

# Validation of Ebola LOD

## **Team Members:**

Jonathan Downs

Yash Pant

**Instructor:** Dr. Edward Fox

**Client:** S.M.Shamimul Hasan

**School:** Virginia Tech - Blacksburg, VA

**Date:** 4/28/2016

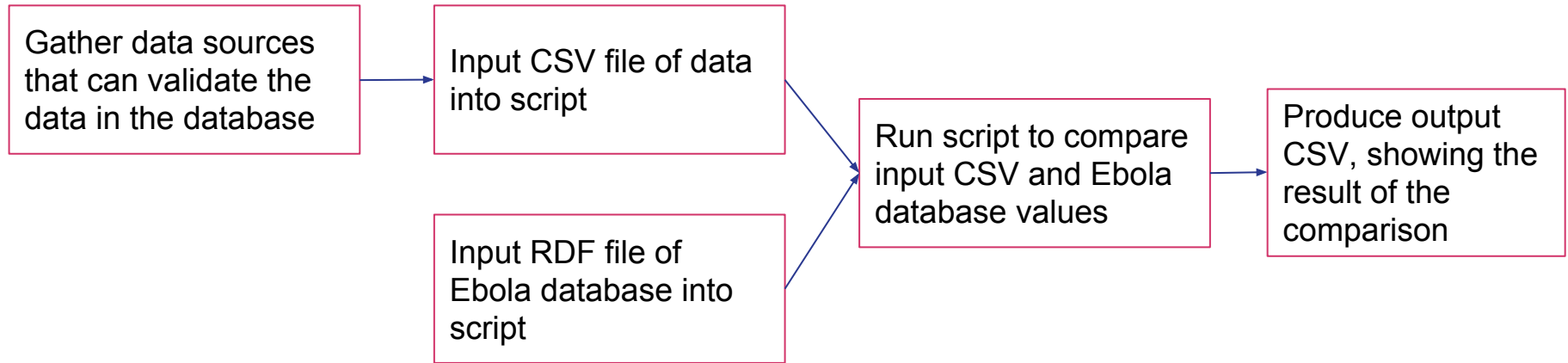
# Background

- Ebola epidemic of 2014
- Largest database of Ebola data
- Large amount of data that needs to be verified
- Validate data

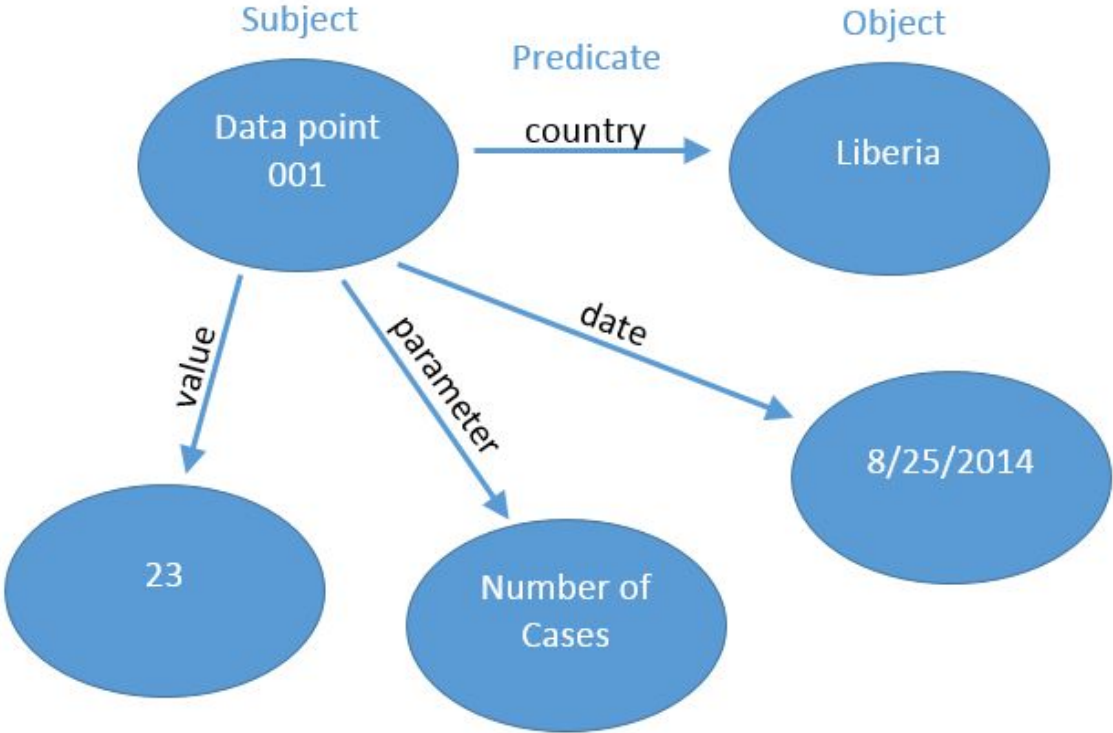


# Project Goal

Goal: To create script that allows a user to validate data in the Ebola database

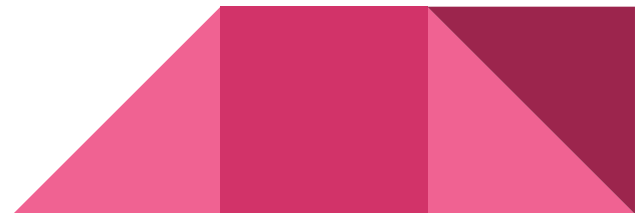


# RDF



# Prototype - Overview

- Written in Python
- Focused on specific dataset (time-series data in Guinea)
- Focused on one validation file



# Prototype - Algorithm

- Matched every row in an input CSV file to one data point in the RDF database
- Searched for data points by looking at country, date, and parameter
- Compared the value at each data point to validate it

Example of row in CSV:

	A	B	C	D
1	Country	Date	Parameter	Value
2	Guinea	8/4/2014	Probable Cases	0
3	Guinea	7/12/2014	Deaths	2

# Prototype - Sample Results

	A	B	C	D	E
1	Datapoint ID	Ebola DB Value	Input CSV Value	Exact Match	Difference
2	9380	28	28	TRUE	0
3	9383	1022	1008	FALSE	14
4	9388	468	465	FALSE	3
5	9378	162	162	TRUE	0



# Final Script - Overview

- Final script generalizes prototype to take various input sources
- Allows a user to input any relevant CSV file as an input, using a GUI
- Delivers an output CSV showing a comparison of input and Ebola DB

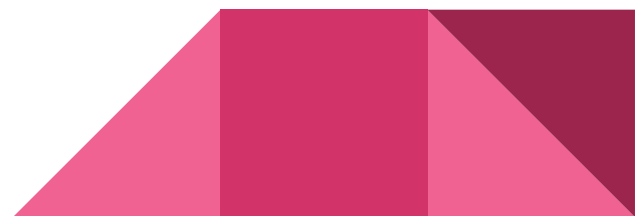


# Modifications to Prototype - Parsing CSV's

- Changed script to parse any CSV inputted
- CSV to RDF relationship:

	A	B	Predicate C	D
1	Country	Date	Parameter	Value Object
2	Guinea	8/4/2014	Probable Cases	0
3	Guinea	7/12/2014	Deaths	2
4	Liberia	8/12/2014	Confirmed Cases	5

Subject

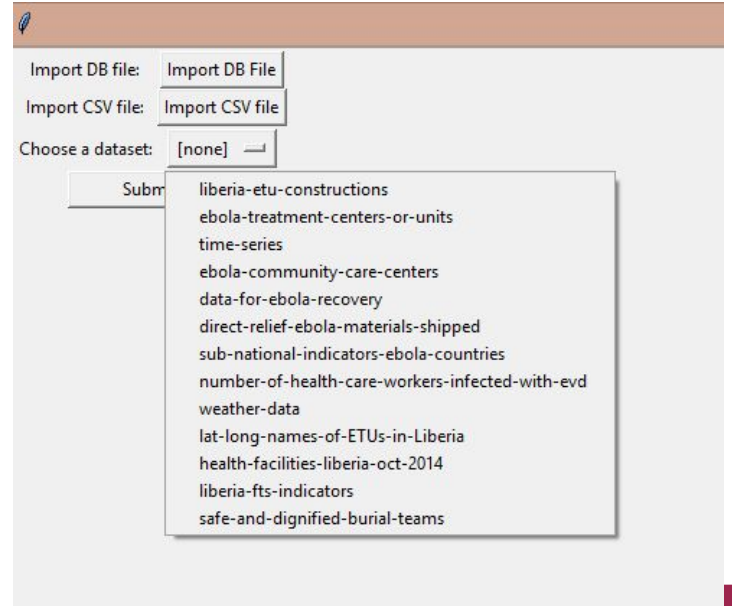
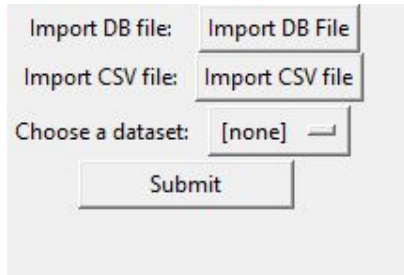


# Modifications to Prototype - Searching

- Create a composite key to search the database for data points
  - Take pairs of predicate and object to find subject
  - Validate one predicate (data field) at a time
- Validating: 0 Cases in Guinea on 8/4/2014
  - Create composite key with the following pairs of (predicate, object):  
(Country, Guinea) (Incident Date, 8/4/2014) (Time Series Parameter, New Cases of Probables)
  - Run search based on this key to find the associated RDF subject (data point)
  - Get the number of cases from the subject found

```
<http://ebola.vbi.vt.edu/dataset/time-series/observation-index/2069> a <http://purl.org/linked-data/cube#Observation> ;  
<http://purl.org/linked-data/cube#dataSet> "http://ebola.vbi.vt.edu/dataset/time-series" ;  
<http://ebola.vbi.vt.edu/def/time-series/statistical-dimension/country> <http://ebola.vbi.vt.edu/time-series/guinea> ;  
<http://ebola.vbi.vt.edu/def/time-series/statistical-dimension/INCIDENT_DATE> "8/4/2014" ;  
<http://ebola.vbi.vt.edu/def/time-series/statistical-dimension/TIME_SERIES_PARAMETER> "New cases of probables" ;  
<http://ebola.vbi.vt.edu/def/time-series/statistical-dimension/VALUE> "0" .
```

# GUI Elements



Dataset Predicates	CSV Columns	Point of Validation?
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/OBJECTID">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/OBJECTID</a>	Choose a mapping. ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Status">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Status</a>	status ▾	<input checked="" type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Accuracy">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Accuracy</a>	accuracy ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Beds_Plan">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Beds_Plan</a>	beds_plan ▾	<input checked="" type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/ADM1_CODE">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/ADM1_CODE</a>	adm1_code ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Comment">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Comment</a>	comment ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Lead_Donor">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Lead_Donor</a>	lead_donor ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/ECF_Code">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/ECF_Code</a>	ecf_code ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Partner">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Partner</a>	partner ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Open_Date">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Open_Date</a>	updated ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/country">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/country</a>	Choose a mapping. ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Longitude">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Longitude</a>	longitude ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Type">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Type</a>	type ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Plan_Open">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Plan_Open</a>	plan_open ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Updated">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Updated</a>	updated ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Beds_Open">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/Beds_Open</a>	beds_open ▾	<input checked="" type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/DSTR_NM">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/DSTR_NM</a>	dstr_nm ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/the_geom">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/the_geom</a>	the_geom ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/FID">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/FID</a>	fid ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/ebola-treatment-centers-or-units/statistical-dimension/country">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/ebola-treatment-centers-or-units/statistical-dimension/country</a>	Choose a mapping. ▾	<input type="checkbox"/>
<a href="http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/AdminCode">http://ebola.vbi.vt.edu/def/ebola-treatment-centers-or-units/statistical-dimension/AdminCode</a>	adm1_code ▾	<input type="checkbox"/>

# Sample Results

	A	B	C	D	E	F	G
1	Name	City	MGRS	STATUS	Beds - EbolaDB Value	Beds - Input Value	Exact Match?
2	Bopolu	Bopolu	TBD	Site TBD	100	100	TRUE
3	ELWA 2	Paynesville	29NLG1236189998	Operational	40	40	TRUE
4	ELWA 3	Paynesville	29NLG1189090530	Operational	200	180	FALSE
5	ELWA 4	Paynesville	29NLG1204390376	Under construction	300	300	TRUE
6	Elwa Hospital	Paynesville	29NLG1230389948	Operational	Unk	Unk	TRUE
7	Firestone Hospital	Firestone	29NLH3743702463	Operational	100	100	TRUE
8	Foya	Foya	29PLK6735966135	Operational	75	73	FALSE
9	Sinoe	Greenville	TBD	Site TBD	50	50	TRUE
10	Island Clinic, Oniyama Specialist Hospital	Bushrod Island	29NLH0238206106	Operational	160	160	TRUE
11	JFK Hospital	Monrovia	29NLG0385195257	Operational	40	40	TRUE
12	Kakata Save the Children	nan	29NLH5384525578	Under construction	50	50	TRUE
13	Ministry of Health	Congo Town	29NLG1051492777	Under construction	100	100	TRUE
14	Old Ministry of Defense	Congo Town	29NLG0814993399	Under construction	300	300	TRUE
15	Redemption Hospital	Monrovia	29NLH0175904425	Operational	Unk	Unk	TRUE
16	Sinje	Sinje	29NKH6360653975	Under construction	100	100	TRUE
17	Samuel Kanyon Doe Sports Complex	nan	29NLG1169891921	Under construction	600	600	TRUE
18	Tubmanburg	Tubmanburg	29NKH9768059470	Under construction	100	100	TRUE
19	Zorzor	Zorzor	TBD	Site TBD	100	100	TRUE

# Sources

<http://fadyart.com/en/images/stories/rdf.png>

