Reference ID: 11219085462_Herzfeld

Submission Date and Time: 12/12/2019 11:00:28 PM

Author Names & Affiliations

Submitting author: Ute Herzfeld - University of Colorado Boulder

Additional authors: None

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

Cyberinfrastructure: Computer Science and Arctic Natural Sciences (Glaciology, Sea Ice Research)

Title of Response

RFI Response: Physically driven versus discipline-agnostic classification

Abstract

The RFI prompts had me brainstorming over the topic of data heterogeneity/ scale invariance/ scale dependence and discipline-agnostic approaches. My main thoughts are the following: (1) The topic of data heterogeneity includes, but is not limited to, that of scale invariance versus scale dependence of physical processes. These problems have been known for a coupe decades, and while not solved, have

seen several approaches to their solution. Future research warranted. (2) I think that physically-driven approaches to cyber-infrastructure work better than discipline-agnostic ones, because an understanding (and incorporation) of the physical process to be studied / classified/ will lead to greater advance of knowledge than a comut-science-only/discipline-agnostic/problem-independent approach. Generalizations are feasible either way.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

A main source of data-intensive/data-driven research is satellite remote sensing, including image data but also many other data sources, such as altimetry (ICESat-2,Sentinel-3 and others). I am writing this from the AGU Fall Meeting. Connecting data analysis and numerical modeling of physical processes is becoming a more and more widely recognized challenge, as both data collection and modeling see large advances. The role of Machine Learning in this context is rapidly increasing, but as results this week have shown, often more a flashy term than substantial reality. i.e. ML methods in geophysics are *still * in their infancy. We are currently funded for a CI software element project, to connect modeling and rem sensing observations through parameterization (using ML) with the goals of (a) providing a community resource (software) for classification and (b) pulling out parameters that will (and in test cases have) optimized parameters in numerical models (of ice dynamics). Especially in Arctic Sciences, this is an emerging problem.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

The largest missing capability I see lies in approaches that are based on an understanding of physical processes in glaciology and other environmental / Arctic science / geophysical disciplines. The gap to be bridged exists, because each discipline - computer science and an applied discipline, say, geophysics in our case, uses different "stories" = sets of problem solving strategies. Within geophyscis, and even within glaciology, different work strategies exist in modeling and in data analysis and in observation/instrumentation/engineering for Earth observation. Comparisons are being done, but creating a continuous work flow is missing (and we have proposed to find a solution to a specific problem. "Specific" means here, ice-dynamics/ glacial acceleration, which is a large class of problems in itself). The main point to be made at the abstract level/ in short form here is that I see the largest potential for advances towards convergent science in building physically-driven, or say, problem-focused/ discipline-centered solutions, building a CI for those, and then going on to generalize to other disciplines - as opposed to: overly abstracting the problem first, building a solution entirely in Computer Science, and then hoping it will yield specific results for a number of disciplines and problems. The problems of data heterogeneity and scale exist and require future study, however, solutions have been proposed and i see the community (incl myself) able to solve this, integrating data, working across sensors, using scaling models or CNNs, while recognizing potentially existing limits on scale specific for given physical processes.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

I like the idea of element, then framework, and so on, building and growing a community. Of course the topics of organization, learning, student participation in research at every level, workforce building are all important, but largely understood. One small thought: We may have to work harder on acceptance of people who think differently, or use different approaches [current state is to work towards being more accepting of people by difference in class, such as gender, race, background - all correct but known paths]. The community has a wide-spread weakness of going with the flow, trying to attach themselves the "popular" folks, or going by popularity expressed in social media (including citation counts). There can be strength in new/ daring/ experimental or controversial ideas. Challenging the status quo is not always accepted. NSF already embraces the concept of high-risk research. If we are to move more towards open-source/open science, we may need to be more open to alternative approaches and different lines of thinking. - Again, this is to be considered a side note.

-- End Submission –