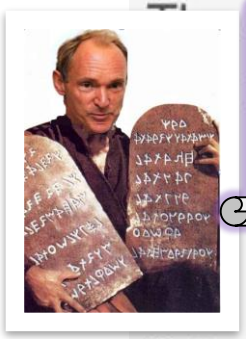
WEB SÉMANTIQUE ET ONTOLOGIES
WEB DES DONNÉES
DONNÉES LIÉES (LINKED DATA)

2 – NOMMER LES DONNEES
URIS UNIFORM RESOURCE IDENTIFIERS
IRIS Internationalized Resource Identifiers

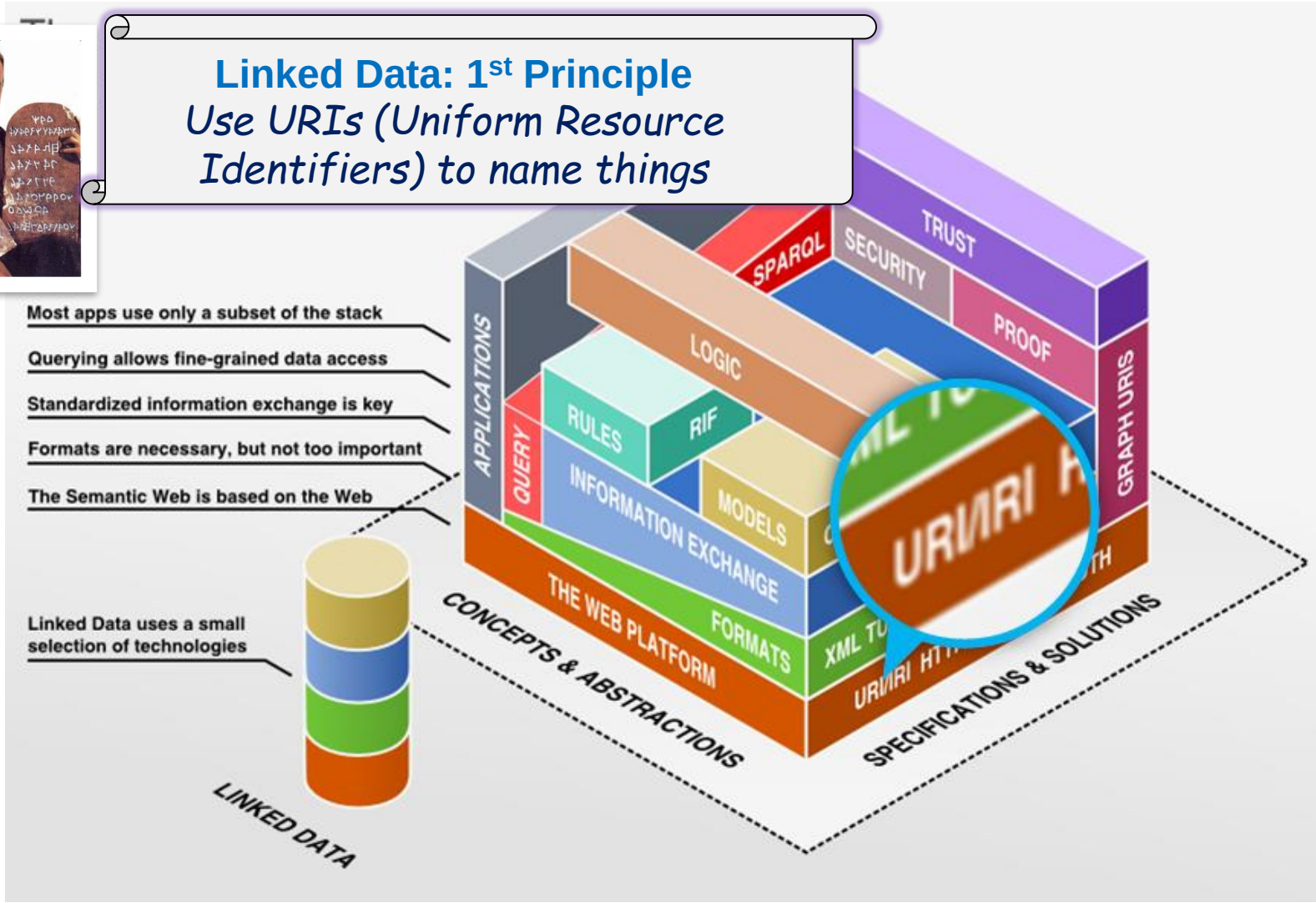
Philippe GENOUD – Danielle ZIEBELIN - LIG-
Prénom.Nom@imag.fr
Mohamed Wadhah Mabrouk

Uniform Resource Identifiers (URIs)



Linked Data: 1st Principle

Use URIs (Uniform Resource Identifiers) to name things



(**Crédit** Benjamin Nowack)

URI: syntax

- URI : generic syntax

scheme ":" ["//" authority "/"] [path] ["?" query] ["#" fragment]

- **scheme:** http, ftp, mailto, …
- **authority:** [userinfo@]host[:port]
  - **userinfo:** authentication section e.g: username:password
  - **host:** domain name, IP address
  - **port:** port number, ex: 80 for HTTP standard port
- **path:** a sequence of segments separated by slashes, e.g. : a path in the hierarchical file system of the HTTP server.
- **query:** a query string of non-hierarchical data. (e.g: a sequence of attribute–value pairs separated by a delimiter (&) for HTTP requests)
- **fragment:** a fragment identifier providing direction to a secondary resource (e.g.: anchor id in a HTML document)

Internationalized Resource Identifier (RFC 3987): extension to support Universal Character Set (Unicode/ISO 10646))

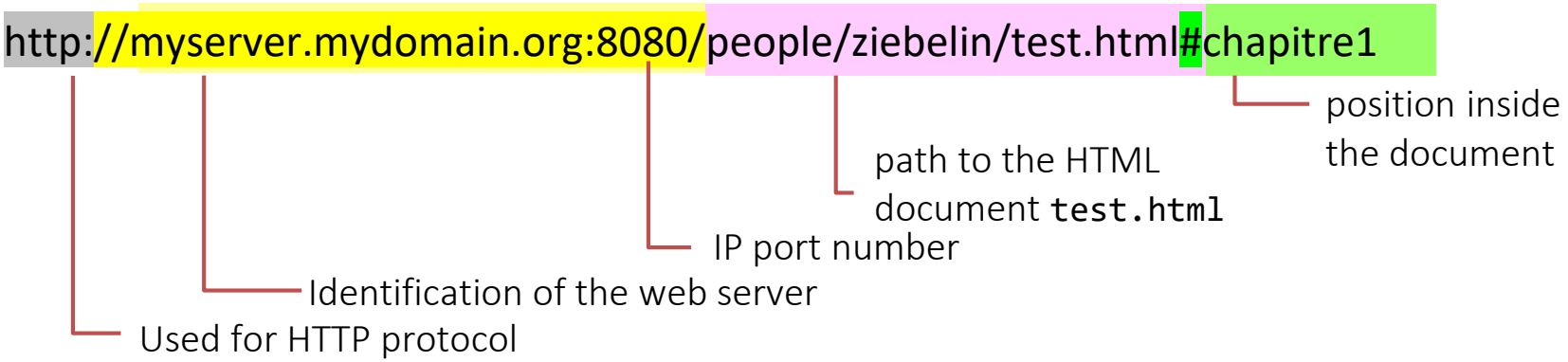
http://fa.dbpedia.org/resource/ژرژ_برسنس

URI: examples

- URI : generic syntax

scheme ":" ["//" authority "/"] [path] ["?" query] ["#" fragment]

example :

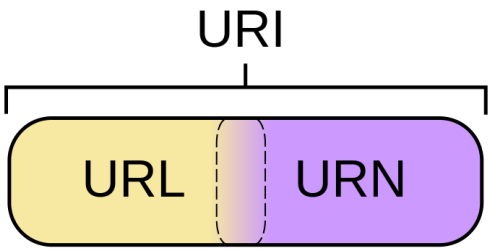


other URIs examples:

ftp://server.example.com/foo
mailto:person@example.fr
urn:isbn:978-0553283686

URI - URL - URN

- an URI/IRI **doesn't necessarily** identifies a resource that is resolvable on the web



- Address (Locator)
  - **U**niform **R**esource **L**ocator (RFC 1738)
  - Tells where and how a resource can be found in the internet
  - *Can change during the life cycle of a resource*
- Identity (Name)
  - **U**niform **R**esource **N**ame (RFC 2141)
  - identifies a resource by name in a particular namespace. A URN can be used to talk about a resource without implying its location or how to access it.
  - *Remains unchanged during life cycle of the resource*

URIs, URLs, and URNs: Clarifications and Recommendations 1.0
Report from the joint W3C/IETF URI Planning Interest Group- W3C Note 21 September 2001
http://www.w3.org/TR/uri-clarification/

Resources in the Web of data

- Data describe elements (things) for a domain of interest through their properties and relationships.
- Theses elements can be:

RESOURCES

- documents
  - HTML pages, images, data accessed through a web service…

URLs

- but also:
  - real world entities:
    - place: Sète
    - a person: Georges Brassens
    - …
  - abstracts concepts:
    - Set of French songwriters of years 60-80,
    - *Author of* relationship
    - …

URNs

URIs

- All these elements are resources identified by an URI

Examples of resources about Georges Brassens available in the web

Traditional web

Documents

identified by URLs

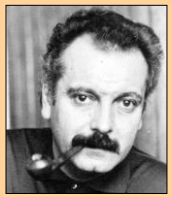
DBpedia

XML file containing structured data about Georges Brassens
http://dbpedia.org/data/Georges_Brassens.xml

Web page (HTML file)

http://fr.wikipedia.org/wiki/Georges_Brassens

Picture (jpeg file)



http://culturetheque.org.uk/media/item/17545/800/brassens.jpg

Video (mp4 file)
https://www.youtube.com/watch?v=rslShTbqNbo

Examples of resources about Georges Brassens available in the web

Traditional web

Documents

identified by URLs

Web of Data

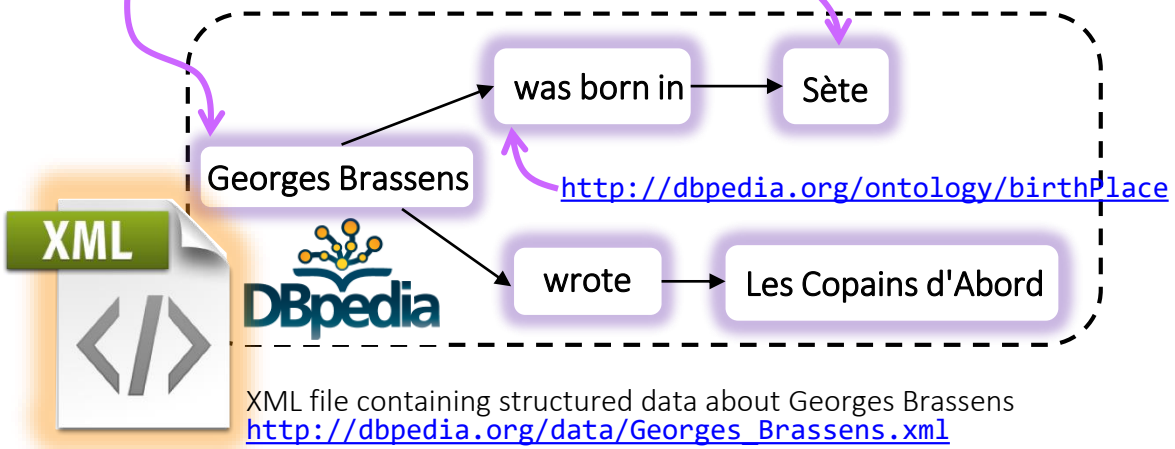
Real world entities

Abstract concepts

identified by URNs

http://dbpedia.org/resource/Georges_Brassens
DBpedia resource representing Georges Brassens

http://dbpedia.org/resource/Sète

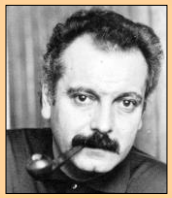


XML file containing structured data about Georges Brassens
http://dbpedia.org/data/Georges_Brassens.xml

Web page (HTML file)

http://fr.wikipedia.org/wiki/Georges_Brassens

Picture (jpeg file)

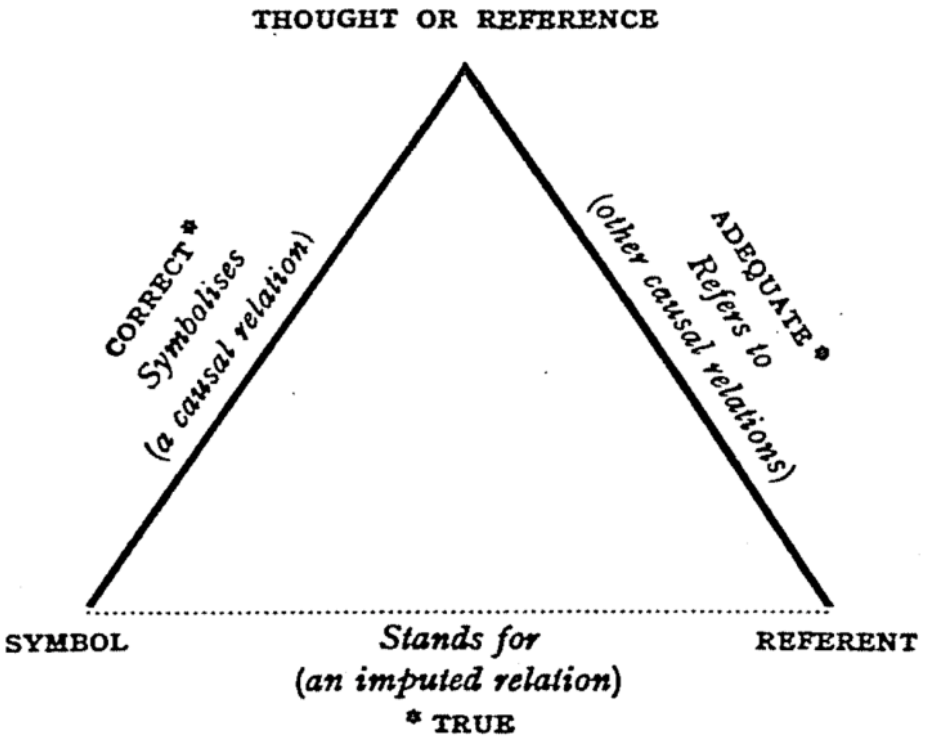


http://culturetheque.org.uk/media/item/17545/800/brassens.jpg

Video (mp4 file)
https://www.youtube.com/watch?v=rslShTbqNbo

Semiotics

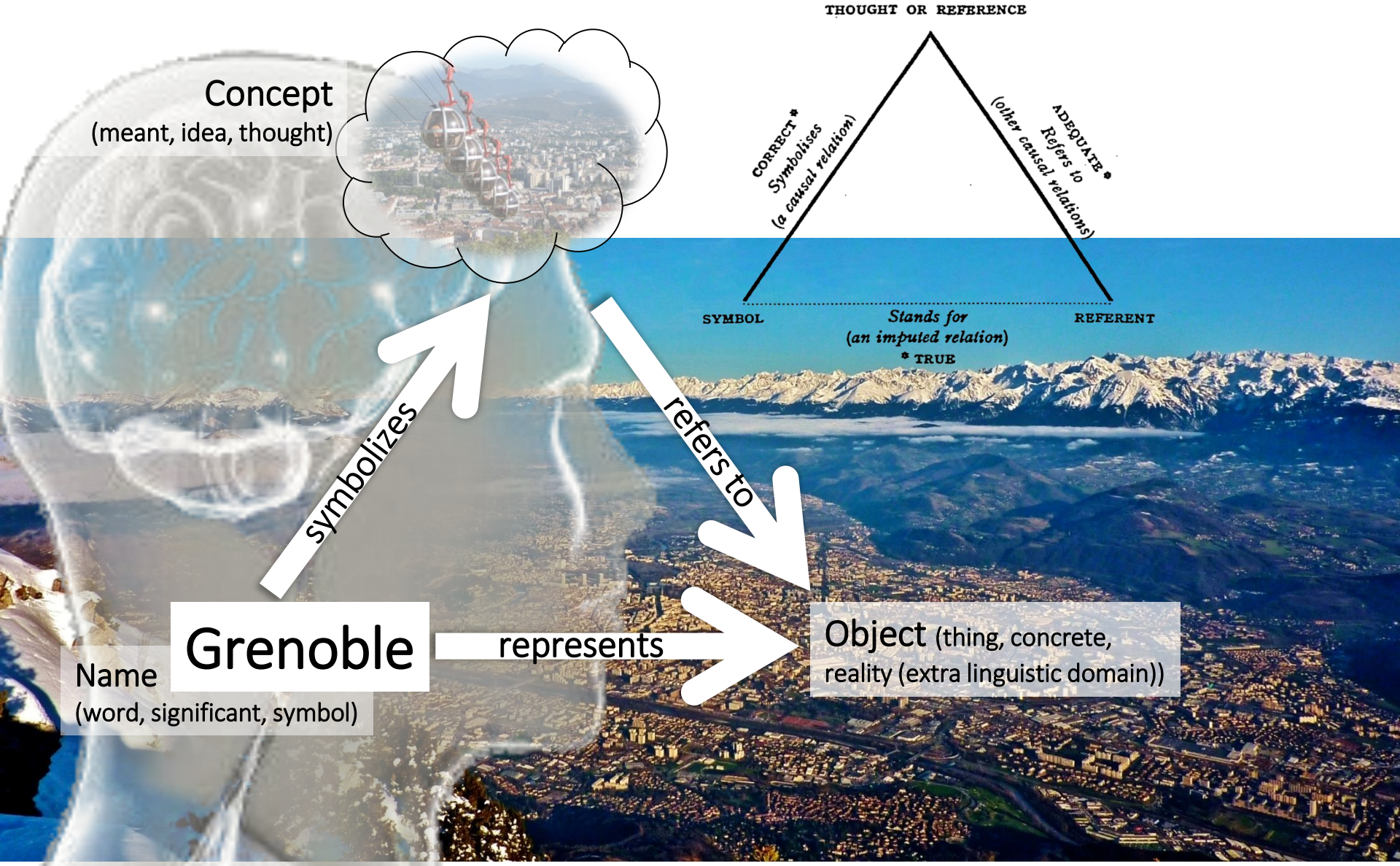
- **Semiotics** : The science of communication studied through the interpretation of signs and symbols as they operate in various fields, esp. language. Oxford English Dictionary (2003).



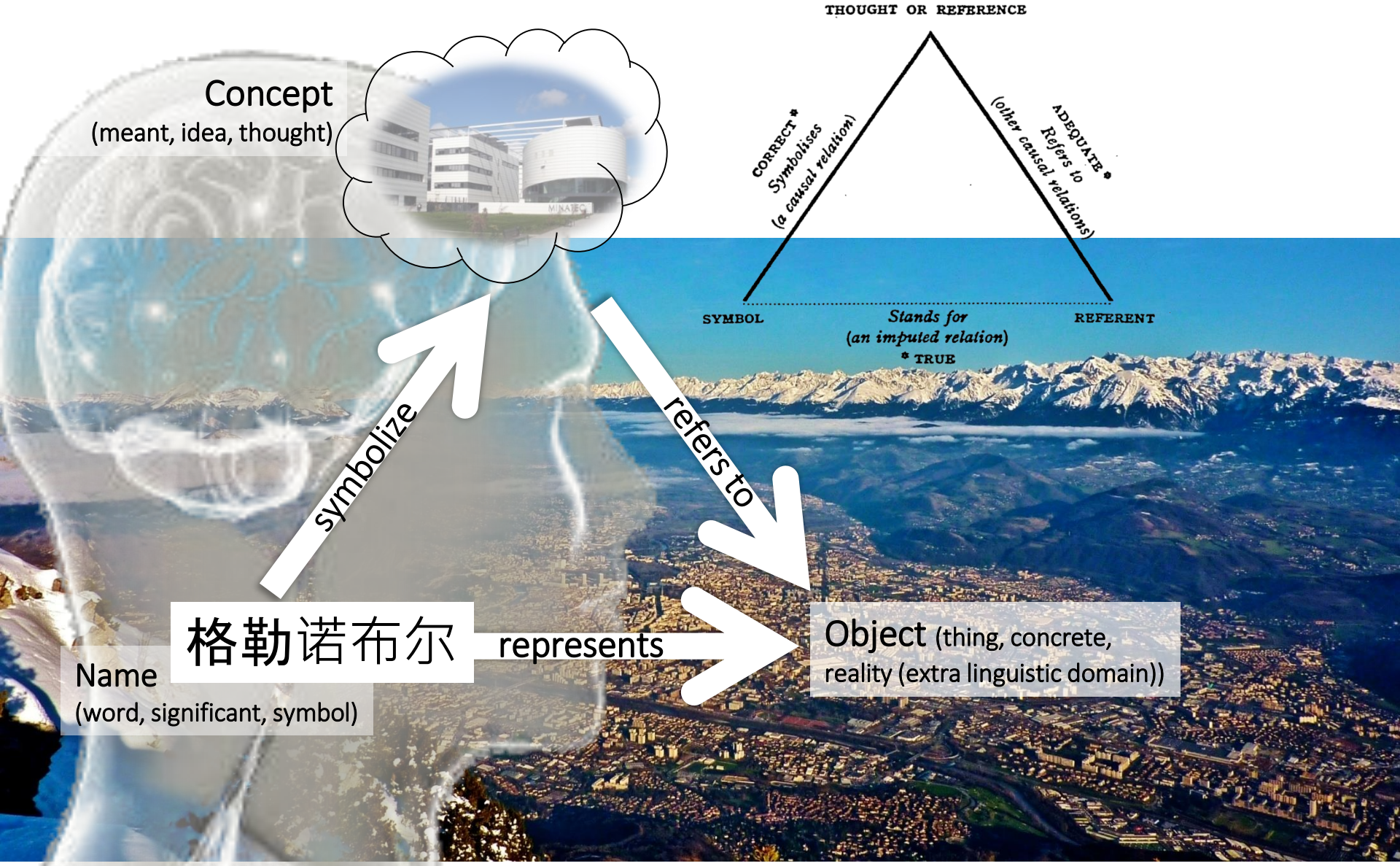
Semiotic Triangle by Odgen & Richard

([*The Meaning of Meaning*](#) -A Study of the Influence of Language upon Thought and of the Science of Symbolism, 1923)

Semiotics


Concept
(meant, idea, thought)
THOUGHT OR REFERENCE
CORRECT *
Symbolises
(a causal relation)
ADEQUATE *
Refers to
(other causal relations)
SYMBOL
Stands for
(an imputed relation)
* TRUE
REFERENT
symbolizes
refers to
Grenoble
represents
Name
(word, significant, symbol)
Object (thing, concrete, reality (extra linguistic domain))

Semiotics


Concept
(meant, idea, thought)
THOUGHT OR REFERENCE
CORRECT *
Symbolises
(a causal relation)
ADEQUATE *
Refers to
(other causal relations)
SYMBOL
Stands for
(an imputed relation)
* TRUE
REFERENT
symbolize
refers to
格勒诺布尔
represents
Name
(word, significant, symbol)
Object (thing, concrete, reality (extra linguistic domain))

Uniform Resource Identifiers (URIs)

represents

represents

DBpedia

GeoNames

`http://dbpedia.org/resource/grenoble`

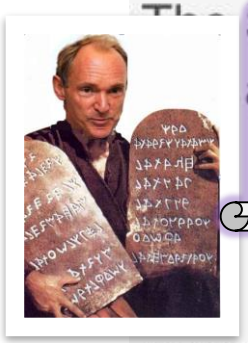
`http://sws.geonames.org/3014728`

The name (URI) that represents the city of Grenoble in DBpedia

The name (URI) that represents the city of Grenoble in GeoNames

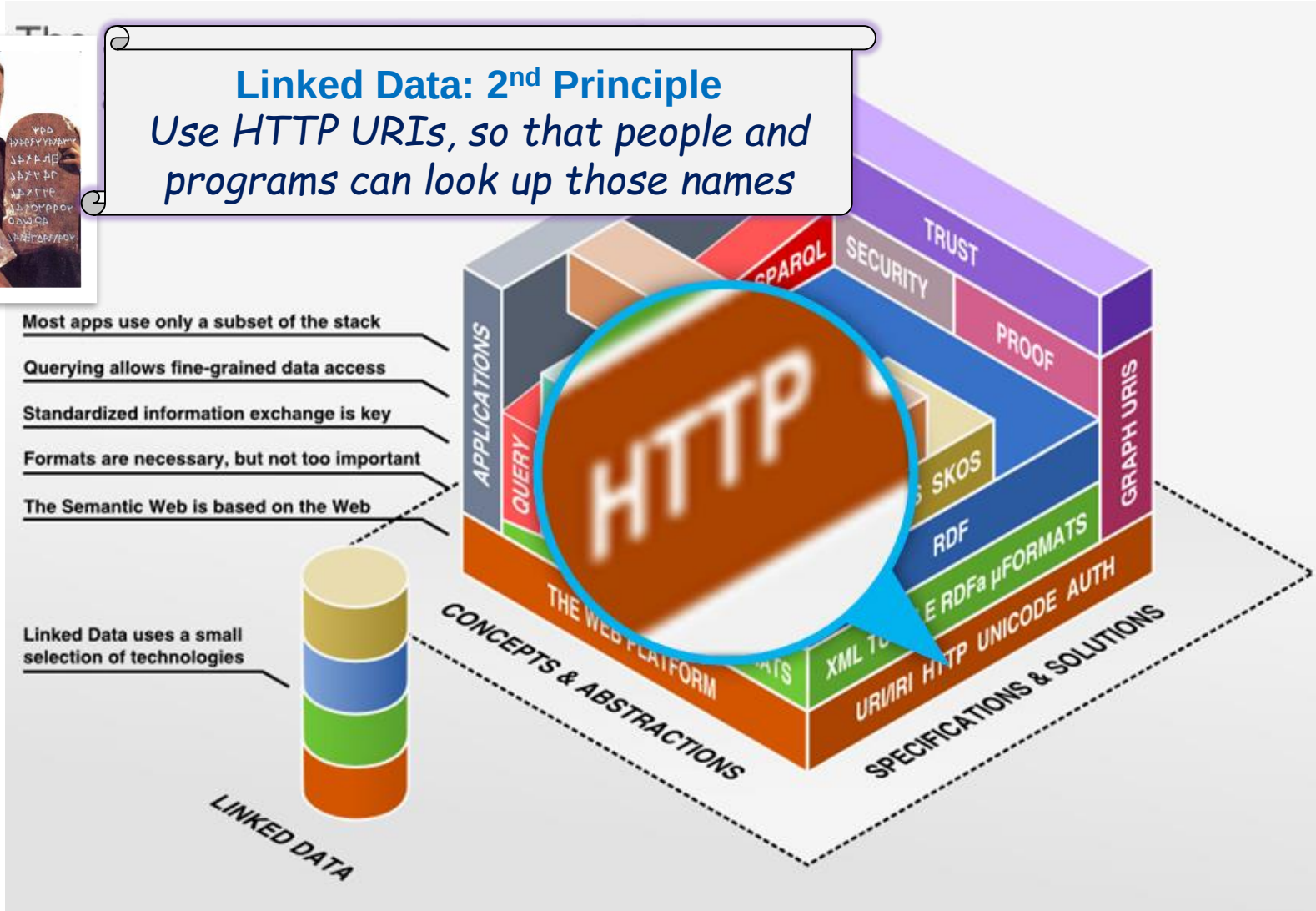
Different URIs (URNs) in different namespaces can represents the same thing

HTTP URIs



Linked Data: 2nd Principle

Use HTTP URIs, so that people and programs can look up those names



(Crédit Benjamin Nowack)

HTTP URIs

- HTTP (Hyper Text Transfer Protocol) protocol is the Web's universal access mechanism.

- HTTP URIs make good names for two reasons*:
  - They provide a simple way to create globally unique names in a decentralized fashion, as every owner of a domain name, or delegate of the domain name owner, may create new URI references.
  - They serve not just as a name but also as a means of accessing information describing the identified
    - HTTP clients can **dereference** (i.e., look up) the URI using the HTTP protocol and retrieve a description of the resource that is identified by the URI.

* Tom Heath and Christian Bizer (2011)
Linked Data: Evolving the Web into a Global Data Space (1st edition).
Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
http://linkeddatabook.com/editions/1.0/

HTTP URIs

- Where HTTP URIs identify real-world objects or abstract concepts, it is essential to **not confuse** the objects or concepts themselves with the Web documents that describe them.
  - A real world object or abstract concept can have different representations
  - It allows separate statements to be made about an object and about a document that describes that object.

URI of DBpedia resource representing Georges Brassens

http://dbpedia.org/resource/Georges_Brassens

a URN **but not** a URL

http://dbpedia.org/page/Georges_Brassens

URL of HTML Page

Description intended to be read by humans

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:dbpprop="http://dbpedia.org/property
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    version="XHTML+RDFa 1.0" xml:lang="en">

<!-- header -->
<head profile="http://www.w3.org/1999/xhtml/v
    <title>About: Georges Brassens</title>
    <link rel="alternate" type="application/rdf+xml" href="http://dbpe
    <link rel="alternate" type="text/rdf+n3" href="http://dbpedia.org/
    <link rel="alternate" type="application/json+rdf" href="http://dbp
    <link rel="alternate" type="application/json" href="http://dbpedia
    <link rel="alternate" type="application/atom+xml" href="http://dbpe
    <link rel="alternate" type="text/plain" href="http://dbpedia.org/da
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph

    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph

    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="alternate" href="http://dbpedia.org/sparql?default-graph
    <link rel="timegate" type="text/html" href="http://mementoarchive.la
    <link rel="stylesheet" type="text/css" href="/statics/style.css" />
```

http://dbpedia.org/data/Georges_Brassens.xml

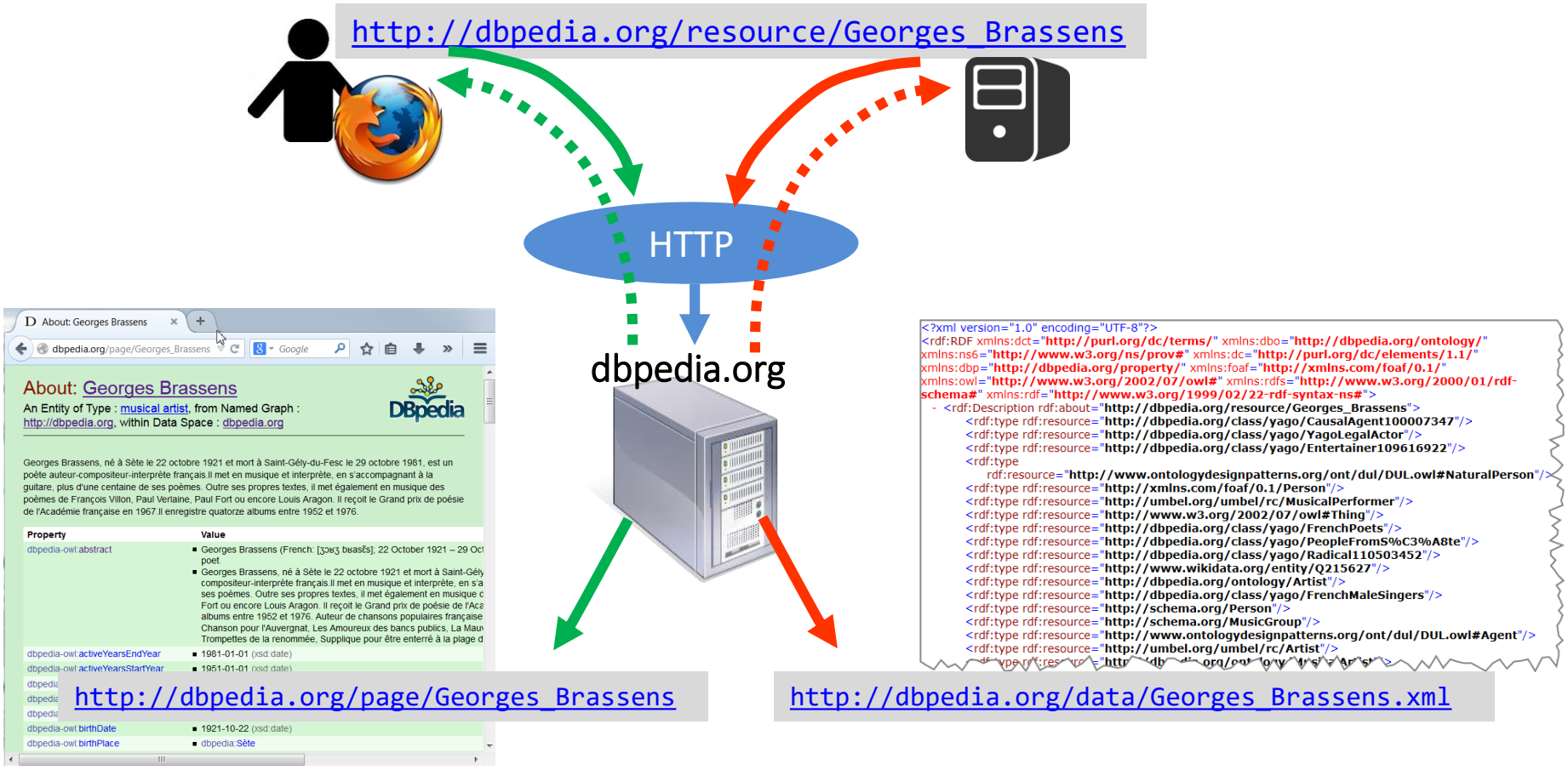
URL of RDF/XML document

Description intended for consumption by machines

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:dct="http://purl.org/dc/terms/" xmlns:d
xmlns:ns6="http://www.w3.org/ns/prov#" xmlns:dc="h
xmlns:dbp="http://dbpedia.org/property/" xmlns:foaf="
xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:
schema#" xmlns:rdf="http://www.w3.org/1999/02/22
 - <rdf:Description rdf:about="http://dbpedia.org/resou
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/CausalAgent100007347"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/YagoLegalActor"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/Entertainer109616922"/>
      <rdf:type
        rdf:resource="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#NaturalPerson"/>
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
      <rdf:type rdf:resource="http://umbel.org/umbel/rc/MusicalPerformer"/>
      <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/FrenchPoets"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/PeopleFromS%C3%A8te"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/Radical110503452"/>
      <rdf:type rdf:resource="http://www.wikidata.org/entity/Q215627"/>
      <rdf:type rdf:resource="http://dbpedia.org/ontology/Artist"/>
      <rdf:type rdf:resource="http://dbpedia.org/class/yago/FrenchMaleSingers"/>
      <rdf:type rdf:resource="http://schema.org/Person"/>
      <rdf:type rdf:resource="http://schema.org/MusicGroup"/>
      <rdf:type rdf:resource="http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#Agent"/>
      <rdf:type rdf:resource="http://umbel.org/umbel/rc/Artist"/>
```

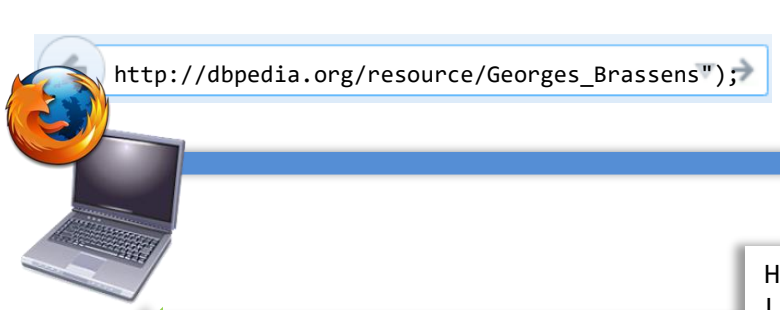
Making URIs Dereferenceable
Content negotiation (303 URIs)

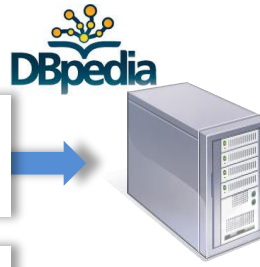
- The same URI can be used to retrieve different representations.



Making URIs Dereferenceable
Content negotiation (303 URIs)

- Contents negotiation uses HTTP headers to a retrieve the resource description

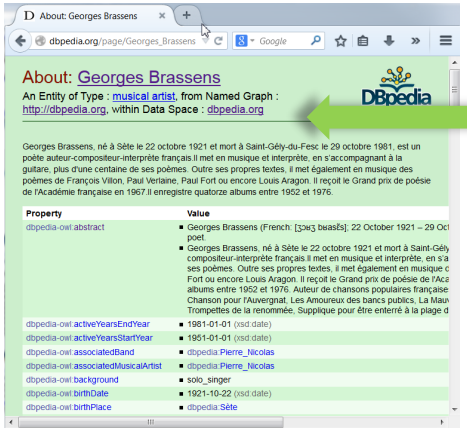





```
GET /resource/Georges_Brassens HTTP/1.1
Host: dbpedia.org
Accept: text/html
```

```
HTTP/1.1 303 See Other
Location: http://dbpedia.org/page/Georges_Brassens
Vary: Accept
```

```
GET /page/Georges_Brassens HTTP/1.1
Host: dbpedia.org
Accept: text/html
```

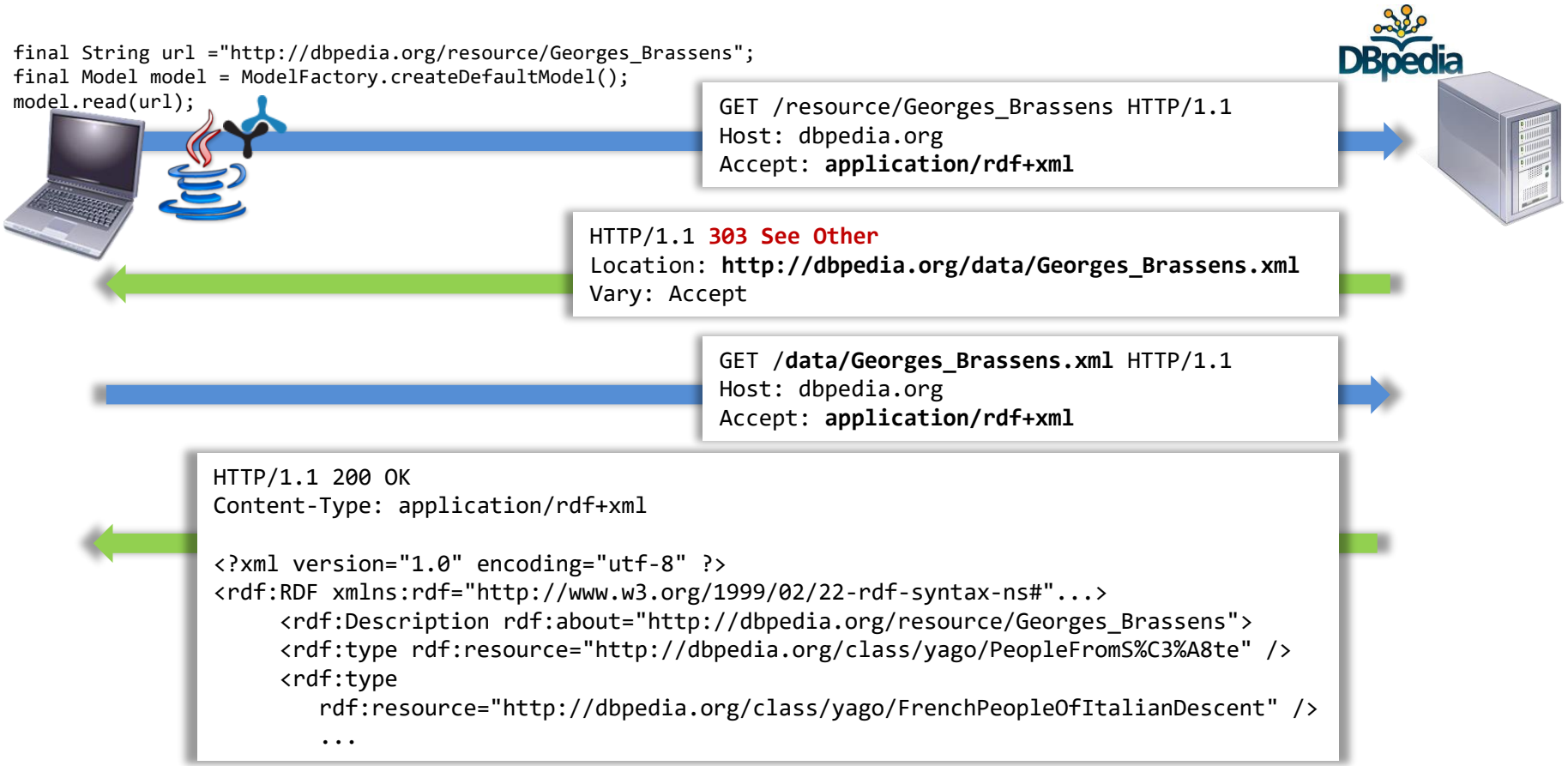



```
HTTP/1.1 200 OK
Content-Type: text/html

<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE html>
   <head>
      <title>About: Georges Brassens</title>
      ...
   </head>
   <body onload="init();">
      <div id="header">
      <h1 id="title">About: Georges Brassens</h1>
       ...
```

Making URIs Dereferenceable
Content negotiation (303 URIs)

- Contents negotiation uses HTTP headers to a retrieve the resource description



Making URIs Dereferenceable
Hash URIs

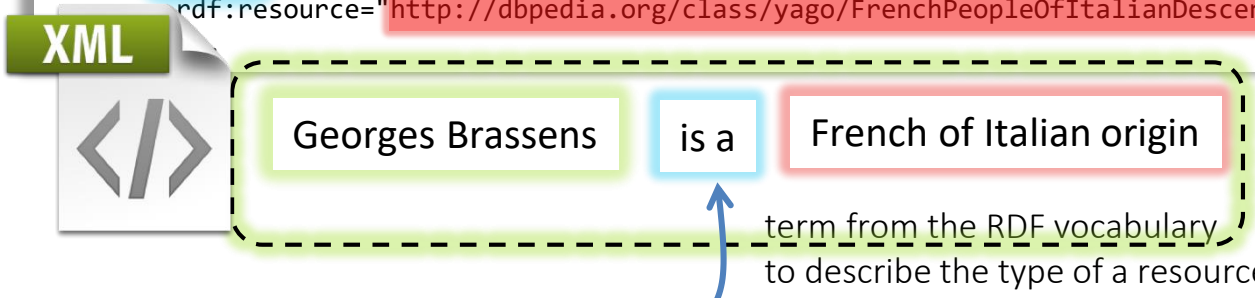
- An other way to identify real world objects or abstract concepts without creating ambiguity with the document that contains it 's description is to use hash URIs.

Example of a hash URI used by DBpedia RDF description of Georges Brassens

http://dbpedia.org/data/Georges_Brassens.xml

```
HTTP/1.1 200 OK
Content-Type: application/rdf+xml

<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"...>
     <rdf:Description rdf:about="http://dbpedia.org/resource/Georges_Brassens">
     <rdf:type rdf:resource="http://dbpedia.org/class/yago/PeopleFromS%C3%A8te" />
     <rdf:type
          rdf:resource="http://dbpedia.org/class/yago/FrenchPeopleOfItalianDescent" />
```



Hash URI `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`

base | fragment identifier

Making URIs Dereferenceable
Hash URIs

When a client wants to retrieve a hash URI, the HTTP protocol requires the fragment part to be stripped off before requesting the URI from the server

http://www.w3.org/1999/02/22-rdf-syntax-ns#type

URL the RDF/XML document containing a description of the whole RDF vocabulary

```
GET /1999/02/22-rdf-syntax-ns
Host: www.w3.org
Accept: application/rdf+xml
```


```
HTTP/1.1 200 OK
Content-Type: application/rdf+xml

<?xml version="1.0"?>
<rdf:RDF   xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">

...
<rdf:Property rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">
  <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
  <rdfs:label>type</rdfs:label>
  <rdfs:comment>The subject is an instance of a class.</rdfs:comment>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
  <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdfschema#Resource"/>
</rdf:Property>
...
<rdfs:Class rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag">
     <rdfs:isDefinedBy rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
     <rdfs:label>Bag</rdfs:label>
     <rdfs:comment>The class of unordered containers.</rdfs:comment>
```

Client is in charge of extracting the information associated to the fragment

How to chose good URIs ?

- Let's consider the book "Fire on the Mountain" written by Edward Abbey (ISBN 0-8263-0457-57). Can you find a cool URI for this resource ?

— uncool URIs

keep out of namespaces you do not control

- `https://www.amazon.com/Fire-Mountain-Edward-Abbey/dp/0062193902`
- `http://artemisbookstore.com/book.php?title=Fire+on+the+Mountain&author=Edward-Abbey&format=rdf`

abstract away from implementation details

- `http://artemisbookstore.com:3333/Fire-on-the-Moutain-Edward-Abbey`

this local name must be unique and persistent.
i.e. ISBN or generated UUID (Universal Unique Identifier)

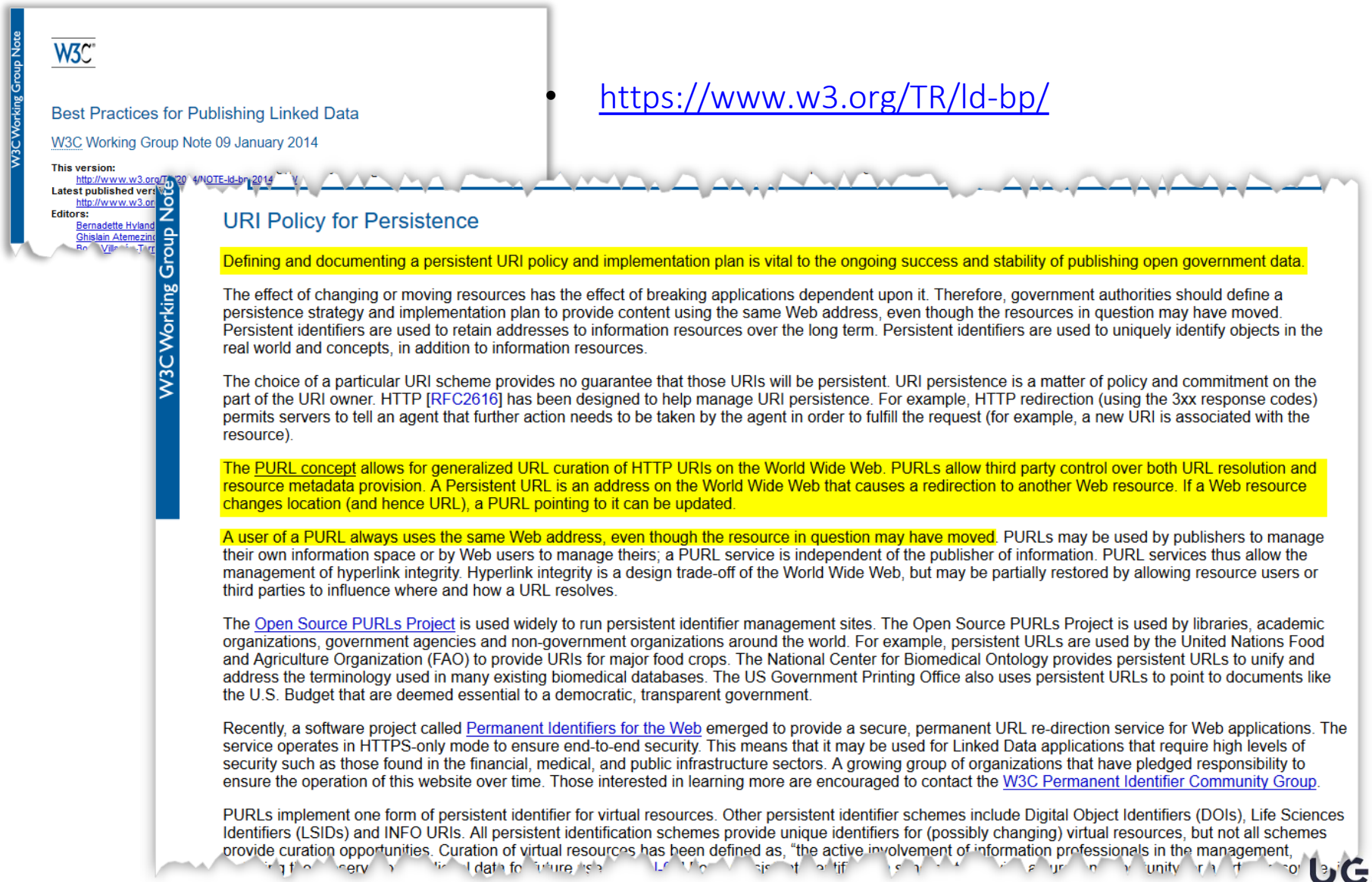
— cool URIs

- `http://artemisbookstore.com/resource/0-8263-0457-57`
- `http://artemisbookstore.com/page/0-8263-0457-57.html`
- `http://artemisbookstore.com/data/0-8263-0457-57.rdf`

- `http://id.artemisbookstore.com/book/0-8263-0457-57`
- `http://page.artemisbookstore.com/book/0-8263-0457-57.html`
- `http://data.artemisbookstore.com/book/0-8263-0457-57.rdf`

Persistent URIs

- https://www.w3.org/TR/ld-bp/



W3C Working Group Note

W3C

Best Practices for Publishing Linked Data

W3C Working Group Note 09 January 2014

This version:

Latest published version:

Editors:

Bernadette Hyland

Ghislain Atemezing

W3C Working Group Note

URI Policy for Persistence

Defining and documenting a persistent URI policy and implementation plan is vital to the ongoing success and stability of publishing open government data.

The effect of changing or moving resources has the effect of breaking applications dependent upon it. Therefore, government authorities should define a persistence strategy and implementation plan to provide content using the same Web address, even though the resources in question may have moved. Persistent identifiers are used to retain addresses to information resources over the long term. Persistent identifiers are used to uniquely identify objects in the real world and concepts, in addition to information resources.

The choice of a particular URI scheme provides no guarantee that those URIs will be persistent. URI persistence is a matter of policy and commitment on the part of the URI owner. HTTP [RFC2616] has been designed to help manage URI persistence. For example, HTTP redirection (using the 3xx response codes) permits servers to tell an agent that further action needs to be taken by the agent in order to fulfill the request (for example, a new URI is associated with the resource).

The PURL concept allows for generalized URL curation of HTTP URIs on the World Wide Web. PURLs allow third party control over both URL resolution and resource metadata provision. A Persistent URL is an address on the World Wide Web that causes a redirection to another Web resource. If a Web resource changes location (and hence URL), a PURL pointing to it can be updated.

A user of a PURL always uses the same Web address, even though the resource in question may have moved. PURLs may be used by publishers to manage their own information space or by Web users to manage theirs; a PURL service is independent of the publisher of information. PURL services thus allow the management of hyperlink integrity. Hyperlink integrity is a design trade-off of the World Wide Web, but may be partially restored by allowing resource users or third parties to influence where and how a URL resolves.

The Open Source PURLs Project is used widely to run persistent identifier management sites. The Open Source PURLs Project is used by libraries, academic organizations, government agencies and non-government organizations around the world. For example, persistent URLs are used by the United Nations Food and Agriculture Organization (FAO) to provide URIs for major food crops. The National Center for Biomedical Ontology provides persistent URLs to unify and address the terminology used in many existing biomedical databases. The US Government Printing Office also uses persistent URLs to point to documents like the U.S. Budget that are deemed essential to a democratic, transparent government.

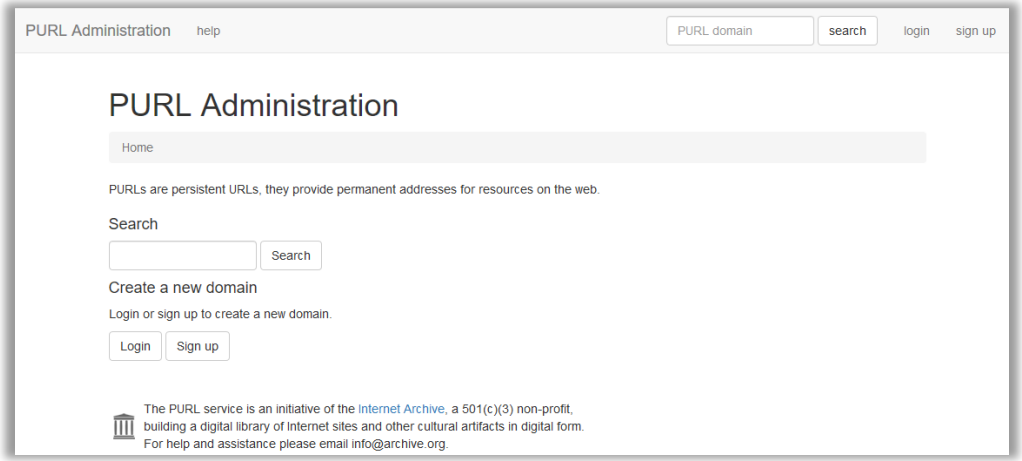
Recently, a software project called Permanent Identifiers for the Web emerged to provide a secure, permanent URL re-direction service for Web applications. The service operates in HTTPS-only mode to ensure end-to-end security. This means that it may be used for Linked Data applications that require high levels of security such as those found in the financial, medical, and public infrastructure sectors. A growing group of organizations that have pledged responsibility to ensure the operation of this website over time. Those interested in learning more are encouraged to contact the W3C Permanent Identifier Community Group.

PURLs implement one form of persistent identifier for virtual resources. Other persistent identifier schemes include Digital Object Identifiers (DOIs), Life Sciences Identifiers (LSIDs) and INFO URIs. All persistent identification schemes provide unique identifiers for (possibly changing) virtual resources, but not all schemes provide curation opportunities. Curation of virtual resources has been defined as, "the active involvement of information professionals in the management,

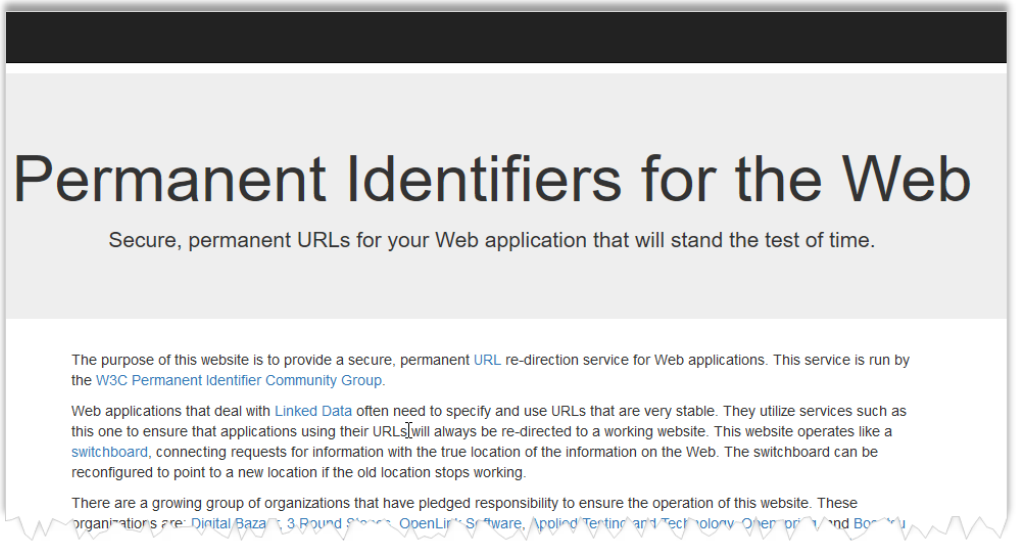
Persistent URIs

- examples of service offering persistent URIs

https://archive.org/services/purl/

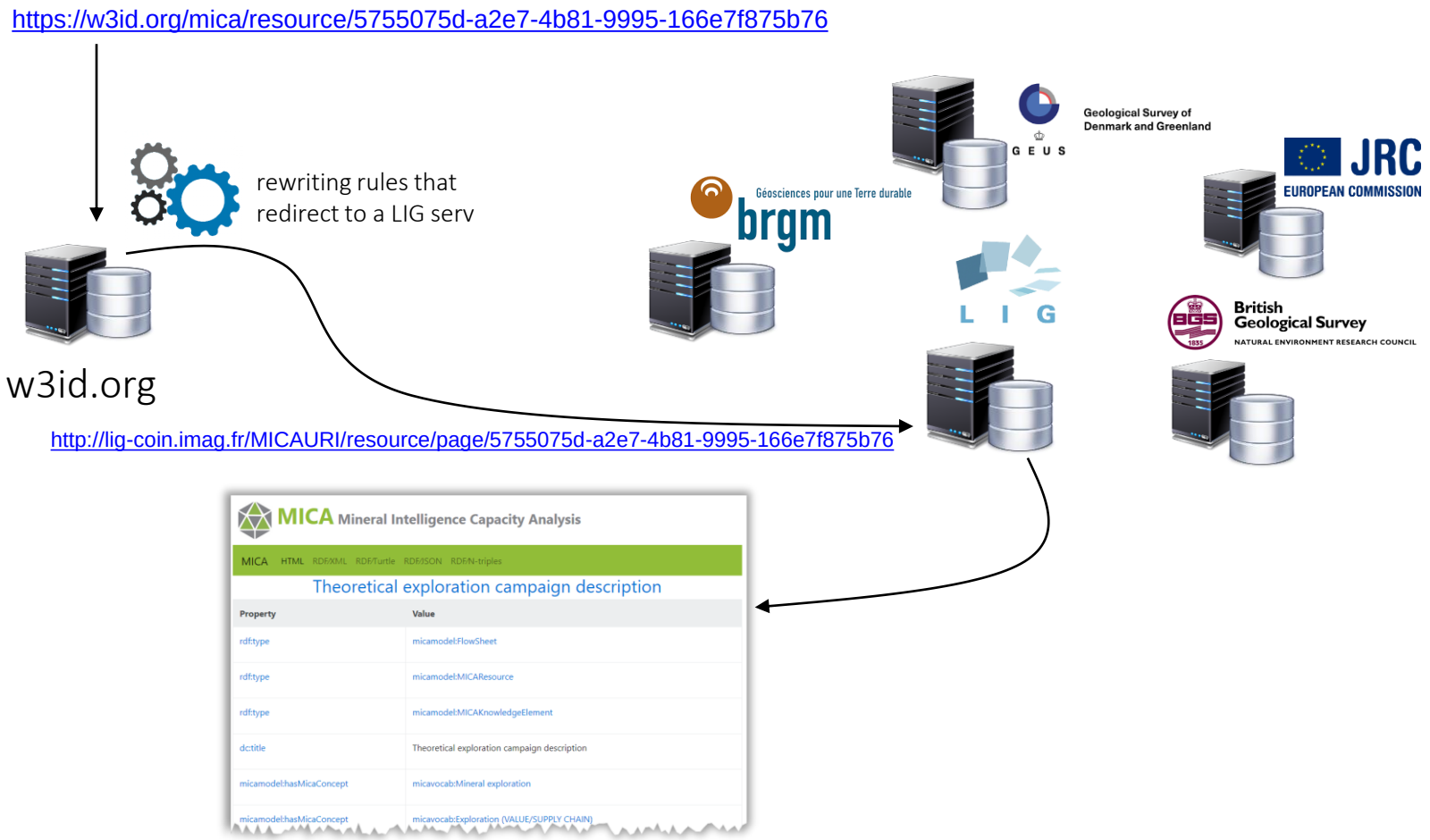


https://w3id.org/



Persistent URIs example

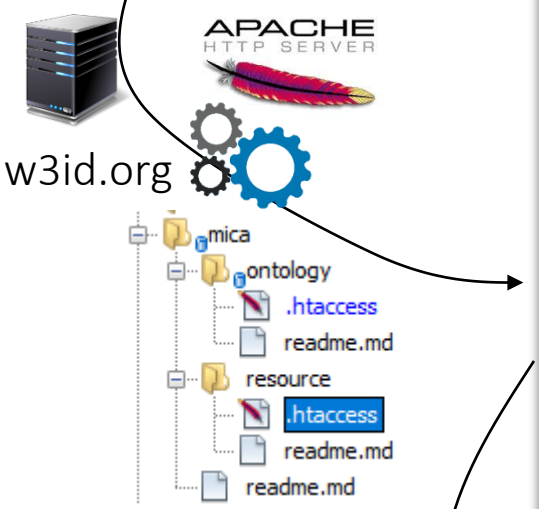
- Example : MICA European project



Persistent URIs example

- Example : MICA European project

https://w3id.org/mica/resource/5755075d-a2e7-4b81-9995-166e7f875b76




```
Options +FollowSymLinks
# Turn off MultiViews
Options -MultiViews

# Directive to ensure *.rdf files served as appropriate content type,
# if not present in main apache config
AddType application/rdf+xml .rdf
AddType text/turtle .ttl

RewriteEngine on

###############################################################################
#                 Resource rules for web negociation                          #
###############################################################################
#Rewrite rules for Resource HTML content from the resource URI if requested
RewriteCond %{HTTP_ACCEPT} !application/rdf\+xml.*(text/html|application/xhtml\+xml)
RewriteCond %{HTTP_ACCEPT} text/html [OR]
RewriteCond %{HTTP_ACCEPT} application/xhtml\+xml [OR]
RewriteCond %{HTTP_USER_AGENT} ^Mozilla/.*
RewriteRule ^(.+) http://lig-coin.imag.fr/MICAURI/resource/page/$1 [R=303,NE,L]

# Rewrite rule to serve RDF/XML content from the resource URI if requested
RewriteCond %{HTTP_ACCEPT} application/rdf\+xml
RewriteRule ^(.+) http://lig-coin.imag.fr/MICAURI/resource/data/$1?type=rdf [R=303,NE,L]

# Rewrite rule to serve TURTLE content from the resource URI if requested
RewriteCond %{HTTP_ACCEPT} text/turtle
RewriteRule ^(.+) http://lig-coin.imag.fr/MICAURI/resource/data/$1?type=ttl [R=303,NE,L]

# Rewrite rule to serve JSON-LD content from the resource URI if requested
RewriteCond %{HTTP_ACCEPT} application/json
RewriteRule ^(.+) http://lig-coin.imag.fr/MICAURI/resource/data/$1?type=json [R=303,NE,L]

#default response
RewriteRule ^(.+) http://lig-coin.imag.fr/MICAURI/resource/data/$1?type=ttl [R=303,NE,L]
```

http://lig-coin.imag.fr/MICAURI/resource/page/5755075d-a2e7-4b81-9995-166e7f875b76