Linked Data

Adila Krisnadhi

Hello!

I am Adila Krisnadhi

Faculty member at the Faculty of Computer Science, Universitas Indonesia.

Co-director of Tokopedia-UI AI Center of Excellence

Ontology engineer and Semantic Web enthusiast

Credits

◈ Presentation template by SlidesCarnival

◈ Photographs by Unsplash

◈ Backgrounds by SubtlePatterns

◈ Figures are taken from Allemang, Hendler, Gandon, "Semantic Web for the Working Ontologist", 3rd Ed.

The Layer Cake: The Technology Components

Smart (Cognitive) Applications & Services

Trust

Proof

Unifying Logic
First-Order Logic (FOL)

Rules
SWRL, SPIN, R2RML, SHACL

Dictionaries
(Ontologies)
RDF, RDFS, OWL,
SKOS, Schema.org

Transmission
Security
(Crypto)

Query
SPARQL, SPASQL

Abstract Language
RDF Subject->Predicate->Object Sentences

Sentence Part Identifiers

HTTP IRIs & URIs
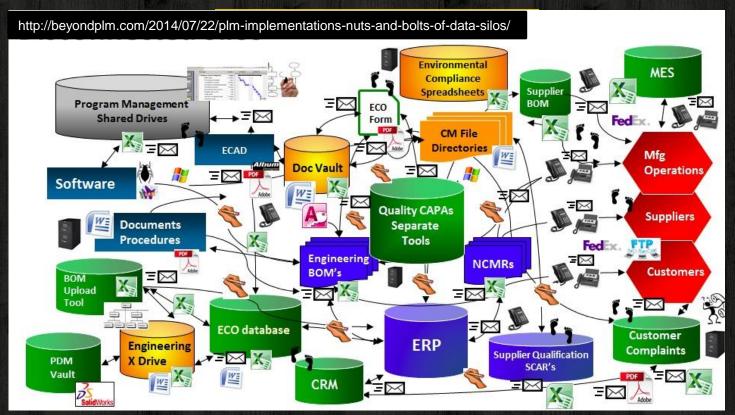
Document Types

RDF-NTriples, RDF-Turtle, RDF-XML, RDF-JSON, JSON-LD, others

Semantic Web of Linked Data

Traditional model: Data silos

http://beyondplm.com/2014/07/22/plm-implementations-nuts-and-bolts-of-data-silos/

Data on the Web vs. Web of Data

- ◈ Data on the Web = make data available on the Web, e.g., spreadsheets, web tables, PDFs.
- ◈ Web of data = data (from around the world) linked together so that it can be found, browsed, crawled, integrated, etc.
- ◈ Semantic Web → Web of data, not just data on the Web.
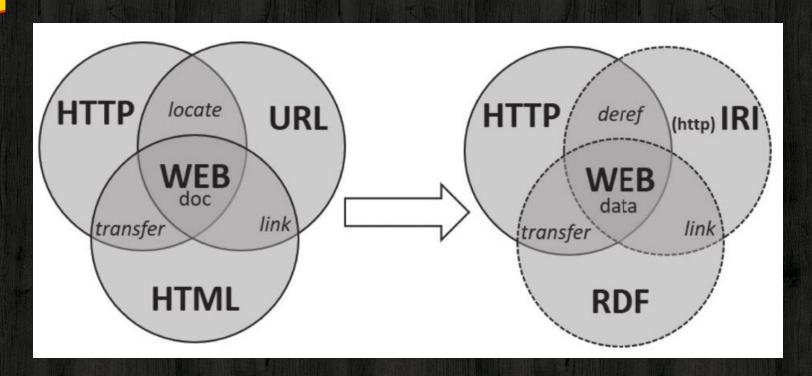
Open data vs. linked data

◈ **Open data**
  ○ Openly available (on the Web) under an open license.
  ○ But not necessarily free to (re)use, e.g., due to proprietary file formats.
  ○ "Open data and content can be freely used, modified, and shared by anyone for any purpose" – http://opendefinition.org
  ○ Open data portals: data.gov, data.gov.uk, data.id, etc.

◈ **Linked data = Web of data**
  ○ Practical and simple Web of data
  ○ Employs W3C Semantic Web standards focusing on simple semantics.
  ○ Data may be closed.
  ○ Linked Open Data = linked data that is also open data.

How do we make (hypertext) Web allowing Web of data?

Web of Data: Content production

- ◈ **Embedded in web pages**
  - ○ Microformats
  - ○ Microdata
  - ○ RDFa
  - ○ JSON-LD

- ◈ **Standalone RDF data**
  - ○ RDF files
  - ○ APIs (backed by triple store, RDB, etc.)

Linked Data, Semantic Web

Microformats

◈ Structured data in embedded web pages.

◈ **Existing HTML tags** are used for information representation.

◈ Information represented with a set of properties and their values

◈ Formats tailored at specific need

◈ Enables more focused search, more accurate classification and visualization of search results.

◈ Can be transformed into RDF with GRDDL (XSLT)

◈ See examples at https://microformats.org/wiki/microformats2

Microdata

◈ Extends HTML5 specification with primitives for information representation.

◈ Simple semantics (compared to Semantic Web standards)

◈ W3C Working Draft: https://www.w3.org/TR/microdata/

◈ Representative example vocabulary: schema.org
  ○ Schemas for structured data mark-up on the Web
  ○ Launched together in Juni 2011 by Google, Yahoo, Microsoft

Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for data types.

- Thing
  - Action
    - AchieveAction
      - LoseAction
      - TieAction
      - WinAction
    - AssessAction
      - ChooseAction
        - VoteAction
      - IgnoreAction
      - ReactAction
        - AgreeAction
        - DisagreeAction
        - DislikeAction
        - EndorseAction
        - LikeAction
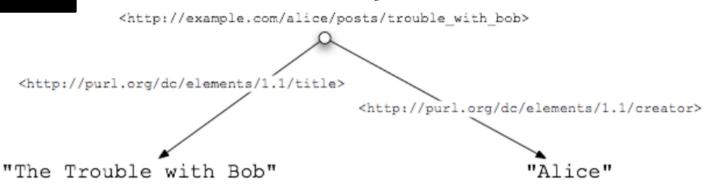
See examples at https://schema.org/docs/documents.html

12

RDFa

- ❖ W3C Recommendation to embed RDF on an XHTML page
  - ○ Bridging human and data webs
  - ○ Software can extract RDF graph for the machine.
  - ○ See playground at http://rdfa.info/
  - ○ RDFa distiller and parser:
    http://www.w3.org/2012/pyRdfa/
- ❖ Schema.org support since RDFa 1.1
- ❖ Open Graph Protocol (OGP)
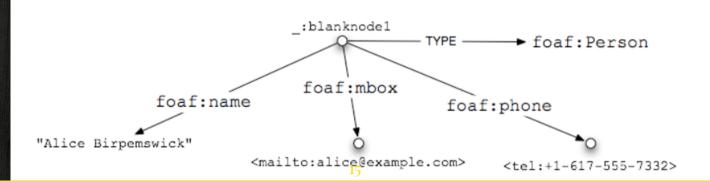 - ○ Major user of RDFa
  - ○ Used in Facebook

RDFa Example

```
<div>
    <h2>The trouble with Bob</h2>
    <h3>Alice</h3>
    ...
</div>
```

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/">
    <h2 property="dc:title">The trouble with Bob</h2>
    <h3 property="dc:creator">Alice</h3>
    ...
</div>
```

<http://example.com/alice/posts/trouble_with_bob>

<http://purl.org/dc/elements/1.1/title>

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

"The Trouble with Bob"

"Alice"

Literal Properties: RDFa lets Alice connect not just one URL to another—for example to connect her blog entry URL to the Creative Commons license URL— but also to connect one URL to a string such as "The Trouble with Bob". All arrows are labeled with the corresponding property name, which is also a URL.

```
<div typeof="foaf:Person" xmlns:foaf="http://xmlns.com/foaf/0.1/">
    <p property="foaf:name">
      Alice Birpemswick
    </p>

    <p>
      Email: <a rel="foaf:mbox" href="mailto:alice@example.com">alice@example.com</a>
    </p>

    <p>
      Phone: <a rel="foaf:phone" href="tel:+1-617-555-7332">+1 617.555.7332</a>
    </p>
</div>
```
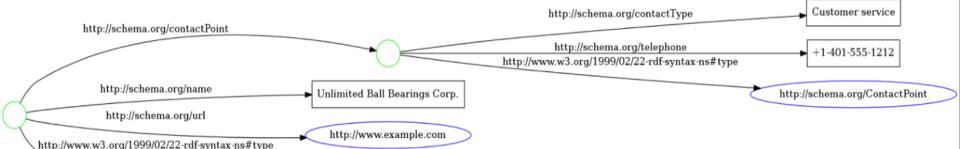
JSON–LD

◈ W3C Recommendation:
https://www.w3.org/TR/json-ld11/

◈ JSON-based RDF serialization format

◈ Can be directly embedded on an HTML page

◈ Schema.org supports

◈ Google recommends JSON-LD over RDFa and Microdata

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "Organization",
  "url": "http://www.example.com",
  "name": "Unlimited Ball Bearings Corp.",
  "contactPoint": {
    "@type": "ContactPoint",
    "telephone": "+1-401-555-1212",
    "contactType": "Customer service"
  }
}
</script>
```


http://schema.org/contactPoint

http://schema.org/contactType — Customer service

http://schema.org/telephone — +1-401-555-1212
http://www.w3.org/1999/02/22-rdf-syntax-ns#type

http://schema.org/name — Unlimited Ball Bearings Corp.

http://schema.org/url — http://www.example.com

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

http://schema.org/ContactPoint

http://schema.org/Organization

17

Linked Data: Main ideas

◈ Creation of (as many as possible) datasets in RDF

◈ Linking datasets together
 ○ Cross referencing data in other datasets
  e.g., place "Indonesia" in GeoNames to president "Joko Widodo" in DBPedia
 ○ Identifying same concepts in different datasets
  e.g., "Jakarta" in GeoNames and DBPedia

◈ Employing lightweight semantic technologies

◈ Linked Open Data (LOD) community

Linked data principles

1. Use URIs to name everything.
2. Use HTTP URIs to allow people looking up those names.
3. When someone/agent looks up a URI, provide useful information using the standards (RDF and its syntaxes).
4. In the returned information, include links to HTTP URIs of other things so that agent can discover more things.

5–star rating in linked data publishing

★	make your stuff available on the Web (whatever format) under an open license[1]
★★	make it available as structured data (e.g., Excel instead of image scan of a table)[2]
★★★	make it available in a non-proprietary open format (e.g., CSV instead of Excel)[3]
★★★★	use URIs to denote things, so that people can point at your stuff[4]
★★★★★	link your data to other data to provide context[5]
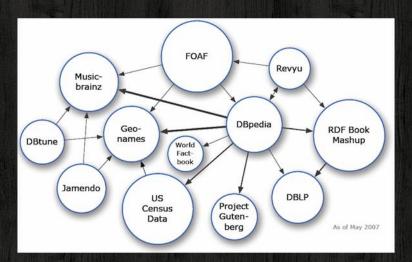
https://5stardata.info/en/

LOD Cloud (of datasets)

2007

As of September 2010

LOD Cloud

◈ 2022?
   ○ Too big to display here...
   ○ See https://lod-cloud.net

Knowledge graph in Google search

Linked Data: How do we name things with URI?

Naming Things with URIs to reach 4 stars

◈ Use HTTP URIs
  ○ Domain name system makes distributed URI minting easy
  ○ HTTP URIs can be used to access more information; so, avoid URNs, DOIs, etc.
◈ Need to avoid confusion between the objects themselves and the Web documents that describe them.
  ○ http://www.cs.ui.ac.id is a university or a Web document about it?

Cool URIs

◈ See: Tim Berners-Lee, "Cool URIs don't change", http://www.w3.org/Provider/Style/URI and https://www.w3.org/TR/cooluris/

◈ Cool URIs are **simple**
  ○ Prefer short, mnemonic URIs (not easily broken when copy-pasted and easier to remember)

◈ Cool URIs are **stable**
  ○ Once set-up to identify certain resources, URIs should live forever.
  ○ Keep implementation-specific out of URIs (e.g., .php, .asp)

◈ Cool URIs are **manageable**
  ○ Issue your URIs in a way that you can manage,
  ○ e.g., include current year (so can change URI schema each year without breaking older URIs)
  ○ Keep all 303 URIs on a dedicated subdomain, e.g., http://id.example.com/alice to ease later migration of the URI-handling system.

URI dereferencing (3rd LD principle)

◈ **Content negotiation**: Returning the right information about a URI using HTTP
  - As specified by HTTP header request fields
  - E.g., Accept: text/plain
  - E.g., a web browser requests a HTML page.

◈ Server decides what to return.

◈ Two dereferencing strategies (to allow distinction between objects:
  - 303 URIs
  - Hash URIs

303 URIs

◆ Problem: real world objects/abstract concepts cannot be returned; only documents can.

◆ 303 URI strategy:
  ○ Server return a 303 See Other response code (HTTP redirect)
  ○ Client can find the related document by following the redirect, e.g., to obtain a HTML page or an RDF description.

◆ Example:
  http://data.finlex.fi/eli/sd/2008/521/luku/1/pykala/1/ajantasa/20160101
  is redirected to
  http://data.finlex.fi/eli/sd/2008/521/luku/1/pykala/1/ajantasa/20160101.html

◆ Good for accessing parts of RDF descriptions (including the large ones).
  ○ Requires one extra HTTP request.
  ○ But only relevant data is transmitted and no further filtering is needed.

Hash URIs

◈ 303 URI needs two requests when dereferencing

◈ Hash URI: use # and fragment identifier.
  ○ E.g., http://example.org/vocab/course#Seminar

◈ Procedure:
  ○ Client truncates URI at # (remove #Seminar)
  ○ Clients sends a GET request for http://example.org/vocab/course
  ○ Server returns either an RDF or HTML document.
  ○ The whole document is returned, not just info about #Seminar (because the fragment is not seen by the server).

◈ Good for small RDF descriptions (e.g., schemas)
  ○ Can save extra HTTP request
  ○ But the client must filter the response further.

Linked data interfaces: How one interacts with it

- ❖ Human application interface.
  - ○ Normal browsing/searching
- ❖ Reading RDF data of a URI, based on URI dereferencing
- ❖ Linked data browsing: Linked data browser interface based on URI dereferencing
  - ○ Browsing based on RDF properties rendered in HTML (e.g., in DBPedia)
  - ○ Using Linkd data browsers for the Web of data (e.g., http://uriburner.com)
- ❖ SPARQL endpoint: querying data in a standard way for mashups, etc.
- ❖ Download data as an RDF data dump.

DBPedia.org URI model

Use several URIs

◈ URI for the real-world object itself.
http://dbpedia.org/resource/Jakarta

◈ URI for a related information resource that describes the real-world object and has an HTML representation
http://dbpedia.org/page/Jakarta

◈ URIs for a related information resource that describes the real-world object and has an RDF representation
http://dbpedia.org/data/Jakarta
http://dbpedia.org/data/Jakarta.ttl
http://dbpedia.org/data/Jakarta.json
...

In addition ....

◈ Wikipedia pages for Jakarta
https://en.wikipedia.org/wiki/Jakarta
https://id.wikipedia.org/wiki/Daerah_Khusus_Ibukota_Jakarta
...

◈ Jakarta as Wikidata item
https://www.wikidata.org/wiki/Q3630

Data linking across datasets to reach 5 stars

Add as many (high quality) links as possible (manual or automated):

◈ **Relationship links**: pointers to additional information
  ○ E.g., persons to places where they lived.

◈ **Identity links**: pointers to similar resources in other datasets

◈ **Vocabulary/ontology links**: pointers to vocabulary terms in metadata models or ontologies in related datasets
  ○ E.g, use Dublin Core vocabulary to describe provenance information such as dates, authors, etc.
  ○ E.g., use keyword thesauri for subject descriptions.

Linked data publishing examples

◈ Wikidata, DBPedia, GeoNames, ...

◈ (Some) open data portals: data.gov.uk, data.gov

◈ ...

◈ See also W3C's Working Group Note: Data on the Web Best Practices Use Cases & Requirements, https://www.w3.org/TR/dwbp-ucr/

Fin.