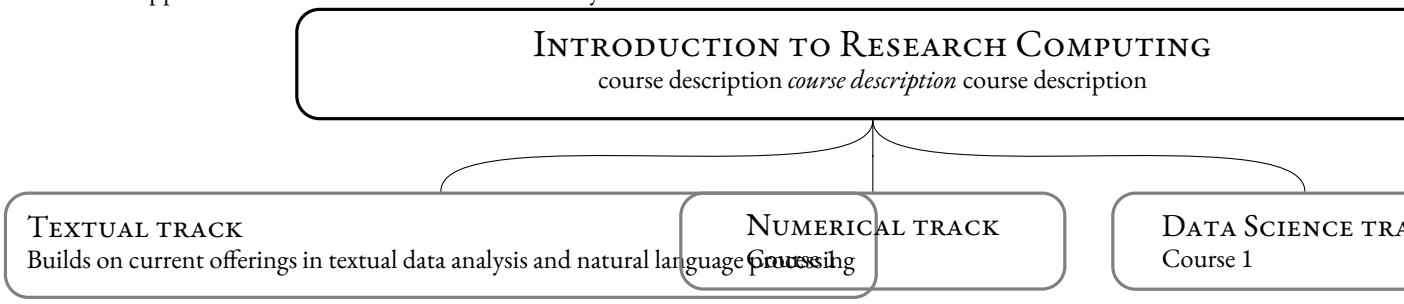
Proposal for a Graduate Certificate in Research Computing

Dennis Y. Tenen, *Associate Professor of English and Comparative Literature, Columbia University*
Roger Creel, *PhD Candidate, Department of Earth and Environmental Sciences, Columbia University*

The growth of interdisciplinary, cutting-edge research that relies on research computing (RC), coupled with the increasingly computationally driven job market for doctoral students, presents an opportunity for Columbia University to produce PHDs who are ready and able to meet these needs, regardless of their career track. In offering a graduate certificate in research computing to doctoral students across the university, Columbia will aid their students as they prepare for distinguished careers. Additionally, the certification process invites participating PhD students to join the interdisciplinary community of scholars at Columbia, enabling new opportunities for distinctive research and collaboration.

Building on the popularity of the Foundations for Research Computing two-day sessions, we propose a certification process where all applicants will take an introductory course on RC with Python before branching off to continue their coursework with approved courses closer to their fields of study.



Summary

Considering the growing prevalence of interdisciplinary, cutting-edge research that relies on research computing (RC) as well as an increasingly computationally driven job market outside of the professoriat, Columbia University must produce PhDs who are ready and able to meet these needs, regardless of their career track. In offering a graduate certificate in research computing, Columbia will continue to meet the needs of world-class PhD students as they distinguish themselves with their research. Additionally, the certification process invites participating PhD students to join the interdisciplinarily minded community of scholars at Columbia, enabling new avenues for distinctive research and collaboration.

With this document, we propose that the Graduate School of Arts and Sciences, with Office of the Executive Vice President for Research, Columbia University Information Technologies, and the Columbia University Libraries, establish a formal graduate certificate in research computing for PhD students.

Our suggestions are preliminary and provide only a launchpad for the work of a more formal committee featuring the members of the bodies mentioned above. We imagine, however, a set of learning objectives similar to what follows:

- Describe and extend the qualities of exceptional research computing
- Analyze, critique, and question inherent assumptions characteristic of research computing
- Interpret, assess, and argue for strategies for managing research data
- Synthesize, defend, and support an ethically driven, critical computation
- Design, develop, and generate a computationally intensive doctoral research project
- Produce a presentation that illustrates and models exceptional research computing

Students will meet the objectives above through four modules:

- A stand-alone, required, for-credit course introducing research computing in theory and practice
- Supplemental, elective courses with a research computing focus drawn from already existing offerings
- Participation in non-credit workshops and training provided by the Libraries and/or CUIT
- Presenting, in public, aspects of their computationally intensive doctoral research project

Need

During the hiring year of 2021–2022, we identified 29 assistant professor openings across the world in earth and environmental science (climate science, geophysics, geoinformatics, etc.). Of these, 26 listings (90%) strongly prefer candidates with a research computing background and 9 (31%) require that expertise. Similarly, at Columbia, research computing fluency is a de facto requirement for graduation in earth and environmental science, but curricular opportunities for acquiring skills in RC are limited and ad hoc.

In a related vein, during the hiring year of 2020–2021, we counted 28 positions (15 tenure-track) across the world that were hiring for someone in the humanities with research computational skills, such as Assistant Professors of Digital Humanities, Assistant Professors in Social Justice Informatics, or non-faculty positions supporting computing and digital scholarship in the humanities, such as Digital Humanities Project Officers or Digital Scholarship Coordinators. Even though fluency in RC is not expected among graduate students in the humanities at Columbia, these findings indicate that the University could support its students better in equipping them to apply for these positions.

Provisional Curriculum

As stated above, the curriculum would be designed and overseen by a faculty-led advisory board. In the current absence of such a board, we propose a provisional curriculum, where graduate students earn the certificate through a mix of pedagogical moments, including courses for credit, non-credit workshops, and public scholarship.

First, at least one for-credit course would be a required course for all students seeking certification. This course, tentatively named "Introduction to Research Computing," would establish the social and environmental contexts for research computing, especially surrounding research design and data management. Upon completing this course, students will be able to consider the effects of various research computing projects, including their own, with a critical eye toward greater accountability. This course would also incorporate the short, web-based course "The Whys and Hows of Exceptional Scholarship with Research Data," prepared by the Libraries and the Columbia Center for Teaching and Learning, and funded by a Provost Interdisciplinary Teaching Award for 2021–2023.

A second required course, "Research Computing in Public," could reposition the contextual concerns towards doing research computing in the public sphere, in the sense of analyzing and presenting research findings with the public as an intended audience or stakeholder.

Additionally, the certification process would include a public-facing project tied to the students' own research. We envision a public event showcasing the student's work with presentations and guidance on hosting their work online.

We also foresee students' taking a number of elective courses. These courses would come from an annually generated list of courses. They would not be offered explicitly for students pursuing the GCRC, but the faculty would know to expect such students in their classes. The advisory board would oversee the collection of appropriate electives.

Finally, students would enrich their knowledge of RC through various workshops offered on campus, including those offered by Foundations for Research Computing and the Libraries. The content of these workshops can also be integrated into the required, for-credit courses. The workshops, however, will not be considered as sufficient instruction in how to do computational research. That is, they are supplemental to programming instruction in the for-credit curriculum and not a replacement for programming instruction.

Audience

We imagine that the initial population of interested students will be GSAS PhD students still completing their coursework, with some exceptions who are already at the proposal stage or later, enrolling with appropriate permission from their adviser(s). There is no prerequisite level of computational knowledge required. Students with no programming experience should be able to earn certification by completing the program.

We seek students who are engaged in or planning a computationally driven research project as well as students eager to learn about research computing for future projects.

We also seek students who are motivated to consider the social and ethical ramifications of research computing, illustrated in their participation in the public-facing final project.

Supplemental Materials

Potential Advisory Board

- Ryan Abernathey

- Matt Jones (History)
- Dennis Tenen (English/Comp Lit)
- Rebecca Wright
- Saima Akhtar (Assoc. Director of Vagelos Computational Science Center)
- Chris Marianetti
- Adrian Hill, Executive Director for Research Planning and Development
- Reshmi Mukherjee, Vice Provost, Academic Research and Centers, Barnard
- CTL (Caitlin DeClercq)
- Teachers College Digital Futures Institute:
  - Lalitha Vasudevan, Managing Director and Vice Dean for Digital Inovation
  - Rochelle Thomas, Directory of Strategy, Planning and Operation
  - Abdul Malik Muftau, Project Operations Lead
- Mark Santolucito (Barnard)

Current state of introductory instruction in research computing at Columbia

- Foundations' Children:
  - Mechanical Engineering reached out to FRC to create bootcamps to run for incoming masters students. Patrick worked with Arvind..., who ran the workshops, etc. FRC provided a logistic support and trained Arvind in Software Carpentries. Patrick also found teachers who would be paid by MechEe. This has been run a few times and is self-sustaining.
  - LEAP: Learning Earth with Artificial Intelligence and Physics. Pierre Gentine, lead PI. Machine learning and climate science Tian Zheng received a grant to.... biggest grant awarded to Morningside Campus. Idea is to run bootcamps for a currently unknown (to us) audience.
- Other stuff:
  - Computing in Context, etc.
  - Women in STEM in SIPA
  - Women in Science at Columbia. (currently changing around. Who knows.)
- For-credit courses offered:
  - Research Computing in Earth Science (Old link)
  - Statistical Computing with SAS (Public Health course)
  - Computing for Business Research (PhD level)
  - Scientific Computing Mawhinney
  - Introduction to Computing for Engineers and Applied Scientists
  - [Probability & Statistics for Data Science] (http://www.math.columbia.edu/~fts/2021%20W5701%20Probability%20and%20Statis Sang Fung)
  - Foundations of Data Science(Yi Zhang, Xuan Zhang)

List of non-intro-level research computing focused courses in 2021–2022

[distinguish by dept or something]

- New Directions in Computing (Undergrad Courses in CS/Barnard)
  - Computing in the Arts
  - Creative Embedded Systems
- Introduction to Numerical Methods (Spiegelman) (prereqs: calc, differential equations, linear algebra, introduction to computing for engineers)
- Tech Society: Good, Bad, & Other (Wright)
- Privacy in a networked world (Wright)
- Algorithms in Data Science (prereqs: programming, calc, diff eq)
- Object Oriented Programming and Design in Java

- Tools for Analytics (Industrial Engineering) (Julian Berman)
- Data Journalism
- Computational Journalism
- [Time Series, Panel Data, & Forecasting] (https://www.coursicle.com/columbia/courses/QMSS/G5016/) (Gregory Eirich)
- [Statistical Inference & Modeling] (https://www.coursicle.com/columbia/courses/STAT/G5703/) (prereq: prob & stats for data science) (Yunxiau, Marco Avella)
- Linear Regression Models
- Time Series Analysis
- Statistical Modeling for Data Analysis
- Applied Data Science(Ying Liu, Tian Zheng, David Shilane)
- Methods in Computational Sciencegithub(Kyle Mandlii)
- Computational Sound(Mark Santalucito)
- Intro-Computational Learning Theory(Rocco Servedio)
- Evolutionary Computation and Design(Hodd Lipson)
- Computational Linear Algebra(Tony Dear)
- Computational Earth Science(Kerry Key)
- MSFE Quantitative and Computational Bootcamp(Michael Miller, Sebastien Donadio)
- Bayesian Statistics(prereq STAT GU4204 or class on theory of statistical inference)(Ronald Neath)

Analysis of the 2022 job market for tenure-track positions in Earth Sciences and Digital Humanities

(include CSVs)

Comparable programs at peer institutions

Etc.

the opportunity for a university-wide certificate in RC
(why)
who - advisers and workers
what
for whom

Notes:
Problem w/ summer institute is figuring out how to do admin.
how do you pay administrators?
CTL, SPS, and GSAPP have summer options
state-level grad certificate program through GSAS
this would be workshops but also courses
Mellon grant for summer institute? 3 years of startup cash?
Certificate in Research Computing:

- Advisory board
- recurring courses that would satisfy the cert
- contact depts and ask if we can send students to their courses
- contact chairs
- certification:
  - requires a curriculum
  - "if we did this this year," how would it look?
  - a picture that convinces CU and NY
  - recurring accreditation admin lifting
  - needs support letters
  - begets a summer institute

- Pull job lists for stuff that would benefit from this cert (academic plus, etc., data skills needed for more and more tt positions)
- how is this not QMSS?
    - QMSS has alternate tracks, a few reqs, then electives in AMSS

What money is needed for a certificate?
Write for Ann, Jonathan, and Carlos Executive Summary