Climate-Aquatics Blog #73: BIG DATA as an engine for aquatic information creation

(…as long as you don't drown in it…)

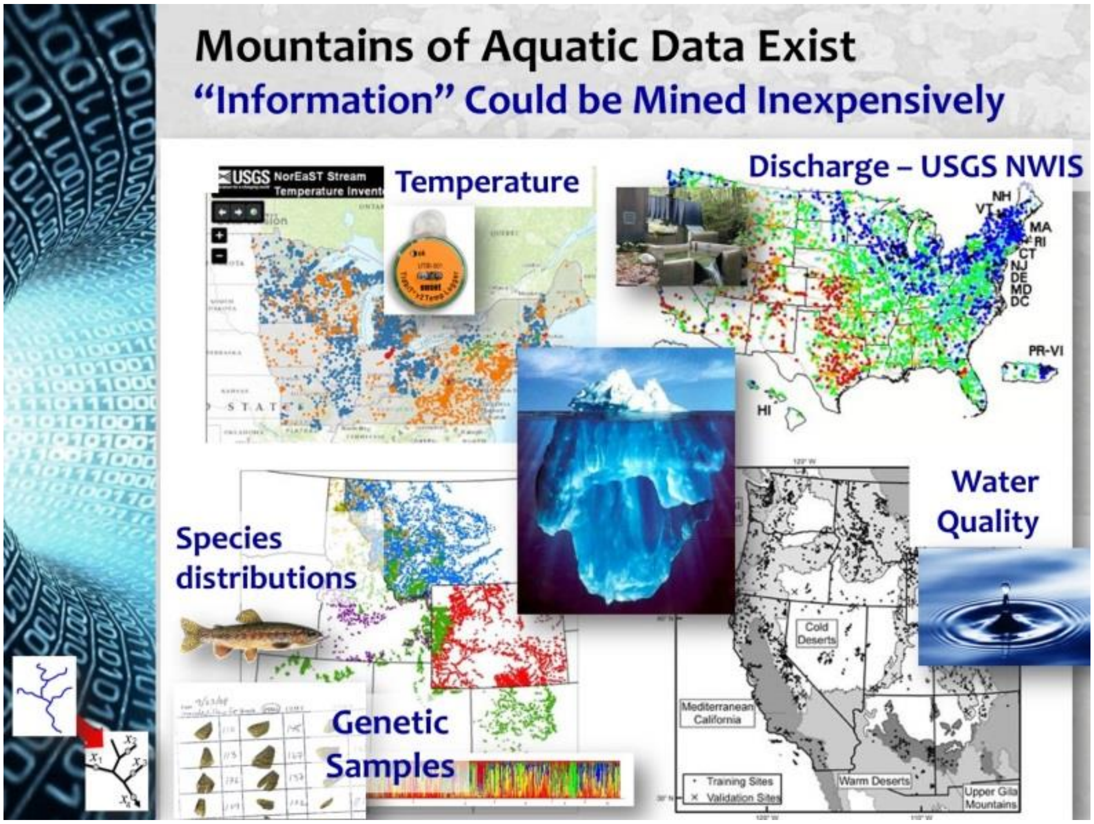
The smartest thing, the only thing really, we can do to conserve & preserve fisheries and aquatic biodiversity as the climate warms this century is to invest our limited resources wisely. Conservation needs will always outpace our ability to meet them, so maximizing returns on investments is crucial. Doing so requires good information to guide investment choices, and data are fundamental to creating that information. Good thing we live in a world that veritably drips with data (graphic 1). Everywhere you look, once you know how to look, it lurks—not at all unlike the classic scene in the Matrix when Neo first perceives Agent Smith's true nature and the digital artifice that surrounds him (click here for the YouTube refresher). As described last time, eDNA will be an important contributor to, and accelerant of, that data drip (blog #72); as will easy-to-use protocols for collecting stream flow and temperature data (blog #60) in what Porter & colleagues call the "sensor data deluge" (study hyperlinked here). But before collecting yet more data, it's also important to realize that we're oftentimes sitting on mountains of the stuff that is seriously underutilized & could be mined to produce high quality information at low cost. Collectively, there are 100s (1000s?) of natural resource agencies and organizations that collect data in, & about, streams and lakes globally, & we have spent many billions of US$ doing so. But a significant impediment to using much of those data is that they are not usable. In fact, it's frequently the case that no one other than the few people involved in specific collection activities even know the data exist because they live in disparate file cabinets (the data, not the people) or hard-drives outside of real databases. Heidorn calls these "dark data" (study hyperlinked here) and discusses their great potential utility to science, and great risk of being lost forever as we transition from the age of paper to one that is digital.

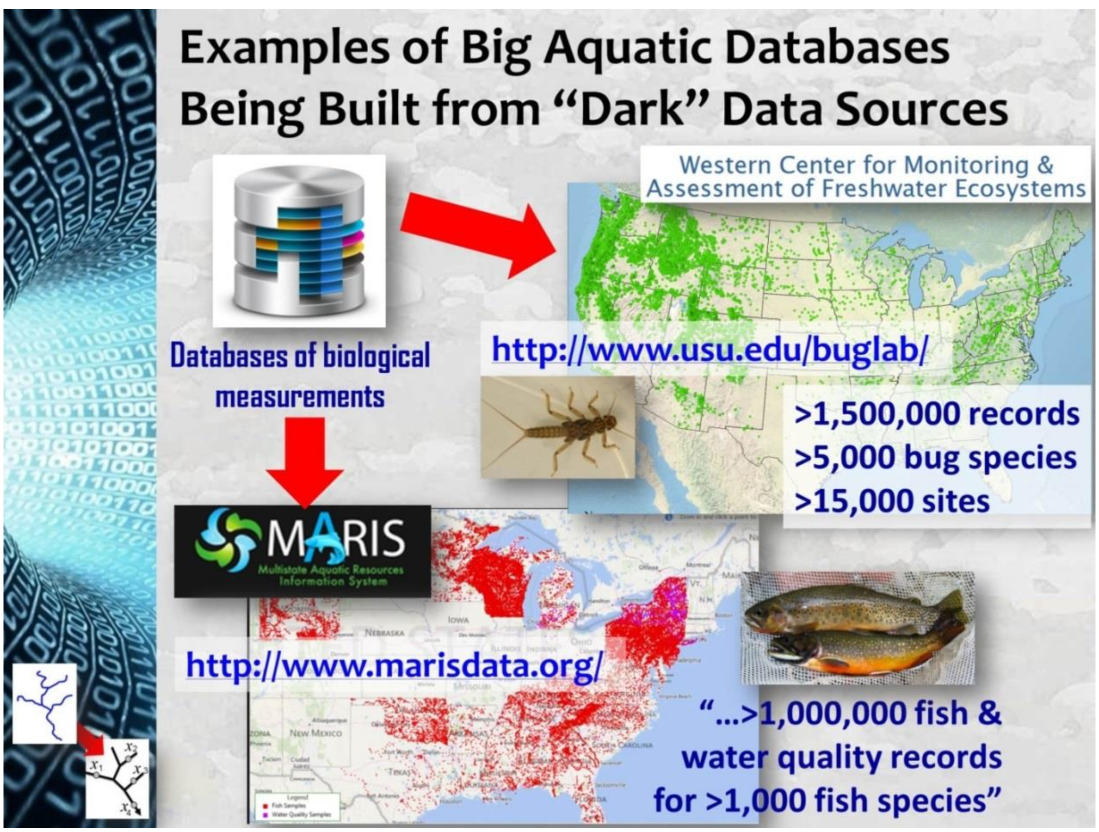
We all know examples of, or have, dark data that could & should be brought into the light. Unfortunately, there are no easy ways of doing so other than rolling up our collective sleeves and diving in. In the case of stream temperature data in the western U.S., it's taken our NorWeST temperature team the past 4 years to clean & organize data from >100 agencies into a functional database & we still have 1 more year & several states to go. And that's just one type of data in one part of the country so there's just a huge amount of work yet to be done to get our legacy datasets up to snuff. But as different groups work their way through dark matter database development, it's rather impressive to see what's out there. Two prime aquatic examples are MARIS (Multistate Aquatic Resource Information System), which hosts >1,000,000 fish sample records for >1,000 species in 25 states; and MapIT (Mapping Application for Freshwater Invertebrate Taxa), which hosts >1,500,000 records at >15,000 national sites for >5,000 aquatic macroinvertebrate species (graphic 2). Also fair aquatic database game to consider are those that consist of environmental descriptors, like the national StreamCat database recently developed by EPA, or a similar precursor developed by the National Fish Habitat Partnership (graphic 3). Those descriptors are typically remotely sensed or calculated from digital elevation models (DEM) but serve as very useful predictor variables when developing models to analyze or predict patterns in streams and lakes. McManamay & Utz (study hyperlinked here) discuss other notable aquatic databases and how open access to large amounts of data presents opportunities and challenges. Then also there are the seriously large databases like GBIF (Global Biodiversity Information Facility), which contains >600,000,000 species occurrence locations across all taxonomic groups globally. Or the U.S.-centric version of that species database recently started by USGS through their BISON initiative (Biodiversity Information Serving Our Nation).

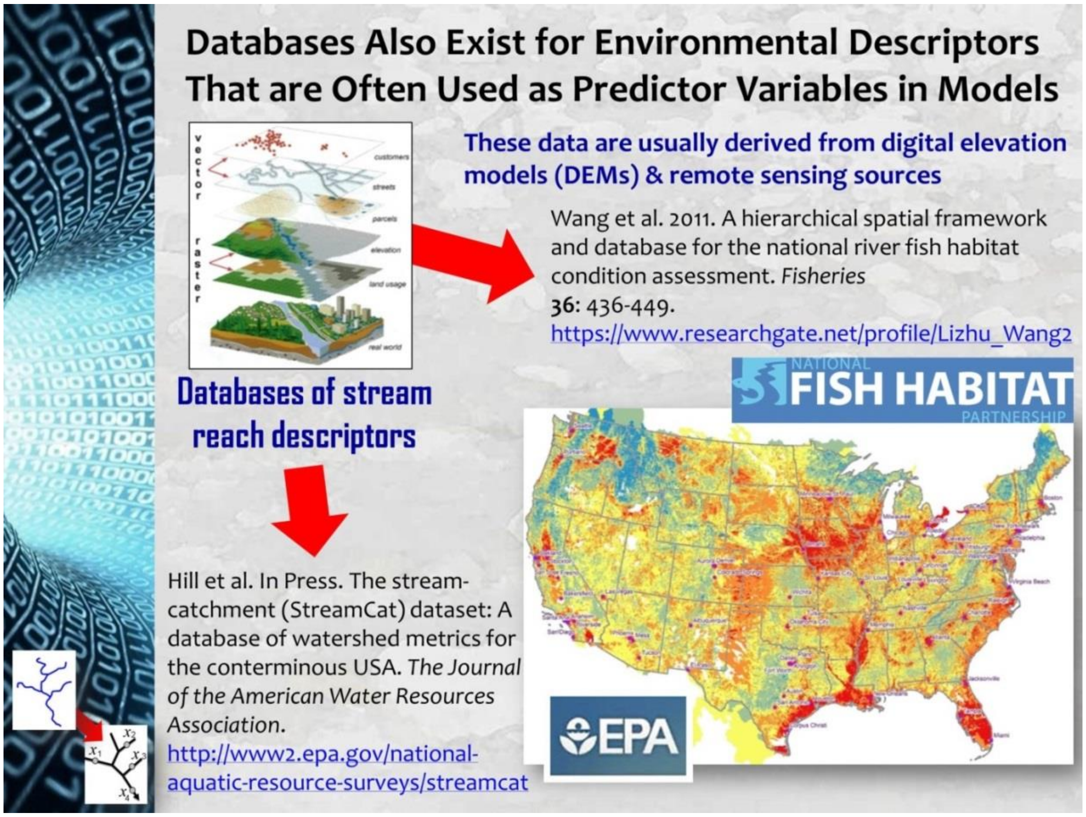
In addition to archiving disparate datasets for posterity, building high-quality databases provides significant value that is quite tangible. In the case of NorWeST, it has taken $1,000,000 in salaries to maintain a database team for 4 years, but that team has created an open-access database from contributions by hundreds of people that would require $10,000,000 & decades of field work to replicate. The true value of any database, however, isn't what it costs to collect the data, but rather that it allows data to be handled & summarized efficiently to provide useful information for decision making. A decision might be as simple as foregoing new data collections (& the associated costs) because the desired data already exist in the database. Or if new data are collected, deciding where to collect them so as to avoid redundancy with existing data locations—a straightforward proposition once data sites are easily queried and mapped (graphic 4). But more exciting is that as more databases are developed, it becomes possible to merge them in interesting ways and conduct novel analyses that yield previously impossible insights. As this process proceeds, we'll increasingly find ourselves not so much limited by data, but by the quality of the questions we ask of data. Poisot & colleagues (study hyperlinked here) argue that if guided by sound hypotheses, such "synthetic datasets" will lead to rapid scientific advances and information creation (graphic 5). That dynamic couldn't happen at a better time given the climate times we live in, and I'm confident that the collective creativity & ingenuity of fish people will find many ways to capitalize on the world of new opportunities opening to us.

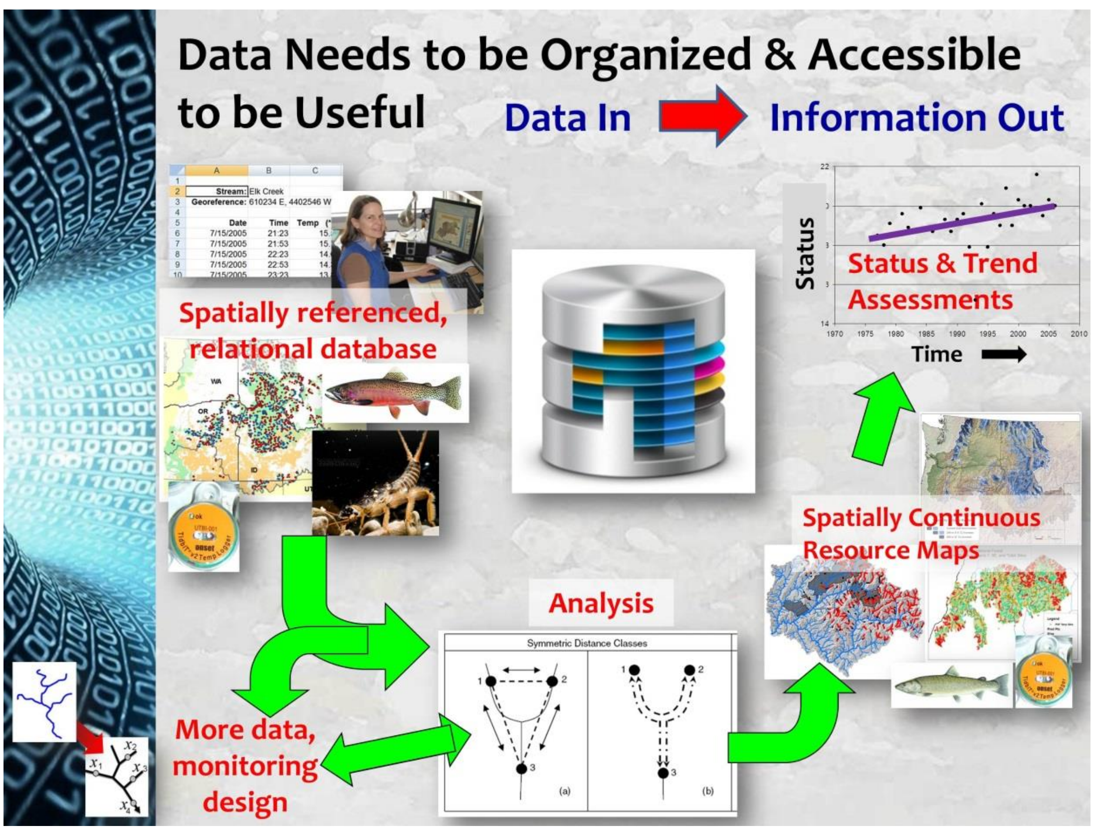
Until next time, best regards. Dan

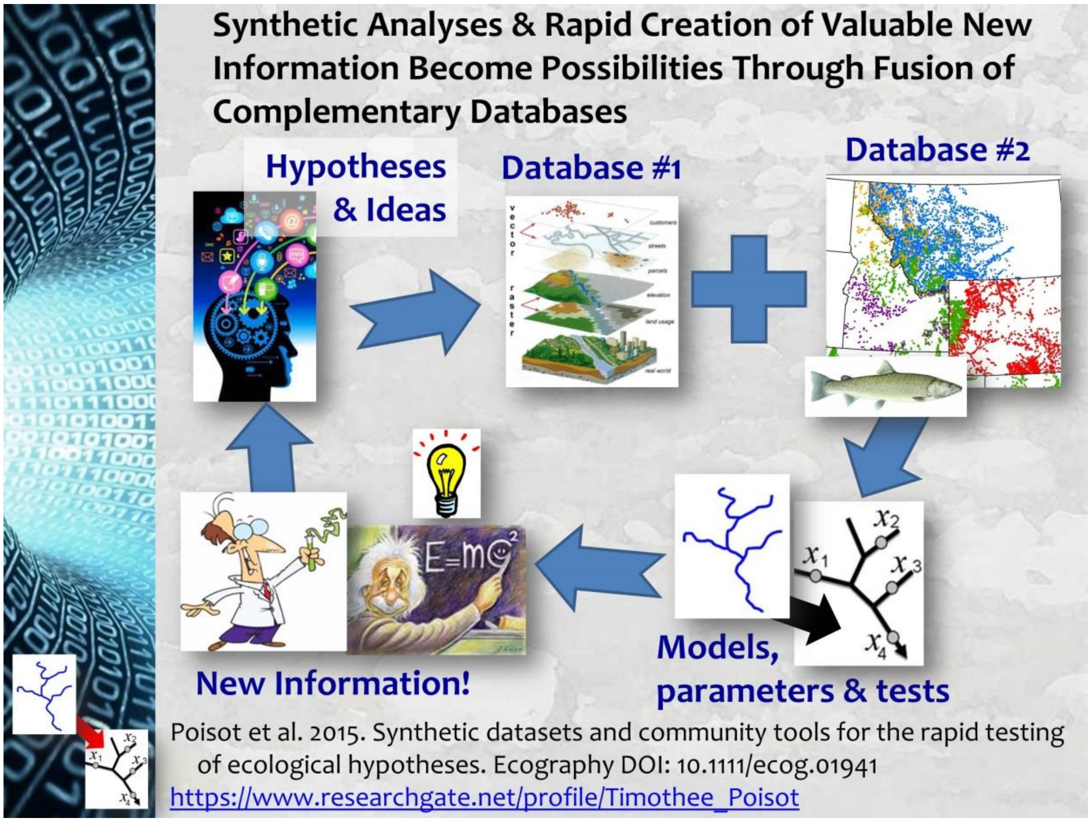
Tweeting at Dan Isaak@DanIsaak


Mountains of Aquatic Data Exist
"Information" Could be Mined Inexpensively
USGS NorEaST Stream Temperature Inventory
Temperature
Discharge – USGS NWIS
NH
VT
MA
RI
CT
NJ
DE
MD
DC
PR-VI
HI
Water
Quality
Species
distributions
Genetic
Samples
Cold
Deserts
Mediterranean
California
Training Sites
Validation Sites
Warm Deserts
Upper Gila
Mountains


Examples of Big Aquatic Databases
Being Built from "Dark" Data Sources
Western Center for Monitoring &
Assessment of Freshwater Ecosystems
Databases of biological
measurements
http://www.usu.edu/buglab/
>1,500,000 records
>5,000 bug species
>15,000 sites
MARIS
Multistate Aquatic Resources
Information System
http://www.marisdata.org/
"...>1,000,000 fish &
water quality records
for >1,000 fish species"


Databases Also Exist for Environmental Descriptors That are Often Used as Predictor Variables in Models
These data are usually derived from digital elevation models (DEMs) & remote sensing sources
Wang et al. 2011. A hierarchical spatial framework and database for the national river fish habitat condition assessment. *Fisheries* **36**: 436-449.
https://www.researchgate.net/profile/Lizhu_Wang2
Databases of stream reach descriptors
Hill et al. In Press. The stream-catchment (StreamCat) dataset: A database of watershed metrics for the conterminous USA. *The Journal of the American Water Resources Association.*
http://www2.epa.gov/national-aquatic-resource-surveys/streamcat
NATIONAL FISH HABITAT PARTNERSHIP
EPA


Data Needs to be Organized & Accessible to be Useful
Data In
Information Out
Spatially referenced, relational database
Status & Trend Assessments
Status
Time
Spatially Continuous Resource Maps
Analysis
Symmetric Distance Classes
More data, monitoring design


Synthetic Analyses & Rapid Creation of Valuable New Information Become Possibilities Through Fusion of Complementary Databases
Hypotheses & Ideas
Database #1
Database #2
E=mc²
New Information!
Models, parameters & tests
Poisot et al. 2015. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. Ecography DOI: 10.1111/ecog.01941
https://www.researchgate.net/profile/Timothee_Poisot


Good Information for Strategic Decision Making Will Be Critical
The 21st-Century will Be a Transitional One
2014 Set New Record
I'm going to invest here…
… not here

Welcome to the Climate-Aquatics Blog. For those new to the blog, previous posts with embedded graphics can be seen by clicking on the hyperlinks at the bottom or by navigating to the blog archive webpage here: (http://www.fs.fed.us/rm/boise/AWAE/projects/stream_temp/stream_temperature_climate_aquatics_blog.html). The intent of the Climate-Aquatics Blog is to provide a means for the ~9,000 field biologists, hydrologists, anglers, students, managers, and researchers currently on this mailing list across North America, South America, Europe, and Asia to more broadly and rapidly discuss topical issues associated with aquatic ecosystems and climate change. Messages periodically posted to the blog highlight new peer-reviewed research and science tools that may be useful in addressing this global phenomenon. Admittedly, many of the ideas for postings have their roots in studies my colleagues & I have been conducting in the Rocky Mountain region, but attempts will be made to present topics & tools in ways that highlight their broader, global relevance. I acknowledge that the studies, tools, and techniques highlighted in these missives are by no means the only, or perhaps even the best, science products in existence on particular topics, so the hope is that this discussion group engages others doing, or interested in, similar work and that healthy debates & information exchanges occur to facilitate the rapid dissemination of knowledge among those concerned about climate change and its effects on aquatic ecosystems.

If you know others interested in climate change and aquatic ecosystems, please forward this message to them. If you do not want to be contacted again in the future, please reply to that effect and you will be de-blogged.

Previous Blogs…

Climate-Aquatics Overviews

Blog #1: Climate-aquatics workshop science presentations available online
Blog #2: A new climate-aquatics synthesis report

Climate-Aquatics Thermal Module

Blog #3: Underwater epoxy technique for full-year stream temperature monitoring
Blog #4: A GoogleMap tool for interagency coordination of regional stream temperature monitoring
Blog #5: Massive air & stream sensor networks for ecologically relevant climate downscaling
Blog #6: Thoughts on monitoring air temperatures in complex, forested terrain
Blog #7: Downscaling of climate change effects on river network temperatures using inter-agency temperature databases with new spatial statistical stream network models
Blog #8: Thoughts on monitoring designs for temperature sensor networks across river and stream basins
Blog #9: Assessing climate sensitivity of aquatic habitats by direct measurement of stream & air temperatures
Blog #10: Long-term monitoring shows climate change effects on river & stream temperatures
Blog #11: Long-term monitoring shows climate change effects on lake temperatures
Blog #12: Climate trends & climate cycles & weather weirdness
Blog #13: Tools for visualizing local historical climate trends
Blog #14: Leveraging short-term stream temperature records to describe long-term trends
Blog #15: Wildfire & riparian vegetation change as the wildcards in climate warming of streams
Blog #23: New studies describe historic & future rates of warming in Northwest US streams
Blog #24: NoRRTN: An inexpensive regional river temperature monitoring network
Blog #25: NorWeST: A massive regional stream temperature database
Blog #26: Mapping thermal heterogeneity & climate in riverine environments
Blog #40: Crowd-sourcing a BIG DATA regional stream temperature model
Blog #60: Bonus Blog: New report describes data collection protocols for continuous monitoring of temperature & flow in wadeable streams
Blog #61: Significant new non-American stream temperature climate change studies

Blog #62: More Bits about the How, What, When, & Where of Aquatic Thermalscapes
Blog #63: Navigating stream thermalscapes to thrive or merely survive
Blog #64: Building real-time river network temperature forecasting systems

Climate-Aquatics Hydrology Module
Blog #16: Shrinking snowpacks across the western US associated with climate change
Blog #17: Advances in stream flow runoff and changing flood risks across the western US
Blog #18: Climate change & observed trends toward lower summer flows in the northwest US
Blog #19: Groundwater mediation of stream flow responses to climate change
Blog #20: GIS tools for mapping flow responses of western U.S. streams to climate change
Blog #21: More discharge data to address more hydroclimate questions
Blog #22: Climate change effects on sediment delivery to stream channels

Climate-Aquatics Cool Stuff Module
Blog #27: Part 1, Spatial statistical models for stream networks: context & conceptual foundations
Blog #28: Part 2, Spatial statistical models for stream networks: applications and inference
Blog #29: Part 3, Spatial statistical models for stream networks: freeware tools for model implementation
Blog #30: Recording and mapping Earth's stream biodiversity from genetic samples of critters
Blog #53: DNA Barcoding & Fish Biodiversity Mapping
Blog #71: Harnessing social & digital network technologies to maximize climate effectiveness
Blog #72: The eDNA revolution and developing comprehensive aquatic biodiversity archives

Climate-Aquatics Biology Module
Blog #31: Global trends in species shifts caused by climate change
Blog #32: Empirical evidence of fish phenology shifts related to climate change
Blog #33: Part 1, Fish distribution shifts from climate change: Predicted patterns
Blog #34: Part 2, Fish distribution shifts from climate change: Empirical evidence for range contractions
Blog #35: Part 3, Fish distribution shifts from climate change: Empirical evidence for range expansions
Blog #36: The "velocity" of climate change in rivers & streams
Blog #37: Part 1, Monitoring to detect climate effects on fish distributions: Sampling design and length of time
Blog #38: Part 2, Monitoring to detect climate effects on fish distributions: Resurveys of historical stream transects
Blog #39: Part 3, Monitoring to detect climate effects on fish distributions: BIG DATA regional resurveys
Blog #41: Part 1, Mechanisms of change in fish populations: Patterns in common trend monitoring data
Blog #42: BREAKING ALERT! New study confirms broad-scale fish distribution shifts associated with climate change
Blog #56: New studies provide additional evidence for climate-induced fish distribution shifts
Blog #43: Part 2, Mechanisms of change in fish populations: Floods and streambed scour during incubation & emergence
Blog #44: Part 3, Mechanisms of change in fish populations: Lower summer flows & drought effects on growth & survival
Blog #45: Part 4, Mechanisms of change in fish populations: Temperature effects on growth & survival
Blog #46: Part 5, Mechanisms of change in fish populations: Exceedance of thermal thresholds
Blog #47: Part 6, Mechanisms of change in fish populations: Interacting effects of flow and temperature
Blog #48: Part 7, Mechanisms of change in fish populations: Changing food resources
Blog #49: Part 8, Mechanisms of change in fish populations: Non-native species invasions
Blog #50: Part 9, Mechanisms of change in fish populations: Evolutionary responses
Blog #51: Part 10, Mechanisms of change in fish populations: Extinction
Blog #52: Review & Key Knowable Unknowns

Blog #65: The Fish Jumble as they Stumble along with the Shifting ThermalScape

Climate-Aquatics Management Module

Blog #54: Part 1, Managing with climate change: Goal setting & decision support tools for climate-smart prioritization

Blog #55: Part 2, Managing with climate change: Streams in channels & fish in streams

Blog #57: Identifying & protecting climate refuge lakes for coldwater fishes

Blog #58: Part 3, Managing with climate change: Maintaining & improving riparian vegetation & stream shade

Blog #59: Part 4, Managing with climate change: Keeping water on the landscape for fish (beaverin' up the bottoms)

Blog #66: Part 5, Managing with climate change: Barrier placements to facilitate fish flows across landscapes

Blog #67: Part 6, Managing with climate change: Assisted migration to facilitate fish flows across landscapes

Blog #68: Part 7, Identifying & protecting climate refugia as a strategy for long-term species conservation

Blog #69: Part 8, Building climate-smart conservation networks (metapopulations + biodiversity + refugia)

Blog #70: Part 9, Restoration success stories that improve population resilience to climate change