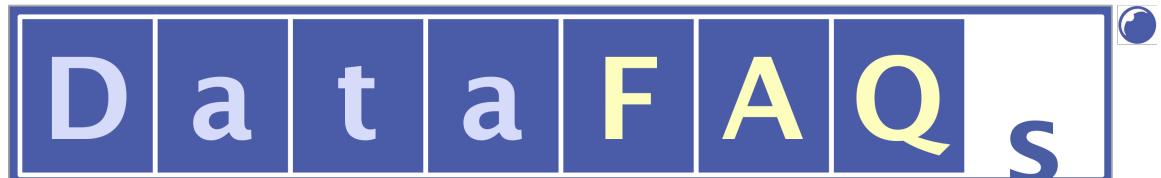




↔ <http://bit.ly/lebo-twed-2012>



LINKED DATA QUALITY REPORTS

Because 5 stars aren't enough.



Timothy Lebo
Tetherless World Constellation
Rensselaer Polytechnic Institute



Rensselaer

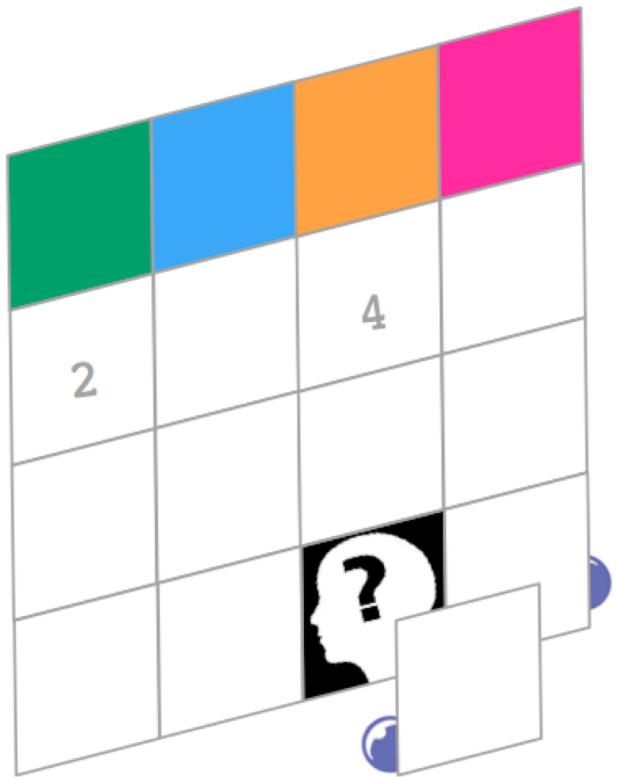


Outline

- Motivation
 - Linked Data production
 - Linked Data monitoring
 - But...
- DataFAQs framework
 - Choose Datasets (CKAN)
 - Choose Evaluation Services (SADI)
 - Explore Evaluations (LODSPeaKr)



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 .



... what other tools ...

Market positioning *

Describe what standards does the tool support

what other tools are in the same market.

decision concerning the tools in the LATC library

* Required
Tool name *

csv2rdf4lod-automatic

?

! ● ●

http://arxiv.org/abs/1202.3667

WDT2

IO informatics Knowledge Explorer

Virtuoso sponger

rdflib's csv2rdf

RDFTOnto

T2LD

S3DB

Talis Tables

Pentaho Data Integration suite

RDF23

EasyRdf

Mapping Master

Anzo

Pubby

Information Workbench

DataLift

H2R

ALOE

XLWrap

M2

Excel2RDF

stuff2rdf

OPPL

Populous

LDIF

Google Refine

Triplify

Stats2RDF

VIVO Harvester

Spotify

OPPL

LDIF

Google Refine

Triplify

Stats2RDF

stuff2rdf

Grinder

Consumption Tool Library

http://latc-project.eu/

Author(s): Guido Cecilio, Stephane Corlosquet, Richard Cyganiak
Website: http://neologism.deri.ie/
License: Gnu General Public License
User Interface: Graphic / Drupal Module

Author(s): Chris Bizer, Richard Cyganiak
Website: http://www4.wiwi.fu-berlin.de/bizer/d2r-server/
License: Apache License, Version 2.0
User Interface: Publishing

Category: Publishing
Author(s): Fadi Mardi, Richard Cyganiak
Website: http://linkeddata.deri.ie/2010/grefine-rdf-extension/
License: BSD License
User Interface: Graphic

Category: Publishing
Author(s): Richard Cyganiak, Chris Bizer
Website: http://www4.wiwi.fu-berlin.de/pubby/
License: Apache License, Version 2.0
User Interface: Command Line



LOD2's Tool Survey

<http://static.lod2.eu/Deliverables/deliverable-3.1.1.pdf>



LOD2

Collaborative Project

LOD2 - Creating Knowledge out of Interlinked Data

Project Number: 257943 Start Date of Project: 01/09/2010

Deliverable 3.1.1 Report on Knowledge Extraction from Structured Sources

Dissemination Level	Public
Due Date of Deliverable	Month 6, 28/2/2011
Actual Submission Date	15/03/2011

VIVO Harvester

OPPL

Spotfire

LDIF

Refine

Triplify

Populous

Stats2RDF

stuff2rdf

el2RDF

on Workbench

DataLift

Grinder



Wrong focus?



Deliverable 3.1.1 Report on Knowledge Extraction from Structured Sources

- Data source
 - CSV, XML, RDB
 - Data Exposition
 - ETL, SPARQL
 - Data Synchronization
 - Static, dynamic, bi-directional
 - Mapping Language
 - XSLT, D2R Map, Visual tool
 - Vocab reuse
 - Yeh, Ney
 - Mapping automation
 - Manual, semi-auto, GUI
- Requires domain ontology
- Yeh, Ney
- Uses GUI
- Yeh, Ney
- How:
How well?
Why?
Publisher
Consumer



How well? (1/3)



SPARQL Endpoints Status

<http://labs.mondeca.com/sparqlEndpointsStatus/>

252 / 362 (69,61%) Public SPARQL Endpoints available / total

KEGG Pathway	75%	76,67%
KEGG Reaction	75%	60%
Kasabi European Election Results	75%	96,67%
Korean Institute of Science Technology and Information (RKBExplorer)	100%	91,67%
LAAS-CNRS (RKBExplorer)	100%	93,33%
LATC Linksets	0%	0%



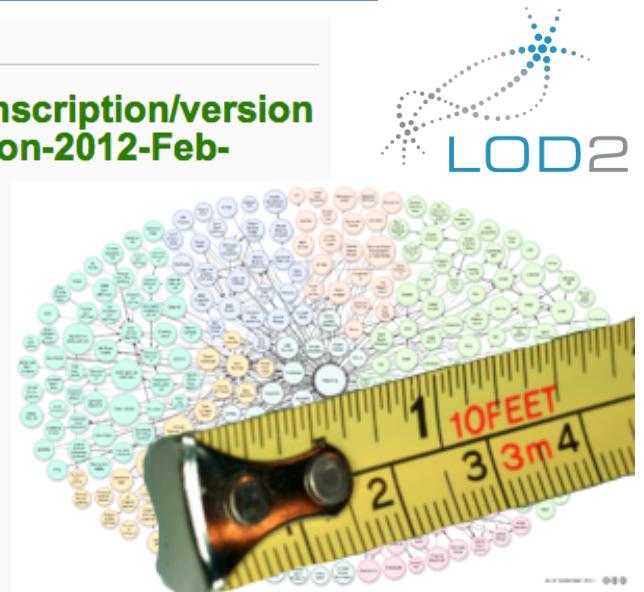
How well? (2/3) LODStats

w3c-provenance-incubator-group-vocabulary-mappings

- <http://logd.tw.rpi.edu/source/twc-rpi-edu/file/prov-xg-mappings-transcription/version/2012-Feb-29/conversion/twc-rpi-edu-prov-xg-mappings-transcription-2012-Feb-29.ttl.gz>
- Last updated: 2012-03-07 10:29:32.103050
- Is being worked on: False
- File last modified: 2012-02-29 17:28:57
- Results last updated: 2012-03-07 10:29:32.103073
- Bytes: 11.8 KiB - 171.5 KiB decompressed

Basic Statistics

Criterion	Value
Triples	2913
Entities	312
Literals	1932
Classes (toggle details)	26
Defined classes	0
Vocabularies (toggle details)	21
Properties (toggle details)	115
Datatypes (toggle details)	3





How well? (3/3) LOV

labs.mondeca.com/dataset/lov



MONDECA

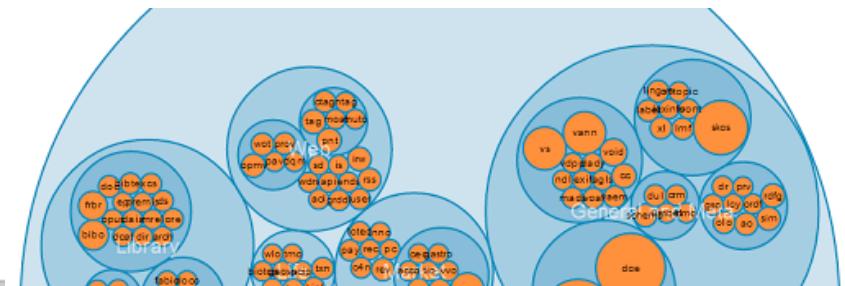


Inserm



Linked Open Vocabularies (LOV)

3747 vocabulary elements used:



LOV distribution	LOV popularity	LOD popularity	Vocabulary Prefix	Element URI
103	307	23087198	dce	http://purl.org/dc/elements/1.1/title
95	592	19346165	dce	http://purl.org/dc/elements/1.1/creator
72	543	9086606	dce	http://purl.org/dc/elements/1.1/description
58	424	25367750	dce	http://purl.org/dc/elements/1.1/date
53	121	13468915	foaf	http://xmlns.com/foaf/0.1/name



Where are we with LOD?

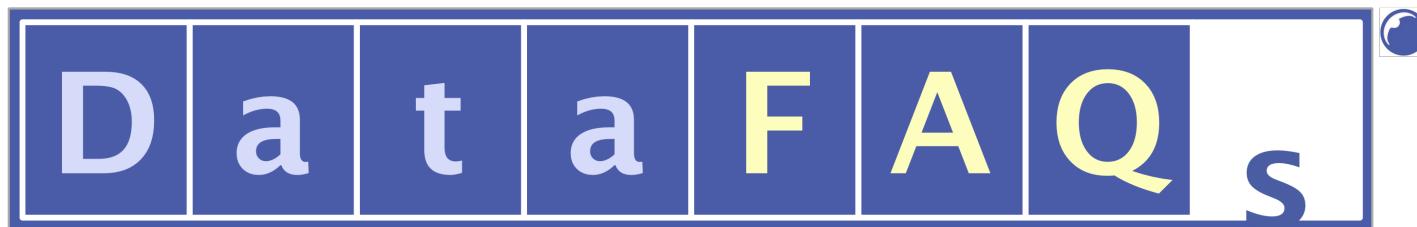
- ✓ • Publishing? 30+ tools
- ~ • Monitoring? Endpoint Status, LODStats, LOV
 - too heavyweight and too centralized
 - **Where's the little guy?** (*... but I need to know x ...*)

		Linked Data	
		Social	Technical
		d	c
Consumer		✗	~
Publisher		b ~	a ✓

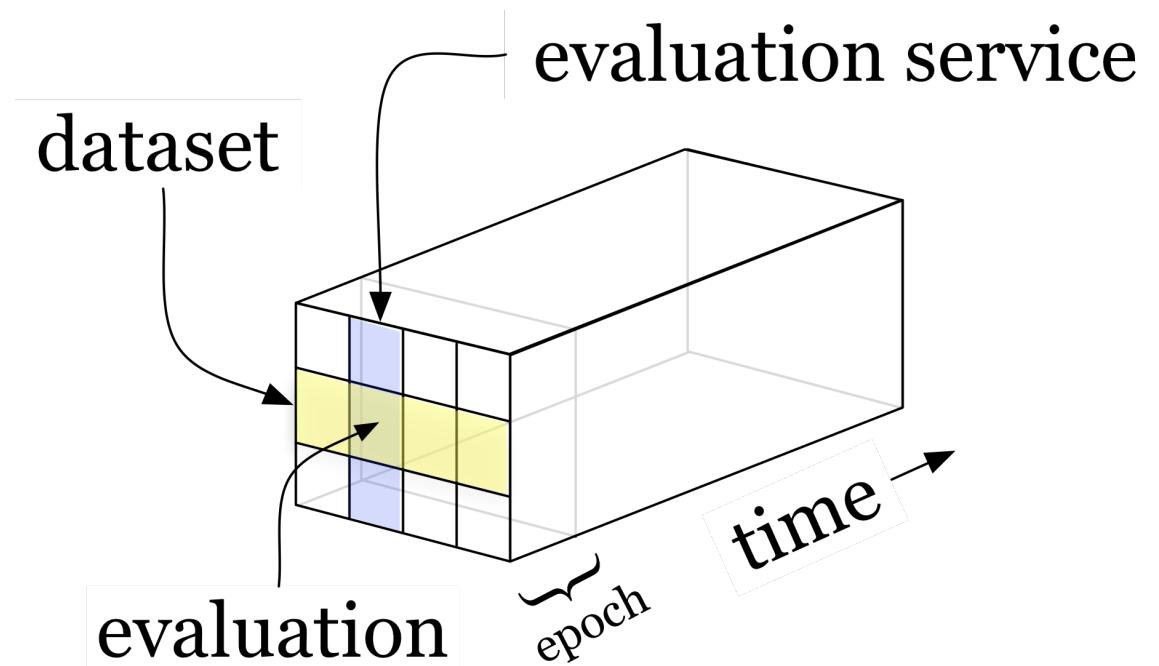
~~System~~
Framework



DataFAQs essentials

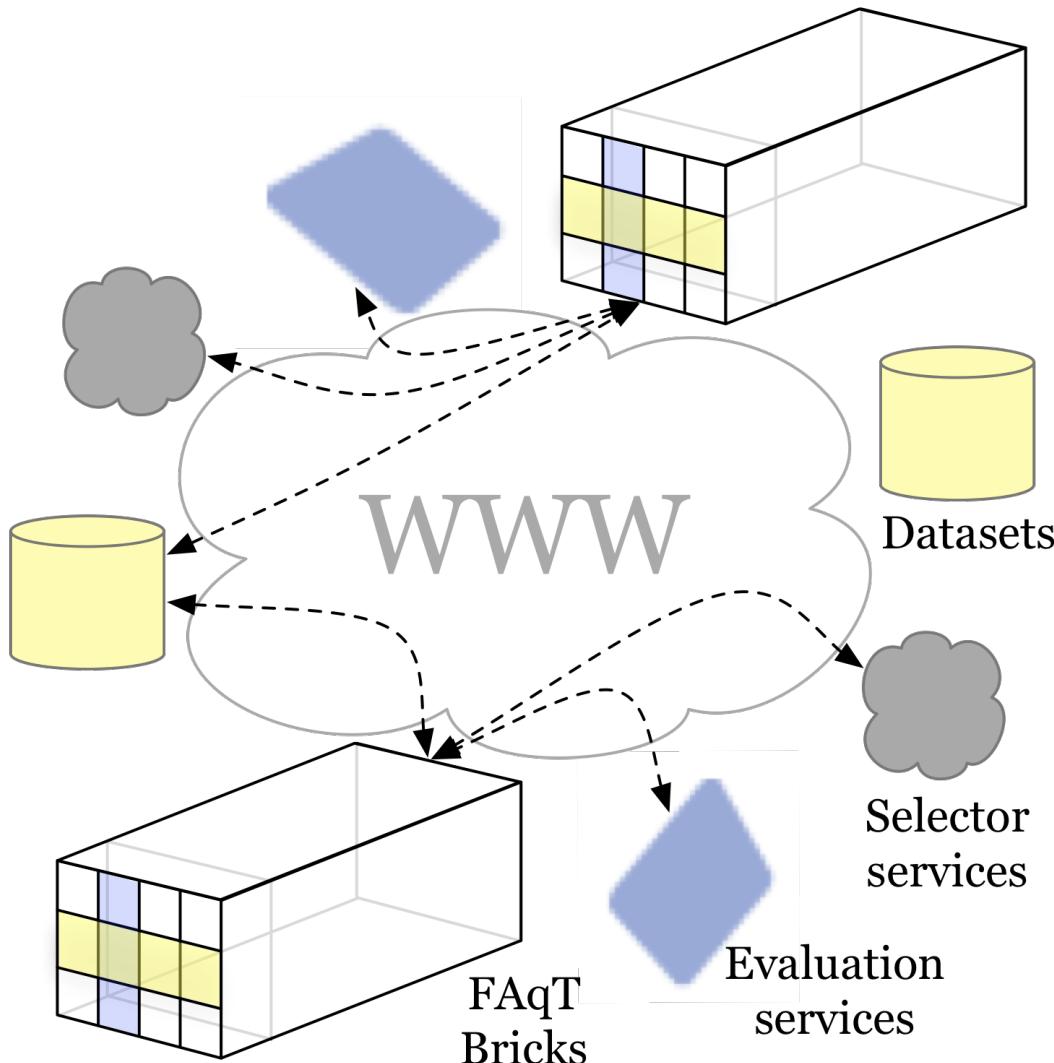


LINKED DATA QUALITY REPORTS





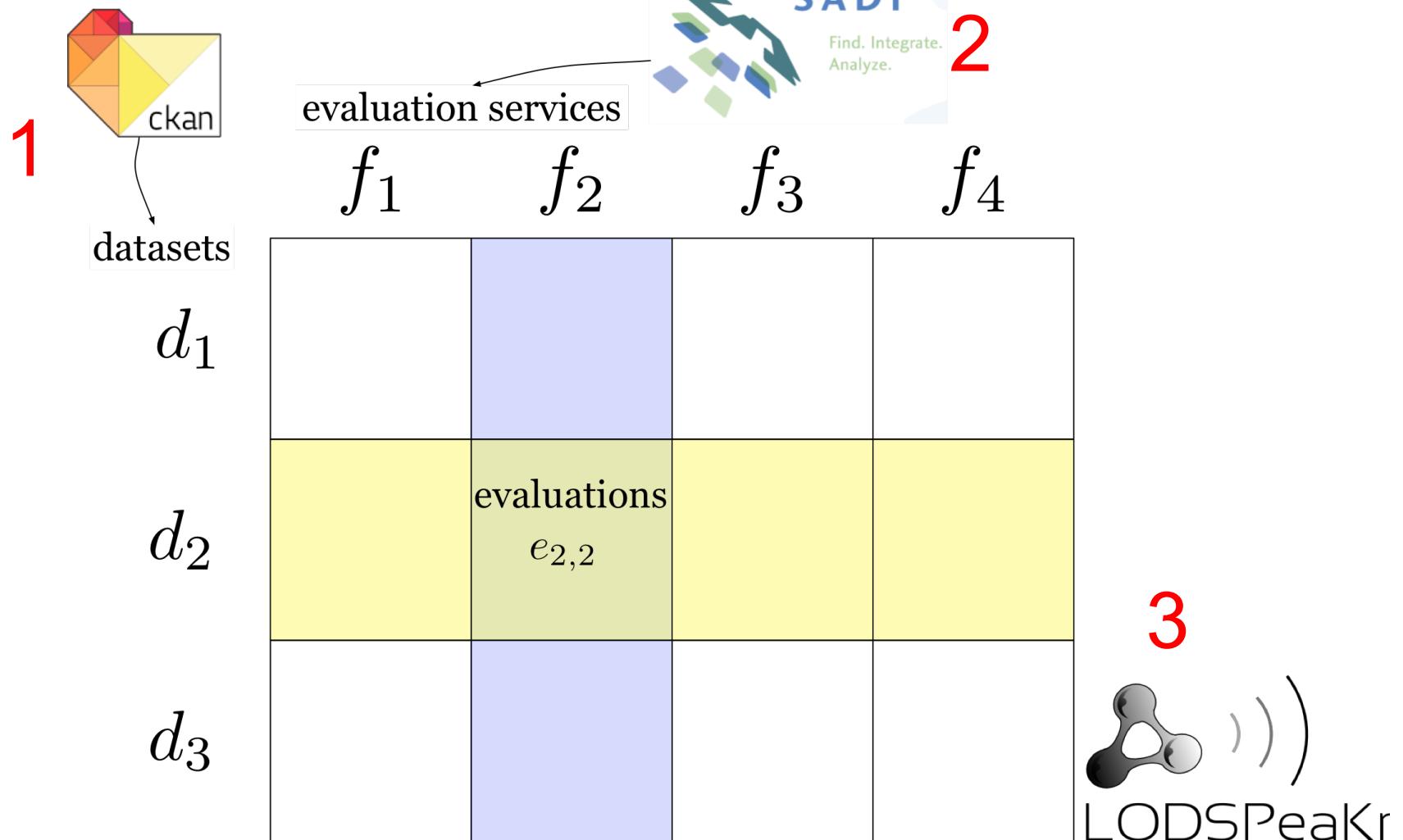
DataFAQs is distributed



- You create and maintain your own **FAqT Brick**
- You evaluate others' datasets
- You use others' evaluation services



The Tech behind DataFAQs





1 thedatahub.org/group/lodcloud

Edit : Groups & Tags : Tags

Edit: Farmers Markets Geographic Data (United States)

View Edit History Authorization

Basic Information

Futher Information

Resources

Groups & Tags

Extras

Groups

Tags

Group: (None)

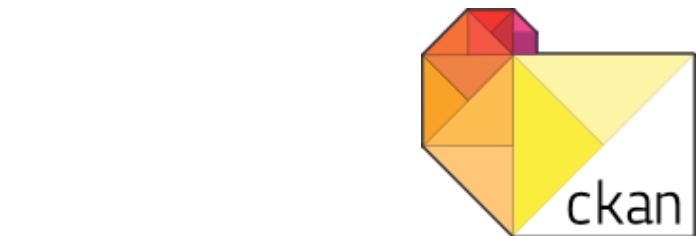
Tags: geographic lod

this dataset is similar to:

media
geographic
lifesciences
publications
government
ecommerce
socialweb
usergeneratedcontent
schemata
crossdomain



published-by-third-party	published-by-producer
no-license-metadata	license-metadata
no-provenance-metadata	provenance-metadata
no-vocab-mappings	vocab-mappings
no-deref-vocab	deref-vocab
no-proprietary-vocab	



- GUI or JSON API in
- SPARQL and deref out

Edit : Extras : New key / with value

Edit: Farmers Markets Geographic Data (United States)

View Edit History Authorization

Basic Information

Futher Information

Resources

Groups & Tags

Extras

New key: with value

New key: with value

New key: with value

triples	130005
shortname	Farmers Markets (US)
license_link	
sparql_graph_name	http://lodg.tw.rpi...
namespace	http://lodg.tw.rpi...



2 SioS (SADI in one Slide)

<http://sadiframework.org/registry/register/>

```
raptor -g -o turtle http://sadiframework.org/examples/hello
```

```
...
mygrid:inputParameter [
    mygrid:objectType
        <http://sadiframework.org/examples/hello.owl#NamedIndividual>;
    ...
];
mygrid:outputParameter [
    mygrid:objectType
        <http://sadiframework.org/examples/hello.owl#GreetedIndividual>;
    ...
];
...
```



```
... # send.ttl
<http://purl.org/twc/instances/TimLebo>
  a hello:NamedIndividual;
  foaf:name "Tim" .
```

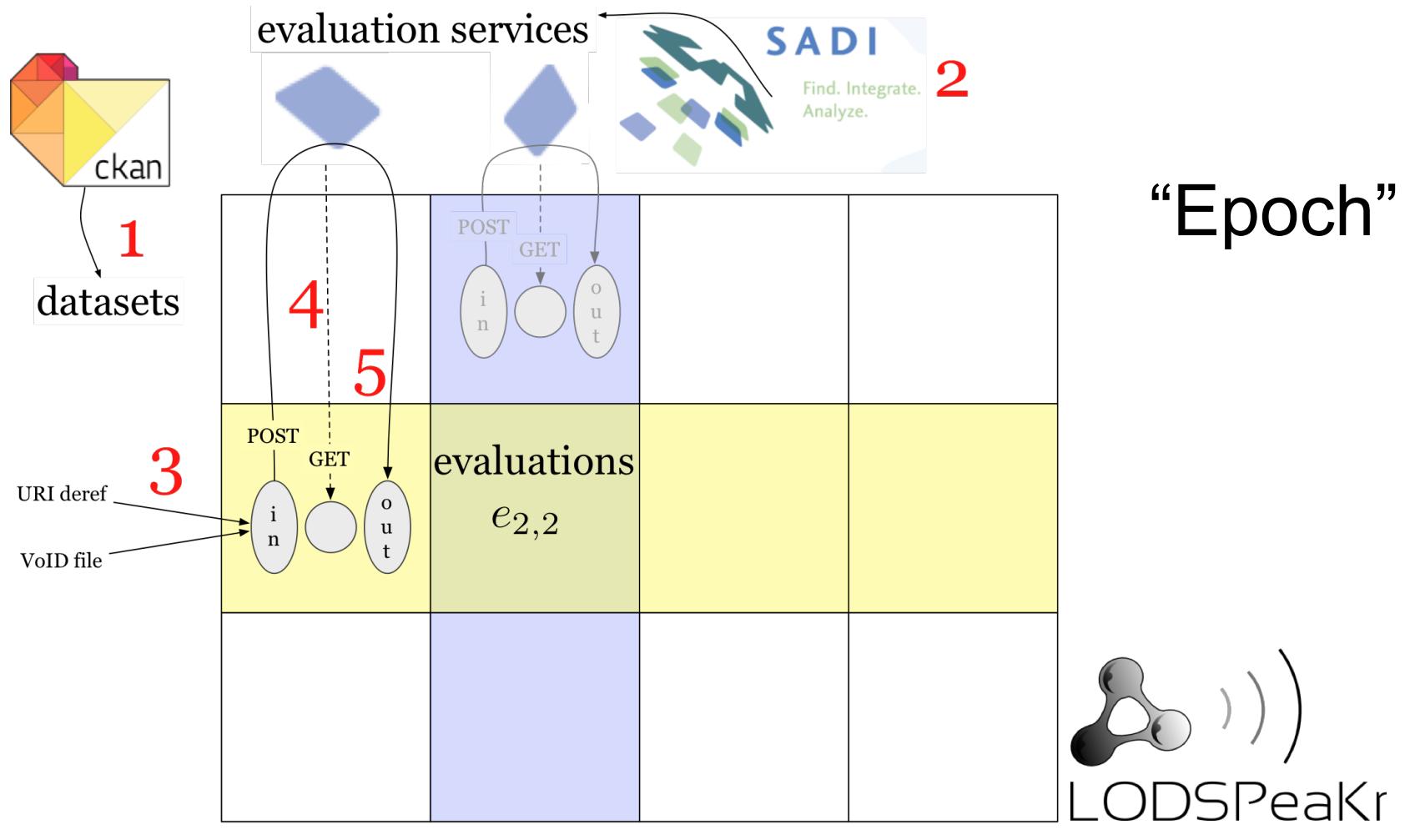
```
curl -d @send.ttl.rdf http://sadiframework.org/examples/hello
```

```
<rdf:RDF    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
               xmlns:hello="http://sadiframework.org/examples/hello.owl#">

  <hello:GreetedIndividual rdf:about="http://purl.org/twc/instances/TimLebo">
    <hello:greeting rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
      Hello, Tim!
    </hello:greeting>
  </hello:GreetedIndividual>
</rdf:RDF>
```



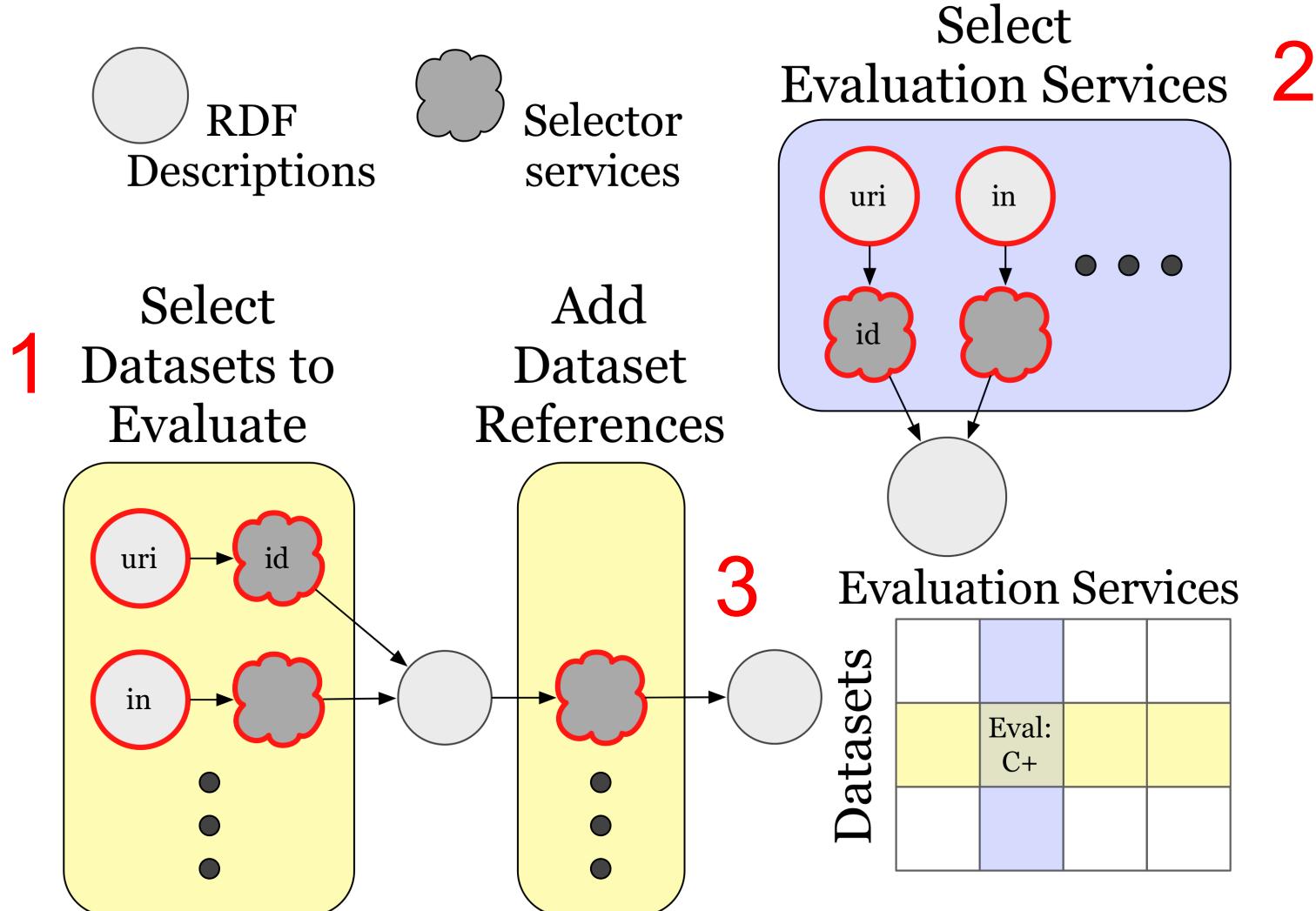
The Tech behind DataFAQs



<https://github.com/timrdf/DataFAQs/wiki/FAqT-Brick>



Configuring an Epoch (with W3C PROV)





DataFAQs Core Services

- <https://github.com/timrdf/DataFAQs/wiki/DataFAQs-Core-Services>
- **select-faqts**
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-faqts/identity**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-faqts/identity)
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-faqts/via-sparql-query**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-faqts/via-sparql-query)
- **select-datasets**
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-datasets/identity**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-datasets/identity)
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-datasets/via-sparql-query**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-datasets/via-sparql-query)
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-datasets/by-ckan-group**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-datasets/by-ckan-group)
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**select-datasets/by-ckan-tag**](http://sparql.tw.rpi.edu/services/datafaqs/core/select-datasets/by-ckan-tag)
- **augment-datasets**
 - [http://sparql.tw.rpi.edu/services/datafaqs/core/**augment-datasets/with-preferred-uri-and-ckan-meta-void**](http://sparql.tw.rpi.edu/services/datafaqs/core/augment-datasets/with-preferred-uri-and-ckan-meta-void)



FAQT Services

- <https://github.com/timrdf/DataFAQs/tree/master/services/sadi/faqt>
- faqt/access/in-sparql-endpoint.rpy
- faqt/sparql-service-description/named-graphs.rpy
- faqt/discuss/contributor-email.rpy
- faqt/lodcloud/max-1-topic-tag.rpy
- faqt/vocabulary/vocabulary-resolves-to-description.rpy
- faqt/redirect-loop.rpy



Running an Epoch

- <https://github.com/timrdf/DataFAQs/blob/master/bin/df-epoch.sh>
 - Follows epoch configuration epoch.ttl
 - Accumulates results to file system
 - Annotates results with metadata
 - Determines named graph organization
 - Publishes dump files to web
 - Loads results to tdb or Virtuoso triple store



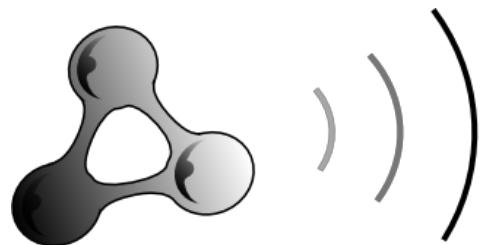
Easy access to dump files

```
PREFIX void: <http://rdfs.org/ns/void#>
PREFIX sd:   <http://www.w3.org/ns/sparql-service-description#>

select distinct ?name ?triples ?dump
where {
    graph ?name {
        [] a []
    }
    optional {
        graph ?meta {
            [] sd:name ?name;
            sd:graph ?graph .
            ?graph void:dataDump ?dump .
        optional {
            ?graph void:triples ?triples
        }
    }
}
}
```



LODSPeaKr



“When in doubt, fork LODSPeaKr!”

LODSPeaKr

- components/types/datafaqs:Epoch
- components/types/datafaqs:Epoch/html.template
- components/types/datafaqs:Epoch/queries
- components/types/datafaqs:Epoch/queries/configurations.query
- components/types/datafaqs:Epoch/queries/sp.query
- components/types/datafaqs:Epoch/queries/subsets.query
- components/types/datafaqs:Epoch/queries/po.query
- components/types/datafaqs:Epoch/queries/evaluations.query
- components/types/datafaqs:Epoch/queries/faqt_service_self_descriptions.query
- components/types/datafaqs:Epoch/queries/dataset_descriptions.query



What's up so far

<http://aquarius.tw.rpi.edu/projects/datafaqs/home>



Conclusions

- Lots of tools to produce Linked Data
- Some tools to evaluate Linked Data
 - But too heavyweight and too centralized
 - ~~How~~ **How well? Why?**
 - ~~Publisher~~ **Consumer**
 - ~~System~~ **Framework**
- Need a framework for micro-contribs (**DataFAQs**)
 - Choose Datasets (CKAN)
 - Choose Evaluation Services (SADI)
 - Explore Evaluations (LODSPeaKr)



Wanna try?



@timrdf

<http://bit.ly/try-datafaqs>

<https://github.com/timrdf/DataFAQs/wiki/Getting-started>



Thanks!

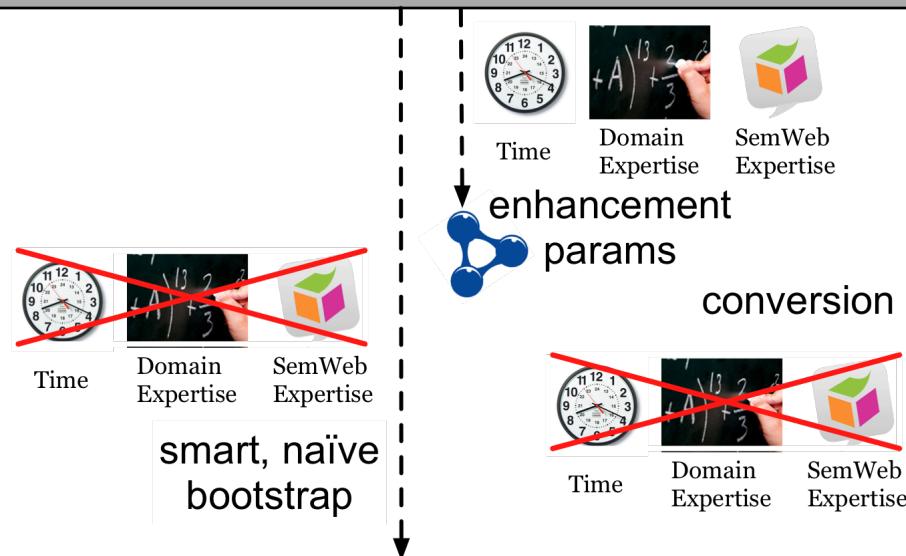


- Yu Chen
- Jim McCusker
- Josh Shinavier
- Alvaro Graves



Raw versus Enhanced

```
"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_1367
  con:preferredURI
ds4383:farmersMarket_1367 .
```

```
ds4383:farmersMarket_1367
  a ds4383_vocab:FarmersMarket;
  con:address :address_1367;
  dcterms:title "Alii Garden Market Place";
  wgs:lat -155.9; wgs:long 19.6 .
```

```
:address_1367
  a con:Address;
  con:stateOrProvince typed_state:Hawaii;
  con:street          "75-6129 Alii Drive";
  con:city            "Kailua-Kona";
  con:zip             "96740" .
```

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