Attribution Key

for more information see: http://open.umich.edu/wiki/AttributionPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV   **Public Domain – Government**: Works that are produced by the U.S. Government. (17 USC § 105)

PD-EXP   **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF   **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

ZERO   **Creative Commons – Zero Waiver**

BY   **Creative Commons – Attribution License**

BY-SA   **Creative Commons – Attribution Share Alike License**

BY-NC   **Creative Commons – Attribution Noncommercial License**

BY-NC-SA   **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL   **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL   **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

FAIR USE   **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

Case Study - HRWC

Site Survey Data Description: HRWC divided the county up into numbered bioreserves which were then combined (geographically) with parcels. They then use this information to try to contact property owners, get permission to survey the property, and then record the results.

Problem: They have two different databases, one for the addresses and contact information and one for the survey information. The two databases do not interact very well.

Section One

What do the IDs represent?

ID's are sequential, but unique ID is linkin records.. this seems like a duplication. It would be good to document what each your IDs represents and perhaps use the same name of ID throughout all your tables (ie. unique ID in wetlands survey table and unique_ID bioreserve in the assessment table... are they the same?)
uniqID - Owner ID

Are the tables flat?

Not all. Some are as flat as can be.

Has uniqueness been maintained where necessary?

No
Perhaps list by address to get a unique ID?

Is the data clean?

parcel numbers needs to be entered the same (assessment table stacks to parcel numbers vs parcel table includes a space between the parcel letter and number → "E -05 (parcel) vs E-05 (assessment)"
- parcel numbers should match the format used by the respective Counties/cvts
- seems like HRWC should have lookup access to all parcel numbers in the respective jurisdictions to validate that field

-Wetlands Survey Table: Time_at_site is unstandardized, ie 2.5, 03:30, 10-12+. Needs to be entered the same. (use cell format to indicate the type of standard format).
-Missing data across tables
-multiple parcel #s in some cells

No, there are no consistencies in the data entry values.
Time at site variable is inconsistent.

Dates are entered in differently, ie. 2008 or 08

What questions would you ask the organization to further elucidate the current structure?

*Why is it important to include the parcel in every table?
*Have you thought of giving parcel it's own table with a unique identifier so that it can be easier to maintain one central location of this info. (this is from a feeling that you are trying to find a way to pre-populate the parcel information, but don't know how to link the tables so it pre-populates so instead the data is being duplicated and manually entered → both time consuming and increases your chance of entered data incorrectly).
*It seems you are looking for a dashboard / one stop shop that compiles all the info. If this is the case, you can streamline the tables to only include data that is pertinent to that table and have it populate on a dynamic dashboard. So this would mean multiple smaller tables that give a unique identifier to the important pieces (parcel, clients, eco_unit, bioreserve, etc).. sometimes used as to validate your data (ie. eco_unit) but ultimately to link all the data.
*What is the base unit/field for your record (what makes one record different from another?)

Section Two

Try to think through how the tables could be rearranged and data entry could be better controlled.
- ## How would you transform the data to the new structure?
- ## Does the new structure lend itself to analysis?
- ## Does it maintain flatness and uniqueness?

Group Discussion:

- What is baseline of data? We thought it was parcel data at first but realized this can have multiple attributes.
- There are some parcels that span bioreserves.
- Wetland units: multiple in parcels.
- Were there any particular problems or issues with how the data is recorded or in terms of flatness or uniqueness?
  - The way the parcel tables are named: township doesn't necessarily line up with site names.
  - Data isn't completely clean.
- Assessment table: 21 and 22 (Sanford vs Sanford2ndVisit): what is the point of this record? To add new parcel? To show that an assessment was done? Not clear. As a result property owner field name gets nonstandardized.
- What is the purpose for why they are collecting the data? What are their intentions?
  - In order to assess which bioreserves they are most interested in surveying, then to record results of survey. Priorities of assessing bioreserves.
- Wetlands Survey Table does not have a standardized "Time_at_site" input.

In parcel table, remove Sitename (is this a concatenation from township, unique ID, and property owner?).. seems unnecessary and repetitive once a proper unique ID is set up (this also allows for more error due to spelling errors and duplication). If you are concerned about human readability in your IDs, then you can use a lookup to change the ID to something readable. Have linked ID's in the different tables titled the same (i.e. unique_ID bioreserve in assessment table and call it the same in the parcel table).

all your dates and time in all your tables should be formatted exactly the same (use table format to specific type of format)

clean up your table structure... break out data into smaller tables (based by properties you are looking for in that table that make it completely different from any other table) and link them into a table that is more of a report and pulls info from your various tables.

is there an exisiting database that you can pull data from? (i.e. parcel database by county records that you can link to to pull address?)

Time should be cleaned up to some standard way.

Eliminate all records with missing "parcel number" info, as they provide no further info.

provide a data entry form to help regulate your entry. this form would have tool tips and suggestions on what type and format of data is valid (ie. Enter Site Visit Date: mm/dd/yy). data validation can further be enforced with drop-down lists, multiple-choice buttons (ie. wetlands A, B, C),, etc.

Think about what you want from your data and then create ways to pull data to create a report (pivot table) as opposed to creating a new table that holds all this info.

There are many tables that include duplicate data from other fields, which could be removed or restructures

Create new table called Owners
- ○ Fields: Owner_id (unique), owner_type (= individual or corporation), and name

Modify Assessment Table
- ● Remove Unique ID, township, property_owner, wetland sheets, grassland sheets
- ● Rename ID to "assessment_id"
- ● Limit "parcel" field to 1 value per cell

Modify Parcel Table
- ○ Remove Sitename
- ○ Rename "uniqueID" to bioreserve_id
- ○ Rename OwnerName to "owner_id"

Modify Wetlands Survey Table
- ○ Remove Unique ID, Township, Property Owner
- ○ Replace "Date" column with "assessment_id"
- ● Limit "eco_unit" field to 1 value per cell. Rename to "wetland_unit" and remove "wetland" from field value (e.g. "Wetland" A becomes "A")

Create new table called Bioreserve
- ○ Fields: bioreserve_id (unique), others that are relevant to Huron River Watershed Council?