

NEH Application Cover Sheet (HT-272418)

Institutes for Advanced Topics in the Digital Humanities

PROJECT DIRECTOR

Dr. David B. Kinney
Research Fellow
1399 Hyde Park Road
Santa Fe, NM 87501-8943
USA

E-mail: david.kinney@santafe.edu
Phone: 505-920-4651
Fax:

Field of expertise: Philosophy of Science

INSTITUTION

Santa Fe Institute
Santa Fe, NM 87501-8943

APPLICATION INFORMATION

Title: *Foundations and Applications of Cultural Analytics in the Humanities*

Grant period: From 2020-09-01 to 2023-08-31

Project field(s): Interdisciplinary Studies, Other; Philosophy, General

Description of project: The use of computational and quantitative tools and approaches in the humanities is rapidly becoming more widespread. At the same time, there are still significant barriers preventing emerging scholars in the humanities from using these tools to generate new insights that make a genuine impact within the humanities themselves. The goal of our proposed advanced institute is to develop an online course and in-person workshop that will empower scholars in the humanities by eliminating the "black box" of computational text analysis. Participants will gain a theoretical and practical understanding of text analysis methods and the interpretation of their outputs. As such, participants will be able to extract content and derive meaning from the growing archives of digital sources, making accessible new directions in humanities scholarship. The in-person workshop in particular will be the springboard for collaborations between the next generation of analytically-inclined humanities scholars.

BUDGET

Outright Request	247,932.00	Cost Sharing	0.00
Matching Request	0.00	Total Budget	247,932.00
Total NEH	247,932.00		

GRANT ADMINISTRATOR

Dr. David B. Kinney
1399 Hyde Park Road
Santa Fe, NM 87501-8943
USA

E-mail: david.kinney@santafe.edu
Phone: 505-920-4651
Fax:

LIST OF PERSONNEL

Kinney, David | Project Director

Santa Fe Institute

DeDeo, Simon | co-Project Director

Carnegie Mellon University

Santa Fe Institute

Complexity Science Hub Vienna (AT)

Chang, Kent | Consultant: Curriculum & Teaching Assistant

Carnegie Mellon University

Gershkovich, Tatyana | Affiliated Faculty and Mentor

Carnegie Mellon University

Guldi, Jo | Affiliated Faculty and Mentor

Southern Methodist University

Lefkowitz, Julia | Affiliated Faculty and Mentor

University of Oxford (UK)

Marghetis, Tyler | Affiliated Faculty and Mentor

Santa Fe Institute

University of California, Merced (as of Fall 2020)

Cowan, Carrie | Staff: Curriculum Advisor, Project Evaluator, Project Coordinator

Santa Fe Institute

Foundations and Applications of Humanities Analytics

PROJECT SUMMARY

We propose an interdisciplinary Institute for Advanced Topics in the Digital Humanities, based at the Santa Fe Institute (SFI) in Santa Fe, New Mexico, to train humanities scholars to understand and apply concepts and techniques central to humanities analytics.¹ The Institute is aimed primarily at graduate students in the humanities who may have no experience with statistical analysis, computer programming, and/or computer science. The curriculum will enable participants to (1) develop a conceptual understanding of the mathematical principles that underlie contemporary data science, and (2) integrate those insights into individual projects relevant to their scholarship, with the goal of peer-reviewed publication. The Institute will be co-led by a philosopher of science and a researcher in cultural analytics, assisted by an early-career scholar in English literature, and supported by a specialist in interdisciplinary graduate training. It will include contributions by faculty in literature, modern languages, history, sociology, and cognitive science; our faculty will themselves gain professionally from an expanded network of humanities scholars.

The proposed institute comes at a critical time. As the use of digital tools and approaches in the humanities grows, humanities scholars need to go beyond out-of-the-box tools in order to engage critically with conceptual justifications of those tools. Only then is new insight into content and meaning possible. Lectures and mentored practice will enable participants to ask hypothesis-driven questions about texts; participants will be empowered, with a deeper understanding of algorithms and their biases, to accurately interpret their results.

The goal of humanities analytics is to combine qualitative and quantitative techniques to generate new insights into human history, arts, and culture in a way that incorporates and enriches traditional research methods. Our curriculum centers around text analysis: e.g., extracting patterns of word use and stylistic and semantic networks among words, and its applications to cross-modal studies that tie these patterns to specific social, geographical, and historical contexts. Our aim is to train humanities scholars to think computationally about text, enabling them to go beyond archival searches or data-for-data's-sake exercises to develop novel hypotheses.

The proposed Institute combines a virtual introductory course, reaching a wide audience, and an in-person workshop for select participants. The former, to be launched in the first year and available thereafter on demand, will offer video lectures and digital exercises, leveraging SFI's existing online learning platform, production team, and long-term infrastructure for data management. The latter, in years two and three, provides a mentored opportunity to apply the tools and theories introduced in the course to projects of interest. Participants benefit in three ways: (1) as scholars, they will gain skills to apply and interpret the meaning of the algorithms on large-scale datasets; (2), as professionals, they will expand and diversify their network of colleagues and collaborators at a critical point in their careers; and (3) as teachers and mentors themselves, they will acquire curricular materials and experience to employ in their classrooms.

¹ We define humanities analytics as the use of computational approaches from information theory and probability to interrogate cultural artifacts at scale and in partnership with the traditional tools and methodologies of the humanities.

PROJECT NARRATIVE

SIGNIFICANCE

Humanities scholars now have access to digital archives and corpora of unprecedented richness and cultural importance. This achievement of digitization presents new opportunities for scholars to use computational tools to generate insights about culturally significant bodies of text. University courses and textbooks now equip scholars with the analysis tools needed to access and query textual data. The state of the field generally encompasses:

- visualizing text(s), for instance, through a character network
- measuring syntactic and semantic features, from tense and parts of speech to conceptual networks
- organizing and classifying texts within a corpus according to context of production
- modeling literary forms and the emergence of genres and practices.

Digital archives and their computational analysis have the potential to expand the range and scope of humanities scholarship. The automation afforded by computation makes feasible the analysis of large and far-reaching corpora. Together, digital archives and computational analysis enhance the capacity to interrogate, characterize, and understand human history, art, knowledge, ideas, and culture.

Scholarship employing computational text analyses has succeeded in inspiring new insights into important questions in the humanities. Computational approaches have revealed a degradation of gender differences in literary characters over the past two centuries², modeled the global flow of stream-of-consciousness in literature³, measured the sphere of influence of anti-slavery resistance writers in mid-19th century America⁴, and tracked social and political life in the capital of the Confederacy during the Civil War⁵, among other examples. Work employing computational tools is appearing in respected scholarly journals, such as *MLQ* and *PMLA*, attesting to the acceptance of the methods and value of the studies within the broader community.

The tools employed in the computational analysis of text, however, may prove opaque to scholars who are not versed in the underlying concepts. The problem is not unique to scholars in the humanities but has faced researchers in all disciplines that have witnessed the emergence of “big data” approaches. In all such cases, practitioners may have not received an adequate foundation in the mathematical theory behind common “plug-and-play” tools and thus face challenges – or skepticism – when interpreting output or when applying those tools in ways that their originators did not explicitly intend.

Many humanities scholars seek computational skills to analyze large collections of text. Especially for advanced graduate students looking toward their future research, teaching, and careers, fluency with data science methods may be especially attractive. Yet humanities departments may lack resources, expertise in, or awareness of such computational methods, and the corresponding computer or data

² Underwood, WE, Bamman, D, Lee, S. (2018) The transformation of gender in English-language fiction. *J. Cultural Analytics* DOI: 10.22148/16.019 ; but see also Bode, K. (2017). The equivalence of “close” and “distant” reading; or, toward a new object for data-rich literary history. *Modern Language Quarterly*, 78(1), 77-106.

³ Long, H, So, R.J. (2016) Turbulent flow: a computational model of world literature. *MLQ* 77:3 DOI 10.1215/00267929-3570656

⁴ Klein, LF, Eisenstein, J, Sun, I. (2015) Exploratory Thematic Analysis for Digitized Archival Collections. *Digital Scholarship in the Humanities* 30:i130-41 DOI: 10.1093/lc/fqv052

⁵ <https://dsl.richmond.edu/dispatch/pages/intro>

science departments may lack the context or capacity to offer appropriate training. To ensure progress, humanities scholars must become active participants in developing new tools for analysis — tools that precisely address important questions in the field.

The goal of our proposed advanced institute is to empower scholars in the humanities by eliminating the “black box” of computational text analysis. Participants will gain a theoretical and practical understanding of text analysis methods and the interpretation of their outputs. As such, participants will be able to extract content and derive meaning from the growing archives of digital sources, making accessible new directions in humanities scholarship, and empowering humanities scholars to go beyond being mere consumers of data science tools.

The proposed institute complements and extends past NEH Institutes for Advanced Topics in the Digital Humanities, and others⁶, in the broad area of digital analysis. Whereas past programs introduced foundational text analysis tools⁷, our institute will use the outputs of those methods as starting points to explore measurements of probability and information that will contribute to understanding. Whereas those programs focused exclusively or preferentially on scholars at advanced stages of their careers, i.e. faculty or professionals, our institute will primarily serve graduate students, especially those underrepresented or underserved and with minimal prior experience in computational approaches, thereby engendering a strong peer scholarly community at a critical stage of professional development.

Computational text analyses based on a “bag-of-words” approach treat text as a set of individual words that are counted, ignoring order and syntax. While these techniques can define the descriptive properties of significant texts, they are limited in their ability to generate original conclusions.

In the hope of developing approaches that can compute meaning from text, significant attention has been focused on machine learning. Machine learning provides a fast and efficient means of extracting statistical patterns from widely varying data sets, including text. Existing machine learning techniques can identify meaningful patterns that can lead to insight.

Unsupervised clustering uses machine learning to identify categories that best represent a given text data set, topic modeling being the most prominent such approach in the humanities. The power of this approach lies in the ability to reveal unexpected themes that distinguish a work. The resulting topics represent a multi-dimensional measurement that can be compared across texts to assess similarity using powerful techniques from information theory.

Transformer-based models of language, such as GPT and BERT, or recurrent neural network systems, such as LSTM, use machine learning to extract the structure of language in a text. The underlying algorithms were designed to help artificial agents predict the next word(s) in a sentence. As for topic modeling, understanding and comparing the outputs of transformer-based models is not trivial, requiring a conceptualization of the underlying probability theory.

SFI brings a unique perspective to the analysis and interpretation of textual sources. SFI is dedicated to the study of complex systems using the tools of mathematics, physics, philosophy and computer science. Complex systems occur across scales, ranging from interacting molecules in a cell to the

⁶ <https://dhsi.org/>

⁷ <https://securegrants.neh.gov/publicquery/main.aspx?f=1&gn=HT-256977-17;>
<https://securegrants.neh.gov/publicquery/main.aspx?f=1&gn=HT-251006-16;>
<https://securegrants.neh.gov/publicquery/main.aspx?f=1&gn=HT-231816-15>

interaction of ideas in a society. The goal of this approach is to reveal universal rules that govern diverse systems, including societies and civilizations as revealed by their cultural artifacts, such as text.

CURRICULUM & WORK PLAN

The proposed Institute addresses the need to build humanities scholars' capacities in the computational analysis of texts. In particular, the Institute will enable scholars to ask hypothesis-driven questions about content and to compare meaning across sources using the tools of probability and information theory. This will be achieved in two interrelated ways:

First, through completion of a **virtual course** that provides the theoretical and technical knowledge for scholars to deeply understand humanities analytics; and

Second, through participation in an **in-person workshop** that will provide guided practice applying the theory and tools from the course to authentic research questions and hypotheses.

The workshop will bring together participants with diverse interests and expertise, and from a range of institutions, to expand their professional networks and establish a community of learning and scholarship that will extend beyond the program. The Institute's curricular materials will facilitate the introduction of this content into classrooms at participants' home institutions.

See ATTACHMENT 5: Course Outline and Work Plan for details.

Virtual Course [8 h instruction; 20 h problem sets, assessments, readings]

The virtual course, delivered on demand online through a standard web interface, will comprise two modules. **Foundations** will impart conceptual skills and computational thinking through a study of the principles behind contemporary artificial intelligence. **Applications** will employ those skills to develop meaningful accounts of literary, historical, and cultural artifacts. A supplementary online tutorial will introduce the necessary steps to get started with Python programming and Jupyter notebooks.

Foundations I | *Uncertainty and Probability* In this unit, students will learn the basics of probability theory, gaining an appreciation of the statistical principles that underlie the computations and outputs of machine learning tools such as topic modeling. Problem sets will introduce computer-assisted probabilistic calculation and modeling using Jupyter (Python) notebooks. Participants will explore examples of probabilities applied to text analysis.

Foundations II | *Information Theory* Building on the previous unit, students will learn the fundamentals of information theory, providing concepts and tools to interrogate the information content of a text or compare similarities across texts. Problem sets will focus on application of information theory measures to probability distributions (previous module). Participants will consider the role of granularity in representations of information and how granularity affects understanding.

Applications The second module invites students to apply their skills to various applications of content analysis. This section of the course will encourage participants to further develop Python programming skills through hands-on, problem-based exercises. Participants will come away with a sense of how to launch their own cultural analytics project. Applications and data sets will reflect the research interests and methodologies of the faculty.

In addition to applying skills to specific case studies in humanities analytics, this module will encourage participants to consider questions in the meta-study of the humanities, questions to which a

quantitative approach, in conjunction with domain-specific knowledge, may provide novel answers. For example, a fundamental question about human culture and society relates to the spread of ideas and beliefs; namely, which ideas and beliefs become accepted and transmitted and why? The methods of content analysis provide a path toward answering such questions.

In-Person Workshop [5 h lecture; 25 h mentored project work; 8 h discussion]

Following successful completion of the online course, participants may apply to attend a one-week, residential workshop at SFI, facilitating a deeper exploration of information theory and probability as applied to questions in humanities analytics. The Core Faculty (see FACULTY & STAFF) will lead two in-person workshops – one each in years two and three – during which they will guide participants in a self-directed research project employing the tools learned. Invited guest lecturers will provide participants with a breadth of examples of scholarship in the broad domains of information theory and humanities analytics.

Workshop participants will learn by doing as well as from discussion and collaboration with fellow participants and the faculty. This immersive approach is designed to impart a level of experience that will allow participants to employ their acquired skills with confidence and awareness once they return to their home institutions. The intensive small-group format affords opportunities to consult with faculty or peers more generally about research, scholarship, and teaching, including the integration of computational text analysis into humanities instruction.

The in-person workshop expects that participants are comfortable with basic text analysis tools in Python. To this end, machine learning tools will be offered in optional online tutorials available prior to workshop attendance and will be reviewed in the workshop itself.

The timing of the online course will allow participants roughly nine months prior to the application deadline for the workshop. Following launch, the course will be available on demand indefinitely. The workshops will occur in summer, when both the intended audience (advanced graduate students, faculty, academic professionals) and workshop faculty are free of most academic duties.

PARTICIPANTS

The proposed Institute, especially the in-person workshop, is intended primarily for advanced graduate students in the humanities who may have no experience with statistics, computer programming, and/or computer science. However, it is also open to faculty or other humanities professionals such as librarians or archivists. The institute is intended to have a catalyzing effect on the careers of early-career scholars, to establish a professional humanities analytics network, to foster further collaborations and intellectual exchange, and to develop the next generation of mentors and teachers in humanities analytics. Enrollment in the virtual course is limited only by interest.

How many participants will benefit? Online course enrollment will be unrestricted. SFI's online learning platform ComplexityExplorer.org (see INSTITUTIONAL PROFILE) includes more than 55,000 registered users. A typical ComplexityExplorer.org course enrolls 1,000 – 2,500 participants per year. We expect the proposed virtual course would reach roughly 1,000 participants in the first year, with similar or greater participation thereafter, due to word-of-mouth and enhanced publicity.

In-person workshop enrollment will be limited to 10 participants in each of its two instances, reflecting the learning goals and teaching methods, which benefit from individualized mentorship and extensive

peer-to-peer exchange. The format and group size is informed by the sustained success of our Graduate Workshop in Computational Social Science (see INSTITUTIONAL PROFILE).

What is the application process? Online course enrollment is free and only requires submission and verification of a valid email address. Workshop applicants will be required to meet several eligibility criteria: (1) successful completion of the virtual course; (2) enrollment in an accredited graduate program, or employment by an accredited academic or non-profit institution; (3) record of scholarship in the humanities. For candidates who meet these criteria, application to the workshop will involve electronic submission of an academic résumé and a brief (1-page) statement describing their motivation for incorporating computational text analysis into their scholarship. An Institute website and application portal will provide an interface for applicants.

How will we solicit applications for the workshop and recruit a diverse cohort? Mindful of a critical lack of diversity in the digital humanities, we will prioritize recruitment of applicants from under-represented groups, in particular women and people of color, and from institutions outside familiar elite circles. The Core Faculty and SFI's Director for Education have established an extensive network of department chairs, program coordinators, and faculty, with an emphasis on those from programs that promote inclusion of diverse scholars. These networks, established through sustained personal outreach and including faculty in both the arts and sciences, underlie our successes in diversifying the participants in SFI's education programs through thoughtful presentation and commitment. For example, this summer, our undergraduate research program, with a computer science emphasis, will have 50% female and 40% underrepresented racial/ethnic minority participation (compared to 18% and 10% national averages⁸). Similarly, through increased efforts to recruit diverse participants for our advanced computational Summer School, 35% of applications received in 2020 (of 163) were from scholars in the social sciences and humanities.

What are the characteristics of a successful applicant? Successful applicants will demonstrate potential to serve as catalysts through both teaching and scholarship, as evidenced by educational efforts and scholarly outputs such as book reviews, circulating articles and preprints. Our evaluation will consider non-traditional methods for disseminating research such as teach-ins, blogs, and special interest groups. Successful applicants will have shown clear motivation for the learning goals of the institute, will advocate a pro-social, open, and tolerant attitude, and will be eager to become active participants in an egalitarian community of learning. We will aim for a cohort that reflects the multiple dimensions of diversity. The PD and co-PD will review applicants and select participants, assisted by SFI Education.

How will we support students during their time at the workshop? To reduce barriers to participation, we will offer the workshop at no cost. Accommodations and meals will be arranged by SFI Education; travel to/from Santa Fe will be reimbursed. We will strive to accommodate participants with special needs or accompanying dependents (see INSTITUTIONAL PROFILE). All SFI programs prioritize the diversity of participants and aim to provide an inclusive and equitable environment for learning. SFI holds a zero tolerance stance towards bullying, harassment, or violence of any nature. All participants must agree to abide by SFI's anti-discrimination and anti-harassment policies, including

⁸ <https://www.nsf.gov/statistics/2017/nsf17310/digest/fod-women/computer-sciences.cfm>
<https://www.nsf.gov/statistics/2017/nsf17310/digest/fod-minorities/hispanics.cfm>
<https://www.nsf.gov/statistics/2017/nsf17310/digest/fod-minorities/blacks-aian.cfm>

Title IX. Allegations of violations of such policies are under the purview of SFI's Human Resources office, in coordination with the VP for Science and the Director for Education.

How will participants learn about the program? In addition to direct digital outreach to our faculty networks noted above, we will use SFI's social media channels, and outreach and advertisement to relevant professional societies and scholarly meetings.

INSTITUTIONAL PROFILE

SFI was established in 1984, the first institution devoted to seeking and defining generalizable principles of complex systems in the physical, biological, psychological and social sciences at a variety of scales. SFI remains the acknowledged thought-leader in this domain. It is led by President David Krakauer, Ph.D. in coordination with a leadership team additionally comprising VP for Science Jennifer Dunne, Ph.D., VP for Advancement Will Tracey, Ph.D., and VP for Administration Janet Gunn.

The SFI faculty includes 13 resident members, located on campus, and 110 external members who spend short research visits at SFI but have primary appointments at institutions elsewhere. The combination of resident and external faculty includes a breadth of expertise and encourages exchange of ideas. The SFI faculty is 30% female and represents a range of academic institutions.

SFI currently hosts 17 Fellows, early-career researchers who spend from one to three years at SFI, developing their research programs and extending their scholarship. Fellows generally use this protected time in order to explore new ideas, complete projects, and learn. Fellows have opportunities for professional development, including mentoring students and teaching in SFI Education programs.

SFI Education has been a part of the institution since its inception. The longest running Education program, Complex Systems Summer School, now in its 31st year, is a four-week residential program that provides a foundation of complex systems. The Graduate Workshop in Computational Social Science (26th year) is a two-week residential program that emphasizes project-based learning. SFI offers a residential summer research experience for undergraduate students.

In 2013, SFI expanded its educational programs to include the online learning platform ComplexityExplorer.org, which currently hosts 8 full-length courses and 12 shorter-form tutorials. All content is available at no cost, on demand; no special software or equipment is required beyond a standard web browser and internet access. Course content, governed by a Creative Commons license, is available for download and re-use. SFI Education contracts for professional subtitling (English) of all video content and supports crowd-sourced translation into non-English languages. Efforts are made to ensure that material is accessible to diverse audiences and 508 compliant.

SFI occupies two campuses in the city of Santa Fe, New Mexico, providing ideal infrastructure for conferences and workshops. Both campuses are ADA compliant and fully accessible and together offer five conference rooms for groups of 5 to 75 participants in addition to a flexible-use classroom space for ~30 individuals. IT and hospitality support are available at both campuses. Conference and classroom spaces contain digital displays or projectors and screens for sharing content; all buildings are equipped with WiFi. Campuses have indoor and outdoor areas for informal discussion and study. The Core Faculty will have access to private-use office space and conference rooms during their stays. Use of SFI facilities will be at no cost to the proposed program.

Santa Fe is readily accessible from destinations around the US, with two airports, SAF and ABQ, within a 20- and 60-minute drive of Santa Fe, respectively. Shuttles serve both airports; Santa Fe itself is well served by ride-share services and is a major tourism center with accompanying infrastructure. SFI Education has existing partnerships with nearby St John's College and Institute for American Indian Arts (IAIA) for the use of their dormitory and dining facilities to house program participants. Both offer accessible single rooms with shared bathrooms; IAIA additionally offers family apartments. SFI's campuses are served by shuttles from off-site accommodations.

FACULTY & STAFF

The Core Faculty, supported by a Curriculum and Teaching Assistant, will provide direction for the curriculum, including developing the online content and leading in-person workshops. The Core Faculty will appoint multidisciplinary Affiliated Faculty and Mentors to provide the necessary breadth of intellectual and technical expertise to achieve the aims of the institute. The Core Faculty will be supported by the Staff indicated in all aspects of the program.

Core Faculty

David Kinney, Ph.D. | Omidyar Fellow, SFI

Dr. Kinney will serve as Project Director and Core Faculty, guiding the project team. Dr. Kinney will oversee the Institute curriculum, developing online course content and tools and metrics for measuring learning outcomes, lead in-person workshops, select participants, and provide subject-matter expertise. Dr. Kinney studies epistemology, in particular, the selective acquisition of knowledge. His research applies the principles of information theory to understand the causal structure of systems in scientific contexts, the formation of group beliefs, the foundations of scientific reasoning, and the development of causal understanding in humans. He has taught logic, philosophy of science, and political philosophy, among other subjects. As Core Faculty, he will leverage his expertise in probability and information theory, helping students without a mathematics background achieve conceptual mastery.

Simon DeDeo, Ph.D. | Assistant Professor, Carnegie Mellon University | External Faculty, SFI

Prof. DeDeo will serve as co-Project Director and Core Faculty, guiding the direction of the curriculum, developing online course content, leading in-person workshops, identifying faculty, selecting participants, and providing subject-matter expertise. Prof. DeDeo leads the Laboratory for Social Minds at Carnegie Mellon University (CMU), where his research group makes use of text sources—from French Revolutionary records and Enlightenment-era scientific communication to online conspiracy theorists and Harry Potter fan fiction—to define factors that influence how and when novel ideas emerge and become accepted. He has taught courses on cognitive and social science, large-scale social phenomena, and research methods in informatics and computing. He leads the major data science research practicum for students in the humanities, psychology, and economics at CMU. As Core Faculty, he will provide expertise in the application of information-theoretic and machine learning techniques to case studies in literature and history.

Kent Chang, M.Sc. | Research Fellow, Carnegie Mellon University

Mr. Chang will serve as Curriculum and Teaching Assistant on the project, developing online course content, assisting with in-person workshops, and providing subject-matter expertise. At the time of the proposed Institute, Mr. Chang will be a Ph.D. candidate in Information Sciences (offers at UCalfifornia Berkeley and Cornell). Mr. Chang's current research focuses on 20th century English literature and queer theory, having studied with co-PI DeDeo and Dr. Richard Jean So (McGill University). Mr. Chang has developed and led workshops on cultural and literary text mining at

University College London and at CMU. He is an experienced instructor, having taught English in Taiwan for six years, and led seminars in Prof. DeDeo's graduate level seminar in humanities analytics and in CMU's undergraduate culture and data science class. As Curriculum and Teaching Assistant on this project, he will provide expertise in the technical aspects of text mining and data science from literary sources.

Affiliated Faculty and Mentors

Tatyana Gershkovich, Ph.D. | Assistant Professor, Carnegie Mellon University

Dr. Gershkovich will serve as a guest instructor, discussing her work on nineteenth- and early twentieth-century Russian prose, Nabokov and Tolstoy in particular. Her current research focuses on how texts are read and the agency of the author therein. Dr. Gershkovich holds an appointment in the Department of Modern Languages, Russian Studies, and has served as PI on a digital humanities project visualizing Tolstoy's intellectual network.

Jo Guldi, Ph.D. | Associate Professor, Southern Methodist University

Dr. Guldi will serve as a guest instructor, discussing the methods and sources she applies to her study of global conflicts about land and water. Her research uses topic modeling to interrogate the history of Great Britain, especially intersections with concepts of property, rent, and eviction. She has created digital tools for historians, exploiting the derivation of probability-based "topics" at different scales. Dr. Guldi holds an appointment in the Department of History.

Julia Lefkowitz | Ph.D. Candidate, Oxford University

Ms. Lefkowitz will serve as a guest instructor, discussing her research using quantitative text analysis of UK newspaper articles from 1968-present to study trends in newspaper values such as tabloidization and sensationalism. Ms. Lefkowitz is a member of the Oxford Internet Institute.

Tyler Marghetis, Ph.D. | Assistant Professor, UCalifornia Merced (Fall 2020) | Omidyar Fellow, SFI

Dr. Marghetis will serve as a guest instructor, discussing his work on moments of rupture in human thought, such as during artistic breakthroughs or religious conversions. His current work uses tools from machine learning and a "big data" corpus of audio recordings to investigate the process of moment-to-moment creativity in improvised jazz.

Staff

Carrie Cowan, Ph.D. | Director for Education, SFI

Dr. Cowan will support the efforts of the core faculty with respect to curriculum design and online content development, advising on learning objectives, assessments, and appropriate pedagogical approaches. She will also serve as the evaluator for the program, working with the Institute PDs and participants to understand the goals of each and advising on changes to optimize learning outcomes. Dr. Cowan brings over ten years of experience in the science education and training realm. She has served as a university lecturer and program evaluator, and has led multiple National Institute of Health-funded professional development and research training programs for early-career scientists. As Director for Education at SFI, Dr. Cowan oversees a staff of three, including Ms. Isa Ruiz and Ms. Carla Shedivy, who will provide project support for the virtual and in-person portions of the Institute, respectively. Ms. Ruiz manages SFI Education's recording studio and the ComplexityExplorer.org platform. Ms. Shedivy oversees all SFI Education in-person programs and will manage the Institute's application and registration processes as well as all on-site logistics for faculty and participants.

IMPACT & EVALUATION

Online Course Outcomes The learning goals for participants in the virtual course are restricted to course content: amassing a conceptual understanding of the information theory and probability principles behind computational text analysis and applying those tools in the context of humanities analytics. All learning outcomes will be precisely articulated and evaluated through formative, evaluative and summative assessments throughout the course. Assessment outcomes that diverge markedly from an expected ~75% success rate will prompt revision or expansion of the relevant content. A major outcome of the online course will be the broad dissemination of the content at no-cost and on-demand through ComplexityExplorer.org.

Workshop Outcomes The learning goals for in-person workshop participants include application of computational theory and tools to an authentic research project. The learning goals will extend to content-independent areas such as synthesis and interpretation of knowledge, providing and incorporating critical feedback to improve scholarship, working effectively in diverse groups, and building and leveraging a professional network. The varied learning goals will be assessed primarily through three means: the success or promise of the workshop project itself, faculty evaluations of participant achievement, and participant self-assessment with respect to learning gains.

Project Outcomes The expected direct outcomes of the project are:

- Participants will gain a theoretical and practical understanding of the computational tools of humanities analytics, their applications, and the interpretation of their outputs.
- Participants will develop skill and confidence in applying these tools to their own research.
- Participants will build a peer community within which to explore applications of computational analysis to questions in the humanities.
- Participants will diversify the community of humanities analytics scholars.

In addition to the outcomes related to the learning goals outlined above, these program outcomes will be assessed by longitudinal tracking of participants through self-reported assessments, scholarly output, implementation in the classroom, participant demographics, and formal collaboration or informal collaboration. Follow-up surveys will be issued at one, two, and three years after completion of the full program cycle (virtual course plus in-person workshop).

The anticipated indirect outcomes of the project are:

- New directions in humanities analytics scholarship through interdisciplinarity.
- Increased computational capacity among humanities scholars, with broad implications for the future of the discipline

These indirect outcomes are anticipated as long-term and involving numerous external factors and thus fall outside the evaluation scope of the proposed Institute.

Outcomes Reporting We will release two white papers describing what we learn from the program, the first after one cohort completes the online course plus in-person workshop (end of year 2) and the second after two cohorts complete the online course, the in-person workshop, and have been surveyed at subsequent one-, two- and three-year intervals. The white papers will be disseminated through our professional networks and made available online through SFI in addition to that of the sponsoring organization.



Applicant Institution: *Santa Fe Institute*

Project Director: *David Kinney*

Project Grant Period: *9/01/2020 through 8/31/2023*

[click for Budget Instructions](#)

	Computational Details/Notes	(notes)	Year 1	(notes)	Year 2	(notes)	Year 3	Project Total
			09/01/2020- 8/31/2021		09/01/2021- 8/31/2022		09/01/2022- 8/31/2023	
1. Salaries & Wages								
Project Director (David Kinney)	Calendar Year Salary: (b) (6) (Year 1); (b) (6) (Year 2); (b) (6) (Year 3)	8.33%	(b) (6)	8.33%	(b) (6)	8%	(b) (6)	(b) (6)
		%		%		%		\$0
		%		%		%		\$0
		%		%		%		\$0
		%		%		%		\$0
		%		%		%		\$0
2. Fringe Benefits								
Project Director		27%	(b) (6)		(b) (6)		(b) (6)	(b) (6)
3. Consultant Fees								
Research Assistant	\$100/day	90 days	\$9,000	30 days	\$3,000	30 days	\$3,000	\$15,000
Vide Production	\$50/day	20 days	\$1,000	4 days	\$200			\$1,200
Animator	\$50/day	10 days	\$500					\$500
Faculty honorariums	\$250/day	4 days	\$1,000	4 days	\$1,000	4 days	\$1,000	\$3,000
4. Travel								
Project Director	To Washington DC to attend PD meeting in Year one							
	R/T airfare \$700		\$700					
	Lodging \$570		\$570					

	Per diem 3*75/225		\$225				
	Transfers \$100		\$100				\$1,595
Research Assistant/Consultant	Project travel to SFI; two trips Year 1 one trip years 2 and 3						
	R/T airfare		\$1,400		\$714		\$728
	Transfers		\$200		\$102		\$104
	Lodging		\$1,400		\$714		\$728
	Per diem		\$840		\$428		\$437
							\$7,796
Workshop Faculty	Travel to Santa Fe for Institute; two trips Year 1 and 3 trips each Years 2 and 3						
	R/T airfare		\$1,400		\$2,142		\$2,185
	Transfers		\$200		\$306		\$312
	Lodging		\$720		\$1,140		\$1,163
	Per Diem		\$360		\$551		\$562
							\$11,040
5. Supplies & Materials							
General project supplies	printing costs, placards, registration materials ,audio-visual, online course web hosting, software.		\$1,300		\$2,350		\$2,350
							\$6,000
6. Subawards							
Subaward to Carnegie Mellon University for Participation of Co-Director DeDeao			\$30,583		\$30,843		\$31,701
							\$93,127
7. Other Costs							
Participant Support Costs	for ten persons per year in Years 2 and 3 for participant travel/ Institute						
	R/T airfare		\$0		\$7,140		\$7,283
	Transfers		\$0		\$1,020		\$1,040
	Dorm style lodging		\$0		\$7,140		\$7,283

	Per Diem		\$0		\$3,060		\$3,121	\$37,087																											
8. Total Direct Costs	Per Year		\$57,975		\$68,457		\$69,736	\$196,168																											
9. Total Indirect Costs																																			
a. Rate: SFI Negotiated rate for all sponsored activities of 57.76 percent of Modified Total Direct Costs, excluding equipment and participant support.																																			
b. Federal Agency: NSF	Per Year	57.76%	\$29,295		\$11,150		\$11,319	\$51,764																											
Effective Period: 05/08/2019-12/31/2021																																			
10. Total Project Costs	(Direct and Indirect costs for entire project)							\$247,932																											
11. Project Funding	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%;">a. Requested from NEH</td> <td style="width: 30%; text-align: right;">Outright:</td> <td style="width: 40%; text-align: right;">\$247,932</td> </tr> <tr> <td></td> <td style="text-align: right;">Federal Matching Funds:</td> <td style="text-align: right;">\$0</td> </tr> <tr> <td></td> <td style="text-align: right;">TOTAL REQUESTED FROM NEH:</td> <td style="text-align: right;">\$247,932</td> </tr> <tr> <td>b. Cost Sharing</td> <td style="text-align: right;">Applicant's Contributions:</td> <td style="text-align: right;">\$0</td> </tr> <tr> <td></td> <td style="text-align: right;">Third-Party Cash Contributions:</td> <td></td> </tr> <tr> <td></td> <td style="text-align: right;">Third-Party In-Kind Contributions:</td> <td style="text-align: right;">\$0</td> </tr> <tr> <td></td> <td style="text-align: right;">Project Income:</td> <td style="text-align: right;">\$0</td> </tr> <tr> <td></td> <td style="text-align: right;">Other Federal Agencies:</td> <td style="text-align: right;">\$0</td> </tr> <tr> <td></td> <td style="text-align: right;">TOTAL COST SHARING:</td> <td style="text-align: right;">\$0</td> </tr> </table>								a. Requested from NEH	Outright:	\$247,932		Federal Matching Funds:	\$0		TOTAL REQUESTED FROM NEH:	\$247,932	b. Cost Sharing	Applicant's Contributions:	\$0		Third-Party Cash Contributions:			Third-Party In-Kind Contributions:	\$0		Project Income:	\$0		Other Federal Agencies:	\$0		TOTAL COST SHARING:	\$0
a. Requested from NEH	Outright:	\$247,932																																	
	Federal Matching Funds:	\$0																																	
	TOTAL REQUESTED FROM NEH:	\$247,932																																	
b. Cost Sharing	Applicant's Contributions:	\$0																																	
	Third-Party Cash Contributions:																																		
	Third-Party In-Kind Contributions:	\$0																																	
	Project Income:	\$0																																	
	Other Federal Agencies:	\$0																																	
	TOTAL COST SHARING:	\$0																																	
12. Total Project Funding								\$247,932																											

Total Project Costs must be equal to Total Project Funding ----> (\$247,932 = \$247,932 ?)

Third-Party Contributions must be greater than or equal to Requested Federal Matching Funds ----> (\$0 ≥ \$0 ?)

6. Subawards									
									\$0
7. Other Costs									\$0
8. Total Direct Costs	Per Year		\$19,504		\$19,670		\$20,218		\$59,392
9. Total Indirect Costs									
a. Rate: 56.8 percent of Modified Total Direct Costs excluding excluding equipment, capital expenditures, charges for tuition remission, rental costs, scholarships and fellowships, internally charged telephone, internally charged copying, and individual subcontract costs in excess of \$25,000.									
b. Federal Agency: ONR	Per Year		\$11,078		\$11,173		\$11,484		\$33,735
Effective Period: 07/01/2019-06/30/2020									
10. Total Project Costs	(Direct and Indirect costs for entire project)								\$93,127
11. Project Funding	a. Requested from NEH								
	Outright:								\$93,127
	Federal Matching Funds:								\$0
	TOTAL REQUESTED FROM NEH:								\$93,127
	b. Cost Sharing								
	Applicant's Contributions:								\$0
	Third-Party Cash Contributions:								
	Third-Party In-Kind Contributions:								\$0
	Project Income:								\$0
	Other Federal Agencies:								\$0
	TOTAL COST SHARING:								\$0
12. Total Project Funding									\$93,127

Total Project Costs must be equal to Total Project Funding ----> (\$93,127 = \$93,127 ?)

Third-Party Contributions must be greater than or equal to Requested Federal Matching Funds ----> (\$0 ≥ \$0 ?)

BUDGET JUSTIFICATION

These requested costs are for support of an interdisciplinary institute, based at the Santa Fe Institute (SFI) in Santa Fe, New Mexico, to train advanced graduate students and others in the humanities in concepts and techniques central to innovative scholarship in humanities analytics. The curriculum will enable participants to develop a conceptual understanding of the mathematical principles that underlie contemporary data science, and integrate those insights into individual projects relevant to their scholarship, with the goal of peer-reviewed publication.

Direct costs in Years Two and Three include a 2% annual cost of living adjustment unless otherwise noted. The Santa Fe Institute (SFI) defines a year as one calendar year.

Salaries and Wages: (b) (6)

Senior Personnel: (b) (6)

David Kinney, PD, 1.0 calendar months

Dr. Kinney is a research fellow at the Santa Fe Institute and one month of salary effort is requested in each year of the project. As PD of the project, Dr. Kinney will lead the PD team and perform a variety of tasks and manage all aspects of the project as described in the proposal, including the Institute curriculum development and virtual institute implementation in year one, in-person Institute implementation in years two and three, and evaluation in all three years.

Fringe Benefits: (b) (6)

Fringe benefits are based on an approved rate of 27% of requested salaries and include health insurance, OASDI, and FICA. Although actual fringe rates vary for each SFI employee, proposal estimates are based on the approved aggregate rate.

Consultant Fees: \$19,700.00

1. *Graduate Research Assistant:* Funds totaling \$9,000 in year one, and \$3,000 per year in years two and three, for a total of \$15,000 are included for Research Assistant Chang (Letter of Commitment included and cv included in Attachments), whose duties will include Institute development and implementation and teaching assistance for a total of 3 months (90 days) in year one, and 1 month (30 days) in years two and three, at \$100 per day with a two percent increase annually.
2. *Video Production:* Funds totaling \$1,000 in year one and \$200 in year two, for a total of \$1,200 at a rate of \$50 per day are included for a to be named video production consultant who will provide services to develop digital content for the virtual course instance, as described in the proposal.

3. *Animator:* Funds totaling \$500 are included in year one at a rate of \$50 per day for a to-be-named animator to contribute to the development of illustrative content to complement faculty lectures in the virtual course instance.
4. *Faculty Honorariums:* Funds of \$1,000 in each year, for a total of \$3000, are included for honorariums for Institute faculty who participate in teaching and curriculum development. Honoraria will be offered at a rate of \$250 per day, with both online course development and in-person workshop instruction requiring one day of service.

Travel: \$20,431.00

All travel activity estimates are based on historical costs and SFI policy and include domestic round-trip airfare (\$700), ground transportation (\$100), and hotel lodging (\$120/night for hotel lodging in Santa Fe off-season; \$190/night for hotel lodging in Santa Fe in-season and in Washington DC; \$100/night for dorm style lodging in Santa Fe at St. John's College and/or the Institute of American Indian Arts), per SFI policy. Travel estimates also include per diem for meals and incidentals (\$60/day in Santa Fe and \$75 for travel to Washington DC, per SFI policy), with increases of two percent per year in years two and three as noted above.

1. *Travel to Program Directors Meeting – Domestic:* A total of \$1,595.00 is included for PD Kinney to attend the Program Director's Meeting in Year One, anticipated to be in the Washington DC area, including domestic airfare, ground transportation, three nights hotel lodging and three days per diem at the estimate amounts indicated above, as required per the Solicitation. Funds for co-PD DeDeo to attend this meeting are included in the sub-award to Carnegie Mellon University.
2. *Project-related Travel for Consultant Chang – Domestic:* A total of \$7,796.00 is included for two trips in year one and one trip in each of years two and three for Consultant Chang to travel to Santa Fe, New Mexico to meet with the PD team for Institute development, as well as to attend and assist with the Institute itself. Costs are requested for round trip domestic economy airfare at \$700 per trip, ground transportation at \$100, seven nights dorm style lodging at the rate of \$100 per night per trip, and seven days per diem at \$60 per day with two percent increases in years two and three, as discussed above.
3. *Travel for Institute Faculty:* A total of \$11,040 is requested for Institute faculty to participate in workshop development and presentation as described in the proposal; funds are requested for two trips in year one and three trips per year in years two and three, including domestic airfare, transfers, hotel lodging for three nights per trip in year one and two nights per trip in years two and three, and per diem for three days per trip at the rates outlined above. SFI Education funds will be used to support travel for faculty traveling internationally. Two percent increases are included in years two and three as discussed above.

Supplies and Materials: \$6,000.00

Funds are requested in the total amount of \$6,000 for general project supplies as well as Institute related supplies including printing costs, placards, signage, registration materials and badges, as well as for audio-visual costs and online course web hosting costs. Funds are also requested for software to support Institute evaluation, applicant tracking and participant surveys, as well as website hosting costs.

Subawards: \$93,127.00

A Subaward to Carnegie Mellon University (CMU) with co-PD DeDeo in the amount of \$93,127.00 has also been included. The details for CMU's participation are included in the included budget and budget justifications for the subaward, attached to this proposal.

Other Costs: \$37,087.00

Participant travel and meals: \$37,087.00

Funds are requested for participant support for ten invited participants to attend the in person Workshop in each of years two and three as described in the proposal. Per-person travel estimates are based on historical costs and SFI policy: domestic round-trip airfare (\$700 per person), ground transportation (\$100 round trip), five nights lodging (\$100/night for dorm style accommodations at St. John's College or the Institute of American Indian Arts (IAIA)) and five nights per diem at \$60 per day; with two percent increases in years two and three.

TOTAL DIRECT COSTS: \$196,168.00**Indirect Costs (F&A): \$51,764.00**

Indirect costs are based on the Institute's current rate for all sponsored projects of 57.76% based on an extension until December 2021 negotiated May 8, 2019 between SFI and the U.S. Department of the Interior on behalf of the National Science Foundation, SFI's cognizant federal audit agency. This rate is applicable to Modified Total Direct Costs of \$89,620.00, and excludes equipment, participant support and sub-award amounts greater than \$25,000.

TOTAL BUDGET REQUESTED: \$247,932.00

BUDGET JUSTIFICATION:

Personnel:

Principal Investigator: Simon DeDeo, Assistant Professor , Department of Social and Decision Sciences, Dietrich College of Humanities & Social Sciences, Carnegie Mellon University (CMU), (1 AY month, all project years) will help manage all aspects of the project as described in the proposal, including the Institute curriculum development and virtual course implementation in year one, in-person Institute implementation in years two and three, and evaluation in all three years.

Salaries: All salaries are based on AY 2020 salaries. Faculty and staff salaries reflect an anticipated annual increase effective July 1. Faculty and Staff Increase Rate: 3.00%

Fringe Benefits: The University's fiscal year is July 1 - June 30. Full-time fringe benefits rate = 24.5%.

Operating Expenses:

Expenditures budgeted under Operating Expenses for this proposal are costs that can be identified specifically with this particular sponsored project and are required in the direct performance of the research. The following expenses can be estimated with a high degree of accuracy for this project and are thus listed individually.

All travel activity estimates are based on historical costs and GSA rates and include mileage for the Pittsburgh to DC drive, and domestic flights for the Santa Fe visits.

1. ***Travel to Program Directors Meeting – Domestic: A total of \$628*** is included for co-PD DeDeo to attend the Program Director's Meeting in Year One, anticipated to be in the Washington DC area, as required per the Solicitation. Costs are based on mileage driven (54 cents/mile) and GSA rates for lodging in DC.
2. ***Project-related Travel for co-PD DeDeo – Domestic: A total of \$3989*** is included for two trips in year one and one trip in each of years two and three for co-PD DeDeo to travel to Santa Fe, New Mexico to meet with the Institute team for curriculum development, as well as to lead the in-person Institute. Costs are requested for round trip domestic economy airfare at \$500 per trip, and five nights each at the GSA lodging rate of \$131 per night.

Overhead

Overhead on this proposal has been calculated at our Office of Naval Research negotiated rate for all fiscal years in accordance with 2 CFR 200. The modified total direct cost base (MTDC) amount used in calculating the indirect costs is the total direct costs, excluding equipment, capital expenditures, charges for tuition remission, rental costs, scholarships and fellowships, internally charged telephone, internally charged copying, and individual subcontract costs in excess of \$25,000. Overhead Rate: 56.8%

COURSE OUTLINE

Virtual Course

[8 h instruction; 20 h problem sets, assessments, readings]

The virtual course will leverage SFI's success in online learning through its course platform ComplexityExplorer.org. The course will include the following learning resources:

Didactic lectures Video-recorded lectures, supplemented with visual aids and illustrations, will cover concepts and their applications. Lectures will be equipped with subtitles and accessible on demand.

Readings Recommended readings will provide participants with an introduction to terminology and concepts before viewing lectures.

Interactive programming problem sets Jupyter (Python) notebooks will allow participants to step through coding and running the computational analyses covered in the course at their own pace.

Discussion forum Participants may engage with each other in an online forum, facilitating a community of scholarship among participants. The Institute staff and Teaching Assistant will monitor the forum to assist with conceptual or technical challenges.

Quizzes Unit and module quizzes will help participants self-assess their understanding of the material and identify topics in need of review or further study.

Exam A comprehensive summative assessment will be available at the conclusion of the course, ie. after all lectures have been viewed and quizzes completed. The exam will require participants to synthesize the course content and implement concepts.

Module 1. Foundations (D. Kinney)

I. *Uncertainty and Probability*

I.1 Introduction to Uncertainty

Lecture topics: Overview of uncertainty; history of probability theory.

Readings: Venn, J. (1876) *The Logic of Chance, 3rd Edition*. New York: Chelsea Publishing Co. pp. 1-19.¹

Problem set: Games of chance, insurance scenarios, decision problems

I.2 Probability Spaces, Events, and Random Variables

Lecture topics: Sets and elements; the operations of complement, union, and intersection; subset relation; functions and inverse functions between sets; probability space; random variables.

Readings: Kinney, D. (2020) "Notes on Set Theory, Probability Spaces, and Random Variables."

Problem set: Assigning probabilities to the different possible parts of speech in a sentence.

I.3 Conditional Probability & Bayes' Rule

Lecture topics: Conditional probability; Bayes' Rule; conditional independence.

¹ <http://www.gutenberg.org/files/57359/57359-h/57359-h.htm>

- Readings:* DeDeo, S. (2018) “Bayesian Reasoning for Intelligent People.”²
Kinney, D. (2020) “Notes on Conditional Probability and Bayes’ Rule.”
- Problem set:* Conditional probability distribution over the number of syllables in a word in a poem given the number of syllables in the previous word.

I.4 Markov Chains

- Lecture topics:* System change over time; vectors, matrices, vector-matrix multiplications; ergodicity.
- Readings:* Grinstead, CM, Snell, JL. (2012) Introduction to probability. *Am. Math. Soc.* pp. 405-413.³
Hayes, B. (2013) First links in the Markov chain. *Am. Scientist.* 101:92-7.⁴
- Problem set:* Linear algebra calculations, modeling a sequence of symbols as a Markov chain to infer whether they encode meaning.

II. Information Theory

- Lecture topics:* Entropy/complexity as information; conditional information; mutual information; Kullback-Leibler (K-L) divergence; information bottleneck; coarse-graining.
- Readings:* DeDeo, S. (2018) “Information Theory for Intelligent People.”⁵
- Problem set:* Calculate K-L divergence as a measure of differences in information content between the topic model probability distributions of two texts

Module 2. Applications (S. DeDeo, with guest lectures from Affiliated Faculty)

This portion of the virtual course uses video lectures and interactive programming notebooks and problem sets to guide participants through various computational tools applied to questions in the humanities. Participants will evaluate the dataset(s) available and appropriate methods given those limitations, apply computational text analysis methods to selected datasets, and interpret the outputs of those methods. The particular methods will focus on measurements of similarity among works, assessed through probability and information theory, as well as information flow through a collection, using probabilities and Markov chains. While machine learning techniques such as topic modeling or BERT may be used to generate the quantitative descriptors of works to be compared, the practical how-to for advanced approaches will not be needed to complete the virtual course. As in the **Foundations** module, the **Applications** module emphasizes theoretical mastery.

Application areas to be discussed in the online course include but are not limited to:

- Applying Hidden Markov Models as a new method for close-reading of and comparing individual texts to extract subaltern and counter-normative desires and modes of relating.
- Applying information theory to reconstructed transcripts of the French Revolution parliamentary debates to study how patterns of speaking are created, adopted, ignored or propagated, and the influences of political ideology, norms, and individual speakers thereon.

² <http://tuvalu.santafe.edu/~simon/br.pdf>

³ <https://math.dartmouth.edu/~prob/prob/prob.pdf>

⁴ <https://www.americanscientist.org/article/first-links-in-the-markov-chain>

⁵ <http://tuvalu.santafe.edu/~simon/it.pdf>

- Applying probability theory to compiled corpora of 18th Century texts to measure innovation and irrelevance in word usage and thus detect the emergence of walking as a leisure activity in Britain.
- Applying K-L divergence to topic models of fanfiction from the Archive of Our Own (AO3) corpus to determine the influence of novelty and similarity on the popularity of new works.
- Applying a computationally-aided critical discourse analysis and probability theory to compare the portrayal of key public figures and events in various media outlets to measure sensationalism.
- Applying probability theory to Tolstoy's extensive letters and diaries during his various literary periods to determine the intersection between the author and his creative output.

Readings: Barron, ATJ, Huang, J, Spang, RL, DeDeo, S. (2018) Individuals, institutions, and innovation in the debates of the French Revolution. *Proc. Nat. Acad. Sci. USA* 115:4607-12. DOI: 10.1073/pnas.1717729115

Chang, KK, DeDeo, S. (2019) Divergence and the Complexity of Difference in Text and Culture. *J. Cultural Analytics*, in review.

Guldi, J. (2012) The History of Walking and the Digital Turn: Stride and Lounge in London, 1808–1851. *J. Modern History* 84:116–44. DOI: 10.1086/663350

Murdock, J., Allen, C., & DeDeo, S. (2017). Exploration and exploitation of Victorian science in Darwin's reading notebooks. *Cognition*, 159, 117-26. DOI: 10.1016/j.cognition.2016.11.012

At the conclusion of the online course, participants will be able to: 1) competently apply basic probabilistic and information-theoretic techniques to formulate hypotheses and answer questions about the content of texts; 2) critically examine scholarly papers in humanities analytics that apply information theory and probability techniques; and 3) conceptually extend the approaches of information theory and probability to a humanities analytics research project of their own definition. This serves as the starting point for the in-person workshop.

In-Person Workshop

[5 h lecture; 25 h mentored project work; 8 h discussion]

The in-person workshop will leverage SFI's success in project-focused advanced training in our long-standing summer schools and workshops. The goal of the workshop is to help participants progress toward a scholarly publication based on a project incorporating the approaches of humanities analytics. The workshop will include the following learning opportunities:

Mentored project work Participants will be asked to bring their original hypotheses or questions that may form the basis of a meritorious research project to be conducted during the workshop. Ideas will be presented, discussed and refined on Day 1, according to input from faculty and peers. Thereafter, each participant will work on their independent project under the individual guidance of the faculty, who will be available to advise on all aspects of the research, including formulation of the hypothesis, available and appropriate datasets, technical/computational how-to, data interpretation, and further readings, both topical and technical. Peer-peer mentoring and teaching will occur during formal and informal discussions.

Didactic lectures Workshop faculty will discuss how they use various tools from humanities analytics in their research, providing concrete examples of questions and their approach. Presentations will focus on applications so that participants can assess the utility of tools to their own research.

Programming tips & tricks Based on participant needs and interests, each day will include a ~30 min session on best practices in authoring code and/or solutions to persistent programming challenges.

Group discussions Workshop days will begin and end with a brief group discussion (~2 min per participant), wherein each participant will present their immediate goals for the upcoming session, as well as any challenges that are preventing progress. Each afternoon will include a more in-depth group discussion about participants' broader goals for their projects. Discussions create opportunities for peer-peer mentorship.

The workshop will proceed as follows, with each day's program running 9AM - 5PM:

Day 1	Days 2 – 4	Day 5
Participant introductions	Group discussion	Group discussion
Initial project proposals	Mentored project work	Mentored project work
Project proposal critique	Programming tips & tricks	Final project presentations
Revised project proposals	Lecture	Group discussion
Lecture	Group discussion	Workshop wrap-up
Group discussion	Mentored project work	
	Group discussion	

Participants will be encouraged to identify and gain access to the appropriate datasets in service of the hypotheses they are interested in pursuing. Examples of available datasets that are amenable to extracting the emergence and transmission of stylistic, conceptual, or topical patterns using the probability and information theory approaches introduced in the virtual course include:

- **Novel450** collection of 450 novels in English, French and German, 1770-1930 | https://figshare.com/articles/txtlab_Novel450/2062002/3
- **History of Black Writing Novel Corpus**, a corpus of 53 publicly accessible novels (450+ in the complete collection) by African American writers | https://textual-optics-lab.uchicago.edu/black_writing_corpus
- **Archive of Our Own (AO3)**, an open fanfiction, fanart, fanvideo database containing over 5.6 million works | <https://www.archiveofourown.org>
- **Tribal Writers Digital Library**, including works by American Indians, Alaska Natives, and First Nations people of Canada, with an emphasis on lesser-known works and writers. | <https://ualrexhibits.org/tribalwriters/>
- **Prozhito Diaries** project, a two-century corpus of ego-document from Russia including both elites and “ordinary” citizens, revealing methods of self-creation and conceptualization of the genre of diary writing | <https://prozhito.org/>

- **Freedom On the Move**, a crowdsourced collection of runaway slave advertisements from North America | <https://freedomonthemove.org/>
- **Hansard parliamentary records** from the United Kingdom, 1803-2005; and letters, reports, speech transcripts, and first-hand accounts from the **French Revolutionary parliament**, 1787-1794 | <https://www.english-corpora.org/hansard/> & <https://github.com/frenchrevdata>
- **Avalon Project**, containing legal, historical, and diplomatic documents, 4th Century BCE – 20th Century, including treaties between the United States government and Native Americans, 1778-1868 | https://avalon.law.yale.edu/subject_menus/ntreaty.asp
- **Aggregated, broad-ranging contemporary corpora** that can reveal the emergence and perpetuation of algorithmic bias
- **Records of Indentures and Apprenticeships** in the Port of Philadelphia, 1771-1773, containing 4-5,000 contracts for indentured servitude and apprenticeship | <https://github.com/AmericanPhilosophicalSociety/Historic-Indenture-Data>
- **Racial Lines**, containing character dialogue and racial/ethnic identity from 780 Hollywood movies produced between 1970 and 2014 | DOI:10.7910/DVN/KERZQY

Others datasets will be introduced during the workshop, pending participants' interests.

At the conclusion of the in-person workshop, participants will have initiated original scholarship in humanities analytics in an area of personal interest. Participants will have the skills and resources, including continued access to the Institute faculty, to advance these projects at their home institutions, ultimately leading to publication and dissemination of the work. Participants will leave the workshop with an enhanced professional network of colleagues with whom they will continue to collaborate and discuss. These peer networks will serve as catalysts to advance digital humanities scholarship by establishing and supporting a platform for continued research and teaching in humanities analytics.

Optional Tutorials

[each approximately 1 h lecture; 3-9 h problem sets]

The following optional online tutorials will be available through SFI's ComplexityExplorer.org platform. The content falls outside the scope of the proposed Institute *per se* but may be helpful or necessary for some participants, depending on prior experience.

Intro to Python programming for text analysis

Topics covered: Installing and running Python (Anaconda); working with Jupyter notebooks; importing packages/modules; navigating directories and files; variables; opening and reading data files; reading webpages; data structures; defining functions; tokenizing text; loops and conditionals; basic outputs, including plotting and writing files.

Basic Concepts in Probability

Topics covered: Review of basic descriptive statistics, including categorical versus continuous data; data dimensionality; distributions; mean, median and variance; t-tests, confidence intervals and P-values.

Topic Modeling

Topics covered: Step-by-step walk-through for topic modeling using Gensim and MALLET in Python; parameters and optimization; assessing model fit (coherence score); outputs.

PROJECT WORK PLAN

	Fall	Winter	Spring	Summer
Year 1 (2020 - 2021)	Virtual course development [content: scripts, slides, coding notebooks] * DK SD KC AF CC	Virtual course development [recording, editing, post-production] DK SD KC AF CC	Online course development [small group pilot; revisions] DK SD KC CC	Launch online course CC
Year 2 (2021 - 2022)	Online course available on demand	Evaluate & revise online course; online course available on demand DK SD CC ----- Solicit applications for in-person workshop CC	Online course available on demand ----- Application deadline & selection for in-person workshop DK SD CC	Evaluate & revise online course; online course available on demand DK SD CC ----- In-person workshop (1 week) DK SD KC AF CC
Year 3 (2022 - 2023)	Online course available (indefinitely) ----- Evaluate & revise workshop DK SD CC	Solicit applications for in-person workshop CC	Application deadline & selection for in-person workshop DK SD CC	In-person workshop (1 week) DK SD KC AF CC ----- Future directions DK SD CC White paper DK SD KC CC

* Core Faculty, Affiliated Faculty, and Staff involved (see FACULTY & STAFF for details). Core Faculty = DK: David Kinney; SD: Simon DeDeo; Curriculum and Teaching Assistant = KC: Kent Chang; Affiliated Faculty (all) = AF. Staff = CC: Carrie Cowan

David B. Kinney

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501
United States

Phone: +1 (505) 960 4651
Email: david.kinney@santafe.edu
(b) (6)
<http://davidbkinney.com/>

Academic Employment

Omidyar Postdoctoral Fellow, Santa Fe Institute, September 2019-Present

Education

Ph.D. Philosophy, London School of Economics, 2015–2019

Dissertation Title: *The Problem of Granularity for Scientific Explanation*.

Advisors: Katie Steele (ANU), Jonathan Birch (LSE), Luc Bovens (UNC Chapel Hill)

Examiners: Christopher Hitchcock (Cal Tech) and Christian List (LSE)

Visiting Student, Australian National University, February–August 2017, August 2018

M.Sc. Philosophy and Public Policy, London School of Economics, 2013–2014 (Distinction)

A.B. Philosophy, Dartmouth College, 2007–2011 (*cum laude*, High Honors)

Publications

2019. [On the Explanatory Depth and Pragmatic Value of Coarse-Grained, Probabilistic, Causal Explanations](#). *Philosophy of Science*. 86(1): 145-167.

2019. [Inductive Explanation and Garber-Style Solutions to the Problem of Old Evidence](#). *Synthese*. 196(10): 3995-4009.

2018. [Imprecise Bayesian Networks as Causal Models](#). *Information*. 9(9), 211.

Awards and Scholarships

LSE Philosophy Class Teacher Award 2016, 2019

LSE Popper Prize (biennial award for distinguished graduate work in an area of philosophy to which Karl Popper made significant contributions), 2018

National Science Foundation Travel Grant 2018

LSE Postgraduate Travel Award 2016 (x2), 2017 (x3), 2018 (x3)

LSE Student's Union Teaching Award: Highly Commended for Feedback and Communication 2016

LSE Ph.D. Studentship 2015–2019

James B. Reynolds Scholarship for Graduate Study Abroad 2013–2014 (provided M.Sc. tuition)

Teaching

Courses Taught at LSE (As Graduate Teaching Assistant):*

*At LSE, graduate teaching assistants have full responsibility for weekly classes and marking formative and summative work.

PH103: Introduction to Philosophy. 2018-19

PH201: Philosophy of Science. 2017-18

PH225: Business and Organizational Ethics. Autumn 2016

PH101: Logic. 2015-16

PH222: Philosophy and Public Policy. 2015-16

Courses Taught at King's College London (As Graduate Teaching Assistant):

6AANA046: Topics in Political Philosophy. Autumn 2016

Teacher Training:

Postgraduate Certificate in Higher Education

Recent Conference and Workshop Presentations

Kolmogorov Complexity and Occam's Razor Assumptions in Inductive Causal Inference. Conference for Philosophy of Science and Formal Methods in Philosophy. December 2019. Gdansk.

Algorithmic Causal Modeling as a More General Model of Inductive Causal Inference. European Philosophy of Science Association Biennial Conference. September 2019. Geneva.

Efficient Information Gathering for Agents with a Causal Model of their Environment. Workshop on Causal Cognition in Humans and Machines. May 2019. Oxford.

Pragmatic Causal Feature Learning. Workshop on Causation vs. Constitution: Loosening the Friction. December 2018. Bergen.

Curie's Principle and Causal Graphs.

Philosophy of Science Association Biennial Meeting. November 2018. Seattle, WA.

Society for Exact Philosophy Annual Conference. May 2018. Storrs, CT.

Bayesian Networks and Multi-Level Causation. APA Central Division Meeting. Feb 2018. Chicago, IL.

Service

Reviewer for *Philosophy of Science* (x2), *Synthese*, *Economics and Philosophy*, *European Journal for Philosophy of Science* and *International Journal of Approximate Reasoning*.

Non-Academic Employment

Hampton Partners. Mergers and Acquisitions Analyst. London. 2014-2015

UK House of Lords. Parliamentary Intern to Baroness Claire Tyler of Enfield. London. 2013-2014 (concurrent with M.Sc. study)

Hughes Hubbard and Reed LLP. Corporate Paralegal. New York. 2011-2013

Curriculum Vitae

Simon DeDeo

sdedeo@andrew.cmu.edu / <http://santafe.edu/~simon> / +1-505-577-2723

I. POSITIONS HELD

- Assistant Professor, Social and Decision Sciences, Carnegie Mellon University. 2017—.
- External Professor, Complexity Science Hub Vienna, Vienna, Austria, 2017—.
- External Professor, Santa Fe Institute, Santa Fe, New Mexico, 2014—.
- Assistant Professor, School of Informatics and Computing; Faculty in Cognitive Science, College of Arts & Sciences. Indiana University, 2014—2016.
- Omidyar Fellow of the Santa Fe Institute, Santa Fe, New Mexico, 2010–2013.
- Postdoctoral Fellow of the Institute for the Physics and Mathematics of the Universe, University of Tokyo, 2009.
- Postdoctoral Fellow of the Kavli Institute for Cosmological Physics, University of Chicago, 2006–2009.

II. EDUCATION

- Princeton University, Department of Astrophysical Sciences. Ph.D., 2006. Thesis: “Dark Energy : Theory and Observational Prospects.” David Spergel (advisor). Thesis committee: Jerry Ostriker (Princeton, Astrophysics), Paul Steinhardt (Princeton, Physics).
- Cambridge University, Department of Applied Mathematics and Theoretical Physics. Part III of Mathematical Tripos (one year taught Masters) *Distinction* grade. 2000–2001. Thesis: “Non-Gaussian Perturbations in the Cosmic Microwave Background.” Thesis Advisor: E. P. S. Shellard. King’s College Scholar, Cambridge University, 2001.
- Harvard University, Department of Astrophysics. A.B., *Magna cum laude* 1996–2000. Thesis: “General Relativistic Constraints on Anomalous X-ray Pulsar Emission Models.” Thesis Advisor: Ramesh Narayan.

III. PUBLICATIONS MOST RELEVANT TO PROPOSAL

- Opacity, Obscurity, and the Geometry of Question-Asking. Christina Boyce-Jacino, **Simon DeDeo**. *Cognition*. Accepted, in Press. <https://arxiv.org/abs/1809.08291> 2019.
- Individuals, institutions, and innovation in the debates of the French Revolution. Alexander T. J. Barron, Jenny Huang, Rebecca Spang, **Simon DeDeo**. *Proceedings of the National Academy of Sciences*, 115.18: 4607-4612. Awarded Cozzarelli Prize for best paper in the Behavioral Sciences in PNAS, 2018.

- State power and elite autonomy in a networked civil society: The board interlocking of Chinese non-profits. Ji Ma, **Simon DeDeo**. *Social Networks*, 54, 291-302. 2018.
- The Evolution of Lossy Compression. Sarah Marzen & **Simon DeDeo**. *Journal of The Royal Society Interface* 14 (130). Pages 20170166. 2017.
- Exploration and exploitation of Victorian science in Darwin's reading notebooks. Jaimie Murdock, Colin Allen, **Simon DeDeo**. *Cognition* 159, 117-126. 2017.
- Conflict and computation on Wikipedia: A finite-state machine analysis of editor interactions. **Simon DeDeo**. *Future Internet* 8 (3), Pages 31–54. 2016.
- The Evolution of Wikipedia's Norm Network. Bradi Heaberlin, **Simon DeDeo**. *Future Internet*, 8(2), 14–35. 2016.
- Major Transitions in Political Order. **Simon DeDeo**. Chapter for *From Matter to Life: Information and Causality* (Cambridge University Press). 393–429. Edited by Sara Imari Walker, Paul C.W. Davies, George Ellis. 2016.
- Sarah Marzen, **Simon DeDeo**. Weak universality in sensory tradeoffs. *Physical Review E* 94 (6). Pages 060101. 2016.
- Wrong side of the tracks: Big Data and Protected Categories. **Simon DeDeo**. Chapter for *Big Data is Not a Monolith* (MIT Press). 31–41. Edited by Cassidy R. Sugimoto, Hamid Ekbia, Michael Mattioli. 2016.
- Optimal High-Level Descriptions of Dynamical Systems. David H Wolpert, Joshua A Grochow, Eric Libby & **Simon DeDeo**. Chapter for *From Matter to Life: Information and Causality* (Cambridge University Press). Edited by Sara Imari Walker, Paul C.W. Davies, George Ellis. 2016.
- Social feedback and the emergence of rank in animal society. Elizabeth Hobson & **Simon DeDeo**. *PLoS Computational Biology* 11(9). Page e1004411. 2015.
- Group Minds and the Case of Wikipedia. **Simon DeDeo**. *Human Computation* 1:1. 5–29. 2014.
- The Civilizing Process in London's Old Bailey. Sara Klingenstein, Tim Hitchcock, & **Simon DeDeo**. *Proceedings of the National Academy of Sciences*. 111 (26) 9419-9424. 2014.
- Robust Compressed Sensing and Sparse Coding with the Difference Map. Will Landecker, Rick Chartrand & **Simon DeDeo**. *European Conference on Computer Vision*, Part III, LNCS 8691, pp. 315329 (2014).
- Demystifying Information-Theoretic Clustering. Greg Ver Steeg, Aram Galstyan, Fei Sha & **Simon DeDeo**. *Proceedings of Machine Learning Research*, 32(1):19-27, 2014.
- Dynamical structure of a traditional South American social network. Paul L. Hooper, **Simon DeDeo**, Ann E. Caldwell Hooper, Michael Gurven, Hillard S. Kaplan. *Entropy*, 15, 4932-4955. 2013.

Kent Kai-hsiung Chang

kentkchang@cmu.edu
<https://www.kentchang.com/>

EDUCATION

UCL (University College London), London, United Kingdom 2017–2018

M.S. in Digital Humanities

thesis: “Literary Modelling Queer Modernity, 1860–1922: Digital Humanities between Literary Studies and Data Science” (advised by Dr. Julianne Nyhan and Dr. Luke Dickens)

National Taiwan University, Taipei, Taiwan 2014–2017

B.A. in Foreign Languages and Literatures (literatures in English, linguistics, and translation)

University of Taipei, Taipei, Taiwan 2011–2014

B.A. program in English Instruction (TESOL), double major in Computer Science; transferred to National Taiwan University in September 2014

SELECTED PROFESSIONAL EXPERIENCE

Intern July 2018–November 2018

Department of Research Services, the British Library (United Kingdom)

(advised by Simon Moffatt)

- created prototype of a web app to facilitate access to the Library’s textual resources
- wrote Python tutorial series in Jupyter and Sphinx

SELECTED TEACHING EXPERIENCE

Teaching assistant September 2016–June 2017

Department of Foreign Languages and Literatures, National Taiwan University (Taiwan)

- TA for FL1009, 1010 Freshman English (with 2-hour Aural Training), session 03, to Prof. Hui-chuan Chang
- created and maintained additional course websites for my TA section: <http://static.kentkchang.com/fe/> and <http://static.kentkchang.com/fe2/>

Teaching assistant February 2014–June 2014

Department of English Instruction, University of Taipei (Taiwan)

- TA for 1018 Sophomore English Listening and Speaking Practicum to Mr. Matthew Werth

Part-time English language instructor July 2011–September 2017

Ying-Da Private Language School

- Teacher of General English Proficiency Test, the most widely recognised English proficiency test in Taiwan
- offered CEFR A2–C1 level classes

SELECTED RESEARCH EXPERIENCE

Predocctoral research fellow December 2018–present
 Laboratory for Social Minds, Social and Decision Sciences, Carnegie Mellon University
 (advised by Prof. Simon DeDeo)

- leading projects focused on the quantitative analysis of queer literature and culture

Research assistant and researcher July 2018–February 2019
 .txtLAB, McGill University and HathiTrust Research Center
 (advised by Prof. Richard Jean So)

- parsed and cleaned OCR'd magazine journal texts of HathiTrust Research Center
- assisted semantic history analysis

Research assistant October–November 2018
 School of Information Sciences, University of Illinois at Urbana-Champaign
 (advised by Prof. Ted Underwood)

- extracted data from periodicals in HathiTrust Research Center
- assisted topic modeling novel review data

SELECTED PRESENTATIONS

INVITED TALKS

“How to Do Queer Cultural Analytics.” World Views and Mythologies, Ancient and Modern Conference, South Methodist University, New Mexico, United States. 4 July 2019.

“Marriage, Meditation, and Mediation.” Natural Language Processing Special Interest Group, the Alan Turing Institute, London, United Kingdom. 27 September 2018.

CAMPUS AND DEPARTMENTAL TALKS

“The Queer Use of Hidden Markov Models.” Language Technologies Institute, Carnegie Mellon University, Pennsylvania, United States. 1 November 2019.

“Advanced Text Analysis with Python.” Department of Social and Decision Sciences, Carnegie Mellon University, Pennsylvania, United States. 9 October 2019.

“Introduction to Text Analysis with Python.” Department of Social and Decision Sciences, Carnegie Mellon University, Pennsylvania, United States. 24 September 2019.

“Digital Humanities: a Historical Introduction.” Social Computing Reading Group, Carnegie Mellon University, Pennsylvania, United States. 20 September 2019.

SELF-ORGANIZED READING GROUPS

Cultural Analytics Reading Group. Carnegie Mellon University, Pennsylvania, United States. September 2019–present. <https://ca.kentkchang.com/>

Cultural and Literary Text Mining Study Group. UCL (University College London). London, United Kingdom. June 2018–August 2018. <https://caltmig.kentchang.com/>

Tatyana Gershkovich
Curriculum Vitae

Department of Modern Languages
Dietrich College of Humanities and Social Sciences
Carnegie Mellon University
180 Baker Hall
5000 Forbes Avenue
Pittsburgh, PA 15213
tgershko@andrew.cmu.edu

Home Address:

(b) (6)
(b) (6)
(b) (6)

ACADEMIC POSITIONS

2016– Assistant Professor of Russian Studies, Department of Modern Languages, Carnegie Mellon University
Fall 2015 Visiting Lecturer, Russian Department, Dartmouth College

EDUCATION

2016 Ph.D., Department of Slavic Languages and Literatures, Harvard University
Dissertation: “Held Captive: Tolstoy, Nabokov, and the Aesthetics of Constraint”
2007 B.A., magna cum laude in Slavic Studies, Harvard University

RESEARCH & PUBLICATIONS

Areas of Interest:

Nineteenth- and twentieth-century Russian prose; modernism; European philosophy and intellectual history; aesthetics; philosophical skepticism; interwar émigré history and culture; digital archives; the epistemology of computational literary analysis; Siberia as a cultural space in Imperial and Soviet Russia

Peer-Reviewed Publications:

2019 “Suspicion On Trial: Tolstoy’s *The Kreutzer Sonata* and Nabokov’s ‘Pozdnyshv’s Address,’” *PMLA*, 134. 3 (May 2019): 459–474.
2019 “Self-translation and the Transformation of Nabokov’s Aesthetics from *Kamera obskura* to *Laughter in the Dark*,” *Slavic and East European Journal* 63.2 (2019): 206–225.
2013 “Infecting, Simulating, Judging: Tolstoy’s Search for an Aesthetic Standard,” *Journal of the History of Ideas* 74 (January 2013): 115–137.
2017 “Хаджи-Мурат, как прообраз эстетической восприимчивости,” (“Hadji Murat’s Exemplary Aesthetic Receptivity”) *Proceedings of the 10th International Leo Tolstoy Conference at Yasnaya Polyana*, ed. by Galina Alekseeva (The Yasnaya Polyana Publishing House, 2018), 141–150. Proceedings of a peer-reviewed conference.

Tatyana Gershkovich
Curriculum Vitae

Other Publications:

- 2019 “In Impossible Proximity: How to Read Like Nabokov,” *The Berlin Journal* 33 (Fall 2019): 55–57.
2016 “News of the Profession: Tolstoy and World Literature 2016,” *Tolstoy Studies Journal* 28 (2016): 143–144.

Manuscripts in Progress:

Book project: *Impossible Proximity: Style and the Problem of Other Minds in Tolstoy, Nabokov*
Book project: *Tolstoy Red and White*
Translation project: *In Praise of Idleness: Selected Essays by Yulii Aikhenvald*, co-translated and edited with Stephen H. Blackwell

DIGITAL HUMANITIES

- 2015– Principal Investigator, *Beyond the Ant Brotherhood: A Digital Visualization of Tolstoy’s Intellectual World*, at www.colloquy.us.

AWARDS & FELLOWSHIPS

- 2019 Berlin Prize, The American Academy in Berlin: Nina Maria Gorrissen Fellow in History
2019 Falk Grant for “Conditions of Creation: Analyzing Creativity with Computational Models,” Carnegie Mellon University
2017 Mellon Digital Humanities (DH) Seed Grant for *Beyond the Ant Brotherhood: A Visualization of Tolstoy’s Intellectual World*, Carnegie Mellon University
2017 Berkman Faculty Development Grant for *Beyond the Ant Brotherhood: A Visualization of Tolstoy’s Intellectual World*, Carnegie Mellon University
2017 David Sloane Memorial Prize awarded for scholarly promise on the basis of “Held Captive: Tolstoy, Nabokov, and the Aesthetics of Constraint,” Harvard University
2016 Princeton Society of Fellows Finalist
2014–2015 Dissertation Completion Fellowship, Davis Center for Russian Studies, Harvard University
2014 The Harvard Graduate Society Merit Fellowship (Term-Time Research Award)
2014 Maurice Lazarus Graduate Research Travel Award, Davis Center for Russian Studies, Harvard University
2013 Setchkarev Memorial Prize for Best Graduate Student Essay, Harvard University
2011 Maurice Lazarus Graduate Research Travel Award, Davis Center for Russian Studies, Harvard University
2011–2012 Foreign Language and Area Studies Grant, U.S. Department of Education
2011 Setchkarev Memorial Prize for Best Graduate Student Essay, Harvard University

Jo Guldi, PhD

Associate Professor in the History of Britain and its Empire
Scholar of Digital History & the Methodology of History

<http://www.joguldi.com>

Appointments

Director, *Think-Play-Hack* (2018-)

Associate Professor of History, Southern Methodist University, Dallas, TX (2018-)

Assistant Professor of History, Southern Methodist University, Dallas, TX (2016-8)

Hans Rothfels Assistant Professor of History, Brown University, Providence, RI (2014-6)

Assistant Professor of History, Brown University, Providence, RI (2012-6)

Junior Fellow, Harvard Society of Fellows, Cambridge, MA (2009-10, 2011-3)

Mellon Postdoctoral Fellow in Digital History, Department of History, University of Chicago, Chicago, IL (2008-9, 2010-1)

Education

PhD History, University of California, Berkeley, 2004-2008. Dissertation: *The Road to Rule: The expansion of the British road network, 1740-1850*. Primary field: British History, 1688-1950. Secondary fields: Urban History, Architecture. Advisor: Professor James Vernon, British History.

AB Literature, Magna cum Laude, Harvard College, 1996-2000.

Completed Publications: Books

(with David Armitage) *The History Manifesto* (Cambridge University Press, 2014; revised edition, 2015; Portuguese translation: Autentica Editora: in progress; Chinese translation, Truth & Wisdom Press: in progress; Japanese translation, Tōsui Shobō: in progress; Korean translation, Hanul Ak'ademi: in progress; Russian translation, *Ab Imperio*, 1–4/2015; Spanish translation, Alianza Editorial: 2016; Turkish translation, Türkiye İş Bankası Kültür Yayınları: 2016), x + 165 pp. [*New Statesman* Book of the Year, 2014; *El País* book of the week, Sep. 2016]; named to “20 Most Influential Books of the Past 20 Years” by the *Chronicle of Higher Education* (2018).

Roads to Power: Britain Invents the Infrastructure State (Harvard University Press, 2012).

Reviews in *The Wall Street Journal*, *Technology and Culture*, *Journal of British Studies*, *Victorian Studies*, *The American Historical Review*, etc.

Digital Projects

These projects are relics of a digital age that compare to books in terms of the grant-writing behind them, the multiple hands that shaped them, the weight of ideas and research that went into them, and their intended wide-ranging readership.

With SMU OIT and Andrew Sempere, *DemocracyLab*, <https://docs.democracylab.io/docs/welcome.html>, digital infrastructure for rendering transparent the debates of any democratic legislature (build in process, 2019-).

With Brown Data Science and Poom Chiarawongse, *HaTORI* (HAnsard TOPic Relevance Identifier), <https://eight1911.github.io/hansard/>, digital software for browsing the Hansard parliamentary debates of the nineteenth century by “topic model” (released 2018).

With Cora Johnson Roberson, *Paper Machines*, digital software for historians (released 2012), <http://www.papermachines.org>

What is the Spatial Turn? (Charlottesville: University of Virginia, Scholar’s Lab, 2012) – digital manuscript only at present (the manuscript has been requested for inclusion in the Harvard University Press book series MetaLABprojects <http://www.hup.harvard.edu/features/metalabprojects/>) : <http://spatial.scholarslab.org/spatial-turn>

Completed Publications: Refereed Journal Articles

“Towards a Set of Principles for Humanistic Infrastructure,” *Digital Humanities Quarterly* (forthcoming)

“A New Atlas of British Empire,” commissioned by Antoinette Burton for a special issue of the *Journal of World History* (in review)

“The Measure of Modernity,” *International Journal for History, Culture, and Modernity* (October 2019)

“The Modern Paradigm of Explanation,” *Isis* (June 2019)

“Topic Modeling the History of Infrastructure in Nineteenth-century Great Britain,” *Technology and Culture* (Spring 2019)

“The Rise of Global Squatterdom,” *Humanity Journal* (March 2019)

“Critical Search: A Procedure for guided reading in large-scale textual corpora,” *Journal of Cultural Analytics* (December 2018). Includes code and data.

(with Ben Williams) “Synthesis and Large-scale Textual Corpora: A Nested topic model of Britain’s Debates over Landed Property in the Nineteenth Century,” in *Current Research in Digital History* 1:1 (2018): <http://crdh.rrchnm.org/essays/v01-01-synthesis-and-large-scale-textual-corpora/>. Includes code, data, and tool.

“Global Questions About Rent and the Longue Duree of Urban Power, 1848 to the Present,” *New Global Studies* (March 2018).

“A History of the Participatory Map,” *Public Culture* 29:1 (January 1, 2017): 79–112. http://publicculture.dukejournals.org/content/29/1_81/79.abstract

JULIA LEFKOWITZ

Address: Pembroke College, St. Aldates
Oxfordshire, Oxford OX1 1DW

Email: Julia.lefkowitz@oii.ox.ac.uk

Phone Number (Mobile): 07768 946634

Education:

DPhil Candidate, Oxford Internet Institute

University of Oxford, October 2016 – Present

Doctoral Thesis: “‘Tabloidization’ in the Internet Age”

- Thesis investigates whether or not a ‘tabloidization’ of British news publications has occurred in the wake of the ascendance of the Internet. ‘Tabloidization’ and its putative causal factors – namely economic, technological, and organizational – are subject to linguistic, content, and ethnographic analysis. Amidst present pivotal shifts in the changing nature of mass media, the role of journalism in society, and the public sphere(s), findings aimed towards media scholars and practitioners.

DPhil Candidate, English Faculty¹

University of Oxford, October 2013 – July 2016

Doctoral Thesis: “Linguistic Change and Continuity in UK Tabloid and ‘Quality’ Newspapers”

Master of Arts in Global Communications

The American University of Paris, January 2011 – January 2013

Distinction

Master’s Thesis: “Mediating Authenticity: French Prestige Press Coverage of Shang Xia”

Bachelor of Arts

Connecticut College, New London, Connecticut, 2002 – 2006

Double Major, English and French, Distinction in French

Publications:

(Forthcoming). Lefkowitz, Julia. 2021. “‘Tabloidization’ in the Internet Age” in *Global Tabloid*, ed. Martin Conboy. Routledge: London.

Blank, G., Dutton, W., Lefkowitz, J. (2020). OxIS 2019: The Rise of Mobile Internet in Britain. Oxford Internet Institute. University of Oxford.

Blank, G., Dutton, W., Lefkowitz, J. (2020). OxIS 2019: Digital Divides in the UK are Narrowing but Deepening. Oxford Internet Institute. University of Oxford.

Blank, G., Dutton, W., Lefkowitz, J. (2020). Perceived Threats to Privacy Online: The Internet in Britain, the Oxford Internet Survey, 2019. Oxford Internet Institute. University of Oxford.

Blank, G., Dutton, W., Lefkowitz, J. (2020). OxIS 2019: Dueling Perspectives on the Internet in Britain. Oxford Internet Institute. University of Oxford.

Lefkowitz, Julia. 2016. “‘Tabloidization’ or Dual-Convergence: Quoted Speech in Tabloid and ‘Quality’ British Newspapers 1970 – 2010.” *Journalism Studies* 16 June 2016: 1 – 23.

¹ In 2016, I switched academic departments due to a number of factors whereby the Oxford Internet Institute emerged as a much better-suited fit for my thesis. The dissertation topic is the same, with an increased focus of the Internet’s possible impact on journalistic values across publication types and mediums.

Lefkowitz, Julia. 2015. "A Comparative Approach to Transatlantic Scandal: *Le Monde* and the *New York Times*' Coverage of Roman Polanski's 2009 Arrest." *Networking Knowledge*, 8(3): 1 – 20.

Lefkowitz, Julia. 2013. "The Dominique Strauss-Kahn Scandal: Mediating the Desire for Authenticity," in *Media and Public Shaming: Drawing the Boundaries of Disclosure*, ed. Julian Petley. London: I.B. Tauris.

Conferences:

"'Tabloidization' in the Internet Age." Rethinking the Press in the Digital Ecosystem. Catholic University of Portugal. Braga, Portugal. July, 2019.

"'Tabloidization' in the Internet Age." European Communication and Research Education Association Annual International Conference. Università della Svizzera, Lugano, Switzerland. November, 2018.

"Scandal in the Internet Era: A Longitudinal, Quantitative Analysis." American Sociology Association, Media Sociology Pre-Conference. Pennsylvania Convention Center, Philadelphia, United States. August, 2018.

"Scandal Pre- and Post-Internet Era: A Quantitative and Qualitative Diachronic Analysis of British Newspaper Scandal Coverage." International Conference in Scandology. University of Bamberg, Bamberg, Germany. March, 2018.

"'Tabloidization' in the Internet Age." American Sociology Association, Media Sociology Pre-Conference. Concordia University, Montreal, Canada., August 2017.

"A Diachronic Study of Linguistic Change and Continuity in British Tabloid and 'Quality' Print Newspapers." *New Directions in Media Research*. University of Leicester, Leicester, UK. June, 2015.

"U.S. and French Prestige Press Coverage of Shang Xia: Examining the Ascension of Journalistic 'Authenticity.'" *The Unlovable Press*. Conversations with Michael Schudson. University of Groningen, Groningen, the Netherlands. June, 2014.

"The Dominique Strauss-Kahn Scandal: Mediating the Desire for Authenticity." *Media and the Boundaries of Disclosure: Media, Morals, Public Shaming and Privacy*. The Reuters Institute for the Study of Journalism, University of Oxford, Oxford. February, 2012.

"The Mediation of Power and Hybridity: *The New York Times* and *Le Monde*'s Coverage of the DSK Scandal." *The Mediation of Scandal and Moral Outrage*. The London School of Economics, London. December, 2011.

Relevant Work Experience:

Oxford Internet Institute, Research Assistant, OxIS 2019 Project, January 2019 – Present

Oxford Internet Institute, Teaching Assistant, Statistics Core Course, Fall 2017 and Fall 2018

Policy & Internet, *Peer Reviewer*, January 2017 - Present

Journal of Communication, *Peer Reviewer*, January 2017 - Present

Networking Knowledge, *Peer Reviewer*, January 2016 – Present

Connected Life Conference, University of Oxford, *Conference Submissions Reviewer*, Spring 2017

Kaliwatch International (Qualitative Media Research Company), *Research Intern*, January 2012 – August 2012

Dana Thomas (Journalist and Novelist), *Research Assistant*, Fall 2011

Organization for Economic Cooperation and Development, Wikiprogress, *Editorial Volunteer*, Summer 2011

Cravath, Swaine & Moore LLP, *Litigation Legal Assistant*, Summer 2006 – Spring 2009

The United Nations, *Research Intern*, Summer 2005

Tyler J. S. Marghetis

1399 Hyde Park Road, Santa Fe, New Mexico, USA 87501

(b) (6)

• www.tylermarghetis.com • (b) (6)

RESEARCH INTERESTS

Regimes and ruptures in human thought. How thought emerges from brains, bodies, and culture. How we make sense of the invisible, imagined, or unfathomably complex

EMPLOYMENT AND EDUCATION

University of California, Merced

2020 – Assistant Professor, Cognitive & Information Sciences

Santa Fe Institute

2019 – 2022 Omidyar Complexity Fellow

Indiana University, Bloomington

2015 – 2019 Postdoctoral Researcher, Psychological & Brain Sciences ('15-'19) & O'Neill School of Public & Environmental Affairs ('17-'19)

EDUCATION

2015 **Ph.D. in Cognitive Science, UC San Diego**

Fall 2011 **Visiting graduate student, Philosophy, UC Berkeley**

2009 **Master in the Teaching of Mathematics, Concordia University**

2007 **Honours B.Sc., Pure & Applied Mathematics, Concordia U., Canada**

SELECTED RECENT PUBLICATIONS

1. **Marghetis, T.**, Landy, D., & Attari, S. (2019). Understanding and correcting perceptions of home energy use. *Nature Energy*, 4, 874-881.
2. **Marghetis, T.** Samson, K., & Landy, D. (2019). The complex system of mathematical creativity. *Proceedings of the Annual Meeting of the Cognitive Science Society*.
3. Hendricks, R., Bergen, B. K., & **Marghetis, T.** (2018). Do metaphors move from mind to mouth? Evidence from new metaphors for time. *Cognitive Science*, 42, 2950-2975.
4. **Marghetis, T.**, Guay, B., Karlapudy, A., & Landy, D. (2018). The psychophysics of society. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

5. Setzler, M.*, **Marghetis, T.***, & Kim, M. (2018). Creative leaps in musical ecosystems. *Proceedings of the Cognitive Science Society*. *joint first authors
6. Cooperrider, K., **Marghetis, T.**, & Núñez, R. (2017). Where does the ordered line come from? Evidence from Papua New Guinea. *Psychological Science*, 28, 599-608.
7. Klein, S.A., & **Marghetis, T.** (2017). Shaping Experiment from the Inside Out: Performance-Collaboration in the Cognitive Science Lab. *Performance Matters*, 3.
8. Tillman, K., **Marghetis, T.**, Barner, D., & Srinivasan, M. (2017). Today is tomorrow's yesterday. *Cognitive Psychology*, 92, 87-100.

SELECTED BOOK CHAPTERS

1. **Marghetis, T.**, & Bergen, B. (2014). Embodied meaning, inside and out: Coupling gesture and mental simulation. *Body-Language-Communication*. Mouton de Gruyter.
2. **Marghetis, T.**, Edwards, L.D., & Núñez, R (2014). More than mere handwaving: Gesture and embodiment in expert mathematical proof. *Emerging Perspectives on Gesture and Embodiment in Mathematics*.

SELECTED INVITED TALKS

1. "A recipe for revolution? Big data and sudden ruptures in math and music." Colorado College, Colorado Springs, USA. Apr. 2020.
2. "Breakthroughs, breakdowns, breakups: The stable regimes and sudden revolutions of human cognition." Concordia University, Canada. Oct. 2019.
3. "Regimes and revolutions in communication and cognition." UC Merced. Feb. 2019.

SELECTED ORGANIZED WORKSHOPS

1. *Origins of time: Insights into the psychological foundations of time*. Canada, Jul. 2014
2. *Bridging Systems Neuroscience and Embodied Cognition*. Invited organizer of 'Advanced Course' for Champalimaud Neuroscience Programme, Portugal, Oct. 2013

OTHER EMBODIED ACCOMPLISHMENTS

Olympic Alternate, 2008 Beijing Olympics, Canadian freestyle wrestling team (74 kg)
National Team Athlete, 2002-2009, and four-time university national champion (Canada)

Carrie R. Cowan
Director for Education | Santa Fe Institute

1399 Hyde Park Road
Santa Fe, New Mexico 87501 USA
office +1 (505) 946 2726
ccowan@santafe.edu

Summary

Carrie Cowan directs the educational mission of the Santa Fe Institute, including development, implementation, and evaluation of in-person and virtual programs. For more than a decade, she has delivered advanced interdisciplinary educational curricula, including research methods training and professional development. She previously led a research group studying molecular pattern formation and information processing in biological systems.

Education

B.A.	Wellesley College , Wellesley, Massachusetts, USA Biological Sciences	June 1995
Ph.D.	University of California , Berkeley, California, USA Plant and Microbial Biology	May 2001

Professional Roles

Director for Education	Santa Fe Institute Santa Fe, New Mexico USA	September 2019 – present
Director, Pre- & Postdoctoral Education	The Jackson Laboratory Bar Harbor, Maine USA	June 2016 – August 2019
Associate Dean	Cold Spring Harbor Laboratory Cold Spring Harbor, New York USA	August 2013 – May 2016
Research Group Leader	Research Institute of Molecular Pathology (IMP) Vienna AUSTRIA	December 2006 – November 2013
Postdoctoral Fellow	Max Planck Institute Molecular Cell Biology & Genetics Dresden GERMANY	June 2001 – November 2006

Selected Professional Service & Teaching

Program Director & co-Director | NIH NIGMS R25 IPERT (PI Paigen), 2017 – 2019

Program Manager | NIH NICHD T32 Institutional Training Grant (PI Braun), 2016 – 2019

Program Consultant & Evaluator | NIH NGMS T32 Institutional Training Grant (PI Gann), 2013 – 2016

Program Consultant & Evaluator | NSF BIO REU, 2013 – 2016 (PI Schatz, Churchland) 2013 – 2016

Assistant Program Coordinator, Vienna BioCenter International Ph.D. Program, 2008 – 2013

Instructor, College of the Atlantic (Bar Harbor, Maine USA) | Developmental Biology, 2018

Lecturer, University of Vienna (Vienna AUSTRIA) | Developmental Biology, Advanced Cell Biology 2007 – 2013

Curriculum Advisor, Computational Training, The Jackson Laboratory (Bar Harbor, Maine USA), 2018 – 2019

Peer Reviewer [journals] *Nature*, *Science*, *Developmental Cell*, *Development*, *Developmental Biology*, *Journal of Cell Biology*, *PLoS Genetics* | [funding agencies] National Science Foundation, European Research Council, European Molecular Biology Organization

Mentoring

[Postdoctoral Fellow]

J. Dobbelaere (2011-2013) | research fellow, Max F. Perutz Laboratory (Vienna AUSTRIA)

[Ph.D. Students]

D. Bienkowska (2007-2010) | CEO, ÜberEnergy (Berlin GERMANY)

A. Goepfert (2006-2010) | senior scientist, AstraZeneca (Cambridge UK)

M. Miki (2010-2013) | postdoctoral fellow, Weizmann Institute of Science (Rehovot ISRAEL)

S. Sanegre Sans (2009-2012) | clinical research manager, U Valencia (Valencia SPAIN)

[M.S. Students]

S. Millonig (2008) | family leave

S. Reiter (2009) | consultant, World Health Organization (Nyon SWITZERLAND)

Research Publications can be found at

https://scholar.google.com/citations?user=_ePj0MEAAAAJ&hl=en



Carnegie Mellon University

Department of Social and Decision Sciences
Carnegie Mellon University
Porter Hall 208
Pittsburgh, Pennsylvania 15213-3890

Fax (412) 268-6938

Simon DeDeo / Laboratory for Social Minds / Assistant Professor, Social and Decision Sciences, Carnegie Mellon University / External Professor, Santa Fe Institute / Santa Fe, NM 505-577-2723 / <http://santafe.edu/~simon> / 3 February 2020

Dear NEH Program Officers and Panelists,

This letter of commitment is to acknowledge my (enthusiastic) commitment to serving as Co-PI at Carnegie Mellon University for the grant “Foundations and Applications of Humanities Analytics in the Humanities”, as outlined in our proposal.

Very sincerely,

A handwritten signature in black ink that reads "Simon DeDeo".

Simon DeDeo

Kent Chang / Research Fellow / Laboratory for Social Minds / Social and Decision Sciences,
Carnegie Mellon University / 20 February 2020

RE: Letter of commitment, “Foundations and Applications in Humanities Analytics”

Dear Dr. Kinney,

Dear Dr. DeDeo,

I am writing to express my commitment to contribute to your Institute for Advanced Topics in the Digital Humanities “Foundations and Applications in Humanities Analytics.” I would be pleased to serve as a Consultant for the Institute and look forward to working with you both on this timely project.

Given my background in English literature and data science and my experience developing courses and workshops in digital humanities, I would be enthusiastic to advise and assist with the development of the virtual course curriculum in year one of the project. In particular, I look forward to developing Jupyter notebooks and Python programming exercises for the course, to guide participants through the computational concepts and analyses discussed in lectures. I will lead the development of an optional online module in topic modeling to prepare participants for the in-person workshop. I will help assemble additional learning resources, such as reading lists for further exploration, and will assist with the online discussion forum that accompanies the virtual course.

Further, I bring my experience as an English teacher and instructor of the aforementioned courses and workshops, and my skill as a data scientist, to the role of teaching assistant for the in-person workshops in years two and three. I will be available throughout the week-long session to assist participants one-on-one with their projects—in particular, the programming and computational aspects – and to assist you both as needed with the curriculum.

While much of my effort on the course can be performed virtually, I understand that for both curriculum development and in-person workshop implementation, I will be expected to travel to Santa Fe to work directly with the Institute team: twice per year in year one and once per year in subsequent years. I understand that my travel-related expenses to/from Santa Fe will be provided by the proposed Institute. I am committed to provide my services for three months (90 days) in year one

and for one month (30 days) in years two and three at the agreed-upon rate of \$100 per day, with a 2% cost of living increase in each of years two and three.

I am thrilled to be a part of this project.

Yours sincerely,

A handwritten signature in black ink that reads "Kent Chang". The signature is written in a cursive style with a large, sweeping flourish at the end of the word "Chang".

Kent Chang

Carnegie Mellon University

**Department of Modern Languages
Marianna Brown Dietrich College**
of Humanities and Social Sciences
Carnegie Mellon University
Baker Hall 160
Pittsburgh, Pennsylvania 15213-3890

Telephone: (412) 268-5669
Fax: (412) 268-1328

Dear David Kinney and Simon DeDeo,

I am writing to indicate my enthusiasm for your proposed Institute for Advanced Topics in the Digital Humanities “Foundations and Applications in Humanities Analytics”. I would be eager to contribute to the curriculum.

Given my expertise in digital approaches to intimate documents (e.g. diaries, letters), I will deliver a lecture on visualizing intellectual biography—specifically, the biography of Leo Tolstoy—as part of the planned online course in year 1. This will be complemented by a more in-depth discussion of my research and methodologies as part of the in-person workshops in years 2 and 3. Further, I will be available to guide participant projects during the in-person workshop. For both, I understand that my travel-related expenses to/from Santa Fe as well as a nominal honorarium will be provided by the proposed Institute.

I look forward to working with you on this valuable program.

Very sincerely,

Tatyana Gershkovich

Prof. Tatyana Gershkovich



DEPARTMENT OF HISTORY

February 17, 2020

Dear David Kinney and Simon DeDeo,

I am writing to indicate my enthusiasm for your proposed Institute for Advanced Topics in the Digital Humanities “Foundations and Applications in Humanities Analytics”. I would be eager to contribute to the curriculum.

Given my expertise in digital history, I will deliver a lecture on temporal analysis as part of the planned online course in year 1. This will be complemented by a more in-depth discussion of my research and methodologies as part of the in-person workshops in years 2 and 3. Further, I will be available to guide participant projects during the in-person workshop. For both, I understand that my travel-related expenses to/from Santa Fe as well as a nominal honorarium will be provided by the proposed Institute.

I look forward to working with you on this valuable program.

Jo Guldi
Associate Professor
Department of History



RE: Letter of commitment, "Foundations and Applications in Humanities Analytics"

Dear David,
Dear Simon,

I am writing to indicate my enthusiasm for your proposed Institute for Advanced Topics in the Digital Humanities "Foundations and Applications in Humanities Analytics". I would be eager to contribute to the curriculum.

Given my expertise in quantitative analysis of language in print media, I will deliver a lecture on the empirical measurement of Tabloidization and sensationalism in UK newspapers from 1968-2016, as part of the planned online course in year 1. This will be complemented by a more in-depth discussion of my research and methodologies as part of the in-person workshops in years 2 and 3. Further, I will be available to guide participant projects during the in-person workshop. For both, I understand that my travel-related expenses to/from Santa Fe as well as a nominal honorarium will be provided by the proposed Institute.

I look forward to working with you on this valuable program.

Warmest Regards,

Julia Lefkowitz

Julia Lefkowitz

February 14, 2020

RE: Letter of commitment, "Foundations and Applications in Humanities Analytics"

Dear David,
Dear Simon,

I am writing to indicate my enthusiasm for your proposed Institute for Advanced Topics in the Digital Humanities, "Foundations and Applications in Humanities Analytics." I would be eager to contribute to the curriculum.

Given my expertise in computational approaches to understanding human cognition and abstract thought, I will deliver a virtual lecture on how statistical techniques can be applied to understand cognitive ruptures — eureka moments, artistic breakthroughs, religious conversions — as part of the planned online course in year 1. This will be complemented by a more in-depth discussion of my research and methodologies as part of the in-person workshops in years 2 and 3, at the Santa Fe Institute. I will also be available to guide participant projects during the in-person workshop.

I look forward to working with you on this valuable program.



Tyler Marghetis

Omidyar Fellow (2019 – 2022)
Santa Fe Institute

Assistant Professor (starting July 2020)
Cognitive and Information Sciences
University of California, Merced

Carrie R. Cowan
Director for Education
Santa Fe Institute
ccowan@santafe.edu

February 17, 2020

RE: Letter of Commitment | NEH DH IATDH "Foundations and Applications of Humanities Analytics"

Dear David,
Dear Simon,

I herewith offer my enthusiastic support for your application to the NEH's Digital Humanities program to establish an institute on "Foundations and Applications of Humanities Analytics." This institute would align with the Santa Fe Institute's educational mission and leverage our existing expertise in both online and in-person training, especially in the rigorous application of mathematical and computational concepts to problems across disciplines. Given the recent emergence of the digital humanities, in general, and computational text analysis, in particular, as major interests of the participant groups that our educational programs serve, the proposed institute fills a growing need for scholars and would enable SFI to expand its reach among broad and diverse audiences.

I am especially excited about the opportunities afforded by this institute to reach scholars from groups that are historically underrepresented in computational fields including the digital humanities, and look forward to working with you to recruit and support women and people of color. Our ongoing efforts in SFI's Education group to enhance our outreach to these demographics are yielding great results, so I am confident that we can help you achieve your goals. We are already in the process of strengthening our networks with humanities departments at institutions around the country through direct, personal contacts. To this end, we continue to refine our focus on underserved or under-resourced institutions and those with strong participation of underrepresented scholars, offering in-person visits from our Educational team and/or faculty to these programs to raise awareness of opportunities at SFI, outline our commitment to diversity and equity, and invite participation.

I offer my own service with respect to curriculum development and program evaluation, as well as the services of my team, including Isa Ruiz (Online Education Coordinator) and Carla Shedivy (Education Program Manager). Isa and Carla will manage the virtual and in-person portions of the institute respectively, bringing their experience with SFI's existing educational programs. I will ensure that Isa and Carla have sufficient time away from their other obligations to assist with the proposed institution, estimating the required time at roughly one month of full-time effort per year of the project. I have set aside three months of full-time effort to provide the support necessary to ensure a successful institute that achieves the goals you have outlined. My experience developing, implementing, and evaluating transdisciplinary training curricula for graduate-level (and beyond) audiences provides me with a valuable perspective to assist in these efforts.

I look forward to working with you on this exciting and timely project.

With best regards,



February 17, 2020

Dear David,

I write in support of your application to the NEH's Institutes for Advanced Topics in the Digital Humanities program, to establish a course on "Foundations and Applications of Humanities Analytics" with Dr. Simon DeDeo (Carnegie Mellon University and Santa Fe Institute external faculty). I will ensure that you have sufficient time away from any obligations to the Santa Fe Institute (SFI) to conduct the project you propose.

The topic of humanities analytics, and in particular the focus on information theory and probability and how these are employed in service of text analysis, is particular resonant at SFI and reflects a growing interest among researchers globally to use data science and machine learning tools to understand human arts, culture and history. I applaud your efforts to bring these foundational concepts to humanities scholars and look forward to seeing the success of the proposed institute.

Sincerely,



Jennifer Dunne
SFI Vice President for Science

Evaluations and description of updates

This attachment is not applicable to this submission. (There were no previously funded Institutes).

**Nonprofit Organization
Indirect Cost Negotiation Agreement**

EIN: 85-0325494

Organization:

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

Date: May 8, 2019

Report No(s) .: 19-A-0709 (FY20E)
19-A-0710 (FY21E)

Filing Ref.:

Last Negotiation Agreement
dated September 15, 2017

The indirect cost rates contained herein are for use on grants, contracts, and other agreements with the Federal Government to which 2 CFR Part 200 apply for fiscal years beginning on or after December 26, 2014 subject to the limitations contained in Section II.A. of this agreement. Applicable OMB Circulars and the regulations at 2 CFR 230 will continue to apply to federal funds awarded prior to December 26, 2014. The rates were negotiated by the U.S. Department of the Interior, Interior Business Center, and the subject organization in accordance with the authority contained in applicable regulations.

Section I: Rates

Type	Effective Period		Rate*	Locations	Applicable To
	From	To			
Predetermined	01/01/20	12/31/20	57.76%	All	All Programs
Predetermined	01/01/21	12/31/21	57.76%	All	All Programs

1/ One-time rate extensions were based on the FY 2019 predetermined rate.

***Base:** Total direct costs less capital expenditures, participant support costs, and the portion of subawards in excess of the first \$25,000.

Treatment of fringe benefits: Fringe benefits applicable to direct salaries and wages are treated as direct costs; fringe benefits applicable to indirect salaries and wages are treated as indirect costs.

Treatment of paid absences: Vacation, holiday, sick leave, and other paid absences are included in salaries and wages and are claimed on grants, contracts, and other agreements as part of the normal cost for the salaries and wages. Separate claims for the costs of these paid absences are not made.

Section II: General

Page 1 of 3

A. Limitations: Use of the rate(s) contained in this agreement is subject to any applicable statutory limitations. Acceptance of the rate(s) agreed to herein is predicated upon these conditions: (1) no costs other than those incurred by the subject organization were included in its indirect cost rate proposal, (2) all such costs are the legal obligations of the grantee/contractor, (3) similar types of costs have been accorded consistent treatment, and (4) the same costs that have been treated as indirect costs have not been claimed as direct costs (for example, supplies can be charged directly to a program or activity as long as these costs are not part of the supply costs included in the indirect cost pool for central administration).

B. Audit: All costs (direct and indirect, federal and non-federal) are subject to audit. Adjustments to amounts resulting from audit of the cost allocation plan or indirect cost rate proposal upon which the negotiation of this agreement was based will be compensated for in a subsequent negotiation.

C. Changes: The rate(s) contained in this agreement are based on the organizational structure and the accounting system in effect at the time the proposal was submitted. Changes in organizational structure, or changes in the method of accounting for costs which affect the amount of reimbursement resulting from use of the rate(s) in this agreement, require the prior approval of the responsible negotiation agency. Failure to obtain such approval may result in subsequent audit disallowance.

D. Rate Type:

1. **Fixed Carryforward Rate:** The fixed carryforward rate is based on an estimate of the costs that will be incurred during the period for which the rate applies. When the actual costs for such period have been determined, an adjustment will be made to the rate for a future period, if necessary, to compensate for the difference between the costs used to establish the fixed rate and the actual costs.

2. **Provisional/Final Rate:** Within six (6) months after year end, a final indirect cost rate proposal must be submitted based on actual costs. Billings and charges to contracts and grants must be adjusted if the final rate varies from the provisional rate. If the final rate is greater than the provisional rate and there are no funds available to cover the additional indirect costs, the organization may not recover all indirect costs. Conversely, if the final rate is less than the provisional rate, the organization will be required to pay back the difference to the funding agency.

3. **Predetermined Rate:** A predetermined rate is an indirect cost rate applicable to a specified current or future period, usually the organization's fiscal year. The rate is based on an estimate of the costs to be incurred during the period. A predetermined rate is not subject to adjustment. (Because of legal constraints, predetermined rates are not permitted for Federal contracts; they may, however, be used for grants or cooperative agreements.)

E. Rate Extension: Only final and predetermined rates may be eligible for consideration of rate extensions. Requests for rate extensions of a current rate will be reviewed on a case-by-case basis. If an extension is granted, the non-Federal entity may not request a rate review until the extension period ends. In the last year of a rate extension period, the non-Federal entity must submit a new rate proposal for the next fiscal period.

F. Agency Notification: Copies of this document may be provided to other federal offices as a means of notifying them of the agreement contained herein.

G. Record Keeping: Organizations must maintain accounting records that demonstrate that each type of cost has been treated consistently either as a direct cost or an indirect cost. Records pertaining to the costs of program administration, such as salaries, travel, and related costs, should be kept on an annual basis.

H. Reimbursement Ceilings: Grantee/contractor program agreements providing for ceilings on indirect cost rates or reimbursement amounts are subject to the ceilings stipulated in the contract or grant agreements. If the ceiling rate

is higher than the negotiated rate in Section I of this agreement, the negotiated rate will be used to determine the maximum allowable indirect cost.

I. Use of Other Rates: If any federal programs are reimbursing indirect costs to this grantee/contractor by a measure other than the approved rate(s) in this agreement, the grantee/contractor should credit such costs to the affected programs, and the approved rate(s) should be used to identify the maximum amount of indirect cost allocable to these programs.

J. Other:

1. The purpose of an indirect cost rate is to facilitate the allocation and billing of indirect costs. Approval of the indirect cost rate does not mean that an organization can recover more than the actual costs of a particular program or activity.

2. Programs received or initiated by the organization subsequent to the negotiation of this agreement are subject to the approved indirect cost rate(s) if the programs receive administrative support from the indirect cost pool. It should be noted that this could result in an adjustment to a future rate.

3. This Negotiation Agreement is entered into under the terms of an Interagency Agreement between the U.S. Department of the Interior and the cognizant agency. No presumption of federal cognizance over audits or indirect cost negotiations arises as a result of this Agreement.

4. Organizations that have previously established indirect cost rates—exclusive of the 10% de minimis rate—must submit a new indirect cost proposal to the cognizant agency for indirect costs within six (6) months after the close of each fiscal year.

Section III: Acceptance

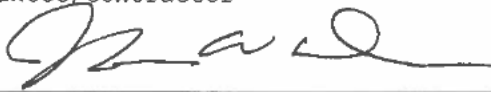
Listed below are the signatures of acceptance for this agreement:

By the Nonprofit Organization:

By the Cognizant Federal Government Agency:

Santa Fe Institute
Grantee/Contractor

National Science Foundation
Cognizant Agency

 /s/

CRAIG WILLS Digitally signed by CRAIG WILLS
Date: 2019.05.09 10:09:17 -0700 /s/

Signature
Jennifer Dunne
Name (Type or Print)

Signature
Craig A. Wills
Name

Vice President, Science
Title

Division Chief
Indirect Cost Services Division
Title

5/7/2019
Date

U.S. Department of the Interior
Interior Business Center
Agency

Negotiated by Muberra Guvenc
Telephone (916) 930-3816

Agreement Date: July 31, 2019

NEGOTIATION AGREEMENT

**INSTITUTION: CARNEGIE MELLON UNIVERSITY
 PITTSBURGH, PA 15213-3890**

The Facilities and Administrative (F&A), Cost of Money and Fringe Benefits rates contained herein are for use on grants, contracts and/or other agreements issued or awarded to the Carnegie Mellon University by all Federal Agencies of the United States of America, in accordance with the provisions and cost principles mandated by 2 CFR Part 200. These rates shall be used for forward pricing and billing purposes for the Carnegie Mellon University Fiscal Year 2020. This rate agreement supersedes all previous rate agreements/determinations for Fiscal Year 2020.

Section I: RATES - TYPE: FIXED (FIXED)

Facilities & Administrative (F&A) Rates

<u>Type</u>	<u>From</u>	<u>To</u>	<u>Rate</u>	<u>Base</u>	<u>Applicable To</u>	<u>Location</u>
Fixed	7/1/19	6/30/20	56.8%	(a)	Organized Research (1)	On-Campus
Fixed	7/1/19	6/30/20	59.0%	(a)	Organized Research (2)	On-Campus
Fixed	7/1/19	6/30/20	26.0%	(a)	Organized Research (1)	Off-Campus
Fixed	7/1/19	6/30/20	26.9%	(a)	Organized Research (1)	Off-Campus NREC ¹
Fixed	7/1/19	6/30/20	25.1%	(a)	Organized Research (2)	Off-Campus NREC
Fixed	7/1/19	6/30/20	60.2%	(a)	Organized Research (1)	Off-Campus Silicon Valley ²
Fixed	7/1/19	6/30/20	10.7%	(a)	All Programs	SEI ³

Cost of Money Factor

<u>Type</u>	<u>From</u>	<u>To</u>	<u>Rate</u>	<u>Base</u>	<u>Applicable To</u>	<u>Location</u>
Fixed	7/1/19	6/30/20	.00157	(a)	All Programs	SEI

¹National Robotics Engineering Center

²CMU Silicon Valley Campus

³Software Engineering Institute

Fringe Benefit Rates

<u>Category</u>	<u>Type</u>	<u>From</u>	<u>To</u>	<u>Rate</u>	<u>Base</u>	<u>Applicable To</u>	<u>Location</u>
Domestic Full Time	Fixed	7/1/19	6/30/20	24.5%	(b)	All programs	Pittsburgh
International Full Time	Fixed	7/1/19	6/30/20	24.9%	(c)	All programs	International
Domestic Part Time w/ Benefits	Fixed	7/1/19	6/30/20	23.9%	(d)	All programs	Pittsburgh
Domestic Part Time w/o Benefits	Fixed	7/1/19	6/30/20	7.8%	(e)	All programs	Pittsburgh

DISTRIBUTION BASES

(a) Modified Total Direct Costs consisting of salaries and wages, fringe benefits, graduate student stipends, materials and supplies, services, travel, and subawards up to the first \$25,000 of each subaward (regardless of the period covered by the subaward). Equipment, capital expenditures, charges for tuition remission, rental costs, scholarships and fellowships, participant support costs as well as the portion of each subaward in excess of \$25,000 shall be excluded from modified total direct costs.

(b) Total full time faculty, staff, and union salaries and wages.

(c) Total full time faculty, staff, and union salaries and wages working internationally.

(d) Total part time salaries and wages with benefits for employees who work at least 17.5 hours per week, in an appointment of four months or longer.

(e) Total part time salaries and wages with mandated benefits for employees who work less than 17.5 hours per week, or in an appointment of less than four months.

APPLICABLE TO

(1) Applies to DOD contracts awarded before November 30, 1993, all Non-DOD Instruments, and all DOD grants and other agreements (See Section II, paragraph F). (Capped)

(2) Applies to only DOD contracts awarded on or after November 30, 1993 in accordance with and under the authority of DFARS 231.303(1) (See Section II, paragraph F). (Uncapped)

SECTION II - GENERAL TERMS AND CONDITIONS

A. **LIMITATIONS:** Use of the rates set forth under Section I is subject to availability of funds and to any other statutory or administrative limitations. The rates are applicable to a given grant, contract or other agreement only to the extent that funds are available and consistent with any and all limitations of cost clauses or provisions, if any, contained therein. Acceptance of any or all of the rates agreed to herein is predicated upon the following conditions: (1) that no costs other than those incurred by the institution were included in this indirect cost pool as finally accepted and that such costs are legal obligations of the institution and allowable under governing cost principles; (2) that the same costs that have been treated as indirect costs are not claimed as direct costs; (3) that similar types of costs have been accorded consistent accounting

treatment; and (4) that the information provided by the institution which was used as a basis for acceptance of the rates agreed to herein, and expressly relied upon by the Government in negotiating and accepting the said rates is not subsequently found to be materially incomplete or inaccurate.

B. ACCOUNTING CHANGES: The rates contained in Section I of this agreement are based on the accounting system in effect at the time the agreement was negotiated. Changes to the method(s) of accounting for costs, which affect the amount of reimbursement resulting from the use of these rates require the prior written approval of the authorized representative of the cognizant agency for indirect costs. Such changes include but are not limited to changes in the charging of a particular type of cost from indirect to direct. Failure to obtain such approval may result in subsequent cost disallowances.

C. FIXED RATES WITH CARRY-FORWARD PROVISIONS: The fixed rates contained in this agreement are based on estimates of the costs for FY 2020. When actual costs for this fiscal year are determined, adjustments will be applied to a subsequent rate negotiation to recognize the difference between the FY 2020 estimated costs used to establish the fixed rates and the negotiated actual FY 2020 costs.

D. CARRY FORWARD AMOUNTS: The following FY 2006-2007, FY2011-2017 final and FY 2018 estimated carry-forward amounts were considered in the establishment of the FY 2020 rates:

() indicates over-recovery

F&A Rates	Liquidated in FY 2020 Rates		
	Fiscal Year	Capped	Uncapped
Organized Research On-Campus	2011	(\$157,012)	(\$55,005)
	2012	(\$135,565)	(\$103,500)
	2013	\$55,952	\$7,804
	2014		\$2,776
	2015	\$137	
	2017	(\$2,000,000)	
	2018	\$2,031,646	\$118,669
Organized Research Off-Campus	2018	(\$206)	
NREC Off Campus	2010		(\$140,000)
	2011	\$8,758	(\$18,945)
	2012	\$17,443	(\$8,284)
	2013	\$15,639	\$14,692
	2014		\$15
	2018	\$36,190	\$30,695

Silicon Valley Off-Campus	2011	\$58,137
	2012	\$14
	2013	\$78,807
	2014	\$37,231

SEI	2011	\$400,169
	2012	(\$100,651)
	2013	\$2,039
	2014	\$66,590
	2018	\$212,247

<u>SEI Cost of Money</u>	2011	\$17,481
	2012	(\$2,106)
	2013	(\$420)
	2014	\$898
	2018	\$21,107

<u>Fringe Benefits</u>	<u>Liquidated in FY 2020 Rates</u>	
	<u>Fiscal Year</u>	<u>Amount</u>
Full Time Faculty/Staff Domestic	2011	(\$553,464)
	2017	(\$797,660)
	2018	\$1,669,691
Full Time Faculty/Staff International	2011	(\$7,966)
	2012	(\$200)
	2017	\$110,581
	2018	\$82,314
Domestic Part time w/Benefits	2018	\$25,244
Domestic Part time w/o Benefits	2018	\$5,489

E. USE BY OTHER FEDERAL AGENCIES: The rates set forth in Section I are negotiated in accordance with and under the authority set forth in 2 CFR Part 200. Accordingly, such rates shall be applied to the extent provided in such regulations to grants, contracts, and other agreements to which 2 CFR Part 200 applies, subject to any limitations in part A of this section. Copies of this document may be provided by either party to other federal agencies to provide such agencies with documentary notice of this agreement and its terms and conditions.

F. APPLICATION OF INDIRECT COST RATES TO DEPARTMENT OF DEFENSE (DOD) CONTRACTS: In accordance with DFARS 231.303, no limitation may be placed on the reimbursement of otherwise allowable indirect cost incurred by an institution of higher education under a DoD contract awarded on or after November 30, 1993, unless the same limitation is applied uniformly to all other organizations performing similar work. It has been determined by DoD that such limitation is not being uniformly applied. Accordingly, the rates cited (2) of Section I, as explained under the title, "APPLICABLE TO" do not reflect the application of the 26% limitation on administrative indirect costs imposed by 2 CFR Part 200, whereas (1) does so.

G. DFARS WAIVER: Signature of this agreement by the authorized representative of Carnegie Mellon University and the Government acknowledges and affirms the University's request to waive the prohibition contained in DFARS 231.303(1) and the Government's exercise of its discretion contained in DFARS 231.303(2) to waive the prohibition in DFARS 231.303(1) for Organized Research Off Campus, Silicon Valley and Software Engineering Institute rates. The waiver request by Carnegie Mellon University is made to simplify the University's overall management of DoD cost reimbursements under DoD contracts.

Accepted:

FOR CARNEGIE MELLON UNIVERSITY:


Carrie Nelson
Associate Vice President
for Finance and Controller

8/6/19
Date

FOR THE U.S. GOVERNMENT:

WOOD.LINDA.MORGAN.1514688946
A.MORGAN.
1514688946
Digitally signed by
WOOD.LINDA.MORGAN.1514688946
DN: c=US, o=U.S. Government,
ou=DoD, ou=PKI, ou=USN,
cn=WOOD.LINDA.MORGAN.1514688946
Date: 2019.08.08 07:18:52 -04'00'

Linda Morgan Wood
Contracting Officer

8/8/19
Date

For information concerning this agreement contact:

Linda Morgan Wood
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995

Phone: (703) 588-2254
E-mail: linda.m.wood@navy.mil

CERTIFICATION REGARDING LOBBYING

Certification for Contracts, Grants, Loans, and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No Federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of an agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any Federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

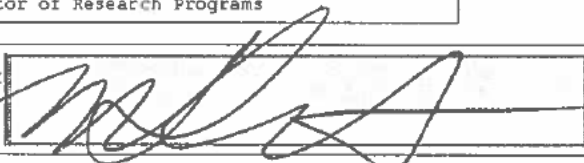
(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers (including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements) and that all subrecipients shall certify and disclose accordingly. This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

Statement for Loan Guarantees and Loan Insurance

The undersigned states, to the best of his or her knowledge and belief, that:

If any funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this commitment providing for the United States to insure or guarantee a loan, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions. Submission of this statement is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required statement shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

* APPLICANT'S ORGANIZATION	
Santa Fe Institute	
* PRINTED NAME AND TITLE OF AUTHORIZED REPRESENTATIVE	
Prefix: <input type="text"/>	* First Name: Susan Middle Name: <input type="text"/>
* Last Name: Carter	Suffix: <input type="text"/>
* Title: Director of Research Programs	
* SIGNATURE: 	* DATE: 02/21/2020