

Reference ID: 11170308792\_Herr

---

**Reference ID:** 11170308792\_Herr

**Submission Date and Time:** 11/24/2019 5:23:03 PM

This contribution was submitted to the National Science Foundation in response to a Request for Information, <https://www.nsf.gov/pubs/2020/nsf20015/nsf20015.jsp>. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

*Consent Statement:* "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publicly available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

**Consent answer:** I consent to NSF's use and display of the submitted information.

#### **Author Names & Affiliations**

Submitting author: Daniel Herr - UNC-Greensboro

**Additional authors:** None

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

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

Chemistry, Materials Science, Nanoscience, Nanoelectronics and Bioengineering

#### **Title of Response**

Brief and Initial Individual Response

#### **Abstract**

This response serves as a place holder for a more extensive submission with my colleagues at the UNCG. Time permitting, I plan to provide a more detailed, supported and polished version of these inputs. Please feel free to edit this brief draft response at will.

**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.

Each of the NSF's 10 Big Ideas requires collaborative and convergent efforts to address. The bringing together of multiple disciplinary perspectives and data-sets to address each of these strategic challenges will require a very data-intensive and data-driven S&E CI infrastructure. Also, it is likely that potential solutions for one Big Idea may have unintended consequences for one or more of the other Big Ideas. Consider the potential interaction matrices between any two or more cross-disciplinary Big Ideas and the size of the required shared data between disciplines could grow exponentially. A disciplinary-agnostic data-related CI is needed to uncover unexpected foundational knowledge near the intersections of these knowledge bases. Time permitting, I will provide specific current foundational research challenges and examples of potentially enabling strategic opportunities.

**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?

Data oriented CI has emerged as a key success factor for addressing many foundational 21st century research challenges. This is especially true in the area of material sciences and engineering. While novel discoveries, e.g., one of a kind reports, have received considerable attention in high impact journals, there is a relative dearth of data and information to enable needed foundational predictive materials models. For example, in the 1990s and early 2000s, the unique properties and capabilities of carbon nanotubes (CNTs) received much press. However, today, we have yet to develop the large data bases needed to predict optimal growth of and to synthesize CNTs with specific dimensions and a single chirality. Another more timely area is the large foundational data sets that would enable the design and controlled directed assembly of soft and hard matter into useful macroscopic structures. These areas represent but two frontiers that will catalyze scientific discovery. While gathering such data may not seem 'sexy' to publishers, it is very much needed. In recognition of this emergent, convergent and transdisciplinary need, the UNC Greensboro and the NC A&T State University are collaborating on a strategic cyber-infrastructure initiative, as noted below. The following July 2, 2019 quote may be found

on the UNCG website at: <https://newsandfeatures.uncg.edu/high-speed-network/>. ""UNCG and N.C. A&T have been awarded a two-year National Science Foundation (NSF) grant of \$499,912 to build a high-speed research data network that will both connect the two largest universities in the Triad with each other and enable faster, easier sharing of research with scientists around the world. The Gate City Research Network (GCRN) is one of only 11 NSF Campus Cyber Infrastructure awards in the state of North Carolina and is the first award in the state for the category of Network Infrastructure. Approximately \$358,000 of the award will be managed directly by UNCG, while approximately \$142,000 will be managed by NC A&T as a "sub-award." The GCRN will create a multi-institutional network supporting research activities through a clean, low-latency, high-speed internet connection. This will give researchers access to dedicated, high performance computing resources while helping to eliminate issues posed by using existing networks that also carry administrative, entertainment (i.e. movie streaming, gaming), and other non-scientific data. The GCRN will enable fast transfers of the enormous amount of data that fuels innovative research. This will significantly increase the fundamental research capacity in disciplines such as chemistry, nano-engineering, nano-, computer-, and data science. "This award represents a significant step in our shared journey to help advance the capabilities and reputation of our universities and our region as world-class leaders in research," said UNCG Chancellor Franklin D. Gilliam, Jr. "A state-of-the-art network will help fuel innovation, spark economic growth in our region, and enable us to deliver meaningful scientific breakthroughs. It will also further deepen the existing partnership between UNCG and N.C. A&T, helping to connect our outstanding scientists with collaborators around the world." Added N.C. A&T Chancellor Harold L. Martin Sr., "This is an exciting opportunity for researchers at N.C. A&T, UNCG and their collaborators from around the globe. My expectation is that this new data network will assist our researchers in strengthening existing partnerships and creating new partnerships that will produce innovative solutions to some of the world's persistent challenges." UNCG's Jeff Whitworth, associate vice chancellor and chief infrastructure architect, served as the Principal Investigator (PI) for this grant, with co-principal Investigators Shyam Aravamudhan, associate professor, N.C. A&T; Geoff Starnes, associate vice chancellor of information technology and deputy chief information officer, N.C. A&T; and Daniel Todd, director, Triad Mass Spectrometry Facility.""

**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.

Seed CI training efforts are needed in K-12 and in informal educational environments.

-- End Submission --