



Database Development: Best Practices Case Study

Jeffrey D. Campbell, Ph.D.

Center for Urban Environmental Research
and Education

University of Maryland, Baltimore County

Originally presented at the Maryland Water Monitoring Council (MWMC) Annual Conference in December 2008. At that time, I was a Research Scientist in the Center for Urban Environmental Research and Education and Chair of the MWMC Information Management Committee.

Slides have been updated and notes were added April 2012.

Current contact information: jeff.campbell@EnvironmentalInformatics.com
www.EnvironmentalInformatics.com

Presentation and notes Copyright 2008-2012, Jeffrey D. Campbell
All Rights Reserved.

Agenda

- Case overview: Fish Survey
- Spreadsheet approach
- Database benefits
- Database approach
- Conclusions

Case Overview: Fish Survey

Objective: Long term monitoring

Citizen scientists – simple field identification key

Variable sites



Fish were caught by seine or dip net, identified, measured and released unharmed.

Case Overview: Fish Survey

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: 0High _____ 0Low _____ Water Temp: _____ Air Temp: _____

Weather Conditions:

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used:

This is a copy of the data sheet used in the field. It seems to be fairly straight forward.

There are some weaknesses to the design:

How are volunteers identified? – first and last name, initial and last name, first name only? How many can there be? How will this information be used?

Probably not a big issue, but time is assumed to be local time (standard or daylight saving) as appropriate. Could become important if later correlated to data such as National Weather Service data that is archived using UTC (Universal Coordinated Time, also known as Greenwich Mean Time).

How is Water Temp measured? In the smaller streams, the water temperature at a given location is fairly uniform, but in the deeper pond, and particularly the river, the temperature can vary greatly with depth.

Weather Conditions is open ended. The original intention seems to be sky conditions (cloudy, sunny, drizzle, rain, etc), but someone could enter “humid” or even less descriptively “sticky”.

The seining method description is entirely open ended. In fact, the fish survey research protocol specified quarter sweeps with the 10m seine, so a check box confirming this was done would have been sufficient.

Case Overview: Fish Survey

Site _____ Date _____

Volunteers _____

Species (Common Name)	Number captured Use hatch marks (ex: IIII for 4 individuals)	Snout-to-tip-of tail length of each fish measured (cm)	Lesions, parasites, sex, other comments

This is the back side of the data sheet.

Site, Date, and Volunteers is redundant with the other side. What happens if the values do not match??? On the other hand, since this is the side where most of the data is actually recorded, having the Site and Date repeated is far better than risking that it is not recorded at all.

The sex of fish is only determined for species where male and females are easily distinguished by coloration, shape, etc. It should also be noted that the sex of juvenile fish often cannot be determined. For very young fish, species identification can be difficult.

Case Overview: Fish Survey

Actual Data

Site FARM Date July 06
Volunteers 10


Fish Survey

Volunteers (10)
Site FARM
Date July 2006 Time: Start 10:20 End 12:00
Seine mesh size 1/4" Seine length 10m
Tide: High Low Water Temp: _____ Air Temp: _____
Weather Conditions: Sky is Wnd S f 2 Wm

Seine Hauls

#	Time haul	Minutes haul	Time haul	Minutes haul	#	Time haul	Minutes haul	#	Time haul	Minutes haul	#	Time haul	Minutes haul	#
1	10:20	10m	2:30m	10m	11			16			21			
2	10:40	10m	7:45m	10m	12			17			22			
3	11:00	10m	8:15m	10m	13			18			23			
4	11:20	5m	10:00m	10m	14			19			24			
5	11:40	10m	10:00m	10m	15			20			25			

Describe seining method used:



Species (Common Name) | Number captured (Use hatch marks for 4 individuals) | Snout-to-tip-of-tail length of each fish measured (cm) | Lesions, parasites, sex, other comments

ALEWIFE	(7)	40, 36	
HOGCHAEKER	(1)	74	
YELLOW PERCH	(3)	12, 17, 6, 5, 3	
WHITE PERCH	(3)	12, 8, 12, 12, 11, 13, 14, 5	
BROOK SILVERSID	(3)	10, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	
BROOK SILVERSID	(1)	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	
BROOK SILVERSID	(1)	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	
MUMMICHOG	(2)	72, 57, 55	
MUMMICHOG	(2)	14, 34, 51, 40, 72, 20, 20, 20	
E. SILVER SID	(1)	5, 5	
WINGLATA LOVER	(1)	46	
BUMP KINDEEP	(2)	37, 81	
LARGE MOUTH BASS	(1)	40, 59, 42, 40, 3	
Atlantic HERRING	(1)	-	
Inland SILVERSID	(1)	72, 72, 72	
American SHAD	(1)	-	

Ok, so much for what the data form was designed to collect. This is a sample of one actually used. The Volunteer names have been blurred to preserve anonymity.

High Tide was simply checked. I believe the intention was to record the time of the high tide. How is the person in the field supposed to know that?

It can be determined later with better accuracy.

My expectation about “Weather Conditions” being vague is confirmed.

The seining method includes a diagram!!!! Who would have expected that. The diagram also explains the unusual data in the table above. Ugh. wow

On the other hand, the additional data is unusually detailed. Since I know the data, I can guess that this must have been the annual field trip for a graduate field method course. However, this would not be obvious to someone less familiar with the data.

Species, number captured, length looks straight forward. However, note that sex (using scientific symbols) is a subdivision of the table, not in the Notes column. At least one species name is spelled incorrectly. More species were found than there were rows in the table. Also, one fish (Needlefish spp.) is identified as NEW, meaning it was not in the local fish guide that includes fish found in earlier surveys. The “spp.” indicates that the species was not determined, just that it was some type of needlefish. That is a database challenge since this means that not every fish observed has a unique genus species binomial name. Another aspect that would be overlooked by a typical IT person is that mummichog and killifish are in the same genus, raising the strong possibility that young of the year may not be identifiable to the species level. For that matter, the term “young of the year” is likely be unknown to the IT person.

Spreadsheet Approach

- Conceptually easy
- One worksheet per year
- Summary worksheet

Having explained the fish survey and looked at the data sheets and one example of actually usage of the data sheets, we will move on to how to handle putting the data into a computer system.

In my experience, if there is any computerized version of ecological data, more often than not it is in a spread sheet. This was the case with the fish survey data, ... before I somewhat foolishly asked “How do you store the data?”

The existing spreadsheet had one page for each year. Each survey during the year was in a column. The fish species were in the rows.

There was a summary spreadsheet to total the number of each species caught each year.

Spreadsheet – 2002 page

Species	Two Run (Apr)	Pindell (May)	Farm (June)	Farm (July)	Otter Point (Jul)	Beaver Pond (Aug)	Species Total
Alewife				2			2
American Eel							0
Banded Killifish			70	25	8		103
Bay Anchovy							0
Black-nosed Dace	14	44					58
Blue Spotted Sunfish	14					24	14
Bluegill	4	1					5
Brown Bullhead Catfish	1					33	1
Chain Pickerel						1	0
Channel Catfish				3			3
Creek Chub	14	9					23
Creek Chubsucker	1	3				1	4
Dace/Creek Chub	1	7					8
E. Silvery Minnow						17	0
Eastern Mudminnow	4	5					9
Gizzard Shad				11			11
Golden Shiner	8	1				3	9
Hogchoker				1			1
Inland Silverside			14	6			20
Largemouth Bass							0
Least Brook Lamprey	4						4
Mosquitofish			2		3	23	5
Mummichog			23	21	12		56
Pumpkinseed		1	2	5			8

Looks like a fairly straight forward way to store the data.

Careful review indicates a row labeled “Dace/Creek Chub”. This is not a species, and not even the same genus. It is fish that could not be determined if they were black-nosed dace or creek chub. The juveniles can appear fairly similar, so it seems these were too small to distinguish.

Spreadsheet – summary page

Species	Total Number of Individuals Captured						
	1999	2000	2001	2002	2003	2004	2005
Alewife	-	26	11	2	2	-	4
American Eel	-	1	2	-	2	1	2
Banded Killifish	42	195	78	103	89	657	321
Bay Anchovy	22	-	10	-	-	19	10
Black-nosed Dace	-	52	57	58	45	92	19
Blue Spotted Sunfish	-	3	13	14	12	5	4
Bluegill	4	17	7	5	9	4	-
Brown Bullhead Catfish	-	17	10	1	13	-	5
Carp	-	-	-	-	-	5	-
Chain Pickerel	-	2	2	-	8	1	-
Channel Catfish	1	1	1	3	-	1	1
Creek Chub	6	16	55	23	12	28	5
Creek Chubsucker	-	5	2	4	5	2	4
E. Silvery Minnow	-	2	2	-	1	3	-
Eastern Mudminnow	1	9	4	9	1	-	10
Gizzard Shad	4	-	2	11	-	-	-
Golden Shiner	-	6	6	9	6	7	14
Hogchoker	1	2	1	1	2	1	-
Inland Silverside	26	39	32	20	-	29	41
Largemouth Bass	-	17	6	-	3	3	-
Least Brook Lamprey	-	-	2	4	-	-	-
Mosquitofish	5	-	5	5	-	15	39
Mummichog	14	108	198	56	5	78	36

The values in this summary page are from the “Species total” column on the annual pages.

The list seems to be alphabetical. The “Dace/creek chub” is not present. What happened to them in the analysis?

It also appears that carp were not found until 2004. I have to conclude that this page is manually updated annually (instead of just using formulas to refer to the appropriate column on the year pages. If it was not manually updated, every time a new species was found, all prior spreadsheets would have to be updated to indicate none were found in those years. Any spreadsheet design that relies upon prior unknowable data is at risk.

The spreadsheet approach is good at addressing the apparent question driving the design – “How many fish of each species were caught each year?”

There is no easy way to view/analyze year-to-year changes at a particular site (stream, river, pond...). There is no easy way to produce subtotals based on other characteristics such as native vs introduced fish, spawning habitat, or role in the food chain.

My suspicion is that the count of fish per species design was motivated by a state requirement to provide this information each year as part of the application process for the scientific collection permit needed to conduct the surveys.

Spreadsheet Approach

- Conceptually easy
- Summary spreadsheet
- One worksheet per year

BUT...

- Most of the data not entered
- Not easy to handle multiple entries (e.g. length, sex)
- Problem adding new species

Summary of using spreadsheets

Database Benefits

- Data stored in “tables”
 - Could think of them as linked, well designed spreadsheets
- Allow entry of multiple values
- Reduces repetitive data entry
- Reduces inconsistent data risk
- Greater analytical capabilities
- Worst case, export data to Excel or statistical package

Database Design Approach

- Identify “things” we need to keep data about
 - Determine the unique identifier (“key”)
- Identify relationships between things
- Retrieve data using these relationships

DB Approach – Identify “things”

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: θ High _____ θ Low _____ Water Temp: _____ Air Temp: _____

Weather Conditions:

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used:

The audience was asked to think about the different types of data things that can be identified from the data entry form.

DB Approach – Identify “things”

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: θHigh _____ θLow _____ Water Temp: _____ Air Temp: _____

Weather Conditions: _____

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used: _____

Survey – info for a particular expedition at a particular location

Survey
SurveyID
LocationID
Date
TimeStart
TimeEnd
SeineMeshSize
SeineLength
WaterTemp
AirTemp
Weather
TideHighTime
TideLowTime
AtHighTide
AtLowTide
Note
Year

One “Thing” is what I call survey. The Survey information is highlighted in yellow. This is the information for a particular expedition at a particular location on a particular date.

The new more nicely formatted column labeled “Survey” is the Microsoft Access representation of the database table that I created to hold the Survey data.

To be most compatible with other databases (MySQL, Oracle, SQL Server, etc), I use name that do not include spaces. I capitalize the first letter in each word to make it easier to read.

SurveyID is an arbitrary unique identifier for each survey. Most simply, it would just be a sequential number. Notice that this number is not on the original data sheet. It is used only internally in the database to relate the various “things” (tables) to each other.

The tide times became four data elements – times for high and low tide, and a yes/no field for whether or not the survey was at low tide or at high tide. (Note both are needed since it is possible a survey could be done at neither high nor low tide. Also for the non-tidal portion of streams, the tide times are irrelevant).

Year was included because for some of the historical data, the specific date was not recorded, only the year. Future data should just use the Date field.

DB Approach – Identify “things”

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: θHigh _____ θLow _____ Water Temp: _____ Air Temp: _____

Weather Conditions:

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used:

Survey

SeineHaul – details for each pull of the seine

SeineHaul

SurveyID

SeineSequence

Time

MetersSeined

The next “thing” is what I call SeineHaul. This corresponds to a single pull of the seine through the water. A survey may have none, one, or many SeineHauls preformed.

The SurveyID is used to link to the Survey data.

The SeineSequence is a sequential number (pre-printed on the data sheet). SurveyID and SeineHaul together uniquely identify every sweep of the seine recorded in the database.

DB Approach – Identify “things”

Fish Survey

Volunteers _____
Site _____
Date _____ **Time: Start** _____ **End** _____
Seine mesh size _____ **Seine length** _____
Tide: θHigh _____ **θLow** _____ **Water Temp:** _____ **Air Temp:** _____
Weather Conditions: _____
Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used: _____

Survey

SeineHaul

SurveyVolunteer
– a volunteer for
a survey



The next “thing” is information about volunteers who helped with the survey that day.

There was an existing database of volunteers that had name, email, and other information needed to manage all of the volunteer programs. I would consider that database to have the volunteer “thing” since it has the data specific to a volunteer regardless of what activities they do. Therefore, I named the table for volunteers at a particular survey to be SurveyVolunteer.

Since that information existed somewhere already, and the historical data for fish survey volunteers was not very consistent, I decided to keep SurveyVolunteer very simple. Again the SurveyID represents the particular one the data is for. The sequence is just a sequential number corresponding to the order in which the name was on the data sheet. VolunteerName is however it was recorded on the data sheet (first, last, initials, or any combination thereof).

VolunteerID was included as a placeholder to translate the varied names for a person (written on different fish survey data sheets) to be the same individual. Presumably the VolunteerID would match the ID for the person in the volunteer database.

DB Approach – Identify “things”

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: θHigh _____ θLow _____ Water Temp: _____ Air Temp: _____

Weather Conditions:

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used:

Survey –

Location

SeineHaul

Survey Volunteer

Volunteer – info about the person regardless of activity

This shows the Volunteer table that is actually in another database

DB Approach – Identify “things”

Fish Survey

Volunteers _____

Site _____

Date _____ Time: Start _____ End _____

Seine mesh size _____ Seine length _____

Tide: θHigh _____ θLow _____ Water Temp: _____ Air Temp: _____

Weather Conditions:

Seine Hauls

#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined	#	Time (sec)	Meters seined
1			6			11			16			21		
2			7			12			17			22		
3			8			13			18			23		
4			9			14			19			24		
5			10			15			20			25		

Describe seining method used: _____

Survey

Location – site details



SeineHaul –

Survey Volunteer

There is one more “thing” that was partially overlooked in the prior search for things represented on the data sheet.

The Site for the surveys were fairly consistent from year to year, with some being added or removed as research objectives changed.

In order to relate data from the same location collected during different surveys, it is useful to create a Location table. The minimum data is shown here with just a LocationID (again a sequential number or other unique identifier) and a name. Presumably additional information that is specific to the location could be added to this table. For example, latitude and longitude, water type – non-tidal stream, tidal, pond, etc.

DB Approach – Identify “things”

Site _____ Date _____

Volunteers _____

Species (Common Name)	Number captured Use hatch marks (ex: IIII for 4 individuals)	Snout-to- tip-of tail length of each fish measured (cm)	Lesions, parasites, sex, other comments

Fish Caught – details (count) of fish of a species that was caught

FishCaught
SurveyID
FishID
NumberCaught
Note
MaleCaught
FemaleCaught
JuvenileCaught

FishCaught provides the counts (and subcounts) for fish observed during the survey.

SurveyID remains the identifier for a survey at a particular site, date, and time

FishID corresponds to a fish species (see next slide for details).

NumberCaught is the total for that fish for that survey even if there are subcounts by sex. This simplifies analysis, particularly the annual count of each species that must be reported to the state.

Note is space for lesions, parasites and any other comments

MaleCaught, FemaleCaught, and JuvenileCaught are used for species where these are discernable and hence are optional.

There is a design issue with those three fields that is not clearly addressed. In a perfect world, the NumberCaught would equal the sum of MaleCaught plus FemaleCaught plus JuvenileCaught, unless sex is not determined in which case the three subcategories would all be zero. The screen for entry into the database could verify these counts. However, if the numbers on the data sheet do not total correctly, the data should be accepted by the system (and possibly flagged as questionable – add another field for the table for QA-QC) and not force the person entering the data to resolve the inconsistency before being able to save the data in the computer. I decided to not perform the verification since there were years of historical data that could not be easily corrected if a discrepancy was found. Furthermore, there was some inconsistency in which species had sex recorded and whether or not sex was actually determined.

DB Approach – Identify “things”

Site _____ Date _____

Volunteers _____

Species (Common Name)	Number captured Use hatch marks (ex: IIII for 4 individuals)	Snout-to-tip-of tail length of each fish measured (cm)	Lesions, parasites, sex, other comments

Fish Caught

FishLength – for each fish measured

FishLength
• SurveyID
• FishID
Sequence
Length

FishLength is a simple table to just relates the SurveyID, FishID, and a sequential number to record the length of a variable number of fish caught for a particular survey and fishID.

Creating this table is a better solution than adding some fixed number of length fields to the FishCaught table. No matter how many were added, there is the possibility of needing more. Additionally, by using the FishLength table, it is very easy to compute the minimum, maximum, mean, and standard deviation of fish lengths for a species for all (or one or a subset of)surveys. These calculations would be far more difficult to perform if there were many length fields in the FishCaught table.

Database experts will also recognize that FishLength correctly implements the 1 to many relationship between FishLength and FishCaught.

DB Approach – Identify “things”

Fish
FishID
CommonName
Family
Genus
Species
Introduced
IntroducedFrom
DorsalCount
DorsalType
TailType
Tolerant
Feeding Structure
LithophilicSpawne
GOI
CountSex
CountJuvenile
DefaultImage
SizeJBAverage
SizeMax

Fish Caught

FishLength

Fish – species
information

The Fish table contains information about fish that does not depend upon the fish being caught during a survey.

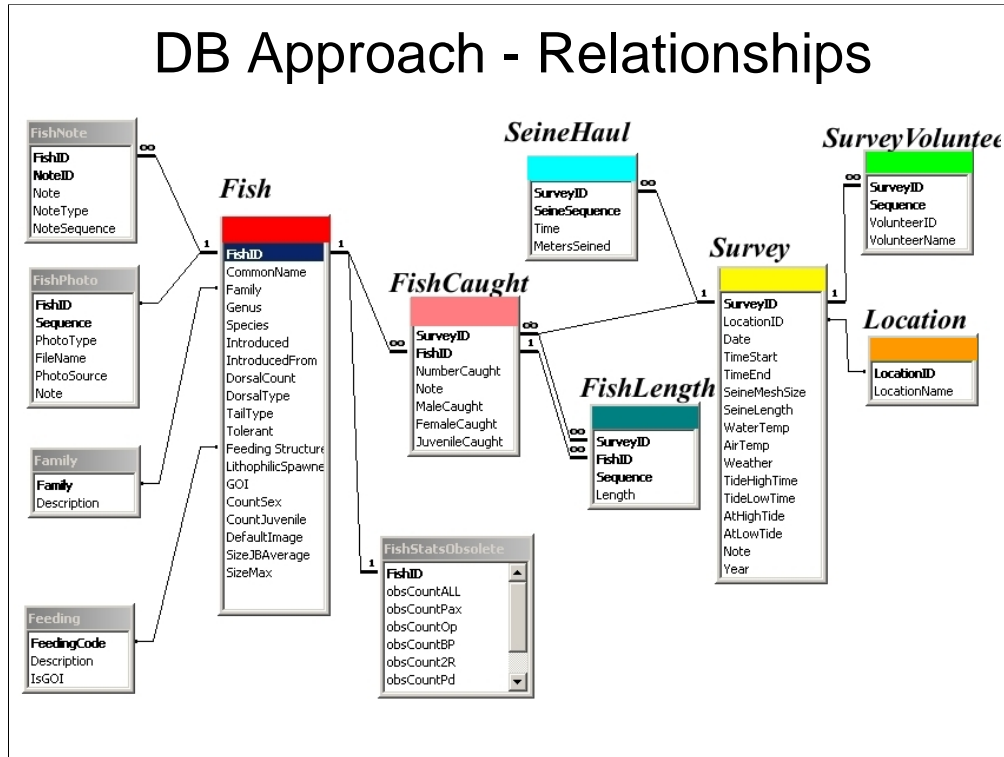
This is a good place to store information that can be used for analytical purposes. For example, LithophilicSpawner is a Yes/No indicator of whether the fish needs a gravel bottom for spawning. FeedingStructure, GOI, and Introduced similarly help to address the shortfalls in the spreadsheet approach that limited analysis.

CountSex and CountJuvenile are yes/no fields to indicate whether or not the user interface for data entry should allow the entry of sex and gender subcounts.

This version of the design partially handles the issue of uncertain identification to the genus species level. Since the unique identifier is FishID, a FishID could be created for “Needlefish spp” a problem noted in the discussion of the sample data. This would work reasonably well as long as the other info (introduced, tolerant, feeding structure, etc was correct or omitted). A FishID for “Dace/Creek Chub” could be created, but would be more problematic since the genus (and maybe family) would be different. The point is that the data be entered. There may be some analytical quirks depending upon how vague the FishID is.

A fancier (and much more elaborate design) could implement the taxonomic tree for fish and allow “FishID” to be linked to any level of that tree, not just the assumed species level. This would also handle subspecies and other lower classification levels if needed. This approach was not deemed to be necessary for this relatively simple fish survey database.

DB Approach - Relationships



This is the database design as represented within Microsoft Access.

The lines represent the relationships between the tables. In other words, the lines connect fields with the same values.

For example the line connecting SurveyID in Survey to SurveyID in FishCaught indicates that SurveyID is the field defining the link. This is equivalent to my earlier discussion that each FishCaught has the SurveyID included to identify which survey the FishCaught is for. The 1 at the Survey end of the line and the infinity sign at the FishCaught end of the line means that this is what computer scientists call a 1 to many (often written as 1:m). This means that for a particular Survey there can be multiple FishCaught records. In plain English, more than one type of fish can be caught during a survey.

Fish to Fish Caught is also one to many, meaning that a species of fish can be caught during more than one survey.

Survey to SeineHaul is one to many, meaning that there can be more than one SeineHaul for a particular Survey.

Survey to SurveyVolunteers is also one to many meaning that a Survey has one or more SurveyVolunteers associated with it.

Location to Survey is not specified, but is also one to many meaning that a Location can have more than one survey done at that location.

FishNote and FishPhoto are other tables in the database that include reference information about species. The Family table is to provide a common description of the taxonomic Family name. Feeding is used to translate the FeedingStructure code into an English description.

FishStatsObsolete is included only because the data was available in the spreadsheet. It IS NOT good design. The fields represent data for each of the locations. The design problem is that whenever a new location is created, the structure of this table would have to be changed. This requires programming changes everywhere this table is used. The table is NOT needed because the same results can be obtained by following the link from Location to Survey to FishCaught (and optionally to FishID) and totaling the number of fish caught. This strategy does not change if the number of locations changes, so it is far better. To repeat, the FishStatsObsolete is included only as a historical artifact, hence the "Obsolete" part of the name. Once I was sure there is no need for this artifact, I would be inclined to delete it.

User Interface - Fish

Fish Entry

FishID: 6

Common Name: Banded Killifish

Genus: Fundulus Species: diaphanus

Introduced: N Native

Introduced From:

Dorsal Count: 1

Dorsal Type:

Tail Type: Broom

Feeding Structure: Invertivore

Lithophilic Spawner: No

Tolerant: Unknown

Count Sex Count Juvenile

SizeJB Ave:

SizeMax:

Record: 5 of 47

Select Image

- 0 ID drawing
- 1 JB photo
- 2 JB photo

File: Killifish Male edited May 12 2007 Fish 1 006 crop.jpg

Source: Jeffrey Campbell

Photo Note: Male

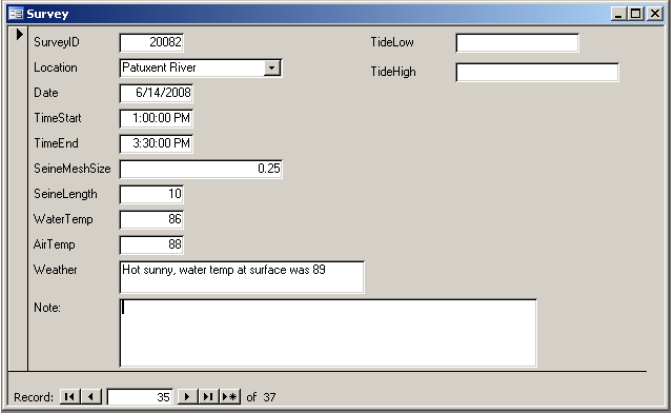
Key design aspects of this data entry screen for Fish data:

Predefined lists for Introduced, Tail Type, Feeding Structure, Lithophilic Spawner, and Tolerant are used to minimize errors in data entry by not allowing free text entry.

Access allows photos to be stored in the database. However, for this application since I wanted to be able to use the photos in various documents, I did not keep the photos in the database. Instead, I included the file name and path in the database and wrote Access VBA code to display the selected image.

This was an initial draft of the screen when I was the only user. I would clean up the spacing, alignment, and other aspects in a screen design for more users.

User Interface - Survey



The screenshot shows a window titled "Survey" with the following fields and values:

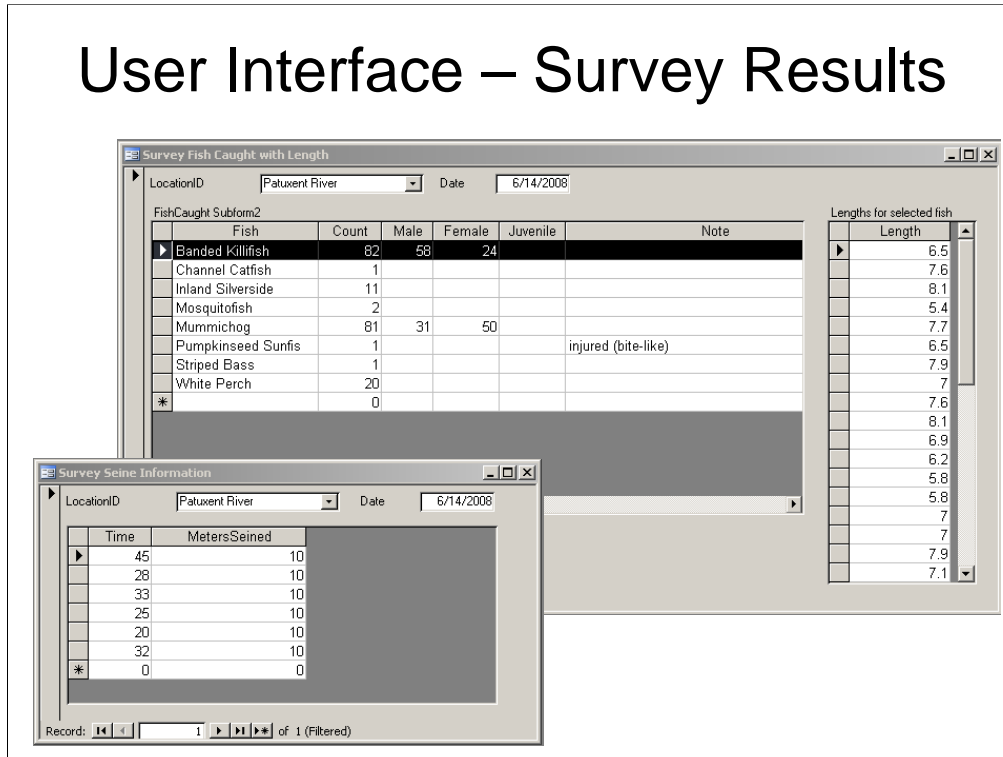
SurveyID	20082	TideLow	
Location	Patuxent River	TideHigh	
Date	6/14/2008		
TimeStart	1:00:00 PM		
TimeEnd	3:30:00 PM		
SeineMeshSize	0.25		
SeineLength	10		
WaterTemp	86		
AirTemp	88		
Weather	Hot sunny, water temp at surface was 88		
Note:			

Record: 35 of 37

A very quick and dirty (automatically generated form) showing the Survey information.

Note that the LocationID is hidden from the user, instead the location descriptions are displayed in the drop down list box. Access stores the corresponding LocationID in the Survey table. This provides the efficiency of a simple LocationID with the user friendliness of a selection list.

User Interface – Survey Results



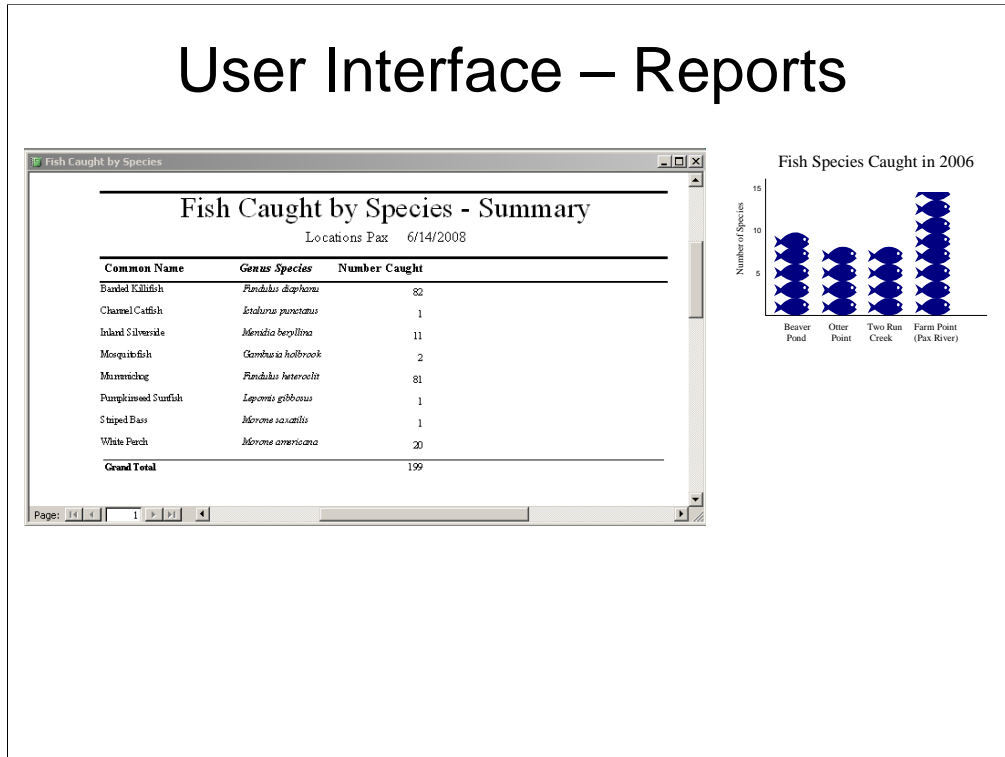
These screens illustrate how the one to many relationships are handled in the user interface

The main screen shows that a row in the main table is selected and then the Length information for that fish is provided in the column to the right.

The Survey Seine Information screen shows how the multiple SeineHall record data can be entered by the users.

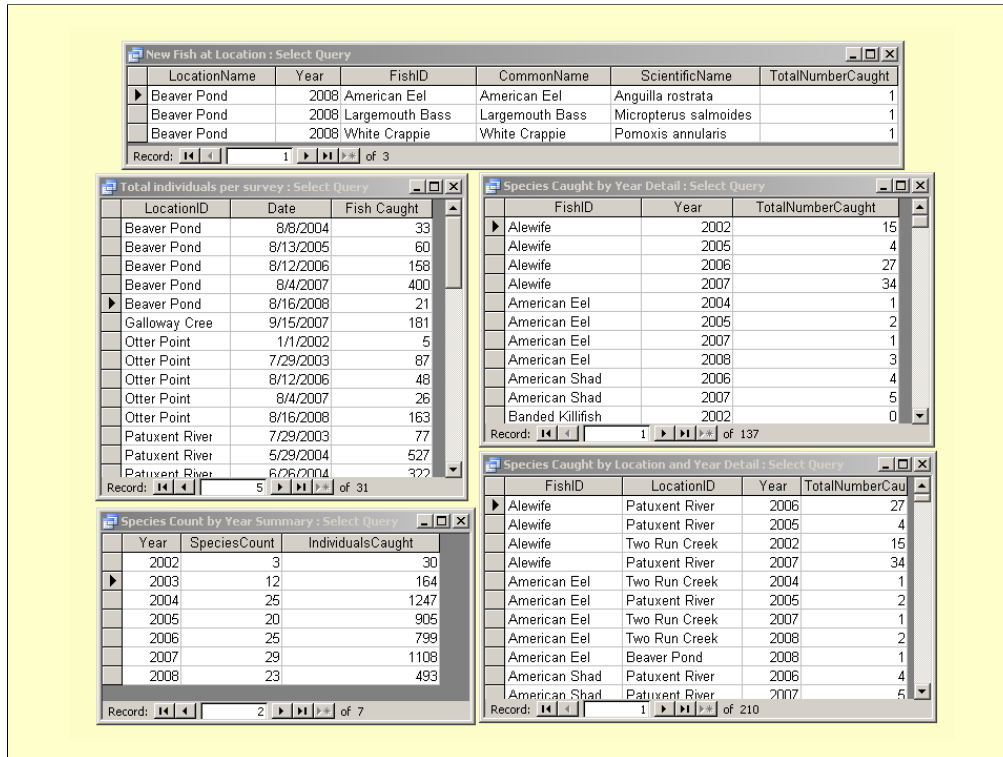
These were initial drafts of the screens when I was the only user. I would clean up the spacing, alignment, labels and other aspects in a screen design for more users.

User Interface – Reports



A simple report showing the fish caught during a particular survey (location and date).

The graphic to the left was used for publication in the quarterly newsletter.



Data in tabular format for various analytical purposes. The headings describe the analysis.

The main point is all of these are the result of using the properly designed database instead of the original spreadsheet.

Conclusion

To get the full benefits from the database:

- Define tables so that no information is repeated
- Link tables together with meaningful relationships
- Add static (“lookup”) tables to store code values

Contact Information



- Environmental Informatics Consulting
 - jcampbell@EnvironmentalInformatics.com
 - www.EnvironmentalInformatics.com