



↔ [http://bit.ly/lebo\\_twed](http://bit.ly/lebo_twed)



# *csv2rdf4lod*

*Long name, amazing results*



Timothy Lebo  
Tetherless World Constellation  
Rensselaer Polytechnic Institute



Rensselaer

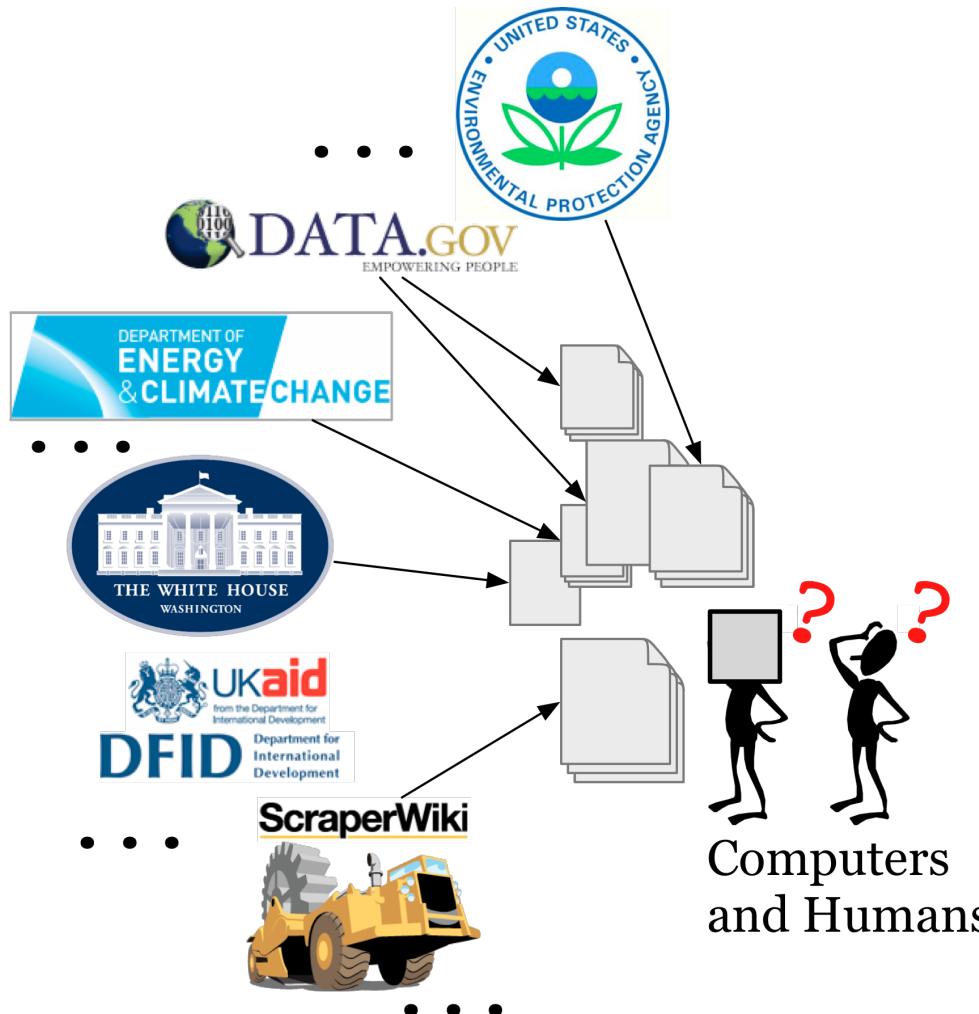


# Outline

- Motivation
- *csv2rdf4lod*
  - Quick and dirty: a smart, naïve bootstrap
  - Time and attention: adding enhancements
  - Breadcrumbs: tracking provenance
  - Reproducibility and distribution: version control
  - Verification: testing
- Future Work
- Conclusions and Questions
- Hands-on tutorial

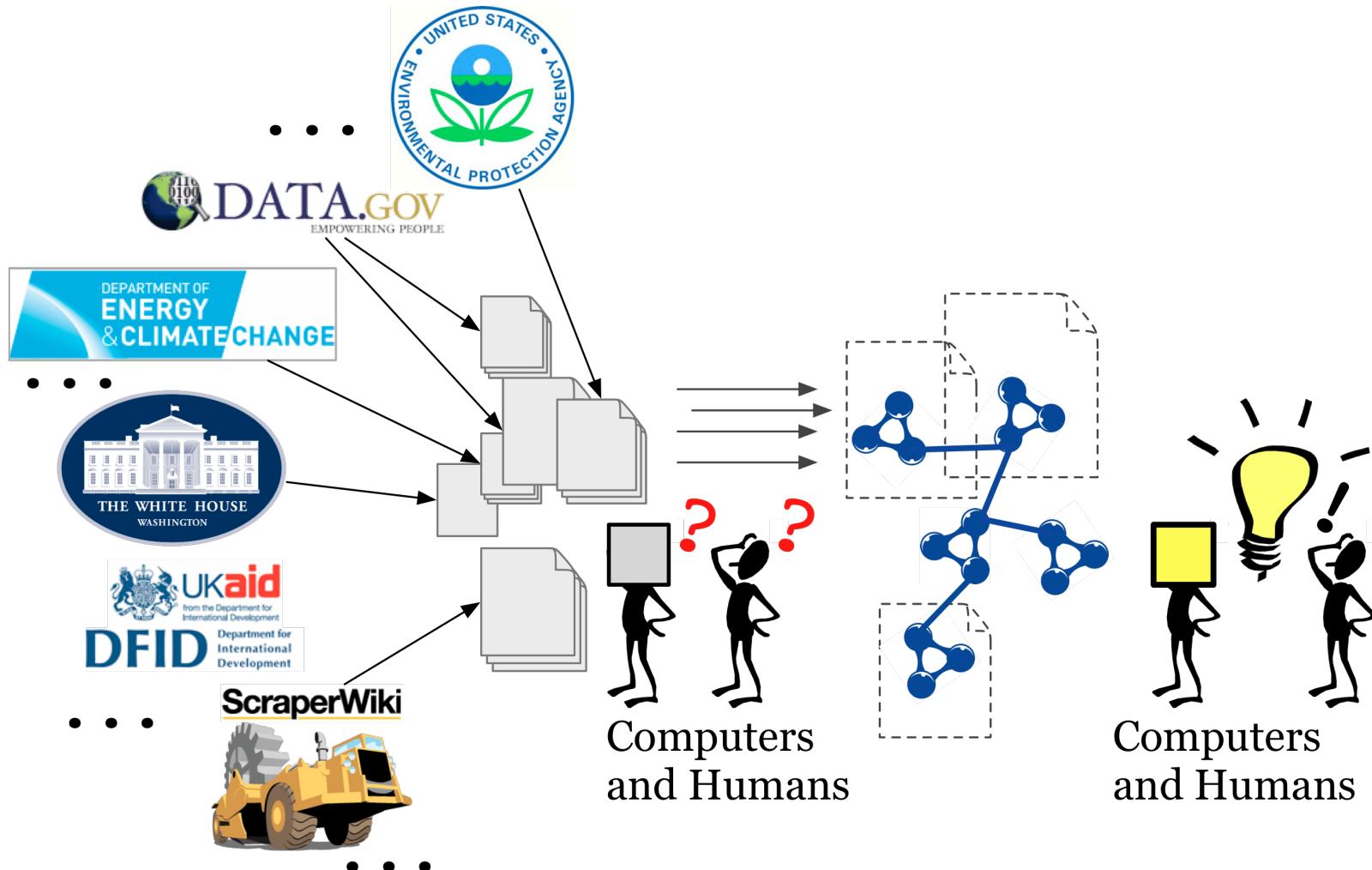


# Motivation



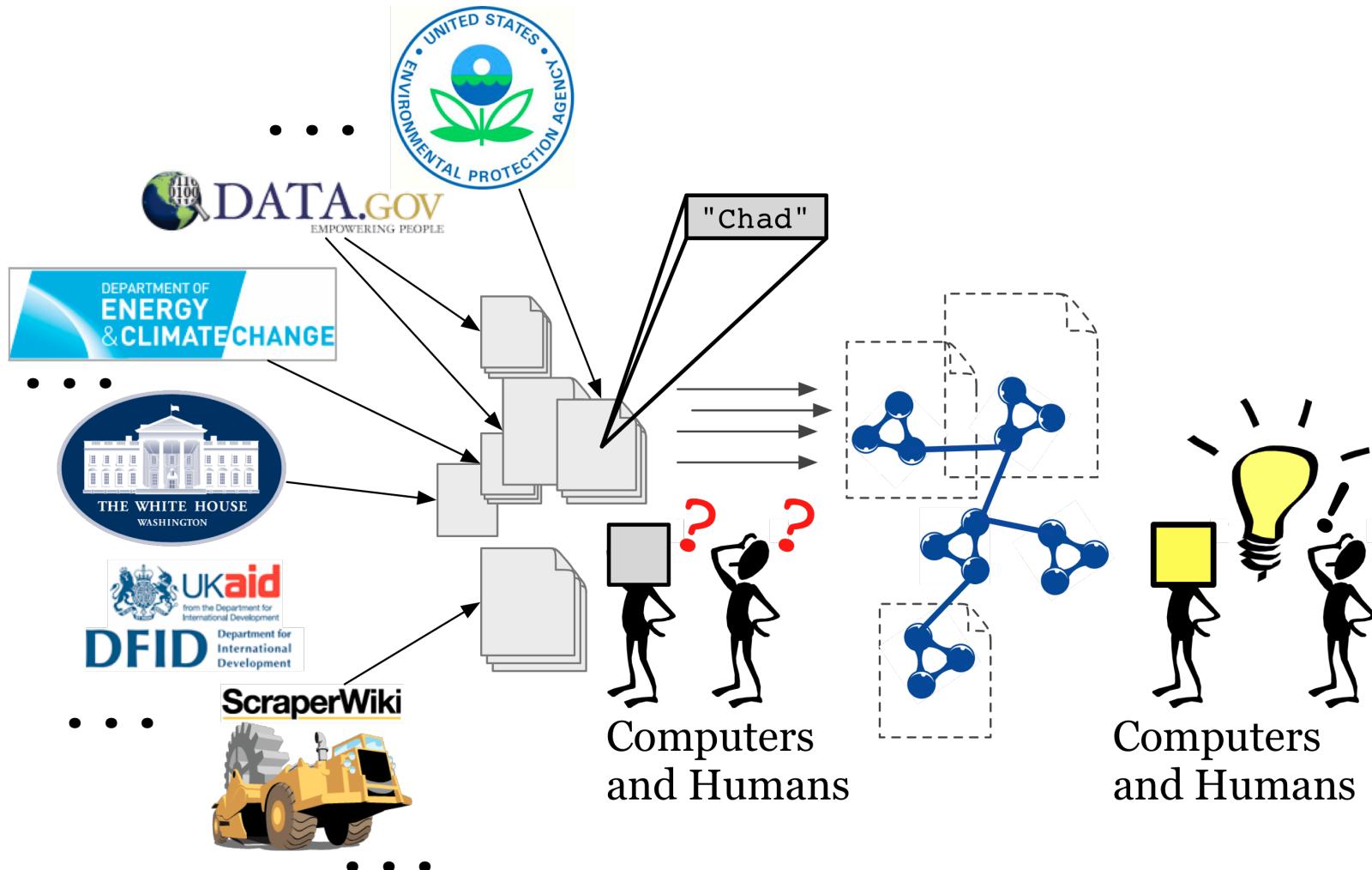


# Motivation



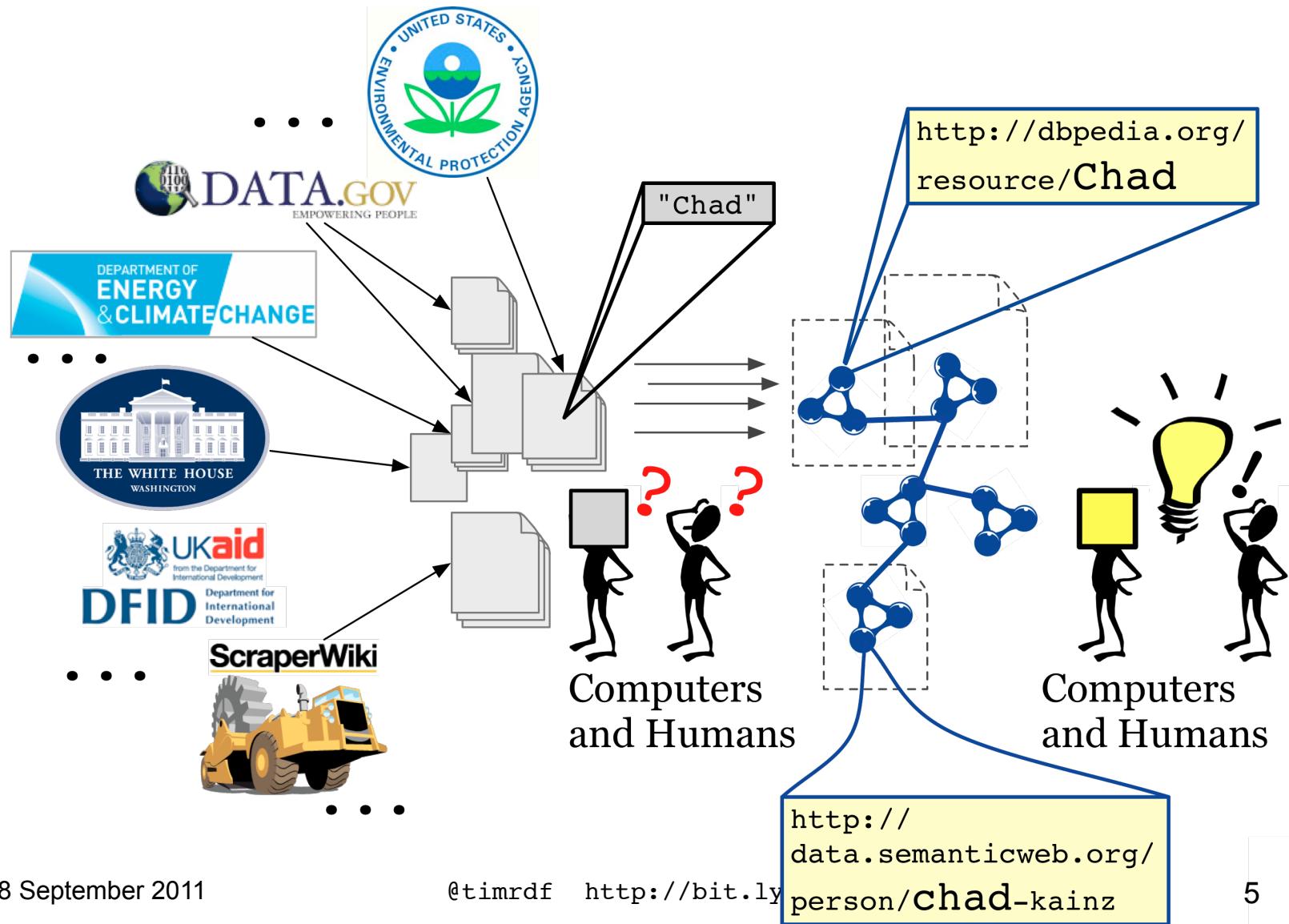


# Motivation



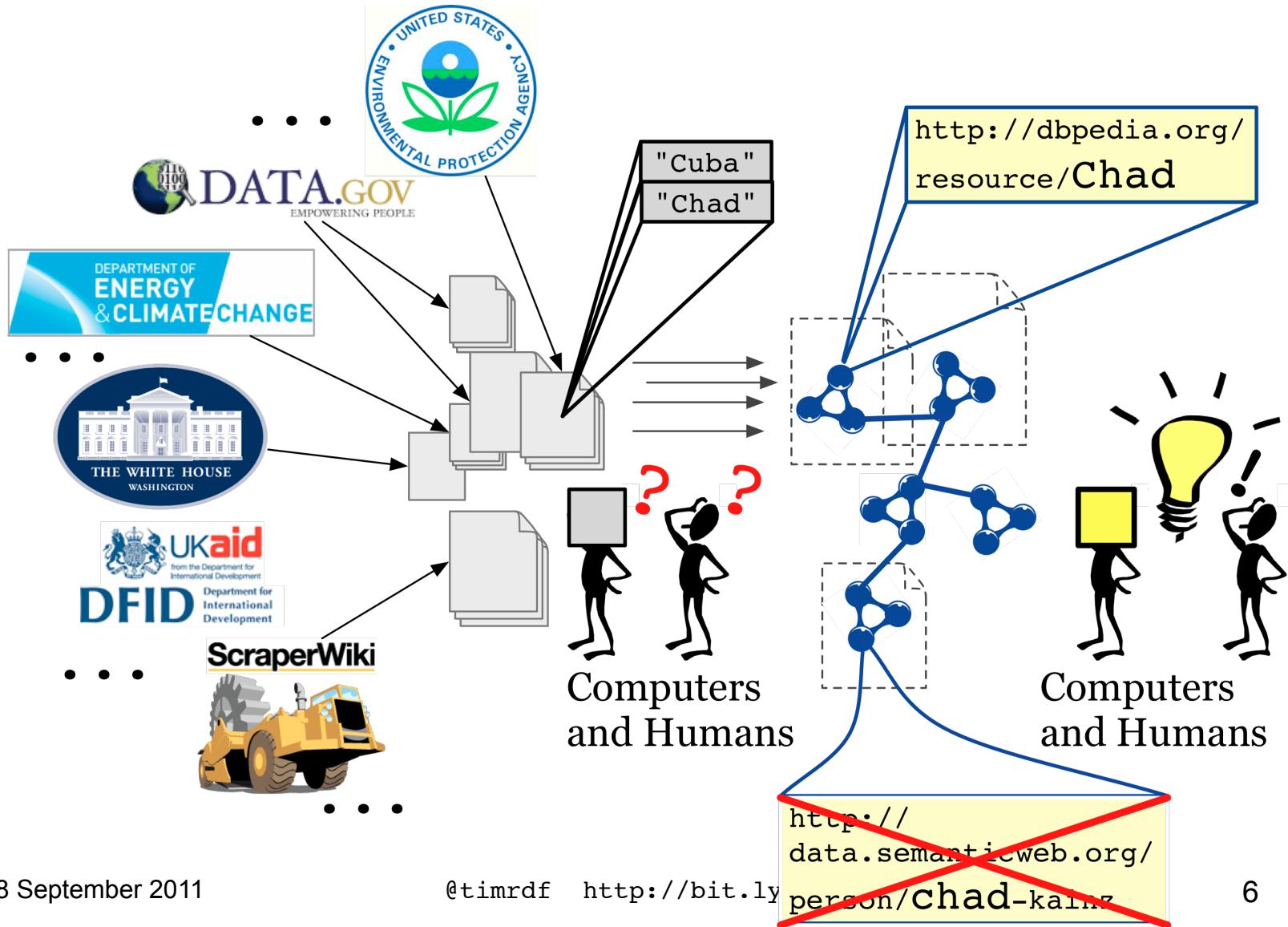


# Motivation



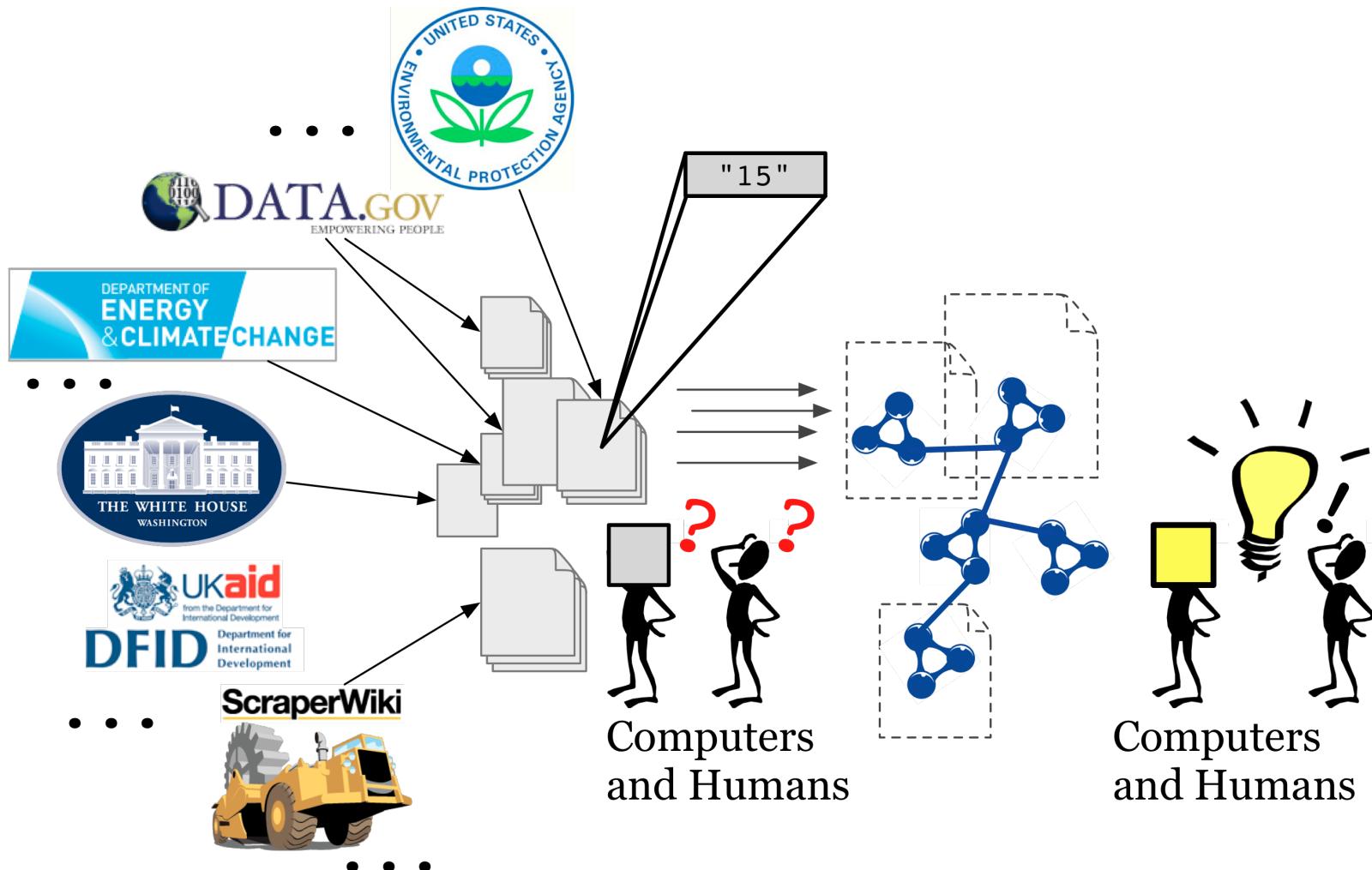


# Motivation



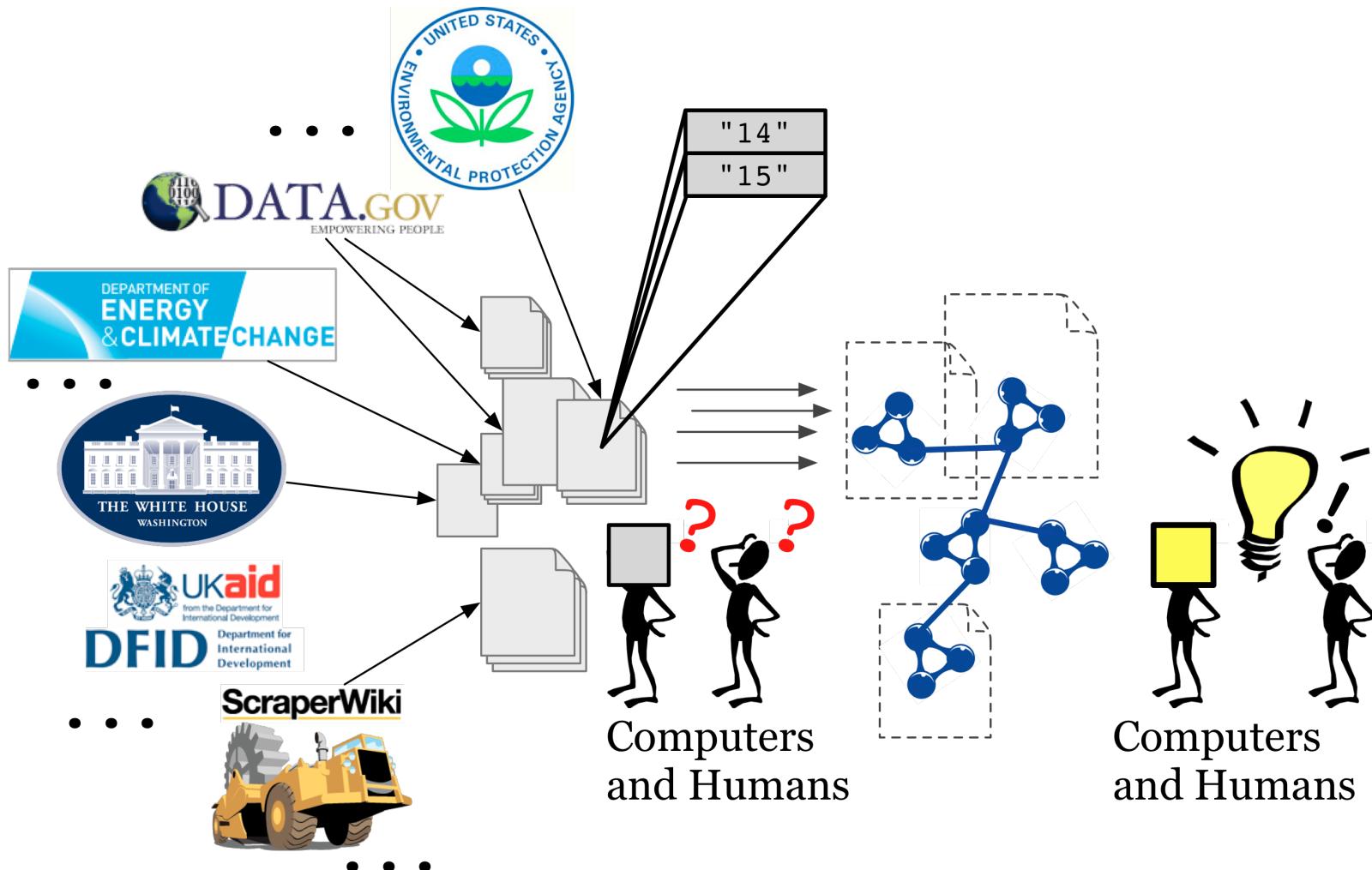


ITWC



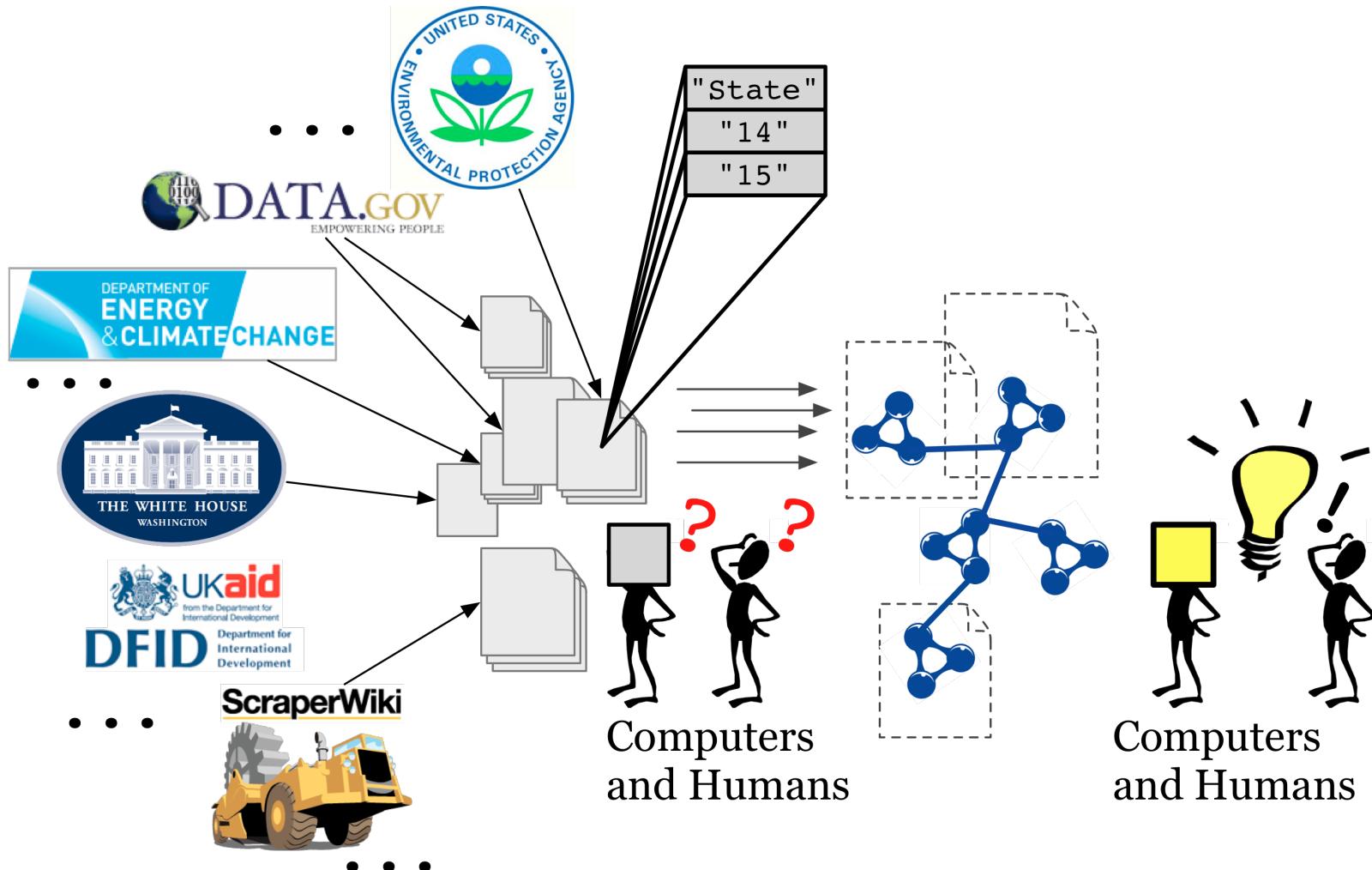


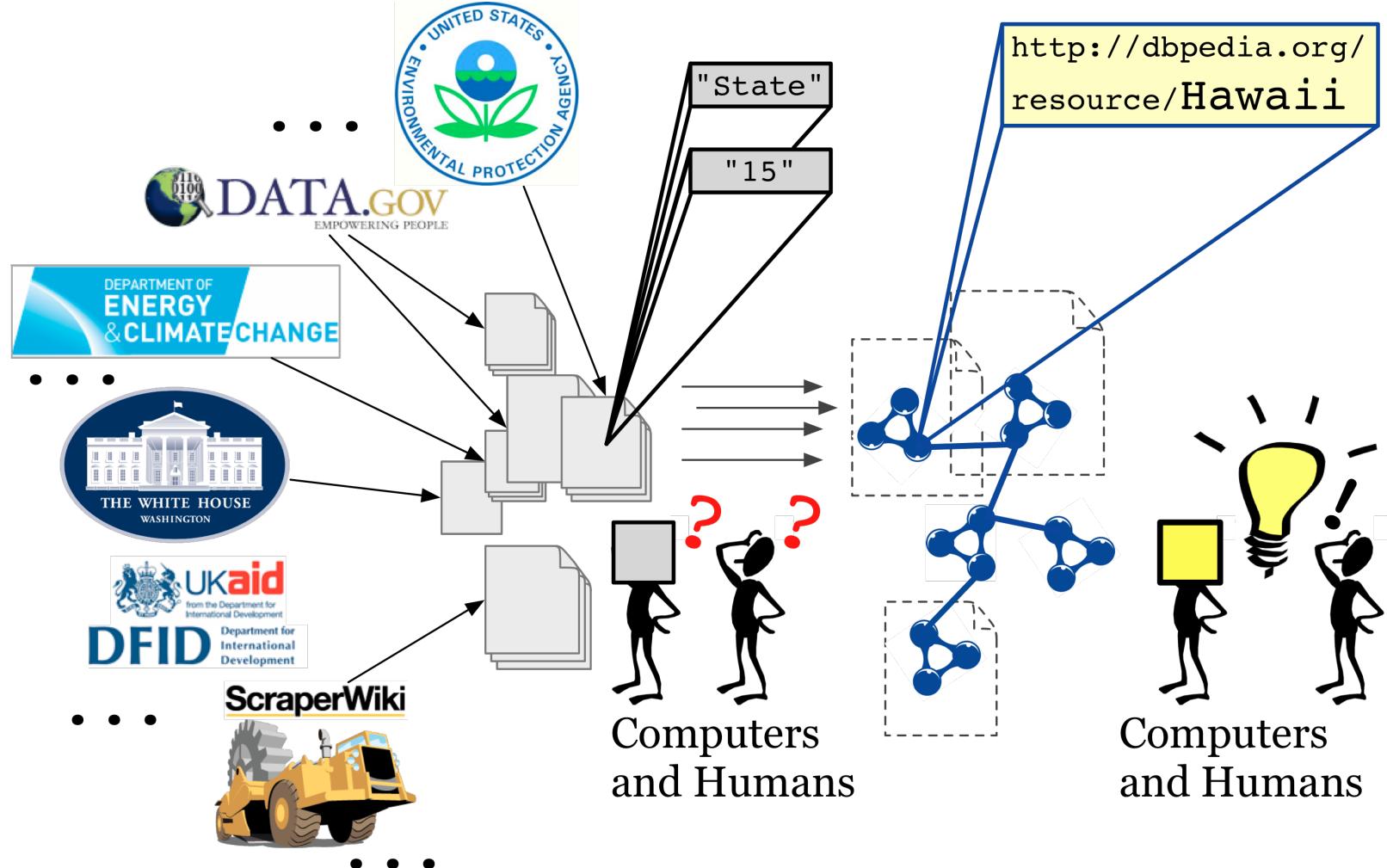
ITWC





ITAWC







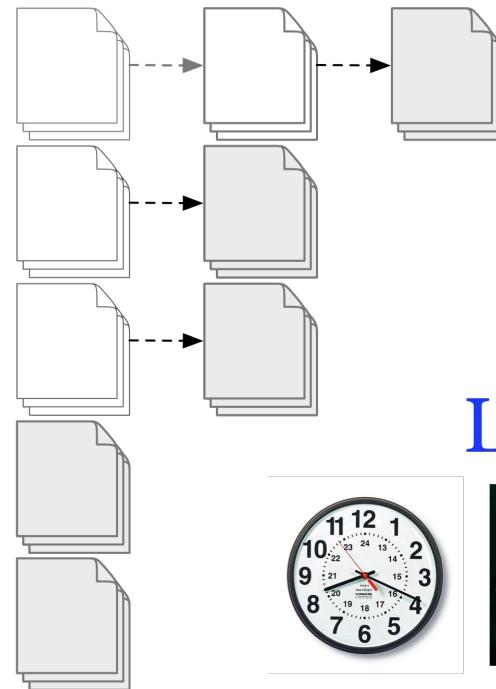
# Economy of Resources

Lots of:

Sources

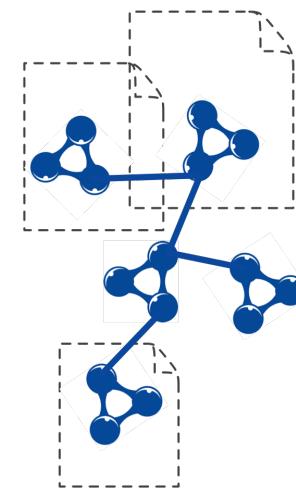
Datasets

Revisions

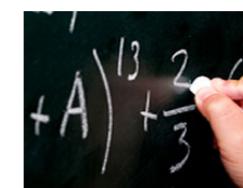


Time

But want:



Low on:



Domain  
Expertise



SemWeb  
Expertise



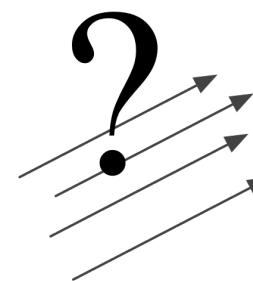
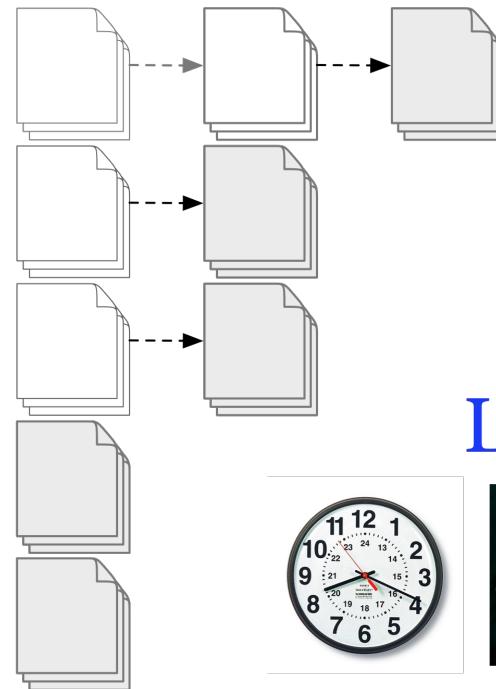
# Economy of Resources

Lots of:

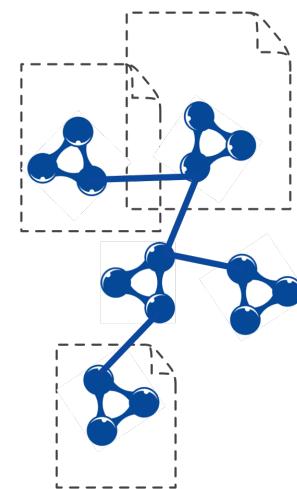
Sources

Datasets

Revisions



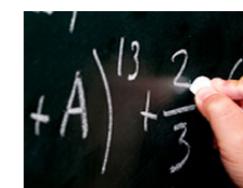
But want:



Low on:



Time



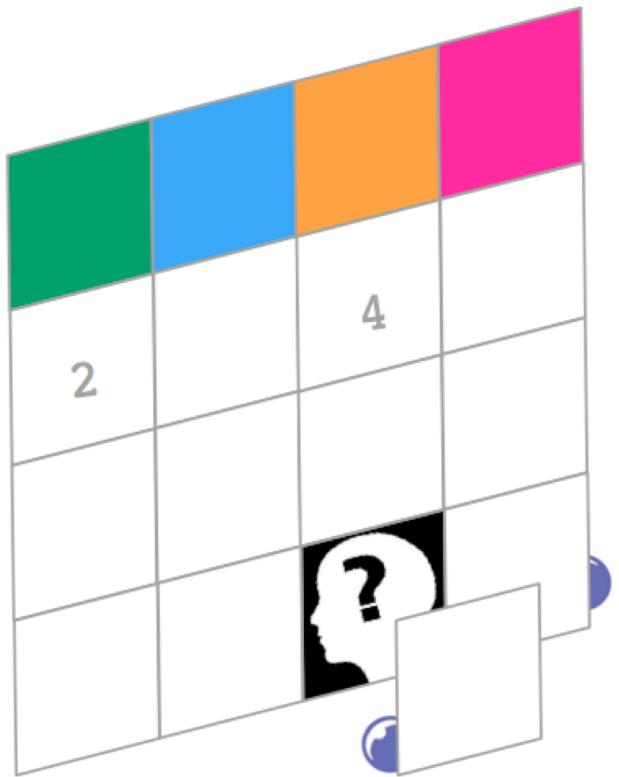
Domain  
Expertise



SemWeb  
Expertise



# csv2rdf4lod-automation



timrdf / csv2rdf4lod-automation

Source   Commits   Network   Pull Requests (1)

Switch Branches (1)   Switch Tags (0)   Branch List



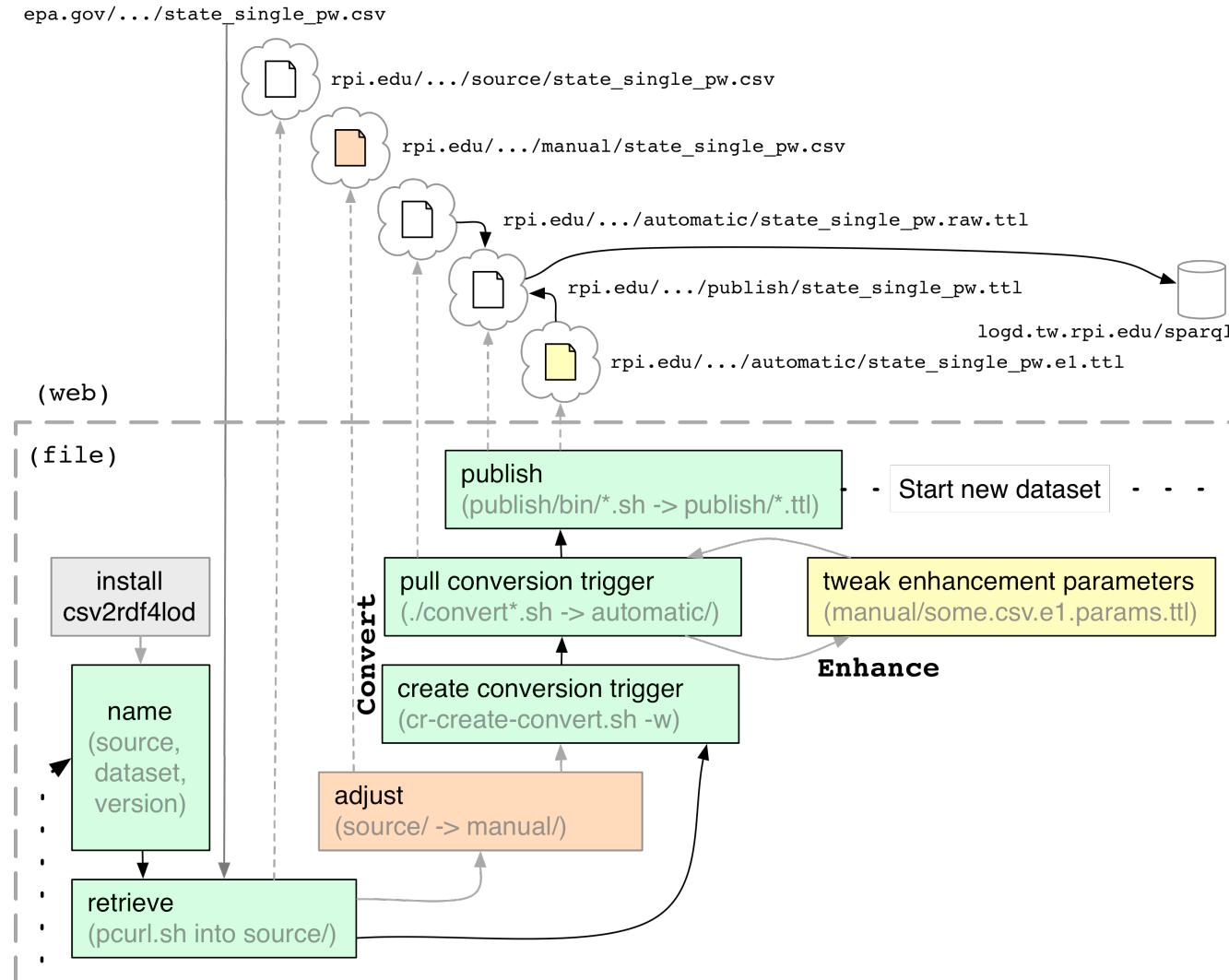
```
$ ./retrieve.sh  
$ vi manual/data.e1.params.ttl  
$ ./convert-foreign-aid.sh  
$ publish/bin/publish.sh
```

<http://purl.org/twc/id/software/csv2rdf4lod> a doap:Software .

<http://purl.org/twc/vocab/conversion/Enhancement> a owl:Class .



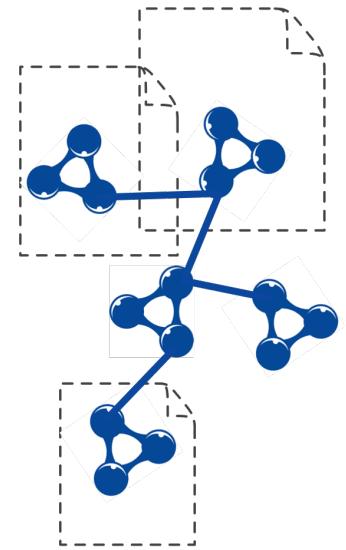
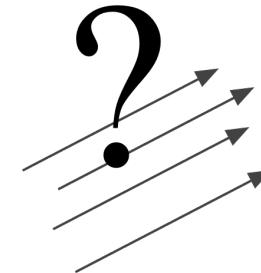
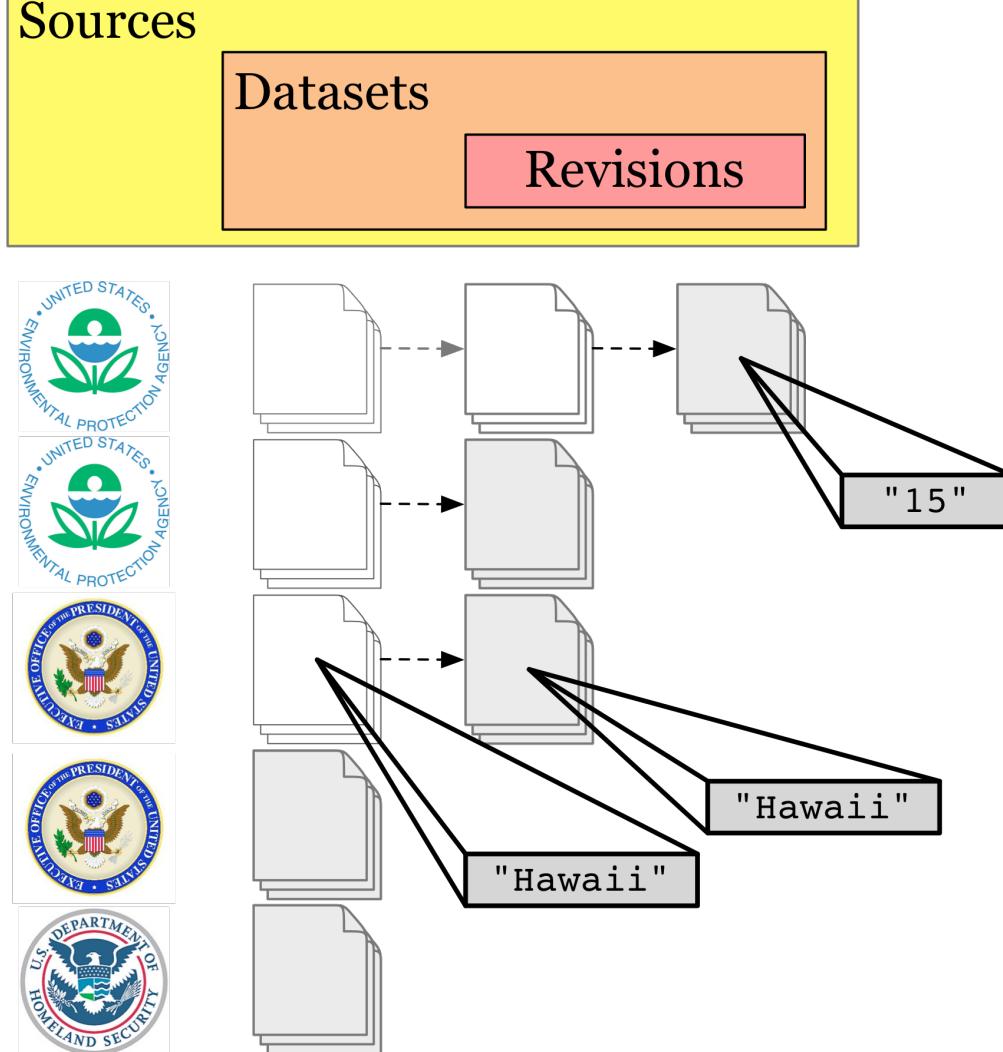
# csv2rdf4lod's Integration Stages





# A smart, naïve bootstrap

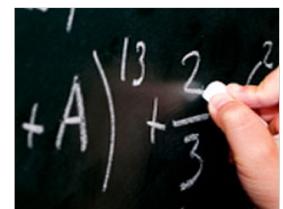
Sources



Low on:



Time



Domain  
Expertise

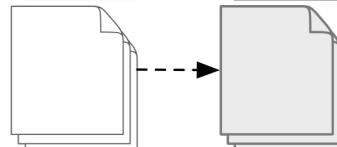
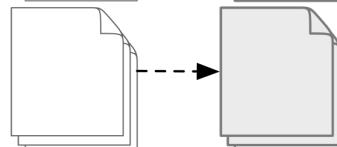
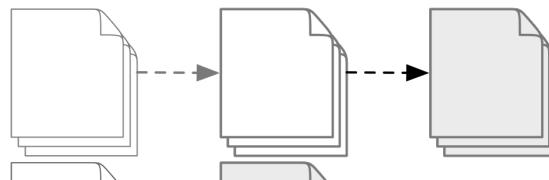


# Contextualizing Data

Sources

Datasets

Revisions



who?	source:	"data-gov"
what?	dataset:	"1095"
when?	version:	"2011-Aug-02"

who?	source:	"data-gov"
what?	dataset:	"4383"
when?	version:	"2011-Sep-27"



# Contexts on the Web

## Contextual Depth Entity URIs

1	http://dbpedia.org/resource/Hawaii
0	http://lodt.rpi.edu /id/us/state/ Hawaii
-1	http://lodt.rpi.edu /source/ data-gov / Hawaii
-2	http://lodt.rpi.edu /source/ data-gov /dataset/ 1095 /typed/state/ HI
-2	http://lodt.rpi.edu /source/ data-gov /dataset/ 1095 /value-of/state_name/ HI
-3	http://lodt.rpi.edu /source/ data-gov /dataset/ 1095 /version/ 2011-Aug-02 / thing_1

Global Context  
Local Context



"Who?"  
"What?"  
"When?"  
-->

who?	source:	"data-gov"
what?	dataset:	"1095"
when?	version:	"2011-Aug-02"

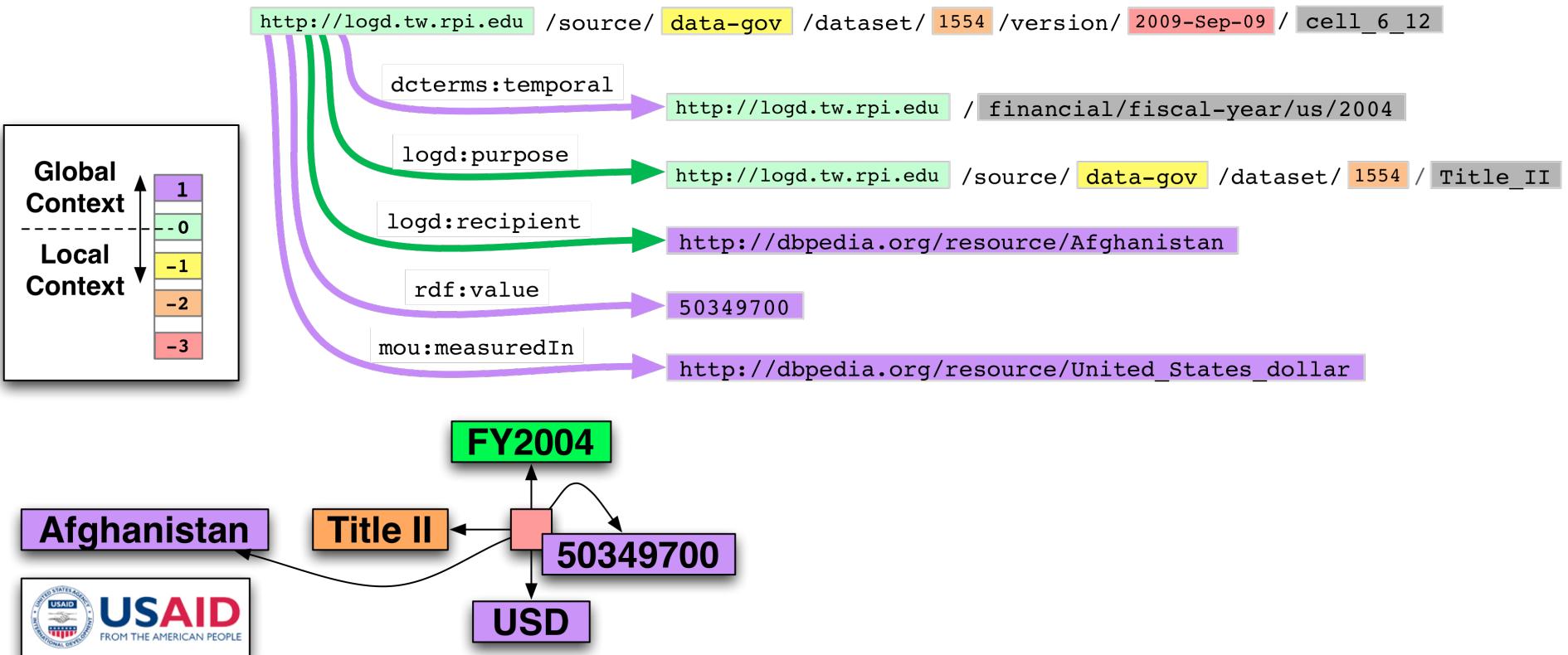


# Using Multiple Contextual Depths



country_name	program_name	FY2004	FY2005
Afghanistan	Title II	50349700	57233800
Albania	Child Survival and Health	1000000	150000

United States Foreign Aid





# Adding Enhancements

```
"Hawaii", "Alii Garden Market Place",
"75-6129 Alii Drive", "Kailua-Kona",
"96740", "-155.9819183", "19.61436844"
```



```
ds4383:thing_1376
con:preferredURI
ds4383:farmersMarket_1376 .
```

```
ds4383:farmersMarket_1376
a ds4383_vocab:FarmersMarket;
con:address :address_1376;
dcterms:title "Get Fresh Maui";
wgs:lat -155.9; wgs:long 19.6 .
```

```
:address_1376
a con:Address;
con:stateOrProvince typed_state:Hawaii;
con:street "Honopiilani HWY";
con:city "Maui";
con:zip "96761" .
```

```
ds4383:thing_1367
raw:column_1 "Hawaii";
raw:column_2 "Alii Garden Market Place";
raw:column_3 "75-6129 Alii Drive";
raw:column_4 "Kailua-Kona";
raw:column_5 "96740";
raw:column_6 "-155.9819183";
raw:column_7 "19.61436844" .
```

```
typed_state:Hawaii
a ds4383_vocab:State;
dcterms:identifier "Hawaii";
rdfs:label "Hawaii";
owl:sameAs
<http://sws.geonames.org/5855797/>,
govtrackusgov:HI,
dbpedia:Hawaii .
```



# Example Enhancement Parameters

```
<http://lodg.tw.rpi.edu/source/data-gov/dataset/4383/version/2010-Oct-22/
  a conversion:LayerDataset, void:Dataset;

  conversion:base_uri          "http://lodg.tw.rpi.edu"^^xsd:anyURI;
  conversion:source_identifier  "data-gov";
  conversion:dataset_identifier "4383";
  conversion:version_identifier "2010-Oct-22";

  conversion:conversion_process [
    a conversion:EnhancementConversionProcess;
    conversion:enhancement_identifier "1";

    conversion:enhance [
      ov:csvCol           1;
      ov:csvHeader        "LocAddState";
      conversion:bundled_by :address_bundle;
      conversion:domain_name "FarmersMarket";
      conversion:equivalent_property con:stateOrProvince;
      conversion:comment "";
      conversion:range     rdfs:Resource;
      conversion:range_name "State";
      conversion:links_via <http://www.rpi.edu/~lebot/lod-links/state-fips-dbpedia.ttl>,
                            <http://www.rpi.edu/~lebot/lod-links/state-fips-geonames.ttl>,
                            <http://www.rpi.edu/~lebot/lod-links/state-fips-govtrack.ttl>;
      conversion:subject_of dcterms:identifier;
    ];
  ];
}
```



# Affecting Subjects

- Enhancements that affect the subject of the triple produced:
  - a conversion:{[Omitted](#), [Only\\_if\\_column](#), [DataStartRow](#)}
  - a conversion:{[ExampleResource](#), [SubjectAnnotation](#)}
  - [conversion:bundled\\_by](#)
  - [conversion:domain\\_template](#)
  - [conversion:domain\\_name](#)
  - [conversion:predicate](#)
  - [conversion:object](#)
  - [conversion:object\\_search](#)

<http://purl.org/twc/vocab/conversion/Enhancement>

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



# Affecting Predicates

- Enhancements that affect the predicate of the triple produced:
  - [conversion:equivalent\\_property\\_subproperty\\_of](#)
  - [conversion:label](#)
  - [conversion:comment](#)
  - [conversion:subproperty\\_of](#)

<http://purl.org/twc/vocab/conversion/Enhancement>

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



# Affecting Objects

- Enhancements that affect the object of the triple produced:
  - [conversion:eg](#)
  - a [conversion:{Repeat\\_previous\\_if\\_empty\\_column, LargeValue}](#)
  - [conversion:repeat\\_previous](#) -
  - [conversion:interpret](#) -
  - [conversion:{date\\_pattern, datetime\\_pattern}](#)
  - [conversion:delimits\\_object](#)
  - [conversion:object](#)
  - [conversion:range\\_template](#)

<http://purl.org/twc/vocab/conversion/Enhancement>

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



# Affecting Objects

- Enhancements that affect the descriptions of the object of the triple produced:
  - [conversion:range](#) - {rdfs:Literal, rdfs:Resource, xsd:integer, xsd:decimal}
  - [conversion:multiplier](#)
  - [conversion:range\\_name](#)
  - [conversion:links\\_via](#)
  - [conversion:subject\\_of](#)
  - [conversion:predicate](#)
  - [conversion:object](#)

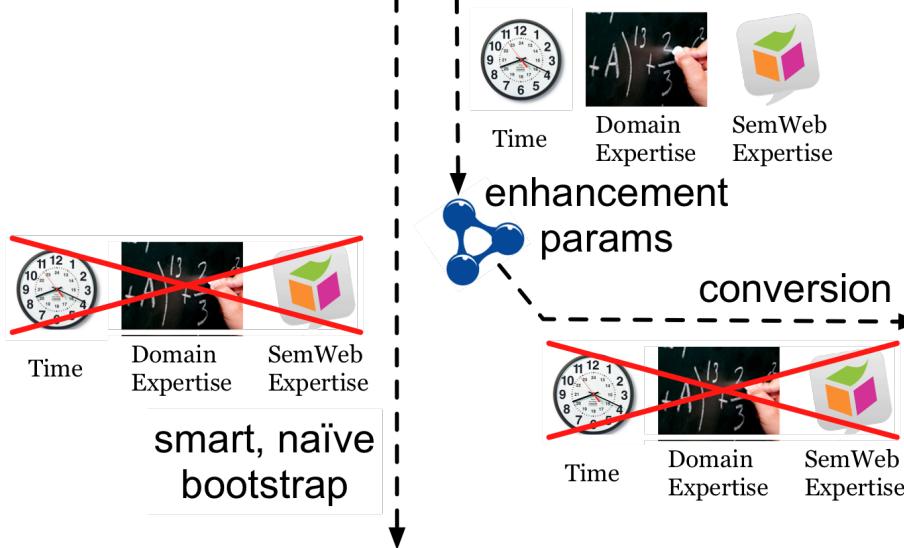
<http://purl.org/twc/vocab/conversion/Enhancement>

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



# One-time cost

```
"Hawaii", "Alii Garden Market Place",
"75-6129 Alii Drive", "Kailua-Kona",
"96740", "-155.9819183", "19.61436844"
```



```
ds4383:thing_1367
  raw:column_1 "Hawaii";
  raw:column_2 "Alii Garden Market Place";
  raw:column_3 "75-6129 Alii Drive";
  raw:column_4 "Kailua-Kona";
  raw:column_5 "96740";
  raw:column_6 "-155.9819183";
  raw:column_7 "19.61436844" .
```

```
ds4383:thing_1376
  con:preferredURI
ds4383:farmersMarket_1376 .
```

```
ds4383:farmersMarket_1376
  a ds4383_vocab:FarmersMarket;
  con:address :address_1376;
  dcterms:title "Get Fresh Maui";
  wgs:lat -155.9; wgs:long 19.6 .
```

```
:address_1376
  a con:Address;
  con:stateOrProvince typed_state:Hawaii;
  con:street          "Honopilani HWY";
  con:city            "Maui";
  con:zip             "96761" .
```

```
typed_state:Hawaii
  a ds4383_vocab:State;
  dcterms:identifier "Hawaii";
  rdfs:label          "Hawaii";
  owl:sameAs
    <http://sws.geonames.org/5855797/>;
    govtrackusgov:HI,
    dbpedia:Hawaii .
```