# Managing Spatial Data: The FlexiGrid Experience

Revital Katznelson, Ph.D. Berkeley, CA

# Today's topics

Timeline Concept and Implementation Lessons learned Options

# **Timeline: History**

2007, June – FlexiGrid concept articulated after working with PHAB protocol authors from USGS and EPA 2007, November – first FlexiGrid system applied to CA interim PHAB protocol 2008 Interim protocol system used in a pilot study with Alameda County Clean Water Program 2008 FlexiGrid applied to field-collection and processing of PHAB and periphyton data according to the developing Surface Water Ambient Monitoring Program (SWAMP) protocols 2009 A crosswalk to the SWAMP database was added to the FlexiGrid data entry template 2010-2012 extensive use by San Francisco Bay Region waterboard 2014 Implementation of the FlexiGrid concept for a customized pebble-count protocol in the North Coast



### Some of the protocols I encountered

Fisheries habitat assessments DFG style (Flossi&Reynolds...) Thalweg profile & cross section surveys Pebble counts Large Woody Debris assessments **Riparian plots & canopy density** Residual pool volume (V\*) Collection of vertical composite samples in water column Sampling storm runoff in outfall & creek networks Bird point-counts in a 50m-radius circle Flow discharge measurements Physical Habitat assessments RBP style Physical Habitat assessments EMAP style Physical Habitat assessments NAWQA style (more)

# My 2000s drawings



Transects Transect points Intertransects Transect Plots Riparian plots Stream segments

Lots of "Crosslines", and "Plots" ...and the internal hierarchy gets more and more complicated

# The Challenges:

Multiple Protocols: Identify the common elements and components among diverse protocols and spatial Sampling Frames

Links: Find a universal way to connect each monitoring Result to the specific spatial component it represents, in a flexible data structure

Hierarchy: Find a way to map multi-dimensional sampling frames with internal hierarchy

Endpoints: Find a simple way to generate assessment Endpoints from desirable aggregates of components

## **Spatial Sampling Frames**

#### Can share the following properties:

- A 'Grid' made of a number of components
- Flexible (variable angles)
- Multi-dimensional
- A structure with internal hierarchy

They can easily be defined in neutral terms: each Frame has

- An origin
- A direction (upstream or downstream)
- A known curvilinear distance.

The **Origin** can be mapped to a Permanent Station and/or Monument

Curvilinear distances can be measured on a tape placed on the thalweg, or the centerline, or the centroid.... Every Frame or grid has a "**backbone**" of sorts

### Mapping is so much easier with a Backbone!

Rock

9

Road

W Bridge

We are talking about "a spatial multi-dimensional flexible sampling frame with a variety of components, some organized with an internal hierarchy" (in short, SMDFAFWVCIH).

#### I called it "A FlexiGrid"

A FlexiGrid with a unique ID is laid out on a stream channel by one Team in one Station Visit. A Team can lay out one or more FlexiGrids in one Visit, each with its own unique ID.

The FlexiGrid refers to permanent locations, but is not permanent itself (FlexiGrids may be laid out differently at different times).

The FlexiGrid concept can be applied to different scales.



The secret: use of neutral terms

# The FlexiGrid is a Concept

It is a way of organizing information about spatial components of complex sampling frames to enable:



(a) Description and mapping of the Frame and each component within the frame

(b) Linking each assessment Result to the component in which it was collected.

#### A plot is a plot is a plot

## Common Data Elements

(These are added to the 'generic' Station-Visit descriptors: Project, Trip, Team, Date, Time, Station, Station-Visit ID, etc.)

List 1: FlexiGrid (Frame) Descriptors

FlexiGrid Origin ID
FlexiGrid distance units (m, ft)
FlexiGrid Origin distance from permanent Station [/monument]
FlexiGrid Type (transects&points, right-angle grid, string of hab units, etc.)
FlexiGrid Backbone (Thalweg, midstream, centerline, centroid)
FlexiGrid Positive-distance direction (Upstream, downstream)

### List 2: FlexiGrid Component descriptors

FlexiGrid Component ID [naming conventions can reflect hierarchy] FlexiGrid Component Type (e.g., Stream-fragment, Crossline, Plot, Point, Vertical, River-Parallel, Crossline-point, etc.)

FlexiGrid Component Name in Protocol (e.g., Reach, Transect, Transect-point, BMI Plot, etc.); [these are protocol-specific]

FlexiGrid Component Pre-determined size [/shape, /boundaries] FlexiGrid Component backbone-distance from Origin FlexiGrid Component place in series FlexiGrid Component distance from Local Reference Point Local Reference Point type (e.g., Left-bank) FlexiGrid Component depth from surface FlexiGrid and FlexiGrid component descriptors

Remember: Any of the spatial components you have defined can be included in the FlexiGrid; you do not HAVE TO snap them to the grid (i.e., to map them) with distances or depth measurements

But if you want to, you CAN!

All you need is ... a spreadsheet.

And I already had a spreadsheet. Actually, I had more than one, I had the entire workbook of the Project File.

# **Timeline: Pre-History**

The Data Quality Management (DQM) system was developed for field measurements in 1998-2001, and implemented by the SWRCB Clean Water Team (Citizen Monitoring Program). In 2006 it was published as part of the Clean Water Team Toolbox and is available on line

http://www.waterboards.ca.gov/water\_issues/programs/swamp/cwt\_toolbox.shtml

## The major spreadsheet in the DQM Project File

	e,	Station Vis	it Informat	ion			Result Information										
Project ID	Team Name	Date	Station Visit ID	Station Visit start time	Permanent Station ID	Instrument ID	Characteristic	Result	Result unit	t Accuracy		Precision					
WIL03	Rkcrew	6/22/2003	T1V1	10:47	WIL070	TTP-STB01	Temperature, water	14.57	С	-1.44	%	0.06	%, RPD				
WIL03	Rkcrew	6/22/2003	T1V1	10:47	WIL070	ECP-STB01	Specific conductance	758.7	uS/cm	-0.14	%	0.40	%, RPD				
WIL03	Rkcrew	6/22/2003	T1V1	10:47	WIL070	DOP-STB01	Dissolved oxygen (DO)	11.08	mg/l	-5.00	%	6.92	%, RPD				
WIL03	Rkcrew	6/22/2003	T1V1	10:47	WIL070	PHP-STB01	рН	8.34	pН	0.71	%	0.12	%, RPD				
SLC95	RD crew	5/3/1995	RD05-v1	10:10	SLC116	EC-2SLC02	Specific conductance	780	uS								
SLC95	RD crew	5/3/1995	RD05-v1	10:10	SLC116	FLO-SLC01	Velocity	0.67	ft/sec								

The Field Result Table is linked to everything else: Location, Instrument, Organization, etc., all present in the Project File workbook

### The basic Results spreadsheet: a schematic



Everything was simple when the Station was a point.

### The FlexiGrid Results spreadsheet, Part 1 (Data Entry Template)



This spreadsheet has been customized several times. Each version served as a **data entry template** for a specific SOP

# In other words: the FlexiGrid information was inserted into the original Result spreadsheet...

Project ID	Team Name	Date	Station Visit ID	Station Visit start time	Permanent Station ID	Characteristic	Result	Result unit	Result type	Endpoint type
			And the second sec	-		nd				
SYCM09	PW crew	6/15/2010	PW05-v1	9:50	SYC050	Index of biological integrity (IBI)	67	(none)	Calculated endpoint	Compound endpoint
SYCM09	PW crew	6/15/2010	PW05-v1	9:50	SYC050	Average slope	4.5	%	Calculated endpoint	Simple endpoint
SLC95	RD crew	5/3/1995	RD05-v1	10:10	SLC116	Stream width	4.8	m	Individual measurement	n/ap
SLC95	RD crew	5/3/1995	RD05-v1	10:10	SLC116	Average Stream Widt	5.3	m	Calculated endpoint	Simple endpoint
SYCM09	PW crew	6/15/2009	PW05-v1	9:50	SYC050	<b>O</b> <b>C</b> <b>C</b> <b>C</b> <b>C</b> <b>C</b> <b>C</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>C</b> <b>D</b> <b>D</b> <b>C</b> <b>D</b> <b>D</b> <b>C</b> <b>D</b> <b>D</b> <b>D</b> <b>C</b> <b>D</b> <b>D</b> <b>D</b> <b>D</b> <b>D</b> <b>D</b> <b>D</b> <b>D</b>	10	%	Estimated number or numeric range	n/ap
WL-yr3	MD crew	5/12/2009	MD02-v2	13:50	WL32	Big Tree Canopy cove	10-40	%	Estimated number or numeric range	n/ap
SUIS06	LM crew	4/7/2009	LM21-v1	13:40	SUI085	Particle d50	34	mm	Calculated endpoint	Quantile
WL-yr3	MD crew	5/12/2009	MD02-v2	13:50	WL32	Stream width	13.5	m	Individual measurement	n/ap
WL-yr3	MD crew	5/12/2009	MD02-v2	13:50	WL32	Water depth	0.43	m	Individual measurement	n/ap

....to produce the FlexiGrid Results spreadsheet

Station Visit Information						ion	FlexiGrid Information						FlexiGrid Component Information										Result Information				
Projec	t ID Te Na	ame	Date	Station Visit ID	Station Visit start time	Permanent Station ID	FlexiGrid Origin ID	FlexiGrid distance units	FlexiGrid Origin distance from permanent Station	FlexiGrid type	FlexiGrid backbone	FlexiGrid positive distance Direction	FlexiGrid Component ID	FlexiGrid Component Type	FlexiGrid Component Name in Protocol	PHAB Protocol Used	FlexiGrid Component backbone-distance from Origin	FlexiGrid Component place in series	FlexiGrid Component distance from Local ref point	Local ref point type	Local ref point distance from Backbone	FlexiGrid Components Aggregate	Characteristic	Result	Result unit	Result type	Endpoint type
01/01			0/45/0040	DIMOS	0.50	01/0050	5005			transects	Thebuse	1 to a to a set	5005.04	Stream-	Death	1		-	-	-	-	DMI alata	Index of biological	07	(2222)	Calculated	Compound
STUN	109 PW	crew	6/15/2010	PVV05-V1	9:50	510000	FG35	m	23	&points	i naiweg	Upstream	FG35-RT	Fragment	Reach	Jones 2010	U	n/ap	n/ap	n/ap	n/ap	BINI plots	integrity (IBI)	67	(none)	endpoint	endpoint
SYCM	109 PW	crew	6/15/2010	PW05-v1	9:50	SYC050	FG35	m	23	transects &points	Thalweg	Upstream	FG35-R1	Stream- Fragment	Reach	Jones 2010	0	n/ap	n/ap	n/ap	n/ap	Slope Segments	Average slope	4.5	%	Calculated endpoint	Simple endpoint
SLC	95 RD (	crew	5/3/1995	RD05-v1	10:10	SLC116	FG-RD05	m	35	String of habitat units	Thalweg	Upstream	FG-RD05-P3- CS2	Crossline	Cross- Section	CCRS 1993	20	2	n/ap	n/ap	n/ap	n/ap	Stream width	4.8	m	Individual measurement	n/ap
SLC	95 RD (	crew	5/3/1995	RD05-v1	10:10	SLC116	FG-RD05	m	35	String of habitat units	Thalweg	Upstream	FG-RD05	Stream- Fragment	Reach	CCRS 1993	n/ap	n/ap	n/ap	n/ap	n/ap	Cross Sections	Average Stream Width	5.3	m	Calculated endpoint	Simple endpoint
SYCM	109 PW	crew	6/15/2009	PW05-v1	9:50	SYC050	FG35	m	23	transects &points	Thalweg	Upstream	FG35-R1- PL08	Plot	BMI Plot	Smith 2003	67	6	5.8	LB	3.9	n/ap	Estimated percent Cobble	10	%	Estimated number or numeric range	n/ap
WL-3	r3 MD	crew	5/12/2009	MD02-v2	13:50	WL32	FGW32	m	-48	transects &points	midstream	Upstream	FGW32-R1- RP2	Plot	Riparian plot	(EMAP)	18	2	n/ap	n/ap	n/ap	n/ap	Big Tree Canopy cover	10-40	%	Estimated number or numeric range	n/ap

....and then the formulae for Assessment Endpoints can be pasted on the right of the data entry columns

### Assessment Endpoints can be

Descriptive statistics (Average, Frequency), Metrics, Indices, or any other Endpoint calculated for an aggregate of raw data.

Examples: Average wetted channel width, Percent cobble, Landfill/Trash Human Influence Index

Trip start date	SV start time	Permanent Station ID	FlexiGrid Compone nt place in series	Characteristic	Result	Result unit	Processe d Result (Formula e!)	Result [Endpoint]	Unit	Characteristic [Endpoint]	Components Aggregate for Endpoint	Number of observations in aggregate (n)
4/13/2005	8:45	201EAS050	1	Boulders cover	1	(code)	5	11	%	Average Boulders cover	Habitat Plots	11
4/13/2005	8:45	201EAS050	2	Boulders cover	1	(code)	5					
4/13/2005	8:45	201EAS050	3	Boulders cover	0	(code)	0					
4/13/2005	8:45	201EAS050	4	Boulders cover	1	(code)	5					
4/13/2005	8:45	201EAS050	5	Boulders cover	0	(code)	0					
4/13/2005	8:45	201EAS050	6	Boulders cover	1	(code)	5					
4/13/2005	8:45	201EAS050	7	Boulders cover	2	(code)	25					
4/13/2005	8:45	201EAS050	8	Boulders cover	0	(code)	0					
4/13/2005	8:45	201EAS050	9	Boulders cover	2	(code)	25					
4/13/2005	8:45	201EAS050	10	Boulders cover	2	(code)	25					
4/13/2005	8:45	201EAS050	11	Boulders cover	2	(code)	25					

### The FlexiGrid Results spreadsheet, Part 2 (with Endpoint Template pasted on the right)



The Endpoint Calculation Template developed for SWAMP has been customized to calculate an array of Endpoint, mostly per the EMAP protocol (Kaufmann et al 1999)

### The FlexiGrid Results spreadsheet, Part 3 (With SWAMP Crosswalk pasted or Built-in)



This spreadsheet has been used in the San Francisco Bay Regional waterboard as a Data Entry Template in 2010-2013, and can be used (in the office or in the Field) for future work Summary Part 1: The FlexiGrid concept supports a data structure that can do the following:

-- Connect each monitoring Result to the exact spatial component it was measured in (or to aggregate thereof);

-- Stage the raw data for easy, streamlined derivation of descriptive statistics, metrics, indices, and any other Endpoint;

-- Place every spatial component on a 'virtual map' as an independent entity while preserving the internal hierarchy and its order in a series;

-- Enable reconstruction of the Frame from the virtual map that is stored in database cells; and allow information transfer into a GIS system;

### Summary Part 2: More good things FG can do

-- Integrate physical habitat (PHAB) results and descriptors with results and descriptors from any other area of inquiry (e.g., chemistry, toxicity, bioassessment, etc.);

-- Enable reporting of any type of Result and Endpoint, generated for any Characteristic in any spatial component and at any scale, on the 'same page'; and

-- Accommodate data generated by any Protocol that has multiple spatial components, including PHAB protocols used by different Agencies and Programs (e.g., NAWQA, EMAP, EPA RBP), DFG Fisheries habitat, etc.

Currently available: All FG templates for SWAMP, plus all SOPs for streamlined data processing and transfer

### Summary Part 3: Drawbacks

-- Tailoring the FlexiGrid spreadsheet to a given protocol requires real understanding of that protocol

-- Development and customizing of the templates requires focused thinking and advanced Excel skills

-- Although data entry and streamlined endpoint derivation can be done by anyone, the Excel Data Entry Template cannot support as many 'fool-proof' safeguards against erroneous entries as some other database systems, and it is harder to protect

-- Users need training and technical support

### Practically speaking ...

-- Users need training and technical support

But this is true for any data management system. FlexiGrid help is available and it does not cost much.

The FlexiGrid is a good option for small monitoring entities that have no funding for massive data management and no access to IT support.

It is just an Excel spreadsheet!

# Thanks!

San Francisco Bay Region (RB2) SWAMP Team 2007-2014: Matt Cover, Karissa Anderson, Carrieann Lopez, Annika Anderson, Peter Otis, Karen Taberski, David Williams, Kevin Lunde, Caitlin White

> Alameda County Clean Water program 2008 Arlene Feng, Paul Salop

North Coast Region (RB1) 2010-2014 Carrieann Lopez, Richard Fadness

> Moss Landing Marine Lab SWAMP Data Management Team 2010-2014: Marco Sigala, Cassandra Lamerdin