



↔ <http://bit.ly/lebo-twed-2013-feb>

# Prizms

*Better Visualizations Catalyzed  
by Better Data*



Timothy Lebo  
Tetherless World Constellation  
Rensselaer Polytechnic Institute



Rensselaer



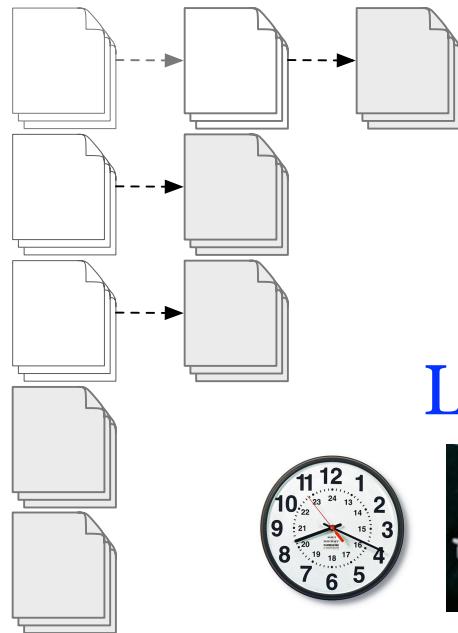
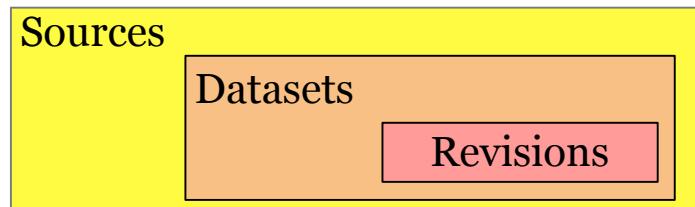
# Outline

- Prizms today
  - csv2rdf4lod-automation
  - DataFAQs
  - LODSPeakR
- Prizms tomorrow
  - Between The Edges / SPO Balance
  - Data Carving
  - Visual Strategies
  - PROVenance + Backlinks

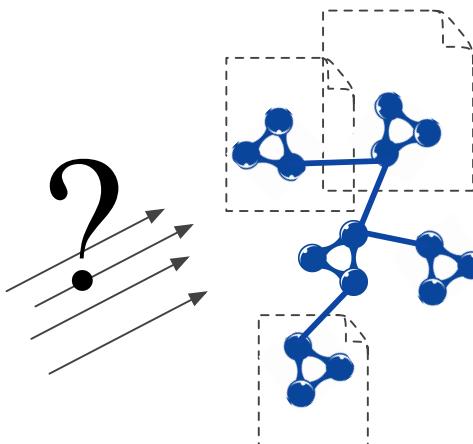


# (1/3) csv2rdf4lod-automation

Lots of:



But want:



Low on:

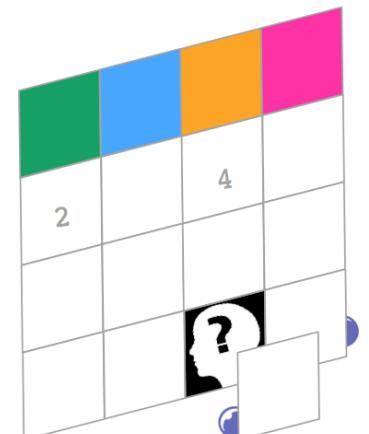


Time

Domain  
Expertise



SemWeb  
Expertise





# (1/3) csv2rdf4lod-automation

- Cataloging, Organizing, Retrieving, Converting, Publishing

The diagram shows a flow from a question mark icon through a process of retrieval, conversion, and publication to a GitHub repository. A yellow starburst labeled "Metadata and Provenance" highlights the central concept.

GitHub repository details:

- Repository: [timrdf/csv2rdf4lod-automation](https://github.com/timrdf/csv2rdf4lod-automation)
- Source tab selected
- Commits: 2
- Network: 1
- Pull Requests: 1
- Switch Branches: 1
- Switch Tags: 0
- Branch List

Terminal commands:

```
$ ./retrieve.sh  
$ vi manual/data.el.params.ttl  
$ ./convert-foreign-aid.sh  
$ publish/bin/publish.sh
```



<https://github.com/timrdf/csv2rdf4lod-automation/wiki>

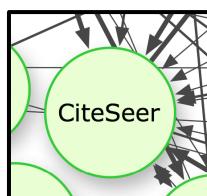
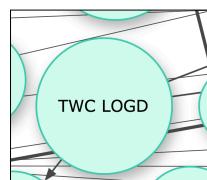


## (2/3) DataFAQs

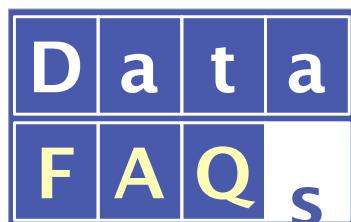
*"I made good Linked Data... so what?"*

evaluation services

datasets



	evals		



<https://github.com/timrdf/DataFAQs/wiki>





## (2/3) DataFAQs example

```
<http://datahub.io/dataset/melagrid>
  a dcat:Dataset ;
  dcat:distribution [
    dct:format [
      a dct:IMT ;
      rdf:value "api/sparql" ;
      rdfs:label "api/sparql"
    ] ;
    dct:title "SPARQL Endpoint" ;
    a dcat:Distribution ;
    dcat:accessURL <http://lod.melagrid.org/>
  ];
.
```



```
curl -H "Content-Type: text/turtle" \
-d @my.ttl http://localhost:9236/prov
```

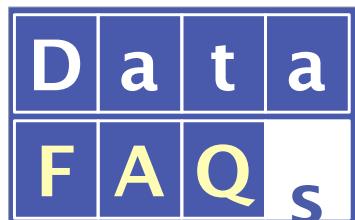
```
<http://datahub.io/dataset/melagrid>
  a datafaqs:Satisfactory;
  void:vocabulary prov: .

prov:actedOnBehalfOf a rdf:Property;
  sio:count 0 .

prov:activity a rdf:Property;
  sio:count 0 .

prov:agent a rdf:Property;
  sio:count 10 .

prov:alternateOf a rdf:Property;
  sio:count 10 .
```



SADI

Find. Integrate.  
Analyze.

<https://github.com/timrdf/DataFAQs/wiki>

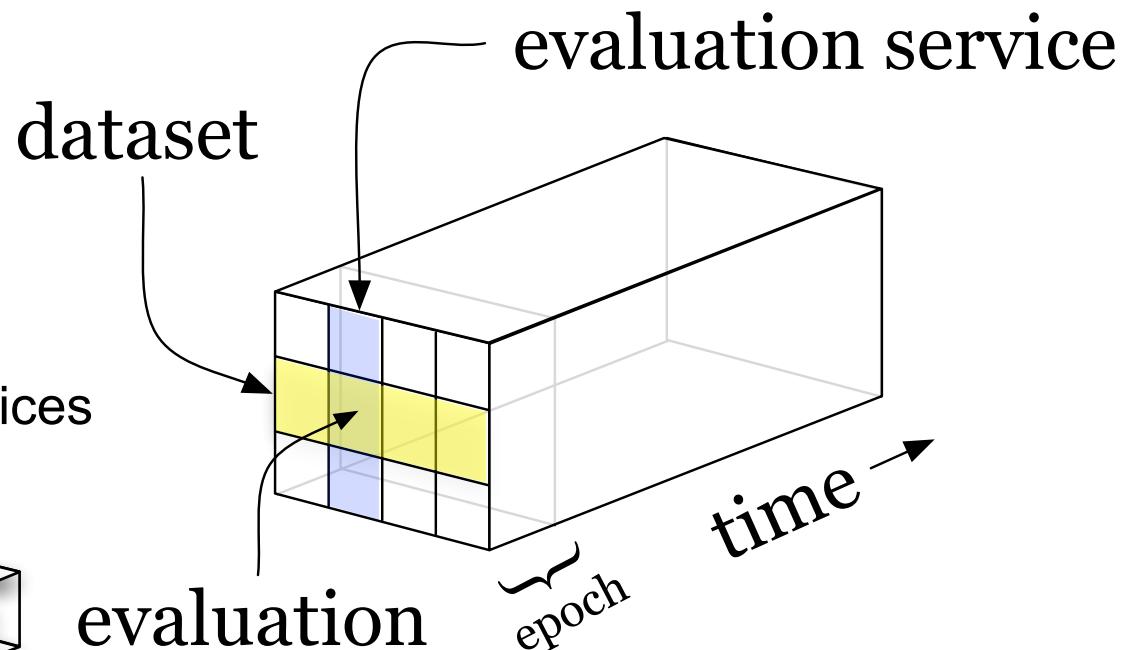
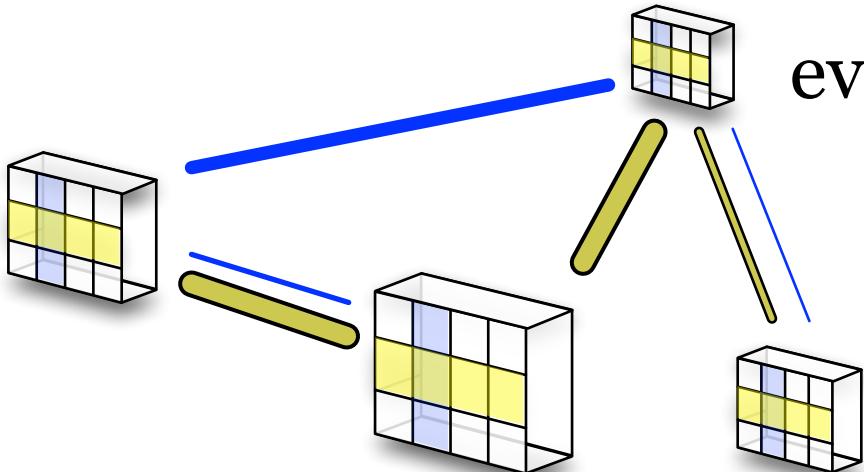




# DataFAQs

LINKED DATA QUALITY REPORTS

1. Choose Datasets
2. Choose Evaluation Services
3. Explore Evaluations

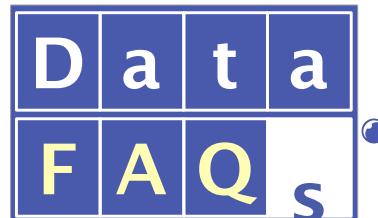
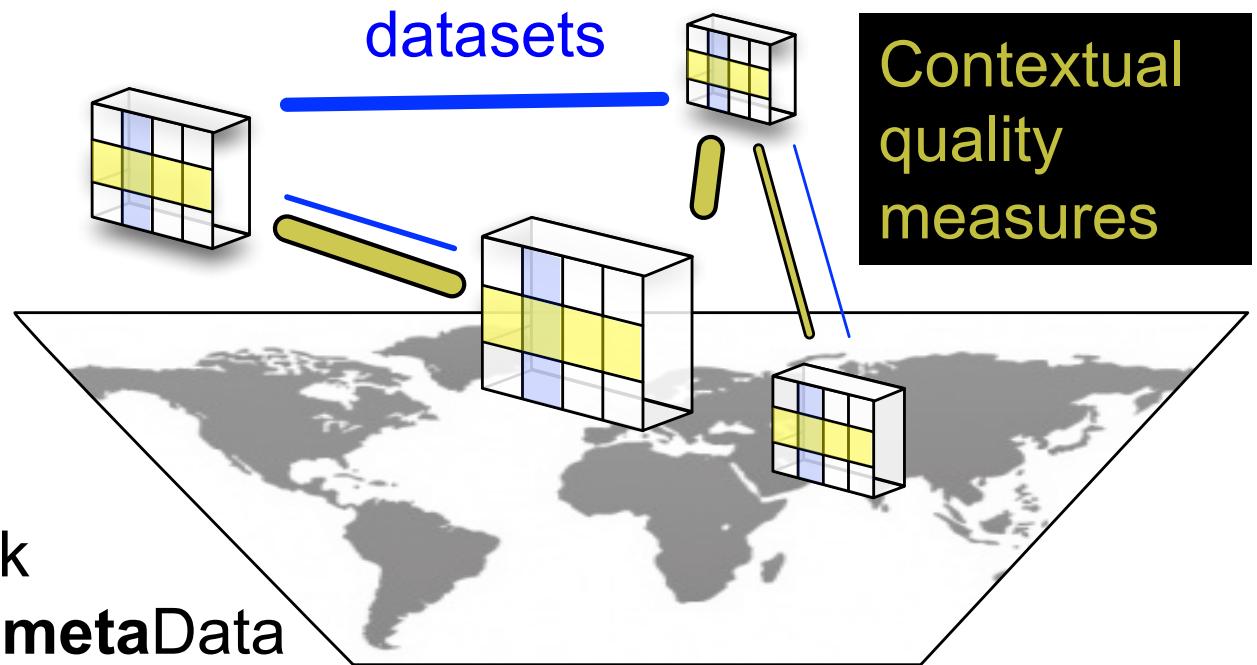


- Individual efforts benefit community
- Automated, quantitative feedback
- Recommendations from the aggregate
- Network effect



## (2/3) DataFAQs

DataFAQs nodes +  
DataFAQs network  
=> Linked Open **metaData**

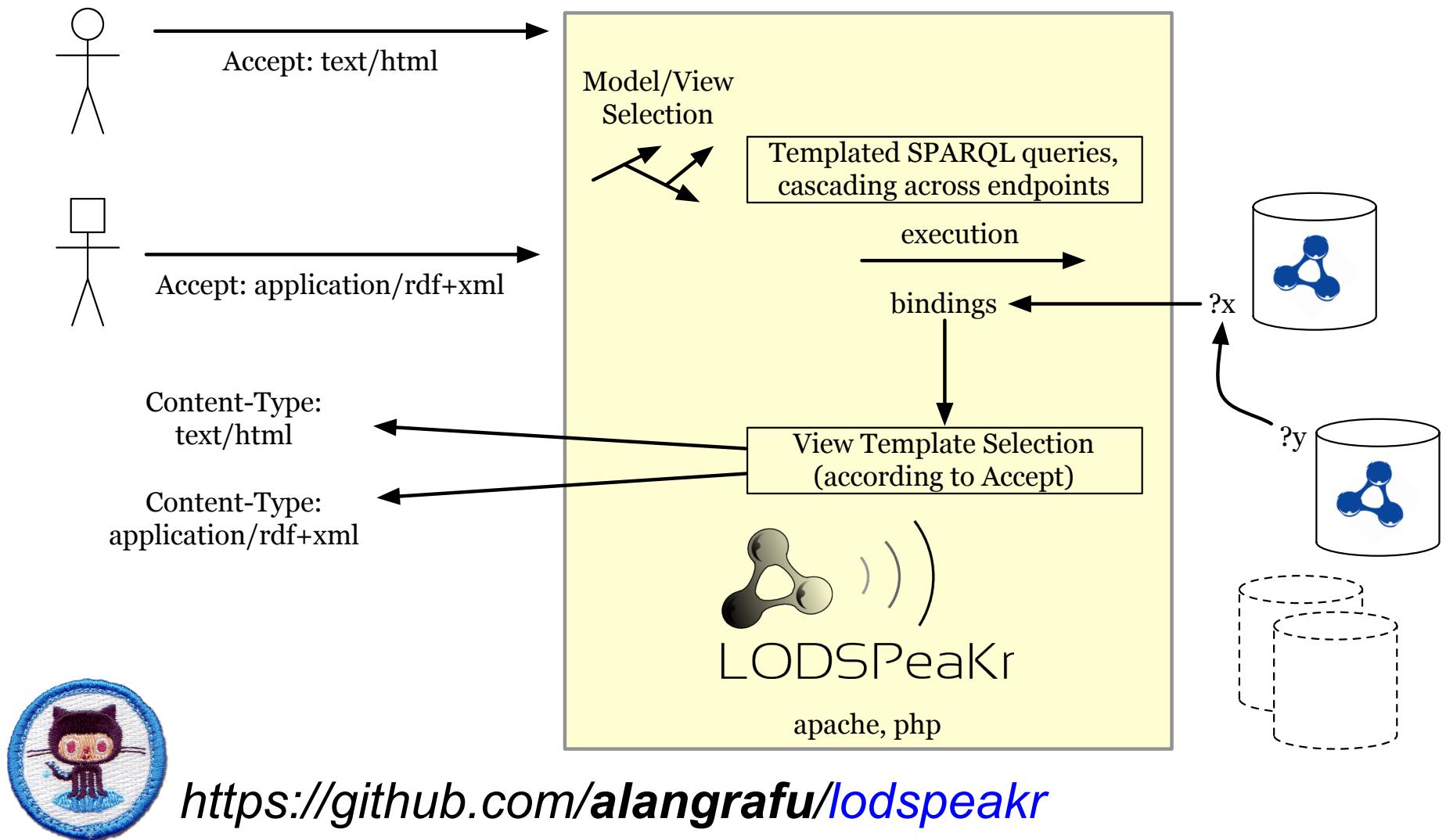


<https://github.com/timrdf/DataFAQs/wiki>

**github**  
SOCIAL CODING

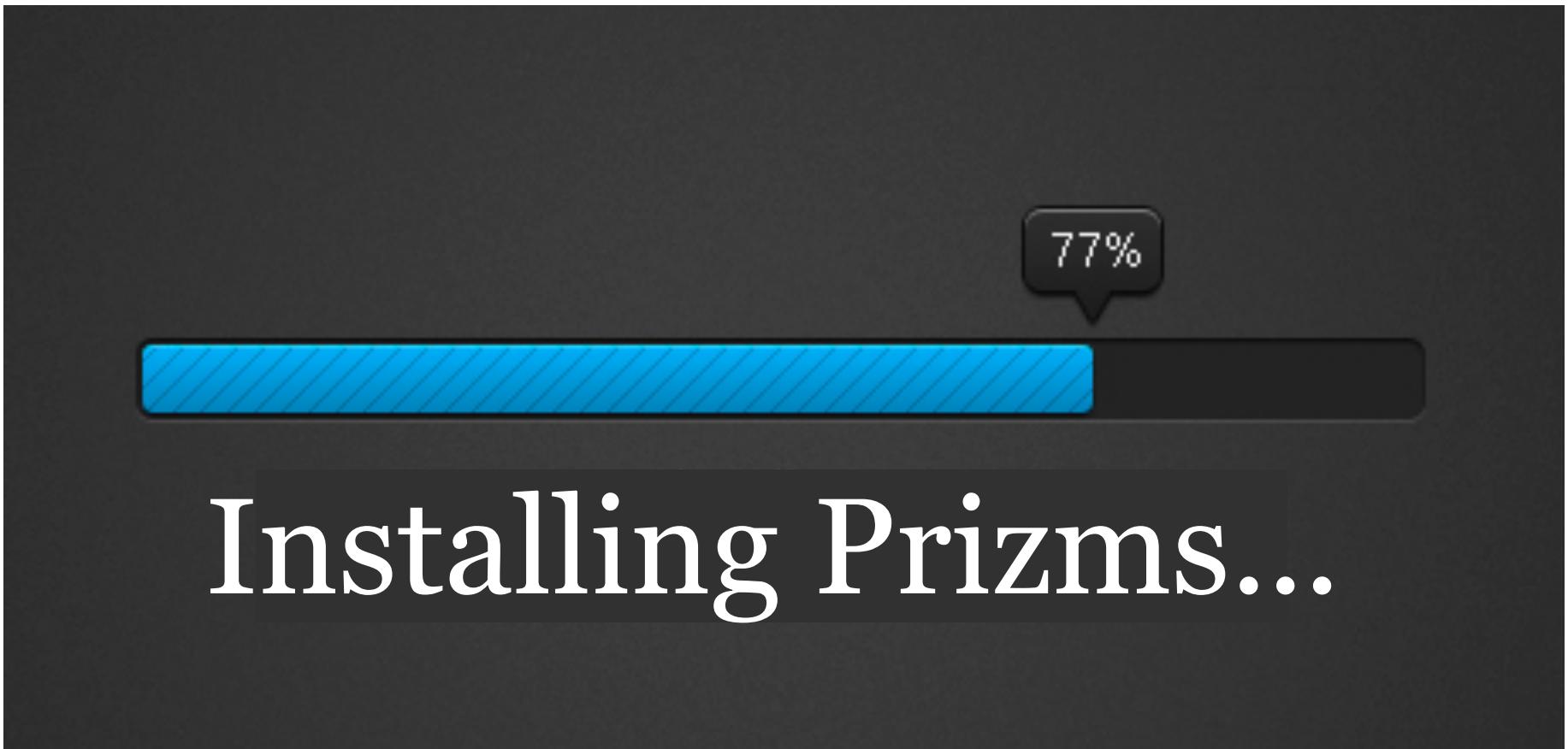


## (3/3) LODSPeaKr





# Prizms Today



csv2rdf4lod-automation, DataFAQs, [Virtuoso], LODSPeakR  
+ their inter-wiring...



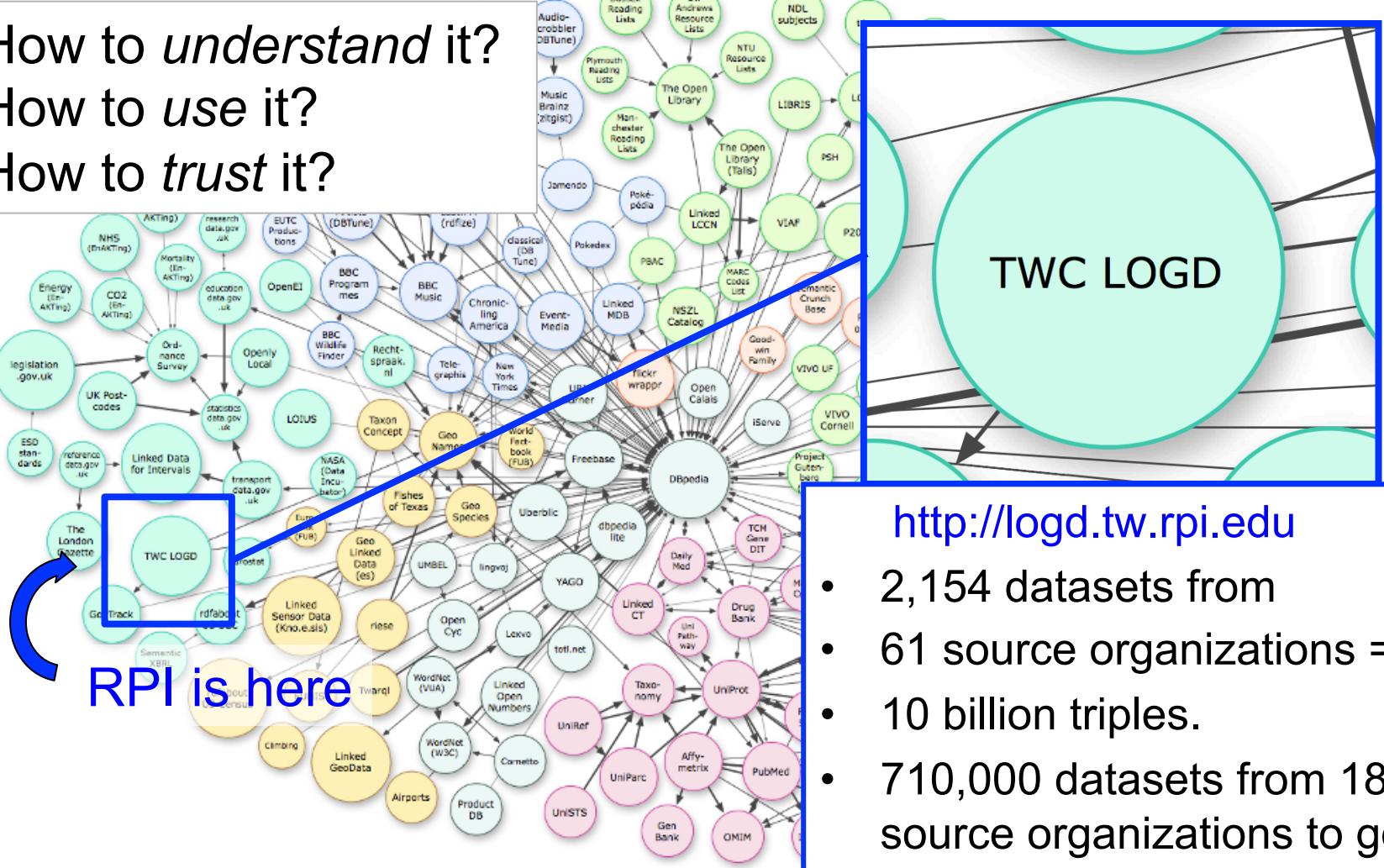
# Outline

- Prizms today
  - csv2rdf4lod-automation
  - DataFAQs
  - LODSPeakR
- Prizms tomorrow
  - Between The Edges / SPO Balance
  - Data Carving
  - Visual Strategies
  - PROVenance + back-links



# LOGD: Linking Open Government Data

- How to *understand* it?
- How to *use* it?
- How to *trust* it?

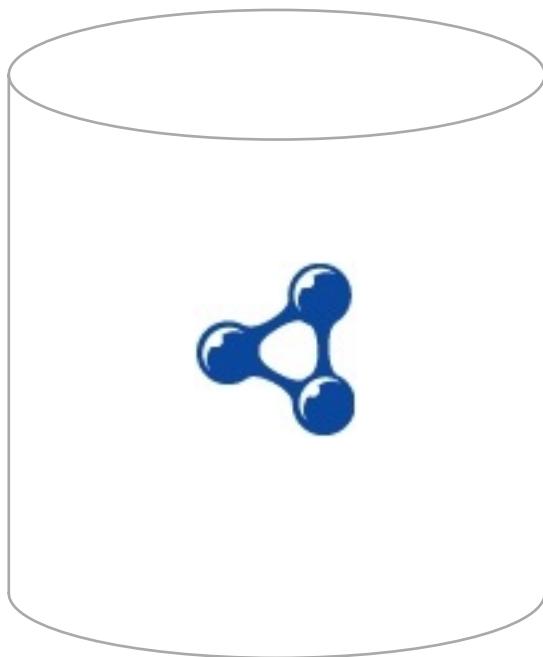




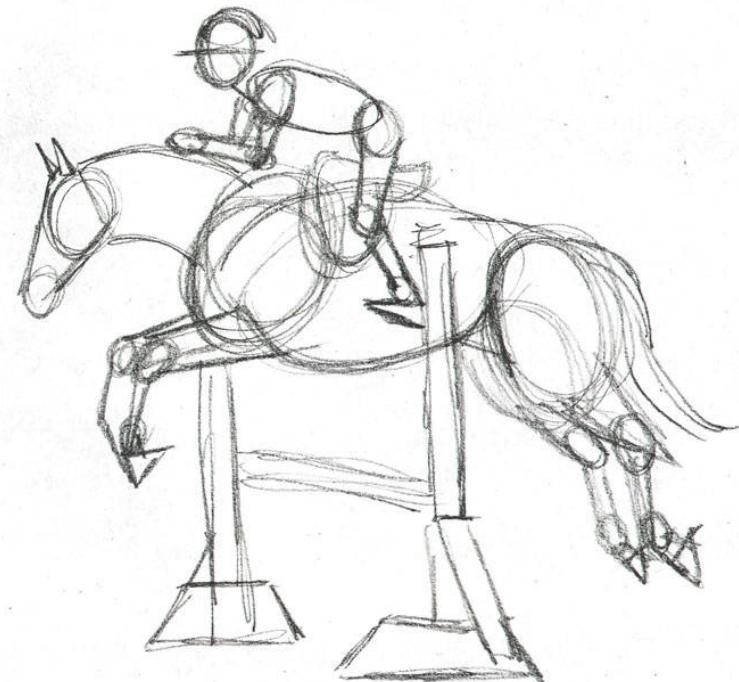
# Disclaimer

What you are about to hear...

...is all very preliminary work.



*“Now what?”*





# @benben's Type by Task Taxonomy

Assumes that users are viewing collections of items, where items have multiple attributes.

- **Overview**: Gain an overview of the entire collection.
- **Zoom**: Zoom in on items of interest.
- **Filter**: Filter out uninteresting items.
- **Details-on-demand**: Get details of an item or group (when needed).
- **Relate**: View relationships among items.
- **History**: Support undo, replay, and progressive refinement.
- **Extract**: 1) sub-collections and 2) query parameters.

VL '96 Proceedings of the 1996 IEEE Symposium on Visual Languages

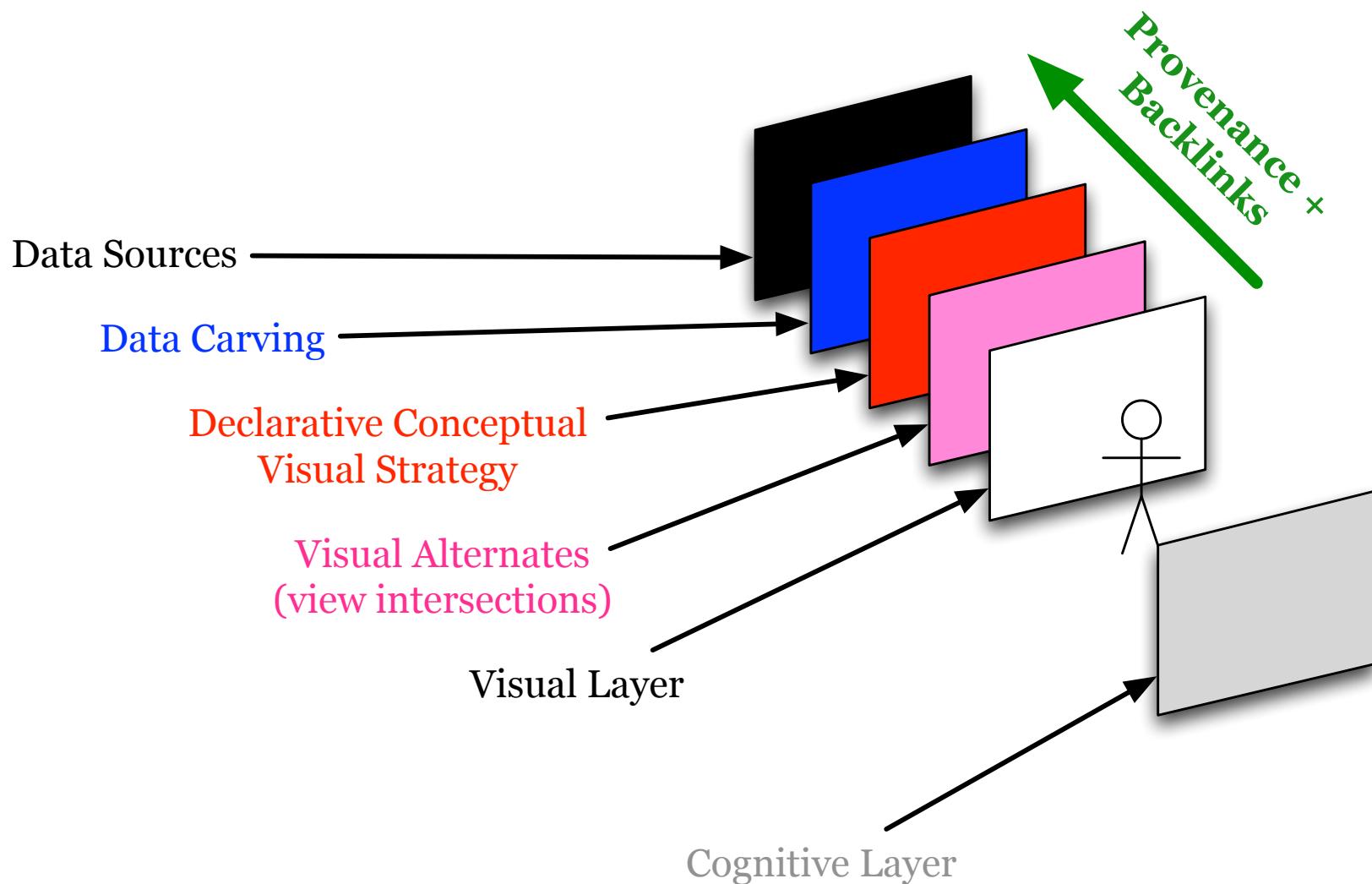


# Fulfilling the Type by Task Taxonomy

- **Overview:** Gain an overview of the entire collection.
  - **Between the Edges, SPO Balance, Data Carving**
- **Zoom:** Zoom in on items of interest.
  - **Visual Strategy**
- **Filter:** Filter out uninteresting items.
  - **Data Carving, Visual Strategy**
- **Details-on-demand:** Get details of an item or group (when needed).
  - **Data Carving, Visual Strategy, Provenance + Backlinks**
- **Relate:** View relationships among items.
  - **Data Carving, Visual Strategy**
- **History:** Support undo, replay, and progressive refinement.
  - **Provenance + Backlinks**
- **Extract:** 1) sub-collections and 2) query parameters.
  - **Provenance + Backlinks**

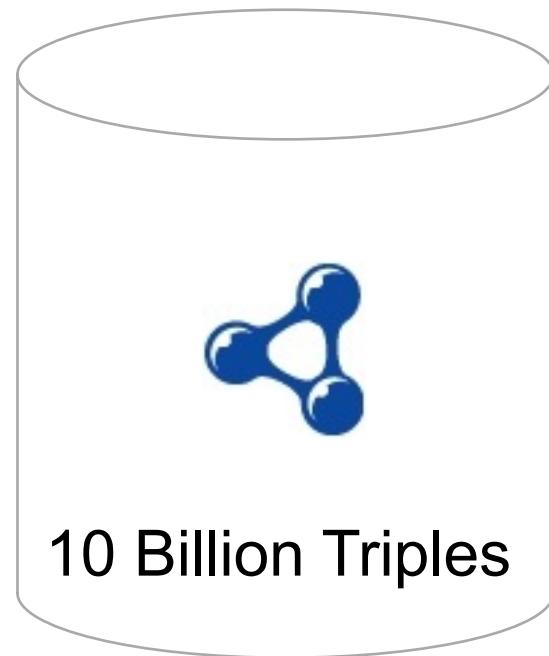


# How these fit together





TWC



*“Now what?”*



# Make More Triples! (with Between The Edges, BTE)



Original Graph

212 triples

```
...  
<http://www.ivan-herman.net/foaf.rdf#me>  
<http://www.johnseelybrown.com/>  
<http://www.kjetil.kjernsmo.net/foaf#me>  
<http://www.koalie.net/foaf.rdf>  
<http://www.lassila.org/ora.rdf#me>  
...
```

→ 116 unique URI nodes



SADI

Find. Integrate.  
Analyze.



New Graph:  
Between The Edges

```
<http://www.koalie.net>
```

```
<http://www.koalie.net/foaf.rdf>
```

1,511 triples with  
323 unique URI nodes

- Vocabulary available at <http://prefix.cc/bte>
- Implementation in DataFAQs [fact/naming/between-the-edges.py](http://fact.naming/between-the-edges.py)

212 → 1,511 triples

116 → 323 unique URI nodes

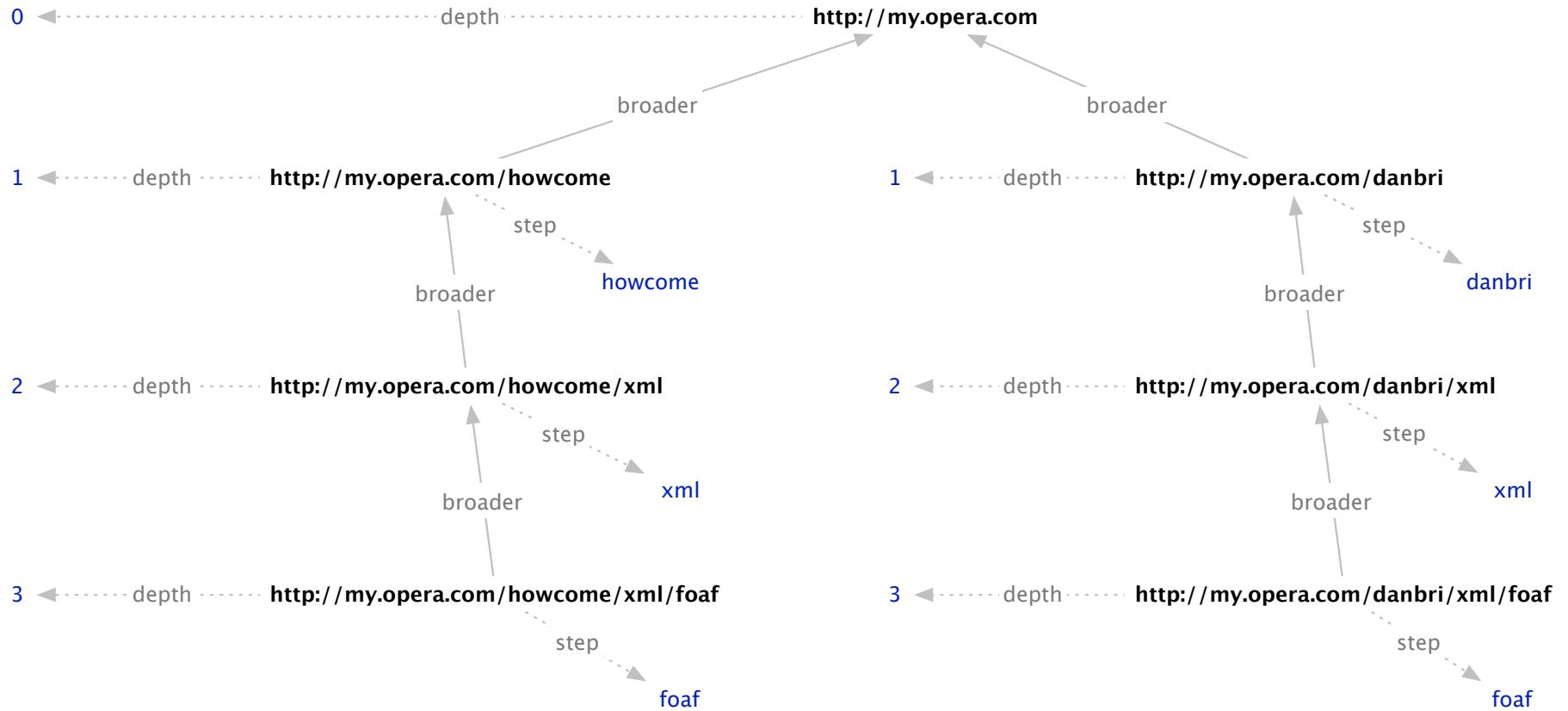


# Between The Edges (BTE)



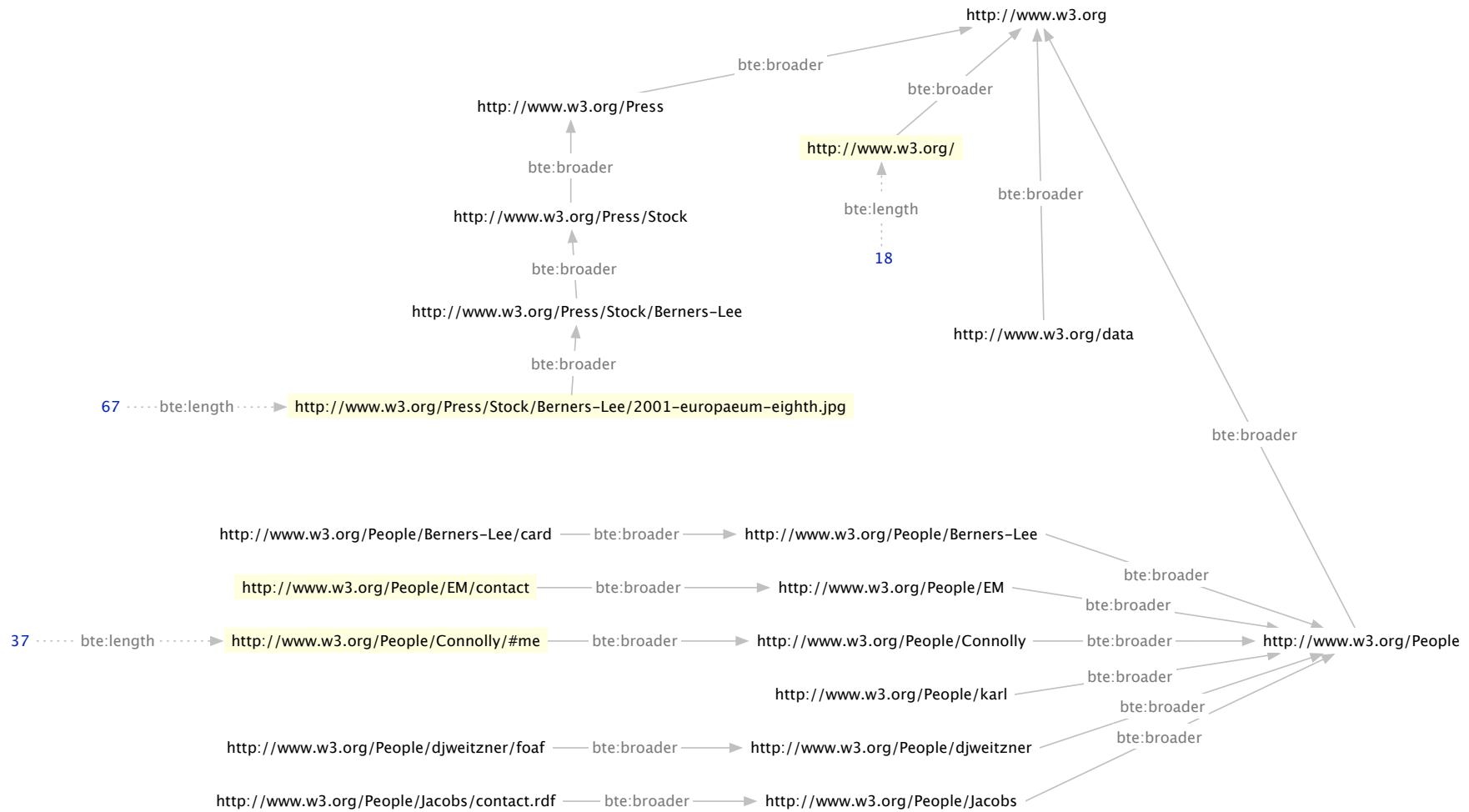


# Between The Edges (BTE)



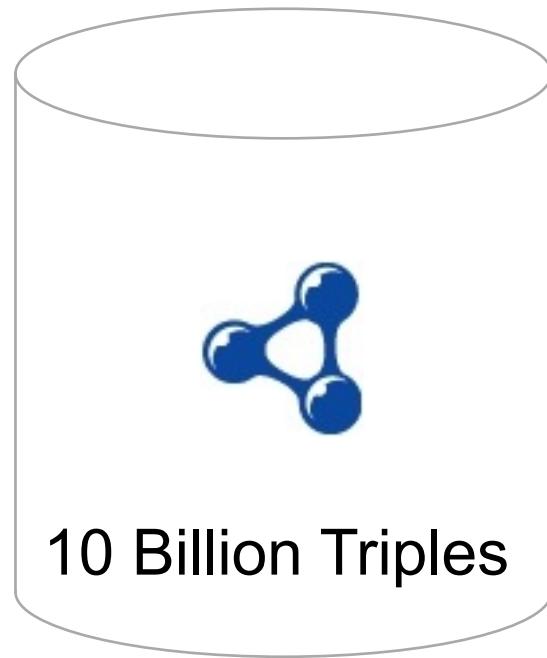


# Between the Edges: Discovering new Neighborhoods





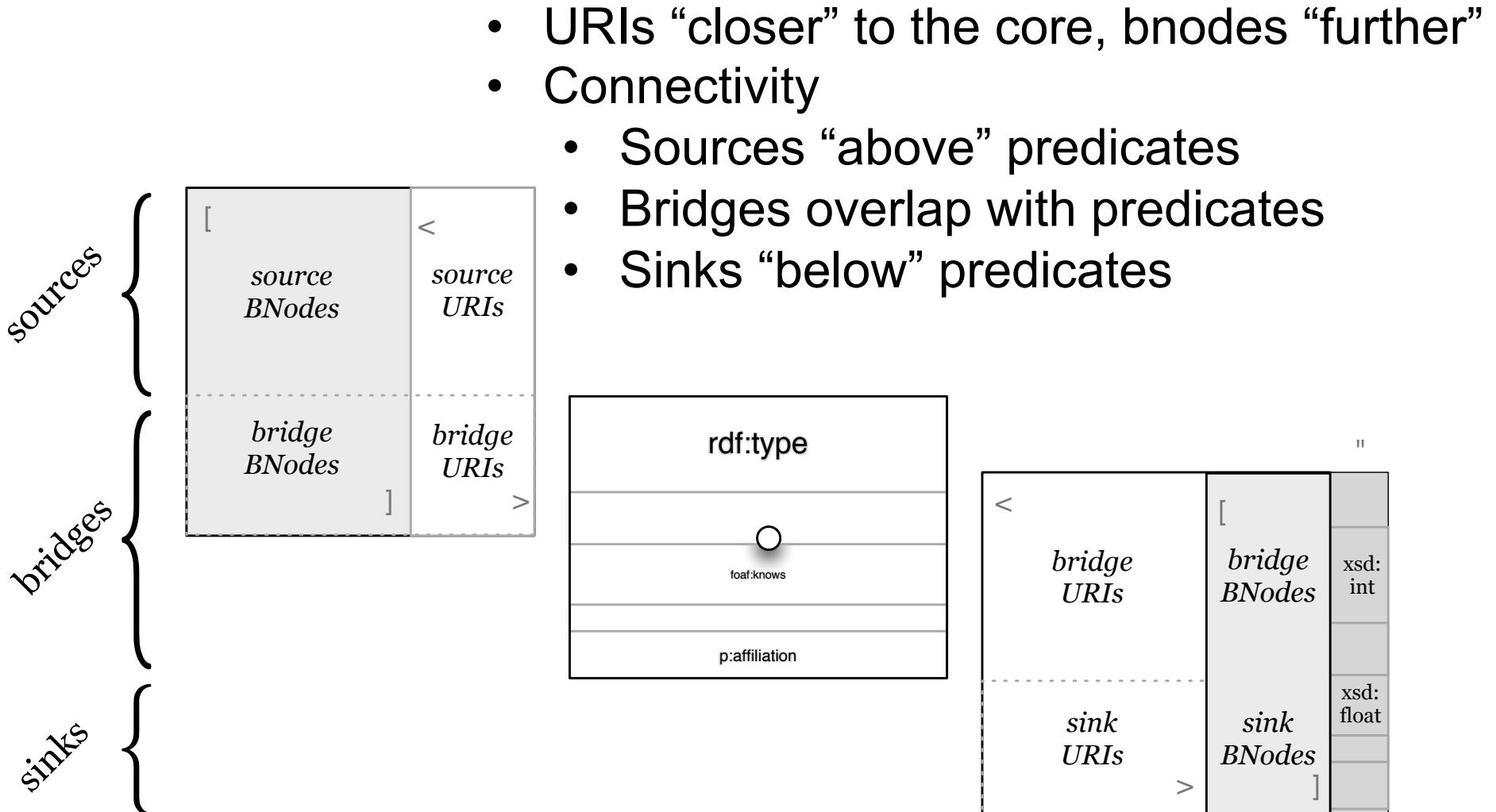
TWC



*“Now what?”*

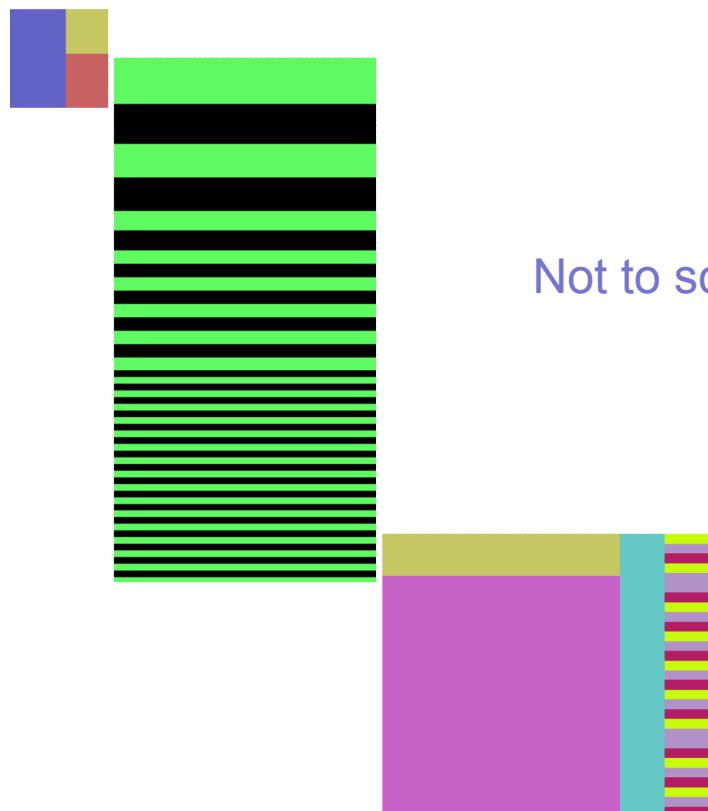


# SPO Balance: Concept



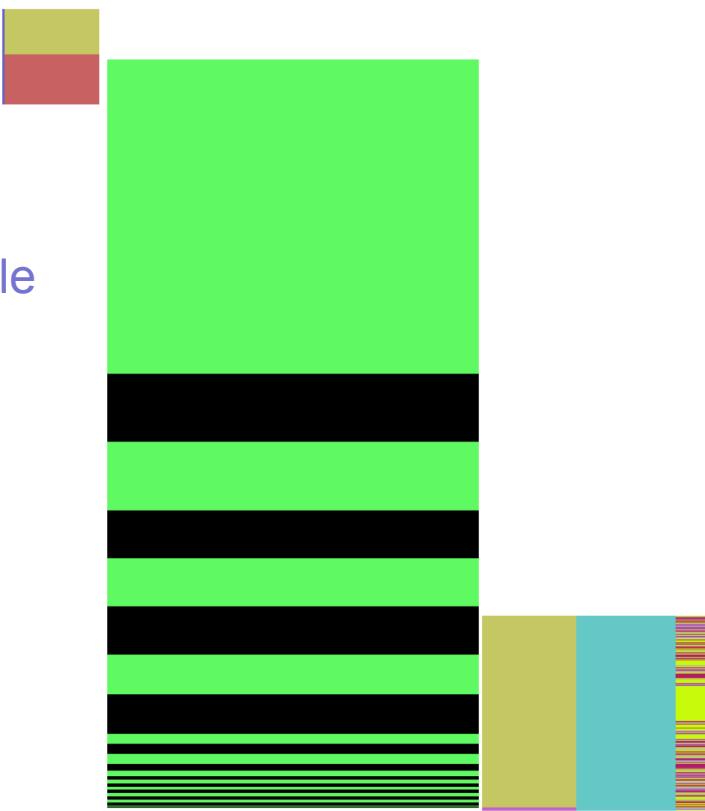


# SPO Balance: Examples



# TBL's FOAF file (212 triples)

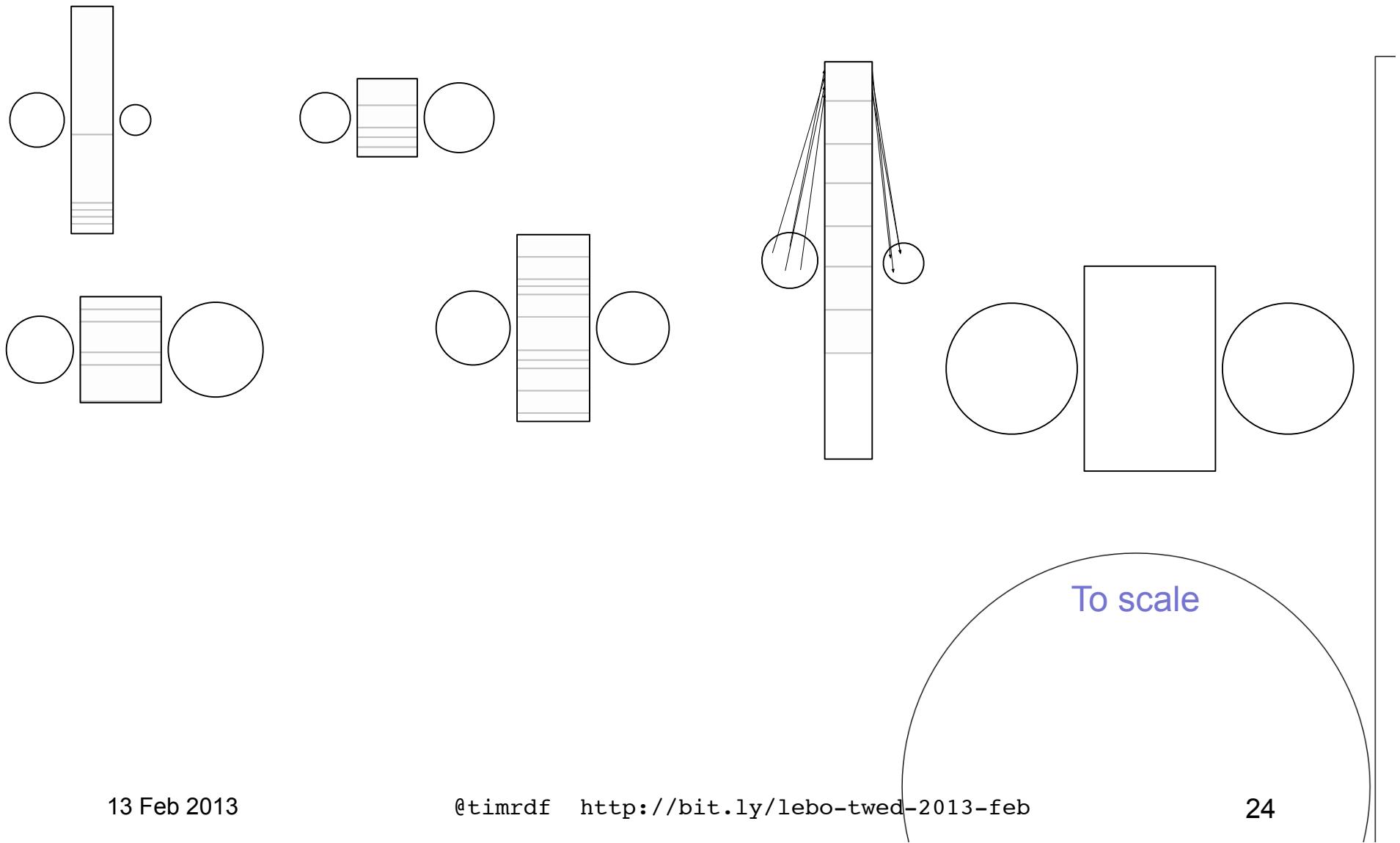
# Not to scale



## Other (89k triples)



# SPO Balance: “Graphscape”





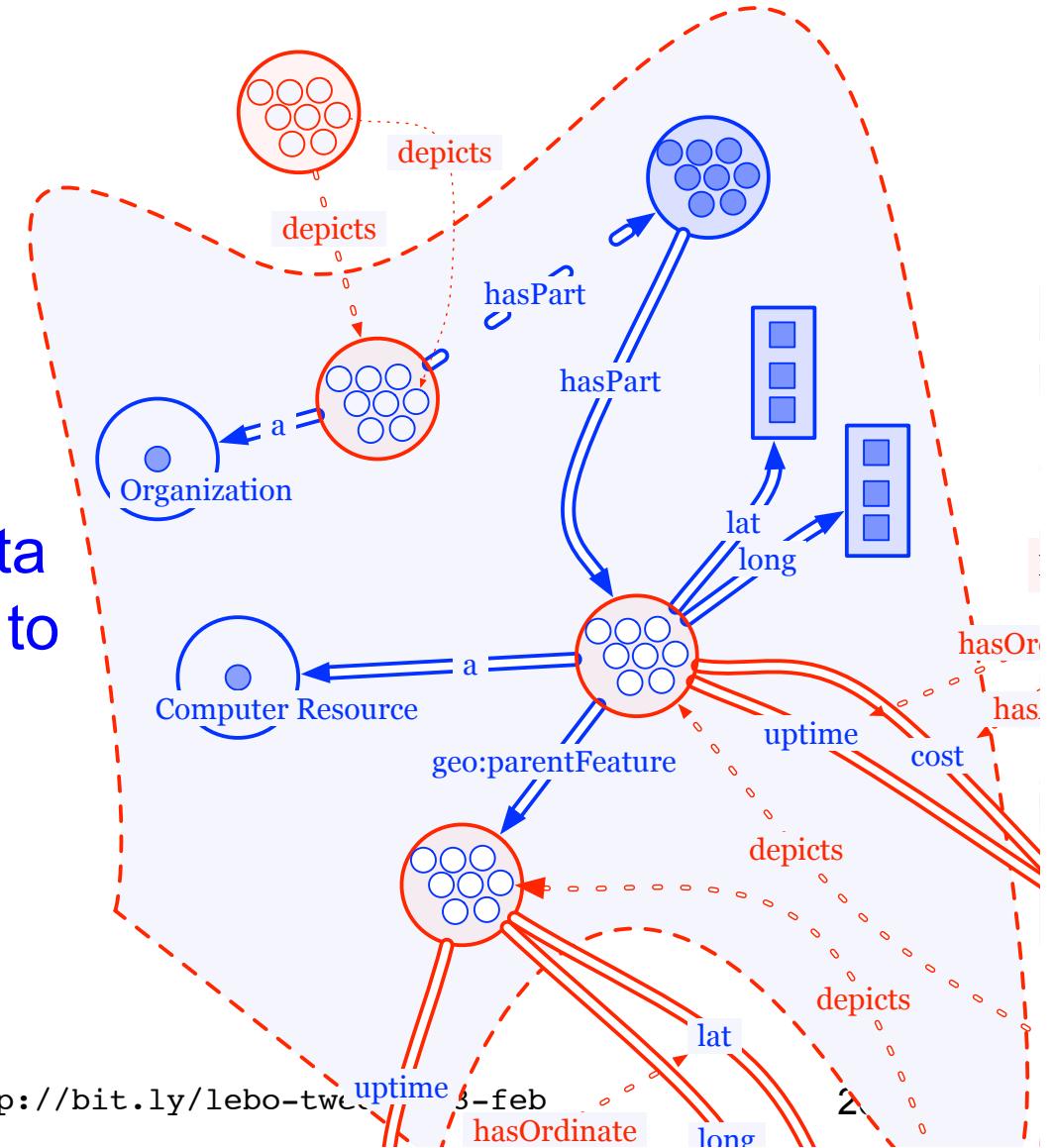
# Visual Strategy

- *“Everything points to the conclusion that the phrase 'the language of art' is more than a loose metaphor, that even to describe the visible world in images we need a developed system of schemata.”*  
-- E. H. Gombrich Art and Illusion, 1959 (p. 76)  
(via @benben)



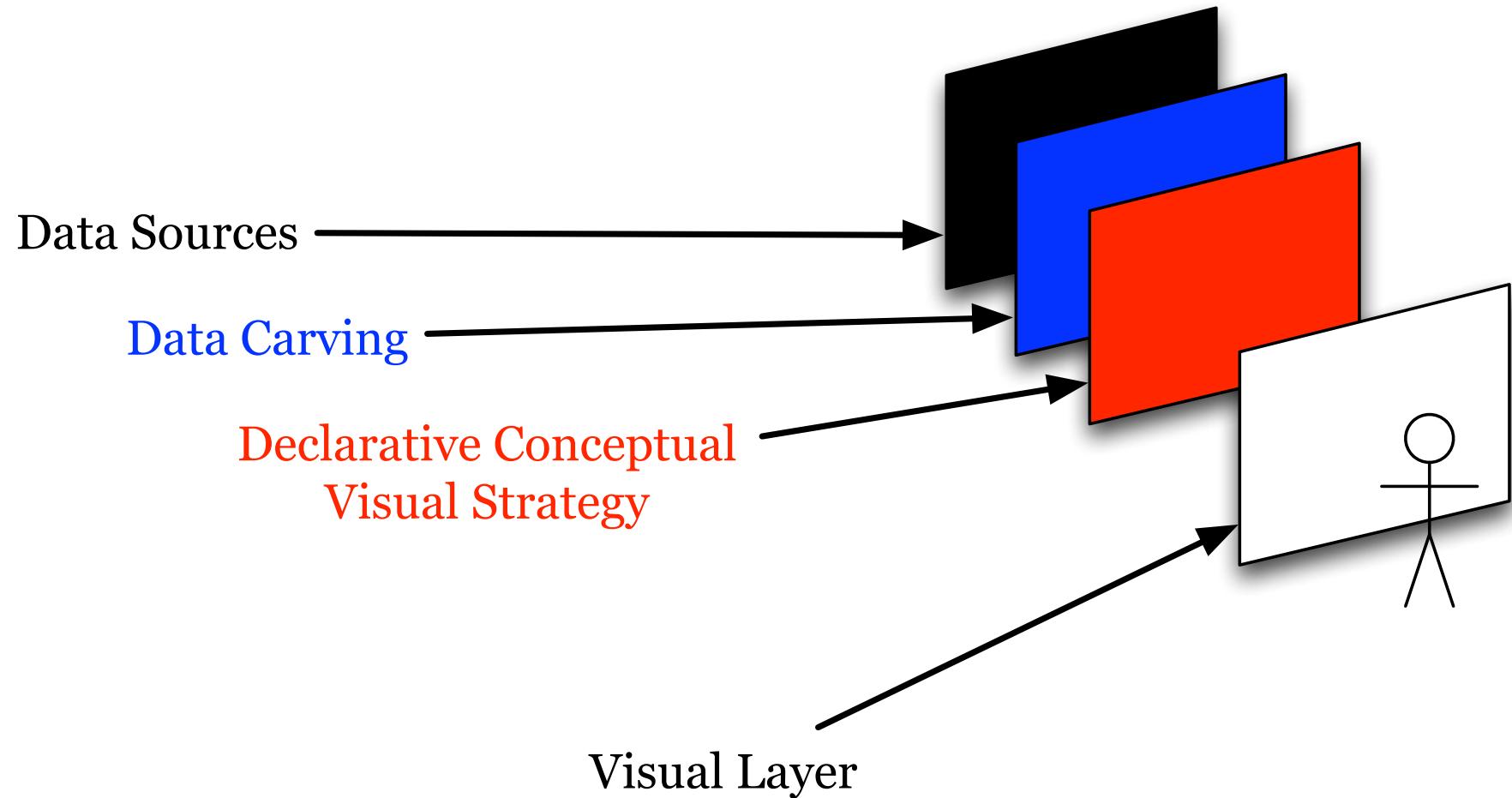
# Data Carving

- Bottom-up data selection
- Working with **all** of the data
- Transition from **overview** to **detail** by self-establishing context
- Establishing the content domain for a visualization

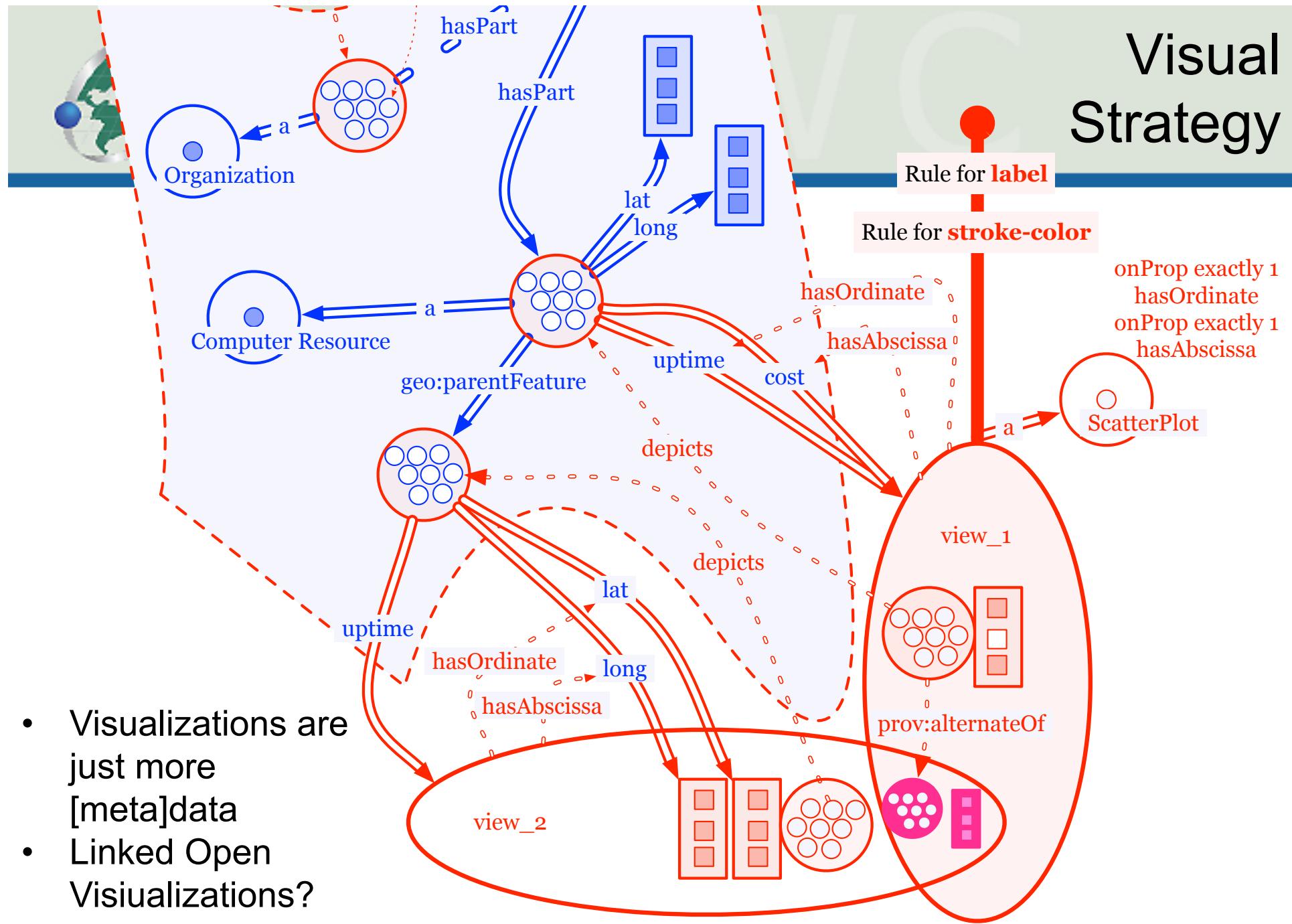




# How these fit together



# Visual Strategy

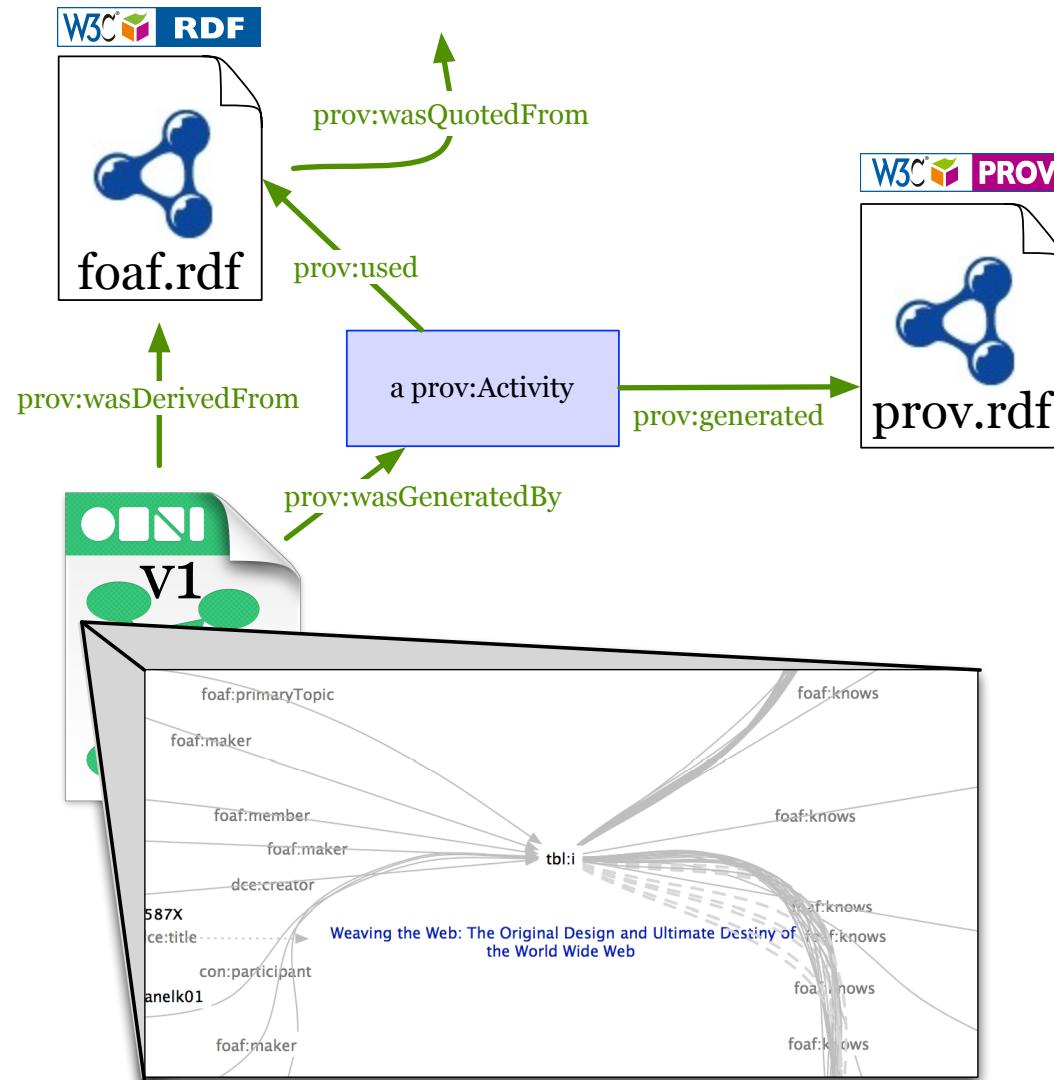


- Visualizations are just more [meta]data
- Linked Open Visualizations?



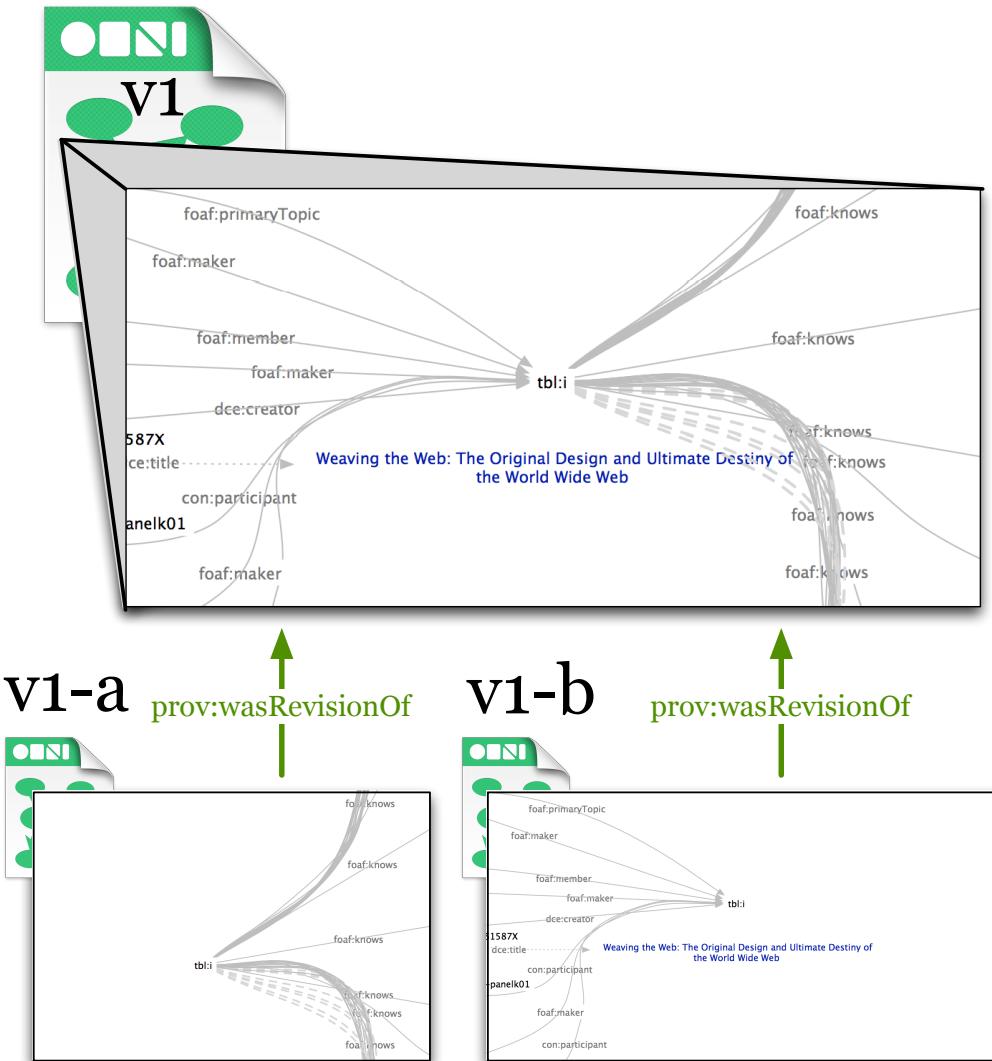
# Provenance + Linkbacks

<http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>



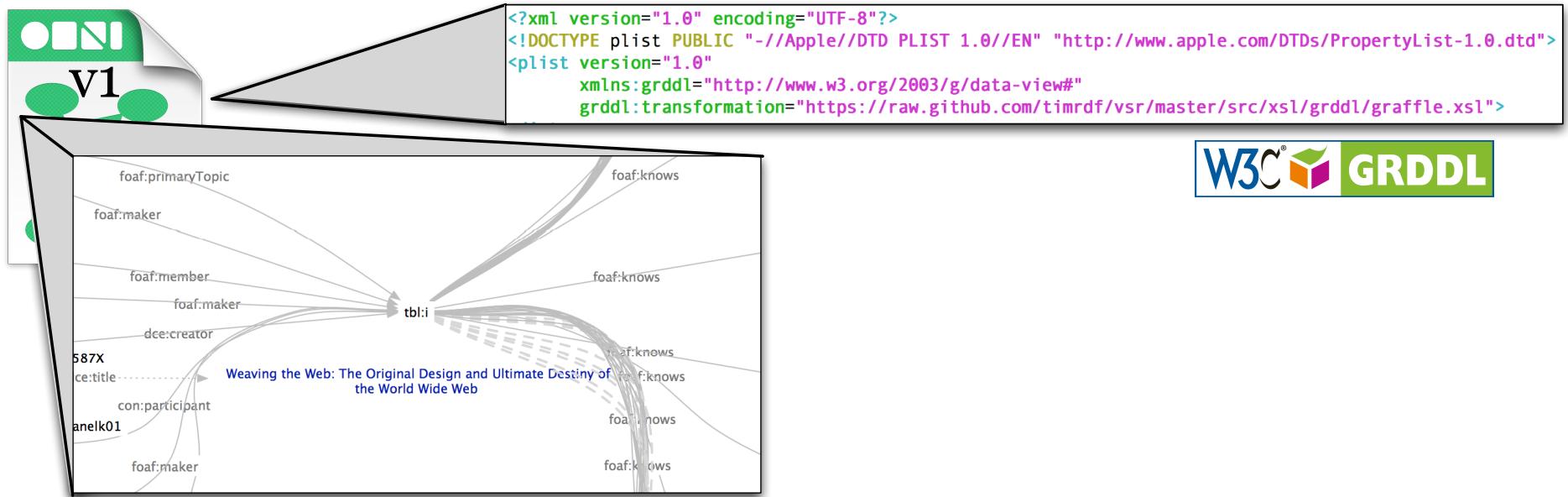


# Two Out-of-band Edits





# Sneaking in a GRDDL Link





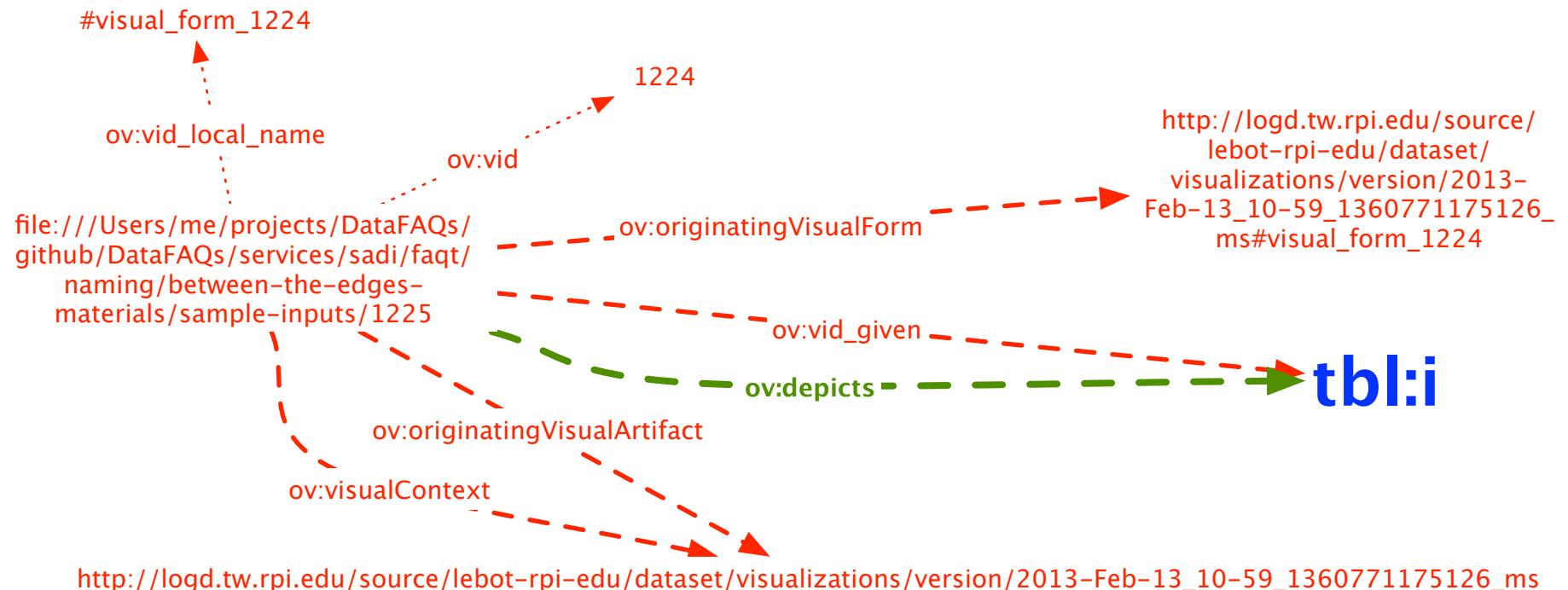
# Combining Data Carving, Visual Strategy, and Provenance+Linkbacks



```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN" "http://www.apple.com/DTDs/PropertyList-1.0.dtd">
<plist version="1.0"
      xmlns:grddl="http://www.w3.org/2003/g/data-view#"
      grddl:transformation="https://raw.github.com/timrdf/vsr/master/src/xsl/grddl/graffle.xsl">
```



*“Give me all visualizations that include TBL”*

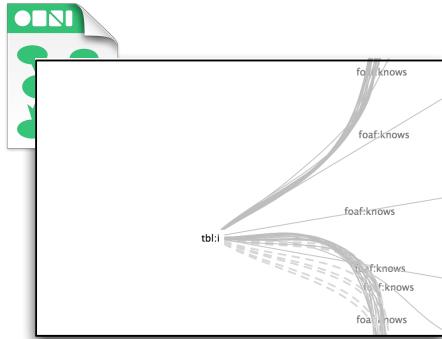


[http://logd.tw.rpi.edu/source/lebot-rpi-edu/dataset/visualizations/version/2013-Feb-13\\_10-59\\_1360771175126\\_ms](http://logd.tw.rpi.edu/source/lebot-rpi-edu/dataset/visualizations/version/2013-Feb-13_10-59_1360771175126_ms)

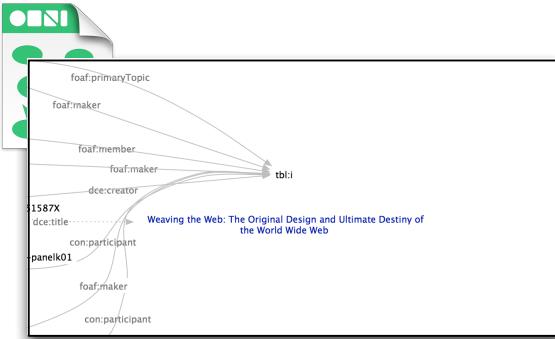


# Visualizations are just Derived Datasubsets

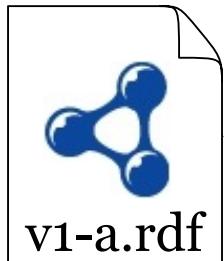
v1-a



v1-b

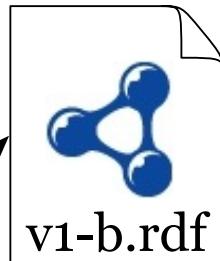


prov:wasDerivedFrom



W3C® GRDDL

prov:wasDerivedFrom



(both reference foaf.rdf, prov.rdf,  
v1.graffle -- and their elements)

*“Which parts of these datasets do people care about?”*

*“Which other unseen parts **should** we be looking at?”*



# Avoiding Archaeological Endeavors

*What are the **things** that you're talking about?  
How do those things **relate**?*

RUBES by Leigh Rubin



At last, the mystery of the Mayan calendar revealed.



A few years later, or the **same moment** somewhere else on the web.



# Conclusions

- Prizms is “half done”
- The semantic web is in dire need of solutions for  
**@benben’s Type by Task Taxonomy**
- I’ve offered several techniques to fulfill these tasks
- **Semantic Recursion:** Using the semantic web to familiarize, understand, and visualize itself
- I hope that these will provide a concrete basis to develop a **Theory of Visualization**
- **Between the Edges, SPO Balance, Data Carving, Visual Strategy, Provenance + Backlinks**



# Thanks!



- Alvaro Graves
- Jim McCusker
- Dominic DiFranzo



Deborah McGuinness

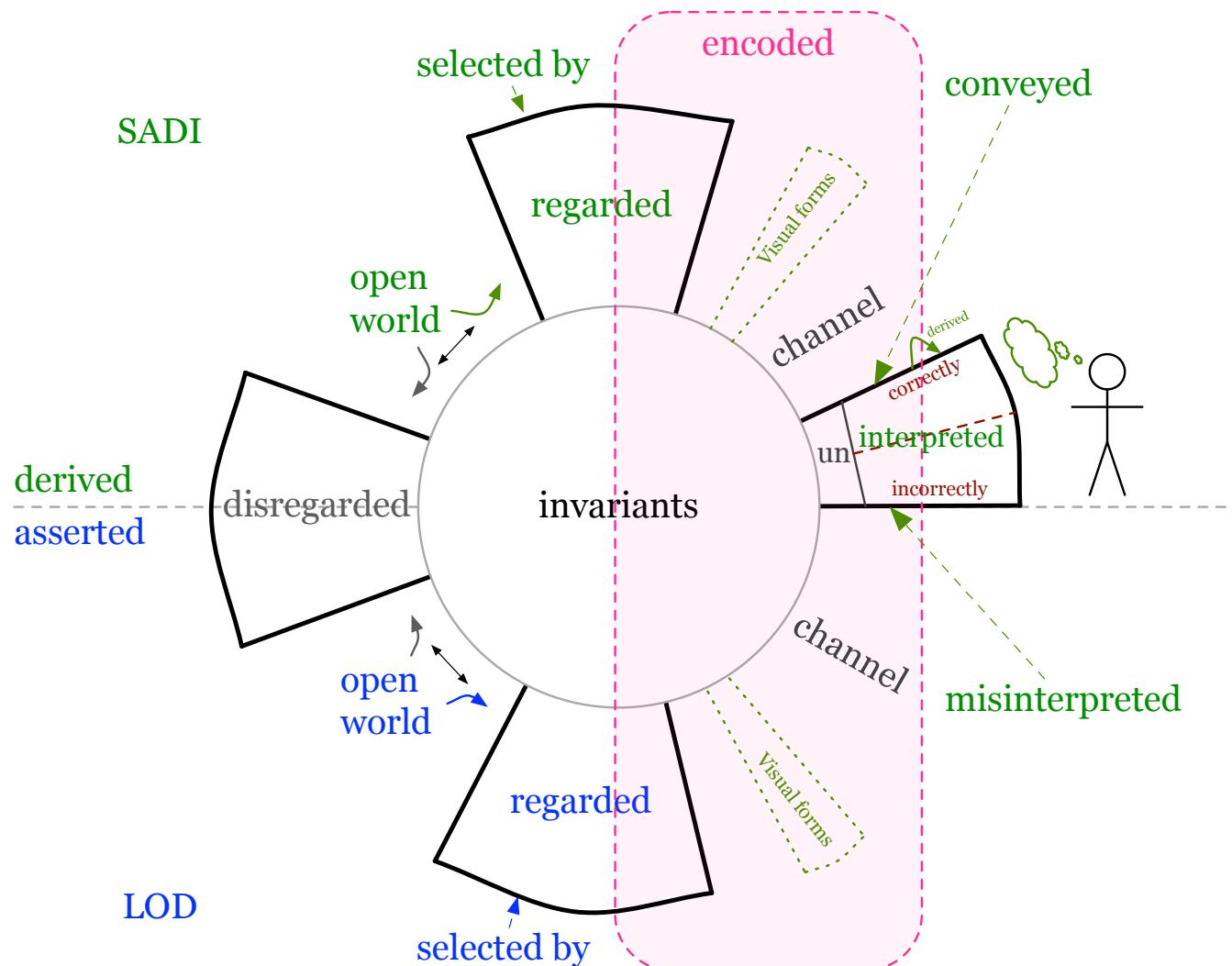


TWC

- backups



# Perspectives for Visual Strategies





# Benefit of Contextualized Data: A Race to Visuals

