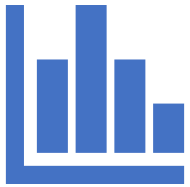
General Data Management Tips

Trainers: Dwane Young, Adam Griggs, Laura Shumway



Overview

Data Management Principles – Dwane

- Data Structure
- Allowable Values
- Who Needs the Data?
- Other considerations

Tools: What to Use and Why – Adam

- Excel vs. Access vs. Enterprise Databases

Becoming a ‘Data Slugger’ - Dwane

- Choosing a tool
- Demo using Access and Excel to answer questions
- Other more advanced capability

Data Structure: Items to focus on

Concepts: What are the concepts that define your data?

Attributes: What are the attributes of those concepts?

Relationships: How do those concepts relate with each other?

Rules: What are the rules that define which attributes are required and under which conditions and which are not?

Some Process Thoughts

- **Step 1: "What are the concepts?"**
  - Leverage Subject Matter Experts
  - Read the governing regulation or guidance
  - Talk to practitioners
  - Create a 'mental model' of how this 'idea' exists in the natural world
- **Step 2: Draw those concepts out on paper** or whatever works for you, but there's no need to try to learn a new technology to do this step. It's all about letting your brain work freely
- As you draw out the concepts ask: "What are the attributes of these concepts?"
  - What 'attributes' are common across concepts?

Example

- Monitoring Location (Concept) has several attributes:
  - Name
  - Water Type
  - Location
    - Latitude/Longitude
    - Metadata

Exercise

With a piece of paper, start identifying the ‘concepts’ and ‘attributes’ of a tree

- Visualize the tree in your mind; draw it if you need to
- There is no wrong answer at this phase
- **Hint: Abstraction is ok – sometimes we need to stay out of the fine details**
- **Hint: Try to stay away from the ‘Functions’ of a tree**
- **Hint: Think about what’s necessary to describe a tree to someone else that makes a ‘tree’ a ‘tree’**

Relationships and Rules

- **Step 3: “What are the Relationships between concepts”**
  - This can at times be the most difficult. Is one concept the ‘parent’ of another concept. Are they ‘siblings’? How many iterations of one concept can be related with one iteration of another? What’s the ‘root’ concept?
- **Step 4: “What are the Rules for concepts and attributes?”**
  - Which ones are required?
  - Are there special conditions under which they would or wouldn’t be required?

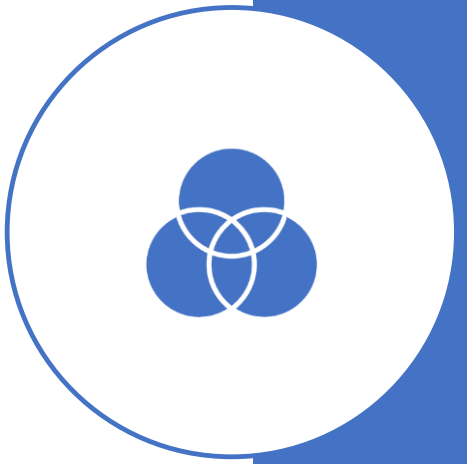
Exercise

Using our prior exercise with our tree: Define the relationships between your concepts and on a separate piece of paper start defining rules.

- You can define the relationship just by drawing lines

Allowable Values

- What are the ‘terms’ that you want to control. Typically, these should be the concepts or attributes that you’ll want to query by later
- From a data management perspective:
  - Dissolved Oxygen <> DO <> Oxygen, Dissolved <> DISSOLVED OXYGEN
  - However, left on their own, people will use all these combinations, plus like this: Disslved Oxygene
- You should set up either ‘rules’ or ‘Look-up Tables’ that define what terms can be used
- These terms can then be enforced via relationships



Who Needs The Data

How you approach a data management problem will be determined largely by *who* needs to enter, access, and use the data

If you're the only user, you have less constraints, just make sure that you do it in a way that it could be passed on to someone else when you're gone

If there are other users involved, consider what tools they're comfortable with and how they consume information.

You may use one tool for data management and another for sharing/visualization – THAT'S TOTALLY OK!

Other Considerations

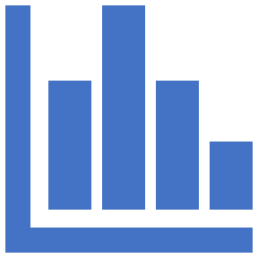
ARCHIVING

SECURITY

INTEGRITY

QUESTIONS?

Data Management Tools

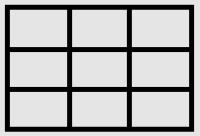


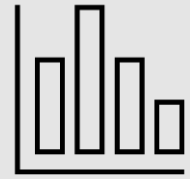
Spreadsheets

Manage datasets (not really a database)

Very useful for manipulating, analyzing, organizing one dataset at a time

Expandable – Macros, external data, XML, data visualization

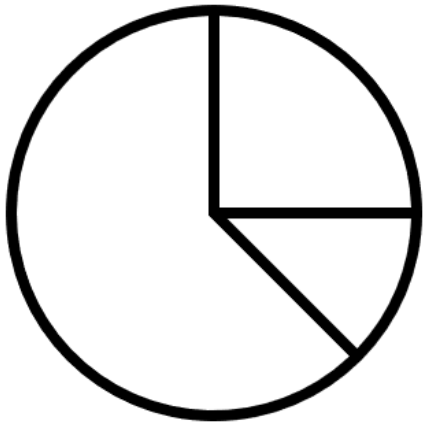


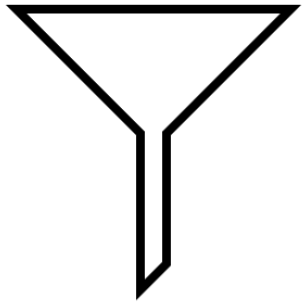


Useful Spreadsheet Functions

- Find and Replace
- Text-to-Columns
- Rapid sorting/filtering
- Analysis
- Data Entry Templates
- Publish/share a dataset
- Copy, paste, drag easily

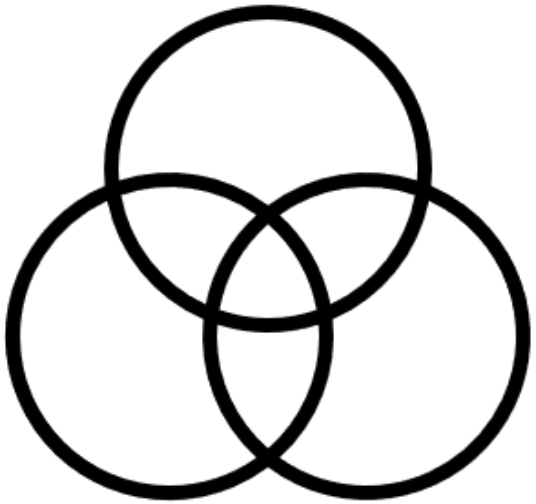
- Ex: Excel, Googlesheets, etc

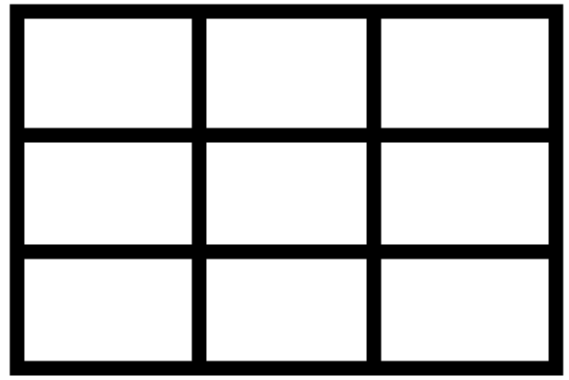


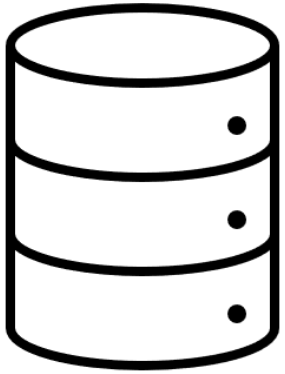


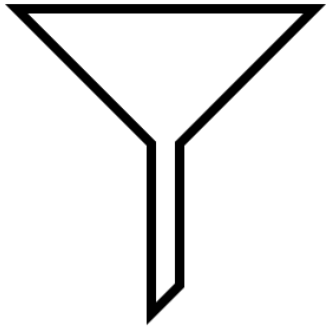
Relational Databases (light)

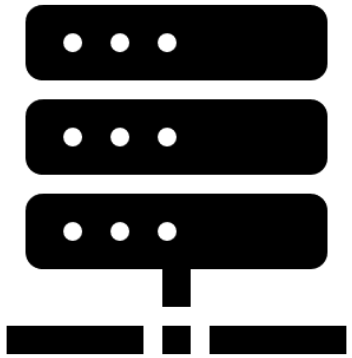
- MS Access (primarily)
- Entry-level database management
- Allows for the management of multiple tables of related data
- Connect, query, filter, update, or append data
- Ensure integrity of data quality/relationships
- Allows for front-end “forms” or “reports” or views of the data











What do we mean by relational?

Data of different types are each managed in their own table

We establish relationships between certain pieces of information in the tables

The related pieces of information are often (but not always) ID or “key” fields

This allows for more detailed information to be stored in separate tables, allowing for useful queries of the database

Ex. This is what allows users to query across place, time, program, and result type in the WQP

All Organizational Data

Org ID	Org Name
001	Minnesota Depart
002	Kentucky Departm
003	University of Was

All Project Data

Org ID	Prjct ID	Project Name
001	CYAN	Cyano Monitoring
001	PRB	Probabalistic mon
001	TRND	Trends Program

All Location Data

Org ID	Loc ID	Location Name
001	DBL	Deep Blue Lake
001	MR	Muddy Run
001	FC	Fishing Creek

Activities & Results Data & Metadata

Org ID	Proj ID	Loc ID	Result ID	Sample Date	Characteristic	Value	Unit
001	PRB	MR	001	5/3/2012	pH	7.6	SU
001	PRB	MR	002	4/17/2016	Temperature	19.1	Deg C
001	PRB	MR	003	4/23/2020	Conductivity	236	µS/cm

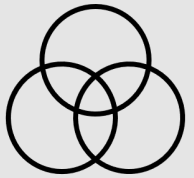
Relational Databases (Heavy)

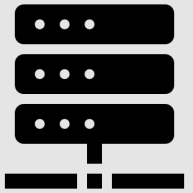
Enterprise

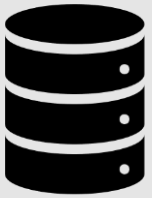
These are fully customized Enterprise database solutions usually built in Oracle, SQL Server, or PostGres, as examples.

These systems are typically operated from a server or more commonly these days, in the Cloud.

These are typically built by developers, at some cost and may include front-ends for staff to access and manipulate.







Selecting a data management tool

Who needs to manage the data?

What resources are available?

How much data needs to be managed?

How often will you need to complete this task?

	Spreadsheets	**Relational dB lite**	**Enterprise dB**
Data Entry	√	√ (Can add forms)	Requires forms
Relational data	Some (Power Pivot)	√	√
Run queries		√	√
Manual data fixes	√	√	Dev req'd
Expertise required	Low	Medium	High
Costs required	Low	Low-Medium	Med-High
Ensure Data Integrity	Some	√	√

QUESTIONS?

Becoming a ‘Data Slugger’

- So… You need to answer a question, you have some data, what do you do?
- Assuming you have the data (getting the data is a whole other training), the first step is to decide which tool to use?
- Will you need to answer the question again later?

DEMO: Using MS Access

- Linking Tables
- Outer Joins
- Using If-Then Statements
- Group By
- Finding Erroneous Values
- Manipulating Data – Make Tables, Append Data, Updates, Deletes

DEMO: More Advanced Capability

WORKING WITH XML

USING EXCEL AND ACCESS TOGETHER

EXCEL PIVOT TABLES AND SLIDERS

QUESTIONS?