Linked Data: Structured Data on the Web

Chapter 1: The Linked Data Web

Steve Baskauf 2016-09-12

Things we would like to enable a "machine" to do with Linked Data

1. Acquire data from other sources, possibly by following links.

2. Merge those data with our data

3. "Learn" things that weren't apparent before the data were merged.

4. Expose our data in a way that would allow others to link to it.

The first three items are what we do with a Web browser and our brains via documents and the World Wide Web. In this group, we want to figure out how to do the same thing automatically with software.

Example of linked resources

image
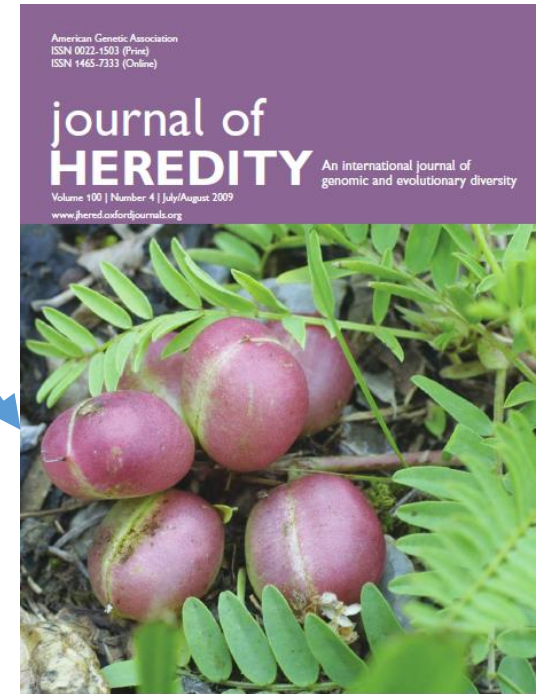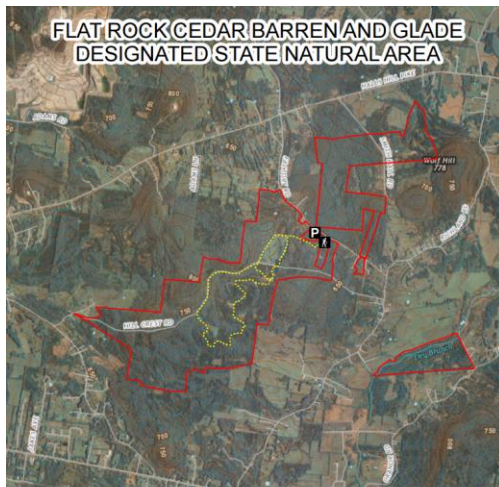
location

photographer

journal article

Human-readable metadata

Baskauf's image number 25041
Photographed at Flat Rock Cedar Glades and Barrens State Natural Area.
Created by Steve Baskauf
See the article "Population Genetics of *Astragalus bibullatus* (Fabaceae) Using AFLPs"
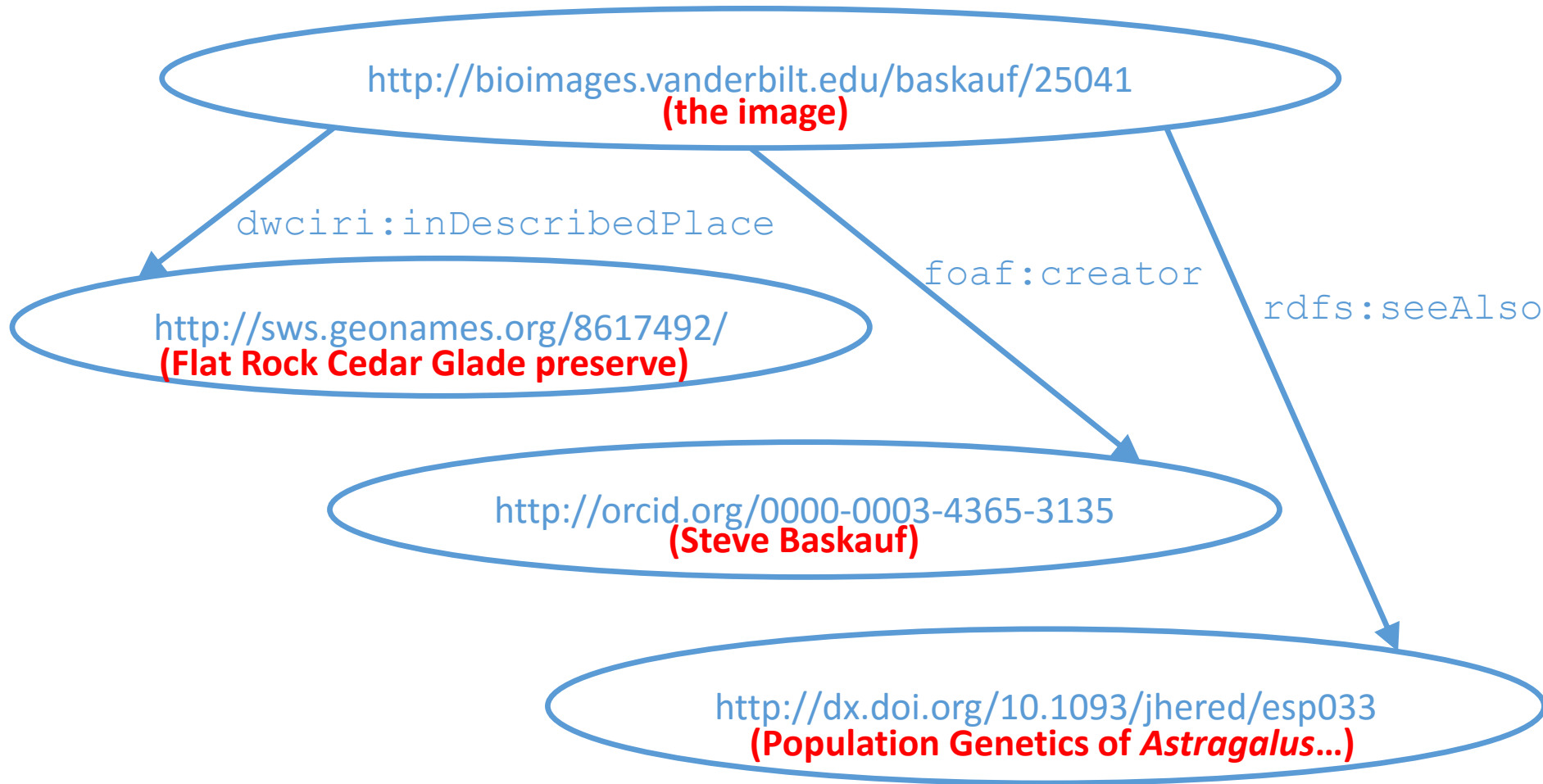
Questions:
What are the licensing terms of the image?
Where does Steve Baskauf work?
What is the latitude and longitude of Flat Rock Cedar Glade?
Who wrote ""Population Genetics of *Astragalus bibullatus* (Fabaceae) Using AFLPs"

Graph-based view of the metadata

Machine-readable metadata

```
<http://bioimages.vanderbilt.edu/baskauf/25041>
    dwciri:inDescribedPlace <http://sws.geonames.org/8617492/>;
    foaf:creator <http://orcid.org/0000-0003-4365-3135>;
    rdfs:seeAlso <http://dx.doi.org/10.1093/jhered/esp033>.
```

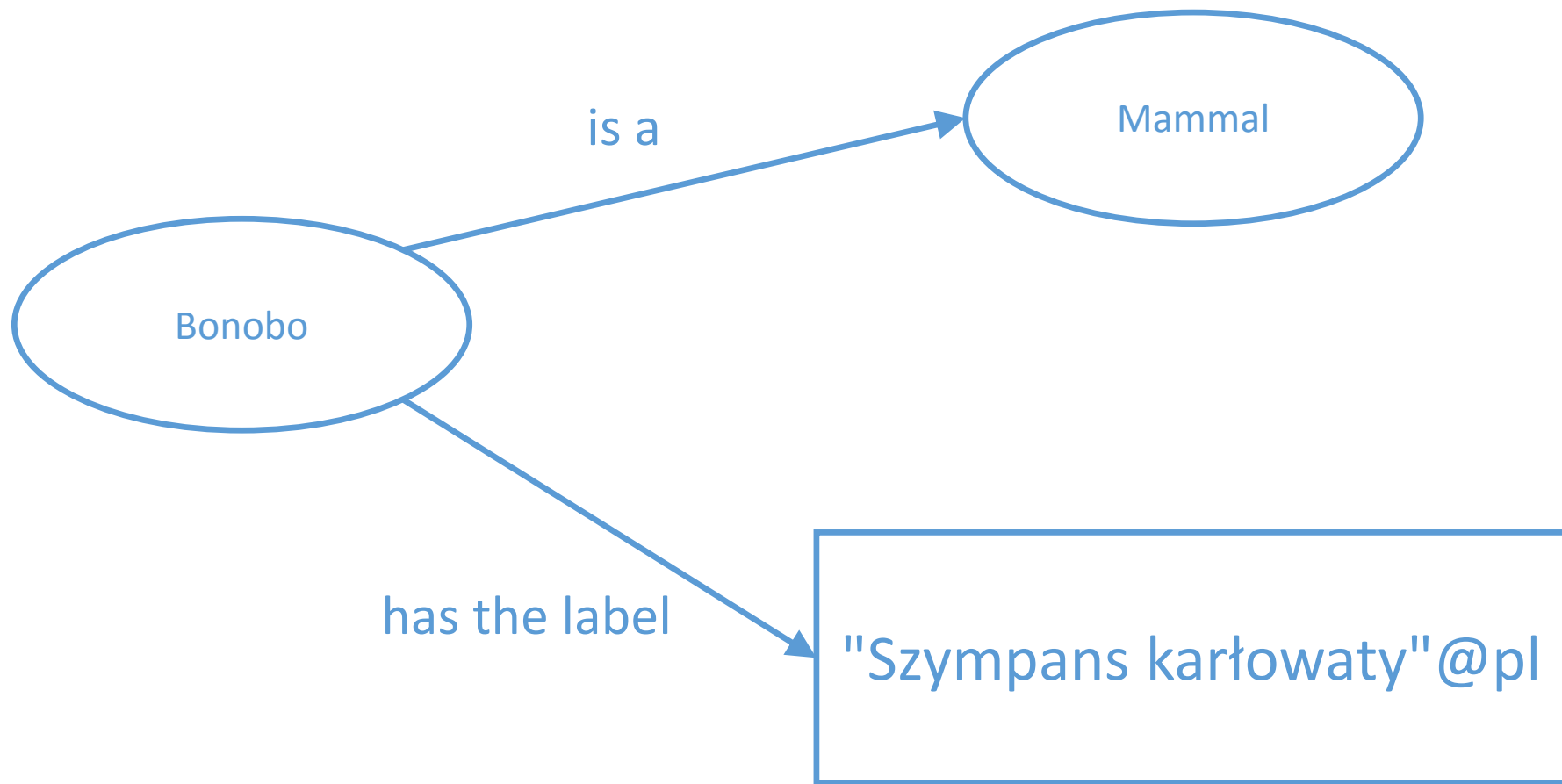**From these data, a machine could answer all of these questions:**

What are the licensing terms of the image?

Where does Steve Baskauf work?

What is the latitude and longitude of Flat Rock Cedar Glade?

Who wrote ""Population Genetics of *Astragalus bibullatus* (Fabaceae) Using AFLPs"

Things we might want to say about bonobos, represented as a graph

How do we know that what you mean by "mammal" is the same as what I mean by "mammal"?

Go to http://umbel.org and find a standard reference concept for mammal. The URI is a globally unique **identifier** for that concept. (It may or may not be a URL that can actually be used in a browser.)
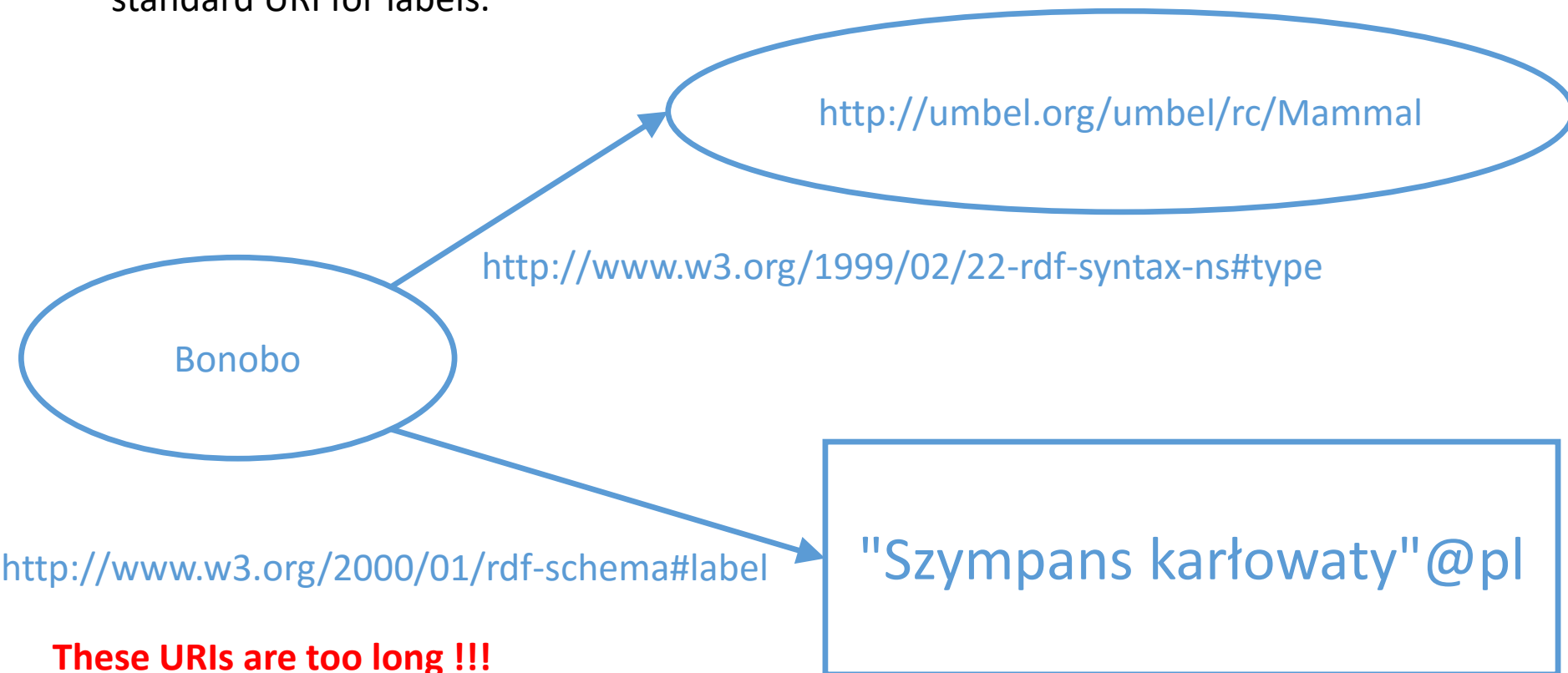
This is Linked Data Principle #1.

Bonobo

is a

http://umbel.org/umbel/rc/Mammal

has the label

"Szympans karłowaty"@pl
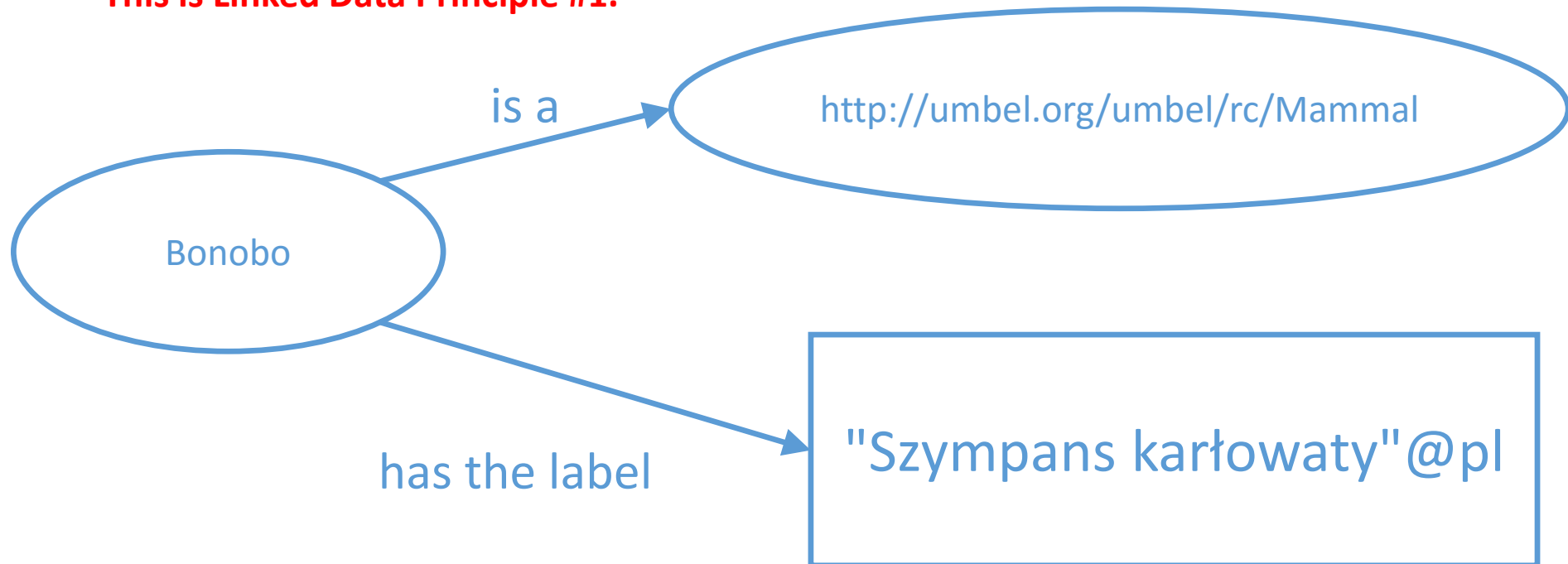
How do we know that what you mean by "is a" is the same as what I mean by "is a"?

Use the standard W3C URI for specifying the class of a thing:
http://www.w3.org/1999/02/22-rdf-syntax-ns#type and another standard URI for labels.

http://umbel.org/umbel/rc/Mammal

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

Bonobo

http://www.w3.org/2000/01/rdf-schema#label

"Szympans karłowaty"@pl

These URIs are too long !!!

We can make a URI shorter and easier to read by abbreviating it

Namespace abbreviations:
rdf: = http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs: = http://www.w3.org/2000/01/rdf-schema#
umbel-rc: = http://dbpedia.org//umbel.org/umbel/rc/

rdf:type
or a
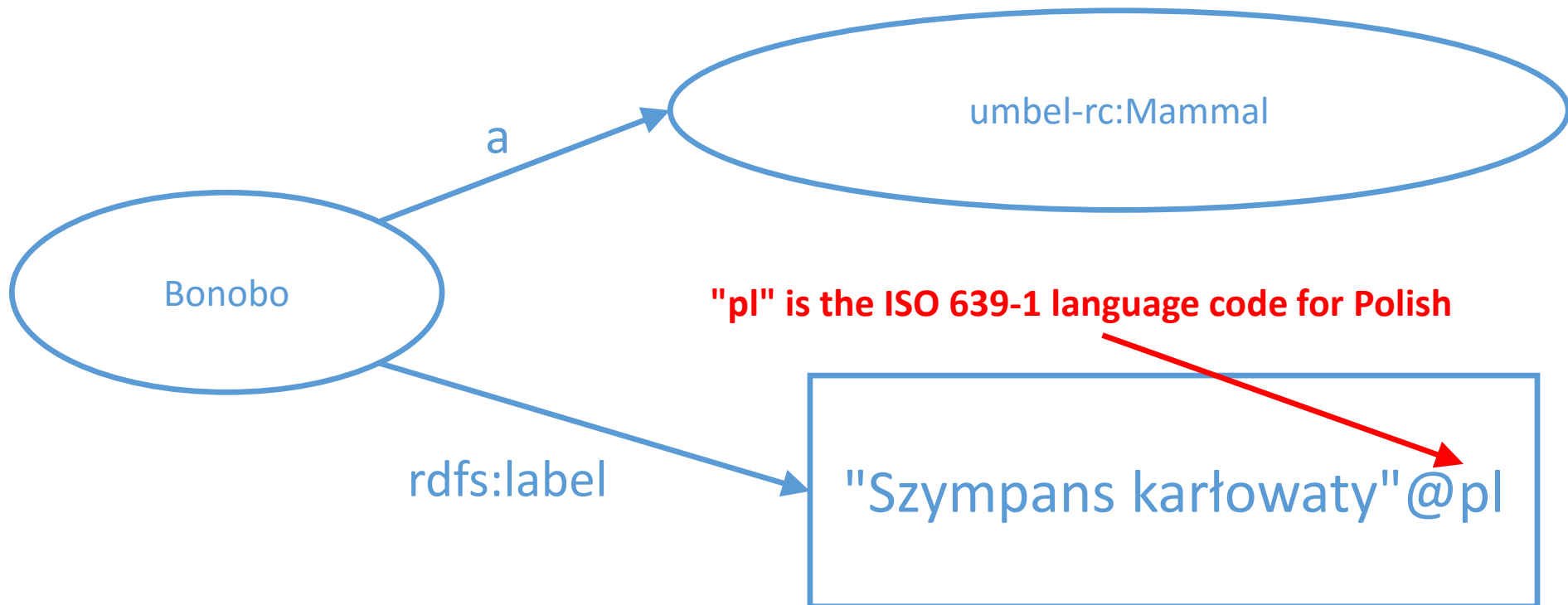
umbel-rc:Mammal

Bonobo

rdfs:label

"Szympans karłowaty"@pl
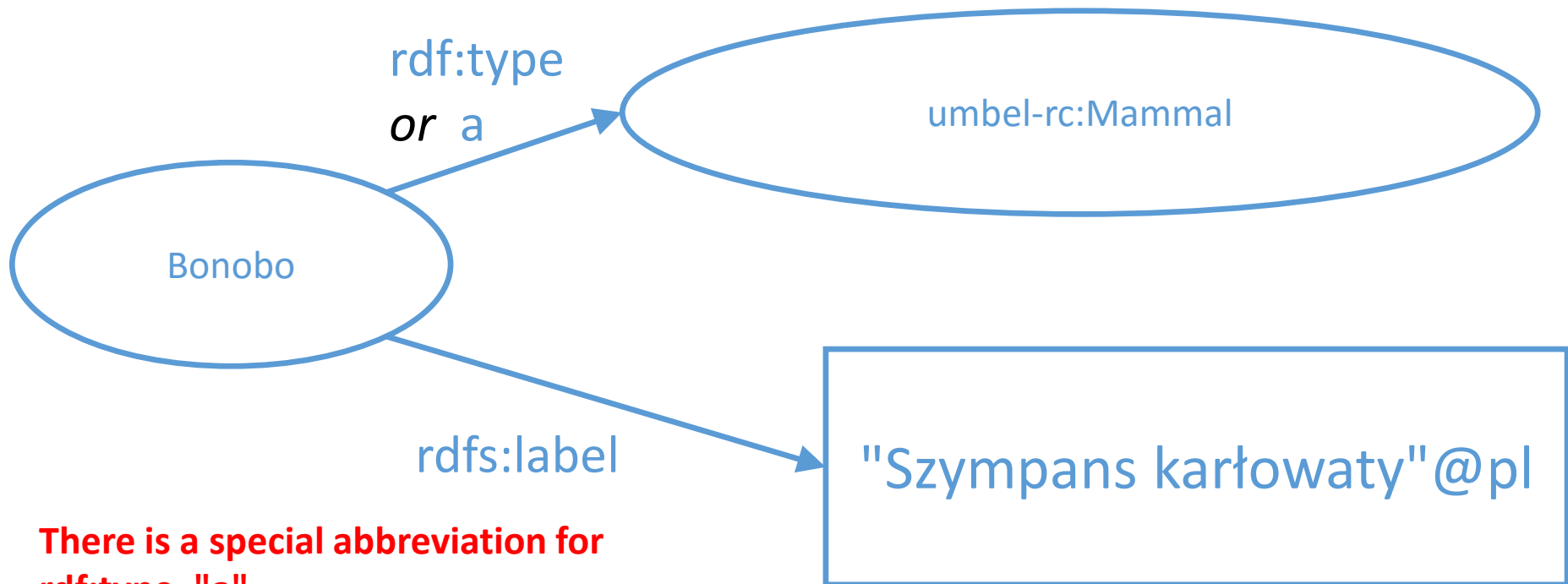
There is a special abbreviation for rdf:type, "a"
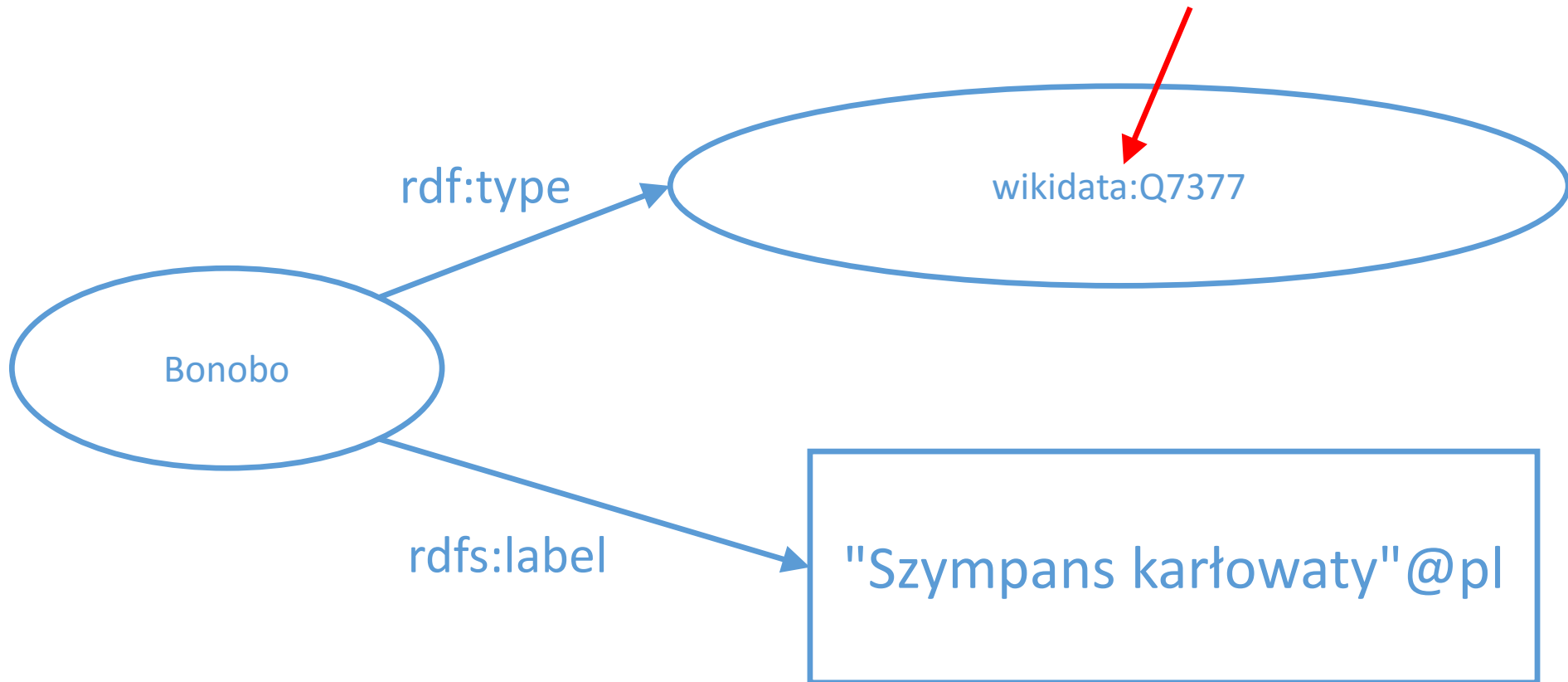
Literals are text strings, perhaps in a specific language

Literals are enclosed in quotation marks and may have a language tag. In this case, one might want to have many labels in different languages.

umbel-rc:Mammal

a

Bonobo

"pl" is the ISO 639-1 language code for Polish

rdfs:label

"Szympans karłowaty"@pl

Keep in mind that when you see a namespace followed by a colon, it's a globally unique URI, not a literal.

URIs may be constructed so that there last part makes sense to humans, but they don't have to. For example:

https://www.wikidata.org/wiki/Q7377 a.k.a. **wikidata:Q7377**

wikidata:Q7377

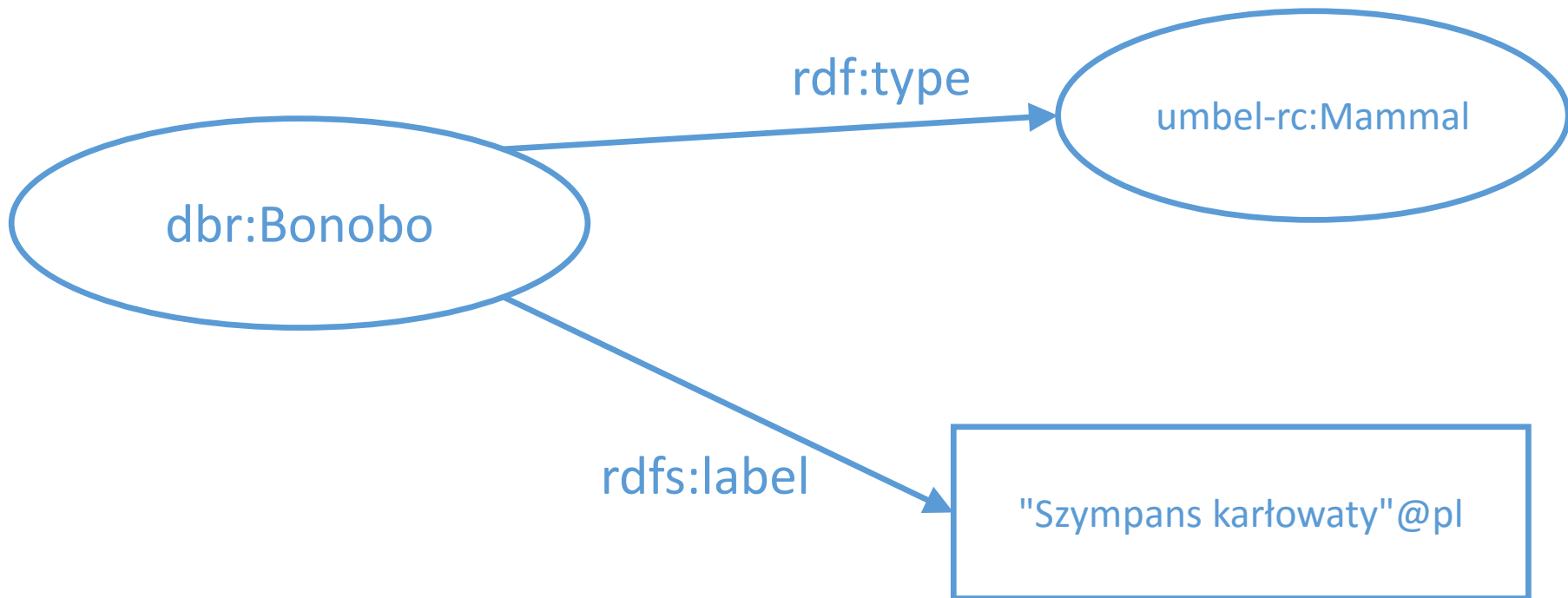rdf:type

Bonobo

rdfs:label

"Szympans karłowaty"@pl

HTTP URIs start with "http:" or "https:"

You might be able to find out something about them in a web browser.
This is Linked Data Principle #2.

Try **http://dbpedia.org/resource/Bonobo**
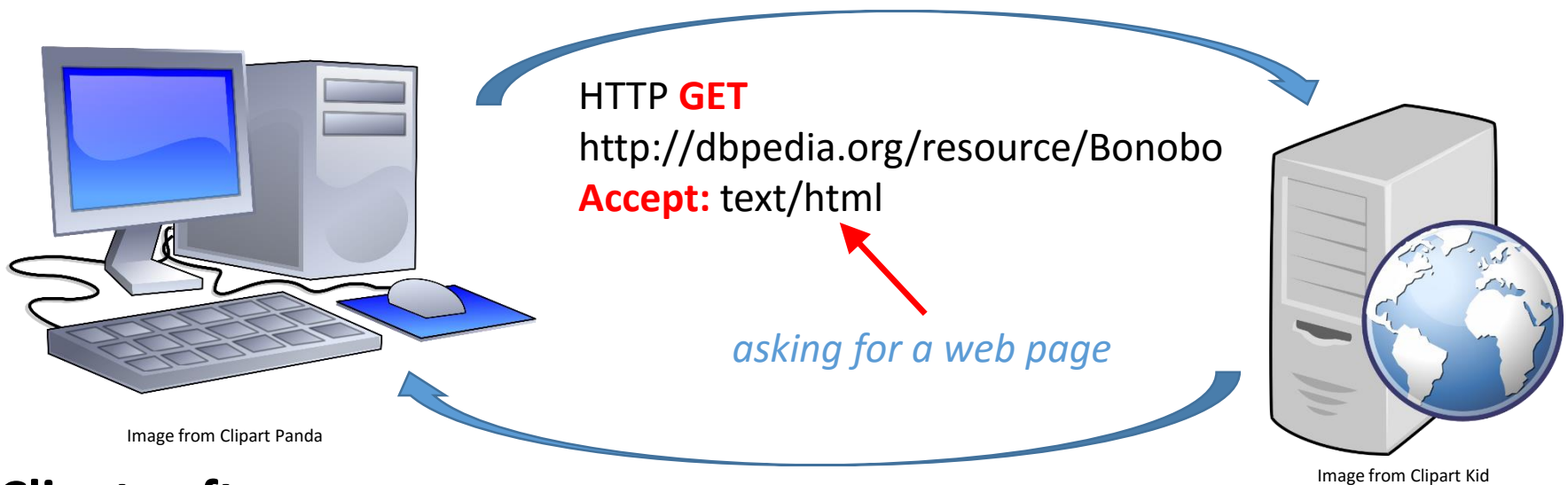
Linked Data Principle #3: Provide useful information when someone looks up a URI.

"Look up" means an interaction across the Internet mediated by Hypertext Transfer Protocol (HTTP). It's called "dereferencing a URI". (People also say "resolving" a URI.)

HTTP **GET**
http://dbpedia.org/resource/Bonobo
Accept: text/html

asking for a web page

Image from Clipart Panda

Image from Clipart Kid

Client software
(a.k.a. the "machine")
In this case, the client is a web browser. It displays the returned body as a web page.

HTTP **Status:** 200 OK
Body:
<?xml version="1.0" encoding="UTF-8" ?><!DOCTYPE html PUBLIC "-//W3C//DTD
XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd"><html
xmlns="http://www.w3.org/1999/xhtml"
xmlns:dbpprop="http://dbpedia.org/property/"
…

Web server

How can we see what's going on when a client interacts with a server?

- The book shows how to use the command-line utility cURL. A more user-friendly option is client called Postman, a Chrome extension (google "postman plugin").

- This time, let's ask for machine-readable data instead of a web page

HTTP **GET**
http://dbpedia.org/resource/Bonobo
Accept: **text/turtle**

Excerpts from the served RDF/Turtle file:

```
# here's all the namespace abbreviations:
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dbr: <http://dbpedia.org/resource/> .

# Here's the actual graph data:
dbr:Bonobo rdf:type umbel-rc:Mammal .
dbr:Bonobo rdfs:label "Szympans kar\u0142owaty"@pl .
```
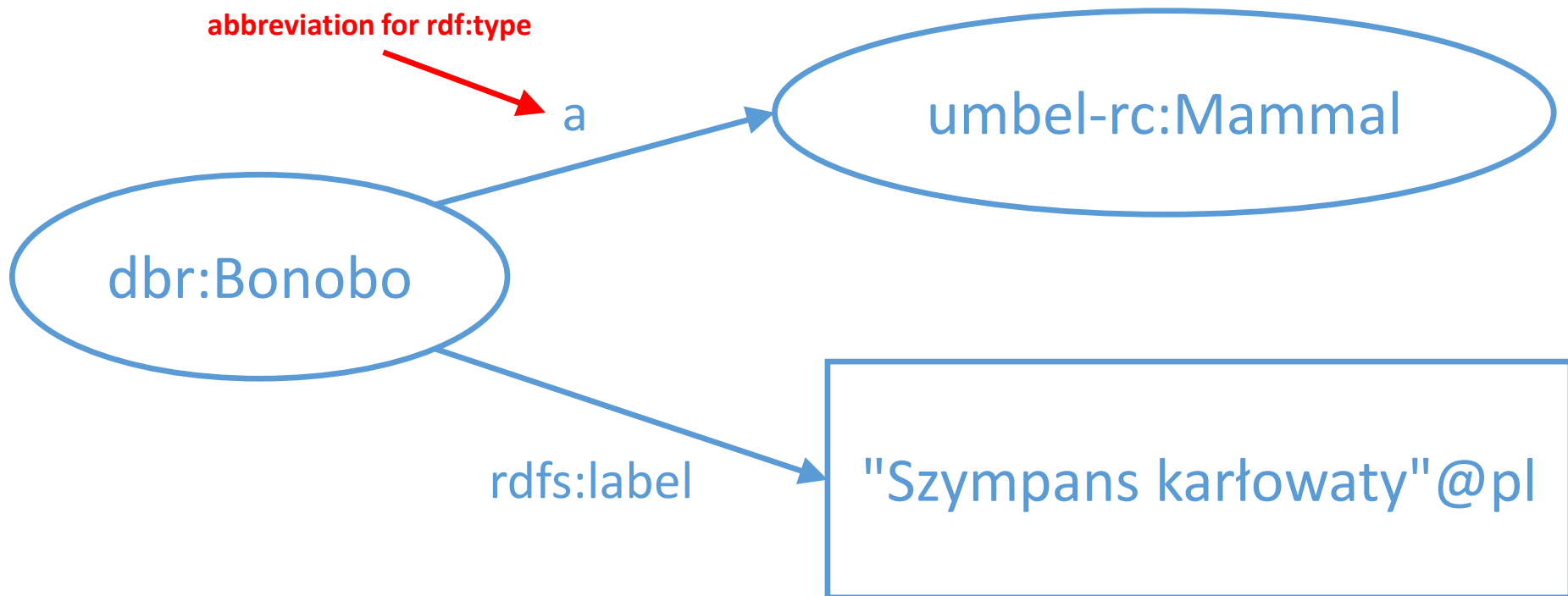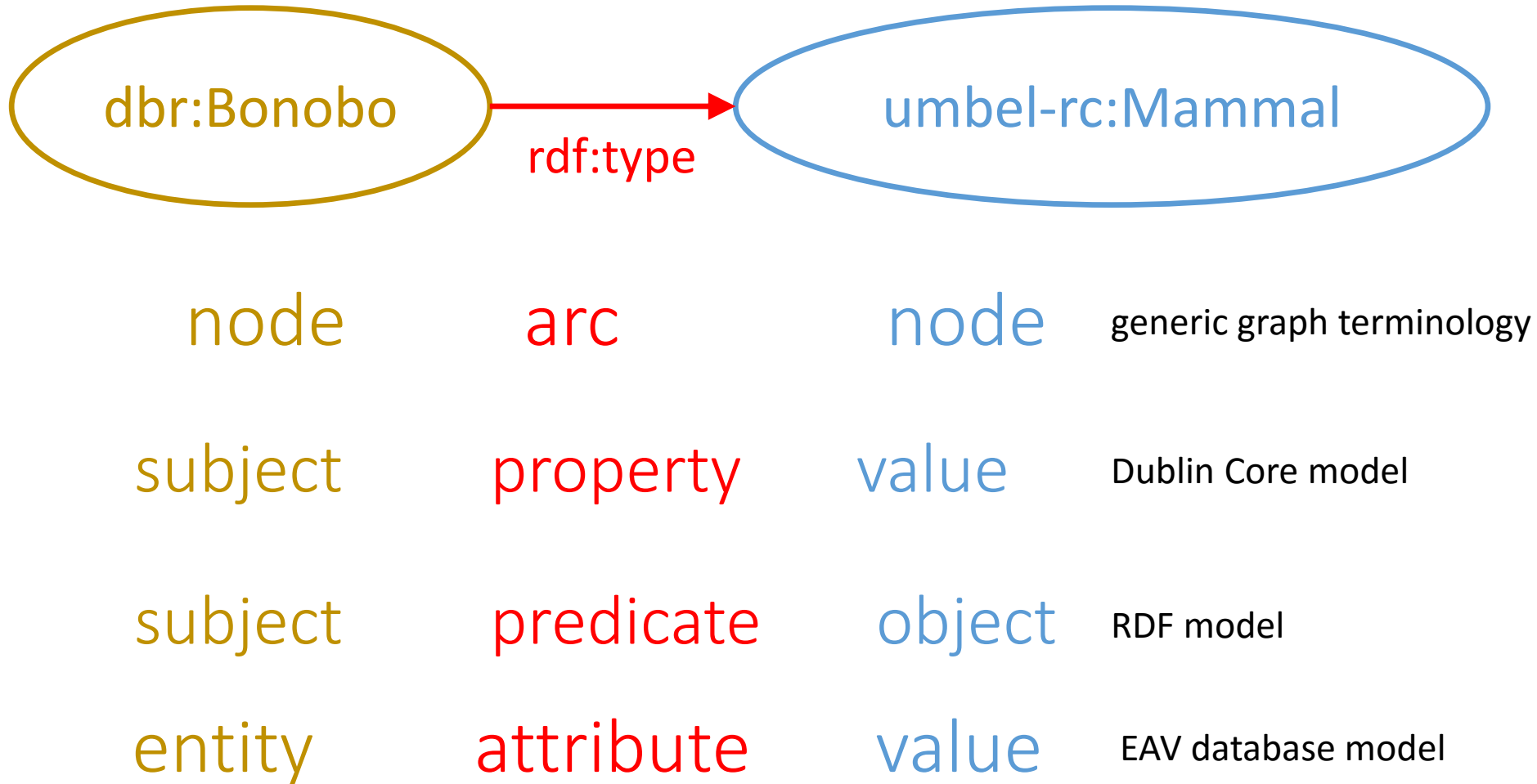
Remember that this is an abbreviation for rdf:type

a

umbel-rc:Mammal

dbr:Bonobo

rdfs:label

"Szympans karłowaty"@pl

Ways we talk about links:

```
# Here's the actual graph data:
dbr:Bonobo rdf:type umbel-rc:Mammal .
```


dbr:Bonobo	rdf:type	umbel-rc:Mammal	
node	arc	node	generic graph terminology
subject	property	value	Dublin Core model
subject	predicate	object	RDF model
entity	attribute	value	EAV database model

Linked Data Principle #4: Include links to other URIs. (Learn new stuff!)

- We know that baskauf:kitty is the same kind of thing as dbr:Bonobo
- We infer that both are also instances of class umbel-rc:Vertebrate