

Followup of OBO Academy's Andra Waagmeester & Tiago Lubiana Introduction to WikiData (BioMed Orientation)

Wikipedia: Maintained independently by >300 language communities

The screenshot displays three Wikipedia articles for 'Malaria' in Dutch, Greek, and English. Each article includes a micrograph of red blood cells and a list of ICD-10 codes. Below the articles, a table summarizes the data:

Language	ICD-10	B50	ICD-10	B50
Dutch	ICD-10: B50.0	B50.0	ICD-10: B50.0	B50.0
Greek	ICD-10: B50.0	B50.0	ICD-10: B50.0	B50.0
English	ICD-10: B50.0	B50.0	ICD-10: B50.0	B50.0

Below the table, a map of the world shows the distribution of malaria. The map is divided into three regions: Dutch, Greek, and English. The Dutch region is highlighted in red, the Greek region in yellow, and the English region in green. The map also shows the number of articles in each language community, with a total of 103,400 articles (197th) as of 2014.

Wikipedia – a
source of knowledge

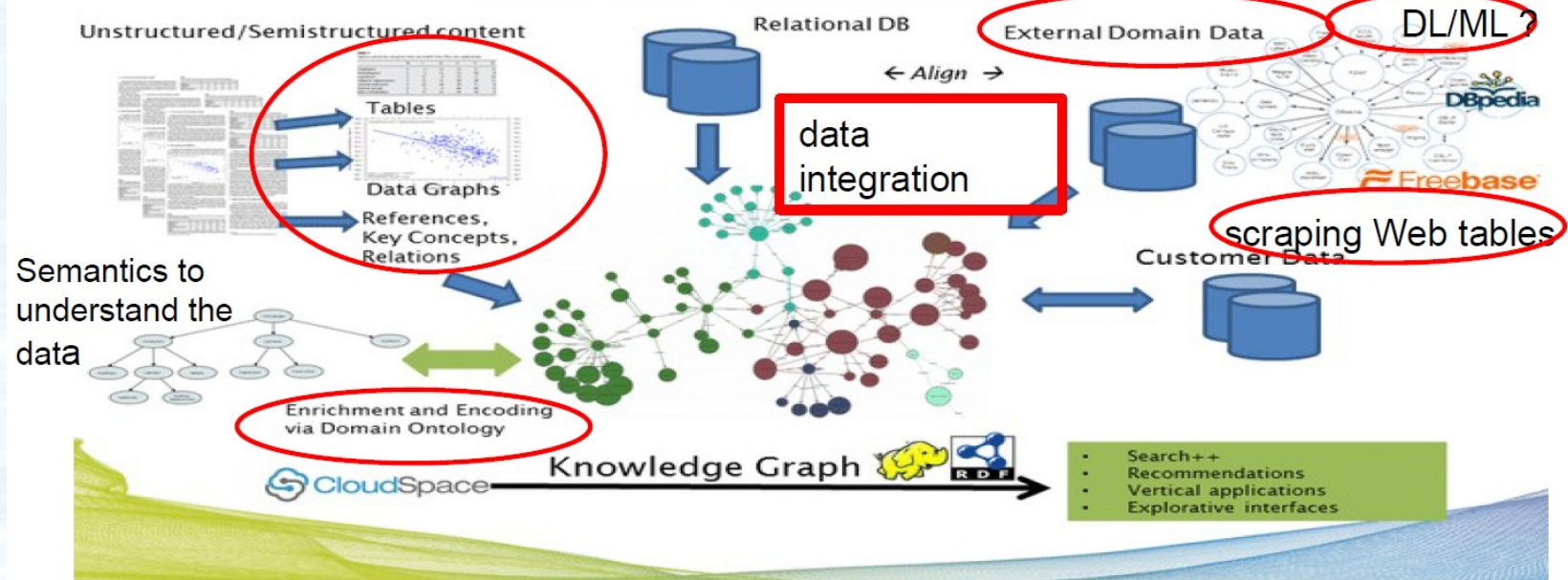
Original Gene Wiki paper:
<https://doi.org/10.1371/journal.pbio.0060175>

Links to DBs & stubbed
templates to develop full
fledged articles.

300 languages so
difficulty updating such
Wikipedia articles

Gary Berg-Cross (gbergcross@gmail.com)
2023 Ontology Summit Co-Chair with James Overton

There are many ingredients to a quality KG



- From https://www.researchgate.net/post/What_is_Knowledge_Graphs
Ajit kumar Roy's What is Knowledge Graphs?

2012 Start BioMedical Example:

Wikidata is a **community** and a **hub** for building community!!!

Label &
header

- Description & Statements like RDF
- Triple is extended with references & qualifiers

aliases

Sidelinks to
different
language
pages

Retinoic acid receptor alpha (Q254943)

mammalian protein found in Homo sapiens
Nuclear receptor subfamily 1 group B member 1 | RARA

Statements

molecular function (P680)
represents gene ontology function annotations

retinoic acid binding (Q14901431)
Interacting selectively and non-covalently with retinoic acid, 3,7
GO:0001972

Statements

subclass of retinoid binding
1 reference

IDA (Q23174122)
Gene Ontology evidence code
Inferred from Direct Assay

Statements

instance of Gene Ontology Evidence code
manual assertion

retinoic acid binding
determination method

1 reference

retrieved 3 January 2017

stated in A human retinoic acid receptor which belongs to the family of nuclear receptors
UniProt-GOA

curator British Heart Foundation

reference URL http://www.ebi.ac.uk/QuickGO/GAnotation?protein=P10276
IDA

transcription corepressor activity
determination method IDA

1 reference

Wikipedia (7 entries) edit

- ar مستقبل حمض الريتينويك ألفا
- en Retinoic acid receptor alpha
- es Receptor de ácido retinoico alfa
- sh Receptor retinoinske kiseline alfa
- sr Receptor retinoinske kiseline alfa
- uk RARA
- zh 视黄酸受体α

British Heart Foundation (Q4970039)
Statements

instance of organization

official website http://www.bhf.org.uk/en/

Identifiers

GRID ID grid.452924.c

Doesn't work with strict idea of definitions

WIKIDATA IN BRIEF

Claim: Item – Property – Value

Item Q number
Property P number

item	property	value
Marie Curie	field of work	chemistry
Q7186	P101	Q2329

Wikidata item basics

[Label, description, aliases](#) – human friendly language-specific strings

[Statements](#) – includes identifiers, sitelinks at bottom of an item

[Qualifiers](#) – adds details to claims (ie. dates, distinctions, etc)

[Copyright](#) - all content in Wikidata is Creative Commons CC0 (public domain)

Critical concepts

[Help:Basic membership properties](#)

Instance (P31) vs subclass (P279)

[Help:Sources](#)

Adding references, use "stated in" or "reference URL"

Displaying items

[Reasonator](#) and [SQID](#) – rich visual display of Wikidata items

[Wikidata Graph Builder](#) – show connectedness of items based on properties

Querying

[Query.wikidata.org](#) – SPARQL interface. Explore example queries and display options. CTRL-Space auto-expands item and property names to Q/P numbers.

[VizQuery](#) – simple query using rules without requiring coding

Simple editing tools

[Distributed Game](#) – contribute through Yes/No/Skip micro-decisions

[Mix-n-Match](#) – connect external identifiers to Wikidata items

Bulk editing

[Quickstatements](#) – use tab-delimited claims to edit

[PetScan](#) – powerful toolbox for Wikipedia categories and Wikidata

[OpenRefine](#) – open source tool for cleaning messy data, reconciliation

[Data donation](#) - how and why to contribute, and guidelines for institutions

Other tools

[TABernacle](#) – inspect and edit Wikidata interactively in table format

[Google Spreadsheet Wikipedia Tools](#) – query Wikipedia and Wikidata from spreadsheet cells

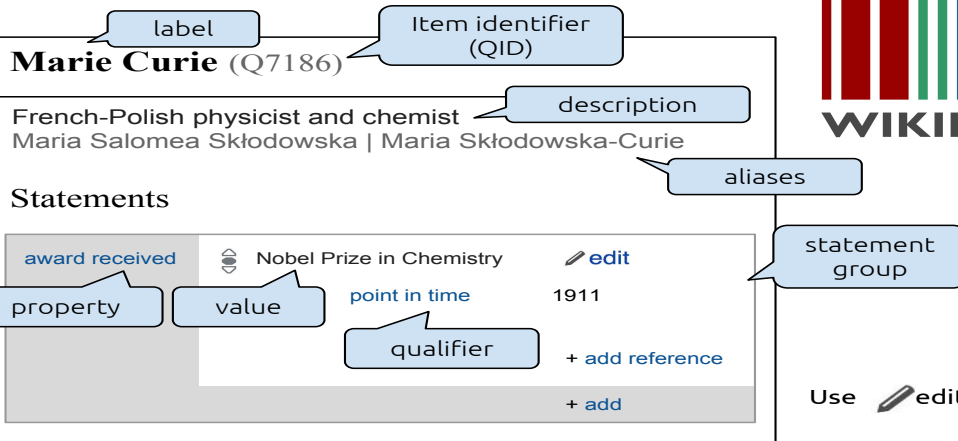
[External tools](#) – comprehensive list of all Wikidata tools

[WikiBase](#) – open source software extension to MediaWiki that runs Wikidata. Available as a Docker container to create a custom structured data repository

Community and news

[Wikidata:Project chat](#) – discussion forum

[Wikidata:Status updates](#) – weekly summary newsletter



Version 1.3 (2018-11)
by Andrew Lih
(User:Fuzheado)
licensed under
CC BY SA 4.0
andrew.lih@gmail.com

Digital version:
bit.ly/wikidata-onepage



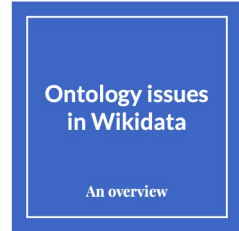
Use edit and +add to modify

See: https://www.wikidata.org/wiki/Wikidata:In_one_page

There is an ontology– though not based on OWL

Overview of types of issues we found

- Semantic drift
- Structural bugs
 - cycles
 - mix-up of meta levels
 - redundant classification
 - redundant generalisation
 - exchanged sub-/superclasses
- Upper level ontology is messy
- Conceptual ambiguity
- Inconsistent modeling
- Overgeneralisation
- Conflicting real-world models
- Unclassified Items



Andra Waagmeester “Wikidata is NOT an ontology, but simply proxy. As long as reliable references exist, there is room for disagreement on Wikidata”

https://www.wikidata.org/wiki/File:WikidataCon_2021_-_Overview_of_ontology_issues.pdf

<https://youtu.be/5AYzuBOkXeI>

Status of Wikidata connection to OBO Foundry

- Sequence types and features ontology - [P3986](#)
- Symptom Ontology - [P8656](#)
- Cephalopod Ontology - [P9334](#)
- Gender, Sex, and Sexual Orientation Ontology - [P9827](#)
- Vaccine Ontology - [P1928](#)
- Relations Ontology - [P3590](#)
- Environment Ontology - [P3859](#)
- Xenopus Anatomical Ontology - [P4495](#)
- Spider Ontology - [P4537](#)
- Protein Ontology - [P4926](#)
- BRENDA tissue / enzyme source - [P5501](#)
- Food Ontology - [P6767](#)
- Gazetteer - [P6778](#)
- Pathway Ontology - [P7333](#)
- Cell Ontology - [P7963](#)
- Hymenoptera Anatomy Ontology - [P9356](#)
- Gene Ontology - [P686](#)
- Disease Ontology - [P699](#)
- Uberon - [P1554](#)
- Cell line Ontology - [P2158](#)
- Evidence & Conclusion Ontology - [P3811](#)
- Human Phenotype Ontology - [P3841](#)
- Mondo Disease Ontology - [P5270](#)

23 dedicated relations for mappings +

- exact match - [P2888](#)
- narrow match - [P3950](#)
- broad match - [P4900](#)

Properties are community created
So not direct mapping - mix N match

How mappings are done? - Manual curation on Mix'n'match

A	B	C
id	name	description
VO_0000000	part of pathogen organism in vaccine	part of organism in vaccine' is a pathogen component in vaccine that contains only a part of the pathogen organism instead of the whole organism.
VO_0000001	vaccine	A vaccine is a processed material with the function that when administered, it prevents or ameliorates a disorder in a target organism by inducing or modifying adaptive immune responses specific to the antigens in the vaccine.
VO_0000002	vaccination	a process of administering substance in vivo that involves in adding a vaccine into a host (e.g., human) in vivo with the intent to invoke a protective or therapeutic adaptive immune response.
VO_0000003	ACAM2000	
VO_0000004	ActHIB	
VO_0000005	Adacel	
VO_0000006	Afluria	
VO_0000007	Attenuvax	
VO_0000008	vaccine-induced lymphocyte activation	
VO_0000009	vaccine-induced B cell antigen processing and presentation	

Vaccine Ontology ID

IDs in the Vaccine Ontology, an OBO Foundry ontology

This catalog appears to be empty, maybe the initial scraping is still running

Imported by user TiagoLubiana | [Refresh](#)

<https://mix-n-match.toolforge.org/#/catalog/5849>

https://github.com/lubianat/obo_to_mixnmatch

Wikidata

A knowledge graph for the earth sciences? ~~Ontology~~

https://docs.google.com/presentation/d/1DeYDfyla8ILk4MGujQa0UL-L2fwwl0stsaSoLbCgT58/edit#slide=id.g10d3f523668_0_96

Charles Vardeman, Albin Larsson and Jan Ainali, Jane Wyngaard & Gary Berg-Cross (Semantic & Harmonization Perspectives)

Wikidata: A knowledge graph for the earth sciences?

Semantic & Harmonization Perspectives
Gary Berg-Cross, Ontolog Forum



For Earth Science Knowledge Graphs (KGs) Data *Vocabularies (including Wikidata's)* should be FAIR & support Interoperability

KG applications provide enhanced (semantic) functions such as search, browsing, improved integrated queries, including data of diverse modalities and from diverse sources.

Wikidata provides an opportunity for Earth Science KGs:

Many data and semantic resources are available for earth science related domains (wildfires, soil...)

BUT they are

- Scattered
- Often shallowly done for a particular purpose and
 - thus somewhat inconsistent and hard to reuse for integrated or new purposes.

Wikidata:List of properties/science/natural science

671 biology-related properties, what about Earth Sciences?

ANPR registration number	P10094	External identifier	person to practise a health profession within Australia	John Greenwood <AHPRA registration number> MED0001048723
FRACS Find a Surgeon profile ID	P10095	External identifier	identifier of a Fellow of the Royal Australasian College of Surgeons in the Find a Surgeon database	Gabrielle <Find a Surgeon profile ID> 4d1-4e6b-8cab
Indonesian Hospital ID	P10115	External identifier	hospital id from Board for Development and Empowerment Human Resources of Health, Indonesian Ministry of Health	RS Umum Meutia <Indonesian Hospital ID> 1171145

suggests the existence of

may prevent

Wikidata property related to mathematics [\[edit \]](#)

Wikidata property related to geology [\[edit \]](#)

Wikidata property related to astronomy [\[edit \]](#)

But established **Earth Science** vocabularies exist, e.g. SWEET, EnvO, GCMD, CF Metadata, SVO, etc....

Title	ID	Data type	Description	Examples	Inverse
constellation	P59	Item	the area of the celestial sphere of which the subject is a part (from a scientific standpoint, not an astrological one)	Andromeda Galaxy <constellation> Andromeda	-
site of astronomical discovery	P65	Item	the place where an astronomical object was discovered (observatory, satellite)	(5611) 1943 DL <site of astronomical discovery> Iso-Heikkilä Observatory	-
minor planet group	P196	Item	is in grouping of minor planets according to similar orbital characteristics	10 Hygiea <minor planet group> asteroid belt	-
spectral class	P215	String	spectral class of an astronomical object	Sun <spectral class> G2V	-
galaxy morphological	P223	String	galaxy morphological classification code	Andromeda Galaxy <galaxy morphological	-

Some of what Wikidata knows about Permafrost

permafrost (Q179918)

soil frozen permanently (practically for a duration of at least two years)

▼ In more languages

Configure

Language	Label	Description	Also known as
English	permafrost	soil frozen permanently (practically for a duration of at least two years)	
Spanish	permafrost	capa de suelo permanentemente congelado	Hielos perpetuos

Nothing about constituents etc.

Permafrost is an instance of soil type

But also

subclass of soil

Many identifiers

E/g. LoC authority ID
sh2005001377

UNESCO Thesaurus

Content language English ▼

Alphabetical

Hierarchy

Groups

Geophysics
├── Geochemistry
├── Geochronology
├── Geodesy
├── Geomagnetism
├── Glaciology
│ ├── Glaciers
│ └── **Permafrost**
├── Seismology
├── Terrestrial heat
└── Volcanology

Compared to Soil has fewer relations like “has quality – texture or pH”

Geophysics > Glaciology > Permafrost

PREFERRED TERM

Permafrost

BROADER CONCEPT

Glaciology

RELATED CONCEPTS

Frost

Ice

Polar regions

Soils

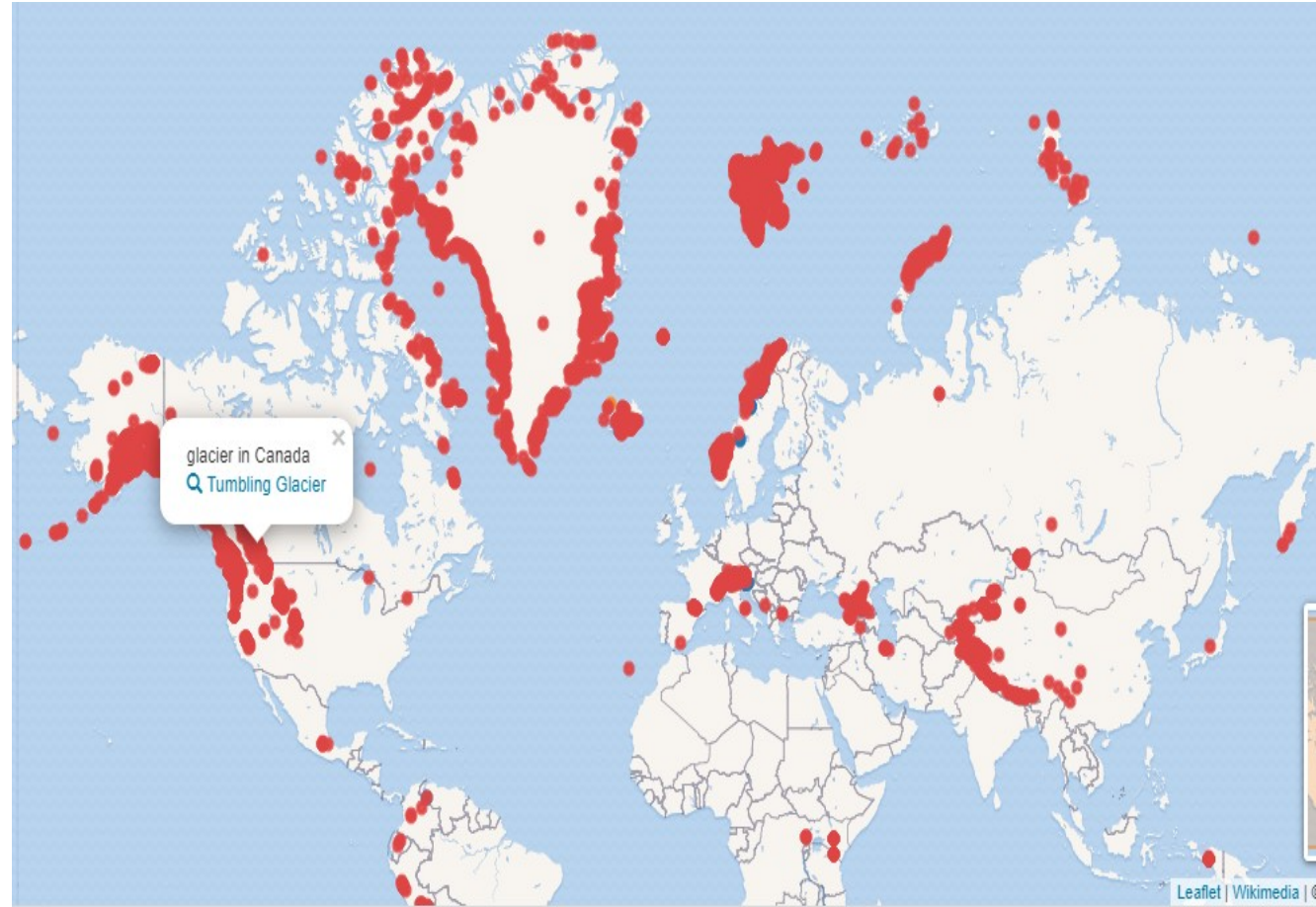
ENTRY TERMS

Frozen ground

BELONGS TO GROUP

Science > Earth sciences

Where Wikidata has info on Soil or Glaciers



Soil (Q36133) in Wikidata's Reasonator Tool



Other properties	
described by source	Brockhaus and Efron Encyclopedic Dictionary comprehensive multi-volume encyclopedia in Russian statement is subject of: ЭСБЕ / Почва
	Brockhaus and Efron Encyclopedic Dictionary comprehensive multi-volume encyclopedia in Russian statement is subject of: ЭСБЕ / Почва, в санитарном отношении
	Small Brockhaus and Efron Encyclopedic Dictionary Russian encyclopedic dictionary statement is subject of: МЭСБЕ / Почва
	Explanatory Dictionary of the Living Great Russian Language book statement is subject of: ТСД2 / Почва
	Granat Encyclopedic Dictionary seventh edition of Russian encyclopedic dictionary statement is subject of: ЭСГ/Почва
	Gujin Tushu Jicheng 10,000-volume Chinese encyclopedia completed in 1725 statement is subject of: 欽定古今圖書集成·方輿彙編·坤輿典·土部
has quality	soil fertility the ability of a soil to sustain agricultural plant growth sourcing circumstances : circa
	soil temperature temperature of the top soil
	soil pH measure of the acidity or alkalinity in soils
	redox potential measure of oxygen availability in soil
	soil texture property of a soil defined by the exact grain size distribution of its finer-grained anorganic matter
	Pore volume
different from	grunt granular material

Not axiomitized

Some Other Relations

studied by pedology study of soils in their natural environment

subclass of granular material (57)
part of pedosphere outermost layer of the Earth

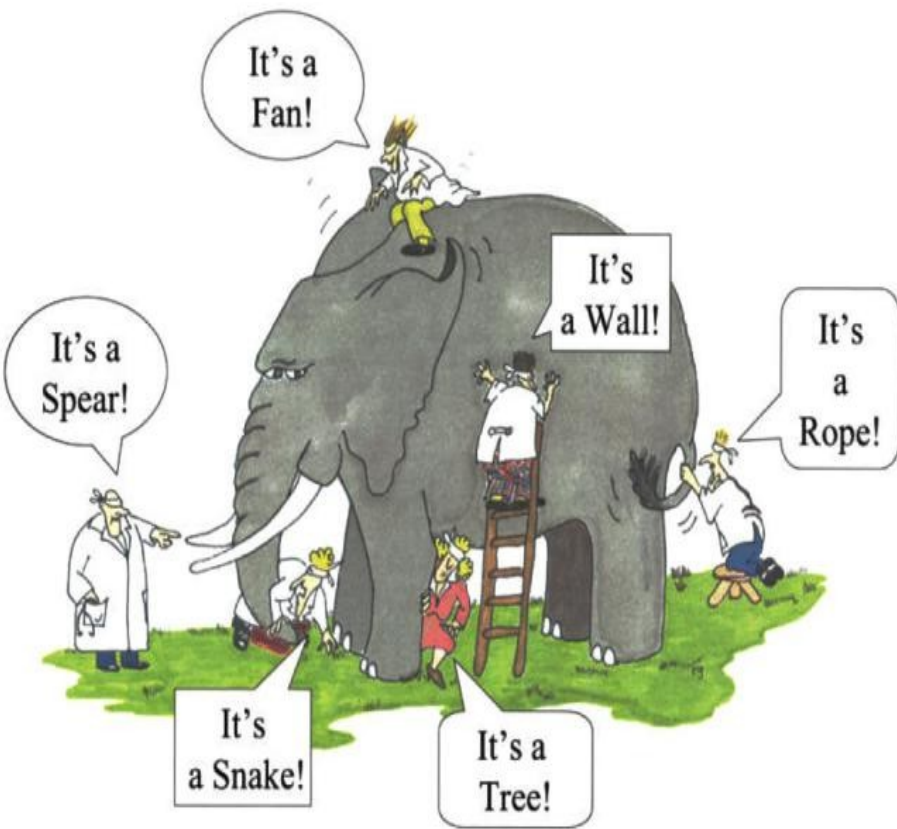
said to be the same as dirt natural surface of the ground

opposite of rock naturally occurring solid aggregate of minerals/mineraloids

criterion used : soil fertility

has part soil horizon soil layer whose physical characteristics differ from the layers above and beneath....

Backing up to the Big KG Picture Issues: Multiple Views of the Elephant



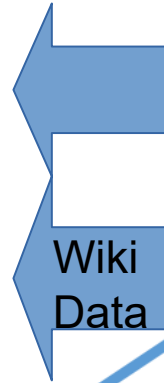
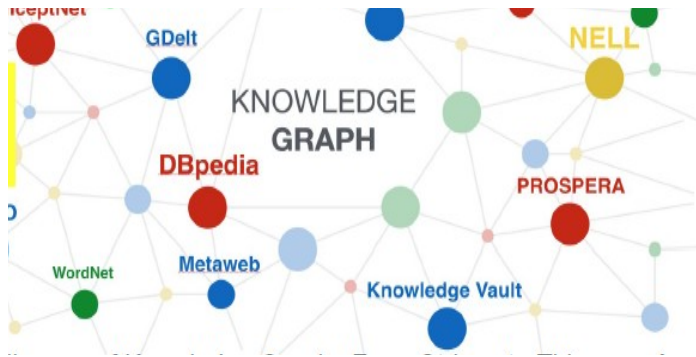
Integration - data available on the Web, of which Wikidata is one, has diverse formats so integrating data from different sources is still an open research area.

- Can KG efforts be integrated or are we building silos at a different level?
- Need common/coordinated models & representation, KR, tools...etc.

What are we harmonizing? Knowledge Graph sources.

Knowledge Graphs have many sources requiring Harmonizations from top to bottom

A rough illustration of the semantic ladder



Stronger semantics

Ontologies

Machine-readable knowledge

Taxonomies

Data models

Linked Data

Thesauri

Glossaries

Pretty much just strings

Controlled vocabularies

Weaker semantics

