

What is RDF?

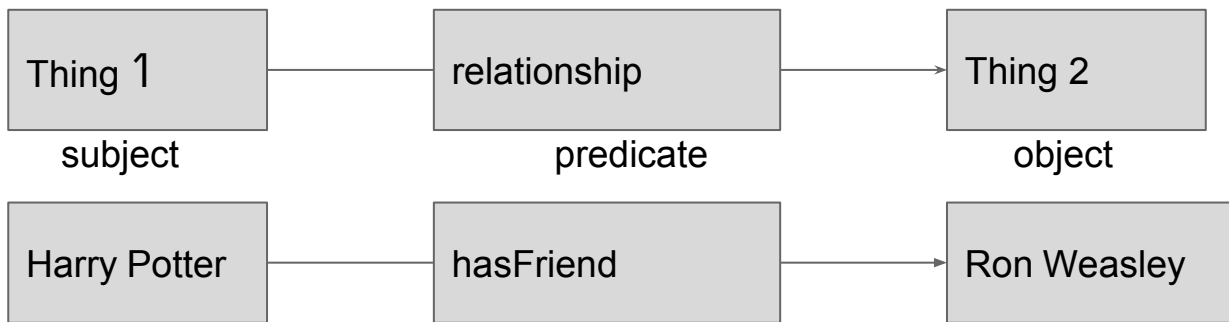


SWWG: *Linked Data* Chapters 1.7-2.6

Resource Description Framework

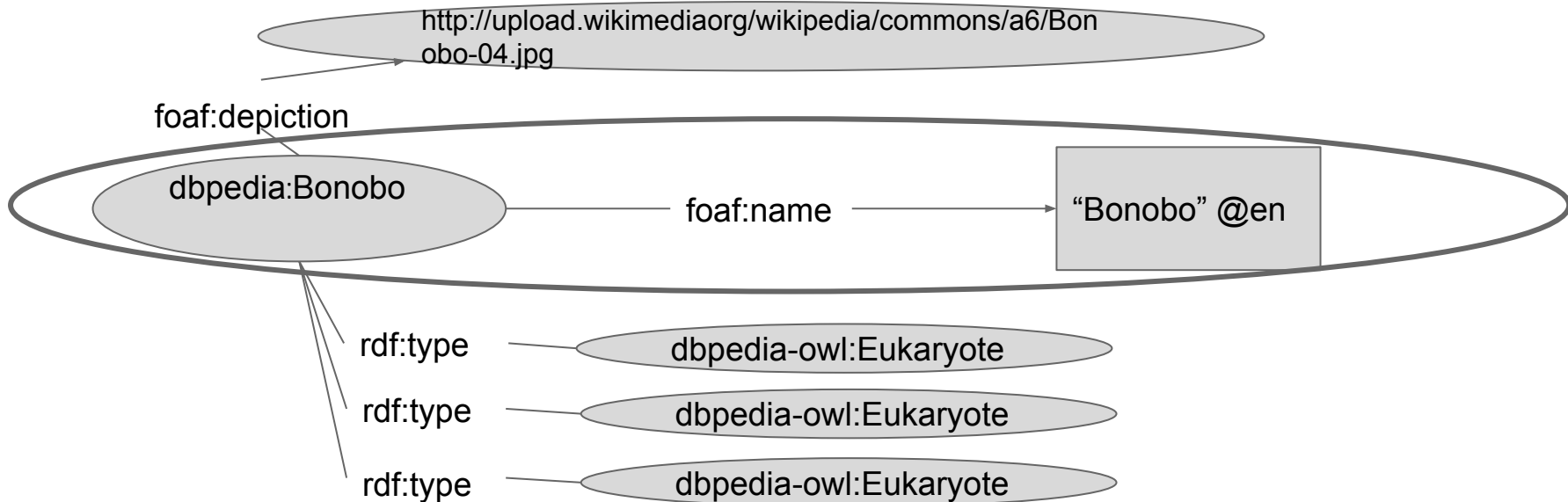
- Linked data uses RDF as a data *model*
- RDF statements describe two things, and a relationship between them. This is referred to as a “triple”

Ex.



RDF: graph

- Multiple triples together make a graph. Inserting authorized URIs for things eliminates ambiguity in who/what you are talking about



Web Example

- Example: Environmental Protection Agency website (EPA)
- Data from lots of sources - not coming from EPA site itself
- Even info that is from EPA is from various sources/databases that weren't designed to work together

The screenshot shows the EPA website page for the Us Tva Browns Ferry Nuclear Plant. The page is titled "Us Tva Browns Ferry Nuclear Plant" and "Facility ID: 110000589355". It features a map of the plant location, a list of data sources, and a list of releases. The page is annotated with arrows pointing to various data sources:

- EPA's Facilities Registry System**: Points to the facility name and ID.
- Wikipedia**: Points to the text "The Browns Ferry Nuclear Plant is located on the Tennessee River near Decatur [...]"
- Open Street Map**: Points to the map of the plant location.
- Wikimedia Commons**: Points to the aerial photograph of the plant.
- Source Data!**: Points to the "View Data as KML" and "View Data as Turtle" links.
- EPA's Facilities Registry System**: Points to the "Screen Address" and "Mailing Address" sections.
- EPA's Toxics Release Inventory**: Points to the "Released" section.

The "Released" section lists various chemicals and their release dates:

- Released in 2009
 - Zinc compounds in 2009
 - Lead in 2009
 - Lead in 2008
 - Lead in 2007
 - Lead in 2006
 - Lead in 2005
 - Lead in 2004
 - Lead in 2003
 - Lead in 2002
 - Lead in 2001
 - Lead in 2000
 - Polycyclic aromatic compounds in 2009
 - Lead in 2009
 - Lead in 2008
 - Lead in 2007
 - Lead in 2006
 - Lead in 2005
 - Lead in 2004
 - Lead in 2003
 - Lead in 2002
 - Lead in 2001
 - Lead in 2000
- Tetrahydrofuran in 1998
- Zinc compounds in 1998
- Tetrahydrofuran in 1999
- Zinc compounds in 1999

Figure 1.10 A Linked Data application published by the U.S. EPA. The page shown describes the Browns Ferry nuclear power plant near Decatur, Alabama. Note the different data sources combined to form the page.

Using shorthand for URIs

- Multiple triples together make a graph. Inserting authorized URIs for things eliminates ambiguity in who/what you are talking about

<http://www.manning.com/dwood>

Book

<http://purl.org/dc/elements/1.1/creator>
or dc:creator

Created by

<http://viaf.org/viaf/266916656>

Wood, David, 1963-

You can use shorthand if you name the URI with a prefix

Prefixes

| Prefix | Namespace URI |
|--------|---|
| dc: | http://purl.org/dc/elements/1.1/ |
| foaf: | http://xmlns.com/foaf/0.1 |
| rdf: | http://www.w3.org/1999/02/22-rdf-syntax-ns# |
| rdfs: | http://www.w3.org/2000/01/rdf-schema# |

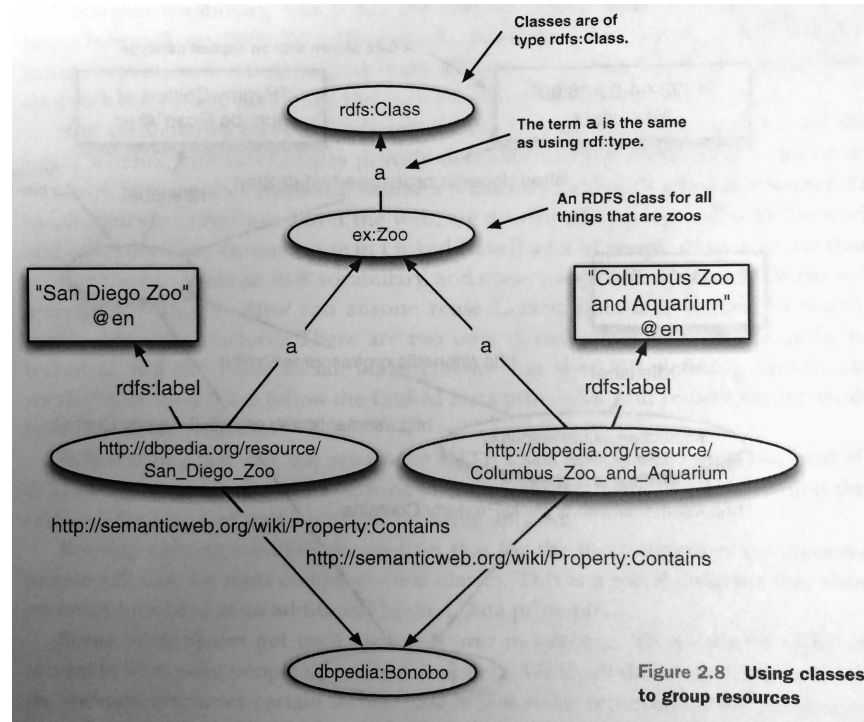
Minting URIs

- Name things with URIs
- Use a DNS domain that you control
- Use natural keys
 - E.g. http://paulsbakery.example.com/baked_goods/bread/rye-12 *versus* <http://paulsbakery.example.com/984d66a>
- Make URIs neutral to implementation details
 - Some sites can have server infrastructure changed by web admins, meaning all their URIs will change (bad for others who may have bookmarked pages or already written RDF using those older URIs)
- Use fragment identifiers with caution

Classes

- Resources can be divided into groups using property `rdf:type`

“San Diego Zoo” is a
literal - a simple string



Blank Nodes

- A URI without a name
- Temporary - URI cannot be relied upon
- Use case: insert into collection of items when you don't want to bother making up your own URI for it
- Many people avoid them because they can't be queried later
- Useful when you don't want to claim responsibility for minting a new URI

Literals

- Strings, or nodes in a triple that are not represented by a linkable URI
 - E.g. “Harry Potter”
- Literals are a dead end: they cannot become the subject of any new triples or queries
- Good for data like numerals, some dates, or other values that don't need to be linked.

Serializations

- RDF data can be expressed in a variety of ways.
- Common serializations:
 - Turtle,
 - N-triples,
 - RDFS,
 - JSON-LD
- Turtle is pretty common bc it's uses a lot of shorthand - the book uses this