⇔ http://bit.ly/lebo-halloween-issues

Towards Data-Centric Digital Communities

An Example for the Department of Health and Human Services

Timothy Lebo
Tetherless World Constellation
Rensselaer Polytechnic Institute

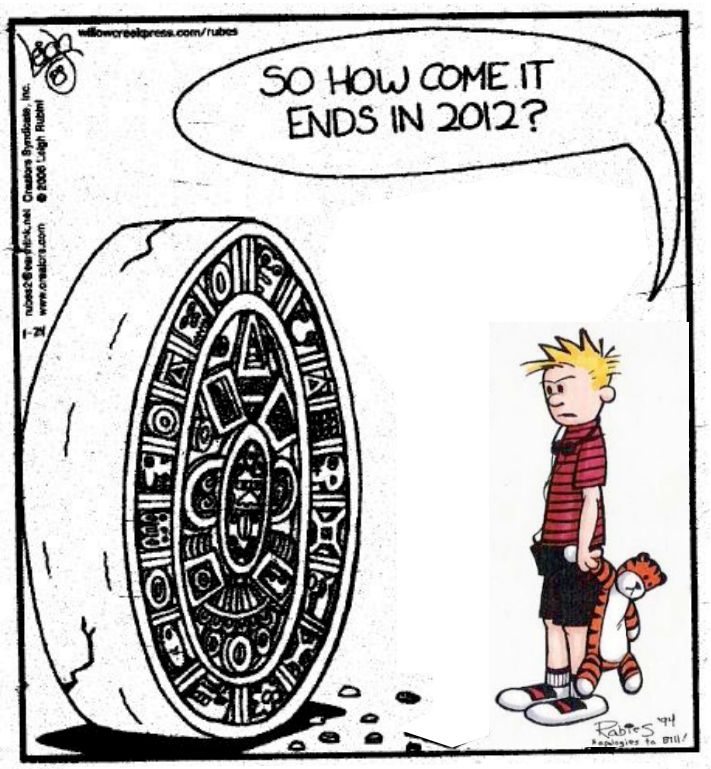
Rensselaer

Motivation: Avoiding Archaeological Endeavors

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

RUBES by Leigh Ruben

At last, the mystery of the Mayan calendar revealed.



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

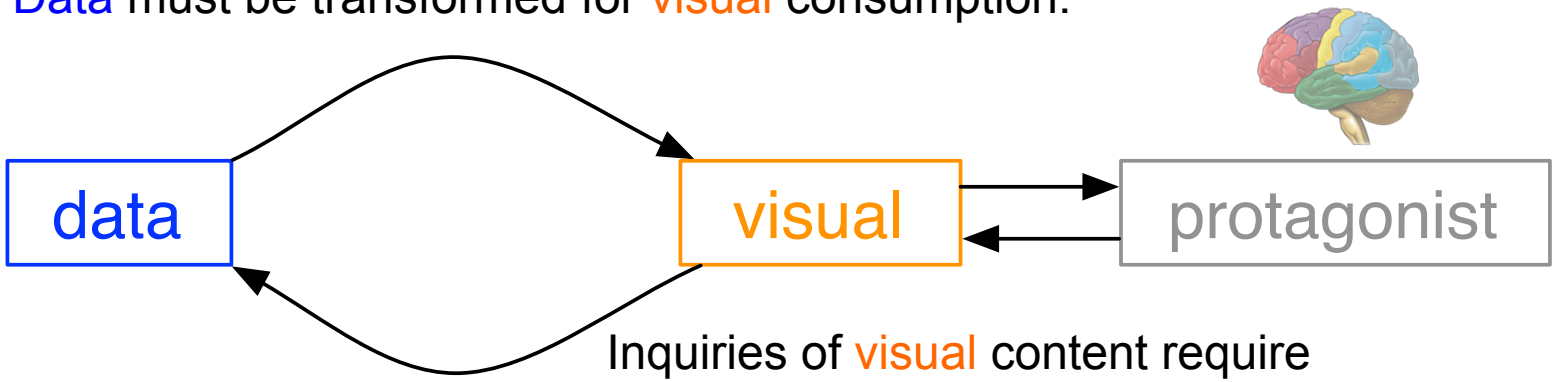
Abstract Objective:

Protagonist is more informed about the world.

Protagonist cannot consume data directly.

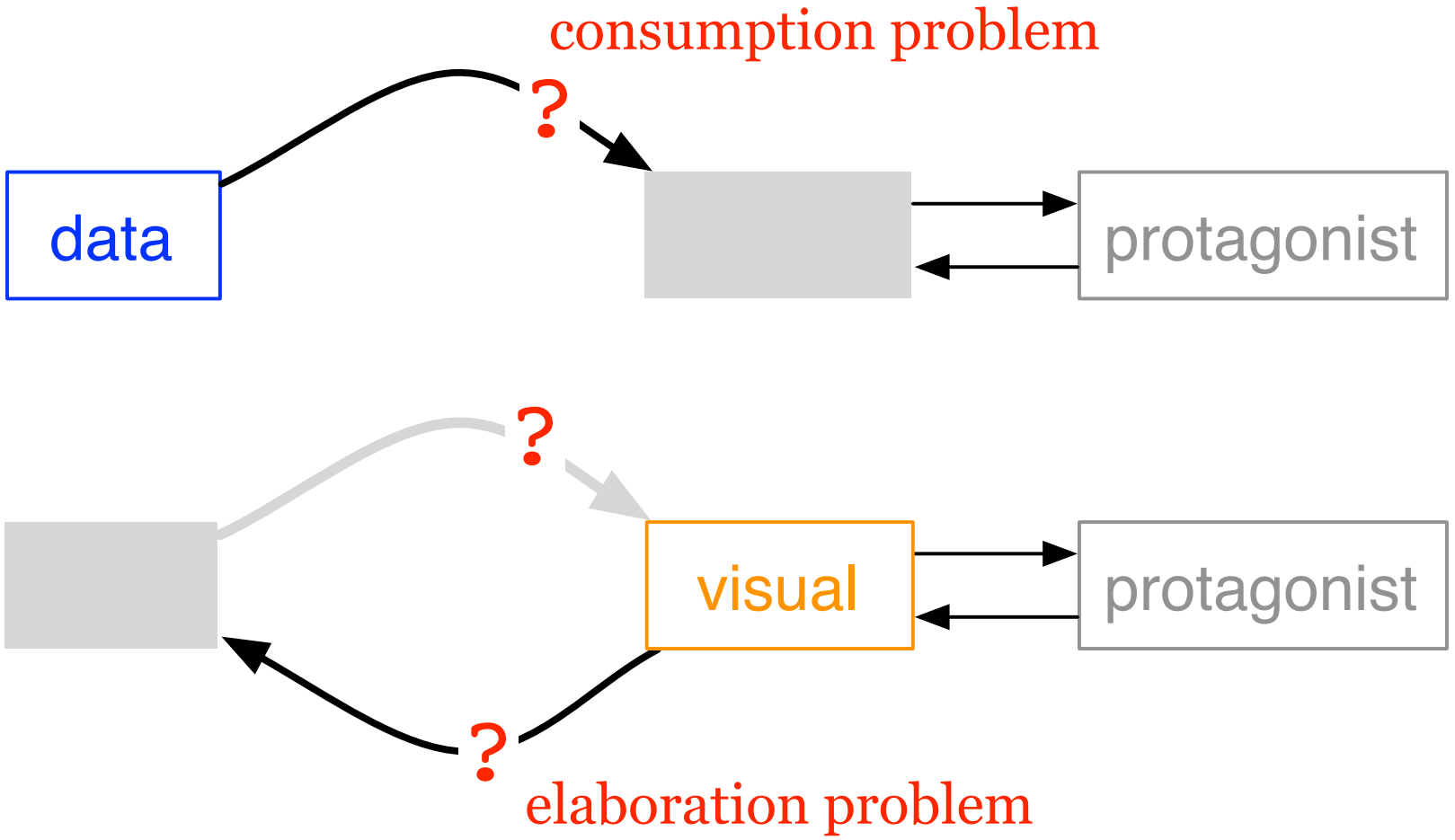
Vision is the predominant method of consumption.

➔ Data must be transformed for visual consumption.

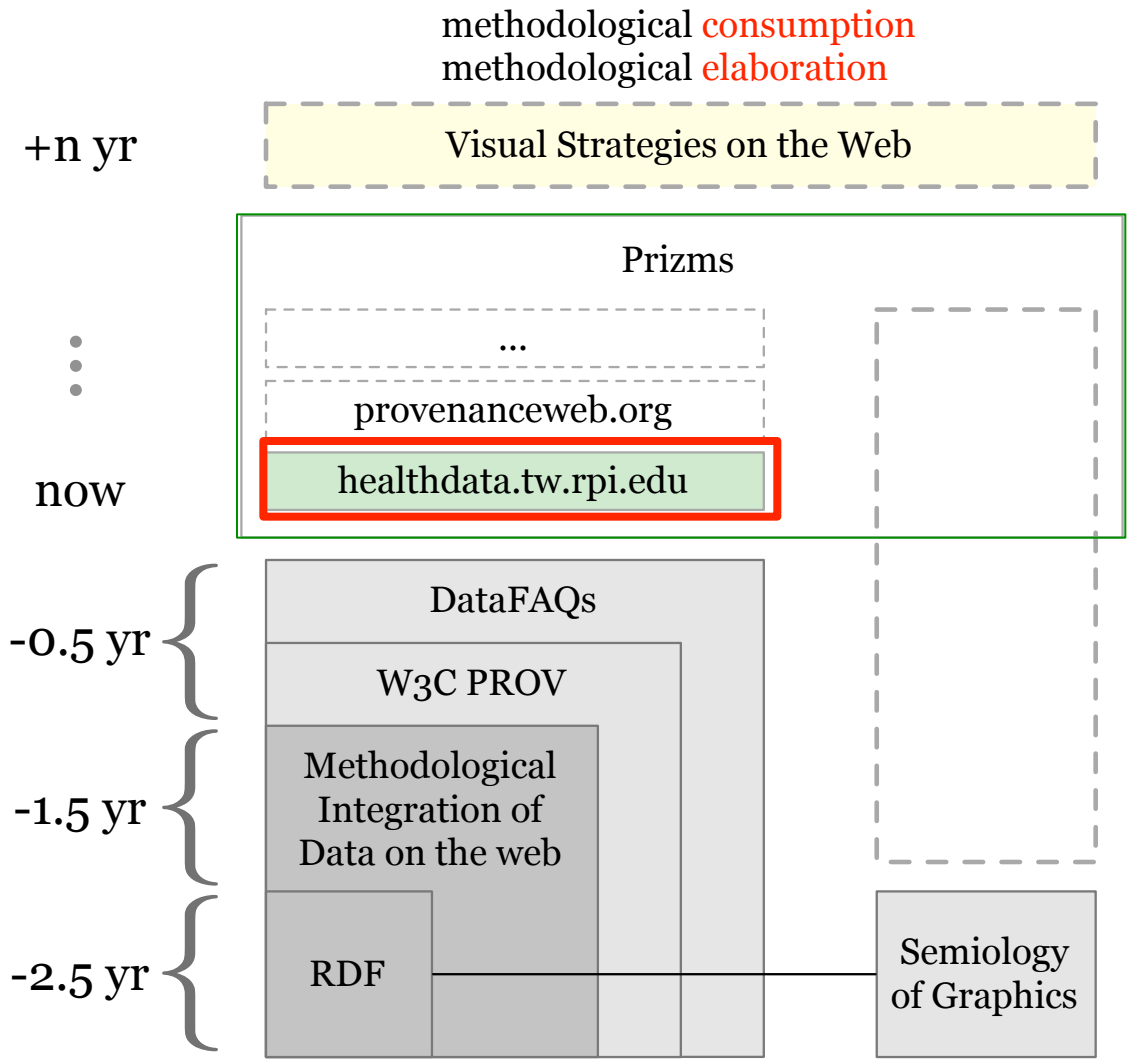


Inquiries of visual content require elaboration of the source data.

Two Symmetric Challenges



Talk Overview



Dept. Health and Human Services Developer Challenges

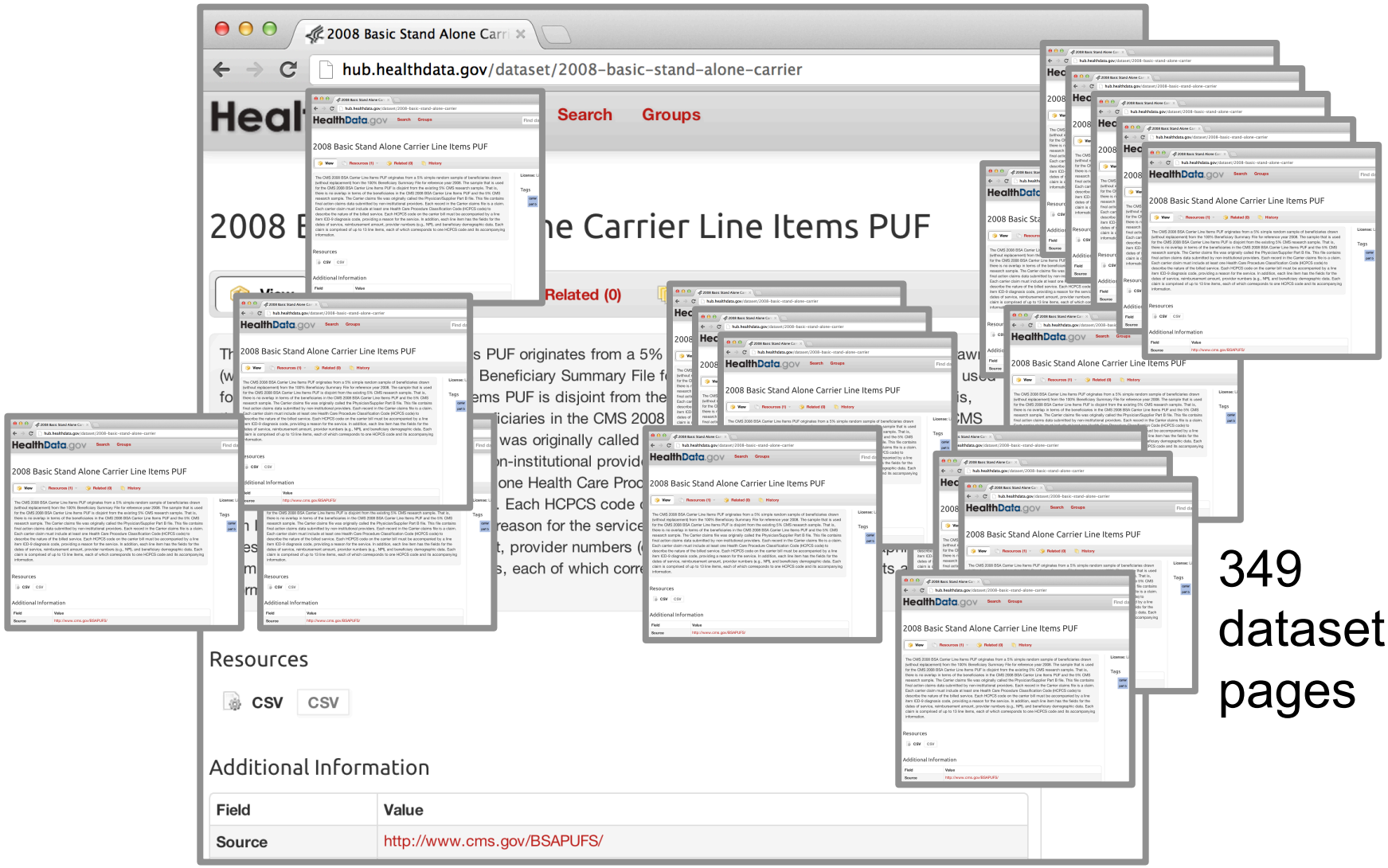
"Establish learning communities that collaboratively evolve and mature the utility and usability of a broad range of health and human service data."

"Facilitate better integration, analysis, and interpretation of our data, helping meet the growing demand for value added information services."

HealthData.gov

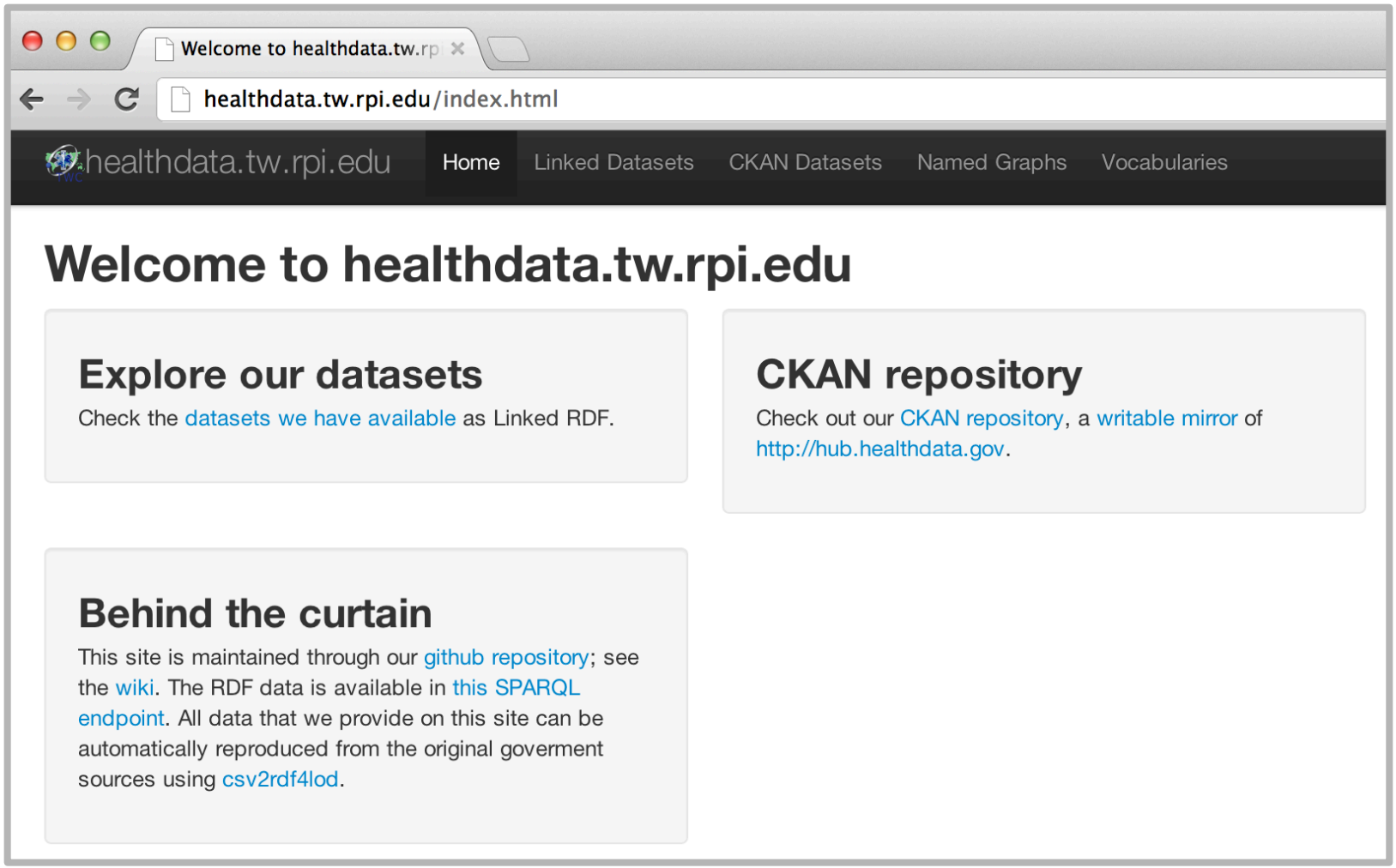
http://www.healthdata.gov/developer-challenges-overview

http://hub.healthdata.gov

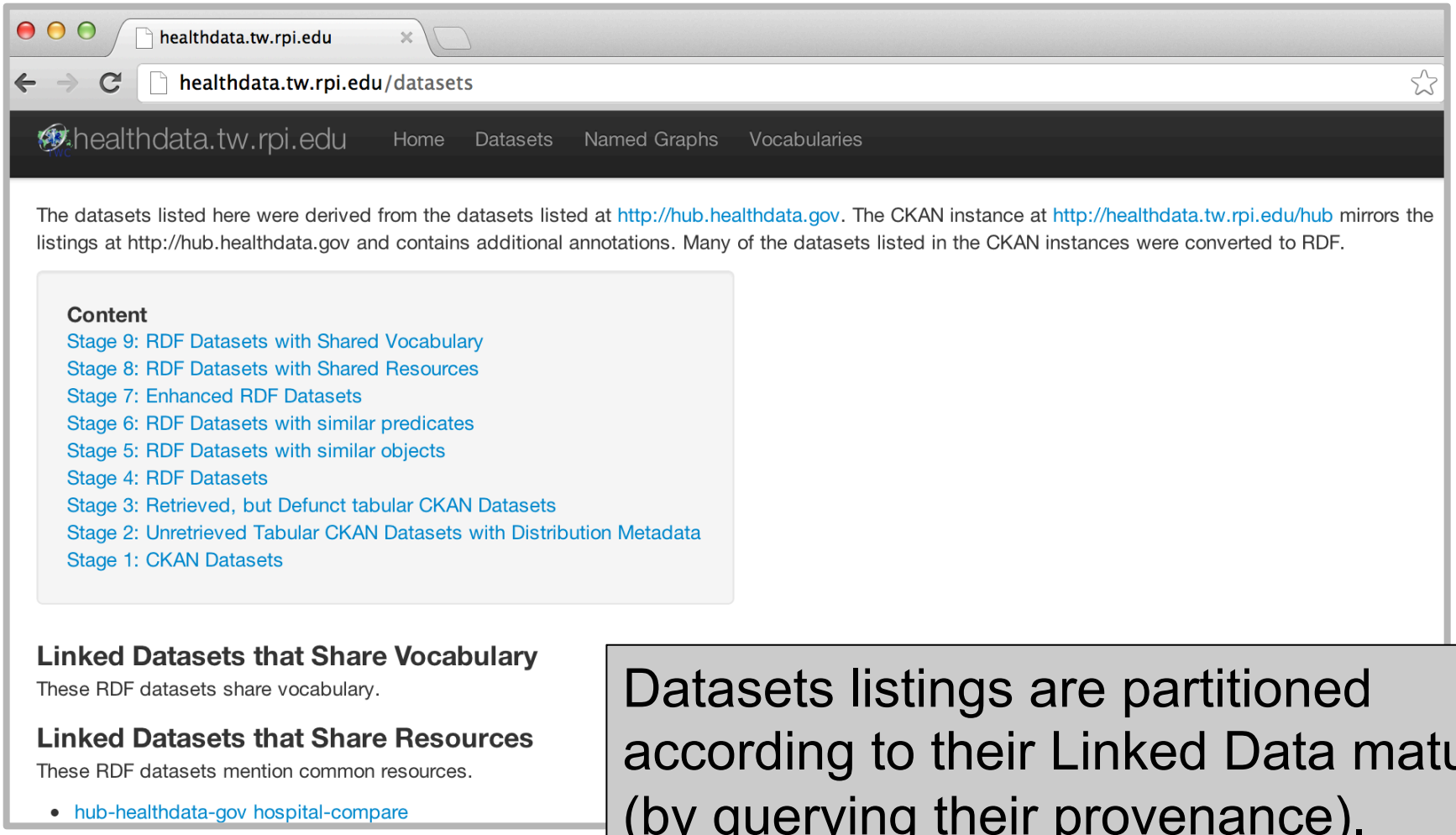


349 dataset pages

http://healthdata.tw.rpi.edu

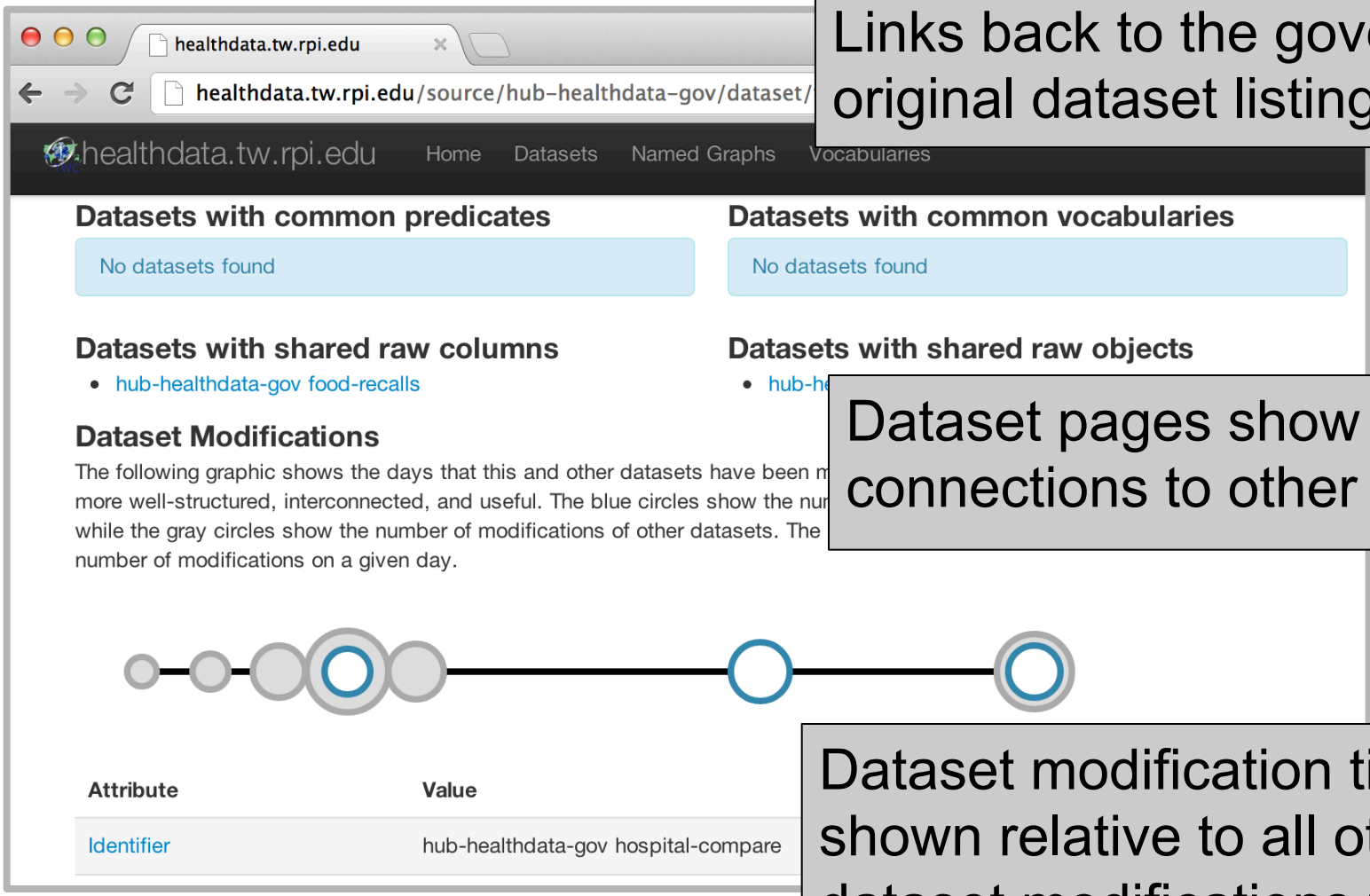


Dataset listings



Datasets listings are partitioned according to their Linked Data maturity (by querying their provenance).

Dataset metadata

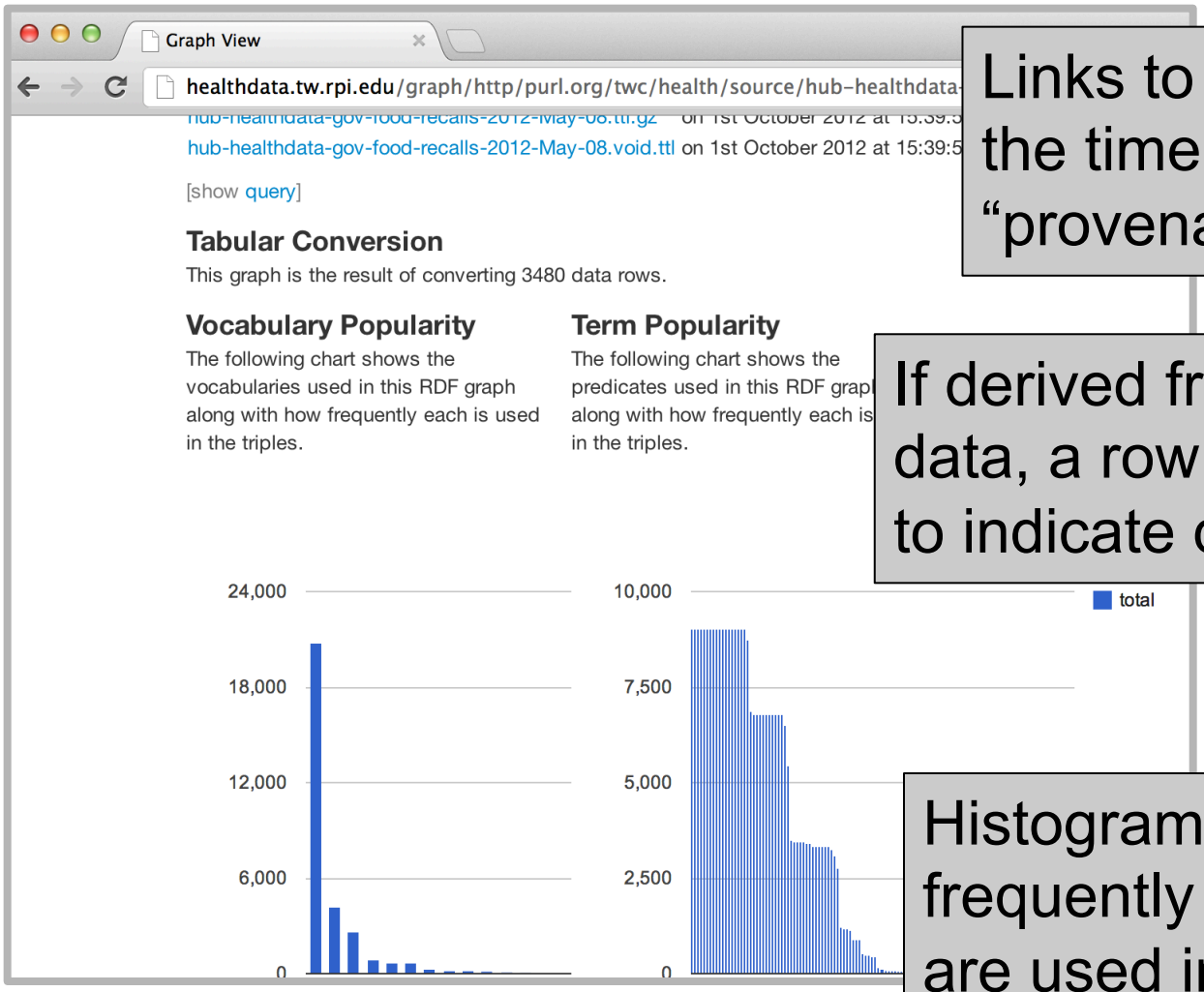


Links back to the government's original dataset listing.

Dataset pages show connections to other datasets.

Dataset modification times are shown relative to all other dataset modifications times.

Overview of loaded graphs

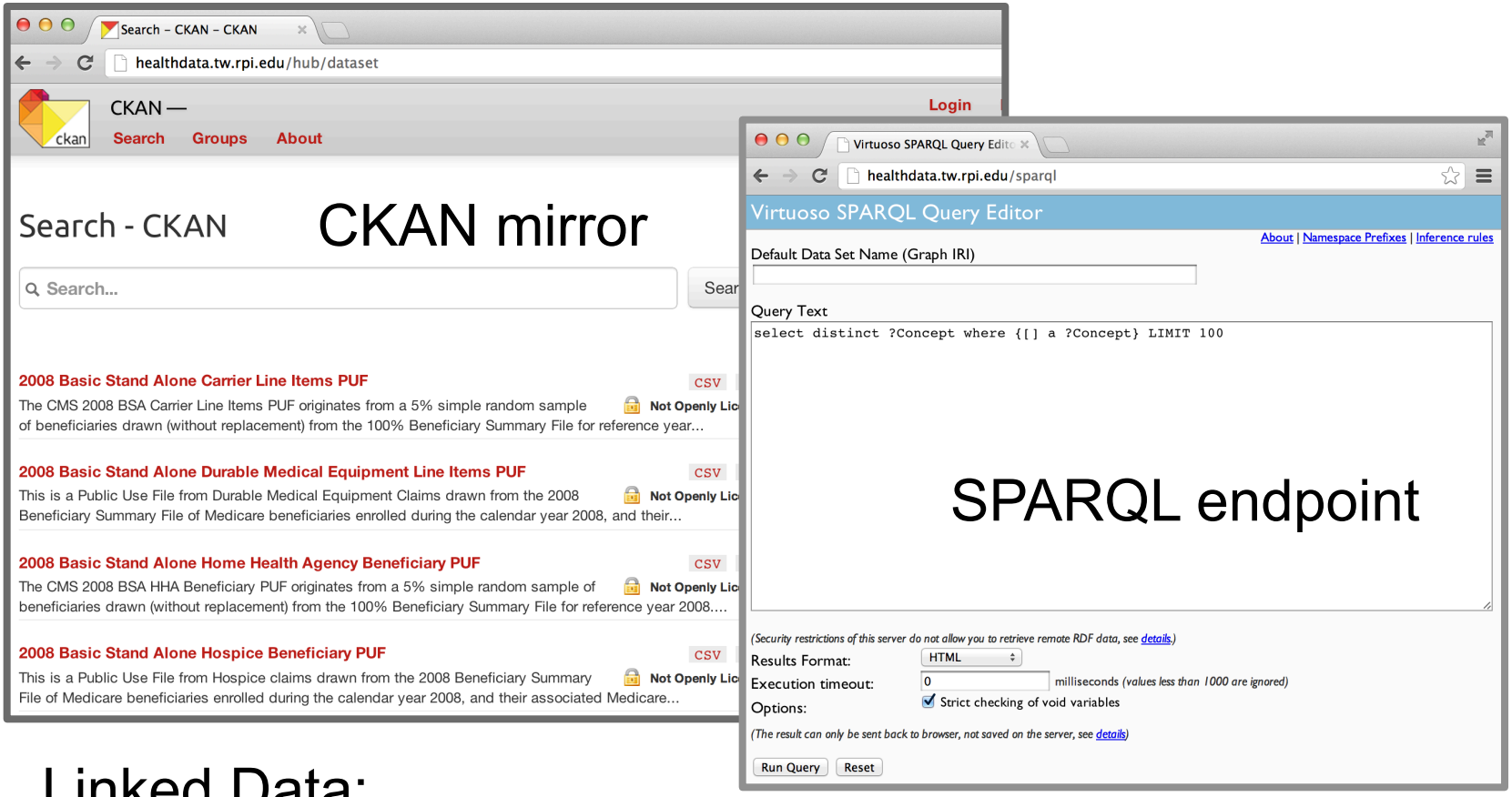


Links to the loaded files, and the time they were loaded (a "provenance freebie").

If derived from tabular data, a row count is given to indicate dataset size.

Histograms show how frequently class and properties are used in the data.

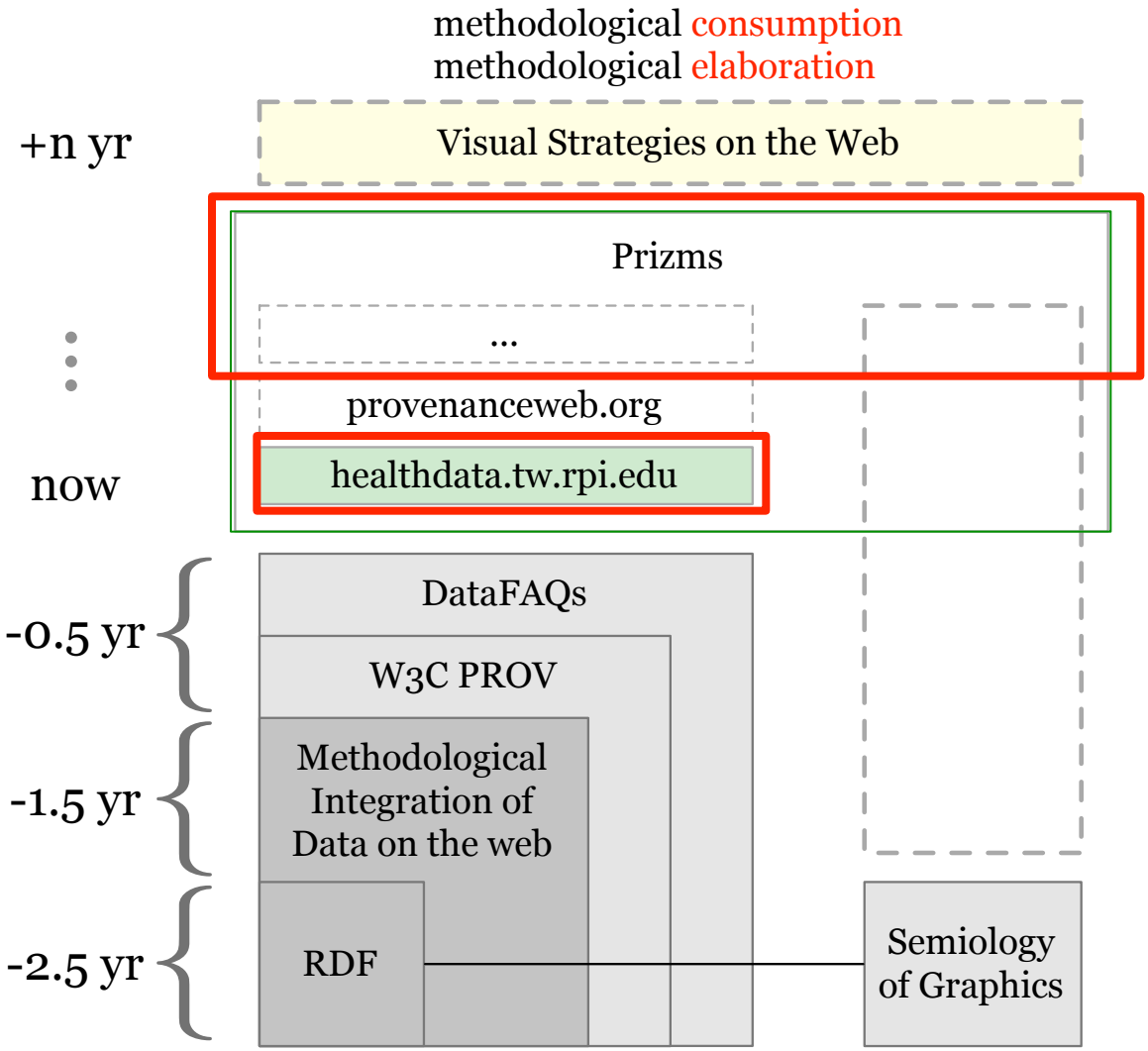
Reusing Others' Tools

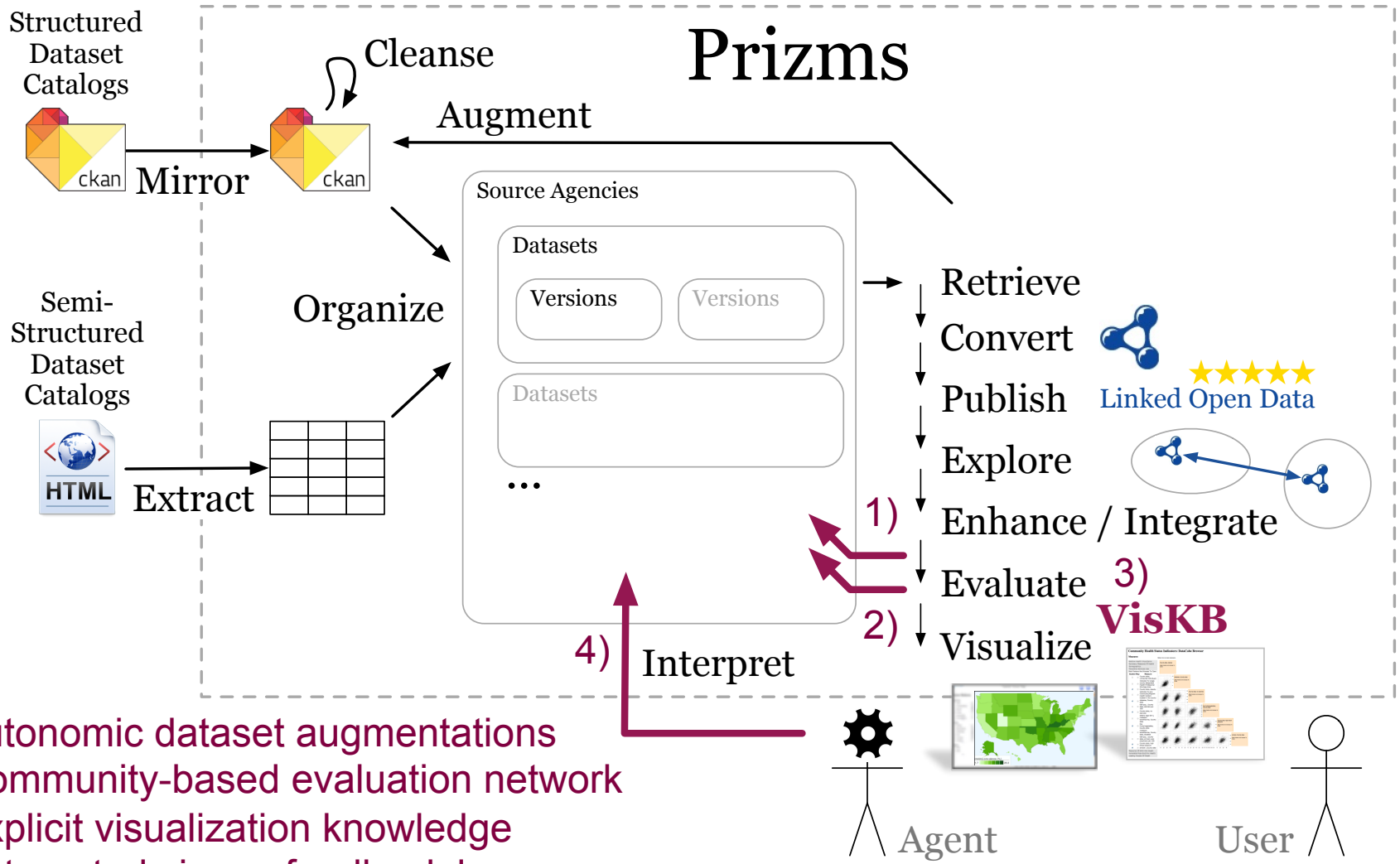


Linked Data:

```
curl -H "Accept: application/rdf+xml" –L http://purl.org/twc/health/source/
hub-healthdata-gov/dataset/hospital-compare
```

Overview

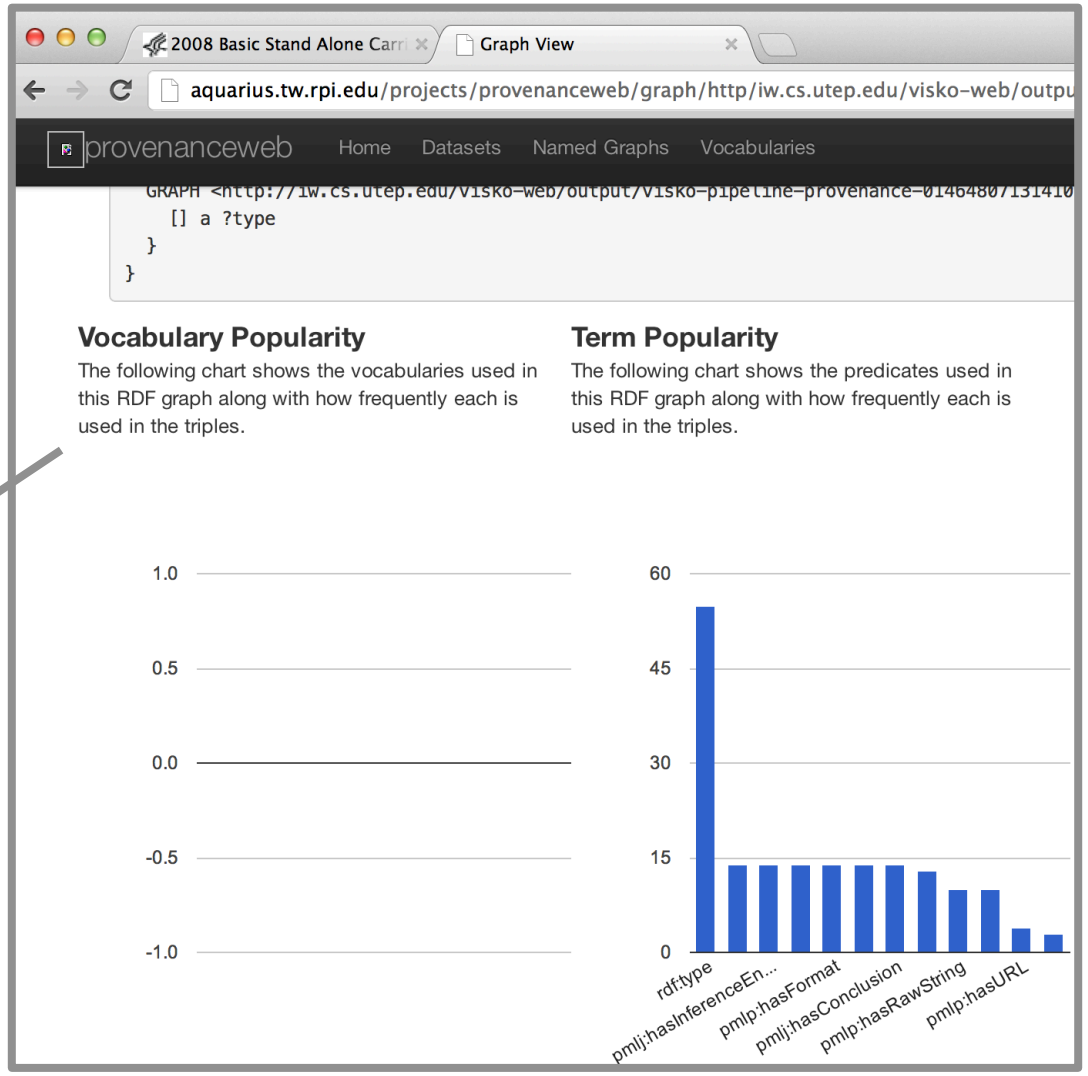

methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
-0.5 yr
DataFAQs
W3C PROV
-1.5 yr
Methodological Integration of Data on the web
-2.5 yr
RDF
Semiology of Graphics



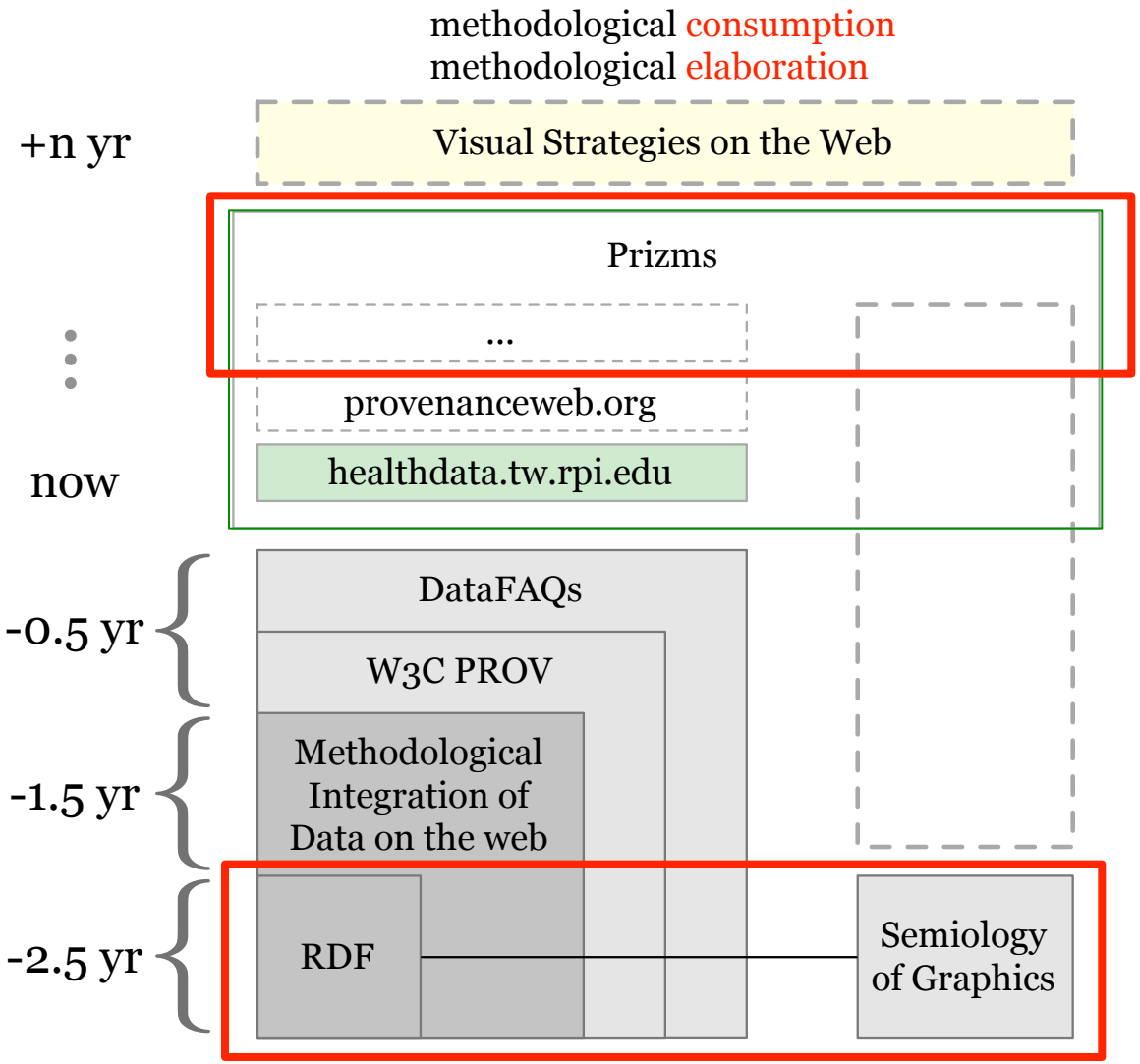
1) Autonomic dataset augmentations
2) Community-based evaluation network
3) Explicit visualization knowledge
4) Automated viewer feedback loop

Applying Prizms

…
http://ieeevis.tw.rpi.edu
http://provenanceweb.org
http://healthdata.tw.rpi.edu
http://logd.tw.rpi.edu
…

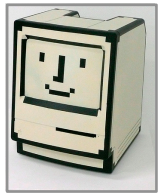


Review

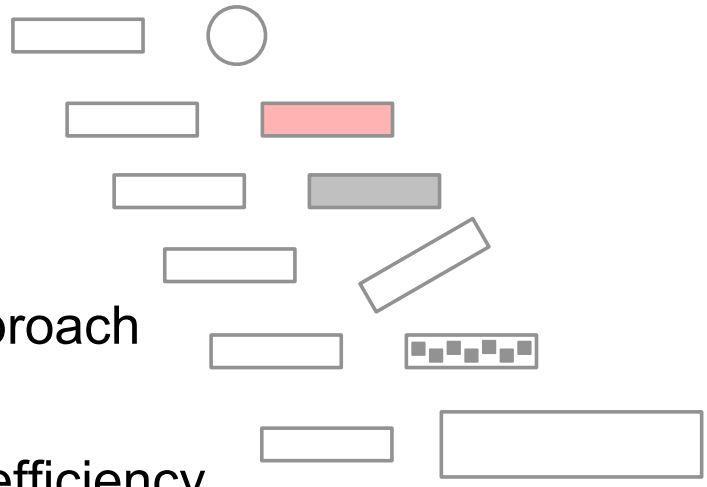

methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
DataFAQs
-0.5 yr
W3C PROV
Methodological Integration of Data on the web
-1.5 yr
RDF
-2.5 yr
Semiology of Graphics

Inspiration for Visual Strategies on the Web

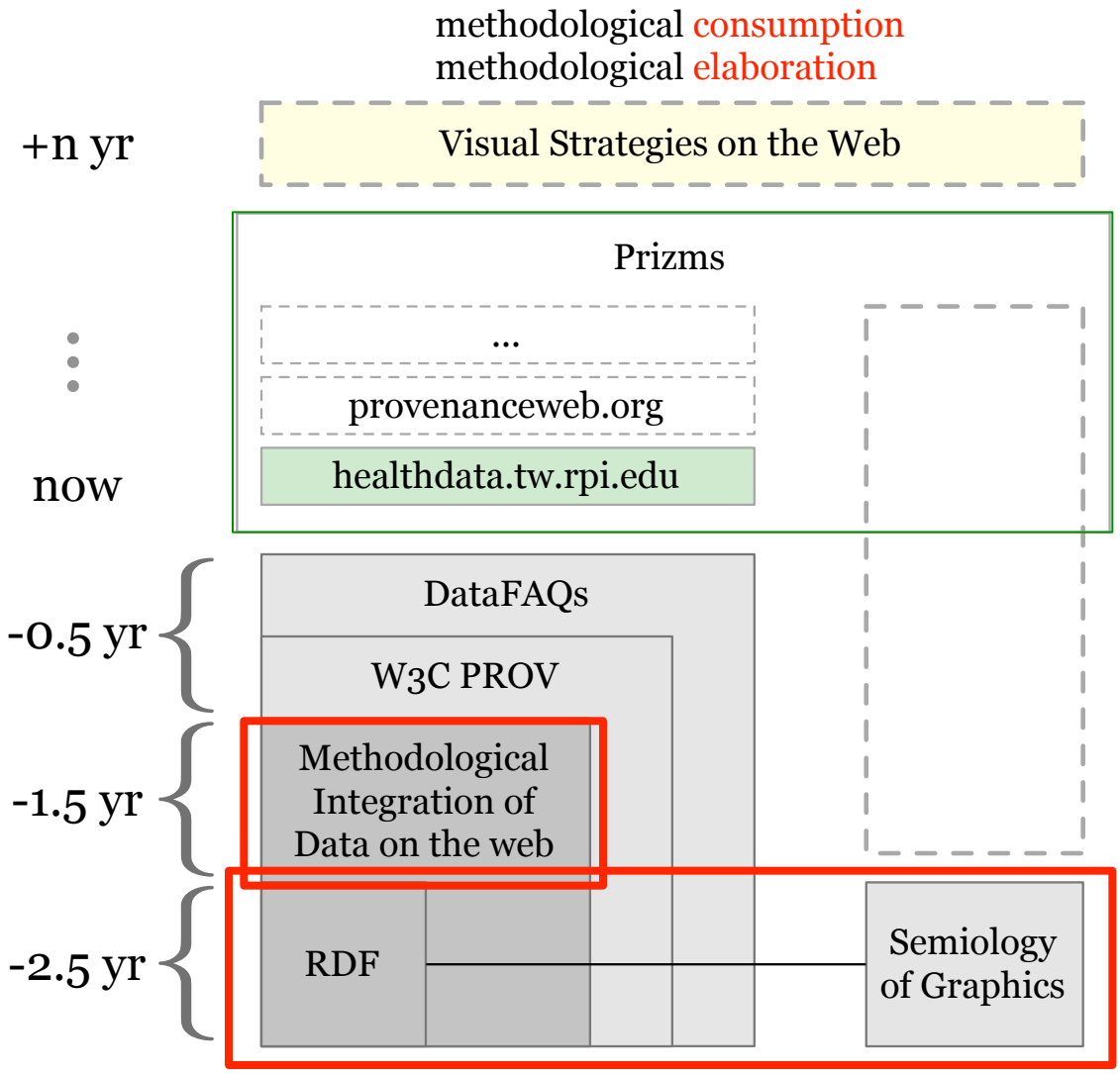
- RDF: My Data Hero
  - What are the things you're talking about?
  - How do those things relate?
  - Can I get more information about a thing with JUST its name?
  - Can I merge your data with any other data in the same format?



- Bertin's Semiology of Graphics
  - Set of Information
  - Analysis of the Information
  - Domain vs. Graphical sign-systems
  - Imposition, implantation, perceptual approach
  - Stages in reading a graphic
  - Level of reading, instant of perception, efficiency



Review


methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
DataFAQs
-0.5 yr
W3C PROV
Methodological Integration of Data on the web
-1.5 yr
RDF
-2.5 yr
Semiology of Graphics

Good RDF and Bad RDF

"Hawaii",
"75-6129
"96740",

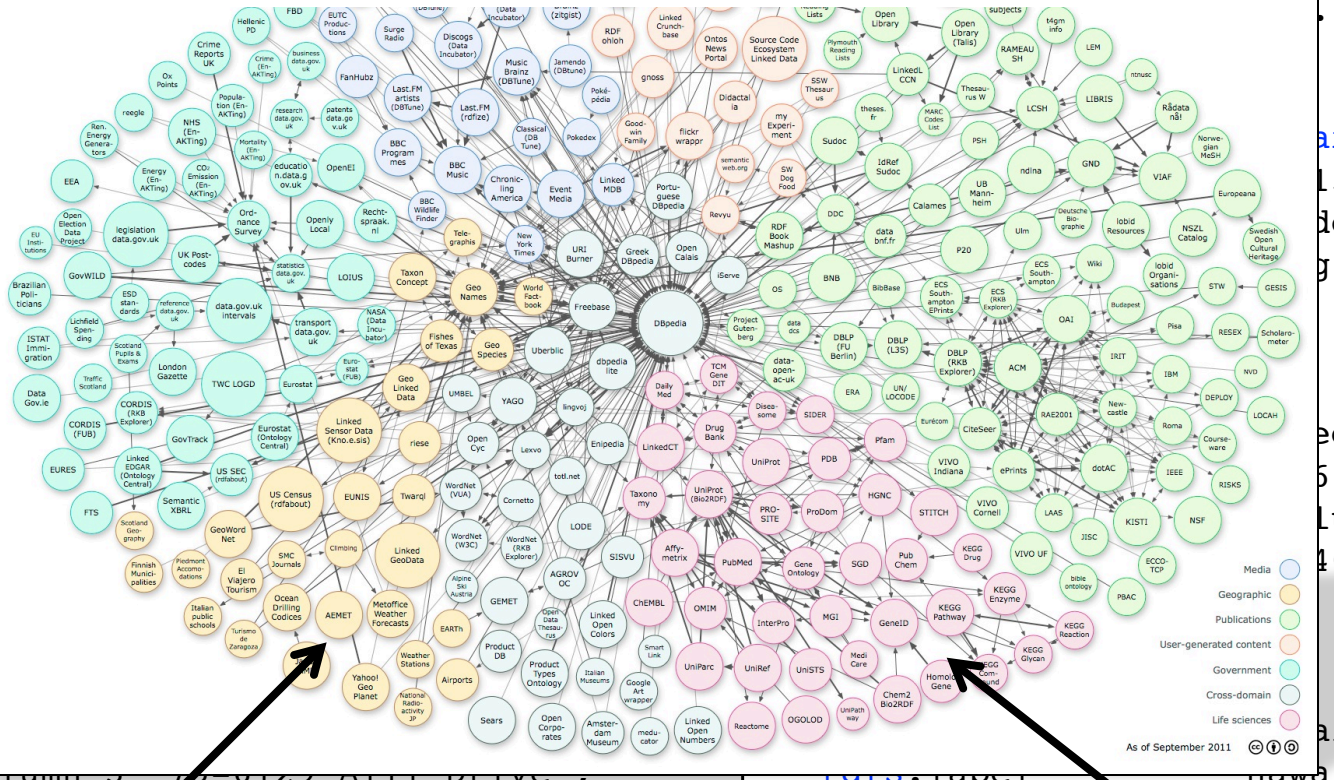
ds4383:th
raw:co
raw:co
raw:co
raw:column_4 "Kailua-Kona";
raw:column_5 "96740";
raw:column_6 "-155.9819183";
raw:column_7 "19.61436844" .

arket;
1367;
den Market Place";
g 19.6 .

ed_state:Hawaii;
5129 Alii Drive";
lua-Kona";
40" .

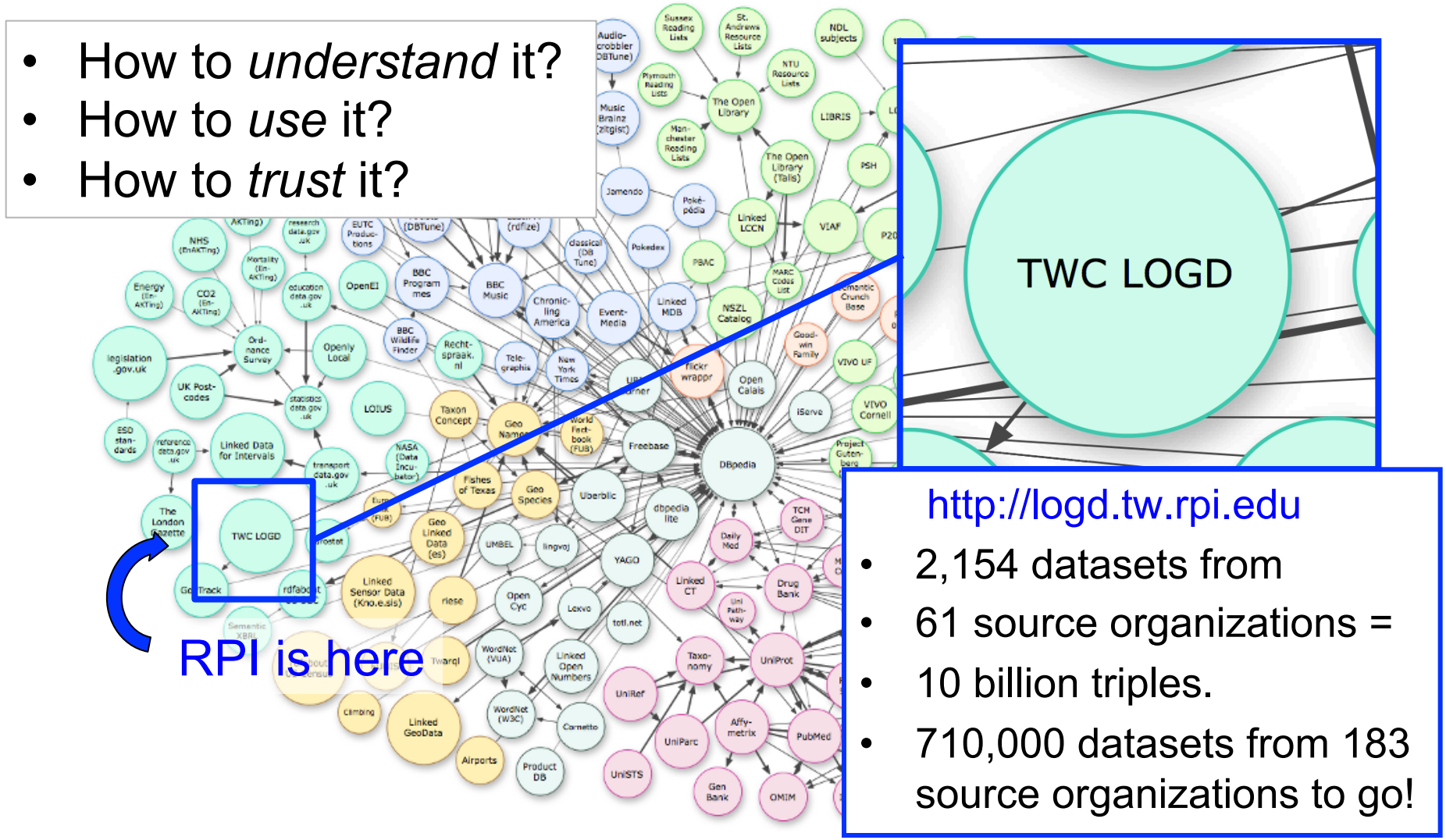
aii";
Hawaii";
owl:sameAs
<http://sws.geonames.org/5855797/>,
govtrackusgov:HI,
dbpedia:Hawaii .

Published as data elements in the WWW



LOGD: Linking Open Government Data

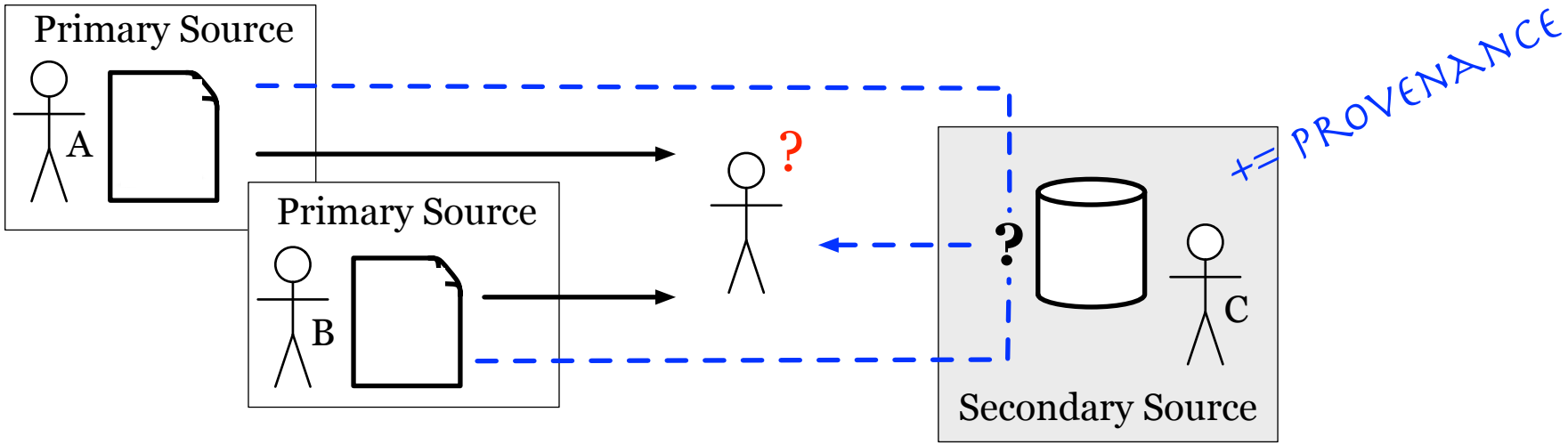
- How to *understand* it?
- How to *use* it?
- How to *trust* it?



http://logd.tw.rpi.edu

- 2,154 datasets from
- 61 source organizations =
- 10 billion triples.
- 710,000 datasets from 183 source organizations to go!

The Integrator's Dilemma



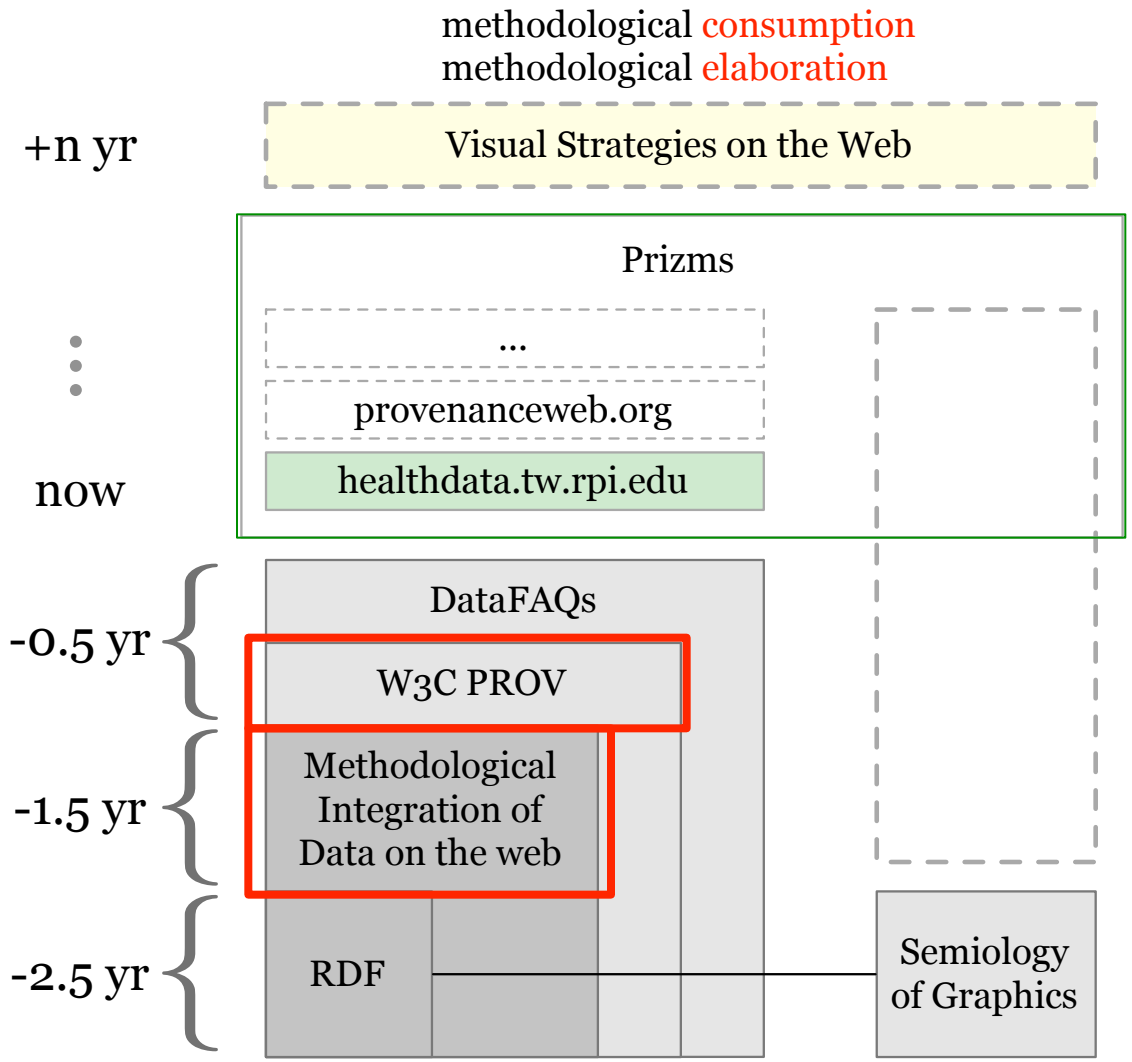
+ "Original"
+ Authoritative
+ Domain Expertise

- "Transformed"
- Non-Authoritative
- No Domain Expertise

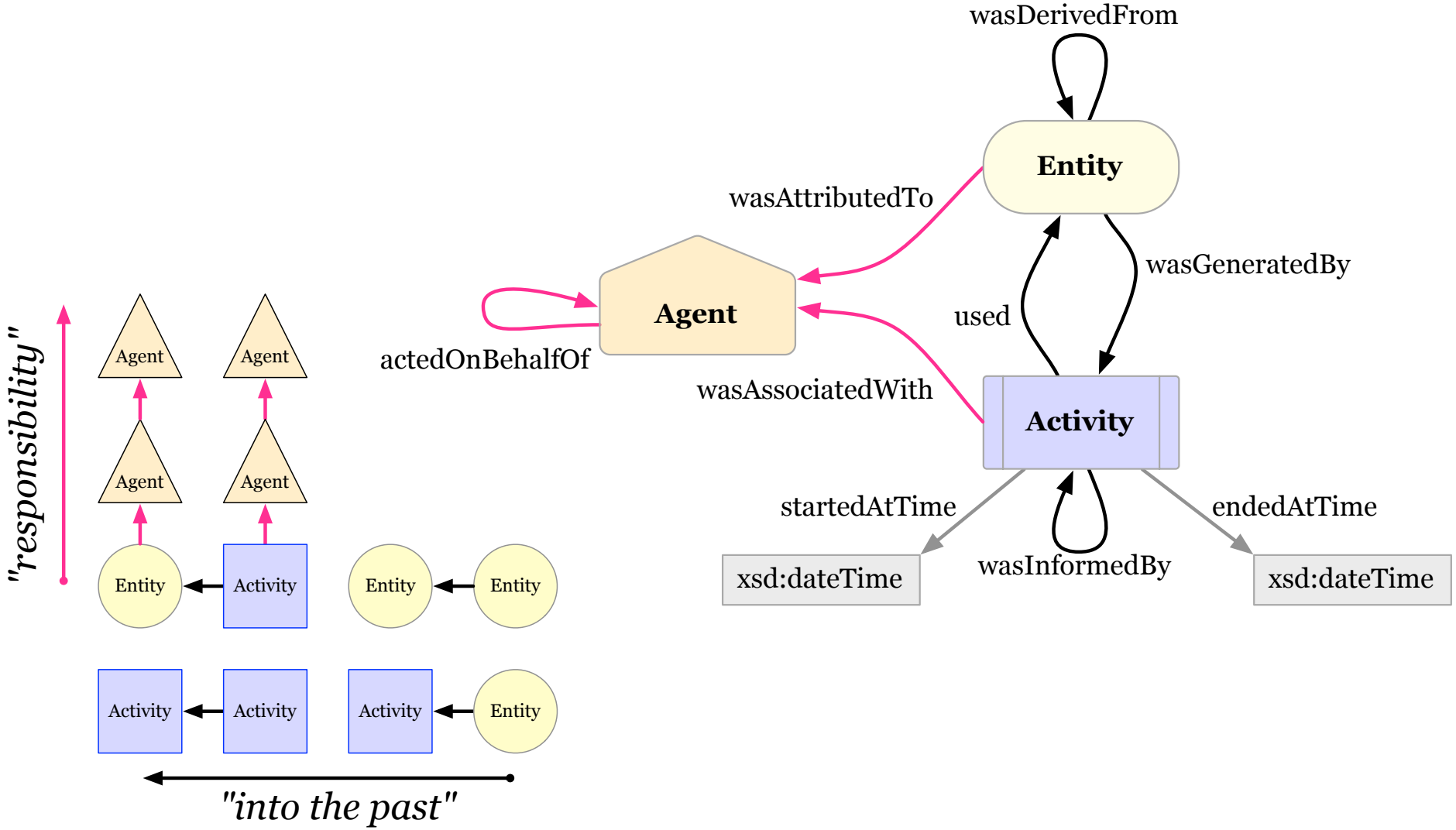
- Not uniform structure
- Not linked

+ Uniform structure
+ Linked (RDF)

Update


methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
DataFAQs
-0.5 yr
W3C PROV
Methodological Integration of Data on the web
-1.5 yr
RDF
-2.5 yr
Semiology of Graphics

W3C Provenance Interchange Model



W3C Provenance Interchange Model

W3C Working Draft

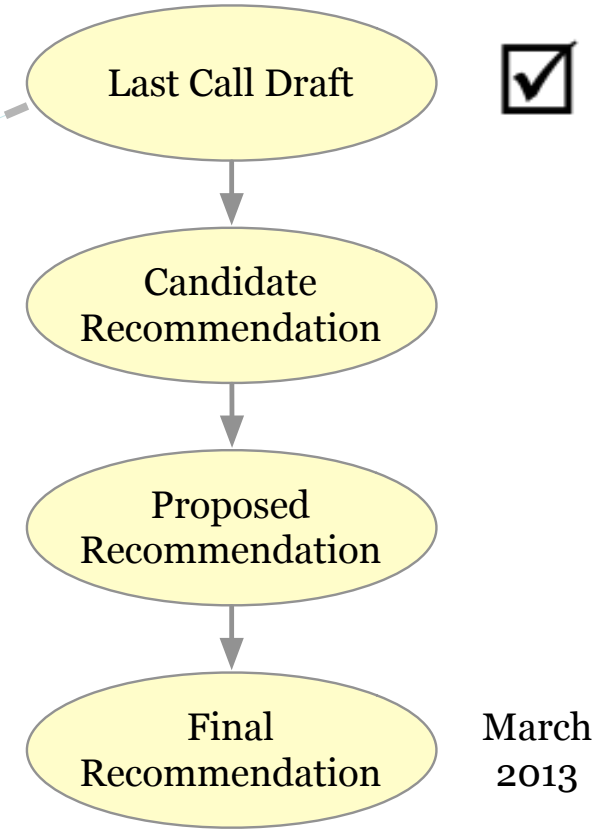
W3C

PROV-O: The PROV Ontology

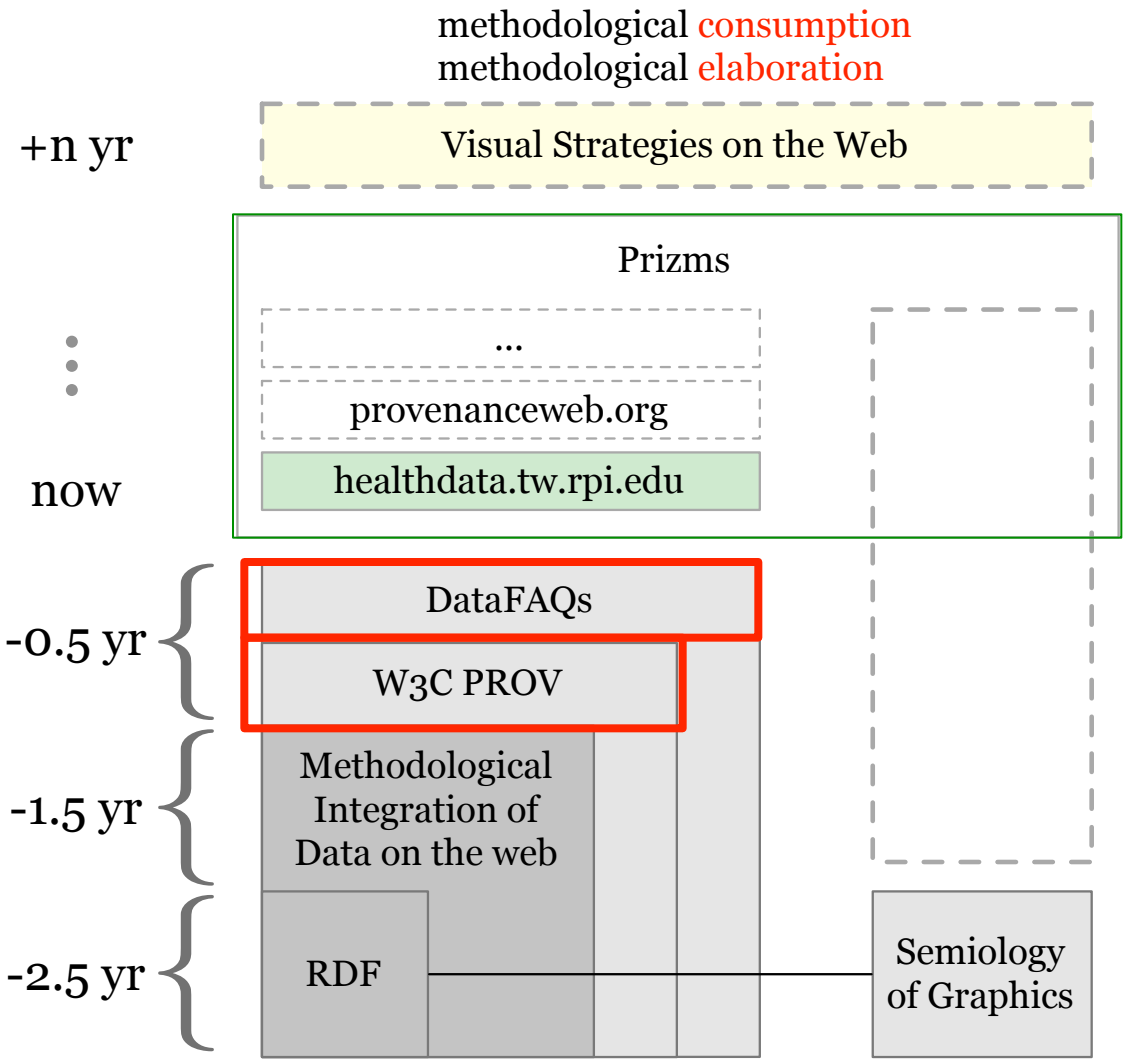
W3C Working Draft 24 July 2012

This version:
http://www.w3.org/TR/2012/WD-prov-o-20120724/
Latest published version:
http://www.w3.org/TR/prov-o/
Latest editor's draft:
https://dvcs.w3.org/hg/prov/raw-file/default/ontology/Overview.html
Previous version:
http://www.w3.org/TR/2012/WD-prov-o-20120503/
Editors:
Timothy Lebo, Rensselaer Polytechnic Institute, USA
Satya Sahoo, Case Western Reserve University, USA
Deborah McGuinness, Rensselaer Polytechnic Institute, USA
Authors:
(In alphabetical order)
Khalid Belhajjame, University of Manchester, UK

Inference WEB

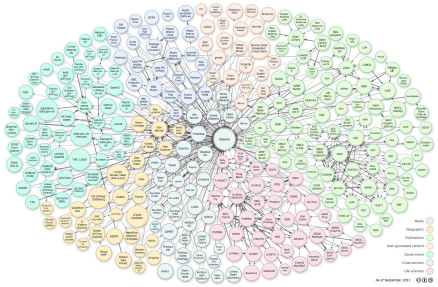


Update

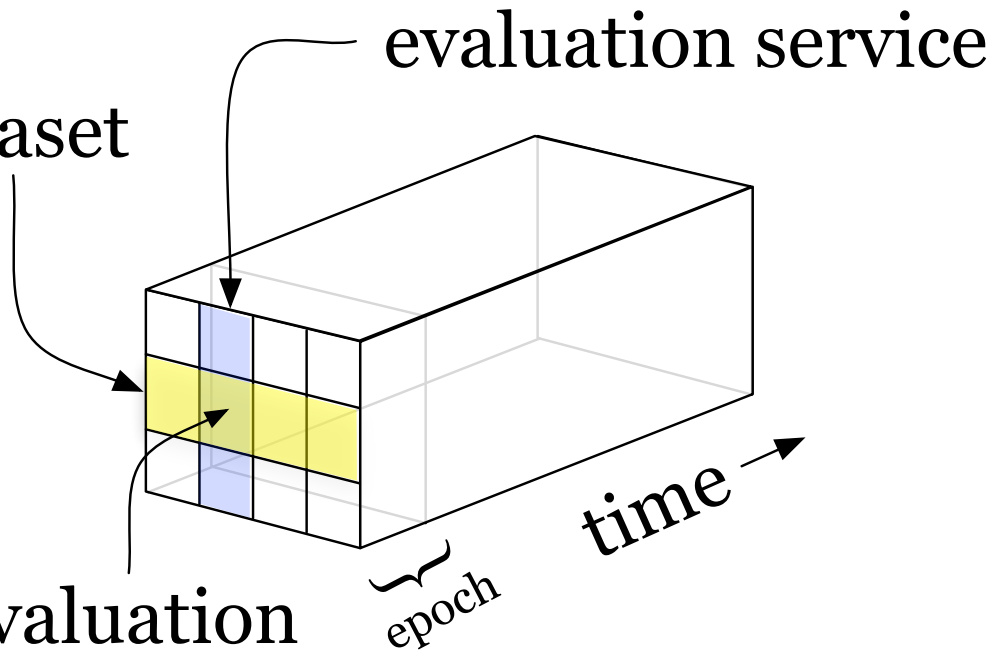

methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
-0.5 yr
DataFAQs
W3C PROV
-1.5 yr
Methodological Integration of Data on the web
-2.5 yr
RDF
Semiology of Graphics

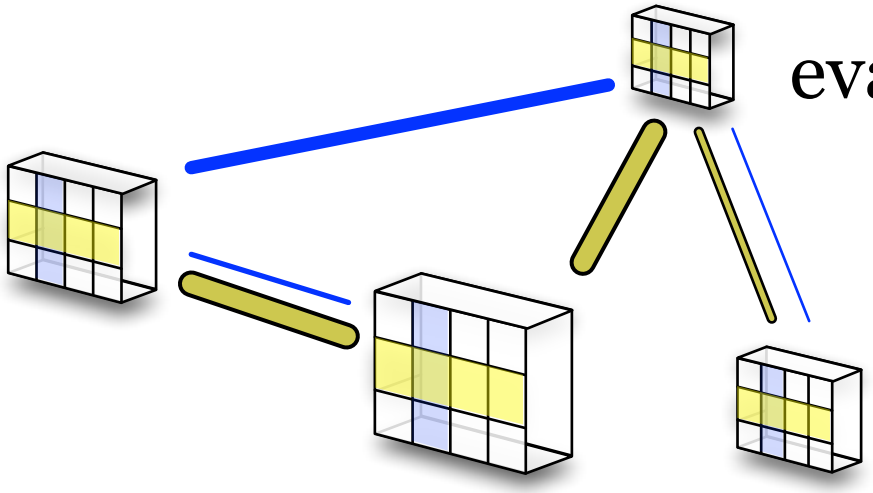
DataFAQs

LINKED DATA QUALITY REPORTS



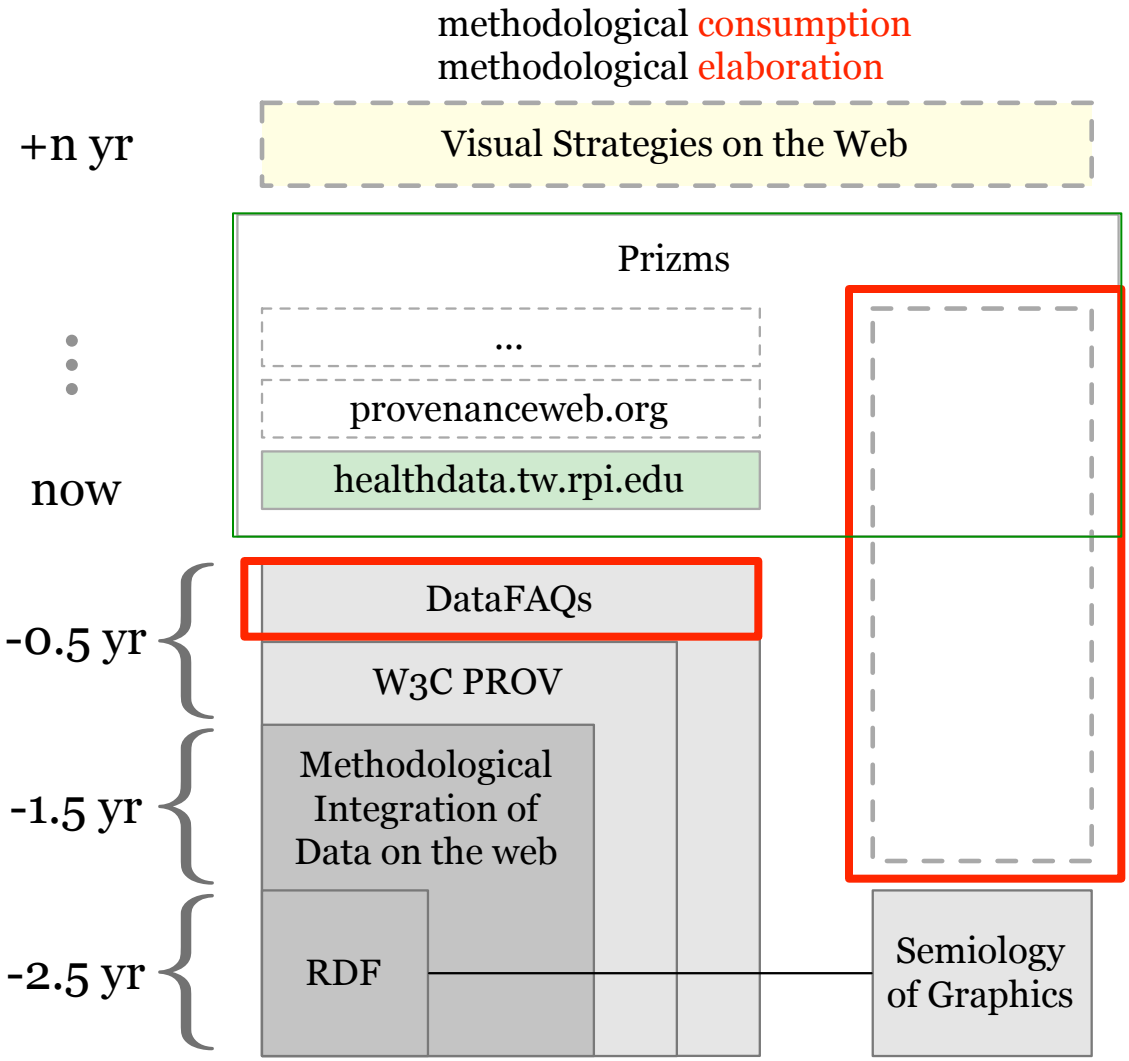
1. Choose Datasets
2. Choose Evaluation Services
3. Explore Evaluations



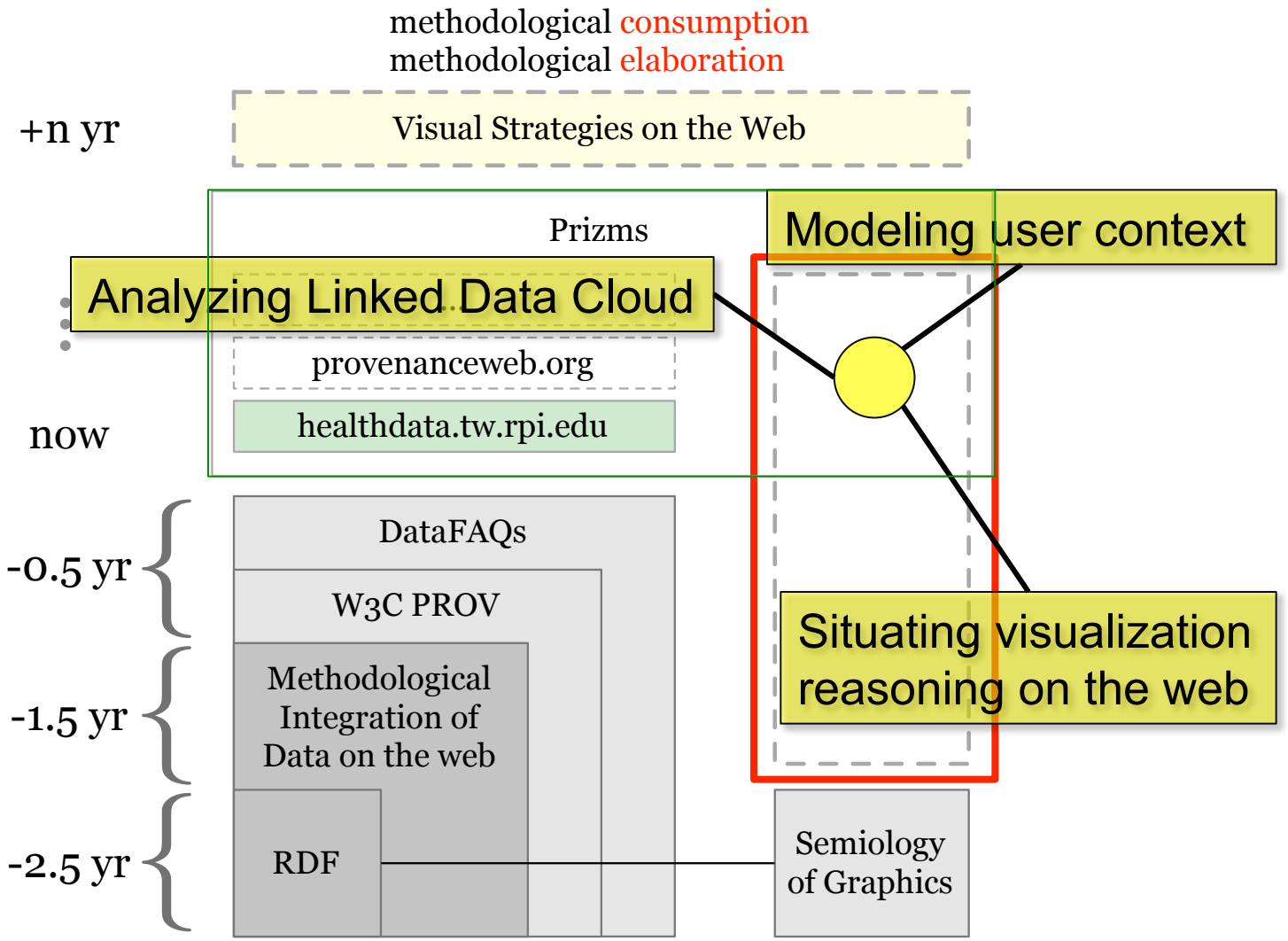


- Individual efforts benefit community
- Automated, quantitative feedback
- Recommendations from the aggregate
- Network effect

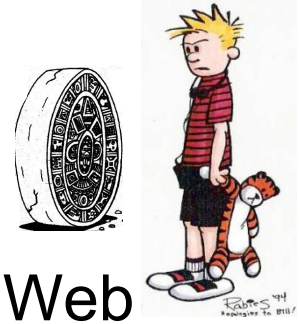
Future Work


methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
...
provenanceweb.org
now
healthdata.tw.rpi.edu
DataFAQs
-0.5 yr
W3C PROV
Methodological Integration of Data on the web
-1.5 yr
RDF
-2.5 yr
Semiology of Graphics

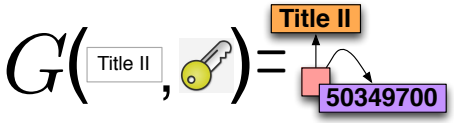
Future Work


methodological consumption
methodological elaboration
+n yr
Visual Strategies on the Web
Prizms
Modeling user context
Analyzing Linked Data Cloud
provenanceweb.org
now
healthdata.tw.rpi.edu
-0.5 yr
DataFAQs
W3C PROV
-1.5 yr
Methodological Integration of Data on the web
Situating visualization reasoning on the web
-2.5 yr
RDF
Semiology of Graphics

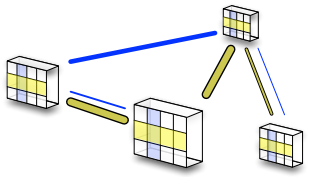
Conclusion



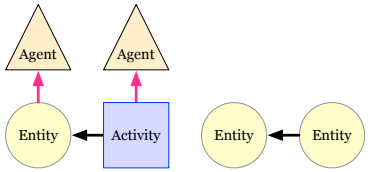
- Motivation with Calvin
- Integrating data on the Web
  - Contextualizing entity names, relations



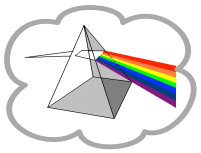
- DataFAQs quality evaluation framework
  - Encapsulating and gathering contextualized perspectives



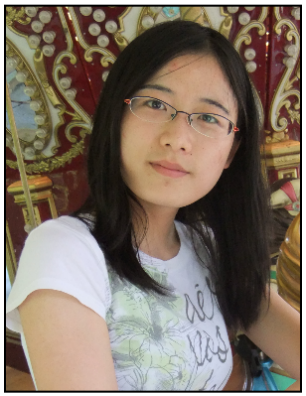
- W3C Provenance Data Model
  - Relating to an historical context

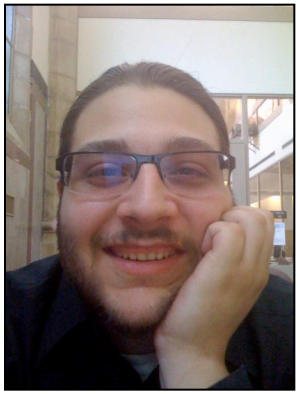


- Prizms (and future work)
  - Visualization knowledge: pairing user context with data context



Thanks!



- Ping Wang
- Alvaro Graves
- Jim McCusker
- Dominic DiFranzo
- Yu Chen
- Josh Shinavier

Deborah McGuinness