

# SEMANTIC DATA SERVICES BY INTEGRATING OPEN AND BIG DATA

Haklae Kim, Ph.D



# AGENDA

- Paradigm shift of data
- Introduction of graph theory
- K# Platform
- Things to do, now



# IN A WORLD OF BIG DATA

## What Happens in an **Internet Minute?**

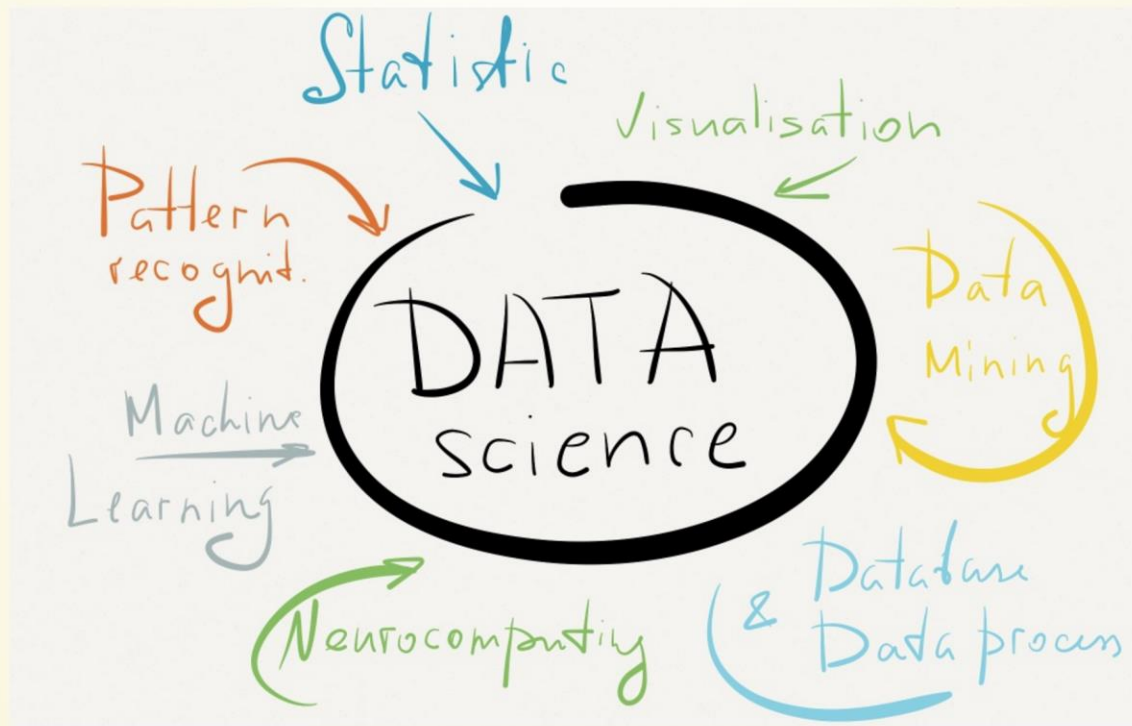


## And Future Growth is Staggering



# DATA SCIENCE AND SCIENTIST

By 2018, there would be between 140,000 and 190,000 unfilled positions in data science. - The McKinsey Global Institute





# SMALL, BUT RIGHT DATA

It's not about how much data your business has,  
it's about [what you're doing with it](#).

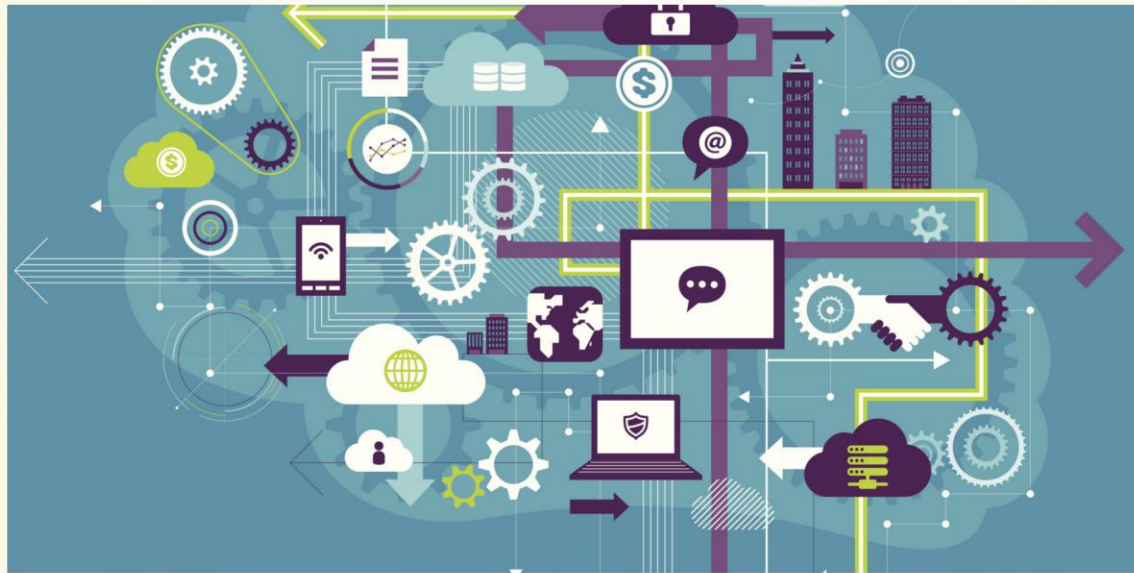


Think about [a wealth of free data](#) that's available on the Web,  
including government open data.



## CONNECTED DATA, DEVICES AND SERVICES

A **graph** that interconnects millions of pieces of data about keywords



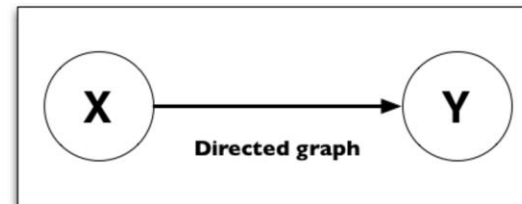
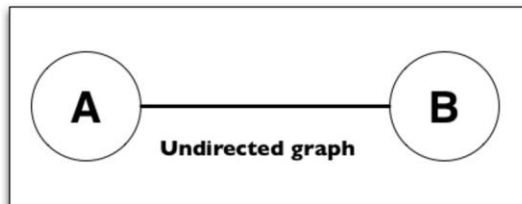
Is it available now? **Maybe**



# GRAPH

In mathematics, and more specifically in graph theory, a **graph** is a representation of a set of objects where some pairs of objects are connected by links.

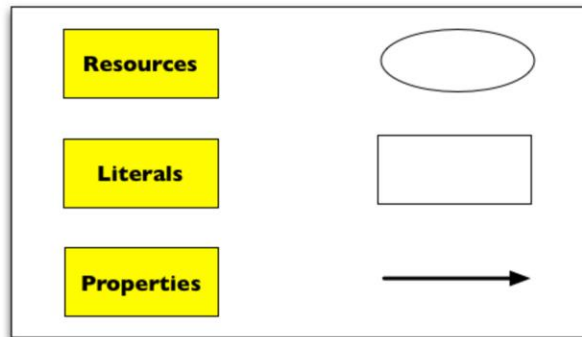
- **Vertices** are called nodes or points
- **Edges** are called arcs or lines



# RDF (RESOURCE DESCRIPTION FRAMEWORK)

A language to describe **Resources**, which are any things or concepts in the world (e.g. a book, a movie, a person, etc.)

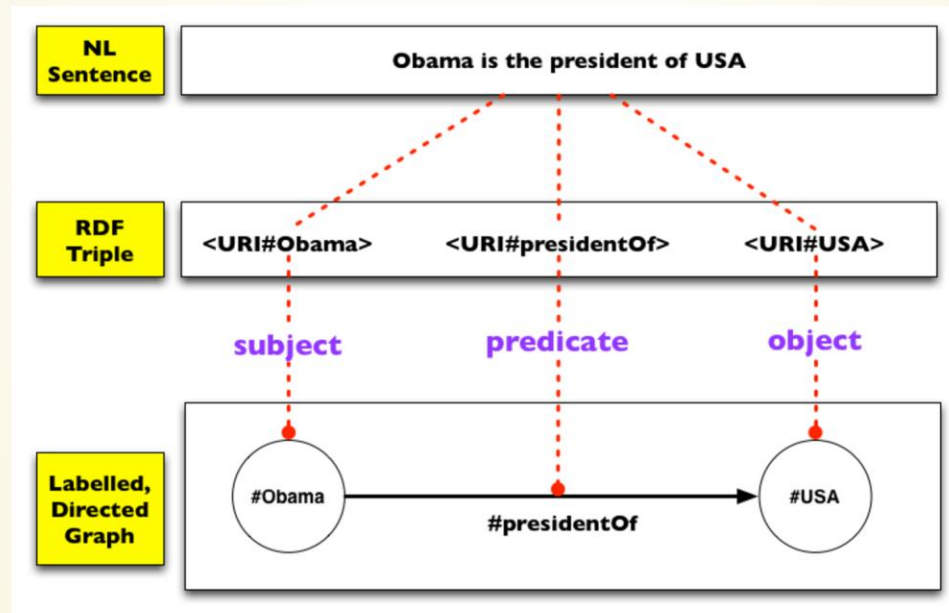
Graph notations





# RDF AS A GRAPH

**Knowledge** is represented in RDF as a set of **Triples**, which is analogous a sentence in natural language.



# RDF AS A SERIALIZATION

An RDF graph is a logical model, several **serialization formats** are standardized, including <sup>RDF/XML, Turtle, N3, NTriples, etc.</sup>

## RDF/XML

```
<rdf:Description rdf:about="URI#Obama">
  <rdf:type rdf:resource="URI#Person"/>
  <URI#presidentOf rdf:resource="URI#USA"/>
</rdf:Description>
```

## Turtle

```
<URI#Obama> <URI#presidentOf> <URI#USA> .
<URI#Obama>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <URI#Person> .
```



# RDF AS A SERIALIZATION

An RDF graph is a logical model, several **serialization formats** are standardized, including <sup>RDF/XML, Turtle, N3, NTriples, etc.</sup>

## RDF/XML

```
<rdf:Description rdf:about="URI#Obama">
  <rdf:type rdf:resource="URI#Person"/>
  <URI#presidentOf rdf:resource="URI#USA"/>
</rdf:Description>
```

## Turtle

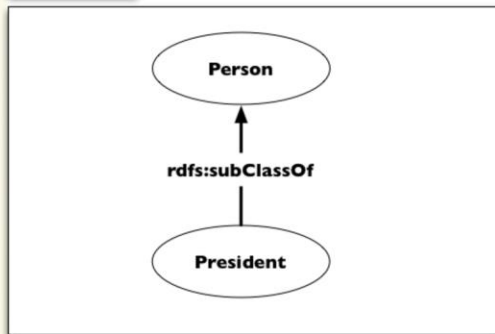
```
<URI#Obama> <URI#presidentOf> <URI#USA> .
<URI#Obama>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <URI#Person> .
```



# RDF SCHEMA

RDFS schema defines the **basic vocabulary** that can be used in RDF document to describe resources and their relationships  
(Class, Property, Hierarchy, Collection, Reification, Entailment for reasoning)

Graph



RDF/XML

```
<rdf:Class rdf:ID="URI#President">  
  <rdfs:subClassOf rdf:resource="URI#Person"/>  
</rdf:Class>
```



# WHAT IS K#?

K# stands for **K**nowledge **S**haring **P**latform.

It is a knowledge base **to enhance intelligent services** with semantic information gathered from a wide variety of sources.

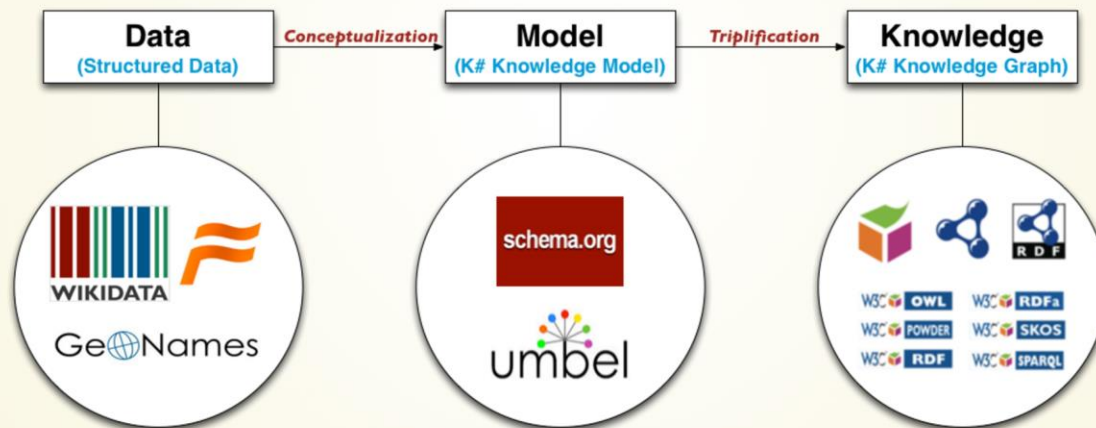
- Extensible Knowledge Graph
- Semantic Search and Reasoning
- Contextual Service Delivery



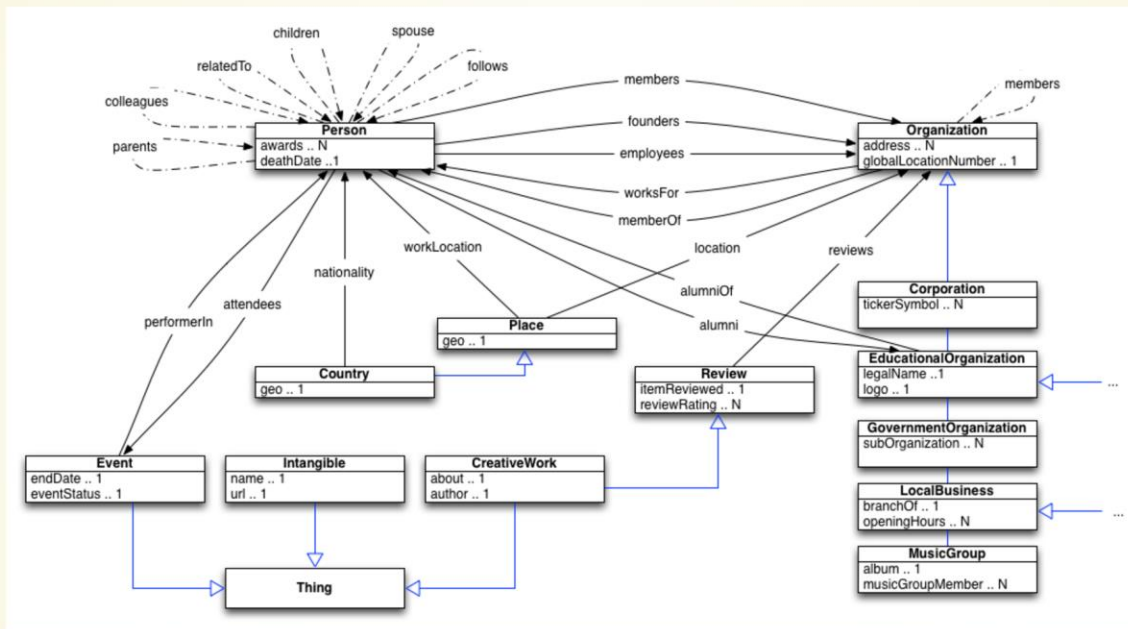


# FROM DATA TO KNOWLEDGE

K# Knowledge Graph creates **structured and detailed information about the topic** in addition to a list of links to other sources.

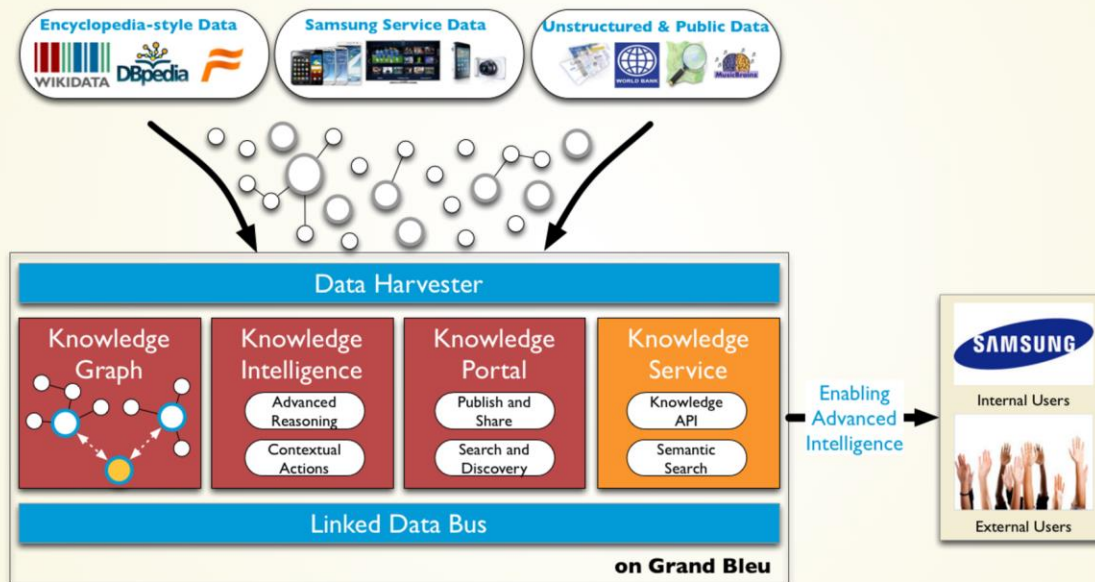


# K# KNOWLEDGE MODEL



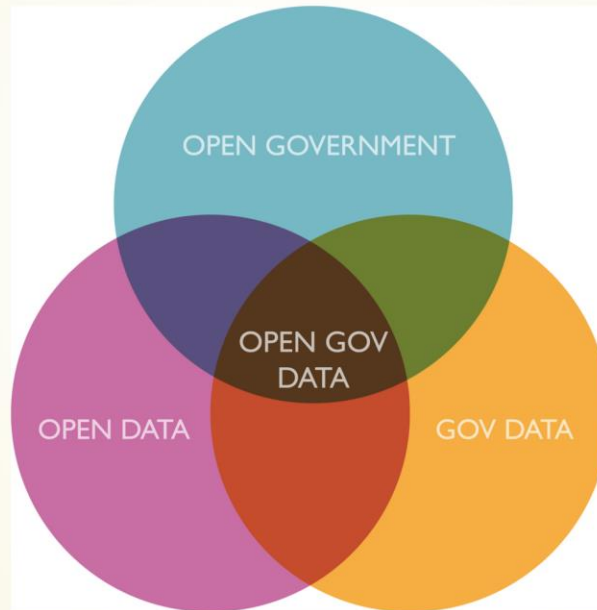
# VISION OF K#

Understand and uniquely identify every "Thing" in the World!



# WHAT DOES OPEN MEAN?

A piece of data or content is open if **anyone** is **free to use, reuse, and redistribute** it.



# USEFUL RESOURCES FOR BEGINNERS

- 공공데이터포털
- 30 places to find open data on the Web
- Big data: 20 free big data sources everyone should know
- Open Data 500 Companies
- Where can I find large datasets open to the public?
- Open data portal in google search



# OPEN KNOWLEDGE KOREA

Korea has been active since 2012, working on Linked Data and encouraging international collaboration to open up data.



**OPEN KNOWLEDGE**

<http://okfn.kr>





# CONCLUDING REMARK

Data scientists take an enormous mass of messy and (un)structured data and use **their formidable skills in math, statistics and programming** to clean, massage and organize them.

- **Problem statement** from our world
- **Implementation** to solve and realise the problem
- **Presentation** for understanding





# THE END

BY HAKLAE KIM / SCOT.KIM@SAMSUNG.COM

