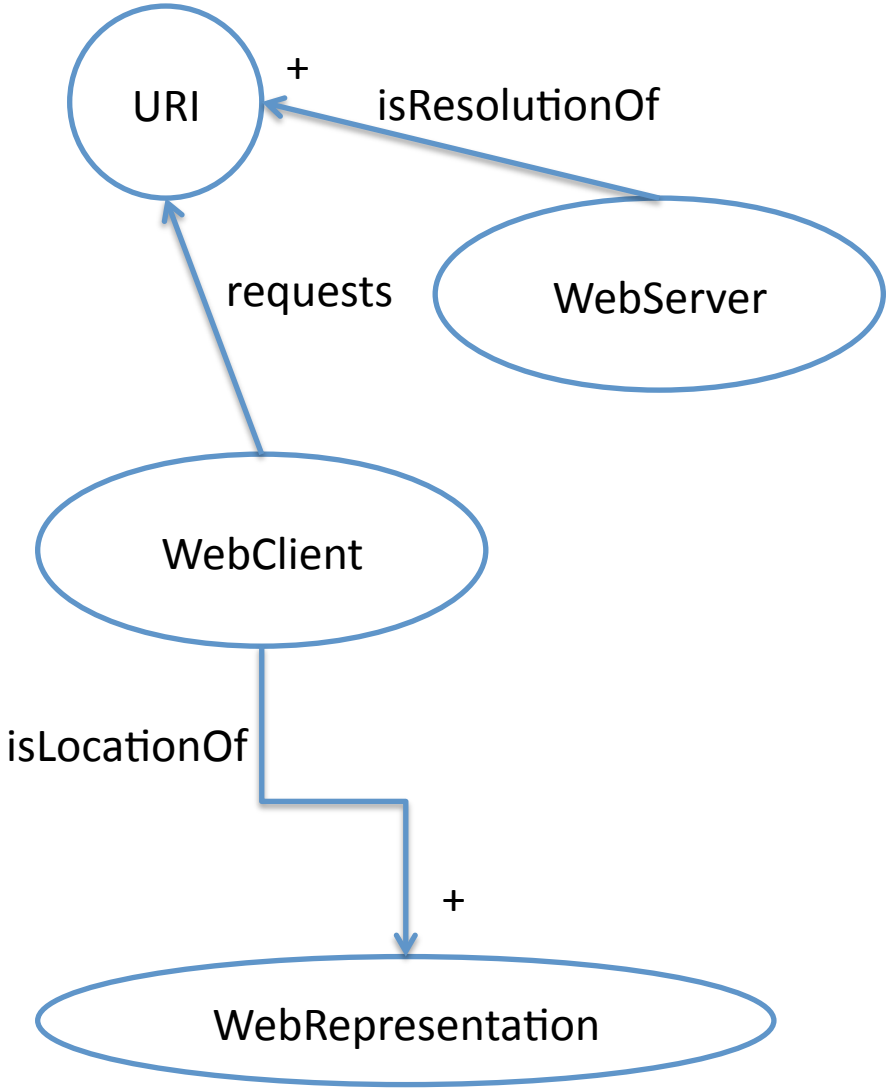
Access and Redirection

- irw:accesses
  - Transitive
  - A URI allows access to the thing it identifies/refers to
  - The notion of access is broader than that of web access
- irw:redirectsTo
  - rdfs:subPropertyOf irw:accesses
  - a URI can redirect to another URI e.g. http 303 redirect
  - The destination URI must identify a resource that is associated with a web representation
  - Two subproperties in a "tag –specific module"
    - tag2irw:redirects303To
    - tag2irw:redirectsHashTo
- irw:SemanticWebURI
  - Must have at least one redirect
  - identify nIR



“Hypertext” Web transactions

- irw:WebClient
  - A client in the Web context
- irw:requests
  - Is a hook to the HTTP in RDF
- irw:WebServer
  - A server hosting web representations
- irw:isResolutionOf
  - Inverse of irw:resolvesTo which is currently implemented by mapping a URI to an IP
- irw:isLocationOf
  - the Web server concretely can respond to an HTTP request with a particular Web representation



Self-describing linked data

- tag2http:redirects303To is ambiguous re the type of resource identified
  - It can occur between two IRs
  - The resource can embed the following in order to self describe itself and become *LOD-enabled*

dbpedia:resource/Eiffel_Tower rdf:type irw:NonInformationResource (*)

- From last Vocamp in Ibiza: can we avoid 303, then?

(*) dbpedia: http://dpedia.org/

Linked Data Transaction

- In the ldow2irw sample module we have defined the class ldow:AssociatedDescription
  - It is an InformationResource
  - It is about at least one NonInformationResource whose URI redirects to its one

dbpedia:resource/Eiffel_Tower rdf:type irw:NonInformationResource
dbpedia:resource/Eiffel_Tower irw:redirects303To dbpedia:data/Eiffel_Tower
dbpedia:data/Eiffel_Tower rdf:type ldow:AssociatedDescription
→
dbpedia:data/Eiffel_Tower rdf:type irw:InformationResource

A simple example of application

Semantic Validation of Linked Data …with very simple reasoning

INPUT → http://dbpedia.org/resource/Eiffel_Tower

Check

1. If HTTP 303 Request with content request type "application/rdf+xml" returns a RDF file
2. If HTTP 303 Request with content request type "text/html" returns a HTML file

OUTPUT (if succeeded) →

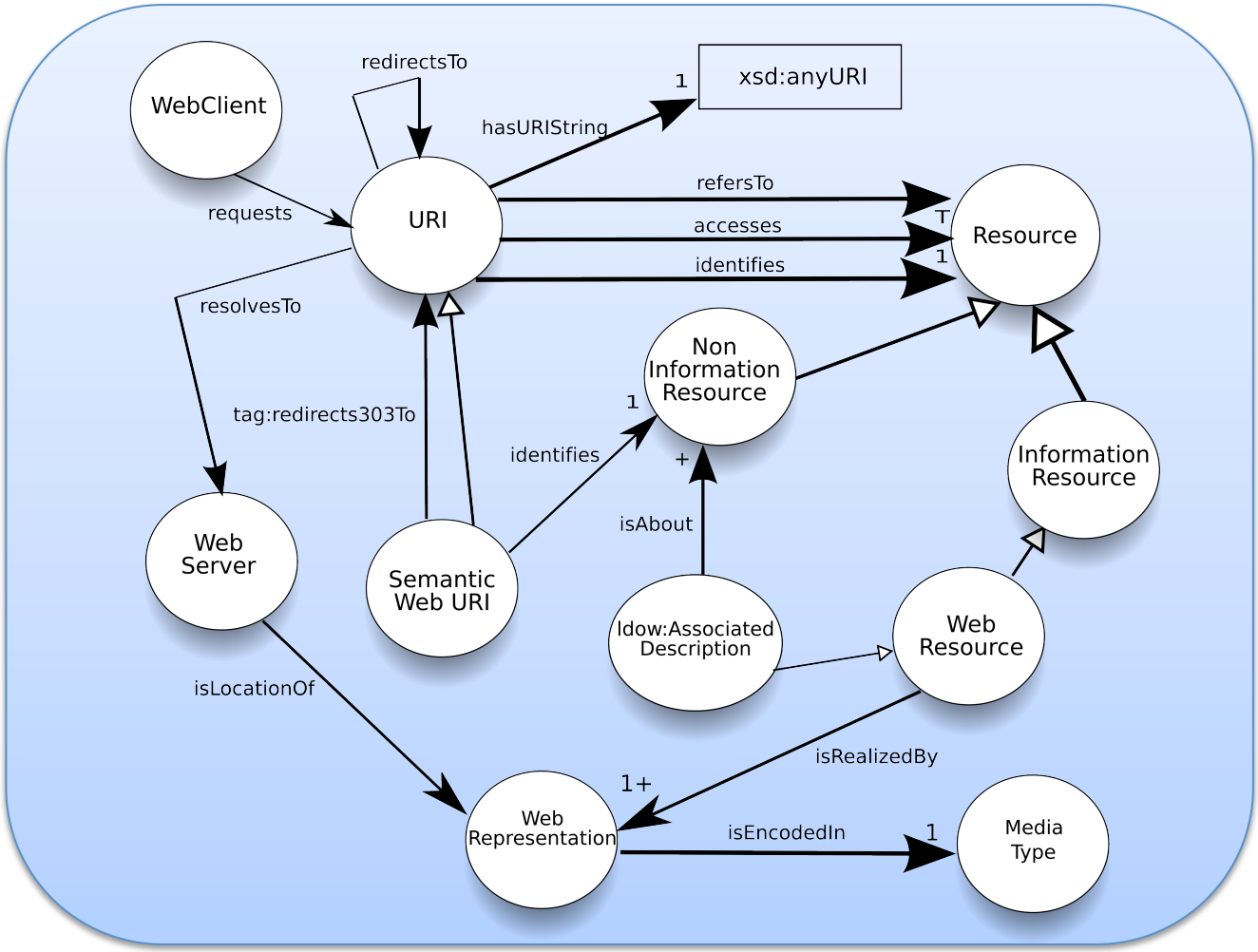
http://dbpedia.org/resource/Eiffel_Tower rdf:type irw:NonInformationResource
http://dbpedia.org/data/Eiffel_Tower rdf:type ldow:AssociatedDescription
http://dbpedia.org/page/Eiffel_Tower rdf:type ldow:AssociatedDescription
http://dbpedia.org/resource/Eiffel_Tower ldow:redirects303To http://dbpedia.org/page/Eiffel_Tower
http://dbpedia.org/resource/Eiffel_Tower ldow:redirects303To http://dbpedia.org/data/Eiffel_Tower

SOME ADDITIONAL ANNOTATIONS CAN BE ADDED:
http://dbpedia.org/data/Eiffel_Tower irw:about http://dbpedia.org/resource/Eiffel_Tower
http://dbpedia.org/page/Eiffel_Tower irw:about http://dbpedia.org/resource/Eiffel_Tower
…
→ Also useful for detecting possible misuses of owl:sameAs

What next?

- To fix/refine IRW based on awwsw discussions/inputs as well as from practical experiences such as those presented today
  - Domain-specific applications can link to IRW as a shared *meta-level* vocabulary
- To align IRE to “Generic Resource” ontology
- Implementations
  - E.g. the ongoing IKS project, SMW extension, etc.

Questions?


WebClient
redirectsTo
1
xsd:anyURI
hasURIString
requests
URI
refersTo
accesses
identifies
T
1
Resource
resolvesTo
tag:redirects303To
1
Non
Information
Resource
identifies
+
Information
Resource
Web
Server
Semantic
Web URI
isAbout
ldow:Associated
Description
Web
Resource
isLocationOf
isRealizedBy
1+
Web
Representation
isEncodedIn
1
Media
Type