**How to Databases**

**Wikidata –**

Wikidata is a large database that stores their data in a Resource Description Framework (RDF) format. This format creates a network of all its datapoints, or items, where each individual item has a unique identifier. Items are connected by their *relationships* to other items, making it easy to find datapoints that share a particular feature or belong to a certain category. In short, items are stored in triples, composed of a **subject** (i.e. Item A), the **predicate** (i.e. proposed relationship), and the **object** (i.e. Item B). That is the basis of how RDF databases are structured and what allows users to explore the relationship between datapoints.

**Wikibase –**

Wikibase is an entity of Wikidata that stores the metadata items found on the database. This can include information such as references for triples, allowing research to validify connections they find. For our project, we wanted to find information about the exact units used for the dosages of each breast/prostate cancer medication found on Wikidata. This information was not readily available as an item, but was stored as metadata on Wikibase.

**Making Sense of Databases SPARQL –** **Sparql Protocol And RDF Query Language**

To find relationships between datapoints in online databases such as Wikidata, we need to write a program, called a query, at will allow users to access the data they need. SPARQL is a query language specifically used for Wikidata, as well as other compatible bases. As an example, we can create a SPAQRL query that looks for all datapoints within a certain category, such as **all biology related cancers** within the general category of **cancer**. Furthermore, we can adapt this SPARQL query to also fetch data from other databases, such as Wikibase, within the same query. Through this we could find our relevant medications on Wikidata and use that list of items to find the units stored in Wikibase.

A summary of how our SPAQRL query retrieves data is as follows:

**Visual Representation (d3.js)**

**Challenges with Wikidata/SPARQL – In depth look into query structure**

Our item (medication daily dosages) 🡪 Statements within that item on Wikibase 🡪 properties of those statements (ie units, values, references).

Array medications, applied to Wikibase to find the dosages 🡪 properties of medication