ATKINS

Making Data Fun & Valuable

Big Data and Decision Making

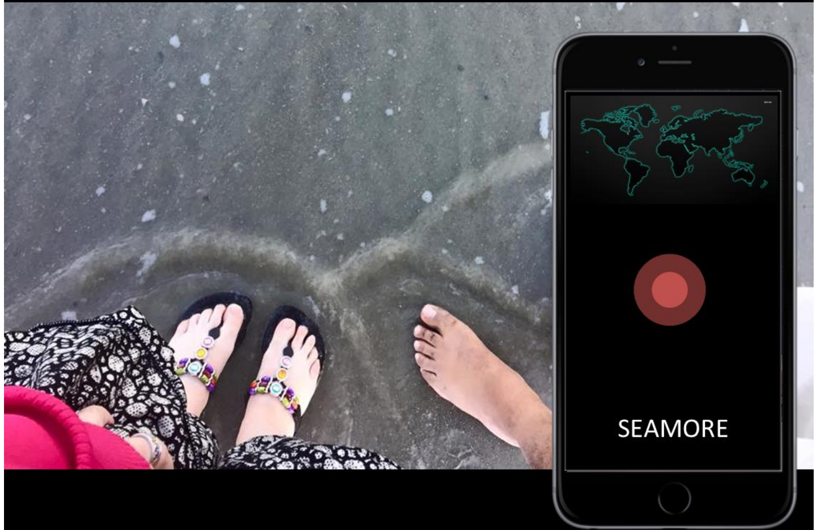
Florida Earth Foundation

Stephen Bourne, P.E.

Project Director, R&D Chair

Resilience Starts With You

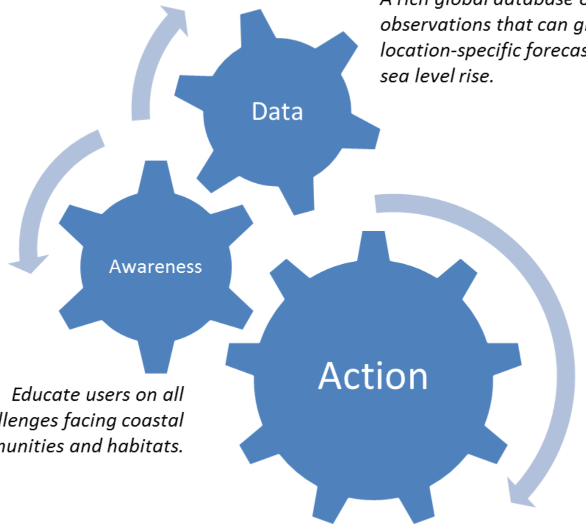
Sea level rise threatens coastal communities and habitats. Help to build our ability to adapt with SEAMORE. *It's a new iPhone App that is building a global record of how the coasts are changing.*



1. Find a Date.*
   You can also do it alone.
2. Stand at the Water's Edge.
3. Press Record.

A product of The Coastal Resilience Collaborative

Benefits of the App

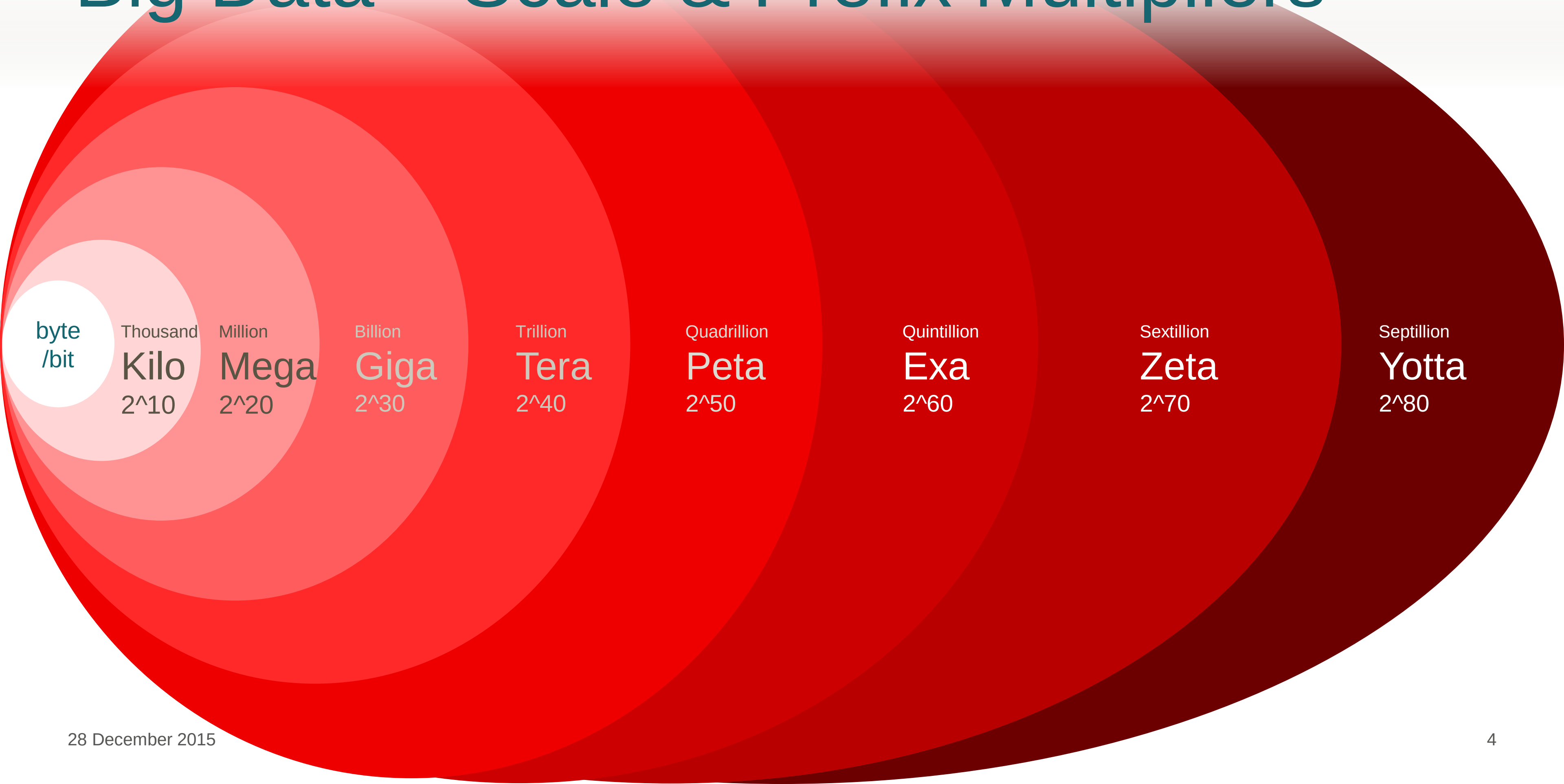


Big Data – the Four V’s

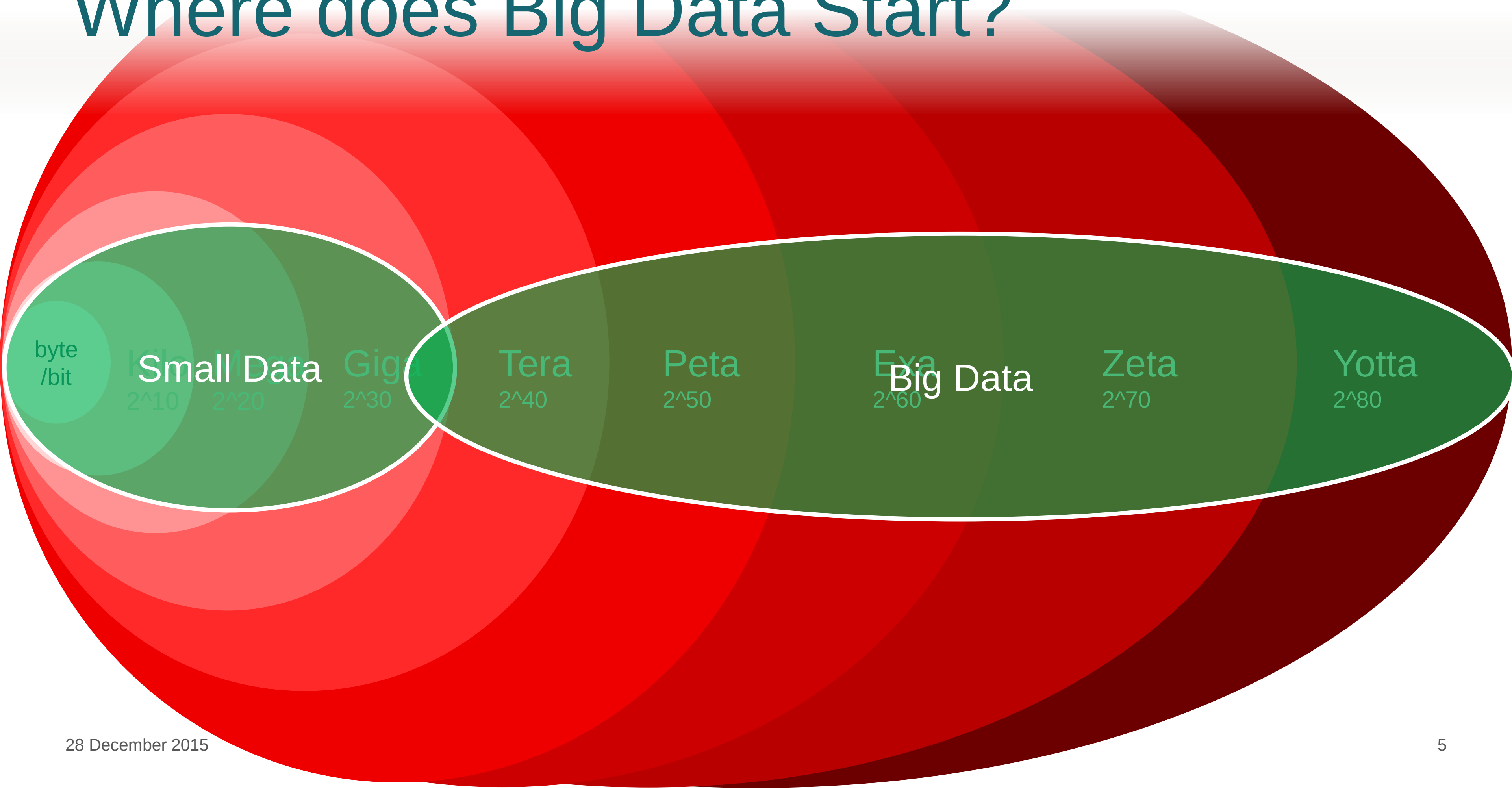
Volume	Velocity	Variety	Veracity
Don’t Sample, Just Observe	*A big spigot, in real time*	*Text, Images, Video, Audio*	*Question the truth*

Doug Laney, Gartner, 2001 then 2012

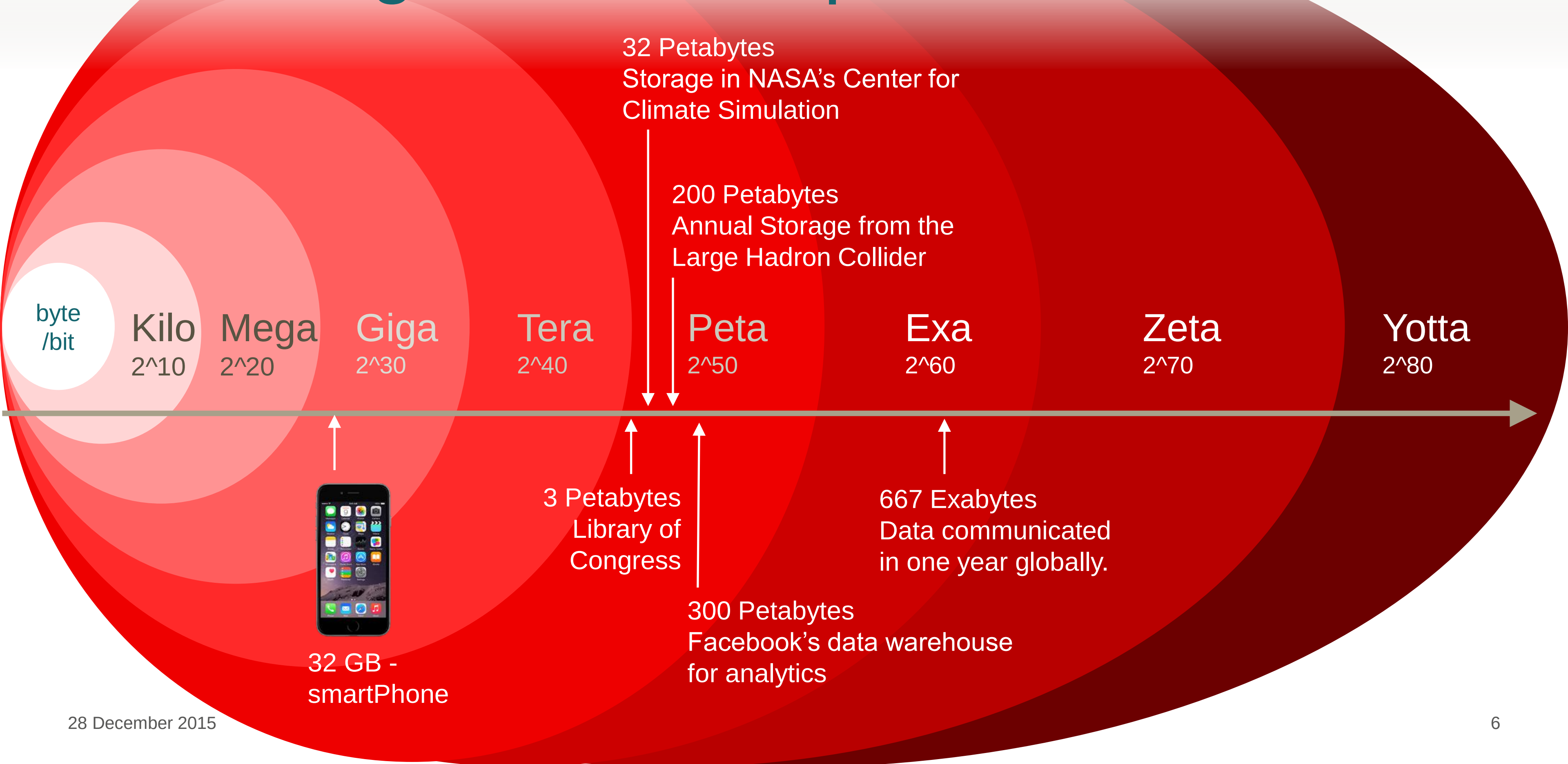
Big Data – Scale & Prefix Multipliers



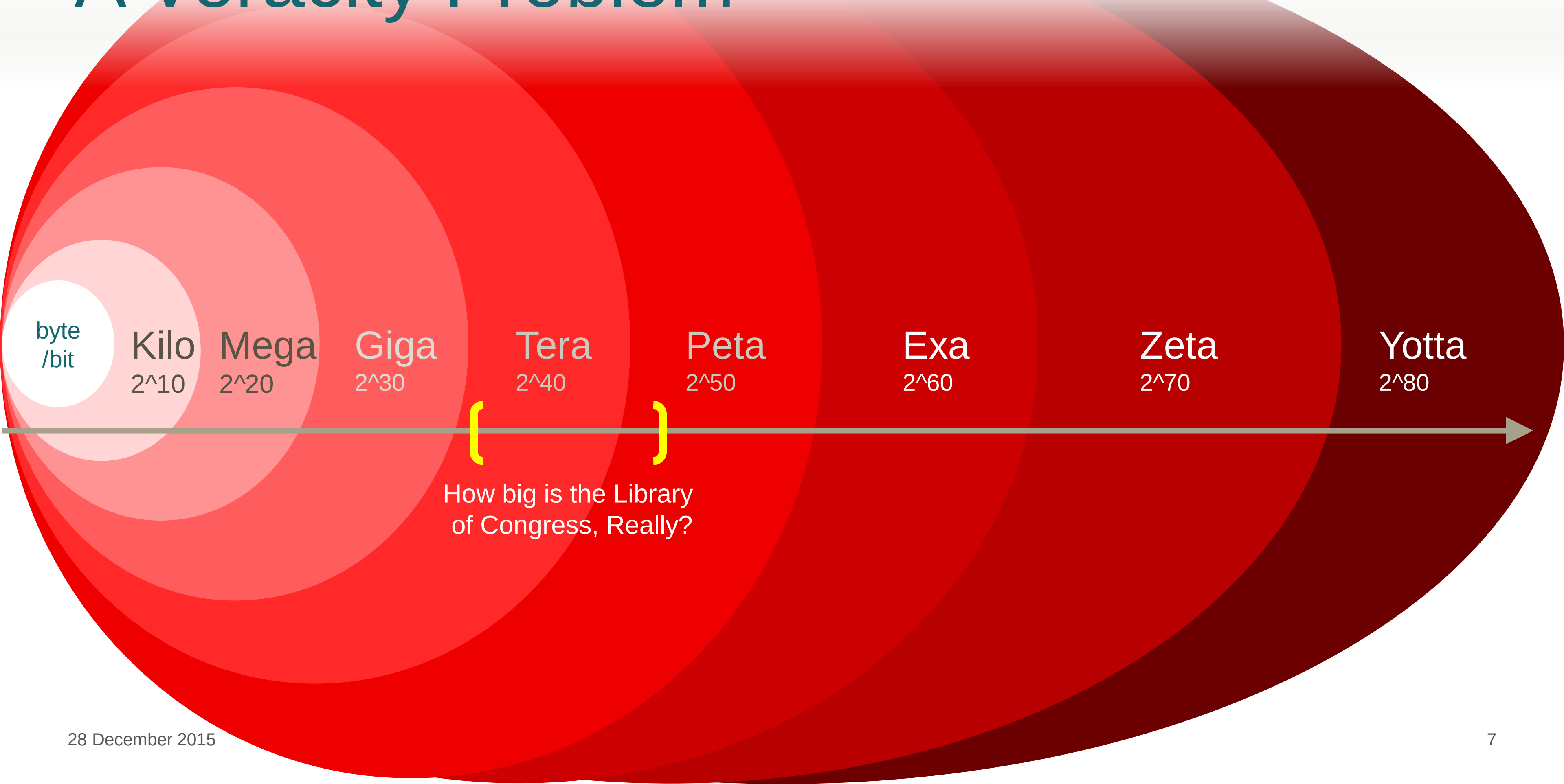
Where does Big Data Start?



Some Big Data Examples



A Veracity Problem



Big Data in Environmental Science

Real-time Deforestation Watch
World Resources Institute

Encyclopedia of Life

CUASHI

SLOSH Models
NOAA

National Hydrography Dataset
USGS

GCMs & Downscaled Models
UN.IPCC & Research Centers

Citizen Science
Oxford University

National Flood Interoperability Experiment

National Elevation Dataset
USGS

Open Water Data Initiative

Big Data in Environmental Science

“Trees can’t tweet”

Angel Hsu, Yale/IRI

Still a long way to go

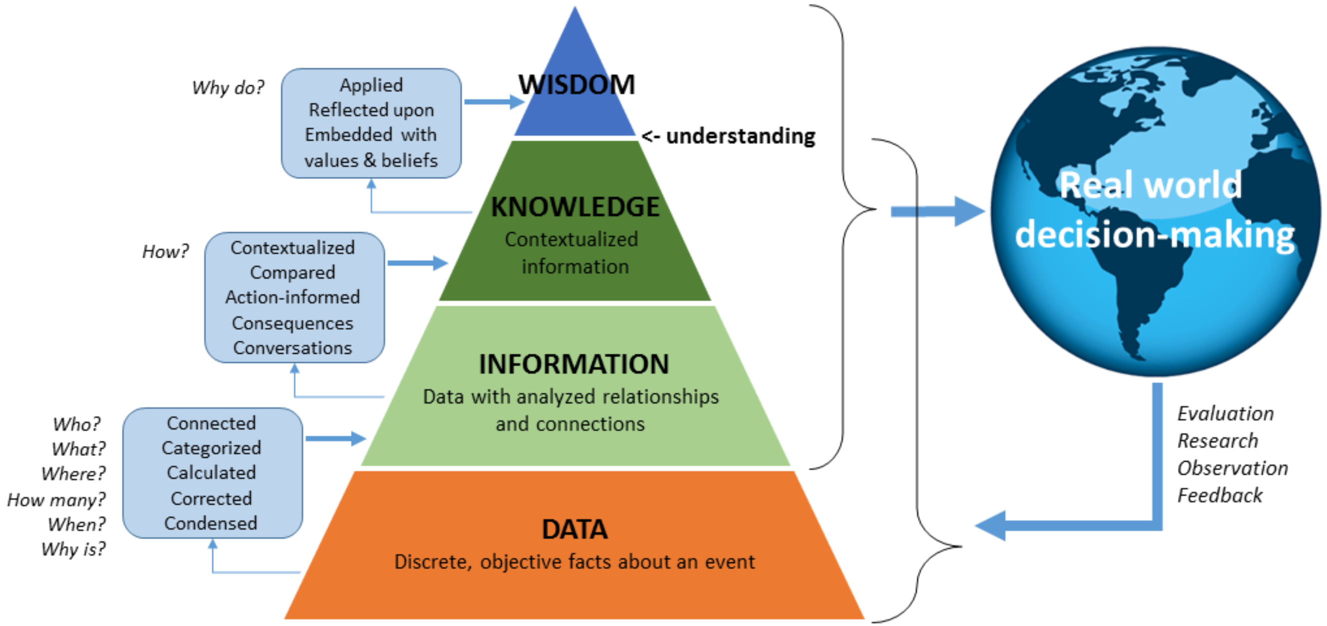
We also need people…

The best minds of my generation are thinking about how to make people click ads. That sucks.

Jeff Hammerbacher

The Challenge

Using the data to build knowledge, and even wisdom



Dbours, https://www.climate-eval.org/blog/answer-42-data-information-and-knowledge

North Slope Decision Support System

Background: Tool funded by U.S. Dept. of Energy in 2008.

Decision: Where should we build an ice road to explore for oil?

Stakeholders: Oil firms, regulators, tribes, enviros, researchers

Key Engineering Questions:

- Cost of road,
- How long do we have to build it?
- Does it avoid sensitive environmental areas?
- Does it avoid cultural resources?

Key Natural Systems Questions:

- Impact on Arctic Lakes under climate change

UNIVERSITY OF ALASKA FAIRBANKS

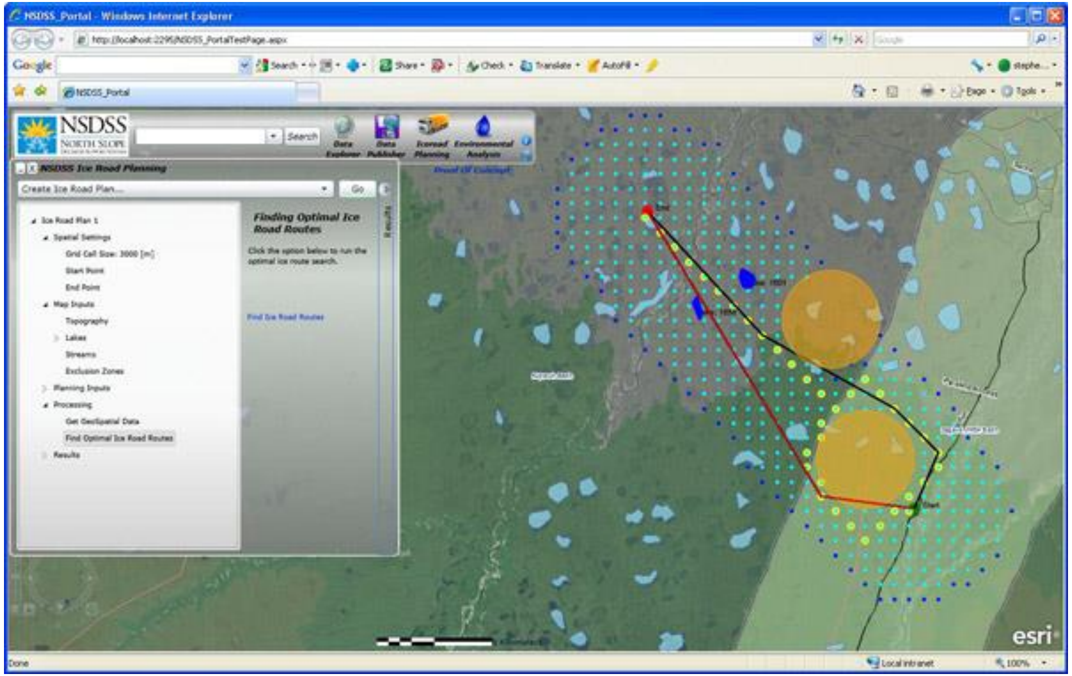
ATKINS

TEXAS A&M UNIVERSITY

North Slope Decision Support System

Data Needed:

- Topography
- Transportation
- Endangered Species
- Tribe lands
- Lakes
    - Bathymetry
    - Species
- Meteorology
- GCMs
- Regulatory Buffers

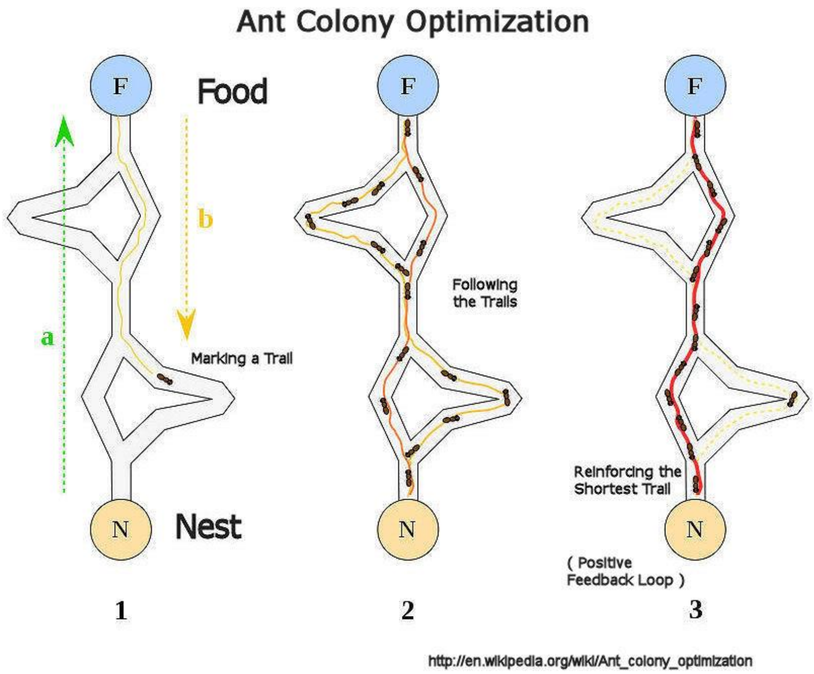


North Slope Decision Support System

Ant Colony Optimization to find best routes

Minimize:
Cost
Travel time
Risk (eg. slope)

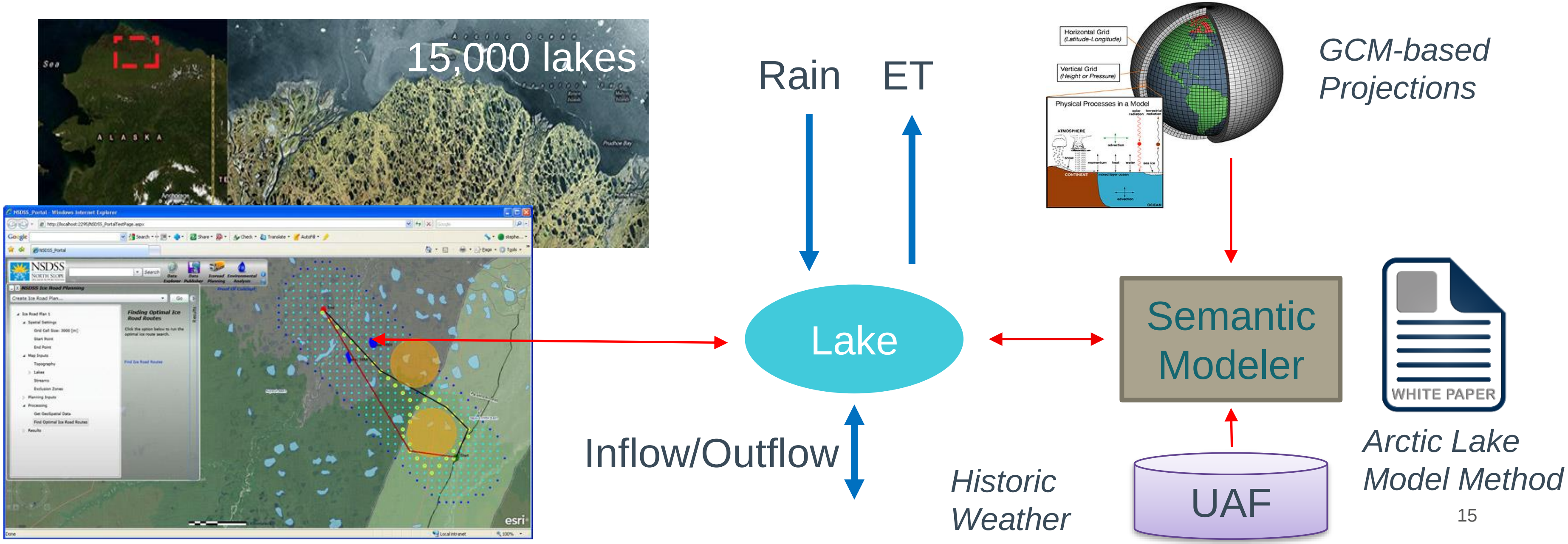
Avoid:
Sensitive Areas of Tundra
Endangered Species
Tribal Areas



North Slope Decision Support System

Suggestive Lake Climate Change Impact Modeling

Automatically create a water budget model and DO model for any lake on the landscape, with climate change forecast

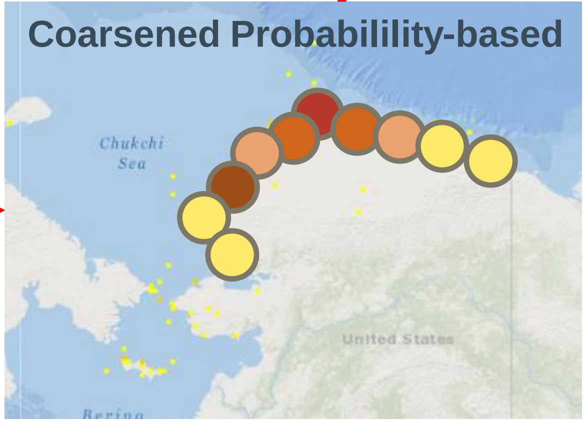


North Slope Decision Support System

Making Polar Bears Fuzzy (increasing veracity)

GBIF → Encyclopedia of Life

Observation "Fuzzer"



NSDSS Route Optimizer

StormCaster

Decision: How much should we grow our stormwater infrastructure?

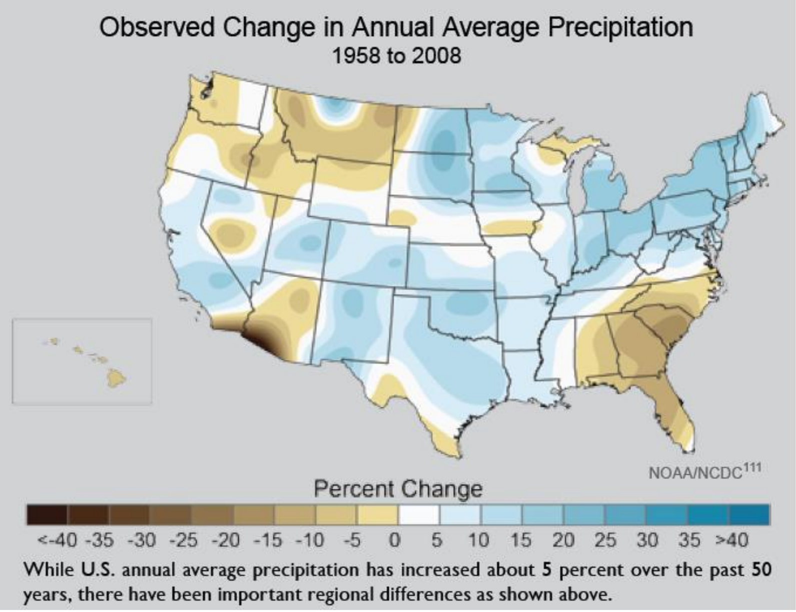
Stakeholders: Cities, Citizens, State and Federal Funders

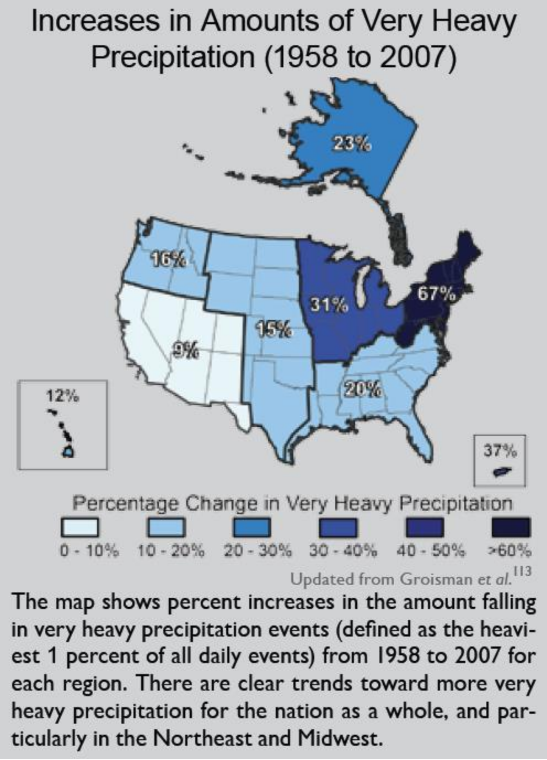
Key Questions:

- Is the 100 year storm changing, and how much?
- What is the best and worst case?
- When will these changes occur?

ATKINS
TEXAS A&M UNIVERSITY®

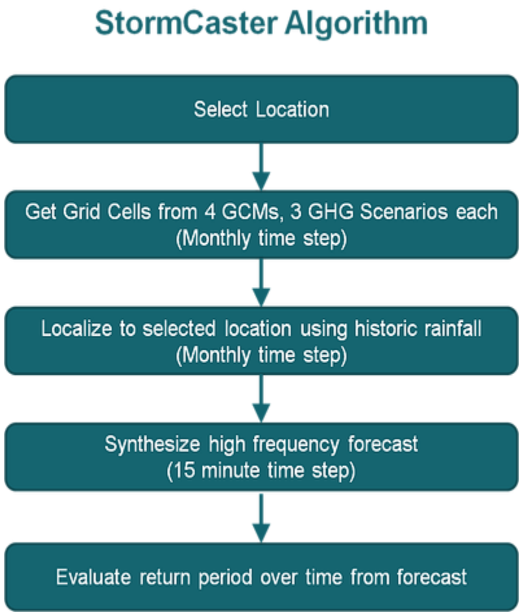
Is the 100 year storm changing?

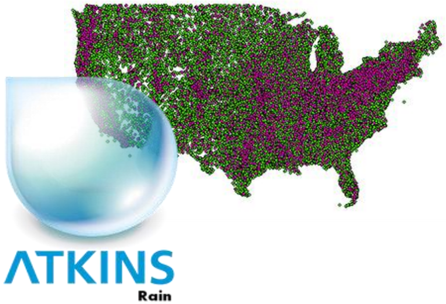




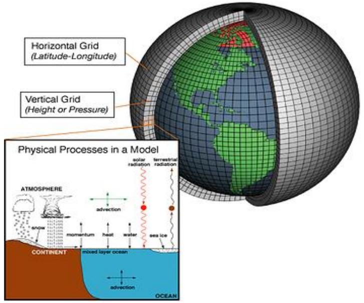
-Global Climate Change Impacts Report (2009)
U.S. Global Change Research Program

Novel climate science to a tool stakeholders can use.

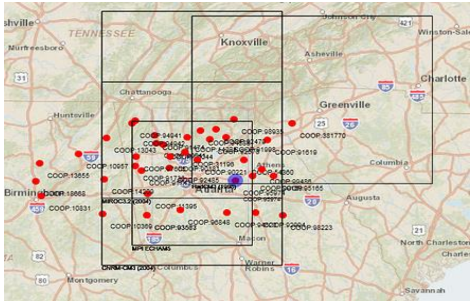


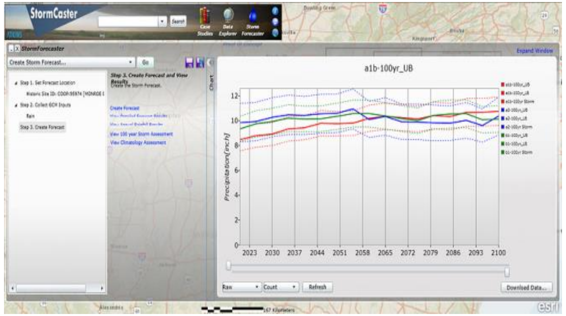


Historic Rainfall at 15 minute time-step provides local climate and high frequency



Precipitation from multiple monthly General Circulation Models provide projections in climate trend.





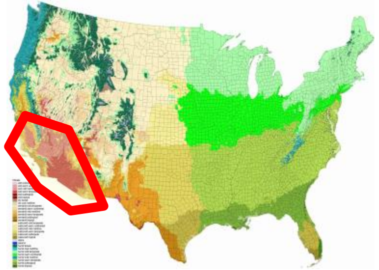
A forecast of Atlanta, GA, USA shows a general increase in 100 year storm over the 21st century. The ensemble of models agree generally in the upward trend, though there is considerable spread in the degree of increase.

StormCaster Network: Create your own algorithm.

Jack Submits his method to the Network

The Code

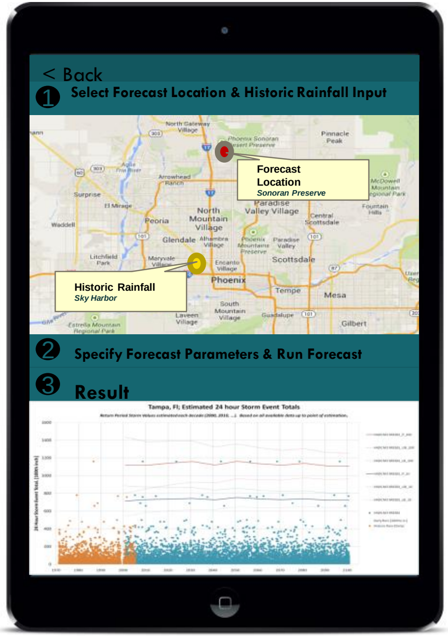
Documentation



Map of where the method is applicable

StormCaster

The method is suggested as one Jill can use when she creates a forecast for Sonoran Preserve.



Transportation Planning in the U.K.

Decision: Where should infrastructure investments in England’s M25 be made?

Stakeholders: Traveling public, business, municipalities

Key Engineering Questions:

- Where is travel demand highest?
- How much should infrastructure be increased?
- What is the effect of teleworking?

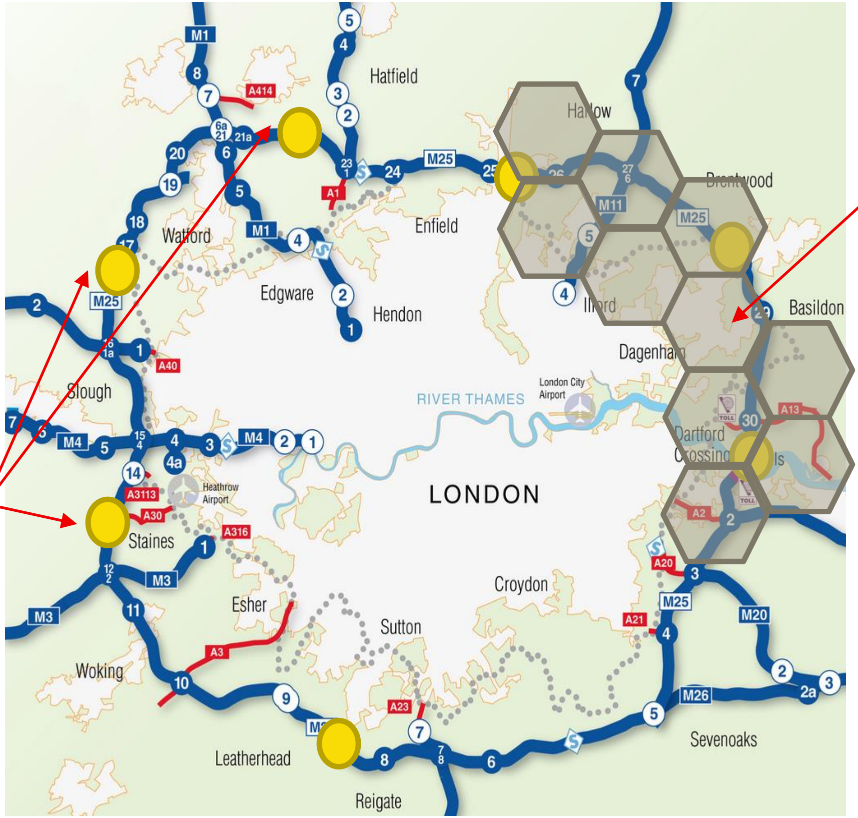
Key Natural Systems Questions:

- GHG emissions, water quality, impervious area

Transportation Planning in the U.K.

Traditional Method

Road-side surveys
Small/Biased Sample
Causes Congestion



Big Data Method

Big Cell Company
Tracks Full Trip Through Cells
Origin and Destination
Much larger sample
(40% of trips)

Transportation Planning in the U.K.

Velocity and Privacy

A velocity problem: 1.2B events per month. This problem will grow as we go from 4G to 5G and introduce the internet of things.

Solution: On-the-fly aggregation and deletion

Also solves the problem of sensitivity: Data is anonymized in the process.

Growth Forecasting

Decision: Where should we enhance the Atlanta Transportation System to reduce future congestion?

Stakeholders: ARC, City Governments, traveling public, business, municipalities

Key Engineering Questions:

- What influences growth in commercial and residential landuse?
- How will growth occur across Atlanta in the future?
- How will travel demand change as result?
- How will demand for other services change?

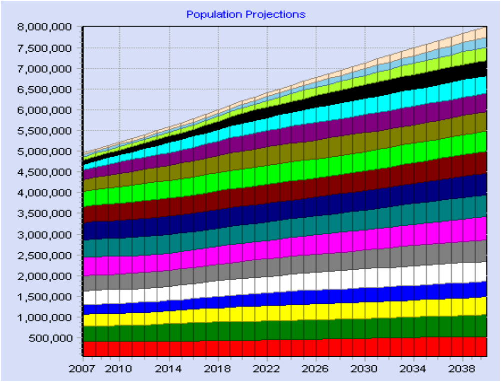
Key Natural Systems Questions:

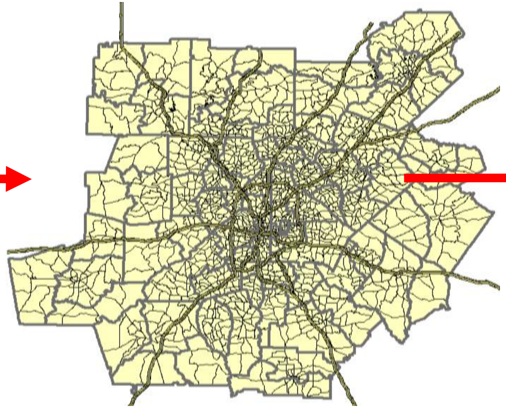
- GHG emissions, impervious area

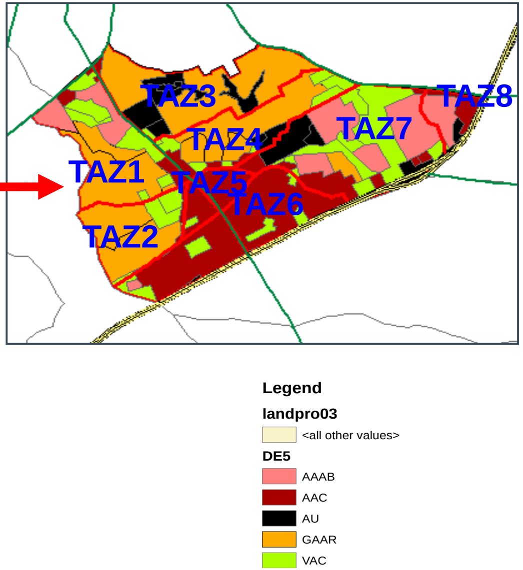
Regional to Local; Population to Landuse

Goal - Disaggregate regional scale jobs/households projection into traffic analysis zone projection.

Population & Employment







Source: REMI

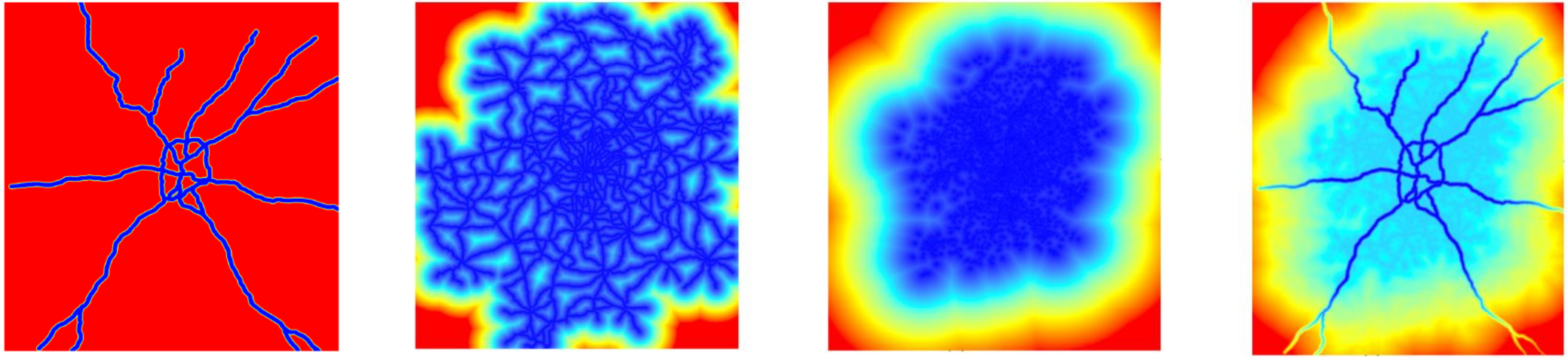
(regional economic models, inc.)

Estimating Likelihood of Development

L raster is composed of several term rasters, 1 for each factor.

$$L(x,y)_{Category} = \alpha_{1,Category} {}^{*}F(x,y)_1 + \alpha_{2,Category} {}^{*}F(x,y)_2 + \cdots + \alpha_{N,Category} {}^{*}F(x,y)_N$$

Example L Raster for Commercial Landuse



α_1 Expressway + α_2Major Roads + α_3Commercial LU = Likelihood, L

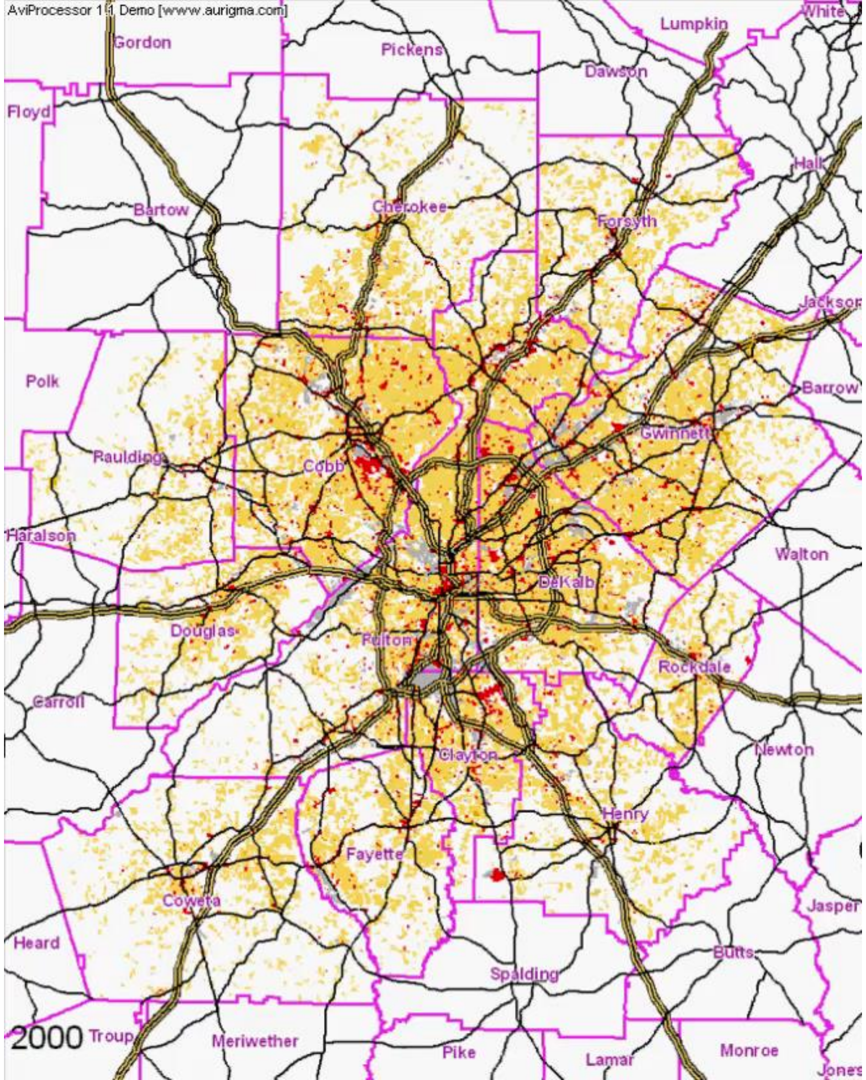
A rich result.

22 Million Acres

6 Million – 10 Million People

Rich source for:

- **Travel Demand**
- **Electrical Demand**
- **Communications Demand**
- **Water/Wastewater**
- **Stormwater**
- **…**



Silos are Dead

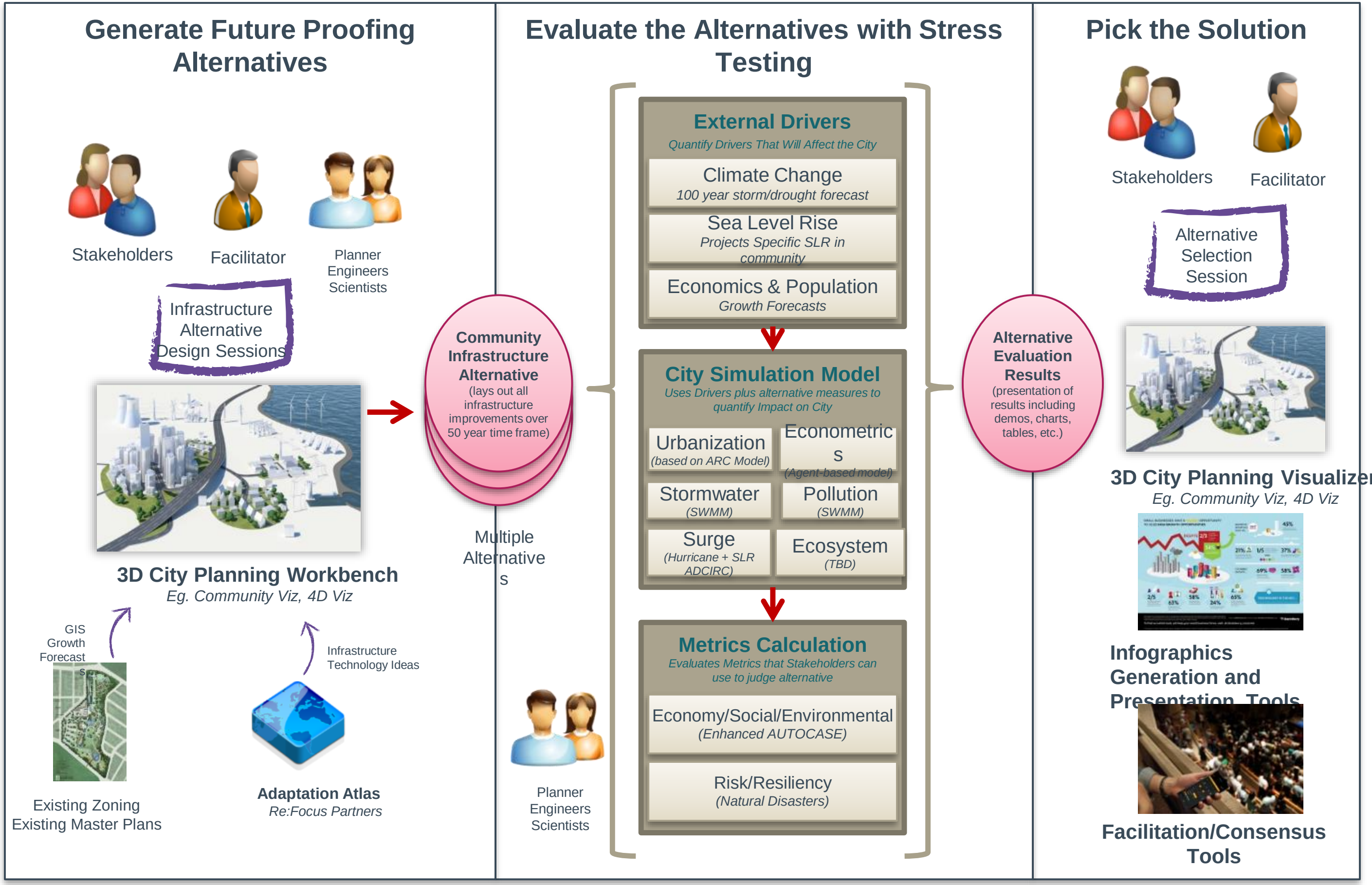
Especially Analysis Ones

Silos are Dead

Especially Analysis Ones

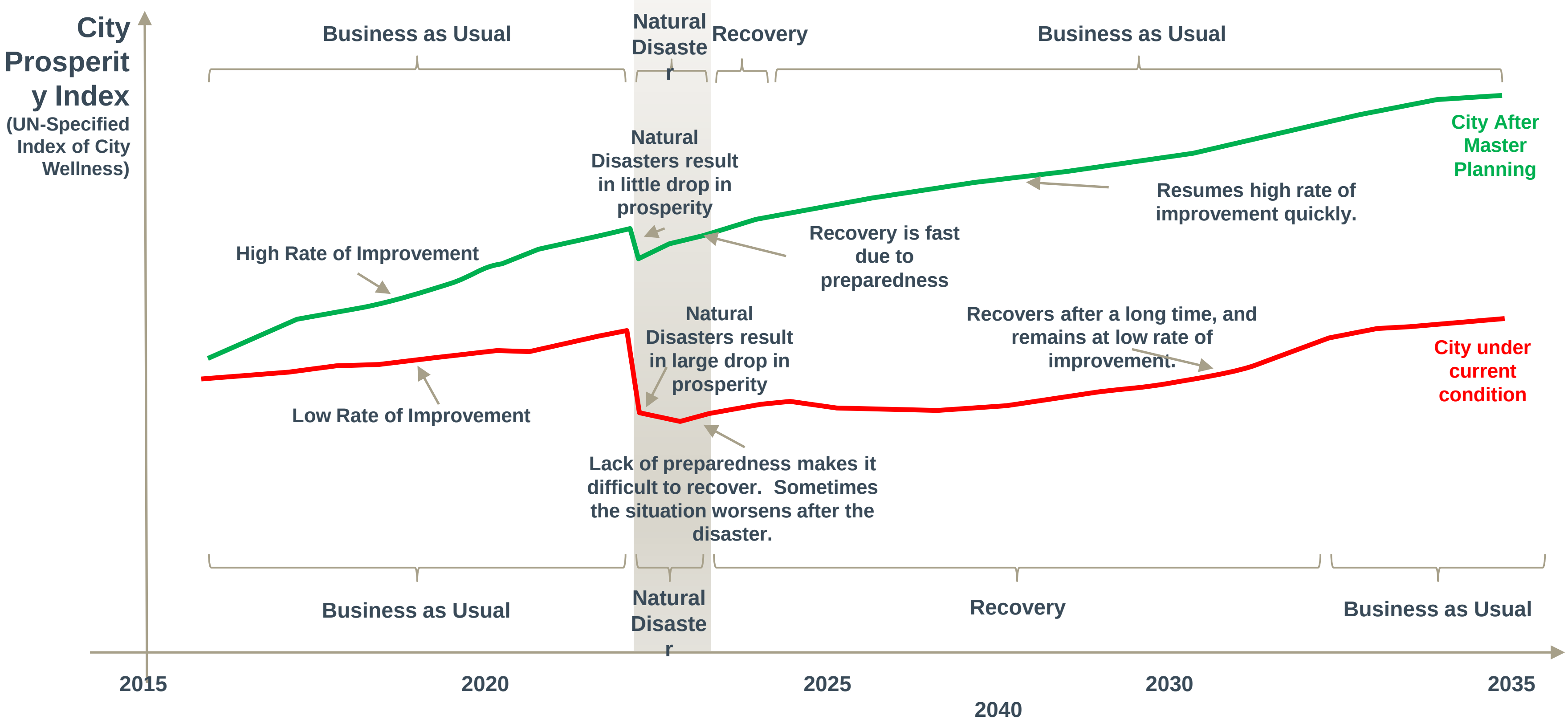
Veracity Problem

Future Proofing Cities

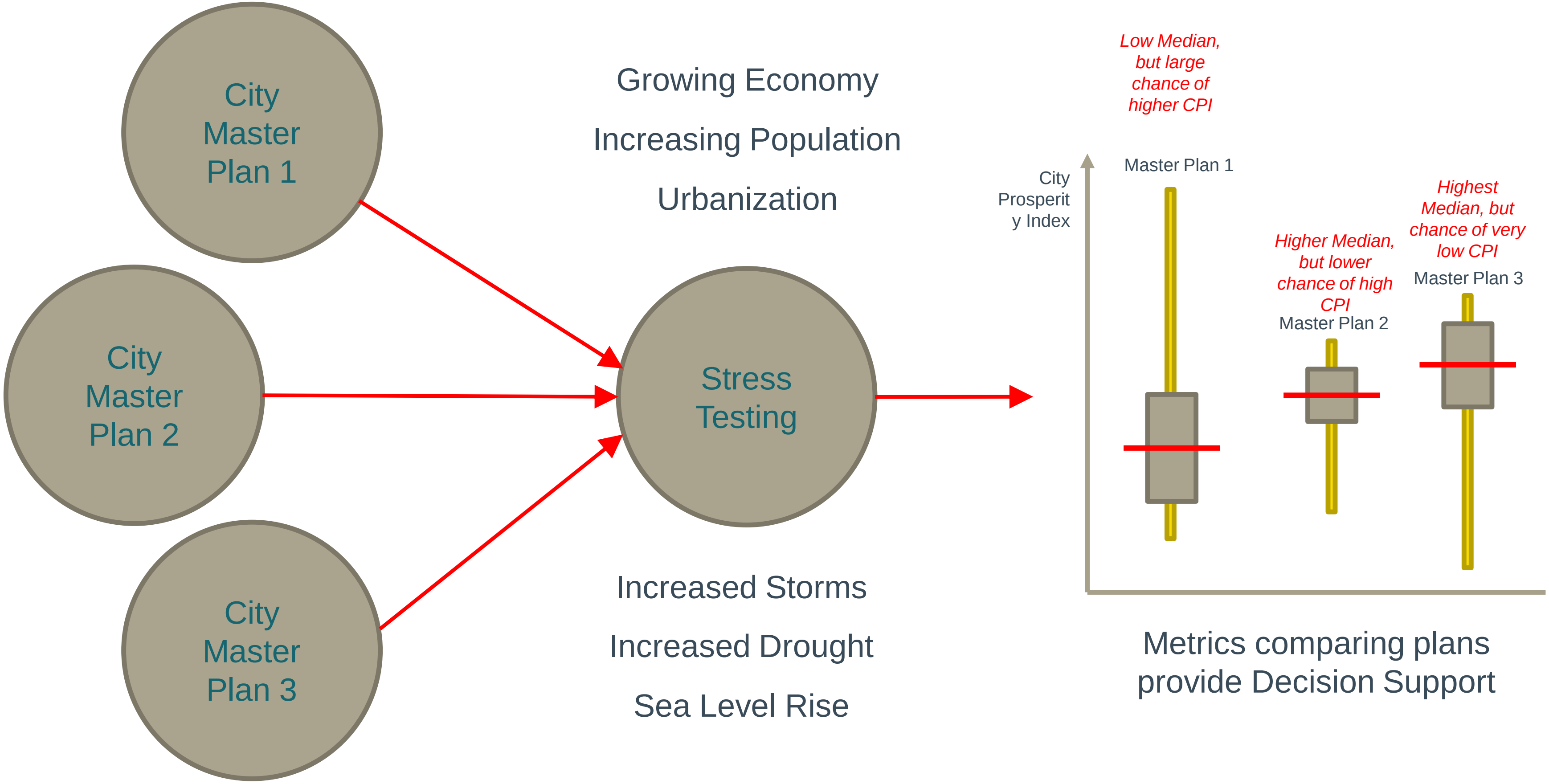


Community Stress Testing

In Stress Testing, we simulate a natural disaster in your city, and evaluate the resiliency of your future proofing measures.



Stress Testing a community under multiple master plan scenarios



Questions?