October 18, 2019
Proposal writing kickoff:

Established Program to Stimulate Competitive Research (EPSCoR) Research Infrastructure and Improvement (RII) Track 1 grant proposal process

Arne Bomblies
State Director, Vermont EPSCoR
Associate Professor, Civil and Environmental Engineering

EPSCoR
Research Infrastructure improvement Track-1 (RII Track 1)

Sensors to Policy: Harnessing the Data Revolution to Transform Management of Vermont's Working Landscape and Promote Wellbeing

Donna Rizzo
Civil and Environmental Engineering

REMINDER 1- NEED A LARGE CROSS-CUTTING CONCEPT THEME THAT:

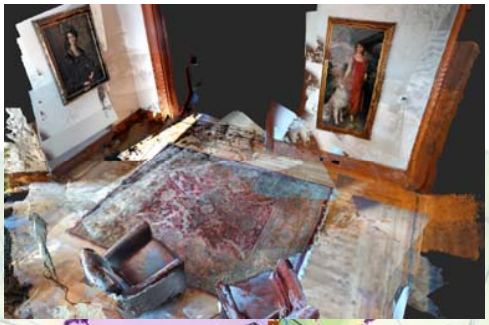
1) addresses one of NSF's Grand Challenges

SENSORS => DATA SCIENCE => POLICY

- **Harnessing the Data Revolution**
- **NSF Includes**
- **Mid-scale Research Infrastructure**
- **Future of Work at the Human-Technology Interface**

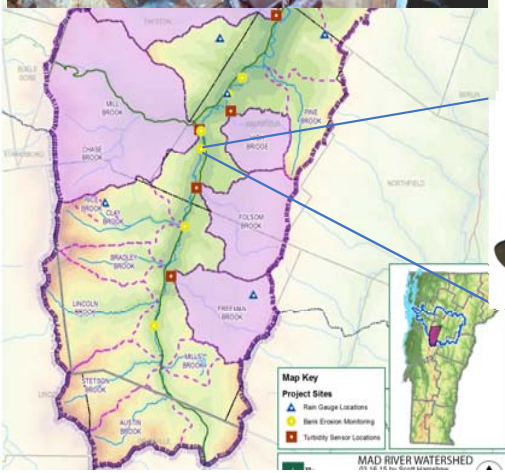
REMINDER 1- NEED A LARGE CROSS-CUTTING CONCEPT:

1) ADDRESSES ONE OF NSF'S GRAND CHALLENGES



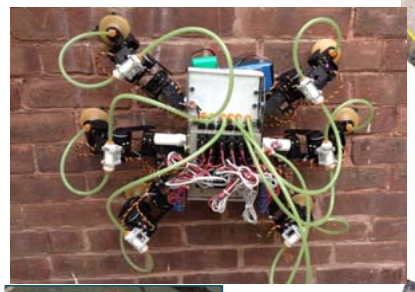
SENSORS => DATA SCIENCE => POLICY

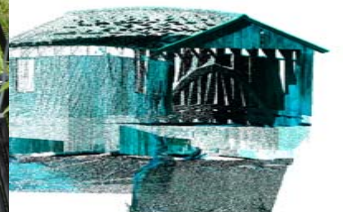
- **Harnessing the Data Revolution**



AUTONOMOUS (IN SITU)

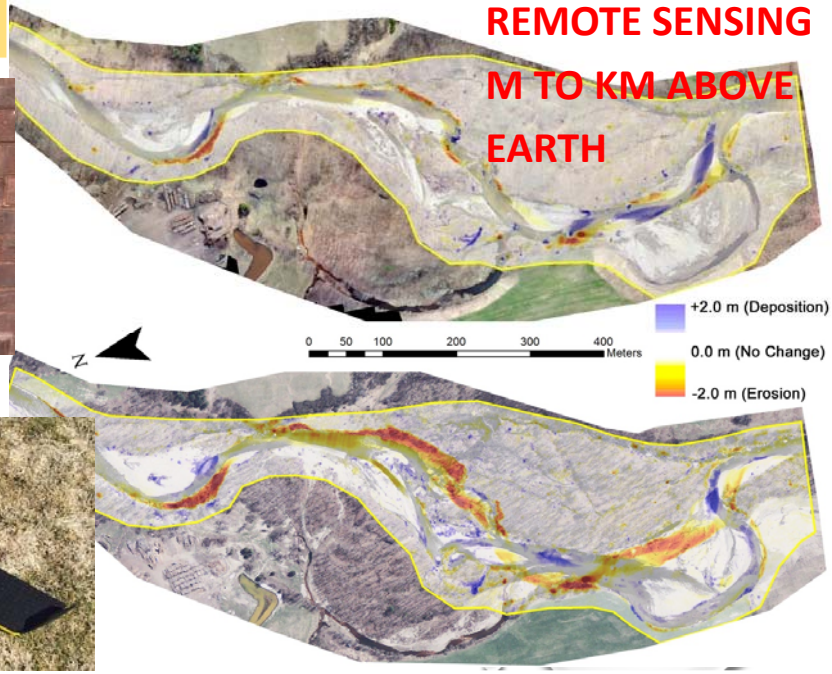
Turbidity monitoring





REMOTE SENSING

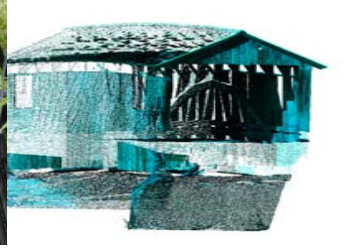
M TO KM ABOVE EARTH

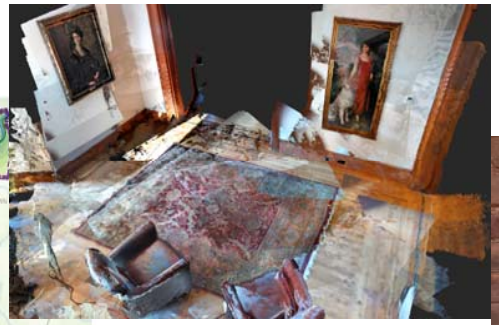


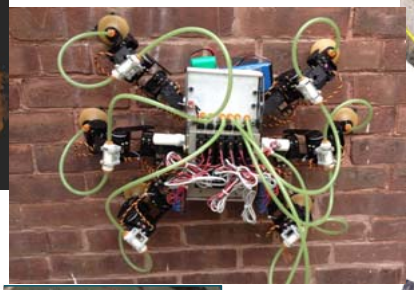
SENSORS => DATA SCIENCE => Policy:

EARTH SYSTEM DATA ALREADY STORES VOLUMES WELL BEYOND DOZENS OF PETABYTES & TRANSMISSION RATES EXCEEDING HUNDREDS OF TERABYTES/DAY

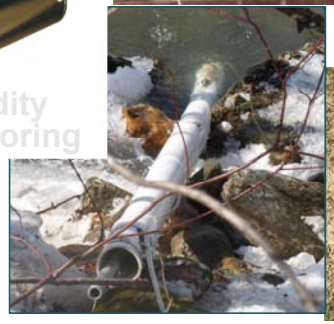
- MANY ADVANCES IN WATERSHED SCIENCE [@ UVM AND ELSEWHERE] HOWEVER, EXAMPLES OF <u>LEVERAGING THOSE DATA TO INFORM OR ASSESS POLICY & WATERSHED MANAGEMENT DECISIONS ARE NONEXISTENT</u>.

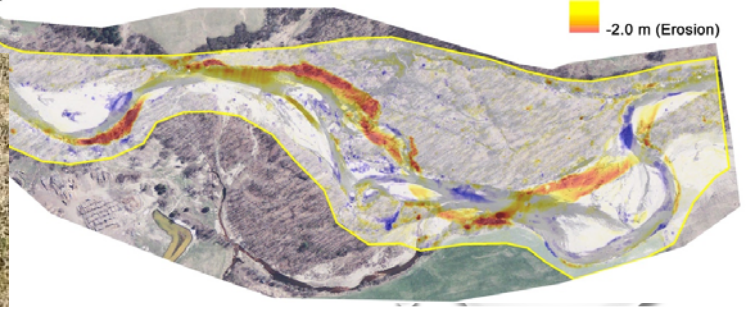






Turbidity monitoring





SENSORS => DATA SCIENCE => Policy:

EARTH SYSTEM DATA ALREADY STORES VOLUMES WELL BEYOND DOZENS OF PETABYTES & TRANSMISSION RATES EXCEEDING HUNDREDS OF TERABYTES/DAY

- MANY ADVANCES IN WATERSHED SCIENCE [@ UVM AND ELSEWHERE] HOWEVER, EXAMPLES OF LEVERAGING THOSE DATA TO INFORM OR ASSESS POLICY & WATERSHED MANAGEMENT DECISIONS ARE NONEXISTENT.

CHALLENGES:

- The temporal and spatial dynamics of ecosystems are difficult to measure simultaneously, either requiring large, costly equipment installations at limited locations that lack the spatial distribution necessary to accurately capture system dynamics for scaling or extrapolation.

- "SMART" SENSORs with respect to TIME

REMINDER 2 - NEED A LARGE CROSS-CUTTING CONCEPT:

1) addresses one of NSF's Grand Challenges, and
2) has <u>regional or jurisdictional importance</u>,

MUST HAVE POTENTIAL TO IMPROVE <u>VT'S FUTURE RESEARCH</u> IN 1 OR MORE OF THE FOLLOWING:

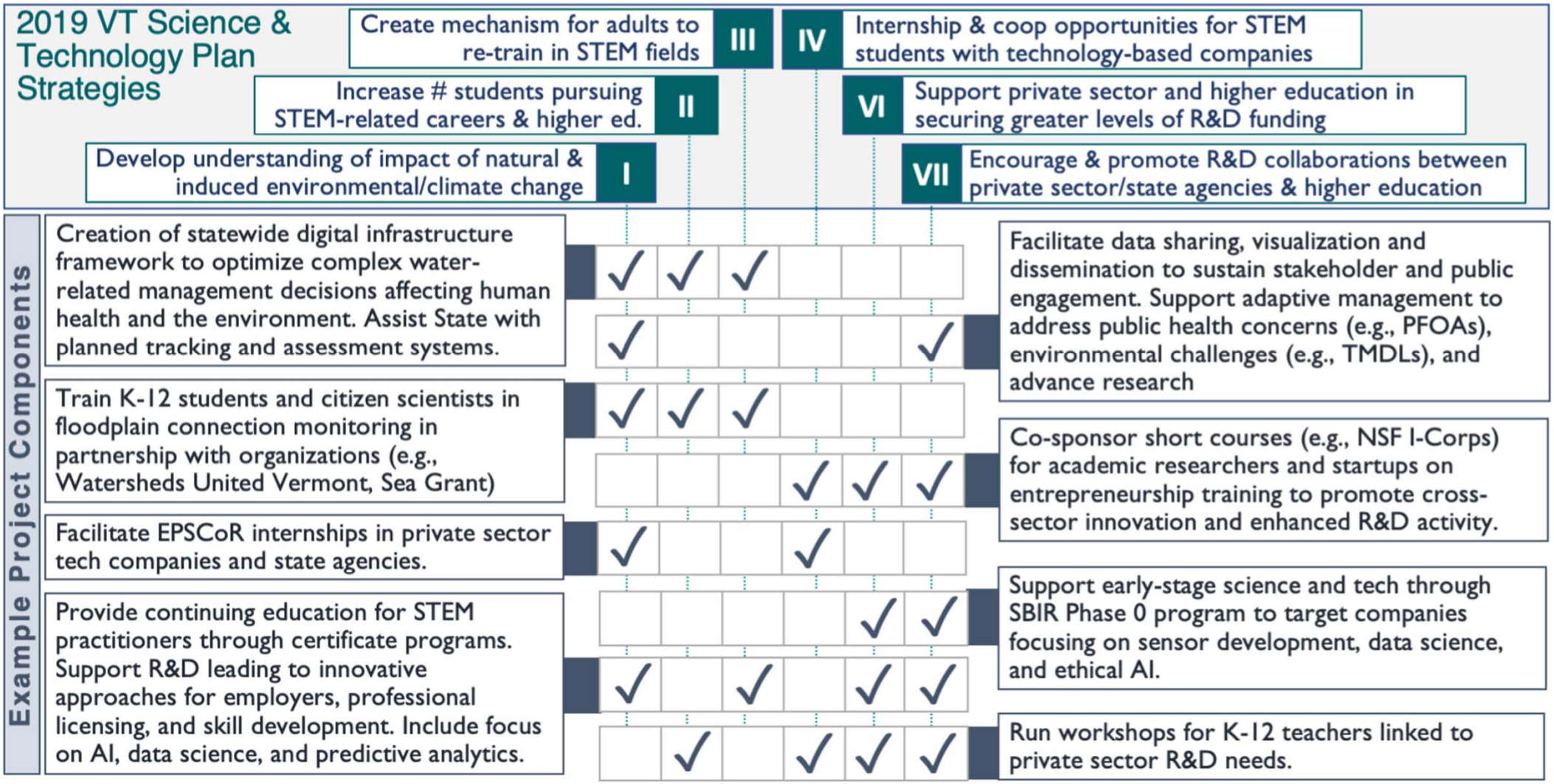
1) PROTECTING THE ENVIRONMENT, (Protection of Water & the Environment)

2) HEALTHCARE TECHNOLOGY AND DELIVERY, (Biomedical & Healthcare)

3) ADVANCED MANUFACTURING AND

4) COMPUTING TECHNOLOGIES (Complex System Center, Big Data, AI, Games)

VT Science & Technology Plan



ALL VT Science & Technology Plan GOALS:

- focus on **increasing the STEM workforce** via training, mentoring and educating Vermonters or relocation of skilled workers to Vermont.
- MuST integrate researchers, institutions and organizations **throughout Vermont** to develop a diverse, well-prepared, STEM-enabled workforce necessary to sustain research competitiveness and catalyze economic development.

PERSPECTIVE

https://doi.org/10.1038/s41586-019-0912-1

Deep learning and process understanding for data-driven Earth system science

Markus Reichstein[1,2*], Gustau Camps-Valls[3], Bjorn Stevens[4], Martin Jung[1], Joachim Denzler[2,5], Nuno Carvalhais[1,6] & Prabhat[7]

Machine learning approaches are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data, but current approaches may not be optimal when system behaviour is dominated by spatial or temporal context. Here, rather than amending classical machine learning, we argue that these contextual cues should be used as part of deep learning (an approach that is able to extract spatio-temporal features automatically) to gain further process understanding of Earth system science problems, improving the predictive ability of seasonal forecasting and modelling of long-range spatial connections across multiple timescales, for example. The next step will be a hybrid modelling approach, coupling physical process models with the versatility of data-driven machine learning.

DATA SCIENCE PORTION OF THIS PROPOSAL WILL FOCUS ON:

(1) EXTRACTING KNOWLEDGE FROM THE EXPLOSION OF EARTH SYSTEM DATA, &

(2) DERIVING NEW COMPUTATIONAL ALGORITHMS (ML, AI, DL, STATS) CAPABLE OF LEARNING MORE FROM DATA THAN TRADITIONAL DATA ASSIMILATION APPROACHES [SPECIFICALLY - AUTOMATIC EXTRACTION OF ABSTRACT (SPATIO-TEMPORAL) FEATURES—(ML =>UNIVERSAL APPROACH IN GEOSCIENTIFIC CLASSIFICATION, AND CHANGE- AND ANOMALY-DETECTION PROBLEMS). DL NOW USED TO EXPLOIT SPATIAL & TEMPORAL STRUCTURES IN THE DATA, FEATURES THAT TRADITIONALLY ARE PROBLEMATIC FOR ML TO EXTRACT).

Description of Proposed Center of Excellence :

Today's Homework:
Identify Groups & Cluster IMPORTANT Science Questions

POSSIBLE SCIENCE QUESTIONS:

How do human activities and behaviors operate in a nonstationary climate to influence catchment function leading to changing rates of water partitioning, storage, transmission and release?

How have these modifications to catchment function manifested in water-related impacts to human health and the environment, as moderated by internal and external thresholds?

Can generalized (or regionalized) patterns of catchment qualities be discerned and uncertainties be quantified from complex and Big Data sets generated by diverse and distributed sensor arrays?

What are the varying capacities for coupled human and catchment systems to mitigate and adapt to impacts from a nonstationary climate to ensure sustainable quantities and qualities of freshwater?

Decision Theory (Scenarios to Solutions)

How does the monitoring of human activities, behaviors (e.g., things to monitor) operate in a nonstationary climate to optimize science-informed policy or practice (built & natural) and human wellbeing?

How does the monitoring of human activities and behaviors operate in a nonstationary climate to influence catchment function (i.e., rates of water partitioning, storage, transmission and release,)?

How to biological and physical systems function across gradients in human-dominated landscapes (detect change & measure dynamics)?

How do we measure sources and sinks across a human-dominated landscape?

How have these modifications to catchment function manifested in water-related impacts to human health and the environment, as moderated by internal and external thresholds?

Can generalized (or regionalized) patterns of catchment qualities be discerned & uncertainties be quantified from complex & Big Data sets generated by diverse and distributed sensor arrays?

What are the varying capacities for coupled human and catchment systems to mitigate and adapt to impacts from a nonstationary climate to ensure sustainable quantities and qualities of freshwater?

RII Track-1 Awards

- NSF's only jurisdiction-based (states and territories) program
- Jurisdiction eligibility is based on percentage of total funds going to that jurisdiction (current cutoff= 0.75%)
- Primary goal is to improve research competitiveness of the jurisdiction to fulfill NSF's mandate
- A jurisdiction can submit only one proposal to the Track-1 program
- 5-year, $20M grants

NSF EPSCoR Track-1 Goals

Catalyze new research capabilities and knowledge

Establish sustainable STEM education, training, and professional development pathways

Broaden diversity in the project team and STEM workforce

Engage partners

Impact research, education, and economic development in academia, government, private sector

Track 1 proposal specifics:

- As per the NSF, PI will be Vermont State Director (me; Arne Bomblies)
- 35 pages total, **22 for research MAX**
- PI will be responsible for coordinating and organizing all elements of the proposal
  - Writers of research component will be core faculty (white paper authors)
  - PI together with EPSCoR staff will write and integrate the other elements of the proposal (data management, budget, hiring, outreach, private sector, etc.)

RII Track-1 Competitiveness:

- Hypothesis-driven, not discovery-based, research is key.
- The infrastructure and increased research capacity are consequences of the cutting-edge research that happens as part of the project
- Review includes non-EPSCoR NSF program staff, Director's Review Board

The goal should not be the RII Track-1 award. The award is the means to achieve a goal.

Sustainability, strategy, execution

Benefits of participation

- Meaningful collaborations
- Mentorship opportunities of postdocs, grad and undergrad students
- Publications
- Summer salary
- Support for travel to conferences
- However, participation is not a way to fund your own lab activities

Homework

1. Form writing groups:
    - Sensor development, CoC's, Data science, Social science/policy
    - Self-identify (to me; abomblie@uvm.edu) as interested in leading
    - We have meeting space and teleconference equipment available for use
    - If not on sign-up sheet, e-mail group preference to me asap (abomblie@uvm.edu).
2. FOCUS on a specific draft research question
    - By November 4
    - Keep it under the umbrella topic that integrates the concept: big data/networks (Vermonitor)
    - Make sure that this draft research question is of interest to a sufficient number of geographically diverse participants (all of VT)
    - Must be fundable by NSF (I will vet the proposal with program officers in directorates)

Timeline

10/18/19- kickoff meeting

11/4/19- writing groups formed; draft research questions done, meeting of writing group leaders. Group leaders meet every two weeks until:

12/20/19- first draft completed

1/15/20- first draft edited

1/31/20- first draft polished and sent to EAC for initial review

3/1/20- reviews back

3/15/20- newest version incorporates EAC comments

4/1/20- NSF visit

5/15/20- second draft done and sent out for another round of review

6/15/20- reviews back

7/1/20- final proposal done

7/20/20- proposal submitted

6/1/21- anticipated project start date

Questions?