Linked Data Generation

Semantic Web

enabled applications rely on
data represented as Linked Data

Linked Data

refers to a method of publishing and connecting structured data on the Web

Linked Data Principles

1. Use **URIs as names** for things
2. Use **HTTP URIs** so people can look up those names.
3. When someone **looks up a URI, provide useful information**, using the standards (RDF*, SPARQL)
4. Include **links to other URIs** so that they can discover more things.

(Berners-Lee, 2006)

Where is Linked Data derived from?

Linked Data

derived from originally heterogeneous (semi-)structured data

How do we employ
Linked Data principles
to obtain Linked Data
originating from raw data?

Linked Data Best Practices

1. Prepare stakeholders
2. Select a dataset
3. Model the data
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

Linked Data Best Practices

1. **Prepare stakeholders**
2. Select a dataset
3. Model the data
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

1. Prepare stakeholders - Linked Data Life Cycles

Linear:

- Hyland et al. or
- Hausenblas et al.

Iterative:

- Villazon-Terrazas et al.
- LOD2 Linked Open Data Lifecycle

(Government Linked Data Working Group, 2014)

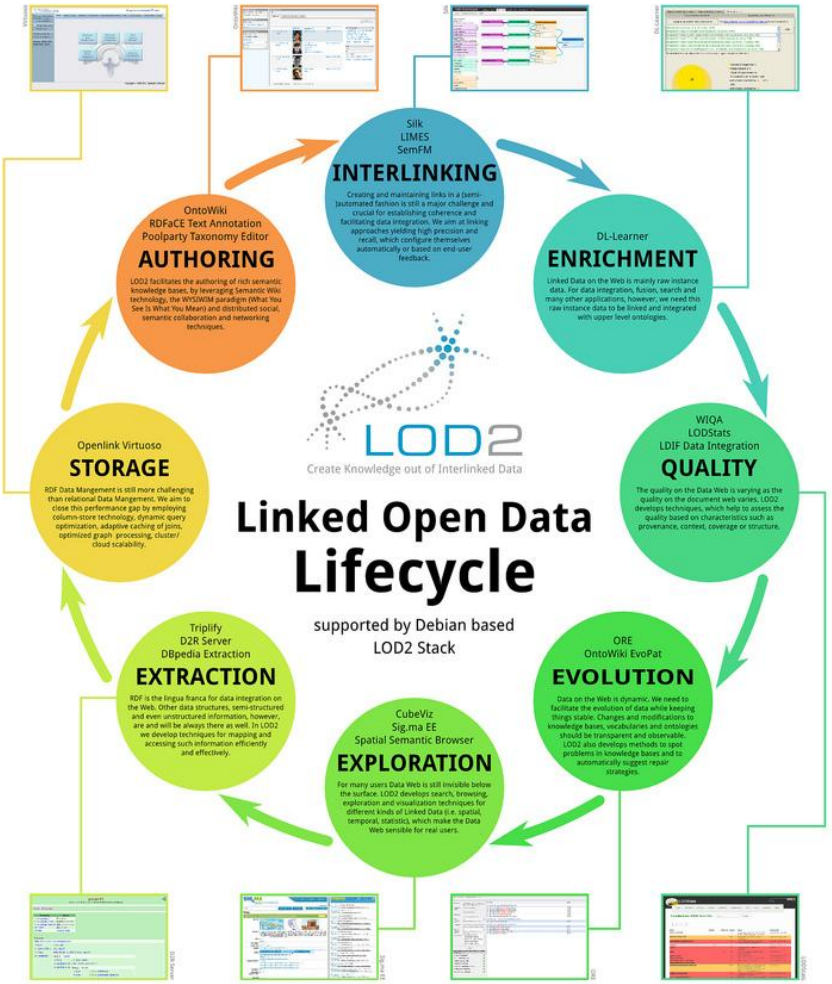
1. Prepare stakeholders - Linked Data Life Cycles

Linear:

(Hausenblas et al.)

1. Prepare stakeholders - Linked Data Life Cycles

Iterative:



(LOD2 LifeCycle)

Linked Data Best Practices

1. Prepare stakeholders
2. **Select a dataset**
3. Model the data
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

2. Select a dataset

2. Select a dataset

2. Select a dataset

2. Select a dataset

Linked Data Best Practices

1. Prepare stakeholders
2. Select a dataset
3. **Model the data**
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

3. Model the data

During the data modeling process, stakeholders are encouraged to

- identify entities, and
- describe how entities are related.

3. Model the data

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

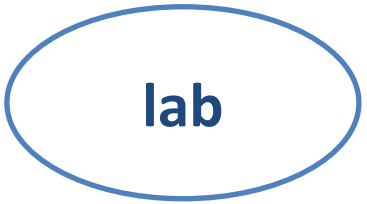
Which entities do you identify here?

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data



id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

How are the entities related?

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

Could the city also be an entity?

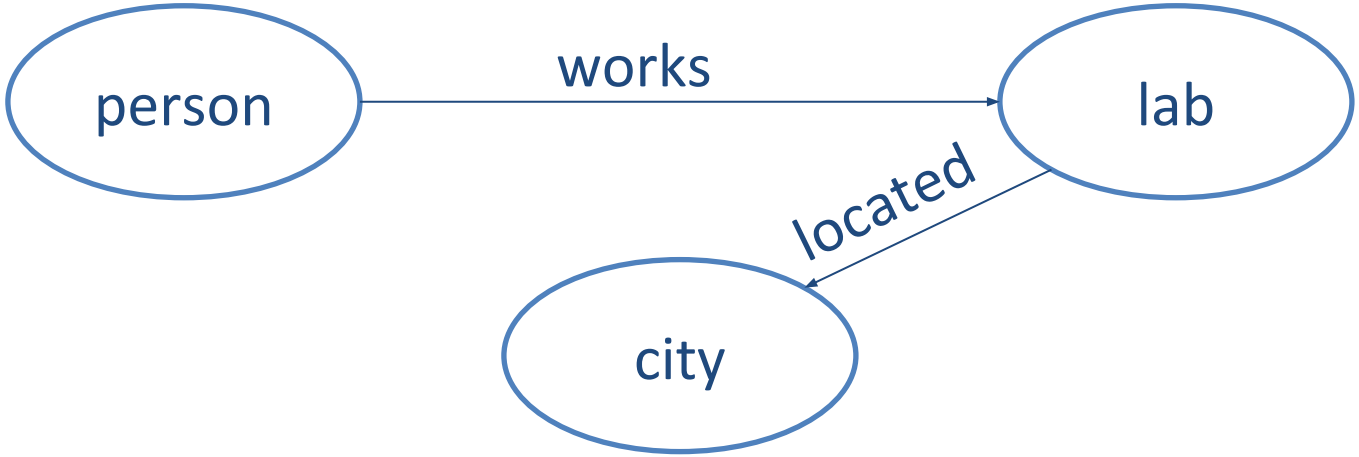
id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

3. Model the data

What if we had another data source with more information about cities?

id	city	country
A	Brussels	Belgium
B	Ghent	Belgium
C	Athens	Greece
D	Paris	France

3. Model the data



id	city	country
A	Brussels	Belgium
B	Ghent	Belgium
C	Athens	Greece
D	Paris	France

Linked Data Best Practices

1. Prepare stakeholders
2. Select a dataset
3. Model the data
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

Linked Data Best Practices

1. Prepare stakeholders
2. **Select a dataset**
3. Model the data
4. Specify an appropriate license
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

2. Select dataset(s)

2. Select dataset(s)

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data
4. **Specify an appropriate license**
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data
4. ~~**Specify an appropriate license**~~
5. Good URIs for Linked Data
6. Use standard vocabularies
7. Convert data

...

(Government Linked Data Working Group, 2014)

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data

~~Specify an appropriate license~~

4. **Good URIs for Linked Data**
5. Use standard vocabularies
6. Convert data

...

(Government Linked Data Working Group, 2014)

4. Good URIs for Linked Data - Linked Data Principles

- Use **URIs as names** for things
- Use **HTTP URIs** so people can look up those names.

4. Good URIs for Linked Data

http://example.com/{id}

http://example.com/{lab}

http://example.com/1
http://example.com/2

http://example.com/IDLab
http://example.com/IDLab

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

4. Good URIs for Linked Data

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	1	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	2	Ghent

4. Good URIs for Linked Data

http://example.com/{id}

http://example.com/{lab}

http://example.com/1

http://example.com/2

http://example.com/IDLab

http://example.com/1

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	1	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	2	Ghent

4. Good URIs for Linked Data

http://example.com/**person**/{id} http://example.com/**lab**/{lab}

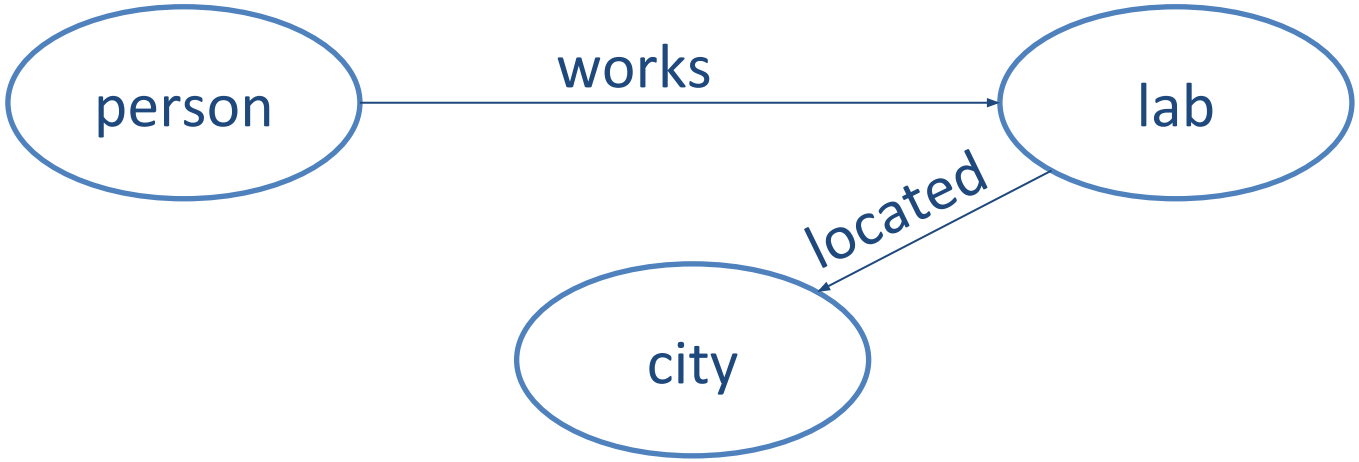
http://example.com/**person**/1 http://example.com/**lab**/IDLab
http://example.com/**person**/2 http://example.com/**lab**/IDLab

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

4. Good URIs for Linked Data

http://example.com/**person**/{id}

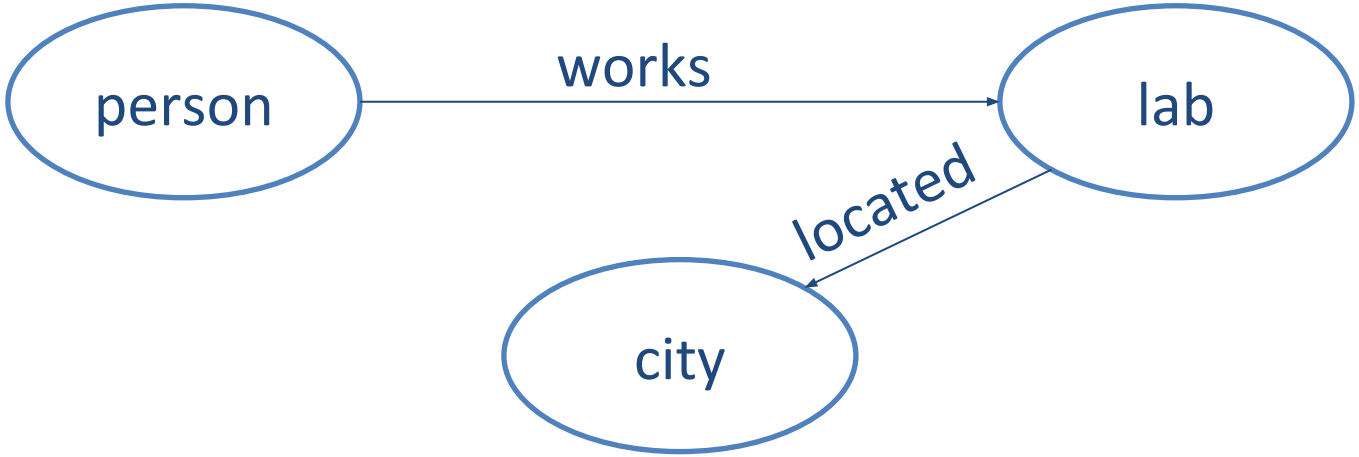
http://example.com/**lab**/{lab}



http://example.com/**city**/{city}

http://example.com/**city**/Ghent

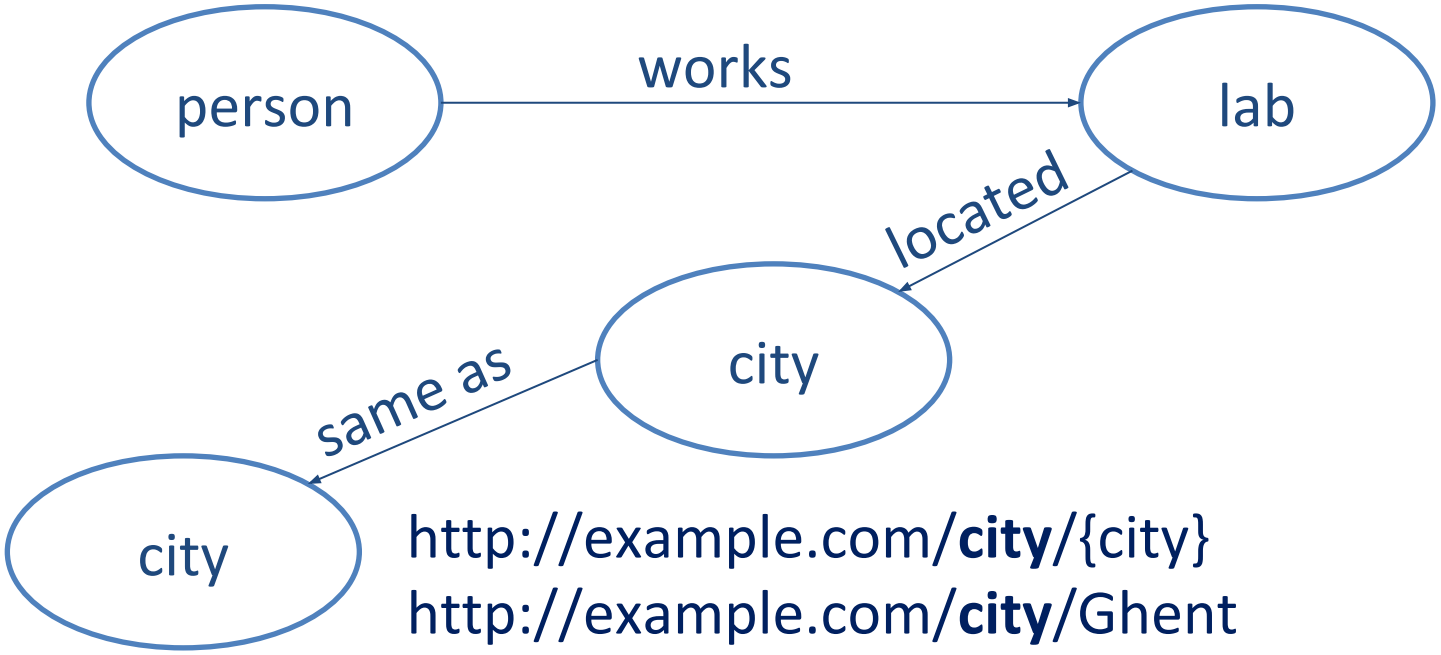
4. Good URIs for Linked Data



id	city	country
A	Brussels	Belgium
B	Ghent	Belgium
C	Athens	Greece
D	Paris	France

4. Good URIs for Linked Data

http://example.com/**person**/{id} http://example.com/**lab**/{lab}

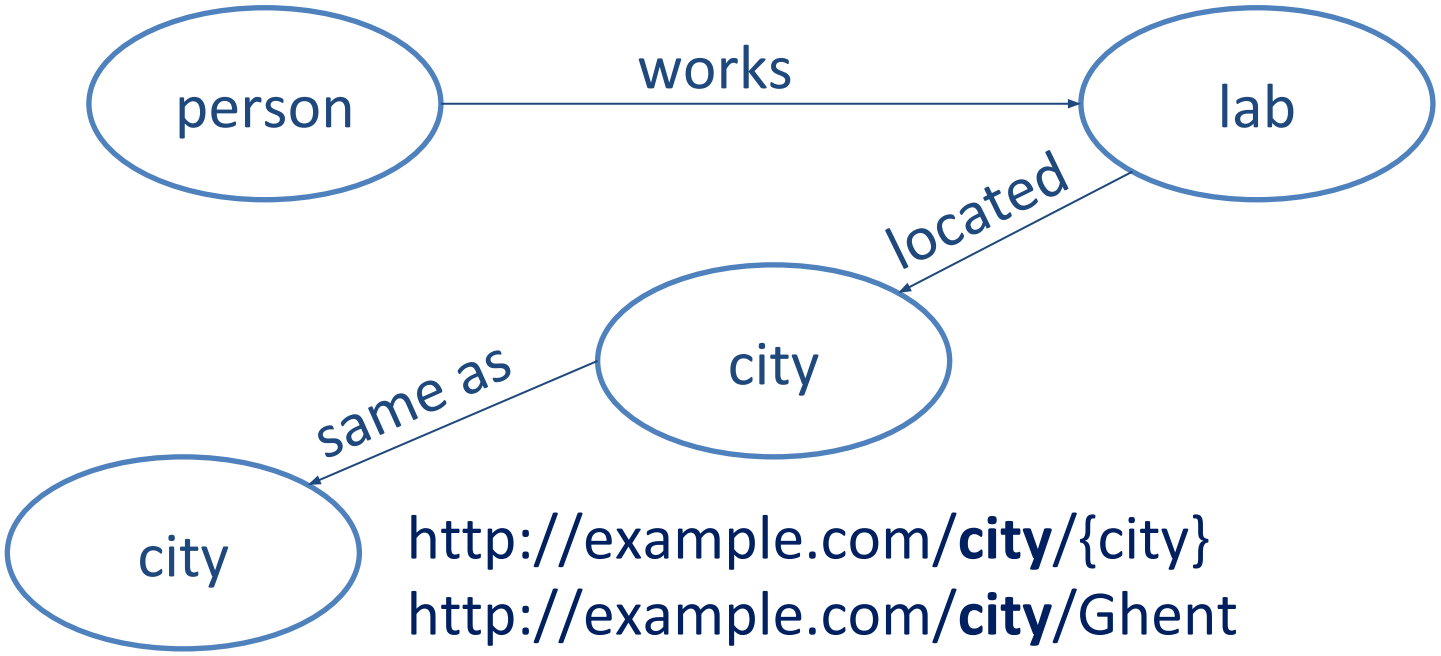


http://example.com/**city**/{city}
http://example.com/**city**/Ghent

4. Good URIs for Linked Data

http://example.com/**person**/{id}

http://example.com/**lab**/{lab}

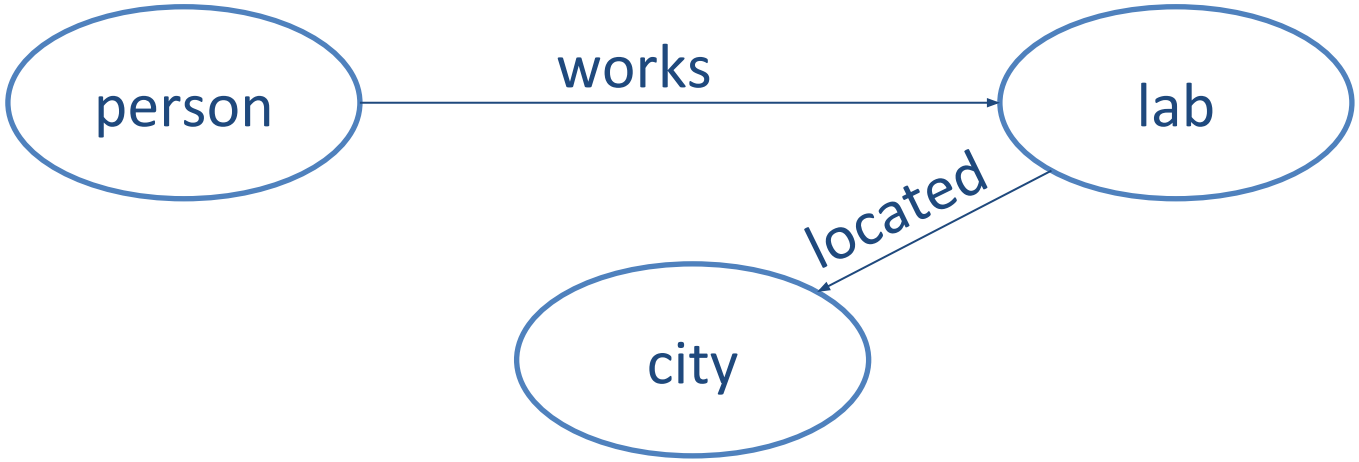


http://example.com/**city**/{city}
http://example.com/**city**/Ghent

http://other.com/**city**/{id}
http://other.com/**city**/B

4. Good URIs for Linked Data

http://example.com/**person**/{id} http://example.com/**lab**/{lab}



http://other.com/**city**/{id}
http://other.com/**city**/B

4. Good URIs for Linked Data - Cool URIs for the Semantic Web

Simplicity

Short, mnemonic URIs will not break as easily

Stability

Once a URI is set up to identify a certain resource, it should remain this way as long as possible

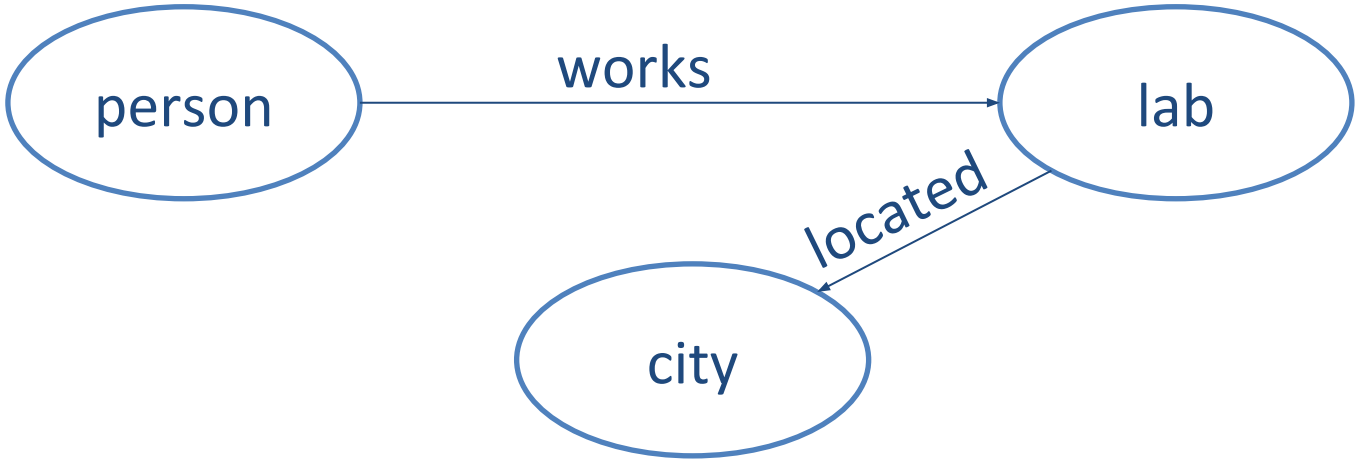
Manageability

Issue your URIs in a way that can be managed

(W3C Interest Group Note, 2008
Cool URIs don't change, Berners-Lee, 1998)

4. Good URIs for Linked Data

http://example.com/**person**/{id} http://example.com/**lab**/{lab}



http://other.com/**city**/{id}
http://other.com/**city**/B

How do I express the relationships between entities?

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data

~~Specify an appropriate license~~

4. Good URIs for Linked Data
5. **Use standard vocabularies**
6. Convert data

...

(Government Linked Data Working Group, 2014)

5. Use standard vocabularies

Vocabularies

a way to organize knowledge

used to tag units of information

Ontologies

vocabularies which are

- expressed in an ontology representation language,
- contain formal constraints on how terms can be used

5. Use standard vocabularies - W3C Recommended Vocabularies

DCAT
2014 - https://www.w3.org/TR/vocab-dcat/
Organization Ontology
2014 - https://www.w3.org/TR/vocab-org/
RDF Data Cube Vocabulary
2014 - https://www.w3.org/TR/vocab-data-cube/
SKOS Simple Knowledge Organization System
2009 - https://www.w3.org/TR/skos-reference/

5. Use standard vocabularies - W3C Group Notes

vCard Ontology
2014 - https://www.w3.org/TR/vcard-rdf/
Asset Description Metadata Schema (ADMS)
2013 - http://www.w3.org/TR/vocab-adms/
Registered Organization Vocabulary
2013 - https://www.w3.org/TR/vocab-regorg/

5. Use standard vocabularies - Most popular

WGS84 Geo Positioning
http://www.w3.org/2003/01/geo/wgs84_pos
DC
http://purl.org/dc/elements/1.1/
FOAF
http://xmlns.com/foaf/0.1/
DCTerms
http://purl.org/dc/terms/
DCMI
http://dublincore.org/documents/dcmi-box/

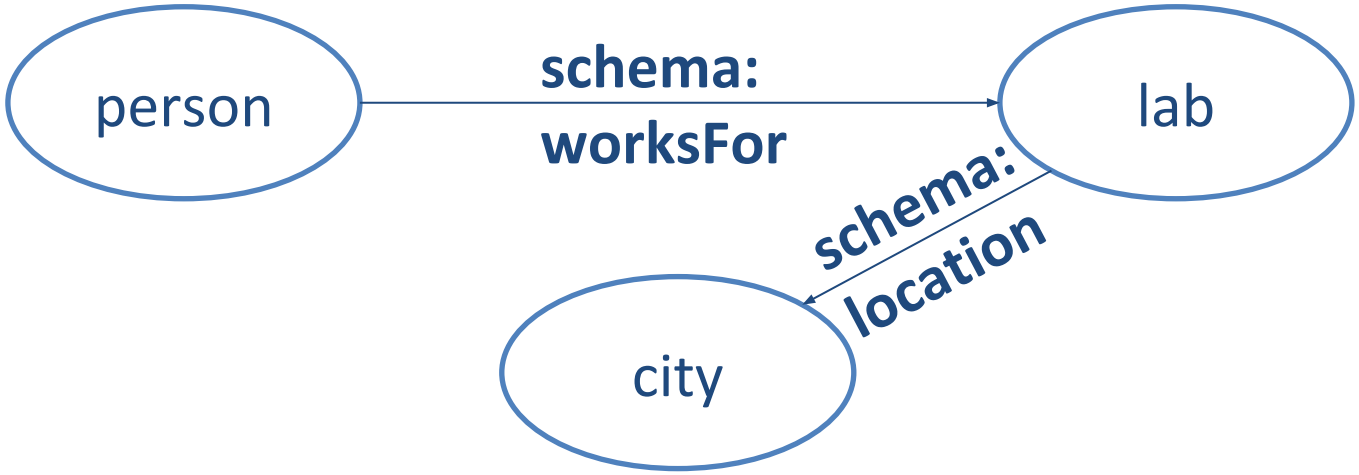
(http://stats.lod2.eu/)

5. Use standard vocabularies - More vocabularies

Linked Open Vocabularies
http://lov.okfn.org/

5. Use standard vocabularies

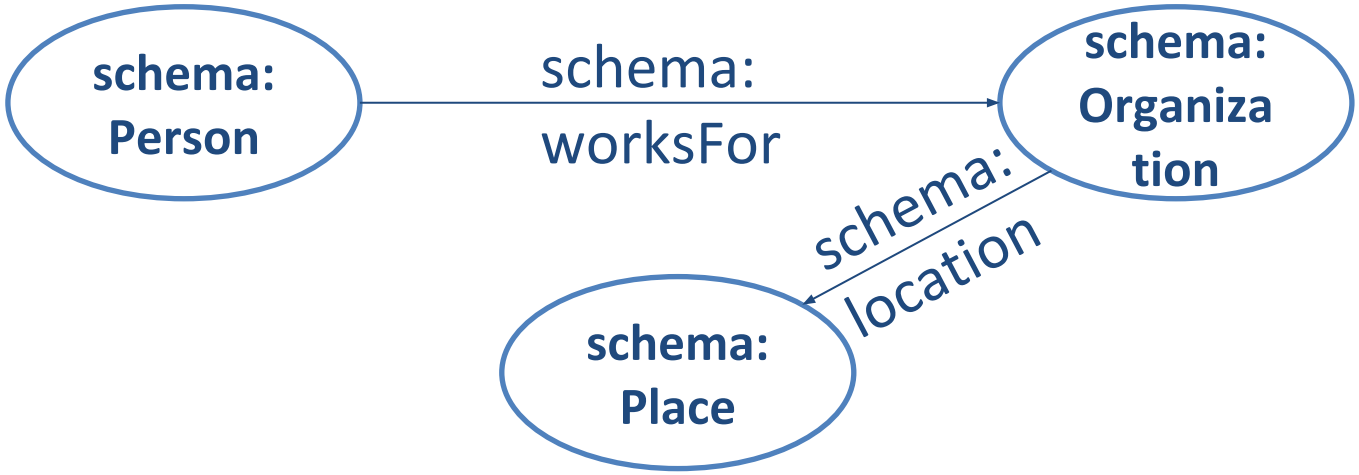
http://example.com/person/{id} http://example.com/lab/{lab}



http://other.com/city/{id}
http://other.com/city/B

5. Use standard vocabularies

http://example.com/person/{id} http://example.com/lab/{lab}



http://other.com/city/{id}
http://other.com/city/B

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data

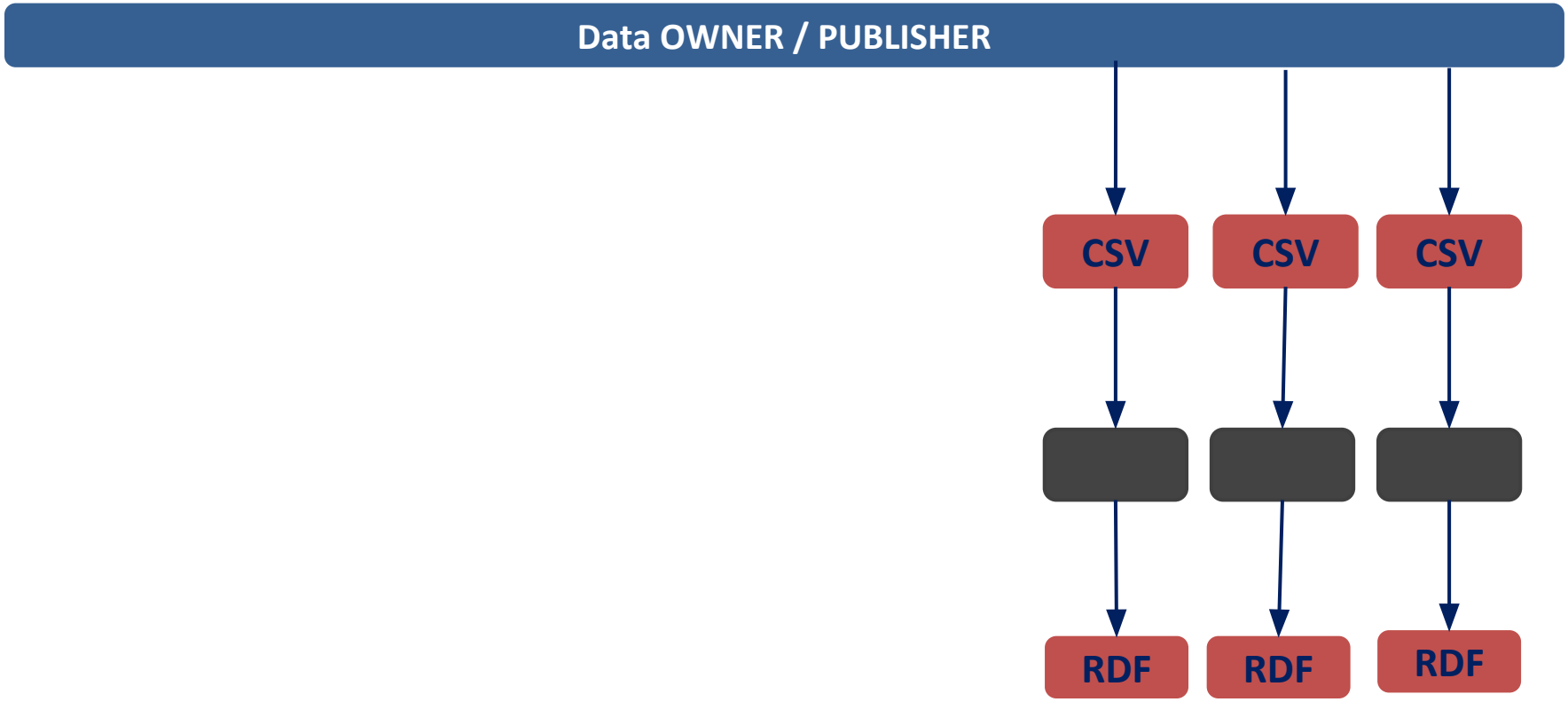
~~Specify an appropriate license~~

4. Good URIs for Linked Data
5. Use standard vocabularies
6. **Convert data**

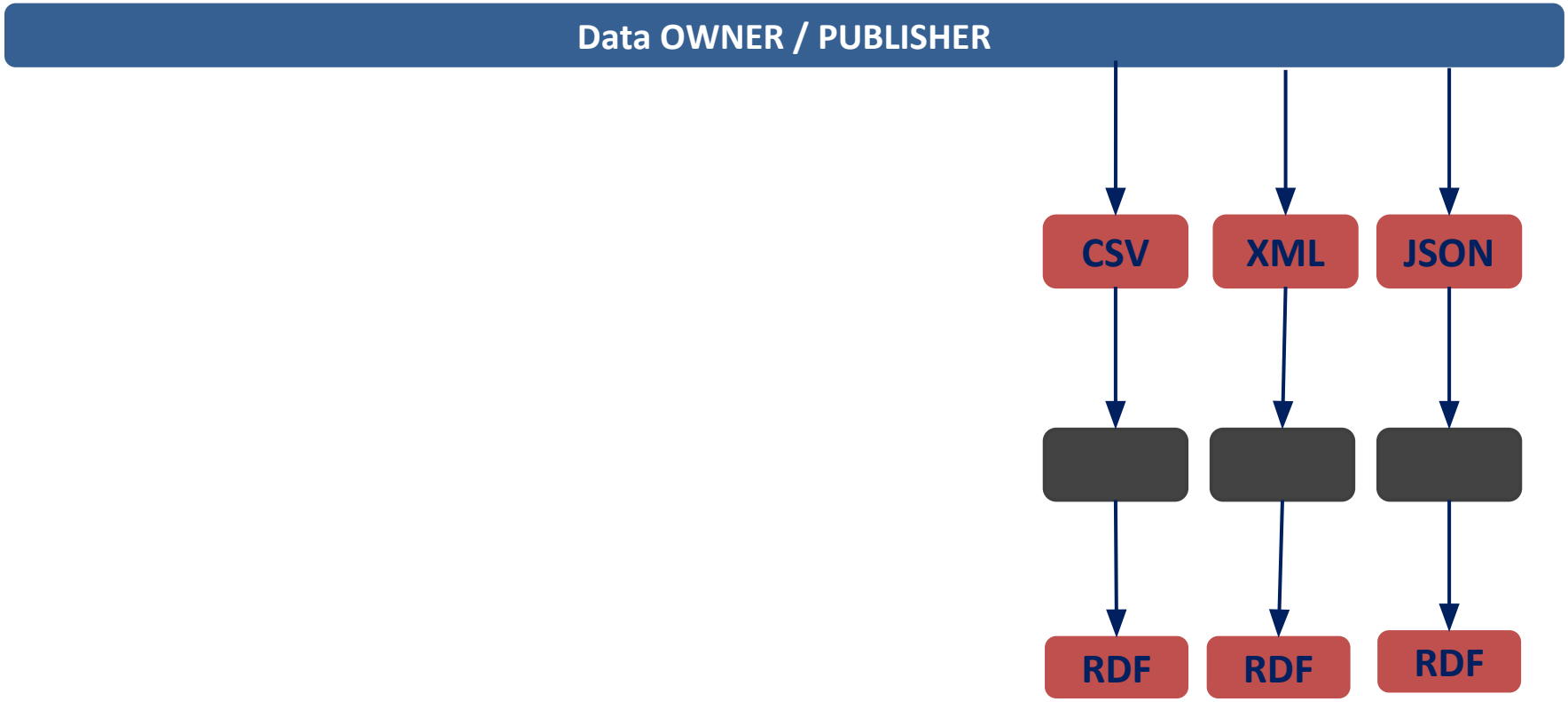
...

(Government Linked Data Working Group, 2014)

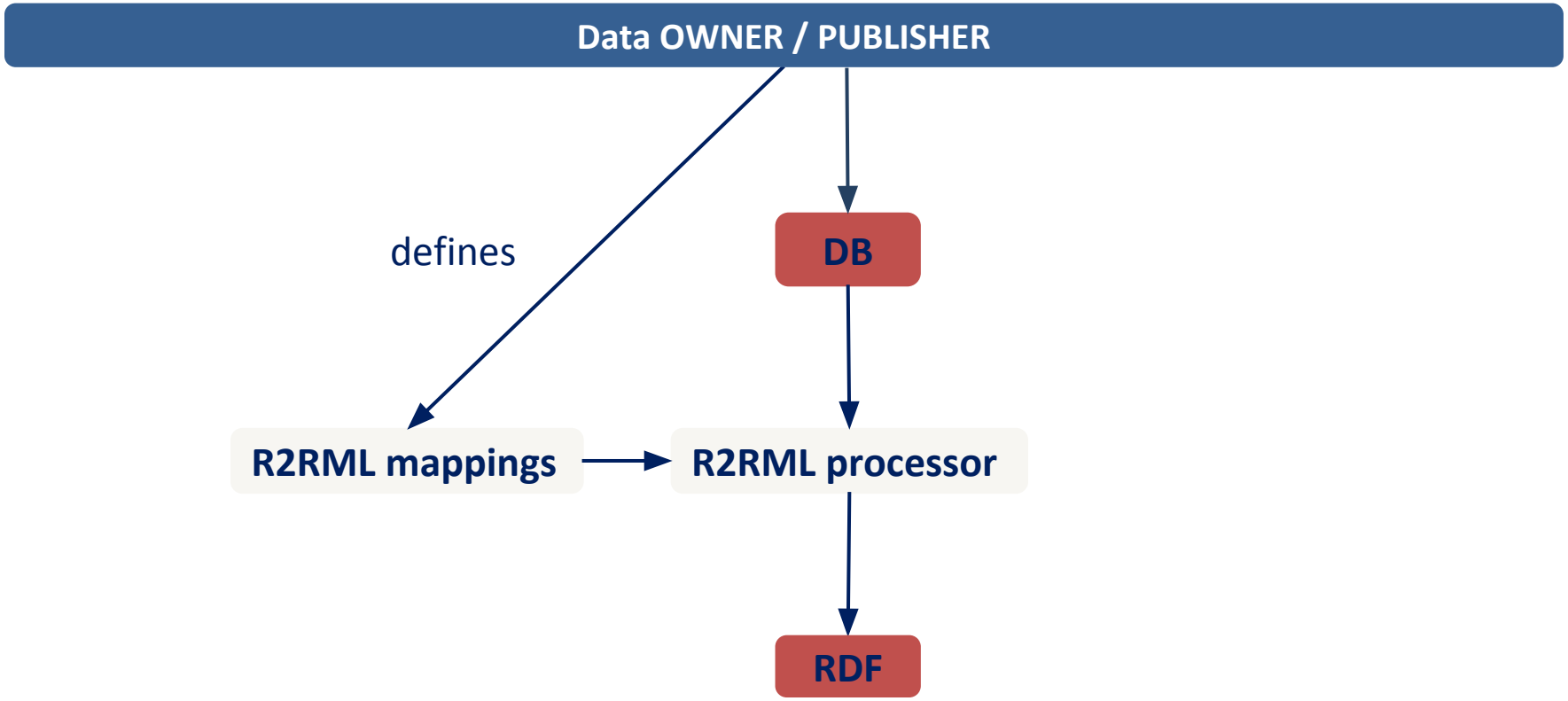
6. Convert data



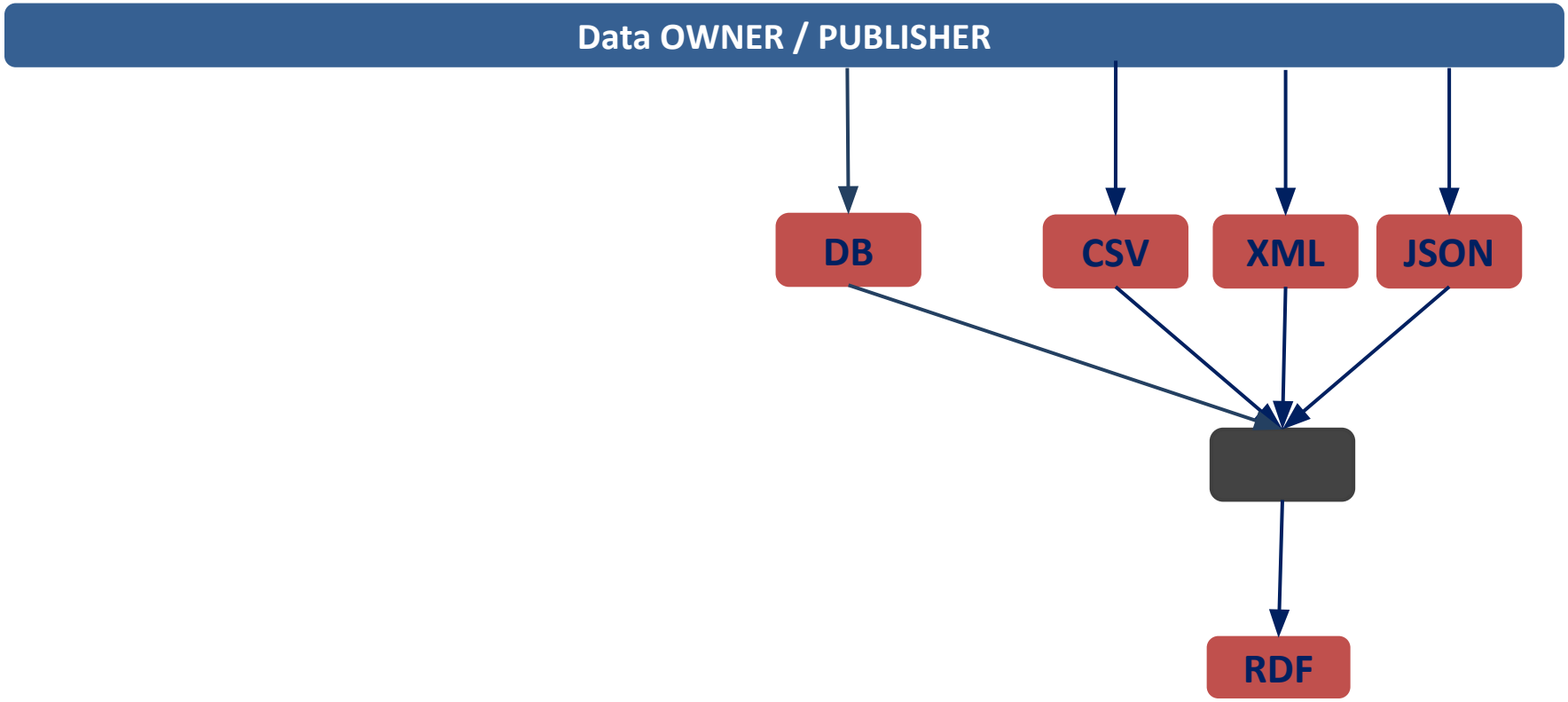
6. Convert data



6. Convert data - W3C recommended R2RML



6. Convert data



Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data

~~Specify an appropriate license~~

4. Good URIs for Linked Data
5. Use standard vocabularies
6. Convert data

...

(Government Linked Data Working Group, 2014)

Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data
4. Good URIs for Linked Data
5. Use standard vocabularies
6. Convert data
7. **Specify an appropriate license**

...

(Government Linked Data Working Group, 2014)

(almost) Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data
4. Good URIs for Linked Data
5. Use standard vocabularies
6. Convert data
7. **Specify metadata**

...

7. Specify metadata

Specify the license

Specify the provenance

PROV-O
2013 - https://www.w3.org/TR/prov-o/

Specify the dataset

VoID
2011 - https://www.w3.org/TR/void/

(almost) Linked Data Best Practices

1. Prepare stakeholders
2. Select datasets
3. Model the data
4. Good URIs for Linked Data
5. Use standard vocabularies
6. Convert data
7. Specify metadata

...

How do we know if we did it right?

Linked Data Quality

Representational dimensions

→ data modeling

conciseness, interpretability, interoperability, versatility

Intrinsic dimensions

→ Linked Data generation

syntactic validity, semantic accuracy,
consistency, conciseness, completeness

~~Accessibility dimensions~~

→ Linked Data publishing

~~Contextual dimensions~~

→ Linked Data consumption

(Zaveri et al., 2015)

Sources

Linked Data principles, https://www.w3.org/DesignIssues/LinkedData.html
Best Practices for Publishing Linked Data, https://www.w3.org/TR/ld-bp/
Cool URIs, https://www.w3.org/TR/cooluris/
Cool URIs don’t change, https://www.w3.org/Provider/Style/URI
Linked Data Glossary, https://www.w3.org/TR/ld-glossary/