Everything is data

Database: an organized collection of data

Data model: a set of rules or guiding principles for how the data is organized

The most basic structure for a data model is a data table.

Data sharing – make sure the data is stored in a reproductive manner so that it can repeated by others

You should share data because it reduces the burden on participants, publications with data are cited more often, to allow new collaborations and more publications, to replicated/extend results, to be discovered and cited, etc.

We hare data via local databases and online databases. Can also use the cloud. Slide 18 = ways to share data. Need to pick one.

Version control = the management of changes to a variety of files. ie. it is a system that lets you save versions of all of your work. Records all changes and shows the differences, allows you to recover older versions, has multiple backups, allows you to add a description to each version, stores versions properties, can be released after a milestone (ex. publication), allows multiple people to work on files simultaneously, and you can see who changed what and when

Data visualization – figure out your intention, then chose the right visualization, then chose the right tool (program, ex. Lumira). Slide 26 shows what visualizations should be used for what.

Metadata = elements = title, description, creator, date, type (slide 30); values = descriptions or administrative information for each element of the metadata

Ontologies 🡪 triples. Rdf triple: includes subject, predicate (relationship or object property) and object

Ontology: a representation of the knowledge of concepts and terms within a domain. These typically include terms and the relationships between the terms.

Interoperability: we need data standards for the purpose of interoperability. So that data can be used by multiple people.

Relational database: a go-to application for data store. It includes data modelling because data modelling helps us understand and manage data.

Relational database management systems: have structure, consistency, integrity, efficiency, types, performance, and are easy to *create update and delete*

SQL: structured query language

Includes:

DDL: data definition language used to create, alter, and delete tables

DML: data manipulation language used to insert, update, and delete data

Primary keys: 1+ columns whose data is used to uniquely identify each row in the table

Foreign keys: 1+ columns in a table that refer to a primary key in another table

Research Data Management: exists because major funding agencies worldwide are pushing researchers to archive data and make it available.

Research Data: primary sources that support research and that can be used as evidence to validate findings and results.

FAIR = findable, accessible, interoperable, and reuseable

Copyright: protects all original work that is created and capture in a tangible medium of expression

Copyleft; a general method for making work free, and requiring all modified and extended version of the program to be free