Data Challenges in Environmental Research

John McGee, Brian Blanton,
Jeff Heard, Oleg Kapeljushnik

RENCI, University of North Carolina at Chapel Hill

Projects with common requirements

Current

- NWS WxEM, Big Board, NOAA Decision Support Prototype
- NCB Prepared
- CI-BER prototypes
- SPH grant tracking app
- SPH farmers' market locator
- Coastal Risk Analysis

Potential

- Public health informatics
- Environmental science
- Decision support / Situational understanding
- Geographic collection browsing / Library Sciences
- NC First, version 2

Exploring a GeoAnalytics Framework

- Rapidly develop and deploy prototypes and solutions
- Reduce the barrier of entry to using geographic data for science
- Scale horizontally to Big Data, its update frequency, access patterns, and management requirements
- Vet, integrate, and federate open source geography tools
- *Led by Jeff Heard at RENCI*
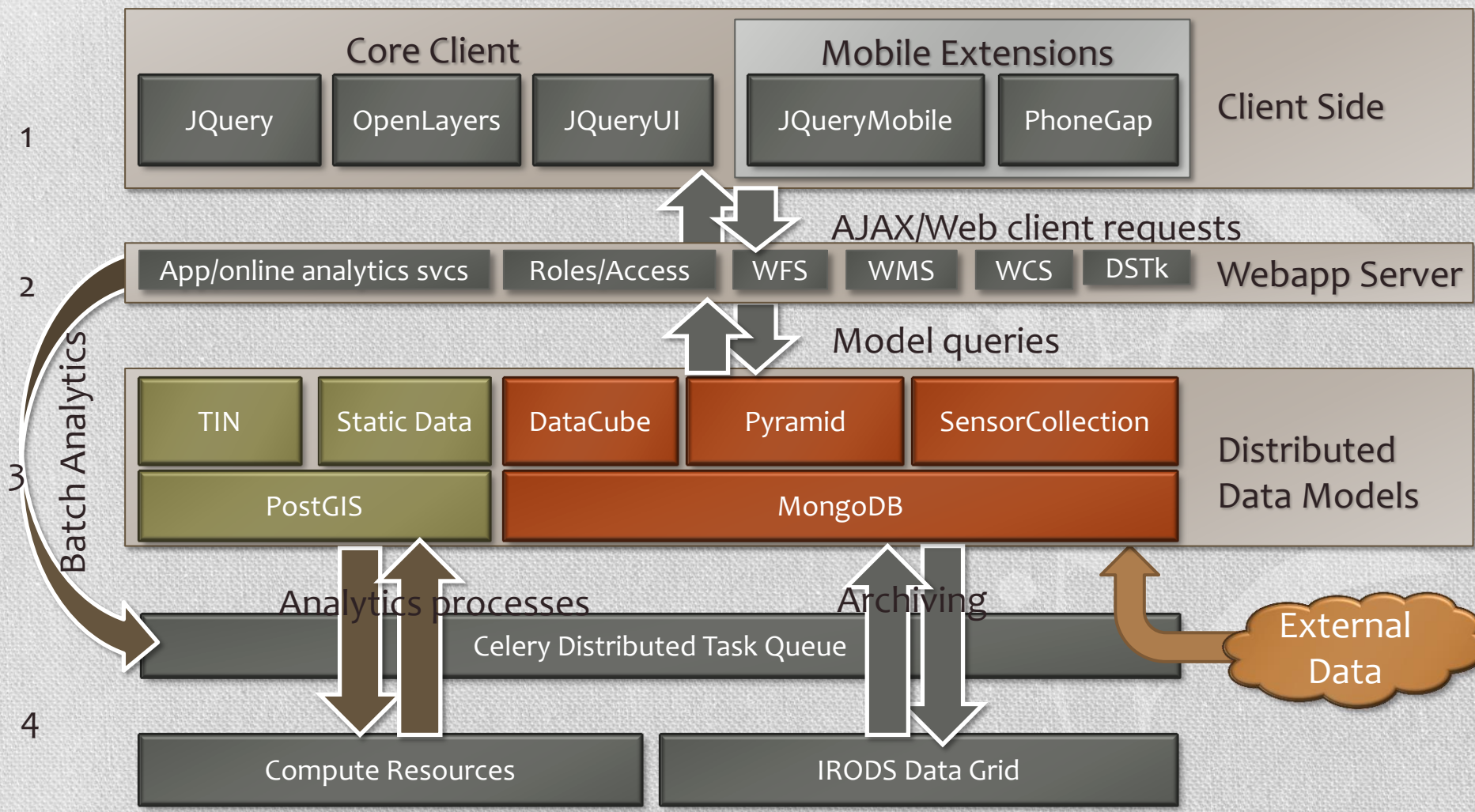
Data Cube

Store/query versioned, 4d regular-gridded data: X/Y, Elevation, Time

- Stores multiple elements as a set
- High level query routines support
  - re-projection
  - filtering
  - masking
- Server-side queries return Fortran-style arrays that can be processed by BLAS, native code, or Python's Numpy/Scipy
- Client-side queries can be filtered through OWS or custom services

Distributed Task Queueing

- Schedule tasks for regular execution
- Defer long-running tasks for batch processing
- Tasks are developed per-application and automatically discovered and scheduled by Geoanalytics, or can be used in server-side program as functions with deferrable execution
- Example tasks include:
  - Sunsetting ADCIRC and sensor data
  - Running batch analytics on an offline dataset

Communicating Coastal Risk Analysis

science lead: Brian Blanton, RENCI

Coastal Risk Analysis

- Significant worldwide populations live along coastlines
- Risks include storm surge inundation, precipitation driven flooding, waves, coastal erosion
- Increased risk likely as impacts of a changing climate are felt through elevated sea levels and potentially increased storm intensity and frequencies
- Involves complex science and large amounts of diverse data
- New methods required for communicating to the public and government/business decision makers

Microsoft® Research
FacultySummit

Some questions of interest

- Do flood levels in the current climate (eg 100-year surge level) undergo a change in frequency such that they occur more often in a plausible future climate?

  - Under what conditions does the 100-year surge level, currently used to set flood insurance rates, become the 50-year surge level?

- Given anticipated coastal land use and development changes over the next century, under which climate scenarios are impacts expected to be greatest?

  - Which areas of the coast should expect the largest changes in risk?

ADCIRC Coastal Modeling

- Statistically meaningful 50 and 100 year flood plain mapping
  - Development of a high quality TIN
    - http://www.renci.org/wp-content/pub/techreports/TR-08-05.pdf

  - Serial monte-carlo pre-processing step
 - Hundreds of 256-way parallel ADCIRC runs
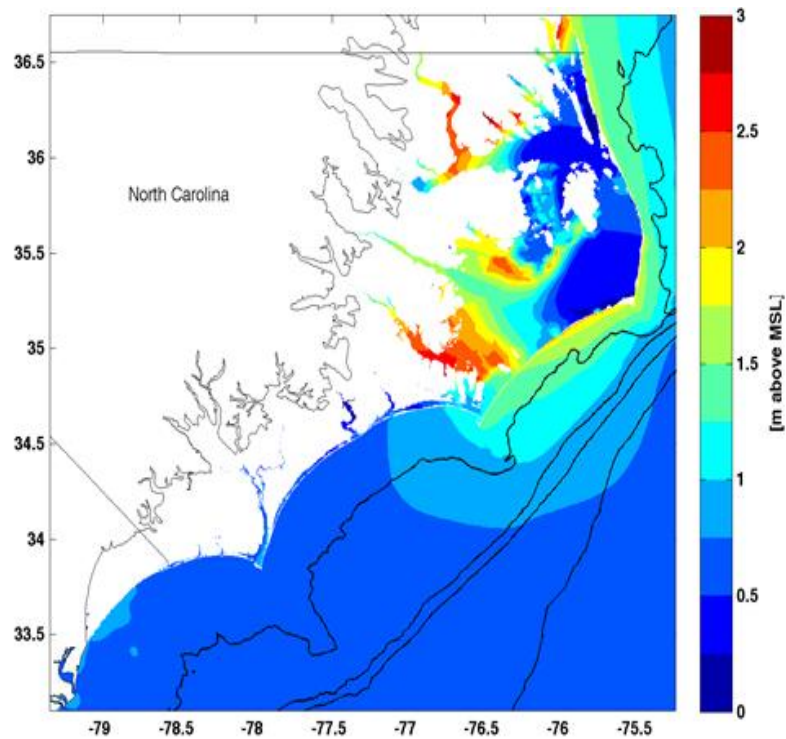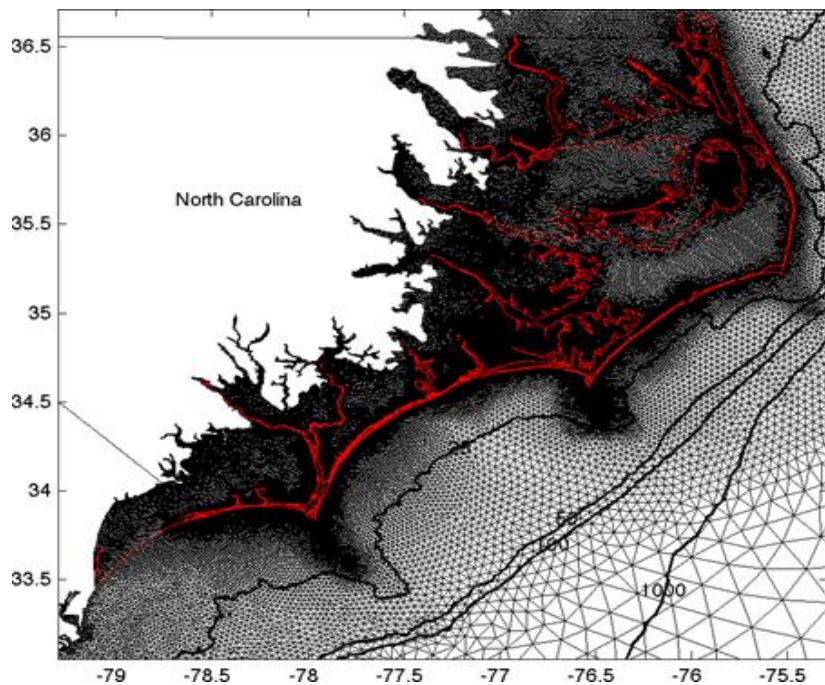
Exploring WWT for Viz and Communication

- Using the SDK, bring data into the environment from:

  - Tiled orthophotos
  - OGC Sensor Observation Service
  - OData Service
  - NetCDF
  - LIDAR data

- NC Emergency Management's Geospatial Technology Management Office

  - statewide: 805 x 298 kilometers

  - orthophotos 0.15 meters/pixel
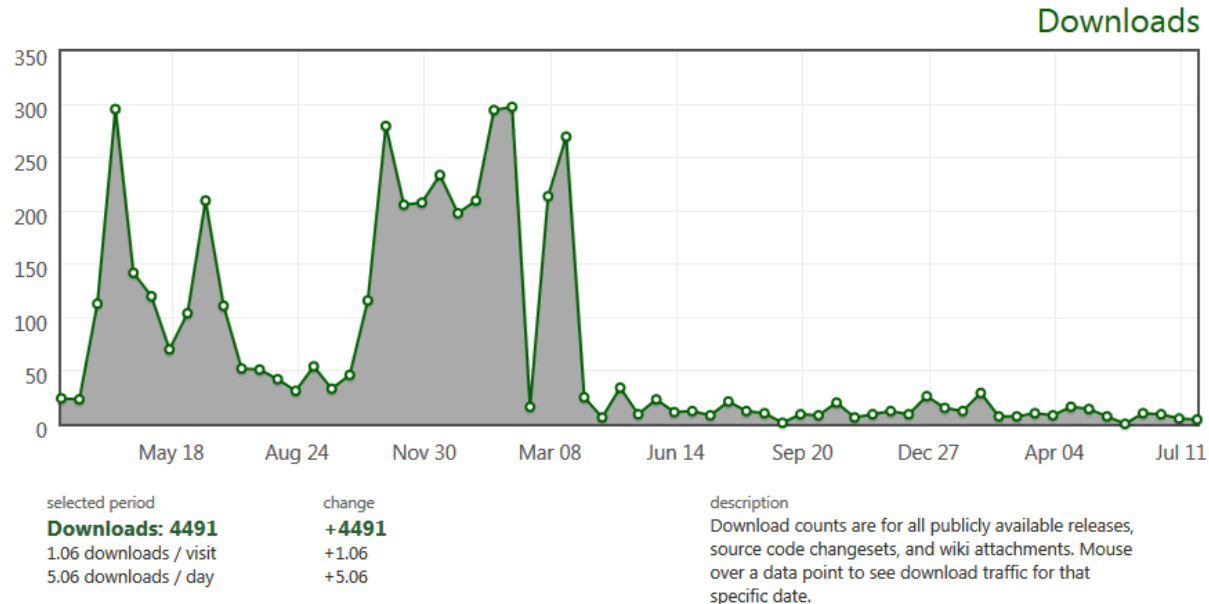 - LIDAR at 3 to 6 meters/pixel
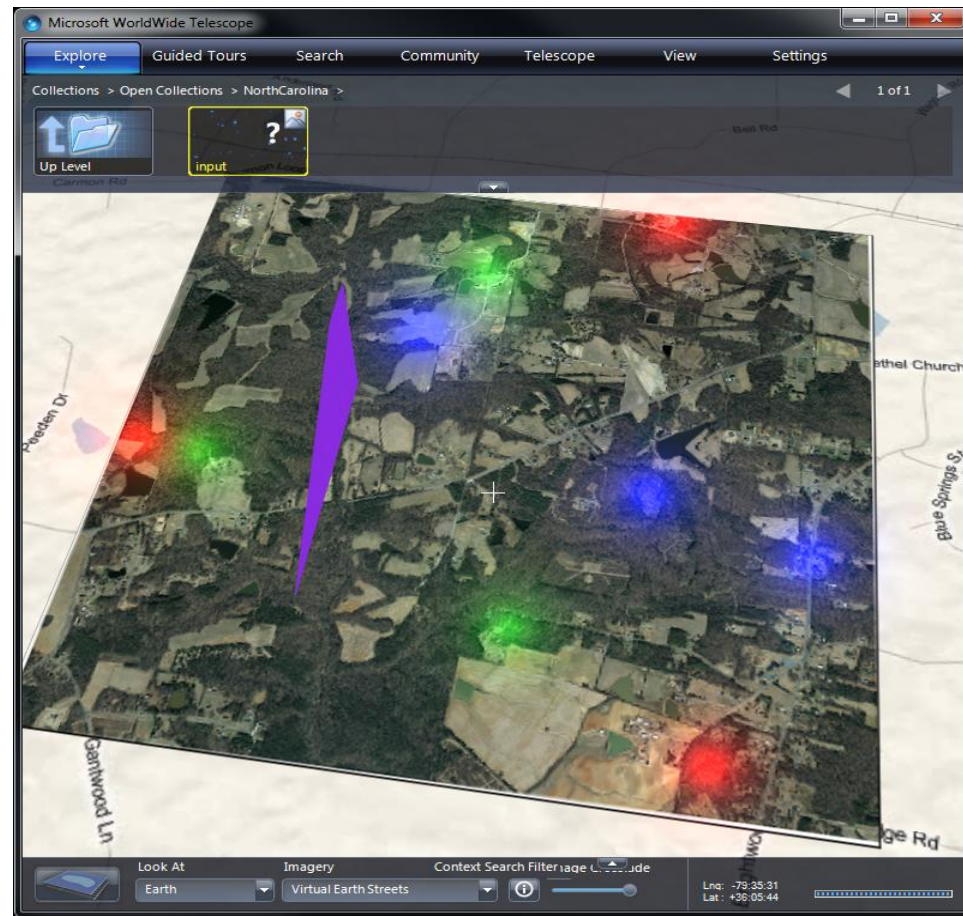
OGC service implementation
Implemented Sensor Observation Service with .Net

- .NET class library and web service template for implementation of the OGC

- http://ogc.codeplex.com

Downloads

selected period
Downloads: 4491
1.06 downloads / visit
5.06 downloads / day

change
+4491
+1.06
+5.06

description
Download counts are for all publicly available releases, source code changesets, and wiki attachments. Mouse over a data point to see download traffic for that specific date.

Unintended Consequences

Required programmatic access to Linux cluster for ADCIRC results

- Created an SSH Client Library for .Net
- Has become an active community, especially around providers of devices supporting SSH for command and control
- http://sshnet.codeplex.com

Stats for the last... | 6 months

Downloads

selected period	change	previous period	description
Downloads: 6559	**+4791**	**1768**	Download counts are for all publicly available releases, source code changesets, and wiki attachments. Mouse over a data point to see download traffic for that specific date.
0.82 downloads / visit	+0.30	0.52 downloads / visit	
36.24 downloads / day	+26.47	9.77 downloads / day	

DEMO

environmental data visualization
with WWT

- model input/output
- sensed data
- image/LIDAR assets

Microsoft® Research
FacultySummit

FUTURE/WORLD
2011   2031