Linked Data

The World is Your Database
Agenda

1. What is Linked Data, and why is it good for you (15 mins)
2. How it works (tech talk) (15 mins)
3. How you can get started (5 mins)
4. Discussion and Q&A (5-10 mins)
What is Linked Data
Before Linked Data: Information was locked inside proprietary databases, each of which used custom database schema and every database record was accessed by identifiers that were only unique and intelligible within the system where they originated.

The guardians of these data ‘fortresses’ didn’t like sharing data. If it had to be done at all then data was extracted and delivered under lock and key with cryptic instructions on how to use it.
After Linked Data: every resource (concept, name, database record, etc.) has a globally Unique Resource Identifier (URI) that is intelligible to anyone else on the planet.

- MeSH (Medical Subject Headings): *Internal Medicine*  

- Getty Art & Architecture Thesaurus: *Renaissance*  
  [http://vocab.getty.edu/aat/300021140](http://vocab.getty.edu/aat/300021140)

- Library of Congress Name Authority: *Barack Obama*  
  [http://id.loc.gov/authorities/names/n94112934](http://id.loc.gov/authorities/names/n94112934)

- GeoNames: *London*  
  [http://www.geonames.org/2643743](http://www.geonames.org/2643743)

- European Environment Agency: *Inland surface waters*  

- WordNet Lexical Database: *finance*  
  [http://wordnet-rdf.princeton.edu/wn31/101136358-n](http://wordnet-rdf.princeton.edu/wn31/101136358-n)

**World Wide Web Database**
**Before Linked Data:** web hyperlinks were mere sign posts to other web pages.
After Linked Data: links become semantic, they express the specific reason why two entities are related.

These semantically expressive links, called predicates, assert factual statements and support machine reasoning. They also have their own URIs identifying their place in ontological schema.
Is Linked Data Always Open Data?

If I create Linked Data is it automatically exposed to the public?

**no**

*open is optional*

All Linked Data is *capable* of being shared but Linked Data can also reside behind the firewall.
Why it is good for you
1. Adopting and re-using Linked Open Data taxonomies
2. Semantic Enrichment
Example 1: Adopting and re-using Linked Open Data Taxonomies
New mantra for taxonomy projects:

ADOPT first

ADAPT second

CREATE third
Linked Open Data Taxonomy Sources

- Trusted authorities
- Many different subject domains
- Millions of concepts
- Many sources in the public domain
- Standard electronic format
- Live query and/or download

Jump start your taxonomy project
When Do Third Party Taxonomies Work Best

Corporate & Enterprise Taxonomies
- Products & Services
- Commodities
- Finance
- Legal & Regulatory

Products & Services
- Commodities
- Finance
- Legal & Regulatory

STEMs: Science, Technology, Engineering & Mathematics
- HCLS: Health Care & Life Sciences
  - Cultural Heritage
  - News Media
  - Geospatial
  - Person Names

© 2016 Synaptica, LLC  www.synaptica.com
Example 1: Getty AAT, IconClass & DBpedia Imported for Search / Browse

DBpedia
- 1.3M primary resources
- 31M relationships & properties

Getty AAT
- 42K primary resources
- 14.7M relationships & properties

IconClass
- 40K primary resources
- 3.4M relationships & properties

LCNAF
- 9.5M primary resources
- 80M relationships & properties

LCSH
- 419K primary resources
- 3.9M relationships & properties

Total in this system
- 11.3M primary resources
- 133M relationships & properties
Example 1: Getty AAT, IconClass & DBpedia Imported for Search / Browse

Process of indexing a visual detail inside a painting to the DBpedia category resource for **Lutes**

http://dbpedia.org/resource/Category:Lutes

via the FOAF predicate **Depicts**

foaf:depicts
Example 1: Semantic Indexing Using Linked Open Data Taxonomies
Example 1: Linked Open Data – a World Wide Database

Bacchus puts a starry crown on Ariadne’s head...
http://iconclass.org/92L12112

Bacchus, Venus and Ariadne
Tintoretto, 1576 to 1577 CE
Doge’s Palace, Venice

Love at first sight
http://iconclass.org/33C211

Lover deserted woman
http://iconclass.org/33C312

Ship of Theseus

Bacchus
https://www.wikidata.org/wiki/Q645312

Ariadne
https://www.wikidata.org/wiki/Q184874

Bacchus and Ariadne
Titian, 1520-1523 CE
National Gallery, London

Theseus abandons Ariadne
Pompeian Fresco, 1st C. CE
National Archaeological Museum, Naples

For image attributions please visit www.linkedcanvas.org
Example 2: Mapping to Linked Open Data resources for Semantic Enrichment
Example 2: Semantic Enrichment – Internal record has minimal data

Use Case
An internal taxonomy of person names stores only minimal information: First Name, Last Name and a field for Biographic Notes.
Example 2: Semantic Enrichment – search for equivalent in external LOD source

By clicking on the Linked Data button we can automatically search for our person in DBpedia or any other Linked Data source such as VIAF or LCNAF.

Matching results are returned and the chosen entity selected.
Example 2: Semantic Enrichment – mapping and choosing properties

A mapping relationship connects the internal taxonomy record to the external Linked Data resource. In this instance the OWL predicate `sameAs` is used.

Lastly selections from an array of descriptive properties can be made.
1. Before semantic enrichment we had only our name and some brief biographic notes.

2. After semantic enrichment we had these plus an authoritative name, a detailed biographic abstract, birth date, birth place.

3. The new data can be seamlessly blended with internal data for presentation in screens, reports and export files.

4. URI links are stored in the background to provide authoritative provenance information on where this data came from and when.
How it works (tech talk)
Fundamental Concepts

Semantic Web, RDF, Triples, URIs, SPARQL

How is Linked Data different from other data and traditional relational databases?
name: Charles Babbage influenced name: Ada Lovelace

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://id.loc.gov/authorities/names/n50031102">http://id.loc.gov/authorities/names/n50031102</a></td>
<td>prov:influenced</td>
<td><a href="http://id.loc.gov/authorities/names/n78030997">http://id.loc.gov/authorities/names/n78030997</a></td>
</tr>
<tr>
<td><a href="http://id.loc.gov/authorities/names/n50031102">http://id.loc.gov/authorities/names/n50031102</a></td>
<td>foaf:name</td>
<td>&quot;Charles Babbage&quot;</td>
</tr>
<tr>
<td><a href="http://id.loc.gov/authorities/names/n78030997">http://id.loc.gov/authorities/names/n78030997</a></td>
<td>foaf:name</td>
<td>“Ada Lovelace”</td>
</tr>
</tbody>
</table>
knownFor
influenced
countryOfBirth

France

Blaise Pascal
Charles Babbage

United Kingdom

Mary Somerville
Ana Lovelace

Astronomy

Astronomy
Maths

Computer Science

Maths

Physics

Astrophysics

Physics

Maths

NT

NT

NT

NT
Pattern Matching and Nested Queries with Triplestores

Computational Time

Relational Database

Graph Database

Size of Dataset
Why use triple stores instead of other types of graph database?

RDF-XML
JSON-LD
N-Triples
N-Quads
Turtle
TriG
N3
Linked Open Data Vocabularies


LOD sources can be available for download or for live query... which method of access is best?
Importing and Exporting Linked Data Vocabularies

curl -H Content-type:text/turtle -upload-file ./authoritiesnames.nt
http://localhost:1234/repositories/linkedcanvas?graph=graphName

curl -X GET -H Accept:application/x-trig
http://localhost:1234/repositories/linkedcanvas
Describing a Primary Resource in Linked Data
Graph databases aren’t supposed to work that well with free text search… how can this be overcome?

1. http://vocab.getty.edu/queries#Full_Text_Search_Query
Free-Text Search in Linked Canvas
Graph Store HTTP

SPARQL endpoint

Static files in RDF formats
How to get started
Option 1: Get an Open Source Triple Store and Some Good Reads

http://graphdb.ontotext.com/

https://www.blazegraph.com/

Heath & Bizer
Morgan & Claypool (pub.)

Ruth, Wood & Zaidman
Manning (pub.)

Allemang & Hendler
Morgan Klaufman (pub.)
Option 2: Adopt, Adapt and Create Linked Data as part of your Taxonomy Management

- Trusted authorities
- Many different subject domains
- Millions of concepts

- Many sources in the public domain
- Standard electronic format
- Live query and/or download

Jump start your taxonomy project
Adopt any external ontology

List of Ontologies

- CIOOC CRM
- DBO
- DCT
- FOAF
- GEO
- NEDO
- OASIS
- OWL
- PROV
- RDFS
- SKOS

Edit Ontology

- Ontology URI: http://xmlns.com/foaf/0.1/
- Ontology Label: FOAF
- Custom URI: http://xmlns.com/foaf/0.1/
- Description: FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked.
Start minting HTTP-URI’s for everything you create

https://demo.linkedcanvas.com/concept/ug5q9mcsut2
Linked Data

1. A way to jump-start taxonomy projects and reduce costs
2. A way to tap into external knowledge that can help answer your business questions
3. A powerful tool for performing complex searches and building smart applications
Linked Data
The World is Your Database

Thank you!

Dave Clarke
Synaptica CEO

Gene Loh
Synaptica Software Architect

Visit us at Booth 201

© 2016 Synaptica, LLC
www.synaptica.com