



Rensselaer

why not change the world?®

Knowledge Graphs, Webs of Data

Ahmed Eleish

Data Science – ITWS/CSCI/ERTH 4350/6350

Week 4, September 23rd, 2026

Tetherless World Constellation
Rensselaer Polytechnic Institute



Contents

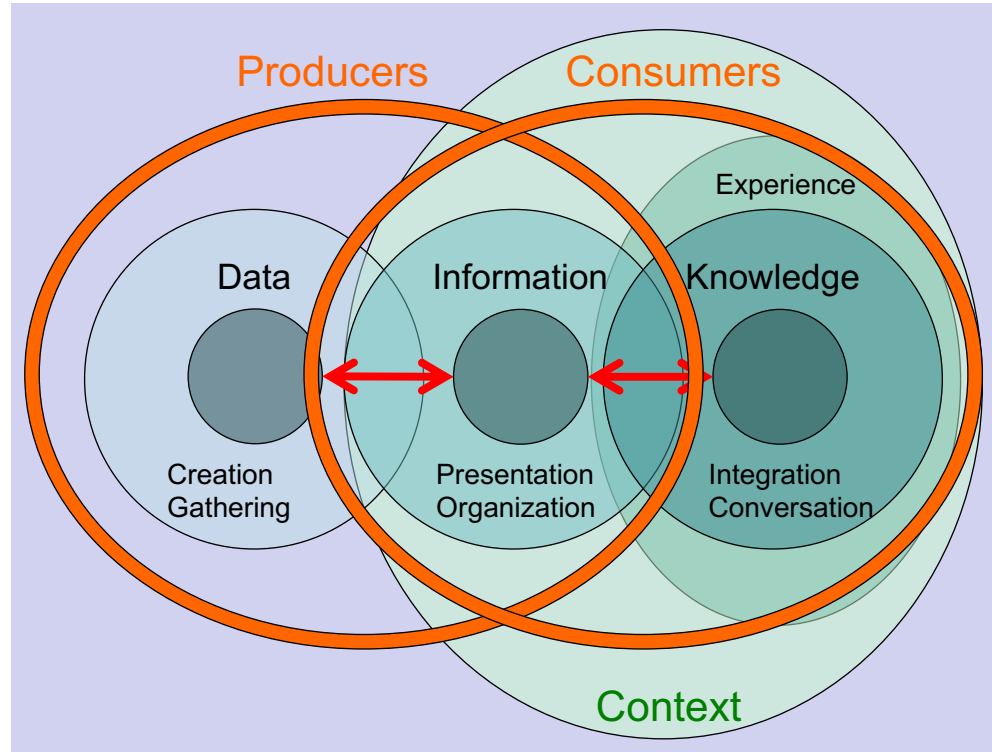
- METS
- Knowledge Graphs
- OWL, RDF & The Semantic Web
- Assignment 2



Reading

- Metadata Encoding and Transfer Standard - METS [\[24\]](#)

Data-Information-Knowledge Ecosystem



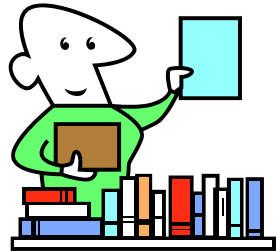
Modes of collecting data, information

- Observation
- Measurement
- Generation

- Driven by
 - Questions
 - Research idea
 - Exploration

Data Management

- Creation of logical collections
- Physical data handling
- Interoperability support
- Security support
- Data ownership
- Metadata collection, management and access
- Persistence
- Knowledge and information discovery
- Data distribution and publication



Data Management

- Creation of logical collections

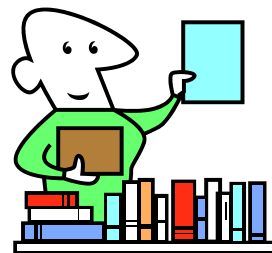
e.g. project files and directories follow consistent naming convention; metadata contain referend to data.

- Physical data handling

e.g. filenames listed in metadata/documentation with brief descriptions, backups in clearly named directory.

- Interoperability support

e.g. data available as CSV/JSON/XML with Dublin Core metadata keys.



Data Management

- Security support

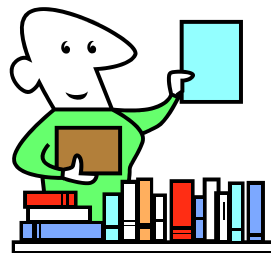
e.g. data stored in cloud service with permissions set individually for authorized individuals.

- Data ownership

e.g. team member assigned to fill quality metadata/provenance.

- Metadata collection, management and access

e.g. metadata collected in spreadsheets/plain text notepad are integrated into a JSON file.



Data Management

- Persistence

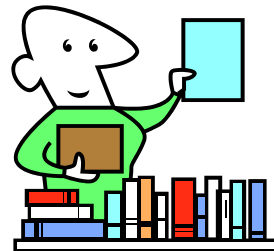
e.g. Data package will be submitted to RPI [Institute Archives and Special Collections](#) for archival, archival package must conform to the [OAIS](#) standard and utilize the [METS](#) metadata schema.

- Knowledge and information discovery

e.g. A description is published along with of the dataset listing keywords, detailing funding, and citing papers related to the data.

- Data distribution and publication

e.g. Dataset is stored on a publicly accessible university ftp server; a postdoc sends an email to a list whenever the data are updated.

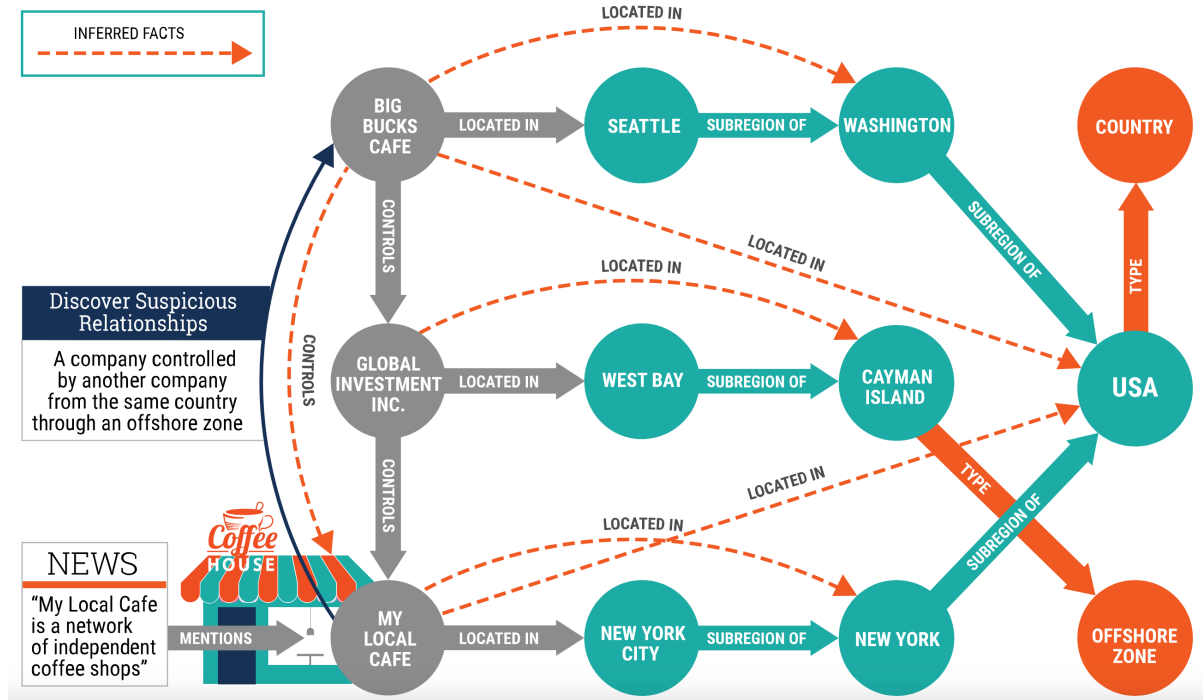


Provenance

- Who?
- What?
- Where?
- Why?
- When?
- How?

Knowledge Graphs

What is a Knowledge Graph?



<https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>

What is a Knowledge Graph?

The knowledge graph (KG) represents a collection of interlinked descriptions of entities – real-world objects, events, situations or abstract concepts – where:

- Descriptions have a formal structure that allows both people and computers to process them in an efficient and unambiguous manner;
- Entity descriptions contribute to one another, forming a network, where each entity represents part of the description of the entities, related to it.

Where are Knowledge Graphs used?

- Web Search (e.g. [Google](#))
- Data Integrations
- Answering Questions ~ before LLMs :’(

Google Knowledge Graph

- Introducing Knowledge graphs:

- Watch:

https://www.youtube.com/watch?v=mmQl6VGvX-c&t=17s&ab_channel=Google

- Read:

<https://blog.google/products/search/introducing-knowledge-graph-things-not/>

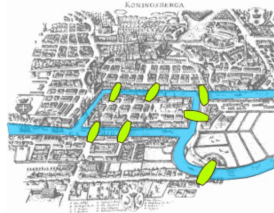
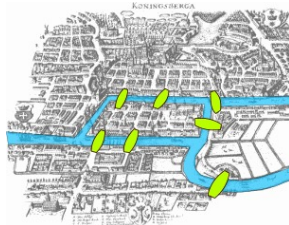
Knowledge Graphs (KGs)

Knowledge Graphs are a type of **graph**.

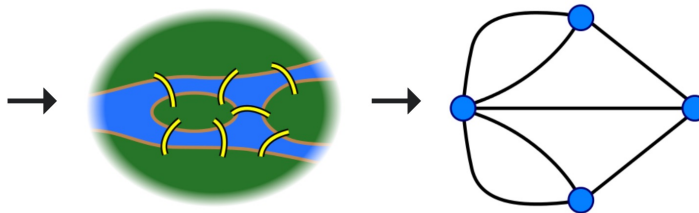
- What are graphs?
- Graphs are simple structures consisting of **nodes (or vertices)** connected by **links (or edges)**.
- Mathematically, a graph is a set of vertices V and a set of edges E where each edge in E connects 2 vertices from V .
- The nodes represent entities (things) and the links represent relations/connections between the entities.

Graphs

- Graphs are sometimes referred to as networks.
- They are a simple but very powerful way of describing how things are connected.
- Graph Theory was introduced by the Swiss mathematician Leonhard Euler during the 18th century with his famous problem known as the Seven Bridges of Königsberg.



Euler used this graph to answer the question: Can you walk through the city and cross each bridge only once?



Images/Resource Credits: https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Knowledge Graphs (KGs)

- KGs are graphs where the nodes represent entities from a particular domain of interest, and the links represent qualified relations between these entities.
- KGs can be explored with structured queries.
- KGs can be analyzed as networks of data.
- Formal semantics can be defined for KGs by using ontologies.

<https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>

The Semantic Web

- The Semantic Web is a **web of data** (not (just) documents)
- “It is about common formats for integration and combination of data drawn from diverse sources..”
- “It is also about language for recording how the data relates to real world objects.”
- “The goal of the Semantic Web is to make Internet data machine-readable”
- How do we encode data on the web? → RDF!

Semantic Web: <https://www.w3.org/2001/sw/>

Web of Data: <https://www.w3.org/2013/data/>

Resource Description Framework (RDF)

- A framework for expressing information about resources. Resources can be anything, including documents, people, physical objects, and abstract concepts.
- RDF encode data in the form of triples:

`<subject> <predicate> <object>`

EXAMPLE 1: Sample triples (informal)

```
<Bob> <is a> <person>.  
<Bob> <is a friend of> <Alice>.  
<Bob> <is born on> <the 4th of July 1990>.  
<Bob> <is interested in> <the Mona Lisa>.  
<the Mona Lisa> <was created by> <Leonardo da Vinci>.  
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```

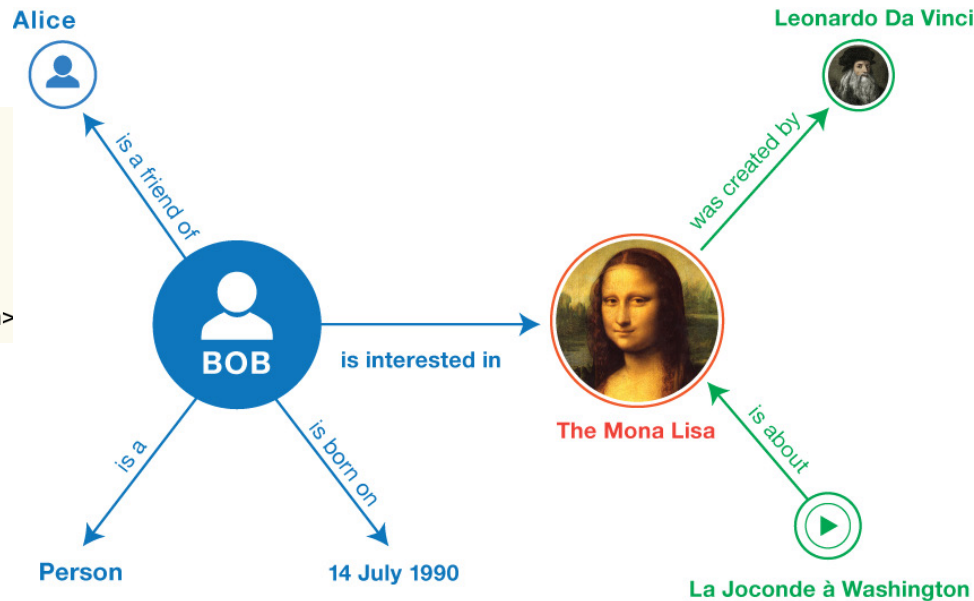
RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Resource Description Framework (RDF)

- RDF encode data in the form of triples

EXAMPLE 1: Sample triples (informal)

```
<Bob> <is a> <person>.  
<Bob> <is a friend of> <Alice>.  
<Bob> <is born on> <the 4th of July 1990>.  
<Bob> <is interested in> <the Mona Lisa>.  
<the Mona Lisa> <was created by> <Leonardo da Vinci>.  
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```



RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Resource Description Framework (RDF)

- In RDF everything is a resource.
- Every resource has Uniform Resource Identifier (URI).
- Classes and properties must have namespaces (e.g. FOAF).
- RDF can be serialized (written in a file) in several ways:
 - N-Triples, Turtle, JSON-LD, RDFa, RDF/XML

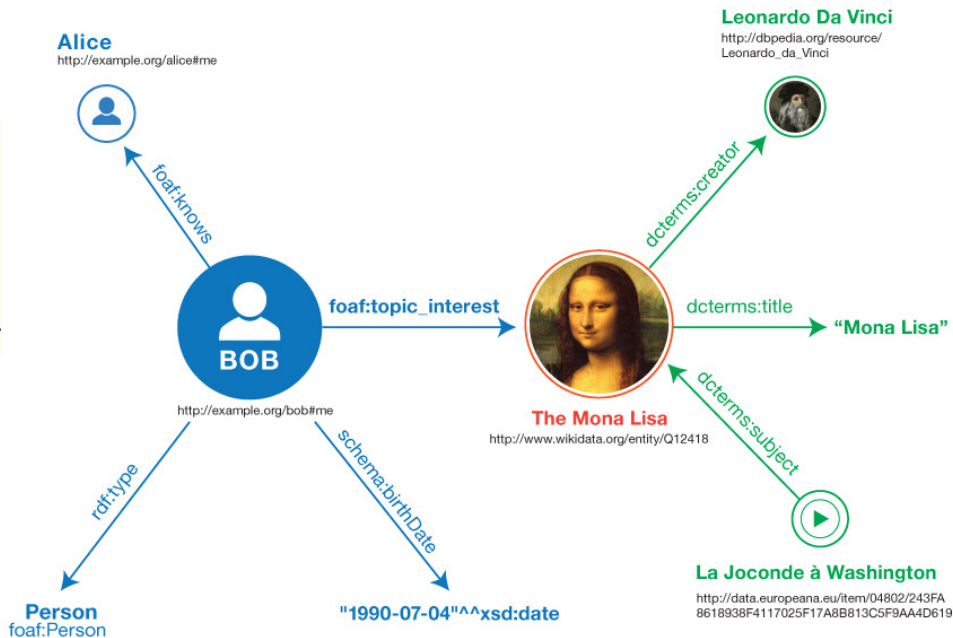
RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Resource Description Framework (RDF)

- RDF encode data in the form of triples

EXAMPLE 1: Sample triples (informal)

```
<Bob> <is a> <person>.  
<Bob> <is a friend of> <Alice>.  
<Bob> <is born on> <the 4th of July 1990>.  
<Bob> <is interested in> <the Mona Lisa>.  
<the Mona Lisa> <was created by> <Leonardo da Vinci>.  
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>
```



RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

RDF/XML

- Serialized RDF contains namespaces, prefixes, URIs

EXAMPLE 15: RDF/XML

```
01 <?xml version="1.0" encoding="utf-8"?>
02 <rdf:RDF
03     xmlns:dcterms="http://purl.org/dc/terms/"
04     xmlns:foaf="http://xmlns.com/foaf/0.1/"
05     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
06     xmlns:schema="http://schema.org/"
07   <rdf:Description rdf:about="http://example.org/bob#me">
08     <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
09     <schema:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1990-07-04</schema:birthDate>
10     <foaf:knows rdf:resource="http://example.org/alice#me"/>
11     <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/Q12418"/>
12   </rdf:Description>
13   <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
14     <dcterms:title>Mona Lisa</dcterms:title>
15     <dcterms:creator rdf:resource="http://dbpedia.org/resource/Leonardo_da_Vinci"/>
16   </rdf:Description>
17   <rdf:Description rdf:about="http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
18     <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
19   </rdf:Description>
20 </rdf:RDF>
```

RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

JSON-LD

- Serialized RDF contains namespaces, prefixes, URIs

EXAMPLE 12: JSON-LD

```
01  {
02    "@context": "example-context.json",
03    "@id": "http://example.org/bob#me",
04    "@type": "Person",
05    "birthdate": "1990-07-04",
06    "knows": "http://example.org/alice#me",
07    "interest": {
08      "@id": "http://www.wikidata.org/entity/Q12418",
09      "title": "Mona Lisa",
10      "subject_of": "http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619",
11      "creator": "http://dbpedia.org/resource/Leonardo_da_Vinci"
12    }
13  }
```

RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

N-Triples

- Serialized RDF contains namespaces, prefixes, URIs

EXAMPLE 6: N-Triples

```
01 <http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
02 <http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
03 <http://example.org/bob#me> <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date> .
04 <http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest> <http://www.wikidata.org/entity/Q12418> .
05 <http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa" .
06 <http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/creator> <http://dbpedia.org/resource/Leonardo_da_Vinci> .
07 <http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619> <http://purl.org/dc/terms/subject> <http://www.wikidata.org/entity/Q12418> .
```

RDF Primer: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>

Ontologies (Information Science)

Ontologies define the structure/organization of knowledge in a domain.

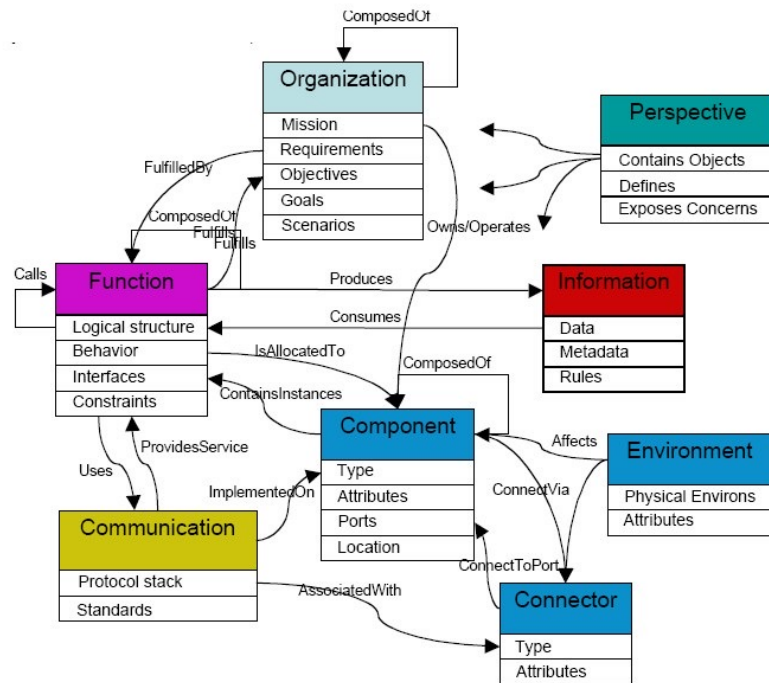
- “An ontology is a formal description of knowledge as a set of concepts within a domain and the relationships that hold between them.”
- “an ontology encompasses a representation, formal naming, and definitions of the categories, properties, and relations between the concepts, data, or entities that pertain to one, many, or all domains of discourse”
- An ontology can be considered the data model or schema of a knowledge graph, it represents the semantics (meanings) behind the data.
- Ontology are similar to, but more comprehensive than taxonomies and controlled vocabularies.

Ontologies

- An ontology contains at the least classes, properties, and relations.
- Ontologies may optionally contain restrictions, rules, and axioms which enable a person or machine (program) to draw inferences from a knowledge graph.

e.g.

- A person's age must be a positive number
- Siblings must share one or more parents
- Taking the Data Science course is necessary for success



Example Ontology - FOAF

- Friend Of A Friend (FOAF) is a “a machine-readable ontology describing persons, their activities and their relations to other people and objects.”
- “FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF) and the **Web Ontology Language (OWL)**.”
- Contains classes: Agent, Person, Group, Organization, Project, OnlineChatAccount, ...

e.g.

```
<foaf:Person rdf:about="#danbri" xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:name>Dan Brickley</foaf:name>
  <foaf:homepage rdf:resource="http://danbri.org/" />
  <foaf:openid rdf:resource="http://danbri.org/" />
  <foaf:img rdf:resource="/images/me.jpg" />
</foaf:Person>
```

FOAF Specification <http://xmlns.com/foaf/spec/>

Why Semantic Web, RDF, Ontologies, FOAF, etc.??

- Linked Open Data:

“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.”

- [Linked Open Data Cloud](#)

Linked Open Data: <https://www.w3.org/DesignIssues/LinkedData.html>

Assignment 1 due Friday night!

Assignment 2 instructions out..

No readings this week!

Thanks!

Work on your data collection!!