Welcome to E2C2

- 2nd Tuesday of the month: Journal Club/Works in Progress/Seminar
  - Facilitated by Jeanette and Robbie
- 4th Tuesday of the month: T32 Interdisciplinary WorkGroups
- November 8th is Election Day so no meeting
- Please sign up to lead a slot **tinyurl.com/49828kak**

Facilitating data sharing in the Environmental Health Sciences

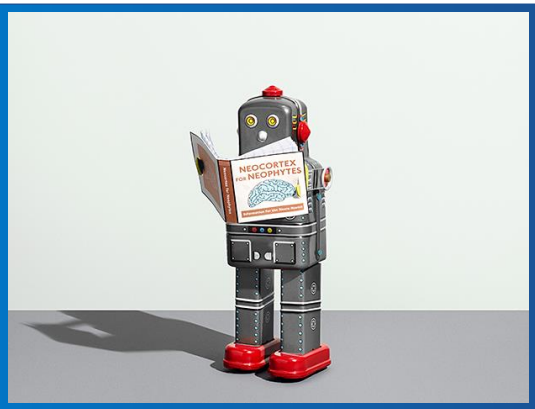
Using ontologies and standardized vocabularies to get the computers to do the hard part for us!

DATA

employee_id	first_name	last_name	nin	department_id
44	Simon	Martinez	HH 45 09 73 D	1
45	Thomas	Goldstein	SA 75 35 42 B	2
46	Eugene	Cornelsen	NE 22 63 82	2
47	Andrew	Petculescu	XY 29 87 61 A	1
48	Ruth	Stadick	MA 12 89 36 A	15
49	Barry	Scardelis	AT 20 73 18	2
50	Sidney	Hunter	HW 12 94 21 C	6
51	Jeffrey	Evans	LX 13 26 39 B	6
52	Doris	Berndt	YA 49 88 11 A	3
53	Diane	Eaton	BE 08 74 68 A	1
54	Bonnie	Hall	WW 53 77 68 A	15
55	Taylor	Li	ZE 55 22 80 B	1

DATA DICTIONARY (METADATA)

Column	Data Type	Description
employee_id	int	Primary key of a table
first_name	nvarchar(50)	Employee first name
last_name	nvarchar(50)	Employee last name
nin	nvarchar(15)	National Identification Number
position	nvarchar(50)	Current postion title, e.g. Secretary
department_id	int	Employee deparmtnet. Ref: Departmetns
gender	char(1)	M = Male, F = Female, Null = unknown
employment_start_date	date	Start date of employment in organization.
employment_end_date	date	Employment end date. Null if employee sti



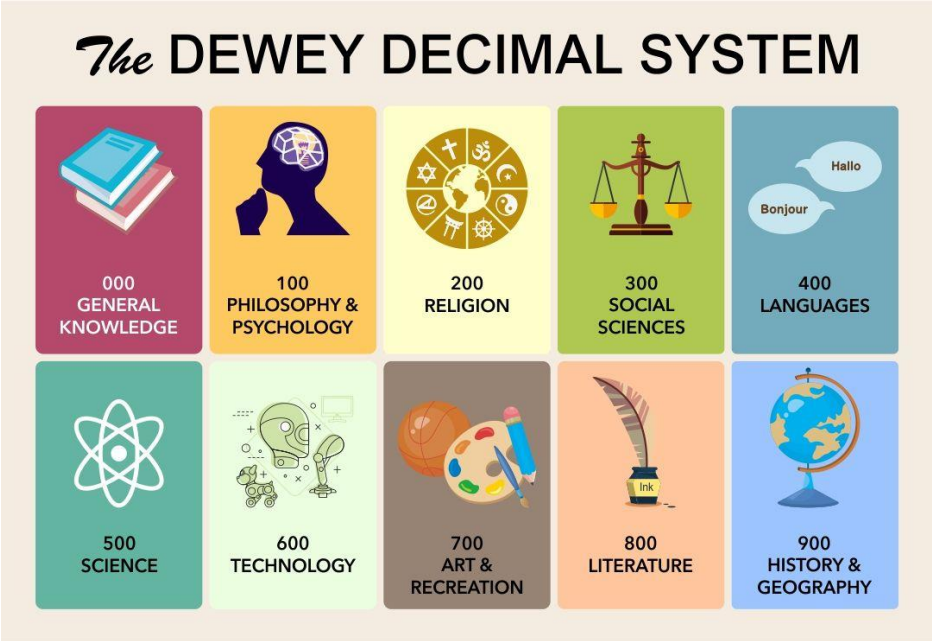
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Jeanette A Stingone PhD MPH
Assistant Professor, Department of Epidemiology
Data Science Translation and Engageent Group, HHEAR Data Center

Overview

- Recap of Ontologies and Standardized Vocabularies
- Rationale for using them to describe our data
- How to find tools and resources
- Hands-on Activity

Data Science for Better Data: Compiling and organizing our existing knowledge and resources



One of the first KOS's we're exposed to in school

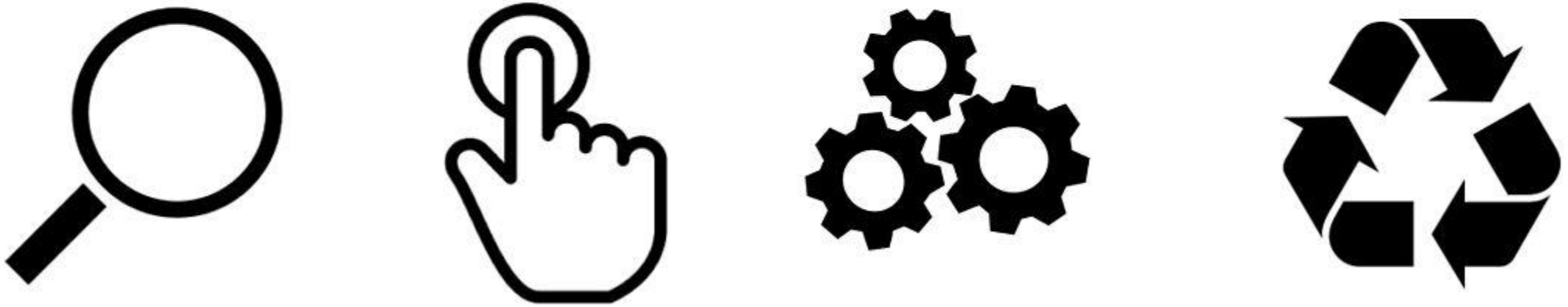
Hey Jarvis…. Can you look through the literature and see if you can find any potential biological mechanisms between pesticides and Parkinson's disease?

Hey Jarvis…. Can you find all of the existing data on pesticide exposure and Parkinson's disease funded by NIH in the last 15 years?

Hey Jarvis…. Can you harmonize data across these ten cohorts so I can conduct a pooled analysis?

Learning to Play FAIR with our Data

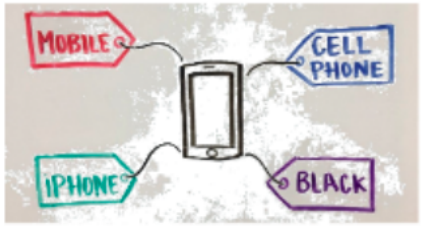
Findable Accessible Interoperable Reusable



Ontologies and Standardized Vocabularies

Free-text tags.

FOLKSONOMY



CONTROLLED LIST

List of pre-defined terms.
Improves consistency.

Pre-defined terms & synonyms.
Hierarchical relationships.
Improves consistency.
Allows for parent/child content relationships.

TAXONOMY

THESAURUS

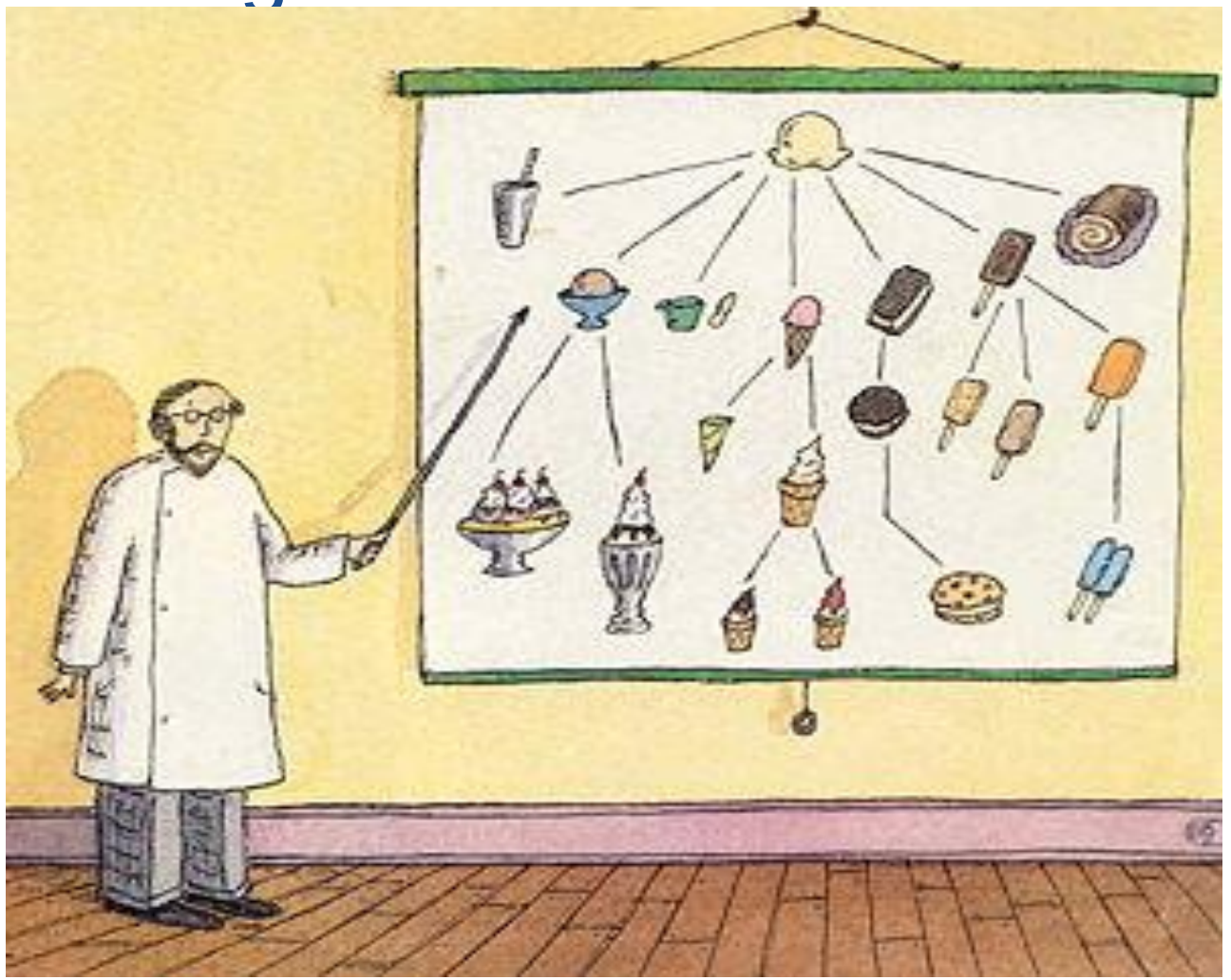
Pre-defined terms & synonyms.
Hierarchical relationships.
Associative ("related to") relationships.
Scope notes.
Increased expressiveness.

Scope notes.
Pre-defined classes & properties.
Expanded relationship types.
Increased expressiveness.
Inference.

ONTOLOGY

Ontologies to Represent our Data and Organize Our Knowledge

Ontology: a formal, explicit specification of a shared conceptualization (Tom Gruber)

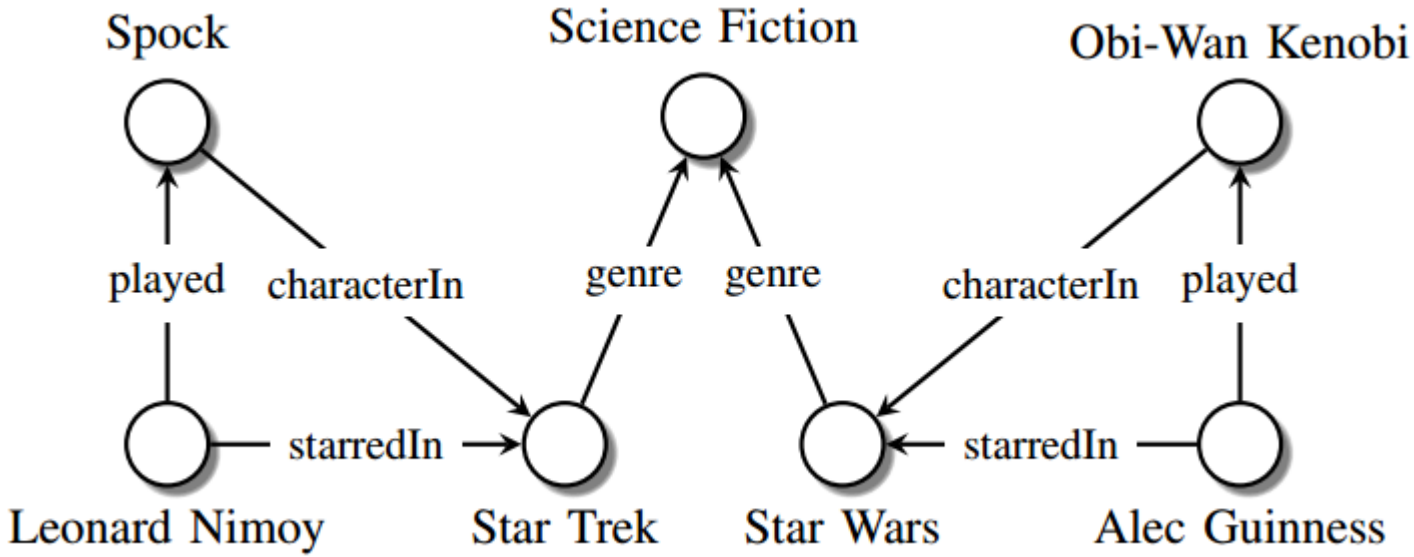


Common Language Enables Knowledge Graphs

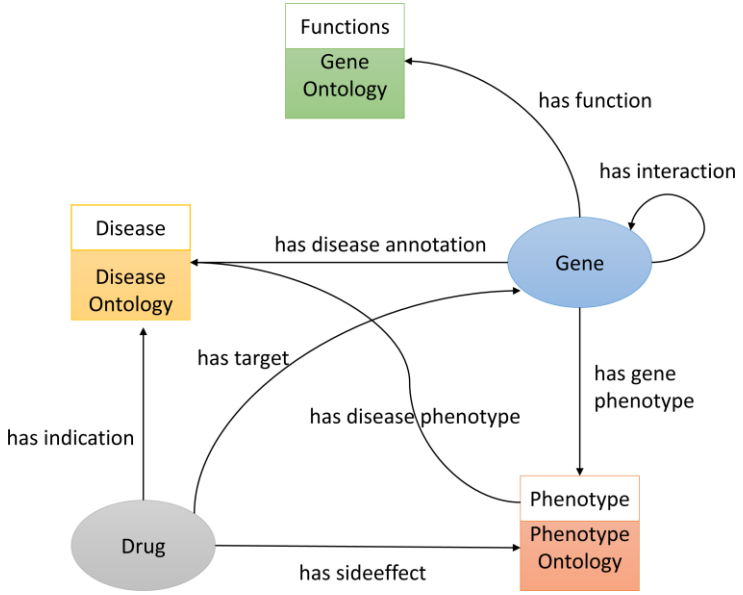
Genre	Movie
Science Fiction	Star Wars
Science Fiction	Star Trek

Movie	Cast
Star Trek	Leonard Nimoy
Star Wars	Alex Guinness

Actor	Character
Leonard Nimoy	Spock
Alex Guinness	Obi-Wan Kenobi



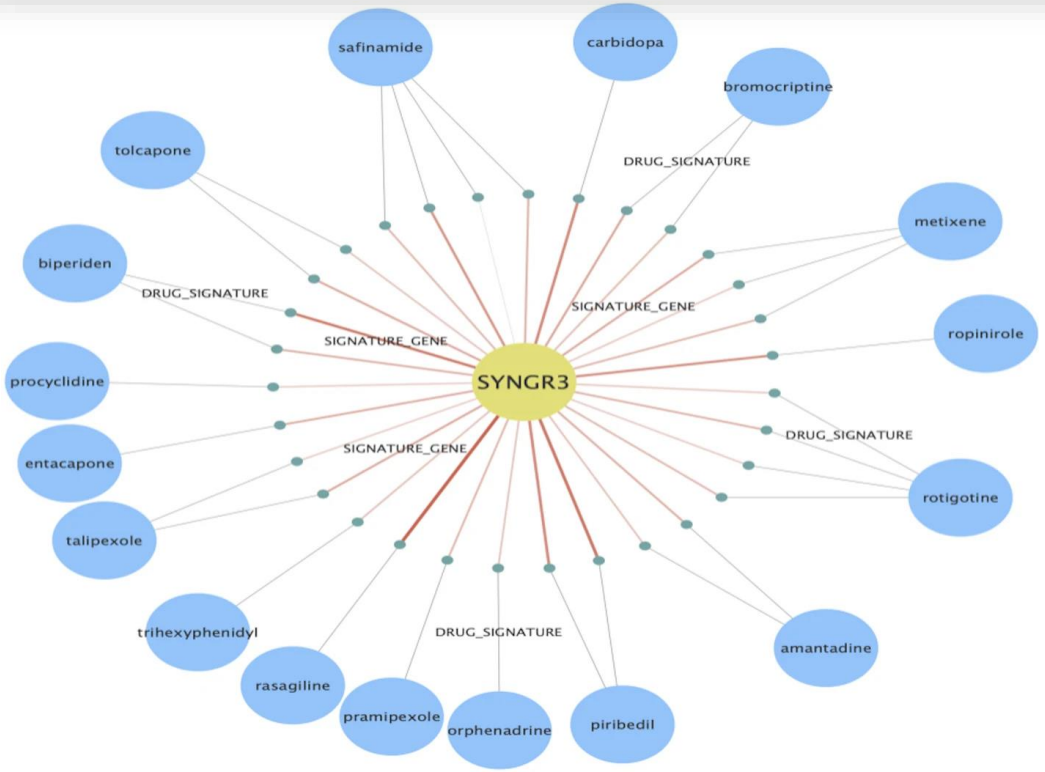
Using Ontologies to Build Knowledge Graphs



Source: doi.org/10.7717/peerj13061

Knowledge graph analytics platform with LINCS and IDG for Parkinson's disease target illumination

Jeremy J. Yang, Christopher R. Gessner, Joel L. Duerksen, Daniel Biber, Jessica L. Binder, Murat Ozturk, Brian Foote, Robin McEntire, Kyle Stirling, Ying Ding & David J. Wild



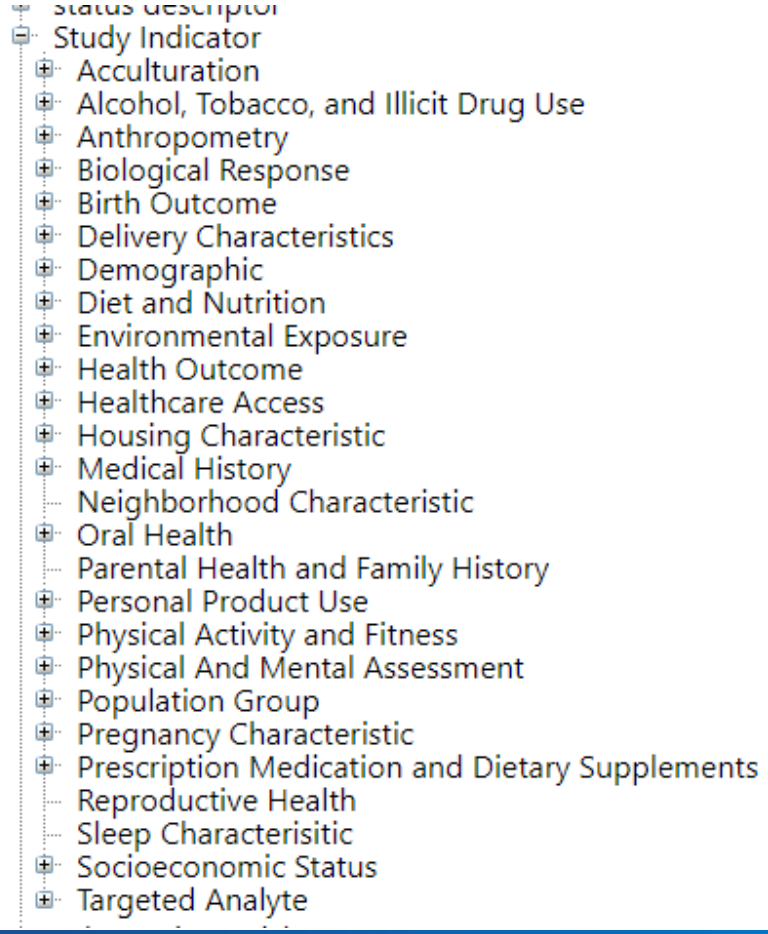
Source: doi.org/10.1186/s12859-021-04530-9

How to find ontologies and standard vocabularies in Environmental Health Sciences

- Clinical Data: Many standardized vocabularies/taxonomies used in clinical settings
  - LOINC (Logical Observation Identifiers Names and Codes)
  - ICD-10
  - RxNorm
- Research Data: Not as standard; Often many overlapping ontologies/vocabularies exist
- Different Resources to Search and Annotate Research Data
  - BioPortal
  - OntoBee
  - EMBL-EBI Ontology Lookup Service
  - GitHub - ISA-tools/OntoMaton: OntoMaton facilitates ontology search and tagging functionalities within Google Spreadsheets.

Cross-Ontology Resources: HHEAR ontology

https://bioportal.bioontology.org/ontologies/HHEAR/?p=summary



Activity: Find Ontology/Standard Vocab Terms for your Data

Use your own data/ data dictionary or use publicly available data posted here:

Step 1: Try to find an ontology term for each variable in your dataset. Add a column to your data/data dictionary to record the term that best matches your variable

Step 2: What are some limitations you notice when trying to map your variables to existing ontology/standard vocabulary terms?

We'll recap and discuss in the last 15 minutes of the session.