UCLA

REPORT OF THE DATAX STRATEGY TASK FORCE

A TEMPLATE FOR TRANSFORMATION

California NanoSystems Institute

Prepared by Jacob G. Foster and Mark L. Green
Special Assistants to VCRCA for Strategic Initiatives in Data Science

A MESSAGE FROM THE INTERIM DIRECTOR

Dear Colleagues,

As the Interim Director of UCLA DataX, I want to present the Report of the Vice-Chancellor for Research and Creative Activities' Task Force on Strategic Initiatives in Data Science and to thank Professors Mark Green and Jacob Foster for the arduous work of listening widely, learning about our shared commitments across many disciplines at UCLA for using data and knowledge to positively transform the world and synthesizing these consultations into a coherent vision for the future.

What resonates most from their work is the possibility of shaping the DataX Initiative to become an interdisciplinary academic institute that can amplify civic engagement and transform education at UCLA by bringing together a robust and highly collaborative community of humanists, social scientists, artists, legal scholars, researchers in health and medicine, researchers pushing forward the boundaries of data science alongside those who understand its limits, and scientists and engineers from many disciplines who address societal problems and care about the impact of data and technology in society.

This report is a great start to laying the foundation for the initiative, notable for the breadth of its vision and the richness of its detail. I want to thank the faculty advisory committee who have been involved in furthering the vision for DataX, with support from the Office of the VC for Research and Creative Activities, the Academic Senate, UCLA Deans, the EVC/Provost's Office, and the Chancellor's Office.

UCLA is poised to remain a global leader at the intersection of data, knowledge, technology, society and justice as we build academic programs and foster stronger research collaborations in the coming years through DataX. There is much to look forward to in supporting the faculty, students, and community members at UCLA whose important research and teaching are central to the mission of the public university.

Sincerely,

Safiya Umoja Noble, Ph.D.

Professor, Department of Gender Studies

Interim Director, UCLA DataX

ABOUT SAFIYA NOBLE

Dr. Safiya Umoja Noble is an Associate Professor at UCLA in the Departments of Information Studies and African American Studies. She is the author of a best-selling book on racist and sexist algorithmic bias in commercial search engines, entitled Algorithms of Oppression: How Search Engines Reinforce Racism (NYU Press). Dr. Noble is the co-editor of two edited volumes: The Intersectional Internet: Race, Sex, Culture and Class Online and Emotions, Technology & Design. She currently serves as an Associate Editor for the Journal of Critical Library and Information Studies, and is the co-editor of the Commentary & Criticism section of the Journal of Feminist Media Studies. She is a member of several academic journal and advisory boards, including Taboo: The Journal of Culture and Education.

Safiya is the recipient of a Hellman Fellowship and the UCLA Early Career Award. Her academic research focuses on the design of digital media platforms on the internet and their impact on society. Her work is both sociological and interdisciplinary, marking the ways that digital media impacts and intersects with issues of race, gender, culture, and technology. She is regularly quoted for her expertise on issues of algorithmic discrimination and technology bias by national and international press including The Guardian, the BBC, CNN International, USA Today, Wired, Time, and The New York Times, to name a few. She holds a Ph.D. and M.S. in Library & Information Science from the University of Illinois at Urbana-Champaign, and a B.A. in Sociology from California State University, Fresno where she was awarded the Distinguished Alumni Award for 2018.

Education

Ph.D., University of Illinois at Urbana-Champaign

M.S., University of Illinois at Urbana-Champaign

B.A., California State University, Fresno

Research Interests

Search engine ethics, Racial and gender bias in algorithms, Technological redlining, Socio-cultural, economic and ethical implications of information in society, Race, gender and sexuality in information communication technologies, Digital technology and Internet policy development, Privacy and surveillance, Information and/as control, Critical information studies

ABOUT THE AUTHORS

Jacob G. Foster is an Associate Professor of Sociology at the University of California, Los Angeles. He is interested in the social production of collective intelligence, the evolutionary dynamics of ideas, and the co-construction of culture and cognition. His empirical work blends computational methods with qualitative insights from science studies to probe the strategies, dispositions, and social processes that shape the production and persistence of scientific and technological ideas. He uses machine learning to mine the cultural meanings buried in text, and computational methods from macro-evolution to understand the dynamics of cultural populations. Foster also develops formal models of the structure and dynamics of ideas and institutions, with a particular focus on the rich nexus of cognition, culture, and computation. He is currently writing a book on knowledge as an emergent feature of complex adaptive systems. Foster is co-Director of the Diverse Intelligences Summer Institute, a program that aims to build community, collaboration, and creative thinking among early career scholars interested in the study of mind, cognition, and intelligence of diverse forms and formats—from ants and apes to humans and AI. After studying mathematical physics at Oxford as a Rhodes Scholar, Foster received his Ph.D. in Physics (with a specialty in Complexity Science) from the University of Calgary. He then spent three years as a postdoctoral scholar in the Department of Sociology at the University of Chicago before moving to UCLA in 2013. His work has appeared in the Proceedings of the National Academy of Sciences, American Sociological Review, NeurIPS, Science, Phil Trans B, Poetics, Sociological Science, and Social Networks, among other venues. He was an Infosys Member at the Institute for Advanced Study, School of Social Science in 2020-2021.

Mark L. Green is a Distinguished Research Professor in the Department of Mathematics at the University of California, Los Angeles. He received his B.S. from the Massachusetts Institute of Technology and his M.A. and Ph.D. from Princeton University. After teaching at the University of California at Berkeley and MIT, he came to UCLA as an assistant professor in 1975. He was a founding co-director and later Director of the NSF-funded Institute for Pure and Applied Mathematics. Dr. Green's research has taken him into different areas of mathematics: several complex variables, differential geometry, commutative algebra, Hodge theory, and algebraic geometry. He received an Alfred P. Sloan fellowship, was an invited speaker at the International Congress of Mathematicians in Berlin in 1998 and gave the Chern Medal plenary laudation at the International Conference of Mathematicians in Seoul in 2014, and is a Fellow of the American Academy of Arts & Sciences, of the American Association for the Advancement of Science and of the American Mathematical Society. Prof. Green served as vice-chair of the BMSA study on The Mathematical Sciences in 2025, and served on the International Advisory Panel for the Canadian Long Range Planning Study for Mathematics. He was part of the US Delegation to the General Assembly of the International Mathematical Union in Bangalore in 2010 and Chair of the Committee of Visitors for the Division of Mathematical Sciences at NSF in 2013. He has served on the scientific boards of the Institute for Pure and Applied Mathematics, the Centre de Recherches Mathematiques and the Banff International Research Station, and was a Trustee of the American Mathematical Society. He served on the Mathematical Advisory Panel for the exhibition "Man Ray: Human Equations" at the Phillips Collection in Washington, DC and on the Advisory Committee of the Association for Women in Mathematics. He serves on the Board of Governors of the group Transforming Postsecondary Education in Math. He is Chair of the NAS Board on Mathematical Sciences and Analytics and is a National Associate of the National Academies.

TABLE
OF CONTENTS

EXECUTIVE SUMMARY

Executive Summary

The world has been transformed by the increased availability of data, the development of new tools to understand it, and the proliferation of ways to use it in applications and creative endeavors. Data touches every aspect of the university's mission and potentially every department. We call this phenomenon DataX. DataX acknowledges that data is transforming research, scholarship, artistic creation, education, and the opportunities available to our students in fundamental ways. A DataX initiative is urgently needed if UCLA is to maintain its position as the premier public university.

We crafted this strategy from the bottom-up, consulting with over a hundred UCLA faculty and campus leaders, as well as the leaders of data science efforts at peer institutions. We received further input through a series of town halls, 43 white papers submitted by units across campus, and consultation with 5 committees of the Academic Senate. From these consultations, we defined a number of guiding principles. UCLA's DataX Initiative must be: inclusive; pan-disciplinary, embracing fundamental data science, innovative applications, and scholarship on social, ethical, and policy impacts; driven by faculty interest; distributed where possible, but centralized where needed; committed to meeting the urgent demand for DataX education; dedicated to diversity and inclusion from the beginning; and attuned to the strengths of UCLA and the remarkable city where it makes its home, acknowledging its key role as an engine of opportunity.

Critically, we found that UCLA already hosts many points of excellence in DataX research and education. Our strategy aims to enable, empower, enhance, and coordinate that excellence, while building new excellence across campus. UCLA's location in Los Angeles constitutes an unparalleled source of comparative advantage when it comes to community engagement, industry partnership, and opportunities for education and outreach to underrepresented groups. Despite these strengths, our interviews consistently suggested an urgent need for a campus-wide effort to consolidate and build on those strengths.

In regard to research, the most direct formulation of the problem we are trying to address is as follows. The analysis and understanding of large, complex datasets have become a fundamental modality of research for a wide variety of disciplines spanning the UCLA campus; they have become an important resource for creative activities as well. This naturally has profound social, ethical, cultural, and policy ramifications, as does the widespread use of data in governance and the economy. To be a first-rank research university, it is increasingly necessary to be a leader in this area. Peer institutions are ahead of UCLA in taking action to build faculty lines and suitable connective tissue to establish a major presence and a rich research and educational environment. Some of them have been able to attract major gifts.

This is not a problem that can be solved on the scale of individual departments. It is interdisciplinary and campus-level in scale.

Our interviews informed this analysis. The need to build up UCLA's faculty in this area, the need for coordination and team building, the need for entryways and support for researchers interested in incorporating these techniques in their work, the need for an inward and outward-facing research portal—all of this emerged from these interviews.

Because of the breadth of our interviewing process, we became aware of opportunities in three

general areas: (1) Fundamental Data Science, (2) Innovative Applications of Data Science, (3) Ethics, governance, data cultures, and societal impact. We concluded that excellence in all three of these areas is both possible and essential at UCLA. Excellence in each of these areas reinforces and enhances excellence in the other two. This entails an unusual degree of interdisciplinarity and coordination.

The desire for an entity linking together the various research groups at UCLA was manifest in our interviews and in the white papers. This was especially apparent when it came to applications for large, complex grants, which call for interdisciplinary teams, often with representation from all three of the cohorts just mentioned; indeed, many grants in this space *explicitly* require involvement from all three areas. This desire for coordination applied as much to high-level research contacts as to the nitty-gritty details of access to datasets and computing resources. Creating a web portal for DataX research at UCLA, both inward-facing and outward-facing, was perceived as an important step.

While there are many islands of excellence at UCLA, the overall number of DataX researchers is thin on the ground, especially for a university of our size. This is especially apparent in some of the white papers we received from campus units, which expressed the hope that DataX could help them in finding collaborators; such collaborators simply do not exist on campus in sufficient numbers to meet the demand. It is also confirmed by the pace of hiring in data science and closely allied fields like data ethics or critical data studies at peer institutions, which substantially outpace hiring at UCLA to date.

By contrast to those whose research and creative activities are steeped in DataX, researchers and creators who use data science less intensively—and those who need access to technical expertise in data science or a user-friendly way to begin using data science techniques—expressed the need for some form of support network. It was repeatedly emphasized that the different disciplinary cultures across campus precluded a centralized, "one-stop-shop" for such research support; instead, our interviewees wanted support by people who speak the language of their discipline and are aware of the distinctive challenges presented by the sort of data they typically use.

In regard to education, it also became clear that existing, localized efforts would benefit from some coordination. There is a need for a new entry point that will provide a broad range of students a welcoming path to data fluency, including both technical aspects and social, ethical, and critical considerations. The simplest way to frame the educational problem is to observe that the career landscape for our students has changed dramatically in recent years. It continues to change at a rapid pace, driven by the widespread availability and use of large quantities of data.

This problem, too, is not amenable to being solved by individual departments. It is fundamentally interdisciplinary, and campus-wide.

As we interviewed faculty from a wide range of departments, our conversations informed us about emerging career opportunities for university graduates whose training mixes some level of data fluency with a disciplinary expertise. There are a number of promising initiatives across campus to offer students the opportunity to acquire this mixture of expertise. It became apparent to us in the course of our interviews that there is a need for new courses, majors and minors—and that UCLA is far from having offerings at a scale to meet this sea change in the opportunities available to students when they graduate.

There is a need for a portal for students trying to navigate a somewhat bewildering buffet of courses, majors and minors. There is a need for students to be able to certify their knowledge of data science, as

well as social, ethical, and critical studies of data, at different levels of depth, and for such certificates to be uniform across departments.

More generally, it would be helpful to have a body working actively to enable this major transformation of the curriculum, providing guidance (to departments that want it) about how proposed programs fit into the tapestry of campus offerings, and to provide educational opportunities that do not fit neatly into one department or division. We note that many colleagues mentioned how valuable it would be to have some campus-wide coordination on this point; in other words, this was not something we decided on our own, but rather a request that we heard repeatedly.

A recent report of the Academic Senate stressed the critical role of interdisciplinary education, and emphasized the need for suitable structures (e.g., Centers for Interdisciplinary Instruction) to ensure a reliable stream of funding. DataX, with its faculty lines and educational efforts, addresses a critical nexus of interdisciplinary education and can act as an exemplar of such a structure.

How do we propose to address these multifaceted opportunities and challenges?

We propose the creation of a DataX Institute as a focal point and catalyst for research and educational initiatives across campus, paired with a distributed DataX Support Network to build capacity in using data for all who are interested in exploring this space. The creation of these entities should run in parallel with a major hiring effort, recruiting 60 new faculty over a 10 year period. We also see an urgent need for a campus-wide effort to develop courses, certificates, minors, and majors in DataX, appropriate for a wide range of student backgrounds and interests. It is clear that we are nowhere near meeting potential demand; our emphasis must be on developing high quality, externally legible, and internally navigable educational opportunities.

The DataX Institute will focus on advancing cutting-edge data science and applications, understanding its ethical, social and policy implications and its possibilities for enhancing creative expression, and providing leadership for DataX research and education across campus. It will foster interdisciplinary working groups, offer onramps to engage new researchers, launch seminars, facilitate the acquisition of important datasets, and work to forge research partnerships with industry and government agencies. It will encourage and enable educational innovation and seek to rationalize UCLA's offerings in the DataX area. We propose hiring 6 regular faculty each year, shared 50-50 between an existing department and the DataX Institute, up to an eventual size of 60 faculty. After 10 years, the full FTE would revert to the partner department, allowing continuing renewal of the DataX faculty in this rapidly evolving field. Faculty would come equally from the three general areas of fundamental data science, innovative applications, and ethics, governance, data cultures and societal implications. We propose a gradual rollout of 30 three-year interdisciplinary postdoctoral positions associated to DataX. Existing UCLA faculty could become associated with the Institute as Faculty Fellows. The Institute will also host industry and national lab fellows, a staff, and a leadership team. The Institute will be housed in a dedicated space with room for its leadership, staff, faculty and postdocs; for faculty and students incubating new working groups or transitioning to a more active engagement with DataX; and for seminars and other intellectual activities. The Institute will be led by a Director and Associate Directors for Research, Education and Support. The leadership will be supported by an Executive Committee, a Stakeholder Board, an External Advisory Board, and a staff. The Institute will actively engage the campus community through outreach and advocacy efforts, including a program each quarter with a distinguished invited speaker of broad interest, an annual DataX master class, and through the creation of a web portal to DataX research activities on

campus.

The DataX Support Network will focus on building capacity in using data across campus. Staff will be embedded in nodes situated in Schools and Dean-level units. Staff will be fluent in the language and culture of the disciplines comprising the Division or School that they serve, and will aid in data acquisition, curation, manipulation, analysis, sharing, and management. They will facilitate faculty and student use of existing data science techniques and will offer mini-courses where needed. The nodes of the Support Network will be integrated into a larger community through the Institute, which will coordinate Network activities.

New educational initiatives will be a major part of DataX, with diversity and inclusion a centerpiece. A number of new programs are being rolled out across campus, but there is a need to integrate them and to take them to a larger scale. We envisage new DataX cluster courses designed to engage students with a wide variety of interests and backgrounds; new data science offerings in individual departments; data science add-ons to existing courses; data science certifications available to students; new majors and minors; vertically integrated research experiences; summer research experiences; internships with local government and NGO's; and capstone courses. High quality instructional datasets will need to be obtained or created, and then maintained and curated. A DataX portal for students that delineates possible pathways through the university's DataX course offerings will be created, coordinated with a portal connecting them to research opportunities.

Insights from peer institutions support our strategy. Colleagues agreed that data science was an existential issue for all modern research universities. They also emphasized the essential role of people in DataX efforts; most peer institutions coupled faculty hiring with some combination of postdoctoral fellows, research scientists, and support staff. Support services were generally available gratis. The organizational model varied considerably, with campus-wide Institutes being the most common mode. Building community among faculty is a particular challenge, and peer institutions pursued a range of activities to stimulate engagement with data science. No single educational model has emerged; instead, colleagues described offerings with variable duration, at a range of technical levels, from workshops to PhD programs. The most serious challenges faced by peer institutions stemmed from the "capture" of data science efforts by particular disciplines or departments, along with persistent difficulty in attracting a diverse group of faculty and students.

The resources required for this transformation are considerable, and it is essential to plan for financial sustainability from the beginning. It is also essential to build in a mechanism to ensure that the intellectual footprint of the Institute can evolve over time. For these reasons, it makes sense to build up DataX in stages, rolling out funding over time. This allows for growth that builds upon a solid foundation but is nimble enough to incorporate lessons learned. We describe some of the resources required in detail; consistent with the findings from peer institutions, these focus on people and space.

The fundamental principle governing implementation is to move forward with creating a transitional entity, the DataX Homeworld, as expeditiously as possible. The Homeworld will serve as an incubator to create the lineaments of the eventual DataX Institute and DataX Support Network. The present fiscal constraints do not diminish the importance of immediate action. We therefore describe a structure in which resources will begin to flow immediately but can be phased in over a period of a few years. We strongly recommend launching the DataX Homeworld in time for the coming academic year, with elements of its research, educational, support and outreach activities rolling out as quickly as possible.

INTRODUCTION
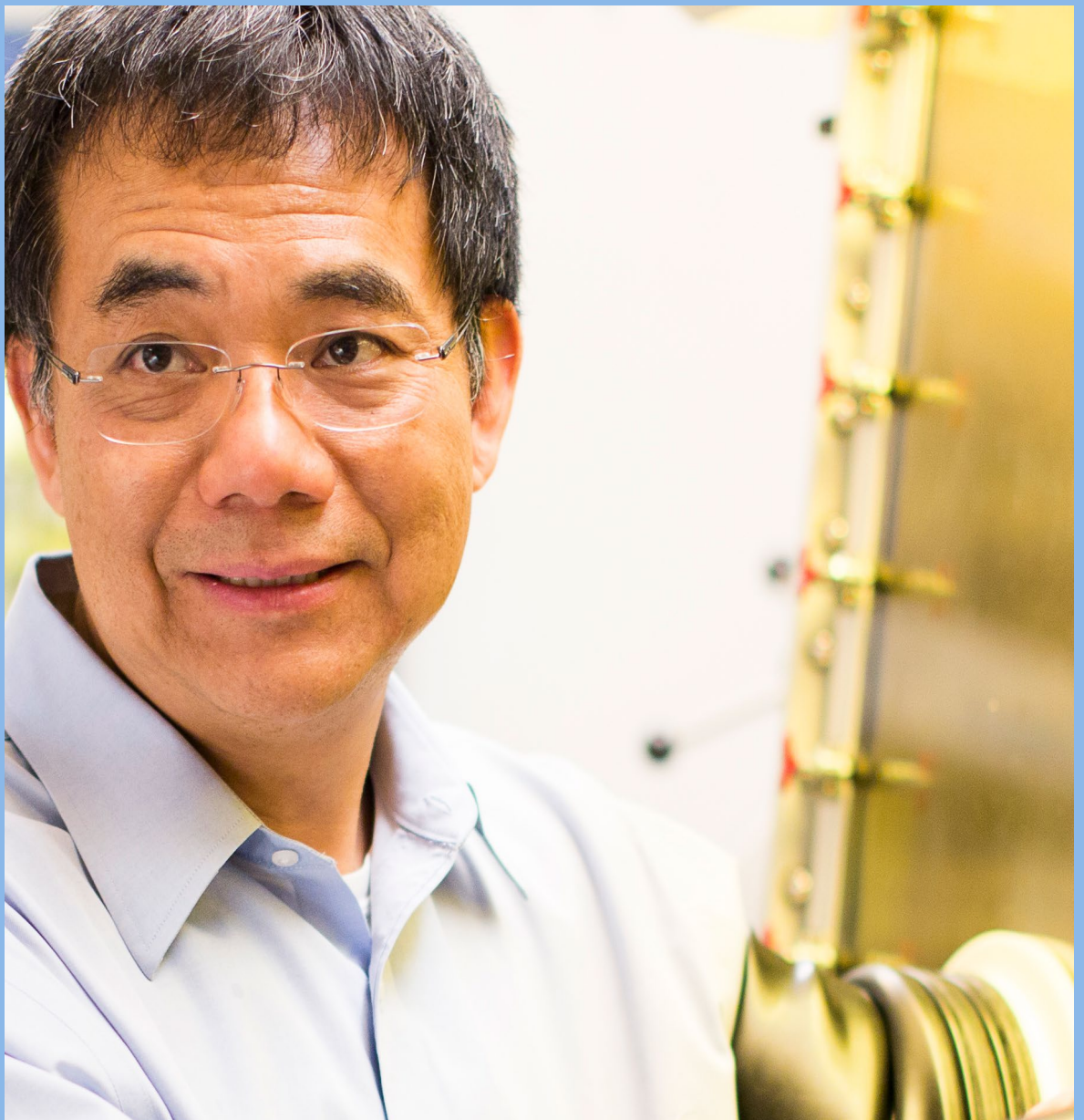
Introduction

This is a report about **transformation**.

The intellectual landscape has been reshaped by recent developments in the availability and analysis of data at scale, as well as the proliferation of ways it is used in applications and creative endeavors. These same developments promise to profoundly alter our society and the world our students will enter. UCLA must change in response, if it is to continue to meet the needs of our students, our city, our state and our nation.

We have adopted the name DataX for this phenomenon. Data touches every aspect of the university's mission and potentially every department. DataX acknowledges that data is transforming research, scholarship, artistic creation, education, and the opportunities available to our students in fundamental ways.

Four factors lead to the current significance of DataX: Plentiful and easily accessed computational power; the collection of staggering amounts of digital data and the increasing importance of open data in research; the rise of the internet, which allows data to be shared across the globe; and remarkable advances in algorithms for analyzing data. Each of these factors has magnified the others, pushing forward the frontiers of possibility in artificial intelligence, personalized medicine, social data science, and countless other fields. These advances pose a suite of unprecedented social problems, from bias and privacy to transparency and governance.

DataX has many faces. Each is marked by challenges and opportunities. Fundamental data science is in itself an exciting field of research. But DataX embraces tools used in research across campus; it is key to applications in almost every discipline. Given the breadth of applications within and beyond the academy, it is essential to study the emerging impact of DataX on society.

DataX is an engine of opportunity for our students—once the barriers to access are breached. Being a data scientist is a growing and attractive career. But DataX also cultivates a knowledge base used at various levels of sophistication in an ever-expanding spectrum of careers, where it is often combined with deep disciplinary expertise.

UCLA is fortunate that our campus is already home to a wealth of outstanding research and creative endeavors in the DataX space. To take this to the next level, we need a campus-wide entity that draws these endeavors together and systematically enables opportunities for collaboration. White papers submitted by units across campus demonstrate that new connections between groups can supercharge research and enable UCLA to make breakthrough contributions to emerging interdisciplinary areas of research and creativity.

The emergence of materials informatics, i.e. materials data infrastructures combined with data-driven materials discovery techniques, holds the promise of revolutionizing the process of discovery and characterization of novel materials. **—Materials Science & Engineering WP**

Over the past six years, JCCC researchers have made 14 discoveries that led to US FDA therapeutic approvals, many driven by Data Science-intensive research...Data science activities at JCCC have been long standing...UCLA Health was recently awarded Level 7 HIMSS analytics designation, making it one of the first organizations in the world to reach this level of data-intensiveness in all aspects of its clinical and cancer care… DataX can serve as a platform to foster multi-disciplinary collaborations that use data science to identify candidate drug targets for laboratory research using models. **—Jonsson Comprehensive Cancer Center WP**

We also underline the opportunity, and the obligation, to center the voices and perspectives of diverse groups in UCLA's DataX scholarship. These groups are often profoundly affected by new data-intensive technologies, but do not play enough of a role in creating or critiquing them. UCLA is already making important contributions here. By working in partnership with important efforts like *Rising to the Challenge* and UCLA's commitment to become a Hispanic-Serving Institution, UCLA's DataX Initiative can be a world leader in inclusive scholarship on data science and its social impacts.

Likewise, our campus has bubbled with creativity in developing new educational programs with a DataX component. To make UCLA a national leader in DataX education, we need both greater integration and a larger scope and scale of offerings. This educational ecology needs well-articulated and varied entry points for new students, alongside clear pathways to majors, minors and certifications—while taking account of the range of student interests, backgrounds, and career plans. The current variety of offerings must become a well-designed tapestry of interconnected educational experiments. At the center of this tapestry must be creative offerings designed to engage and include students from diverse backgrounds and life experiences in DataX education and research—particularly students from groups that are currently underrepresented in this cutting-edge sector of the academy and economy.

Transformational and inclusive change is urgently needed, if we are to build on the excellence of our faculty and existing programs to create a DataX strategy unique to UCLA—one that will allow us to leapfrog our peers to a position of leadership. Peer institutions have been rolling out programs in data science and related fields at a breakneck pace; some have established entire Schools in this area. They have also raised substantial funds to support these programs. This is a challenge: We are behind, in some cases significantly. But it is also a blessing: We can learn from what worked and avoid what didn't.

It is not our place nor within our expertise to attempt to prioritize all of the needs to be addressed and opportunities to be pursued for the campus as a whole. This is clearly not the only opportunity that UCLA should be pursuing. Two questions we hope this report will answer are: Is this something that UCLA needs to do, with considerable urgency? Is the status quo disadvantaging our students and faculty? In both cases, we have come to feel that the answer is a resounding yes.

UCLA is currently the #1 public university in the US. The proposed transformation is essential to maintaining this ranking over the next 10 years.

BACKGROUND & PROCESS
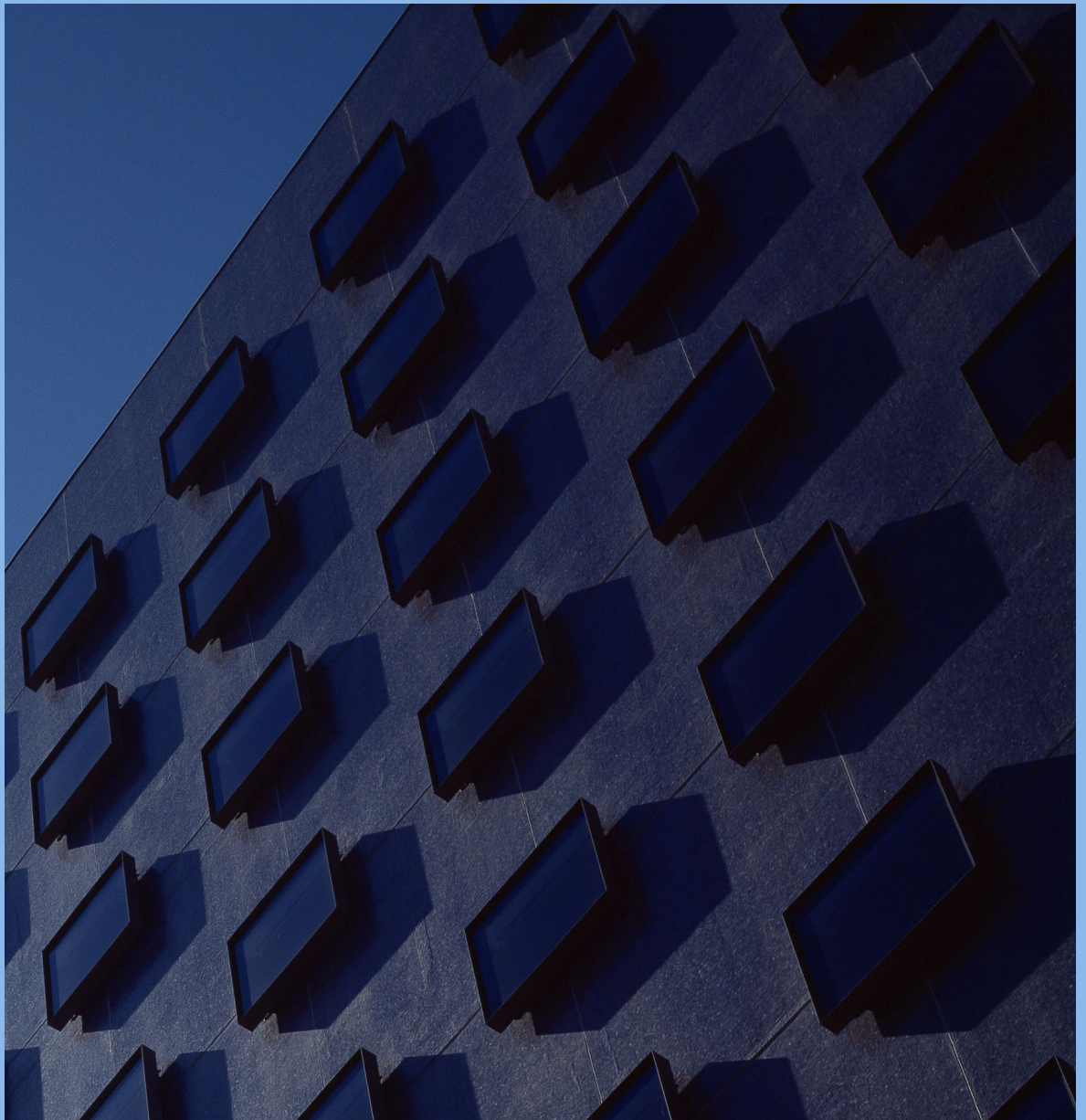
Background and Process

To craft a strategy for UCLA's DataX Initiative, Vice Chancellor for Research and Creative Activities Roger Wakimoto created a two-person task force (us) with a single representative from North and South Campus. Shortly after she was appointed, Executive Vice Chancellor and Provost Emily Carter gave her support to this project, underlining its importance for the future of UCLA.

As the North Campus representative, VCR Wakimoto appointed **Jacob G. Foster**, then assistant (now associate) professor of sociology. Originally trained as a statistical physicist, Professor Foster is a computational sociologist of science and culture who uses data science techniques (e.g., machine learning and network science) to study the social production of collective intelligence, the relationship between culture and cognition, and the evolutionary dynamics of ideas. He has been involved in a number of related efforts on the UCLA campus, including service on the Executive and Leadership Committees of the Institute for Digital Research and Education and the Big Data Initiative of the Division of Social Sciences. He has also led efforts to create multiple transdisciplinary communities, including the Metaknowledge Research Network and the Diverse Intelligences Summer Institute. He is a Rhodes Scholar and 2020 -2021 Infosys Member at the Institute for Advanced Study.

As the South Campus representative, VCR Wakimoto appointed **Mark L. Green**, Distinguished Research Professor of Mathematics. Professor Green is a pure mathematician whose work spans algebraic geometry, commutative algebra, Hodge theory, differential geometry, and complex analysis. He was a Founding Director of UCLA's Institute for Pure and Applied Mathematics, an NSF-funded national mathematics institute focused on fostering interdisciplinary research. He served as Vice-Chair of the committee that wrote the National Academies' forward-looking study *The Mathematical Sciences in 2025* and is Chair of the National Academies Board on Mathematical Sciences and Analytics; he also serves on the Board of Governors of Transforming Postsecondary Education in Math. He is a Fellow of the American Academy of Arts and Sciences, of the American Mathematical Society, and of the American Association for the Advancement of Science, and was a member of the National Academies Roundtable on Data Science Education.

The appointment of such a small committee was predicated on a number of assumptions: That we would consult widely and familiarize ourselves with what is already going on at UCLA and at peer institutions; that we would, by this process, surface a broad range of opportunities; that we would come to the task with open minds; and that a small committee was more likely to be nimble than a large one.

Once constituted, we held numerous meetings across campus to gather information, insight, and ideas: around 50 interviews (lasting from one to two hours) with more than a hundred individuals; meetings with multiple deans; presentations to the Faculty Senate Council on Research, and the Deans' Council; and consultations with EVC/Provosts Scott Waugh and Emily Carter .[1]

Later in the process, we consulted with Founders, Directors, and key participants in data science efforts at peer universities: the University of Wisconsin-Madison, the Santa Fe Institute, the University of Michigan-

---

1   The sequence described here broadly follows the sequence of events (meetings with campus stakeholders, meetings with campus leaders, meetings with leaders at peer institutions, etc.). That said, some meetings occurred outside this general pattern.

Ann Arbor, Caltech, NYU, the University of Chicago, UC Berkeley, and the University of Washington-Seattle. At this point, we had largely completed our extensive consultation on campus and developed the strategy recommended in this report. We were thus pleased to learn from these consultations that our proposal was consistent with what had been done at peer institutions, as well as a prominent report about establishing data science initiatives (funded jointly by the Moore and Sloan foundations), while including a number of innovations that respond to UCLA's distinct challenges, advantages, and opportunities. We offer a detailed summary of the feedback from these consultations later in the report (p. 46).

To enable further consultation with the campus community, we held four town halls in Spring 2020. At each town hall, we presented the main recommendations of the task force, leaving ample time for Q&A. Videos of the presentation and Q&A sessions were posted to the VCR's website and remain available there for viewing (https://www3.research.ucla.edu/datax).

After the completion of the town halls, VCR Wakimoto and former EVCP Carter issued an open call for white papers from Departments, Divisions, and other campus units devoted to research, education, or service. We received 43 white papers, in which responding units describe their current DataX-adjacent activities and intended participation in the DataX Initiative. These documents are attached as an appendix to this report. The white papers helped to deepen our knowledge of what is going on across campus and to further refine our thinking. In many cases, they suggested new activities or modes of organization that we incorporated into the report. We expect that the white papers will also prove valuable in the implementation phase of the Initiative; they highlight a wide range of opportunities in research and education, and suggest the catalytic benefits of bringing together interdisciplinary teams and engaging departments, institutes and centers in cross-campus collaboration. We have chosen quotes from each white paper that collectively illustrate the rich tapestry of efforts and possibilities in this space; each quote is not necessarily representative of the white paper from which it is drawn.

Over the course of writing this report, we received valuable input from 5 committees of the Academic Senate: the Executive Board, Committee on Research, Graduate Council, Undergraduate Council and the Council on Planning and Budget.

PHILOSOPHY & APPROACH

UCLA Alumni Refik Anadol's "Quantum Memories" utilizes computational research data and algorithms

Philosophy and Approach

Research and creative activities in DataX are rapidly evolving, as are the associated career opportunities for students. For that reason, we have tried in this report to steer a middle course between generality and specificity in describing what the opportunities are; in many cases, we drew on the white papers for illustrative examples. The most important thread is the availability of data (broadly construed) at scales or complexities that require the use of computation for its management, analysis, and interpretation; a critical understanding of the ways that data are socially produced and constructed; and an analysis and evaluation of the impact of this development on society and individual lives.

We did not start out with a list of problems to be solved. Instead, our approach was to talk to as many researchers and educators at UCLA as we could manage in Data Science and allied fields (e.g., Digital Humanities, Critical Data Studies, Information Studies, etc.). We spoke with more than 100 ladder faculty. We sought to learn their perspectives—what was holding them back and what would empower them to move forward. Insofar as the report references or addresses problems, those problems emerged from the faculty.

One possible outcome of this process might have been 100 different, mutually contradictory answers. Luckily for us, that did not happen. Instead, **there was broad agreement on what the problems are and what kind of structures would be empowering.**

There are a number of principles that guided us as we formulated a strategy for DataX. We largely derived these principles from our consultations with campus stakeholders. We believe that they reflect principles and values broadly shared across the campus.

Our overriding principle is one of **inclusion**. It is essential that DataX be open to all aspects of science, scholarship, and creative activity on our campus. Only this commitment does justice to the transformational potential of DataX—and our campus values. This commitment has consequences: for inclusion to be meaningful, and for DataX to be truly open, it must provide **pathways** for interested faculty and students at all technical levels. Integral to this commitment is **building a community of researchers and creators with diverse identities, experiences, and backgrounds, with a particular focus on inclusion of hitherto underrepresented groups**. This has implications for hiring, research, teaching, outreach—the full scope of DataX activities.

Inclusion also reaches beyond the campus. As part of UCLA's mission as a public university, we support a campus commitment to developing an **open data policy** (as many funding agencies and foundations have emphasized) and **innovative social and technical infrastructure for sharing data** as widely as possible.

We are also committed to a **broad understanding of DataX**. This implies an approach that balances three important strands of DataX scholarship: **fundamental research** in data science; innovative **applications** and creative activities; and work on the **social, ethical, and policy** dimensions of DataX.

Within fundamental data science, we were drawn to a **strategy that embraces good ideas from all disciplines**, and rejects the idea that data science "belongs" to a particular one. Not only does this build on UCLA's unusual excellence across applied mathematics, statistics, computer science, and

electrical and computer engineering; it is also consistent with advice we received from our contacts at peer institutions, who encouraged us to **give equal billing to those with innovative approaches and applications**, whose home might be in a variety of departments. The increasing public recognition of the broad societal implications of data science[2] leads us to insist on an **equal and critical role for scholarship on social, ethical, and policy impacts**.

Taken together, these positions argue against an approach that is dominated by one department or division, as has been pursued at some peer universities. They also argue against selecting or recruiting someone with a strong and overarching research vision to lead DataX; instead, **DataX requires leadership that will seek out a variety of opportunities and empower interdisciplinary groups of faculty**. Again, the importance of organic and self-organizing scholarship was repeatedly emphasized in our external consultation. It is also consistent with UCLA's long tradition of faculty governance.

Recognizing the staggering diversity of interests, methods, and background on our campus, we favored a **distributed and bottom-up approach that would impact the campus broadly**, while honoring its intellectual diversity. Our strategy must **enhance research opportunities for those who put DataX at the center of their work, as well as for those who are keen to explore the possibilities of DataX in their own research** but need assistance from people with expertise in DataX—people who also speak the language of their discipline.

At the same time, we felt a **keen need to bring people together across campus** around the challenges and opportunities of DataX. Our consultations made very clear that UCLA has a compelling opportunity for community-building and coordination in DataX research and education. This will allow us to make the most of the excellence that already exists on campus, and build for the future in a way that acknowledges the inherent inter- or trans-disciplinary nature of DataX. Our colleagues view DataX as having the potential to unify the campus in hitherto unexploited ways. We agree, and this is one reason why our recommendations completely avoid the idea of setting up a department, division, or school with sole responsibility for DataX.

At its most basic, our proposal is a template for building a community of researchers, educators and students. Building such a community must be done step-by-step; there is no simple formula. We have each had extensive experience in building highly interdisciplinary research and educational communities—Foster at the Diverse Intelligences Summer Institutes (DISI), Green at the Institute for Pure and Applied Mathematics (IPAM). [For example, IPAM is the only national math institute to hold a program sponsored by the National Endowment for the Humanities; DISI has successfully established respectful dialogue between sociocultural anthropologists and AI researchers, to pick one of many intellectual divides bridged.]

Intellectual divides between areas are real and not easily overcome. What is clear is that they will not be overcome without an explicit effort to do so. UCLA has a number of successful interdisciplinary institutions at different scales. We encountered some of these institutions—as well as the intellectual divides between disciplines—in our wide-ranging consultation. We had interviews with both faculty and Deans in the Schools of Business, Engineering, Law, Medicine, Music, and Theater, Film and Television;

---

2   As one important piece of evidence, we can point to the recent policy at major AI conferences like NeurIPS that requires every submission to include "a statement on the broader impact their research could have on society." https://www.nature.com/articles/d41586-020-03611-8

we also met with faculty and Deans across the College. The 43 White Papers we received are full of explicit ways that the originating units would engage with and benefit from DataX. Our interviews revealed a thirst for connection, an openness to connecting, and a perception by both faculty and administration of the benefits of connecting.

This an intellectually vibrant area in part because each type of data has its own unique character and poses its own special set of challenges. Addressing the particularities of the data at hand is a central part of any successful application; so is understanding the limitations of any given method or approach (we emphasize both of these points in our own teaching in this space). Successful applications always involve a dialogue between fundamental data scientists and experts in the application discipline, as well as experts who have thought about the disciplinary and societal context from which the data emerged and the ethical, policy and societal implications. Such conversations are iterative, gradually homing in on a possible solution. The role of DataX is to provide a welcoming environment where such conversations can take place and be facilitated; to proactively bring people together.

Large mergers, such as the ATT-DTV merger, often require regulators to analyze large data sets (in this case, involving consumer cellphone usage data) to understand the competitive relationship of firms. Similarly, financial regulators regularly process terabytes of data in building regulatory actions aimed at protecting consumers from financial fraud. Amazon now employs over 250 PhD economists across all their business areas. **—Economics Dept WP**

The need to consider the provenance, context and assumptions that went into a corpus of data constitutes an additional argument for the essential role of scholarship on the social and cultural dimensions of data in the DataX Initiative. Such scholarship plays a pivotal role in assuring that fundamental data scientists and developers of innovative applications do not forget the epistemological issues, as well as the social, cultural, and economic contexts, that are essential to practicing their craft responsibly.

We approached issues of education in DataX with great urgency. It is essential that we offer our students courses, research experiences, majors, and credentials that will serve them well in their lives and careers. This is true whether they are undergraduate or graduate students; whether they study in the College or in the professional schools. It quickly became apparent from our interviews that **a suitable level of expertise in data science, combined with a deep disciplinary knowledge, was a highly desirable skill-set across multiple domains,** including some we had never imagined. The **level of expertise in data science needed was quite variable;** one size does not fit all. There are a number of highly promising initiatives planned and even underway in both North and South Campus.

It is apparent, however, that we are **far from meeting the demand for DataX education**. This implies that we are not in a regime in which units are "competing" for DataX students, or in which we need to hunt for efficiencies. Instead, our overarching principle for DataX education is **experimentation married to coordination**. New approaches and opportunities should be cultivated, in the context of **an overall plan to align these initiatives with one another** into coherent pathways that students can easily follow.

DataX is very much in its early stages on our campus, rapidly expanding in both research and educational opportunities. This allows us **to build diversity and inclusion into UCLA's DataX strategy from the beginning.** Some of our peer institutions have stumbled on this front. Complacency will lead to disparate

participation, given diversity challenges in data science and related fields. A number of recommendations flow from this commitment; for example:

- giving equal priority to fundamental data science, innovative applications, and social, ethical, and policy dimensions in research programming and faculty recruitment;
- creating (as one part of UCLA's educational offerings in this space) a suite of courses which combine DataX with topics in which there is strong student interest, such as social justice, sustainability, and health care, and which are accessible to students independent of the mathematical and technical background they start with; and
- ensuring that DataX features prominently in UCLA outreach and recruitment efforts.

Recent events—from widespread social movements for racial justice to the disparate effects of the COVID-19 pandemic—have only reinforced the urgency of this commitment. UCLA can rightfully be proud of many remarkable points of excellence in the DataX space, in research, artistic creation, and education. We began our process aware of this excellence. Through our interviews, however, we came to realize that there were many more great things going on—of which neither of us was aware.

From these observations we distilled several further principles. First and foremost, our DataX strategy should **enable, empower, enhance, and coordinate** what is already happening; it shouldn't interfere with or obstruct faculty initiative and excellence. Second, our strategy should be **attuned to the specific opportunities and excellences of UCLA**. UCLA is in the unusual situation of having a great many nationally and internationally recognized departments and research groups; it is also home to world-class professional schools, all a short walk from each other. Third, we wanted a strategy that would **take advantage of UCLA's status as an urban campus.** We are located in one of the most remarkable cities in the world. We have unique potential partners thanks to this location, like the entertainment industry. We can also pursue a panoply of successful projects for social good in partnership with state and local government, such as "Sustainable LA." Finally, we must ground our strategy in the knowledge that **UCLA is an engine of opportunity** for the people of an exceptionally diverse city and state.

In partnership with Center X in SEIS, the department led an NSF-funded program to use data science to improve high school STEM learning. The Mobilize Introduction to Data Science (IDS) curriculum was one product, and to date it has taught almost 15,000 high school students a college-preparatory version of data science in schools that serve primarily low-income students or students from underrepresented groups. IDS is taught in 15 school districts in California, and soon will be taught throughout the State of Oregon. **—Statistics Dept WP**

In doing this work, we were conscious of UCLA's status as America's #1 public university. From the outset, it was clear to us that anything short of a strategy directed at a campus-wide transformation would fail to meet the moment. Our strategy aims to **seize upon UCLA's many strengths and to leapfrog peer institutions**, arriving at a level of excellence in DataX consonant with UCLA's stature.

LANDSCAPE

"L'occhio del cielo" (eye of the sky) - Located in the UCLA Humanities Courtyard area

Landscape

UCLA is exceptionally fortunate to have many faculty, programs, and departments that are already involved in the DataX space. Our strategy is designed to take advantage of these **many points of excellence**. Our goal is to enable, empower, enhance, and coordinate them, while building new points of excellence across the campus.

Growing demand for data-oriented programs with an emphasis on social science is indicated by the recent emergence of degree programs in areas such as Social Data Science (MSc, University of Oxford; MSc, University of Copenhagen); Applied Social Data Science (London School of Economics); Computational Social Science (MA, Univ of Chicago); and in interdisciplinary Data Science (Duke University). Given the concentration of faculty expertise already at UCLA, our campus is well-positioned to become a leading institution providing training in this area. **—MIDAS WP**

UCLA has unusual strength across the core data science disciplines of Mathematics, Statistics, Computer Science and Electrical & Computer Engineering. UCLA also hosts exceptional strength in innovative applications, across the College and the professional schools. And UCLA is a leader in research and teaching activities in social and ethical aspects of data science. UCLA is also home to interdisciplinary entities with significant DataX footprints. Furthermore, there are opportunities to use the tools of DataX to improve campus operations; for example, in transportation or in the improvement of student outcomes.

All faculty in the Department of Statistics conduct research in the cutting edge of Data Science. The Department of Mathematics has a world-renowned Applied Mathematics group whose members have made contributions of historic significance to Data Science. Data Theory, understanding the mathematical and statistical reasons why methods in DataX work, or not, is an essential component of the DataX revolution...At the heart of our approach is the "reciprocal design" philosophy, in which the development of theory and algorithms lives in close feedback with real-world deployment. **—Data Theory Program WP**

In our interviews, we encountered many **exemplars of what the future might look like**. We list some of them here, expanding on the brief summary above. This is by no means an exhaustive list; the goal is to highlight the **variety and excellence** of campus efforts in DataX research and education.

Research and Impact in DataX

- Fundamental Contributions to Data Science by UCLA Faculty: These include fundamental work on Causality by Turing Award winner Judea Pearl (https://amturing.acm.org/vp/pearl_2658896.cfm); the development of Compressed Sensing by Terence Tao and others (https://www.theage.com.au/lifestyle/terence-tao-the-mozart-of-maths-20150216-13fwcv.html); work by Stanley Osher that led to his winning the Gauss Prize, the top international prize in applied mathematics (http://www.ipam.ucla.edu/news/osher-receives-carl-friedrich-gauss-prize/); fundamental algorithms developed by Andrea Bertozzi (https://www.math.ucla.edu/~bertozzi/); and key developments in structural equation modeling by Peter Bentler (https://www.psych.ucla.edu/news/peter-bentler-receive-lifetime-achievement-award)

- The newly launched Social Science Big Data Initiative, which will support faculty research, transdisciplinary working groups, and data science teaching (including a social data science minor), with a focus on diversity and inclusion. https://bigdatasocialscience.ucla.edu/
- Groundbreaking projects like Million Dollar Hoods, led by historian and MacArthur Fellow Kelly Lytle Hernández, which trains dozens of undergraduates (predominantly from racial/ethnic minority groups) in data science each year through mapping the fiscal and human costs of mass incarceration. Lytle Hernández also recently received a $3.65M grant from the Mellon Foundation for the project "Archiving the Age of Mass Incarceration," which will preserve and analyze four decades of data from the LAPD. https://milliondollarhoods.pre.ss.ucla.edu/
- The Digital Archaeology Lab, which underlines that data-intensive scholarship can involve extremely detailed digital representations of specific objects as well as more traditional "Big Data." https://dal.ucla.edu/
- The Center for Critical Internet Inquiry (C2i2), co-directed by MacArthur Fellow Safiya Umoja Noble and Sarah Roberts, which recently received $2.9M in funding to investigate the social impact of digital technology on communities and the broader public good. https://www.c2i2.ucla.edu/
- UCLA Institute for Technology, Law and Policy at the UCLA School of Law (https://law.ucla.edu/academics/centers/institute-technology-law-policy), directed by John Villasenor
- The Program on Understanding Law, Science and Evidence (https://law.ucla.edu/academics/centers/pulse-program-understanding-law-science-evidence), including the AI PULSE program (https://aipulse.org/) headed by Edward Parson.
- The Future Storytelling Summer Institute in Theater, Film and Television, financially supported by Amazon Web Services, which in 2019 considered how media-rich datasets about Los Angeles could be used to train machine learning to curate media experiences in an Olympics pavilion for 2028. https://remap.ucla.edu/future-storytelling-studio/
- A constellation of outstanding data-intensive initiatives in UCLA's highly ranked medical school, including the Department of Computational Medicine (https://compmed.ucla.edu/), the Institute for Precision Health (https://www.uclahealth.org/precision-health/), and the Jonsson Comprehensive Cancer Center (https://cancer.ucla.edu/).
- The Institute for Quantitative and Computational Biosciences, which re-imagines biosciences research, training, and education to meet the shifting emphasis between bench science and computation. https://qcb.ucla.edu/
- The Institute for Pure and Applied Mathematics (IPAM), which has hosted pivotal programs in many areas of data science over two decades. Examples include Functional Genomics; Multiscale Geometry and Analysis in High Dimensions, during which a fundamental new approach to collecting data—compressed sensing—was developed; a pioneering Summer School in Deep Learning; Culture Analytics; and Networks in the Humanities (perhaps the only program at a national math institute sponsored by the National Endowment for the Humanities). https://www.ipam.ucla.edu/
- The Depression Grand Challenge, headed by Nelson Freimer and funded by a major grant from Apple, which has created an app that allows students to self-report indicators of their level of anxiety at frequent intervals, providing data for the study. https://newsroom.ucla.edu/releases/ucla-launches-major-mental-health-study-to-discover-insights-about-depression
- The Institute of the Environment and Sustainability, which is engaged in many data-centric activities; the "Sustainable LA" Grand Challenge envisions ambitious data-intensive research. https://grandchallenges.ucla.edu/sustainable-la/
- Several UCLA faculty serve on the LA County Department of Health Services COVID-19 Predictive Modeling Team, currently focused on healthcare demand at the epicenter of California's pandemic.

Teaching in DataX

- The Data Theory Major, the nation's first undergraduate degree focused on the mathematical and statistical concepts that underlie the data revolution. https://datatheory.ucla.edu/
- An NSF Research Traineeship program in Modeling and Understanding Human Behavior, which will train more than 100 PhD students in the social, biological, mathematical, and computational sciences to perform data-intensive scholarship on human behavior, https://www.math.ucla.edu/~bertozzi/NRT/index.html
- A new PhD program in Communications, reflecting a profound shift in that discipline toward data-intensive scholarship. https://comm.ucla.edu/graduate/research-areas/
- UCLA Anderson's Master of Science in Business Analytics, currently ranked #2 in the world, with 100% job placement six months after graduation. https://www.anderson.ucla.edu/degrees/master-of-science-in-business-analytics
- A thriving Program in Digital Humanities (bringing computational technologies to study culture, history, and society). https://dh.ucla.edu/
- A new, Mellon-funded program in Social Justice and Critical Data Studies, which will launch innovative cluster courses and a Master in Data and Society degree.
- The Cognitive Science Major, a degree with strong data-science elements, which now attracts ten times the majors originally planned.
- The "Better Book Project," led by Jim Stigler and funded by the Chan Zuckerberg Initiative, the California Learning Lab, and the Schusterman Family Foundation, which applies data-science methods to improve the design and delivery of educational material, including real-time evaluation of student performance and instructional effectiveness. https://uclatall.com/
- A variety of venues offer short courses on data science skills: The Institute for Digital Research and Education (IDRE), the Collaboratory in QCBio, the UCLA Library's Data Science Center, the Digital Research Consortium, etc.
- An Introduction to Data Science course currently taught in 45 high schools across California, training thousands of students in the fundamentals of data science. https://www.introdatascience.org/

The city of **Los Angeles is a resource** that can contribute enormously to DataX. No amount of investment by peer institutions can substitute for the unique comparative advantage of our location in one of the great global megacities. The Sustainable LA initiative is a **one-of-a-kind opportunity for community engagement**, thanks to its existing relationships with city and county governments. UCLA faculty have embarked on ambitious projects for social good using data from local governments, e.g., the Million Dollar Hoods project and the California Policy Lab's work on homelessness. Los Angeles also has a singular concentration of compelling industry partners. Most obvious is the entertainment industry in all its forms (theater, film, television, digital effects, animation). But Los Angeles is also home to a significant aerospace presence; the growing tech cluster in Silicon Beach; the biotech giant Amgen; and many others. There are also unique local non-profits, like the RAND Corporation and the Getty. UCLA has an opportunity to **integrate student internships and research experiences in the city as an important feature of its role as an urban campus**, piggybacking on widespread student interest in having an impact on issues of sustainability and social good. Finally, the diverse population of the city presents **many opportunities for education and outreach for underrepresented groups**. For example, UCLA's recent **commitment to become a Hispanic-Serving Institution (HSI)** means that we have the opportunity to become the nation's leading university in training Hispanic data scientists, helping to correct a huge disparity. Combined with the *Rising to the Challenge* initiative, these commitments could have a transformational effect on the composition of the data science workforce within and beyond

California, assuring that access to the transformational potential of data is more equitably distributed and equipping more students from underrepresented backgrounds with the tools to create and critique these technologies.

Arah's group on Causal Inference and Computational Epidemiology uses computationally intensive modeling of causal and bias methods...to investigate effects of social and clinical interventions on obesity, diabetes and pediatric, perinatal and cardiovascular outcomes. In particular, they have designed and built an agent-based virtual model of Los Angeles, currently focused on simulating the natural and interventional development of obesity and type-2 diabetes from birth to age 65. **— California Center for Population Research WP**

Santa Monica has provided several student groups with interesting data, and in turn the students have helped the city design policies around e-scooters.

—Dept of Public Policy, Luskin School of Public Affairs WP

"Mapping African LA": Organizes and analyses social, demographic, ethnographic and historical data on the African diasporic communities in Los Angeles to build an oral history archive of these communities and develop pathways of diversity recruitment at UCLA.

—African Studies and International Development WP

CPL partnered with LA County and LAHSA to use data linked across 8 county departments to develop models to identify county clients at the highest risk of returning to homelessness or experiencing homelessness for the first time. In response, the County Homeless Initiative invited CPL to be a key member of its Homelessness Prevention Work Group and the Board of Supervisors passed a motion directing County agencies to work with CPL to develop strategies to use CPL's risk models to target County department resources to those at highest risk of homelessness. **—California Policy Lab WP**

NEEDS & OPPORTUNITIES

Needs & Opportunities

Given the extraordinary strength reviewed in the previous section, one might reasonably ask whether UCLA *needs* a DataX initiative. Our interviews strongly suggest that it does.

Now is the time to consolidate and build on the strengths outlined in the previous section. Our interviewees expressed (from multiple vantage-points) that a campus-wide effort was necessary to meet the moment and take things to the next level. What is already happening on campus constitutes a great jumping-off point for such a transformation. **On the research side, we can expand the footprint of data science, foster new collaborations, and launch cross-cutting transdisciplinary research efforts that leverage the excellence of our large, multifaceted campus. We can also establish a coherent, external-facing presence for UCLA's DataX activities that appropriately reflects the scale of UCLA contributions. On the educational side, our interviews indicated that educational offerings should expand dramatically to fully reflect the explosion of opportunities for our students in careers that require a combination of disciplinary expertise and data-related skills.** Our interviews consistently articulated a tension between the need to distribute resources and expertise across the campus, in a way that adapts to local needs and opportunities, and the need for campus-wide coordination and community-building, to maximize the potential of our DataX efforts and assure that our research and educational programming is legible and navigable within and beyond UCLA. We also note that some degree of campus-wide coordination is important to align our DataX efforts with UCLA's commitments in equity, diversity, and inclusion.

Our conversations with campus stakeholders revealed needs and opportunities that cut across campus functions: Research and research support, education, and engagement beyond UCLA's campus. In the research space, needs and opportunities cluster into two distinct bundles, which broadly correspond to two distinct campus constituencies. Educational opportunities cut across (and go beyond) both bundles; we survey them briefly below.

The first bundle focuses on **advancing cutting-edge data science and its applications, understanding its social, ethical, and policy implications and its possibilities for enhancing creative expression, and providing leadership for DataX research and education across campus.** The constituency attached to this bundle explicitly orients toward interdisciplinary scholarship; for these scholars and scientists, DataX is a core part of their identity. Addressing this bundle requires some degree of centralization. The appropriate campus structure must bring people together, through convening power and community-building activity. Bringing people together makes them visible to each other and beyond the campus. It also catalyzes the interdisciplinary collaborations that drive breakthrough innovation in fundamental data science; applications that advance discovery, creativity, and scholarship; and our understanding of the social, political, and ethical dimensions of data[3] . Concentrating our campus excellence in DataX innovation also makes it easier to secure external partnerships (with corporations, non-profits, and government) and negotiate or acquire access to high-value data. These various

---

3   Note that these three components are positioned at the heart of recent federal investments, e.g., the AI Institutes program, which explicitly requires a virtuous feedback cycle between fundamental research and applications, and underscores the need to address social and ethical issues within the Institute framework.

requirements together suggest a centralized structure. We therefore recommend the creation of a **DataX Institute** as the appropriate structural innovation.

> *An important CTSI effort is to help connect different researchers together. Specialists in the management and analyses of specific data types (e.g. imaging, molecular) and domains (e.g. pathology, cancer) now exist. Still others (e.g. computer scientists, statistics, mathematics) provide technical expertise in different computational methods (e.g. deep learning, NLP). Finding these subject matter experts for consultations and collaboration at UCLA remains an open challenge, as it requires in-depth knowledge about an individual's specialization...and no resource presently provides the level of characterization needed.* **—Clinical and Translational Science Institute WP**

The second bundle focuses on **building capacity in using data across campus**. The constituency attached to this bundle tends to orient relatively more toward their own discipline and its particular questions. These scholars and scientists are interested in expanding their repertoire to leverage new data resources and in learning existing, validated methods for incorporating large and/or complex data into their current practices. Meeting this demand to build capacity requires expertise that is closely adapted to the diverse data types and scholarly practices used on our campus. It suggests a decentralized and distributed structure. We therefore recommend the creation of a **DataX Support Network (DSN)**, with nodes at the Division or School level and cross-campus coordination (through the DataX Institute) to structure offerings, establish common knowledge of campus resources, share best practices, and continually upgrade skill sets. Many of these service units already exist (as demonstrated by the white papers they submitted); in other cases they need to be created, and where appropriate their scale of offerings expanded.

Through consultation with leaders of Data Science efforts at peer institutions, we confirmed that this dichotomy emerges robustly at both private and public research universities. Peer institutions have developed a range of structures to address these needs and opportunities. The structures proposed in this report are broadly consistent with structures found at peer institutions, but have been tailored to distinctive features of our campus. While the structural tension between centralization and de-centralization also confronts peer institutions, it is especially acute at UCLA because of our large size and comprehensive scope. To resolve this tension, we highlight many opportunities for collaboration and coordination between the DataX Institute and the DataX Support Network.

Turning to **education**, it was clear from our consultations that **the need is acute and the opportunities are manifold**. Not only has there been extraordinary growth in the need for knowledge of data science in existing careers; a large number of new careers have also emerged, which combine data science with a particular disciplinary knowledge. **Such opportunities are not restricted to science, medicine and engineering. They also span the social sciences, humanities, and the arts.** There are opportunities in companies of every scale and stripe, as well as non-profits and government. Clearly, UCLA should do everything it can to inform students about these opportunities and to provide training pathways that make such careers available to all interested students. **Developing courses, certificates, minors, and majors appropriate for a wide range of student backgrounds and interests is thus a high priority.**

These courses also provide an opportunity to highlight the **importance of robust public access to research data.** Open data can accelerate the pace of discovery and its applications to societal problems, as well as heighten the visibility and reputation of UCLA and its scientists and scholars. By ensuring transparency and facilitating the reproducibility of research results, access to data and code is also

important to preserving research integrity and maintaining public trust in science.

The white papers reflect a variety of initiatives planned or underway at UCLA. The Institute and Support Network will energize, amplify, coordinate and rationalize educational developments, while also cultivating DataX internship and career opportunities for UCLA graduates. It must be emphasized again that **we are nowhere near meeting the demand (or potential demand) in this space**. The campus should therefore set aside concerns about "duplication" and resist attempts to monopolize DataX training. Instead, the emphasis should be on assuring the development of high quality, externally legible, and internally navigable educational opportunities.

Dean Hunt in 2018 instituted a Division of Social Sciences "Big Data Initiative"...The Committee formulated four central goals...1. Study the societal impacts and ethical aspects of big data and data science...2. Provide a space for intellectual exchange for researchers working with large-scale and complex data...3. Broaden data science training and its access to diverse students...4. Provide computing and programming support for big data intensive research.
—Division of Social Sciences WP

This leads to the sobering conclusion that the vast majority of students in one of UCLA's largest majors leave the university with a statistical toolkit composed of basic methods developed in the early 20th century (e.g. t-tests, correlation). This toolkit is obviously inadequate for a 21st century workforce...The creation of an undergraduate quantitative methods minor is a natural synergy between the department's 2018 strategic hiring plan and the DataX initiative.
—Dept of Psychology WP

Data Science is now playing a role which is at least on a par with the various natural sciences. It is fair to say that Data Science is transforming both research and practice in Bioengineering, Civil Engineering, Chemical Engineering, Material Science, Mechanical Engineering etc. Yet, except for CS and ECE the broader curriculum at Samueli Engineering has largely been untouched by this transformation. **—CS and ECE WP**

The proposed MS in Public Health Data Science emphasizes data analytic tools, computing, practical data solutions, and close connection with health industry and other fields in public health...For students and health professionals interested in applying data science to medical and public health problems, options are scarce. Among the existing health-oriented data science programs, only Harvard University's MS in Health Data Science and UNC Chapel Hill's MPH concentration in Public Health Data Science are housed in the school of public health. USC's MS in Public Health Data Science is in the Keck School of Medicine and the UW MS in Data Science is an interdisciplinary program involving six departments and schools. The UC system has yet to propose a data science degree program dedicated to medicine and public health. The existing programs are increasingly finding themselves oversubscribed and failing to meet the burgeoning demand.
—Biostatistics Dept WP

In the next sections, we will describe the DataX Institute and DataX Support network in their fully mature form. We acknowledge, however, that current resource constraints will make it difficult to roll these out immediately. As examples at peer institutions suggest, there is need for substantial fundraising to secure the resources and space that these transformational efforts require. That said, we strongly discourage the campus from delaying its efforts until resources and space have been secured. Instead, we recommend the creation of a transitional entity, the **DataX Homeworld**, which can provide an interim home for DataX hiring, research, coordinated educational efforts, training, outreach, institutional research, and other activities (including a modest physical footprint). The Homeworld will transform into the DataX Institute once funds have been secured and an appropriate physical space established (peer institutions often devote entire buildings to their campus-wide data science efforts).

We envision the structures we have recommended as an **enabling resource** for the entire campus, helping people to find collaborators, learn about new problems and techniques, and participate in a campus-wide educational effort. A very frequent request in the white papers was for **assistance in finding collaborators**. There are, of course, researchers who succeed in finding their own collaborators, but the greater the distance between fields, the harder it is to do this. **For highly interdisciplinary collaborations, researchers often do not even know what expertise is out there and what they are looking for.**

In our experience, UCLA has an excellent culture of openness to interdisciplinary collaborations; by proposing mechanisms that build upon this culture to create an exciting interdisciplinary community around DataX, we by no means intend to imply that there is something "broken" that needs to be fixed. Instead, we are responding to feedback in our consultations and in the white papers suggesting that appropriate structures could catalyze collaboration at new scales and levels of intensity, while also reducing entry-barriers and making collaboration less dependent on pre-existing networks (with important implications for inclusion).

What we are proposing is part of the way interdisciplinary collaboration has worked on our campus. There is a long history at UCLA of creating structures to enable interdisciplinary research and education when they would be helpful, e.g., the Institute of the Environment and Sustainability. A recent report of the Academic Senate on Interdisciplinary Education makes the point that new structures are needed at UCLA to enable interdisciplinary education.

We did consider various possible structures, including doing nothing and leaving everything to individual departments. We came to recommend the structures that we did only after extensive consultation with over 100 UCLA faculty who are actively involved in research and creative activities and education in DataX, after reading the white papers, and after looking at what peer institutions have done. We did find that certain needs were best met by a distributed structure, and the DataX Support Network is tailored to those needs. We also found that certain needs were best served by something that was inherently interdisciplinary and was not situated in any one department, school or division. The DataX Institute is designed to meet these needs.

One way to frame the difference between an enabling resource and a "layer of bureaucracy" is whether the structure created makes it possible for faculty, schools and departments to do things they otherwise would find difficult or impossible, versus a structure that prevents or slows down those activities. The DataX initiative, if we have anything to say about it, will be of the former type.

PROPOSED STRUCTURE:

DATAX INSTITUTE

Engineers Week at UCLA

Proposed Structures: DataX Institute

The DataX Institute would provide a research hub and meeting place for scholars who put DataX at the center of their research. It would work to energize activities across campus by connecting people, building community, spurring and coordinating educational innovation, providing an interface with external partners, and establishing a strong public-facing presence for UCLA on the web and the broader media landscape. As the description below suggests, this will require substantial investment, and the campus should commit early to fundraising in support of an eventual Institute. We emphasize, however, that many of these activities can be fruitfully pursued under a transitional entity (the **DataX Homeworld**) and we urge the campus to identify activities that can be initiated in parallel to fundraising efforts. While we usually use the term "Institute" below, in many cases the DataX Homeworld can be substituted when activities are pursued before the formal establishment of the Institute.

Personnel

Institute personnel fall into five groups: **Faculty Fellows**, **Postdoctoral Fellows, Partners from Industry and National Labs, the leadership team** and **support staff**.

Faculty Fellows will have a home in an existing department. Some Fellows will be recruited from existing faculty; others will be recruited externally, and supported with a half-FTE from the Institute and a half-FTE from an existing department. They will be drawn from three equally important **cohorts**: (1) Those doing fundamental research in data science, mostly drawn from Mathematics, Statistics, Computer Science (CS) and Electrical and Computer Engineering (ECE); (2) Faculty who use cutting-edge data science methods and find innovative ways to apply data science in their research, scholarship and creative activities; (3) Faculty working on data cultures, ethics, policy, and societal implications (DEPS). There will be explicit expectations of involvement in Institute activities for becoming and remaining a Fellow; indeed, we expect that the majority of Institute activities will be initiated and led by faculty Fellows. Later in the report, we describe the benefits and expectations associated with this affiliation.

The Department of Information Studies brings a distinctive focus on social justice and diversity to bear on the production and use of data as an ethical and cultural, and not only technical matter... Among the issues we address are ownership, privacy, authenticity, validity, classification, equity and sustainability...Our faculty, students, and programs are renowned among peer schools for their critical, multi-disciplinary, community-driven research, teaching, creative work and professional practice. **—School of Education and Information Studies WP**

UCLA has been awarded a Specialized Program of Research Excellence (SPORE) in Prostate Cancer from the National Cancer Institute, tasking it with leading efforts to drive progress in the diagnosis, prevention, and treatment of prostate cancer. This is a national competitive award...The SPORE includes a dedicated computational core which supports multiple prostate cancer research projects across UCLA, and would be an ideal point for integration with DataX through, for example, quantitative biology working groups and joint analysis of complex datasets. **—UCLA Urology WP**

We *strongly recommend* the proposed approach of **treating these three cohorts as equal in importance;** this reflects the broad understanding of DataX that we observed in our interviews, and avoids a common pitfall in which "core" data scientists dominate a campus-wide data science effort, limiting the broader interdisciplinary reach and discounting policy, ethical, social, and cultural implications. More importantly, we see such a balance among cohorts as underpinning the creation of something unique and exciting. This balance fully capitalizes on the strengths of UCLA and is consonant with its values and mission. It also reflects the emerging international vision of data science, which draws equally on these three strands.

As we describe in detail below, we view the selection and hiring of Faculty Fellows as one essential site for the campus to pursue its equity, diversity, and inclusion commitments in the DataX space. We also believe the Fellows (whether internally recruited or externally hired) will play a pivotal role in spreading expertise in DataX across the campus, through their teaching, mentoring, and other activities in home departments. To give a sense of scale: many peer institutions have launched hiring initiatives in this area, and some—despite being smaller than UCLA—have committed to hiring as many as 50 new faculty at the intersection of computing and the disciplines. Engagement with the Academic Senate is necessary to establish appropriate procedures for hiring and evaluating faculty who hold partial FTEs in the Institute.

The MAE Department views Data Science as a critical area for hiring, educational and research activities… One of the unique aspects of the research and educational endeavors at MAE is the importance of transitioning basic research to applications. Our faculty members have strong track records of working with aerospace, automobile, and power system industries, implementing data-driven technologies into complex engineering systems… Some ongoing research efforts: Data assimilation and machine learning to train sensors to detect and estimate atmospheric disturbances during aircraft flight… Smart Electric Vehicle charging algorithms to satisfy EV charging load and machine learning to deal with EV driver behavior uncertainties.
—Dept of Mechanical and Aerospace Engineering WP

ISG faculty is actively engaged in grant-funded research to ask the fundamental question of how science is changed by new computational and data tools and modes of thought. We have a variety of ongoing research projects looking at the nature of bioscientific work as it becomes more computational. We have conducted research into interdisciplinarity, asking what kinds of socio-technical infrastructure is necessary to productively conjoin different expertise across computer science and cell biology, for example; how publics use big data resources generated by genomics including the transformation of genetic counseling; crises of replicability and the politics of open science. **—Institute for Society and Genetics WP**

We conceive of the **Postdoctoral Fellows** as a prestigious position, comparable to the data-science-oriented fellowships at peer institutions like the University of Michigan. Postdoctoral Fellows will be drawn from the three categories of researchers mentioned above, in the description of Faculty Fellows. **Each Postdoc will have two mentors, drawn from two of these three cohorts.** Postdocs will have some teaching obligations (described in greater detail below); they will also be involved in Institute research, and have opportunities to support data science on campus. These positions will reinforce the Institute's interdisciplinary nature and provide an infusion of energy by making UCLA a magnet for emerging talent in

the DataX space. Postdocs have also been a key component of data science strategy at peer institutions, and often play a critical role in seeding data-intensive scholarship across the campus.

The proposal to offer Graduate Student Researcher and Postdoctoral positions through the DataX Institute aligns with the plans of the Cotsen Institute to offer such positions on projects with significant data components in an effort to expose scholars at an earlier stage in their academic careers to the complexities and importance of data literacy. **—Cotsen Institute WP**

Partners from Industry and National Labs: These positions allow researchers employed in industry or in national laboratories to spend time at the DataX Institute. Such partners would provide a valuable connection between the Institute and their home organization, promoting research collaboration and the cross-pollination of ideas. In some cases, partners could participate in the Institute's educational mission (e.g., by mentoring students, leading research teams, or teaching workshops or classes). We also view them as supplying a critical conduit between UCLA students and internship opportunities in industry and at national labs. The National Lab Partners would contribute to the campus goal of strengthening the relationship between UCLA and the UC-managed national labs (especially Lawrence Berkeley but also Lawrence Livermore and Los Alamos).

The leadership team will act to guide the Institute's activities, make decisions about use of Institute resources, seek out opportunities in the DataX space very broadly, and energize and enable research and education across campus. The Institute **Director** will be appointed by the EVC/Provost and report to the Vice Chancellor for Research and Creative Activities; the detailed reporting structure is likely to evolve as the Institute mission evolves, see below. The role of the Director is to foster an inclusive and inspiring environment in research and education, and to seek out and respond effectively to a wide spectrum of research opportunities. Fundamentally, the role is one of community-building and advocacy, rather than setting and advancing a particular research direction. The Director will be supported by **three Associate Directors, for Research, Education, and Support,** who will lead Institute activities in these three critical functions. The Associate Director for Research will lead the Institute's portfolio of research activities (described in detail below) and will ensure that the Institute remains at the cutting-edge of DataX scholarship. The Associate Director for Education will play a critical role in coordinating campus-wide educational initiatives, and will lead the development of educational programming led by the Institute. The Associate Director for Support will lead the Institute's support team and coordinate the activities of the DataX Support Network; s/he will also lead the development of an ecology of innovative training programs (workshops, hackathons, short courses, online modules, etc.) to build campus capacity in DataX, and (in coordination with the other Associate Directors) lead the acquisition and management of research and educational datasets. We anticipate that these leadership roles will be very intensive—matching or exceeding the scope and complexity of analogous roles at interdisciplinary units like IPAM—and may in the Director's case be closer to Vice-Provost roles in programs like the International Institute or the Institute of American Cultures. We recommend the appointment of interim leadership (e.g., during the DataX Homeworld phase) so that campus-wide DataX activities can begin as quickly as possible; see below for further discussion of implementation.

To assist and advise the leadership team, an **Executive Committee** will be elected by the Faculty Fellows. It will have nine representatives, three from each of the three cohorts; it will also include the Director and 3 Associate Directors *ex officio*. The cohorts' representatives would be elected separately by fellows in (1) Mathematics, (2) Statistics, and (3) Computer Science and ECE; (4) Applications from North

Campus units, (5) Applications from the Physical Sciences and Engineering, and (6) Applications from Medicine and the Life Sciences; and (7) Fairness, Accountability, Transparency, and Ethics, (8) Societal Implications of DataX and Data Cultures, and (9) Policy and Legal Issues in DataX. The Executive Committee will approve the appointment and renewal of the Faculty Fellows; appoint search committees (one for each of the three cohorts) who will review applications from departments to identify candidates to receive partial FTEs from the Institute; approve the allocation of Institute FTE to faculty candidates; and discuss the activities and long-term plans of the Institute, providing guidance, to the leadership team.

To provide broad input to Institute activities from its campus constituencies, the leadership team will also appoint a **Stakeholder Board** with expertise in each of the critical functions (Research, Education, and Support), expertise in equity, diversity, and inclusion; and representatives from the Postdoctoral Fellows and technical staff. The Stakeholder Board should have subcommittees for Research, Education, and Support, providing more detailed advice to the appropriate Associate Director in these areas.

To provide high-level advice and build connections to local partner organizations (e.g., local government, non-profits, community organizations, industry, and relevant national labs), the leadership team will also appoint an **External Advisory Board.**

The **staff** should be appropriate in size to the level of activity at the Institute—small at first, but expanding as the Institute evolves. In addition to typical staff roles like administrative support, financial management, and IT support, the Institute will also require staffing in grant support (so that grants can be run through the Institute effectively). To support the research activities of the Institute, its working groups, and its incubator programs, the staff should include expertise in research computing; in dataset acquisition, curation, access, and management; and in data science. These research support functions will sit inside a "node" of the DataX Support Network, housed at the Institute and connected to the other Support Network nodes as described below.

Activities

The **activities of the Institute** will fall into four broad categories: **Research; Education; Campus Leadership, Outreach and Advocacy;** and a role in **Faculty Appointments.**

Research Activities:

Working Groups, proposed by Institute faculty or suggested by the Institute leadership, will come together around specific themes in research or education. Working groups are a core element of the Institute, and ensure that its activities align with bottom-up interests at UCLA and retain a cutting-edge orientation. Working groups might be initiated by grant opportunities; another interesting possible model is the "Sandpit" funding process, which was recently conducted by UCLA's Sustainable LA Grand Challenge. Working groups will have access to support from the Institute DSN node, and will be open to participation by Industry and Lab Partners as well as Fellows. Faculty Fellows would be expected to participate actively in at least one working group.

Incubator Grants will allow teams of Faculty Fellows to be resident at the Institute for a period of time (usually 1 quarter) to catalyze incipient Working Groups. **"Doorway" Grants** will allow faculty who want to jumpstart a shift in their activities toward DataX to spend a quarter at the Institute; these grants would not be open to Faculty Fellows, as they are intended to "grow" new Faculty Fellows from the broader campus community. Both of these opportunities will have substantial spillover benefits in graduate education; graduate students and postdocs can participate in Incubator projects, and faculty who have benefited from a Doorway Grant will be positioned to enhance training for their students and others in the home department.

In the course of preparing this report, we did encounter some concerns that DataX might constitute a "highly funded and favored layer of research." There is, of course, already a highly funded and favored layer of research—determined by the actions of external funding agencies. **Some of the programs we propose can help to spread this external funding more evenly. DataX would provide some seed funding, via the Incubator Grants and Doorway Grants, to aid in forming the highly interdisciplinary teams needed to go after some of this funding.** The Doorway Grants, in particular, are intended to help faculty with little previous DataX engagement add this dimension to their scholarship, preparing them to compete for relevant grants or join in collaborative funding proposals. We envision this as enabling a wide variety of projects, coming from across campus. We feel confident in our claims that the benefits of DataX will be spread across the campus because everyone we spoke to (and every white paper we received) articulated many ways their scholarship and teaching could be enhanced by an ambitious effort in this space.

Seminars would be organized by Faculty, Postdoctoral Fellows or Industry and Lab Partners. The seminar topics can be distinct from Working Group themes; they might involve fundamental data science, substantive applications, or work on data cultures, ethics, policy, and societal implications. These seminar series would be open to everyone across the UCLA campus. Fellows would be expected to participate in at least one seminar. Some seminars can also be structured as for-credit courses, following a common practice in other units on campus; even without course credit, the seminars will enrich the environment for graduate students, as well as advanced undergraduates.

The Institute would facilitate, on a case-by-case basis, the **acquisition of important datasets.** A particular priority is the **acquisition or creation of datasets that enable research on racial equity and racial justice, with a view toward making UCLA an international center of such research.** This priority is in alignment with the *Rising to the Challenge* initiative and UCLA's commitment to become a Hispanic Serving Institution. Datasets should be maintained and made accessible to the campus community in a "data commons," to which campus researchers could contribute their datasets as well (see details below). The Institute would also **facilitate access to computing** for Institute projects, including dedicated resources & support staff. These compute resources could be provided through the Institute for Digital Research & Education (IDRE), Divisional clusters, and outside providers (e.g., Amazon Web Services, Microsoft Azure, etc.).

The Institute will also work to forge **research partnerships with industry, government agencies, and community organizations.** Such partnerships might involve the sharing of relevant data, code, compute resources, or even personnel, as the case and the collaboration warrants. We envision these partnerships as providing entree to internship opportunities for UCLA's burgeoning cohort of DataX-savvy undergraduate and graduate students. To facilitate partnerships that will lead to positive social impact on our home city, we propose a DataX LA program, which would provide 50% funding for 1-quarter

internships for undergraduates with LA City and County governments and local NGO's. The Center for Community Engagement provides a wealth of expertise and a network of contacts for launching this program.

As a strategy for stimulating collaboration within and beyond campus, the Institute will also host periodic **DataXelerator** events, perhaps initially once a quarter. At these events, individuals from industry, government, and community organizations will give talks about data-intensive or data-related problems that they face, with the expectation that interested Fellows and students might initiate collaborations to address those problems.

We will identify the existing compliance issues of AI use, provide an independent source of public-facing evaluation and knowledge for citizens seeking greater information, protection and redress, and deliver a model legislative package that upholds dignity, equality, and transparency in government's use of algorithmic and human moderated digital and data-reliant systems… Professor Noble developed a new course [that] will explore the moral, social, and ethical ramifications of the choices we make at the different stages in the social construction of data. This will include cultivating a critical analysis of processes of data collection, data mining, data storage and the deployment of these as they affect and are affected by social conditions on a variety of different communities, publics, nation-states, and individuals... Students will develop an acute understanding of the social, historical, and political dilemmas of big data, algorithmic decision-making, predictive analytics, and the distinct challenges associated with ethical, civil, human and sovereign rights models of engaging the modern digital information era. **—Center for Critical Internet Inquiry WP**

In non-profit organization and government agencies the process of data curation is less obvious--data are oftentimes more fragmented, collected for administrative purposes rather than research and analytic purposes, and frequently disconnected from policy decisions...Data curation around well-being measures can offer transparency to communities, accountability to advocates, clarity to government, and opportunities to students. **—Center for Health Advancement WP**

The Institute will also host periodic, UCLA-specific **DataXelerator** events, at which faculty and graduate students with interesting data problems can present their challenges to potential collaborators; this will provide a critical opportunity for faculty to build bridges to DataX and collaborations with DataX scholars, even if they do not want to build personal capacity or expertise in DataX. We can also imagine a fruitful partnership between the **DataXelerator** series and StartUp UCLA, allowing individuals or groups with compelling business ideas to find complementary technical, ethical, or policy expertise.

The UCLA Ethnomusicology Archive...collects field recordings and rare commercial recordings of musical genres from all over the world, and, with over 150,000 holdings, ranks as the second largest audiovisual archive of traditional music in the United States...During Spring 2020, the archivists and GSR's worked overtime to provide digital copies of holdings to many different classes involving hundreds of students. In some cases, the Archive's ability to provide digitized recordings and videos has made the difference in keeping classes viable. **—Ethnomusicology Dept WP**

Education:

> *As early as the 1980's ethnopharmacology emerged as the discipline aimed at translating traditional medicinal practices into laboratory research… This generated a few spectacular successes, such as the discovery of the anti-malarial drug, Artemisin… PharmAntica will generate robust data for the development of new medicines produced on the basis of deep-time traditional uses in environmentally and economically sustainable ways.* **—PharmAntica WP**

> *The project that we are proposing directly supports DataX by collaborating with our partners to acquire, curate, and manage data.* **—CEILS WP**

The Institute, which by its nature will be a gateway to the latest in DataX research, will have primary responsibility for leading the formal educational initiatives envisaged in this report; the Support Network, by contrast, will be responsible for informal training through workshops, hackathons, and one-on-one research support. The Institute will have a standing working group on data science education, chaired by the Associate Director for Education.

The Institute will foster development of cross-campus courses in data science. A top priority is the development of a suite of **easy-entry freshman cluster courses,** to be described elsewhere in the report. We would like to see these expand to a set of **core lower-division data science courses,** which would be suitable preparation for a variety of majors and minors, as well as **interdisciplinary graduate offerings.** These courses would be taught by Institute Fellows; those not receiving part of their FTE from the Institute can receive a course buyout at the internal rate to compensate the department (resulting in a net increase in the courses offered on campus; note that these faculty can also leverage the materials and expertise developed in Institute courses to inject DataX into departmental curricula).

The precise scope of these offerings will depend sensitively on what happens across campus; we envision the Institute educational activities as complementing rather than substituting for departmental offerings, e.g., by preparing students for such courses, offering "bridge" courses to empower students to move to more technical coursework, and providing venues for highly interdisciplinary instruction that would not sit naturally within departments. The Institute will also foster development of data science courses within home departments, with funds available to Fellows to support course development (e.g., through GSR positions to develop course materials).

To stimulate campus-wide investment in DataX education the Institute will develop possible **templates for data science courses and degree programs;** by providing a standard technical infrastructure, skills "checklist," and a robust pool of instructional materials (datasets, computational notebooks, readings, etc.), the Institute can greatly facilitate the task of creating new courses and degree programs, while making courses easier to cross-list and helping to rationalize and commensurate offerings across campus. Institute leadership will advocate for DataX educational initiatives at the Academic Senate, e.g., when units seek approval for new majors, minors, or certifications.

In addition to degree programs, the Institute should propose a set of **certifications in data science** at multiple technical levels, with criteria that can be satisfied by a variety of campus pathways. These

certifications should ultimately be institutionalized so that students can get them as official designations on their transcripts. Finally, the Institute will seek to expand **internship and research opportunities for students,** leveraging its external partners and the outstanding affiliated Faculty Fellows.

SSCERT assists faculty with the design and creation of online course websites and while many of these are straightforward, we have worked with a number of faculty in the development of quite complex learning software...One of the most innovative instructional projects that SSCERT has continued to work on for a number of years is the class gaming platform conceptualized by Professor Lohmann in Political Science that provides the foundation for several of her courses….She has created dozens of multiplayer interactive games and simulations with which students can collect and analyze data, attempt to replicate studies, participate anonymously in transactional activities, test theories, and understand complex concepts as participants in numerous communal activities.— **Social Sciences Center for Education, Research and Technology WP**

Chemical engineers are dealing with more, and more-complex data generated in modern plants every year. To take advantage of the increasing amount of data, chemical engineers need data science tools to do data analysis and incorporate them in various fields of chemical engineering. **—Chemical Engineering WP**

The philosophy of probability and the philosophy of statistics are growing and important fields within philosophy, and they interact heavily with traditional philosophical questions about the relation between chance and responsibility...When does a data set license one to infer causation rather than mere correlation?...One could imagine, for instance, discussing this last question by reading important texts from the history of philosophy like Hume or al-Ghazali, alongside contemporary authors like Judea Pearl and James Woodward. This terrain is also rich in its legal and ethical applications, such as statistical evidence in the law, racial profiling, differential privacy, and the kinds of responsibility borne by self-driving cars. **—Philosophy Dept WP**

21st Century Biosciences have been thoroughly transformed by two profound developments: The quantitative transformation catalyzed by the transition from cataloguing components to studying the connections between them and the systems-emergent properties ("Systems Biology"); The data sciences transformation, catalyzed by technological innovations that generate massive amounts of complex data, allows high dimensional characterization of biological samples and dynamical processes. As a result, biosciences research efforts have shifted from being 95% experimental in 2000 to >50% computational in 2020. This transformation requires dramatic shifts in biosciences research training programs and undergraduate degree programs that ensure that graduates are equipped to enter the 21st century biosciences workforce, or pursue careers in the academic discipline. **—Quantitative and Computational Biosciences WP**

Campus Leadership, Outreach and Advocacy:

The Institute will play a leadership role in **fostering activities in DataX across campus** (e.g., offering advice to departments or providing partial support for DataX-adjacent colloquium series). The Institute will also **engage in DataX outreach and advocacy**, within and beyond the campus. Most critically, it will create and maintain a **comprehensive and compelling website**, which will serve as an entry point for all DataX activities on campus. The website will include a comprehensive **portal for DataX research.** This portal will allow UCLA students and faculty to find potential mentors and collaborators; it will also provide a single platform for showcasing UCLA's activity in DataX to the world beyond campus, assuring that the public perception of its stature and success is commensurate with the quality of our activities. In addition to hosting pages for all Institute Faculty and Staff, as well as Institute programs, the website will share success stories and recent news, be used to post job opportunities, and advertise seminars and events. We recommend that the Institute's precursor, the DataX Homeworld, immediately launch an effort to systematically document DataX research activities across campus, at the level of individual projects. As part of its DataX outreach and advocacy, the Institute will work with the library and the DataX Support Network to build and maintain a **data commons** for both datasets and code; the underlying assets can be hosted on campus resources or elsewhere (e.g., GitHub). Building on successful models in biomedicine, this portal should allow grades of access wherever possible (e.g., description of resource but no access; full access to registered users on campus; full access to registered users anywhere in the world), coupled with easy means of contacting data and code creators. In keeping with UCLA's commitment to equity, diversity, and inclusion, the Institute should make a particular effort to secure and host data relevant to minoritized and marginalized communities, with easy onramps to relevant code and datasets. Finally, the Institute should focus on developing infrastructure for data access and data sharing that are appropriate for data types on which UCLA has some comparative advantage, like health records (through the School of Medicine and its various data-intensive departments and Institutes) or media data (e.g., high-definition video, through the School of Theater, Film, and Television). Such data present particular technical challenges and therefore represent research opportunities for scholars across the three cohorts, e.g., developing algorithms for privacy- or secrecy-preserving analysis (like homomorphic encryption), creating protocols for efficient streaming (for media data), or exploring the complex tradeoffs between data access, data quality, and privacy (as in health data).

There are strong precedents on campus for supporting visualization and sonification of scientific data, but less infrastructural engagement with how digital audiovisual media can both be processed as data--e.g. as a machine learning training set or network flow--and the same media can be experienced "directly" by the senses in other contexts, including the arts. Media created as the result of research activity in the arts can be used as input to research activity in the sciences and media captured within the context of other research across campus could be reinterpreted through artistic processes on campus...Creation of large datasets that serve as both media archives and, for example, machine learning training material may provide interesting opportunities for collaboration. For example, the Future Storytelling Summer Institute at TFT in 2019, supported by Amazon Web Services, considered how media-rich datasets about Los Angeles could be used to train machine learning to curate media experiences in an Olympics pavilion for 2028.
—School of Theater, Film and Television WP

The Institute's outreach role will also be substantial. It will **hold events to raise awareness** of and interest in DataX, including a **DataX Distinguished Lecturer Series**, held once each quarter, and an annual **DataX Master Class**, which will bring a distinguished practitioner to campus for a hands-on introduction to cutting-edge tools. It will also **sponsor community engagement activities** (e.g., career days) and forge **formal partnerships with community organizations**; the Center for Community Engagement will be a critical partner and resource for such efforts. Lastly, it will work to **build industry/ academic partnerships** around data science and data-engaged scholarship.

Role in Faculty Appointments:

Modeled on existing practice in other units (e.g., the International Institute, the Institute for American Cultures, the California Center for Population Research), the Institute will receive FTE to build up in fundamental and applied data science, broadly conceived, as well as in data cultures, ethics, policy, and societal implications. We recommend a bottom-up process, based on the highly successful process used for the *Rising to the Challenge* Initiative, in which departments will propose hires in particular areas and collaborate with the Institute (or, in earlier stages, the Homeworld) in running the search. Proposed hires will only proceed when an existing department or departments vote to hire that person as faculty. A possible mechanism is explained in more detail in the section on resources.

Infrastructure

Space Requirements: These are substantial, particularly for activities at the scale and ambition that we recommend; many peer institutions have built or are building entire buildings devoted to their data science efforts. The Institute will need office space for Leadership, Staff, select Fellows (e.g., those supported by Institute FTEs or those with Incubator or Doorway Grants), Postdocs and Industry and Lab Partners. It will need collaboration spaces (including temporary "Incubator" spaces) and space to hold seminars (although in the short run these can probably be delivered virtually; in the long run, seminar spaces should be designed for hybrid participation and scalable sharing of presentations and other events).

The needs for space will come quickly once the DataX Homeworld is launched, so that the leadership and staff can hit the ground running. It will be critical once the Initiative begins recruiting faculty and postdocs. **It is essential that early faculty and postdocs have office space in close proximity to each other;** this will catalyze collaboration through frequent, informal interaction, and establish an Institute culture of presence and activity, in keeping with the scholarly literature on the importance of propinquity in stimulating interdisciplinary collaboration. The same is true of DataX Working Groups, which will benefit from intense interaction with other working group members as well as Institute faculty and postdocs.

In our consultations with peer institutions, **it became clear that if Institutes fail to create an attractive culture from the very beginning—so that the Institute is an exciting and stimulating place to spend time—it is very difficult to reverse later on.** This early "culture-setting" is necessary to make the Institute an intellectual magnet for campus research on DataX; we therefore recommend the identification of a suitable space for the transitional DataX Homeworld, and careful attention to establishing and sustaining a culture of presence and participation.

Grants: The Institute will work actively to build teams for interdisciplinary grants and put together

competitive proposals. The Associate Director for Research will have a particular mandate to identify opportunities and cultivate teams. In many cases, proposals will grow naturally out of Institute Working Groups. We envisage the Institute pursuing both large scale, capacity building grants (e.g., recent NSF calls for Institutes in AI or Data-Intensive Research or NRT awards) as well as individual research grants. Grants supporting research and education at the Institute would be managed through the Institute, and return of overhead would help to fund Institute activities. The Institute will need one or more staff with experience in applying for and managing grants, who will liaise closely with the Office of Contract and Grant Administration. Depending on the scale of Institute activities, it may also need a staff member with expertise in grant writing, to support Institute Fellows in grant preparation.

Reporting Structure and Formal Character: Because of its campus-wide scope, the Institute will be directly under the Office for Research and Creative Activities (ORCA). In initial phases, leadership will report directly to the Vice Chancellor for Research and Creative Activities; where appropriate, the Director may occasionally join the VCR's meetings with the EVC/Provost. As the Institute and Support Network evolve, we foresee this arrangement evolving as well. In particular, the recent Administration of Teaching, Learning and Services (ATLAS) Committee and the structures it proposes might open new avenues for appropriate reporting on the educational aspects of the Institute; for example, a dual report to the VCR and a Vice Provost with an educational innovation portfolio may be appropriate.

The formal character of the Institute (and its precursor, the DataX Homeworld) should be commensurate with the scope of its activities, which include holding FTEs, developing a range of research and educational activities, and hosting grants. Because we view DataX as a having a campus-wide, transformational scope, it is important that it be positioned administratively to have a broad footprint— hence the recommendation that Institute leadership report to the VCR. The administration should consider models like the International Institute and the Institute of American Cultures when setting up the Homeworld and eventual Institute. Likewise it should establish the Homeworld/Institute in an administrative form that maximizes long-run flexibility (most likely as a Center for Interdisciplinary Instruction).

PROPOSED STRUCTURES:

DATAX SUPPORT NETWORK

UCLA Students - Kerckhoff Coffee House

Proposed Structures: DataX Support Network

The DataX Support Network will complement the DataX Institute, helping faculty and students across campus build capacity in data science and data-intensive research. While the Institute will focus on research and growing the ecology of formal instruction on campus, the Support Network will target a range of capacity-building activities, from one-on-one advice to workshops, short courses, bootcamps, etc.

As our interviews made abundantly clear, the scope and focus of the required support varies tremendously across schools and divisions. For this reason, we propose a distributed model. The DataX Support Network will be dispersed across units, but tied together through a "Hub" at the Institute. This strategy acknowledges a key fact surfaced by our interviews: It is critical that the people providing support be fluent in the disciplinary language and culture of the units they are supporting. This fact alone argues against trying to centralize DataX support. The Support Network will function as a kind of "circulatory system" for knowledge and techniques relevant to DataX, and will enable broad involvement of researchers and creators across campus. We also envision the support network as providing a point of entry for faculty interested in deepening their involvement with DataX, who could start by engaging with their local node, then might later begin to engage with the Institute (perhaps through a Doorway Grant), and eventually seek formal affiliation as a Fellow.

Structure: "Nodes" in the network will be distributed at the "dean-level" (i.e., Divisions and Schools). We also envision some campus-wide nodes with functional focus (e.g., the Office of Advanced Research Computing (OARC, formerly OIT), the Library, or the support staff at the DataX Institute). Nodes will be networked together through shared activities at the Institute, which will be organized by the Associate Director for Support. Note that we envision this networking through the Hub as entirely informal; node leadership will report within their Division or School, and will have no official reporting obligation to the Hub at the Institute.

Personnel: In many cases, staff at individual nodes will be part of existing units such as SSCERT, the QCBio Collaboratory, the Library's Data Science Center, or the Research Technology Groups in OARC. Staff will be **fluent in the language and culture of the disciplines comprising the Division or School they serve.** While the precise distribution of roles and responsibilities will vary based on local needs, we describe a typical node here. Some staff will aid faculty and student researchers with **data acquisition, curation, manipulation, access, and management.** Where appropriate, they will interface with the Institute and make the data they are using widely available. Other staff will facilitate faculty and student use of **existing data science techniques**, and **offer workshops, tutorials, mini-courses, or bootcamps** where appropriate. Nodes may also include consultants to help with **computing**, or refer people to the Office of Advanced Research Computing. Our interviews made clear that this expertise should typically be offered at no charge (i.e., faculty should not have to "buy" time through recharge). Faculty seeking extensive support should have the option to do so, but this should never occur at the expense of free, smaller-scale support.

Since April 2017, DSC has offered over 80 distinct workshops and events (two-day boot camps, short courses, events, or lab-type sessions) to over 1000 learners from over 80 departments, schools, or units on campus...We've taught two data science and coding multi-week workshops series for the Master of Social Science (MASS) program and worked with a Master of Urban and Regional Planning

(MURP) student group to put on a Planning Code Camp in winter 2018 that provided training in data literacy and R programming. **—Library Data Science Center WP**

Network: To create a community for staff in the DataX Support Network, maintain strong lines of communication, and sustain the campus knowledge-base, network staff will convene at "the Hub" for bi-weekly meetings and training sessions, where they can share new techniques, discuss issues, and build community. These meetings will be held at the Institute and involve Institute personnel; they will typically be convened by the Institute's Associate Director for Support. Quarterly, the leadership of network nodes will meet to coordinate and harmonize offerings. **The network will integrate individual staff in a larger community, and ensure that they remain up-to-date with applicable methods**. It will also allow network staff to "route" faculty or students to other network nodes when they require specialized expertise hosted elsewhere. To assess Support Network needs, we recommend that all relevant units (Divisions, Schools, and campus-wide functional nodes) conduct a self-assessment early in the process, in collaboration with the Homeworld/Institute. In cases where nodes already exist, the assessment should focus on emerging needs and opportunities for growth; in cases where nodes do not exist, the assessment should establish unmet needs, potential demand, and recommend a strategy for node-creation (if desired). The Homeworld/Institute can support expansion plans where advisable, and provide advice regarding the creation of new nodes as needed.

Reporting Structure: Each node should report to a representative of their respective Dean (e.g., the Associate Dean for Research) unless an alternative reporting structure is preferred. Node leadership will also work closely with each other, and with the Institute's Associate Director for Support. Again, the "Hub and Spoke" architecture is entirely informal, and the Institute node is distinguished only by its convening role. We encourage "lateral" collaborations and partnerships between Spoke nodes.

EDUCATIONAL INITIATIVES

"Million Dollar Hoods " a big-data research initiative co-led by UCLA Professor Kelly Lytle Hernández

Educational Initiatives

Opportunities have exploded for fulfilling careers that require some form of expertise in DataX. **New educational initiatives are needed** to offer students access to courses, majors and minors **at a suitable scale** to enable them to prepare for these careers, and more generally to be data-savvy informed citizens. A lot of good work has already been done in this direction at UCLA. Still more is in the planning stages. Nevertheless, a coordinated campus-wide effort is needed, to meet the scale of demand and to ensure that offerings are coherent, accessible, and navigable. There is an opportunity, indeed a moral imperative, in this emerging educational space to **ensure that the student cohort it attracts fully reflects the diversity of UCLA's student population.**

OIT offers over 30 non-credit training workshops, tutorials, and mini-courses, along with ongoing walk-in consultation services, to help faculty and graduate researchers increase their technical proficiency with the tools, methods, and techniques for data use across the research lifecycle. Its community of research and data scientists also serve as lecturers and in-class experts for the university's credit-bearing courses in Urban Planning, Digital Humanities, World Arts and Cultures/ Dance, and Urban Humanities. —**Office of Information Technology (now OARC) WP**

A variety of promising initiatives should be pursued in parallel, in addition to existing initiatives. Such efforts should be coordinated and rationalized, and our interviews made clear that—-while most initiatives would properly occur in departments, Divisions, and Schools—some degree of centralization was essential for coordination and planning. We envision the Institute as playing that convening, coordinating, complementary role. The **full panoply of the faculty's creative energy should be mobilized, with a strong level of institutional support** backing it up.

We have a newly developed Ph.D. program that begins this fall. Our students will be trained to conduct research on a variety of communication-related questions using tools and techniques relevant to DataX, including computational and formal modeling, data science and machine learning, experimental design and data collection, social network analysis and large-scale survey research. We see our students entering both academia and industry, possessing critical skills necessary for empirical work in the behavioral sciences.—**Dept. of Communication WP**

The core educational principles and curricular building-blocks of data science are still evolving; this is even more true for the expansive, pan-disciplinary mode of science, scholarship, and creativity indexed by our term DataX. Nevertheless, we can draw on some first efforts in this direction, most prominently a consensus study commissioned by the National Academy of Sciences[4] and a National Academies Roundtable[5] in which one of us (MG) participated. We can also draw on input from colleagues at and beyond UCLA developing courses and curricula in this space.

---

4 https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options
5 https://www.nationalacademies.org/our-work/roundtable-on-data-science-postsecondary-education

We propose that DataX education at UCLA should be guided by the following key principles:

- It should be **experiential and project-based**; above all, it should involve working directly with real data.
- It should take advantage of the ample "low-hanging fruit" accessible to undergraduate and graduate researchers, and prioritize **paths to hands-on student research.**
- It should present DataX tools and techniques **in context and in practice**, so that their benefits—and their limitations—are more apparent. Ideally, these contexts should "meet students where they are" and speak to their interests.
- It should incorporate a **rigorous and critical examination of the social impacts, ethics, and policy implications of how data is collected and used.**
- It should create **strong communities for student learners**, leveraging the well-known benefits of cohorts for retention (especially for students coming from underrepresented groups).
- It should provide some opportunities for **team science**, without sacrificing the value of individual learning.
- It should offer **learning pathways at multiple levels** of mathematical, computational, and social/ethical/policy sophistication, recognizing that there are opportunities and career paths for all such backgrounds.
- It should ensure that **there are educational pathways accessible to students without substantial prior mathematical and computational training.**
- It should provide opportunities for practical learning experiences, via **engagement with industry or with community organizations.**
- It should emphasize two benefits of DataX training: **career-readiness** and more **informed civic participation.**
- It should encourage **experimentation at a modest scale**, recognizing that this area is sufficiently new that the best curricular structures or pedagogical practices may not yet be worked out. Such experimentation should explicitly engage with and develop work on inclusive pedagogy in DataX.
- It should create a campus ecology that is **coherent and navigable** without promoting an overly uniform monoculture.

It will be an early and essential task of the DataX Homeworld to systematically develop consensus around key curricular components for DataX education at a range of technical levels, and to stimulate and cultivate educational experiments on and beyond campus. This is a core mission for the Associate Director of Education and an educational working group (which should include scholars with expertise and research interests in DataX-relevant pedagogy and curriculum development). We are particularly excited about opportunities to pursue federal funding for educational innovation, e.g., programs like the NSF's "Data Science Corps" (associated with the "Harnessing the Data Revolution" Big Idea) or the National Science Foundation Research Traineeship program; this can be done through partnerships between domain scholars and relevant experts from the School of Education & Information Studies.

We turn now to concrete modalities for educational innovation in DataX instruction.

We are particularly enthusiastic about the idea of seeing UCLA roll out several **new DataX yearlong cluster courses,** building on this established modality for interdisciplinary teaching to deliver a unique, UCLA-specific pathway into DataX. These clusters should be carefully designed so that they are accessible to students regardless of their previous mathematical or technical background. They should be available to students as soon as they set foot on campus. Clusters should be taught by charismatic

faculty. Clusters would be interdisciplinary, interweaving an experiential encounter with data science, its use, and its broader implications with a particular substantive setting (e.g., Sustainability and Environment; Social Justice; Inclusive Innovation; Digital Governance; AI & Medicine). Topics should focus on areas with strong student interest; where appropriate, they may involve partnership with community organizations or government (following the model of Million Dollar Hoods). Courses should also emphasize the growing importance of open data and cultivate best practices for curating, managing, and sharing data and code. Courses would be similar in spirit but not in detail to the standard cluster format. Instruction would progress over the three quarters from hands-on instruction with core techniques and ideas in Data X toward small research seminars and collaborative project teams. Regardless of the specific topic, clusters should offer interchangeable training on core methods and questions in DataX, so that students are not limited in their subsequent options by the choice of cluster. We also anticipate that these Cluster courses will be designed to fulfill a mixture of GE, quantitative reasoning, writing, and diversity requirements. Suitable support for creating and staffing these courses is needed; we envision this occurring primarily through the Institute, its Fellows, and its Postdocs, with the addition of graduate TA's and GSR's drawn from departments. The DataX Cluster model is more experiential than approaches used at some peer institutions, more connected to student interest, and more explicitly oriented toward building a community amongst its students. For all of these reasons, it reflects UCLA's distinctive values and offers a competitive difference with other peer institutions.

We also see a niche for a scalable, one quarter invitational course ("**Explorations in Data**"), along Berkeley's Data8 model. Since its primary function is to entice students to further training in DataX, the course should be taught by charismatic faculty with the capacity to attract students to DataX. Like the clusters, it should have extensive teaching support so that it can be hands-on and experiential.

We emphasize, in both of these cases, the importance (and difficulty) of creating compelling courses from scratch. We were cautioned repeatedly in our external interviews against trying to create such courses "on the cheap" by lightly modifying existing courses with a veneer of "data science paint." Doing this right will require a substantial investment of time, effort, and creativity. That said, the entire campus can benefit from the instructional materials (datasets, instructional notebooks, and other resources) produced. These are expensive to produce, but can easily be adapted and repurposed for other DataX courses across campus.

In order to reduce barriers to participation, we propose to establish DataX Computing Scholarships, which will aid students in acquiring a laptop with sufficient memory and storage for contemporary data science applications. The size of this program should scale with the DataX Cluster Courses, and should have as its goal the removal of this critical barrier to participation.

What if, instead of training undergraduates in majors defined around disciplinary specializations, we cultivated an interconnected set of computational, scientific, and cultural literacies? And what if we combined this training with critical problem-solving and ways of thinking, designing, and collaborating to imagine more inclusive, just, and livable futures? The heart of such a program would be project-based learning and collaborative research across a "third space" between the humanities, arts, social sciences, and professional schools on the one hand, and the physical and life sciences, engineering, and medical fields on the other. —**Division of Humanities WP**

In addition to these cross-cutting offerings, there is a clear need—-often expressed in the white papers—-for **new data science courses in individual departments**. Developing these could be incentivized by competitive GSR funds to develop course materials. We also heard considerable support for the idea of **data science "add-ons" for existing courses**, by analogy with foreign language add-ons. These "add-ons" would allow students to augment their standard coursework with some relevant experience in data-intensive scholarship.

To help students build their resumes with legible credentials, UCLA should explore offering **data science certifications** at a range of (technical) levels, for both undergraduate and graduate students. Our colleagues also found a related idea compelling: UCLA could offer "transcriptable options" in DataX, with explicit designations for "level" of technical skill (something along these lines is offered at the University of Washington). The Institute, in its coordinating role, would work to establish consensus on what sets of courses represent a comparable level of expertise.

In these courses, teams of students will exhaustively study individual manuscripts, collecting all available codicological, textual, and paratextual metadata under the guidance of ISP instructors. With support from DataX, they will then explore advanced techniques in the Digital Humanities to analyze and publicize these data. For example, they might map the movements of a manuscript from library to library between the fourteenth century and the present; test software for digital imaging, text recognition, and digital editing; or use network analysis to reconstruct transregional communities of authors, collectors, and readers.—**Islamic Studies Program WP**

The white papers document plans for a range of **new minors and majors** related to DataX. We anticipate that more of these will be planned as the DataX initiative unfolds. As we noted above, we are not (at this point) concerned about supply exceeding demand. Instead, the campus should act to facilitate development of high-quality offerings at the course, certification, and degree levels. In its leadership and outreach and advocacy role, the Institute can support such development through creating and disseminating templates through its working group on DataX education. In its coordinating role, the Institute (or, more likely, its precursor, the DataX Homeworld) can identify gaps in the educational ecology by leading an early effort to systematically document DataX majors, minors, graduate programs, and courses (existing and planned). This mapping effort will reveal gaps where investment might be especially beneficial. The Institute can also support these efforts by offering training opportunities focused on teaching technical material to non-technical audiences.

The general area of ML [Machine Learning] is especially attractive to students due to its rapid growth and plentiful (high paying) job opportunities...Ms. Laura Thapa is a graduate student working with Prof. Pablo Saide on the ML based prediction of wildfire emissions, and has recently been awarded a NASA/FINESST grant to work on this topic...Mr. Gavin D. Madakumbura is a graduate student working with Prof. Alex Hall on using ML in explaining controls of forest drought resistance and for tree mortality modeling. Also, he uses ML and their interpretation techniques to identify systematic biases in climate models. —**Atmospheric and Oceanic Sciences Dept & Earth, Planetary and Space Sciences Dept WP**

Online courses (and even degree programs) **represent both an opportunity and a challenge.** On the one hand, such courses and programs can provide tremendous value to learners (through providing essential skills, critical credentials, etc.); on the other hand, online education requires distinctive approaches and pedagogy to take full advantage of the medium and provide a high quality learning experience. While we envision departments taking the lead in the development of online education in the DataX space, the Initiative can support such efforts in two ways. First, the DataX Homeworld and eventual Institute can help to identify niches in the national and international educational ecology where UCLA has particular comparative advantage (e.g., in medicine, entertainment, urban sustainability, causal inference). Second, the Homeworld and Institute can facilitate the development of resources that make it easier to stand up new courses and programs (e.g., instructional datasets; robust Jupyter or CoLab notebooks) while also convening a working group focused on best practices for online education in DataX.

DataX at UCLA will provide opportunities for student involvement in research at many levels. Indeed, this is one of its most appealing features: it allows our students to take advantage of their enrollment at a world-leading research university and gain transferable, hands-on skills in data science, writing and speaking about data, etc. The campus should encourage the development of **vertically-integrated research experiences**, involving undergraduates, graduates, postdocs and faculty. Funded **summer research experiences** for undergraduates, such Bruins In Genomics or IPAM's Research in Industrial Projects for Students, can grow in scale. **Capstone courses** - including substantial research projects- will become a hallmark of many DataX majors.

Internships with local companies, non-profits and government are another important component of fostering student engagement with Los Angeles. In particular, we propose the establishment of a DataX LA Internship program, which would provide a 50-50 match for student internships with LA City and County government and local NGO's.

Effective instruction in data science depends on **high quality instructional datasets.** Ideally, these should be "real-world" rather than toy data (although they may need to be cleaned more extensively, especially for introductory courses). The Institute and Support Network should work together with faculty to support the **creation, curation, and sharing of instructional datasets**. For example, some data in the data commons may be selected for further cleaning and enrichment so that they can be used for teaching. The same applies to code; we hope that the Institute will serve as a central hub for aggregating and disseminating useful instructional code, particular when that code is offered in a pedagogical format like a Jupyter or Google Colab Notebook. Note that such investment will make it much easier for individual units or instructors to launch DataX courses, as they can draw on a common pool of pedagogical resources. As the DataX curriculum proliferates at UCLA, we imagine an "**ecological succession" of educational offerings**. For example, local development of specific courses or minors could progress to cross-cutting introductory sequences and eventually to the formation of Centers for Interdisciplinary Instruction (along the lines proposed by BioDataX in their White Paper).

Navigation: To empower students to make best use of DataX opportunities at UCLA, the Institute will create a **student portal** for the university's DataX course offerings. This portal should delineate possible pathways that prepare students for the available certificates, majors and minors. Ideally, it should offer "pathway building" tools, so that students can plot out possible paths through the DataX ecosystem. The student portal could also allow for one-stop-shopping for student research opportunities in DataX, drawing on information in the Institute Research Portal and the existing portal for undergraduate research. Skills needed for particular research opportunities can be cross-referenced to pathways in the portal,

so that students can see what opportunities are currently a good fit, and what further coursework would be necessary to pursue opportunities in the future. Development of the portal will be time- and labor-intensive; we therefore recommend that the DataX Homeworld take it up as an early project. At the same time, we recognize that no portal—no matter how sophisticated—can substitute for **effective advising in DataX.** We therefore encourage a close partnership between the DataX Homeworld and the Division of Undergraduate Education. Regular meetings (e.g., every quarter) with representatives from undergraduate advising will ensure that college advisors are aware of the shifting landscape of DataX education and able to effectively assist students in navigating it. If possible, some advisors should develop special expertise in DataX offerings. It is especially important that advising be offered around the cluster courses as students begin their time at UCLA (e.g., during orientation, but also as something to mention during campus tours); this is a critical strategy for recruiting a broad and diverse group of students to these new educational opportunities.

Rationalization: Our interviews strongly suggested the need for campus-wide assessment of the existing ecology of courses, focusing on the identification of gaps and overlaps. Because it draws on the faculty most committed to DataX research and teaching, the Institute is well positioned to perform such an assessment, through its standing working group on education, and informed by the data-gathering efforts described above. When **gaps and overlaps** are identified, the working group can suggest solutions and work together to implement them. In some cases, it will be valuable to develop **bridge courses** that enable a student to transition from one level of expertise to more advanced training in core data science disciplines. Such bridge courses should be offered in Summer Sessions and perhaps through University Extension, in addition to offering them during the academic year. Where appropriate, the Institute will facilitate **cross-listing of data science courses** across campus and advise departments on which courses constitute appropriate preparation to embark on a given course. We envision the Institute as serving a "complementary" role to departments in course development; it is distinctively positioned to develop bridge courses that serve as transition pathways to multiple majors (e.g., from the cluster courses), as well as highly interdisciplinary courses (at the graduate and undergraduate level) that may be outside the scope of a particular department. An important long-term goal is the provision of a substantial number of rigorous DataX courses through Summer Session and UNEX, which will broaden participation in DataX education and lower barriers to entry.

Articulation: While this section has focused on educational innovation at UCLA, we recognize that there is a critical challenge in the articulation between offerings at UCLA and offerings at the K12 schools and community colleges that prepare our undergraduate students. There are exciting opportunities to experiment with articulation by building on the partnerships that UCLA's School of Education & Information Studies has established with its "community schools" (the Geffen Academy and the Mann UCLA Community School) and with some of our most important "feeder" community colleges (Santa Monica College and Pasadena City College). On the K12 side, we envision the development of DataX-intensive classes and project-based learning opportunities, while also taking advantage of an existing SEIS teacher residency program focused on STEM education. On the community college side, it would be exciting to develop transfer pathways that allow students to transition seamlessly into the DataX ecology, including paths into the more technical aspects of DataX education. Reaching beyond our immediate community, we encourage the campus to build on successful efforts like the CenterX "Introduction to Data Science" course and the extensive role of the Department of Mathematics in training high-school mathematics teachers (perhaps by offering courses that incorporate data science suitable for inclusion in high-school math).

In collaboration with the UCLA American Indian Studies Center and the Bunche Center, AIS faculty have created a website showcasing data on COVID-19 cases on American Indian Reservations and Alaska Native Villages using publicly available data. We have transformed these data into a user-friendly, interactive resource for tribal communities, leaders and policy makers...In current work, faculty are also using cellphone traffic to identify movement on and off American Indian reservation locations over the COVID-19 closures and subsequent openings by state, county and tribal jurisdictions. We are the first to use this data in this context and believe it is a useful tool for assessing COVID-19 cases and transmission routes for small and hard to count populations.
—American Indian Studies WP

Diversity and Inclusion in DataX Education: It is essential to make the opportunities implicit in data science careers available to a broadly inclusive group of students, providing a fresh approach to acquiring quantitative skills and reaching out to enhance their awareness of the opportunities offered by training in DataX. Possible models include Kelly Lytle Hernández "Million Dollar Hoods" project and clusters, courses, or themes like "Data Science for Social Good," "Data Science for Sustainability," and other substantive topics that are likely to energize and excite students. The DataX Cluster Courses are inspired by this idea. Events such as the pandemic-postponed visit of Stuart Russell in March 2020, focused on "Human Compatible Artificial Intelligence," can raise the profile of data science around campus and energize faculty and students. The DataX Distinguished Lecturer series is inspired by this idea. The visibility of the Institute will provide opportunities to forge partnerships with industries reliant on data science; those partnerships can catalyze research on campus and provide internship and employment opportunities for students. Similar opportunities are available in partnership with local governments and community organizations. The DataX LA Internship program is inspired by this idea. We emphasize that our goals of inclusive participation in DataX also require recruiting a diverse cohort of faculty mentors, so that students can see themselves in DataX roles. Close partnership with *Rising to the Challenge* and activities related to the campus commitment to becoming a Hispanic Serving Institution can ensure that equity, diversity, and inclusion goals in DataX education are met and exceeded. One broader structural innovation worth pursuing is the development of 3+1 sequences, in which students can complete an undergraduate degree having substantial DataX content in 3 years, and then proceed immediately to a 1-year masters. Because students are still eligible for federal financial aid throughout the 3+1 years, such a structure (currently explored by the Mellon-funded Social Justice Curriculum and the Masters in Data and Society (MIDAS)) makes it much easier financially for students to secure a postgraduate qualification that attests to valuable skills unlocking many employment opportunities. Finally, for some students, it is a financial challenge to acquire a laptop suitable for DataX courses; the DataX Computing Scholarship program is designed to provide help to some of these students.

"Intergenerational Transmission of Stress": The Mothers' Cultural Experiences Study is a longitudinal study that follows Latina women in Southern California across pregnancy and postpartum with the goal of understanding how cultural and biopsychosocial stressors and experiences are transmitted generationally...The extensive biological data from three body fluids reflect endocrine, inflammatory, metabolic, and genetic maternal-paternal-fetal functions. **—Behavior, Evolution and Culture WP**

INSIGHTS FROM PEER INSTITUTIONS

UCLA Powell Library

Insights from Peer Institutions

In this section, we briefly summarize the insights we gleaned from colleagues at peer institutions. We should emphasize that these external interviews largely confirmed and refined the strategy we had developed through broad consultation with stakeholders at UCLA. They also motivate the scale of investment that we propose; although substantial, this seems to be what it takes to achieve campus transformation. Because UCLA is acting "later in the game" and trying to leapfrog peer institutions whose programs are already well-established, the investment required is almost certainly larger than that made initially by earlier entrants. While we must "go bigger," we also have the advantage of investing smarter, learning from the successes and challenges of early movers.

Peer institutions generally use the language of "data science" and we follow that practice here.

Overall assessment and scale: Our colleagues framed data science as an existential issue, and broadly agreed that it was part of the basic intellectual infrastructure of a modern research university. Although the vast majority of these colleagues came from the physical or computational sciences, they emphasized that the biggest opportunities for growth and catalytic change came from fields that currently have a lower level of data science activity (e.g., the social sciences, the arts, the humanities, law). Core investments range widely, but are typically of transformational scale; several universities have raised $100M to $500M in this space and envision overall commitments up to $1B. At some campuses, research and education grew on parallel tracks, but it was widely agreed that these should eventually be united.

Hiring and personnel: It was broadly agreed that people were the most important resource for a successful effort in the DataX space. A frequent strategy involves **postdoctoral fellows** who link method and domain, in both their expertise and their mentorship. Depending on the funding source, postdocs may have a modest teaching load, and range from 10-12 (research only) to 20+ in number. Our colleagues noted their essential role in circulating expertise between interdisciplinary data science units and departments. Colleagues also noted that postdocs and faculty lines were important to get campus buy-in. The **most typical faculty model involves some faculty with a half-time appointment in a data science unit plus a larger group of affiliated faculty**. Some campuses envision full faculty appointments in this space. One campus anticipates a model in which faculty can temporarily move FTE into the data science unit [6]. Attracting excellent faculty can be challenging. Incentives include: funding for PhD students and postdocs; appealing space for scholarly interaction; and intellectual community, coupled with relevant in-house expertise. Colleagues repeatedly emphasized the parallel importance of mechanisms for keeping people engaged, and severing relationships with faculty when their DataX engagement falls off. Hiring models varied, across campuses and across time. In some cases, hires are "Institute-driven," while other models allow units to pitch hires to the data science unit. Many colleagues underlined the importance of non-faculty expertise of various types, with some going so far as to say these were more important than new faculty lines. Peer institutions hired developers, paying them salaries that compete with industry. Some staff will provide research support; these staff must have strong people skills as well as data science skills, and assist faculty and students with access, management, and analysis of data. Many colleagues noted that this expertise is provided free, rather than for hire; this was viewed as essential. We were also encouraged to take advantage of joint hires with industry and national

---

6   We note that some colleagues at UCLA suggested something similar; we think this is worth considering seriously.

labs (e.g., Lawrence Berkeley National Lab). National lab positions may be especially appealing because data science for science is often more intellectually interesting than data science for industry.

Structure & Funding: The most common model is some sort of cross-campus Institute, which provides a campus focus for interdisciplinary research. The scale can vary, from a more modest Center all the way to a Division (Berkeley) or School (MIT). **Many campuses have a reporting structure in which the Institute does not belong to a particular department or division; some colleagues recommend that it report to a campus-wide administrator.** Our informants universally emphasized the value of substantial dedicated space, which can range in size from the floor of a building to an entire building. Such space, and the personal interactions it promotes, lower the activation energy of collaboration. Colleagues emphasized that this needs to be "real" space, not just a meeting room. At many campuses, these spaces are extremely active; it would be useful to model any UCLA investment on the most active spaces.

Across campuses, it was more challenging to build community among the faculty than to involve students in data-intensive education (because the students are hungry for such experiences). At some institutions, grant applications provide an important lever for bringing people together, but this is not effective across all disciplines. A central Institute can also be effective in getting data-adjacent centers off the ground, with one Institute even providing physical space to related efforts during their launch phase. Colleagues generally advised a "grass-roots" approach to research, rather than imposing themes from the top. The greatest success came from amplifying existing strengths. Colleagues also recommended inclusive leadership, drawing on many disciplines and personal backgrounds. These leaders need to have a solid cross-campus reputation, to enhance the legitimacy of decision-making. Several campuses observed the same "two cultures" we saw, with some faculty requiring support and capacity building and others oriented toward interdisciplinary data-intensive scholarship.

Just as in our campus interviews, few colleagues talked extensively about computing infrastructure. In some cases, compute resources were provided centrally, with hardware on campus; in others, cloud options were successfully pursued. There was no consensus on which model was more cost-effective, though colleagues mentioned the importance of easy access to compute.

Funding models varied; some campuses started with substantial infrastructure grants, others with central funds as part of a campus strategic focus on data science. Continued support came from grants, overhead return, and standalone MA programs administered by the data science unit.

Activities: We heard about a range of useful activities for stimulating engagement with data science. Many campuses launched working groups with a range of foci, e.g., teaching, substantive research topics, and methodological topics (e.g., text data, time series, image data). These faculty working groups are further supported by postdocs or Industry partners. Some institutions provide modest internal grants to jumpstart collaborations. Institutes hosted bootcamps, short courses, and hackweeks focused on particular disciplines. They also provided training on how to teach technical material to non-technical audiences. Libraries often played a critical role in helping to manage and serve data, in collaboration with disciplinary expertise. Data needs are multi-faceted; it is often necessary to buy important data (with staff who can acquire, clean, and provide access to such data) but also essential to provide easy access to the data-intensive products of the university.

Education: Many models and modalities have been tried. Student demand is high, beyond the resources of any one department, and 10X enrollment changes are common. There was a consensus that one needs to make new courses, and not "spray paint" existing courses with a veneer of data science. We saw several different models: a model with a single comprehensive intro course followed by more specialized "spokes," exemplified by Berkeley's Data8 course; a model in which many units developed new majors or minors; a model in which a single unit developed a major, a master's degree, and a PhD program; and a model that started with upper division courses and "transcriptable options" with stipulated competencies. Certificate programs, summer programs, and short courses are all part of the picture, as are "connector" courses that help students transition to more demanding technical material. Colleagues noted that centralization can make it easier for great course material to be developed; they also mentioned the value of training instructors to teach technical material to non-technical audiences. It is a challenge to incentivize course development. Some have used GSR's to develop new course material; others use data science teaching fellows or teaching faculty. It is important to firewall the people who provide support for research from demands that they support instruction as well; their time can easily be consumed by such student requests

Challenges: What are some challenges that peer institutions have encountered? There have been a number of issues centered around the definition of data science, with sometimes an overly restrictive conception of who is a data scientist. This has resulted in a very skewed disciplinary representation, and the data science effort has sometimes been subsumed by an existing department. It was in some cases necessary to dispel the myth that data science is the same thing as high performance computing, or that researchers always need massive computing resources to engage in data science. The tension between disciplinary and interdisciplinary research was hard to resolve; likewise the related tension between centralized and decentralized models. Getting stable funding and continuity of support from the administration were ongoing challenges, especially for places that had an initial outside infusion of funding. The rapid evolution of the field made it hard to stay at the cutting edge. Joint industry/ academic appointments do not have an agreed-upon template; at the same time, they were viewed as essential to retain talent. Attracting a diverse group of faculty and students required sustained effort and focus. Colleagues repeatedly warned against strong "top down" models, and urged us to get things started without waiting to recruit a high-profile leader for campus efforts. Finally, it took time and effort to transition faculty from a "zero sum" mentality to a sense that resources directed to data science benefited the campus as a whole.

RESOURCES AND FUNDING
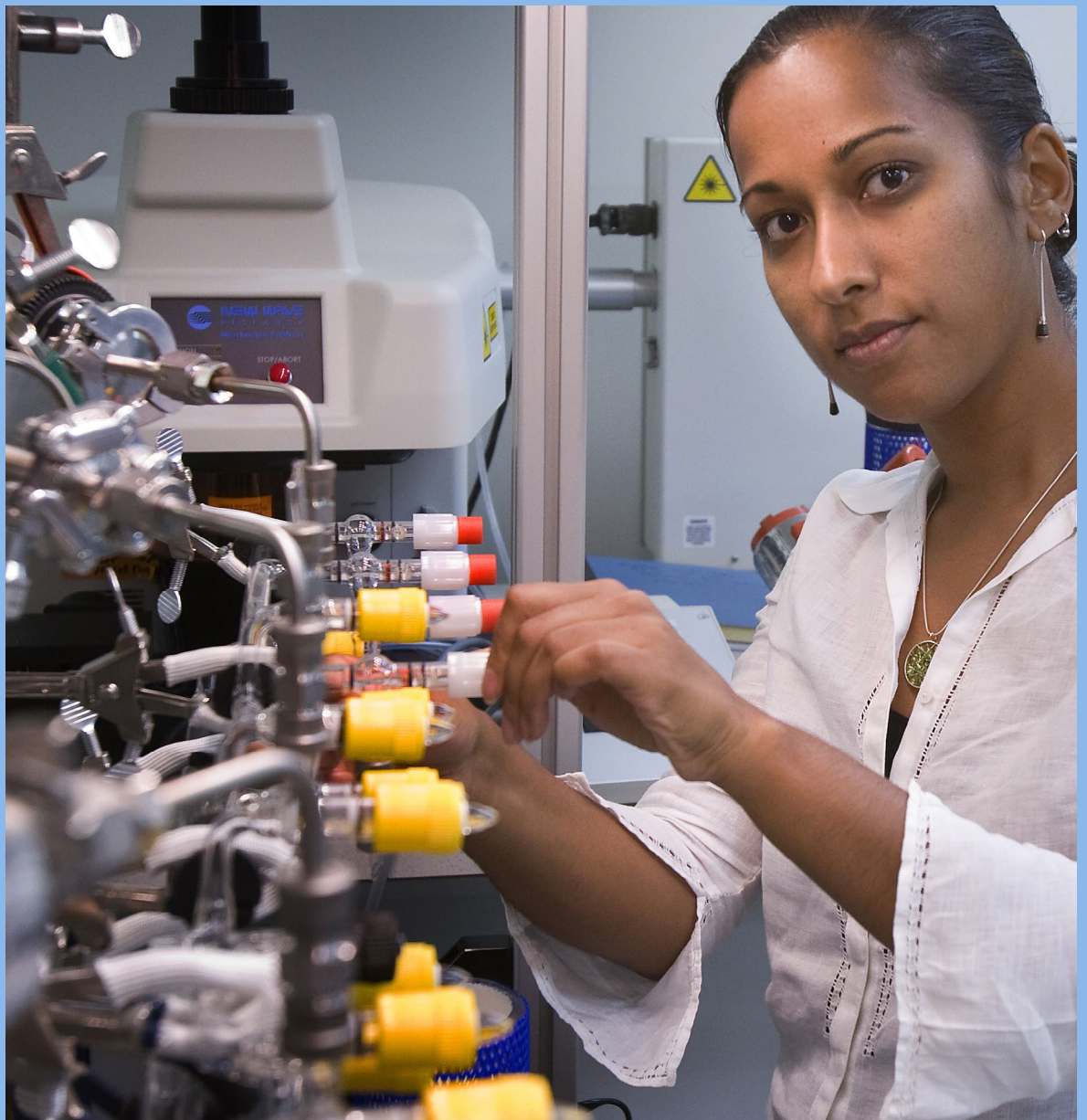
Professor Aradha Tripati - UCLA Department of Earth, Planetary and Space Sciences

Resources and Funding

It is essential to plan for financial sustainability from the very beginning. Sustainability has been a major issue in data science efforts at peer institutions, even—and perhaps especially— at places that started out with a substantial initial infusion of funding such as the Moore-Sloan Data Science Environment grants. **It is also essential to build in a mechanism to ensure that the intellectual footprint of the Institute can evolve over time,** and not become locked into paradigms that are no longer cutting-edge. **Finally, it makes sense to build up DataX in stages, rolling out funding over time. This will allow for growth that builds upon a solid foundation but is nimble enough to incorporate lessons learned.**

Institute Resources

Institute FTE: We have proposed that substantial FTE be allocated to the Institute. This is consistent with activity at many peer institutions, which are hiring either to support expansion in particular areas (e.g., data science in the humanities and social sciences at UW) or across the board (e.g., 50 new positions associated with the Schwartzman College of Computing at MIT). **The DataX Institute FTE would be used to hire faculty with ½ FTE coming from the Institute and ½ FTE from the home department. After 10 years, the home department would have the faculty member full-time, and the Institute's ½ FTE would return to the pool of Institute FTE.** Faculty would do half of their department's normal teaching load in the Institute. This teaching would span the DataX cluster courses, the lower-division core, undergraduate capstone projects, and graduate seminars. Courses will be interdisciplinary in scope and broadly accessible, but some could be "homed" in particular Divisions or Schools, e.g., a seminar on legal dimensions of data offered in the Law School but open to graduate students from across campus. The precise contours of Institute teaching obligations will evolve as the overall DataX educational ecology evolves; the purpose of the Institute teaching is above all to stimulate the creation of a robust set of broadly accessible, interdisciplinary DataX courses at many levels. Faculty would initiate and actively participate in Working Groups at the Institute.

This model of FTE allocation is quite similar to that used by the International Institute. The key difference is that DataX FTEs revert to the Institute after 10 years. Although this model is unusual, we believe it is required by the dynamic nature of the field. There is a risk, in fast-moving fields, that we "lock in" the expertise and focus of a particular moment in time. Having FTE revert after 10 years will support continued hiring and thereby allow the Institute's intellectual focus to evolve over time. Of course, faculty whose FTE have reverted may choose to retain their Institute affiliation as a Fellow.

Each of the three cohorts—foundational data science, innovative applications, and data cultures, ethics, policy, and societal implications —would have a search committee, appointed by the Institute Executive Committee. Each search committee would work in coordination with departments interested in hiring in the DataX space. Again, there are examples of this process on campus, e.g., hiring by the California Center for Population Research or by American Indian Studies. We believe the most promising approach is that pursued by *Rising to the Challenge*. In this mechanism, departments would submit proposals to the Institute Executive Committee, which would select some proposals for support. The appropriate Institute search committee would join the departmental process (either making parallel recommendations or acting as a single, joint committee); hiring would depend on a positive departmental vote and approval

by the Institute Executive Committee. This model balances bottom-up initiative by individual departments with coordination through the Institute to ensure a strategic and continually updated distribution of expertise across campus.

When we surveyed what peer institutions have done in creating a data science program, it was striking how rarely these programs were fully multi-dimensional. **We view it as being of the utmost importance to achieve the correct balance.** As a core intellectual development, the "data revolution" has been driven by interaction between fundamental data science and new applications. More recently, scholars and practitioners have realized the essential importance of critical considerations of the social, ethical, political, policy, and cultural dimensions of data and its uses, in many cases drawing on long-standing critical traditions in science studies, information studies, feminist scholarship, etc. This three-fold emphasis—on fundamentals, applications, and social/ethical implications—has appeared in both federal and philanthropic funders, e.g., the recent NSF AI Institute calls. At the same time, peer institutions that "moved earlier" into the data science space often fail to integrate all three dimensions.

Since we are recommending something that will involve disciplines from across the entire campus, **we view it as essential to have certain guard rails in place to achieve and then maintain a reasonable balance between the three general areas of fundamental data science, applications, and ethics, policy, data cultures and societal impact.** It would be all too easy for the balance to drift in one direction or another, and appropriate structures are essential to restore the balance and avoid further reinforcement of a particular area's dominance. **We place the highest priority on building a widely interdisciplinary community, in which fundamentals, applications, and social/ethical/policy dimensions inform one another continuously.** Infighting over who gets more positions would be inimical to this goal.

We understand and indeed anticipate that many individuals—both those currently at UCLA and those hired with DataX FTEs—will transcend this simple classification. In any case, in looking for candidates to hire for DataX, a key qualification will be excitement about interacting with the rest of the DataX community. We believe that interdisciplinary collaboration and creative thinking will be essential in hiring, not just in the social, ethical, and policy pillar but across all three. We understand and anticipate that in any given year, there may be an opportunity that temporarily throws off the equal balance between the three groups. A certain amount of "horse trading" will be natural, so long as the balance is restored in the relatively near term.

The precise mechanisms by which this balance will be understood and maintained is something that will be part of the charter of DataX, which will be one part of the discussion with the Academic Senate and the EVCP when the exact place of DataX at UCLA is worked out.

Because hiring with DataX FTEs represents a long-term investment and an aspirational vision, we view it as essential that those hires roughly balance across the three areas—that is where we'd like to end up, and that is where the field aspires to be. The postdocs, which represent a key "glue" for the DataX community and a link between the three areas, also need to reflect that aspirational balance across fundamentals, applications, and social/policy/ethical implications.

Other elements of the Institute, e.g., existing UCLA faculty who become DataX Faculty Fellows and who become part of DataX working groups, are more likely to reflect short term interests and opportunities, and must have the flexibility to "drift" from a strong balance as opportunity dictates. That said, the governance

structures we describe in the report are designed to give the three areas equal input on critical aspects of DataX, its activities, and its investment, as a further hedge against the "capture" of DataX by particular disciplines, departments, or areas. Having seen this happen at peer institutions, we thought it essential to build in mechanisms that preserve and protect the role of all three pillars. We intend that the DataX Executive Committee will play a critical role in interpreting and maintaining balance across the three pillars, and have expressly designed its structure to ensure equal representation and voice across those pillars among the elected members of the Executive Committee.

That said, it is important for DataX to be intellectually (and otherwise!) inclusive and to maintain fluid boundaries. **We do not want to rule out people, ideas or research threads preemptively.**

We propose that 30 ladder FTE be allocated to the Institute. These would be rolled out according to the following timetable: Each year, 3 FTE (= 6 hires) would become available, with 1 FTE (= 2 hires) allocated to each of the 3 cohorts. Unfilled hires would roll over and add to the pool available in the next year. While balance among cohorts is essential over the long haul, we anticipate cohorts will on occasion be flexible as to which year their positions will be filled, both to allow unanticipated opportunities to be seized, and to enable cluster hires where appropriate. **Among the very earliest rounds of hiring, we would advocate making at least 2 hires with Institute FTE of faculty whose research focus and educational interests center explicitly on racial equity and racial justice,** in alignment with UCLA's *Rising to the Challenge* initiative and its plan for becoming a Hispanic Serving Institution. Indeed, we encourage close collaboration on hiring between DataX and these initiatives, as well as the recent "Mentor-Professor" program, which aims to recruit faculty with a particular interest in mentoring students from underrepresented groups. Such alignment and collaboration will ensure that DataX faculty who may belong to underrepresented groups themselves are part of "cohorts" or "cluster hires" within the Institute and across the campus; such cohort/cluster mechanisms are known to be particularly effective in recruiting and retaining faculty from underrepresented groups.

Over 10 years the 30 FTE would be filled by 60 faculty associated with the Institute. Because each FTE would eventually revert to the Institute, this pace of hiring could continue indefinitely with no further FTE being added to the Institute's allocation. This arrangement makes explicit what a department is committing to in participating in these shared FTE; somewhat like the Presidential Postdoc program, departments would participate knowing explicitly that they would hold a full FTE after 10 years, gaining all of the colleagues' teaching but increasing their FTE count accordingly. This model does require that Deans ultimately take over the half FTE, but this will occur gradually, and we anticipate that enrollment in the DataX courses created within Schools and Divisions by Institute faculty will help close the gap. We foresee, in the implementation phase, that a template will be crafted that lays out carefully and explicitly the expectations that the Institute and the department would have for these faculty, with a view toward ensuring that faculty can know in advance how these positions will operate. The Academic Senate can play a critical role in creating this template and associated processes around merit increases, tenure, and promotion.

We believe that this plan will allow a **gradual and thoughtful expansion in DataX, at a scale appropriate to the fundamental role we expect it to play in campus-wide research and education at UCLA. Our approach will have a transformative effect over time, without unduly taking resources away from disciplinary hiring.** We have devoted considerable thought to the number of faculty appropriate to this effort, given the intellectual footprint and disciplinary breadth we envisage for DataX, the number of Departments, Schools and Institutes with which it will interact in important ways, and the

level of engagement with students that we foresee. The overall investment proposed is commensurate with that made by peer institutions, some of which do not have Schools of Medicine, Business and Law.

At the same time, our proposed approach— in which no faculty would have an exclusive Institute appointment and all faculty hired through this Initiative eventually integrate into their home departments— emphasizes and cultivates the deep connection between DataX activities and existing disciplinary practice. It also allows UCLA's faculty expertise in DataX to change organically as new techniques, applications, and social dimensions arise. We will always be able to advance cutting-edge scholarship at the Institute through recruiting leading scholars across career stages. We envisage these hires as occurring across career stages as needed; while in many cases a junior hire may be most appropriate (especially to secure expertise in emerging areas of DataX scholarship), at other times a senior hire may be required to build out a new area or secure our position in the face of retirements.

It is important to use these shared FTE to bring to UCLA as many new faculty as possible. DataX needs a core of committed faculty to achieve its mission in research and education. Why share them 50/50 with a partner department? There are several reasons: (1) It is essential that the faculty DataX hires meet the standards expected of UCLA faculty in at least one discipline; (2) The involvement of DataX in these hires promotes strategic coordination in hiring in DataX, taking into consideration the needs of the campus as a whole beyond the needs of any individual department; (3) Sharing the FTE keeps a good balance between centralized and distributed resources; (4) Because of the wide range of fields in which DataX might hire, we need a partner department to advertise the position among the appropriate communities; (5) The cost of set-up funds varies widely among departments, and by making the partner department responsible for these, DataX is free to consider each appointment without taking set-up costs into consideration. As with any split appointment, being evaluated by two units can introduce additional complications, and Institute Leadership must make sure that this does not present an undue burden on individuals recruited with such FTE, e.g., through the details of a standardized memorandum of understanding between DataX and the partner department.

We are aware that the proposal to have the full FTE revert to the partner department after 10 years is new. When administratively allocated FTE are "given" to a department for a particular purpose, it usually happens that after a few years, those FTE are counted in deciding on future allocations. The proposed policy makes explicit how this will work. Because of the rapidly changing evolution of the fields of research embodied in DataX, we felt that this arrangement was appropriate. To give just one example, the "deep learning" revolution is exactly ten years old as of 2022. The initial responses to this proposed reversion mechanism have been quite positive, and we are hopeful that this could be a model for other interdisciplinary efforts. It is our expectation that the appropriate bodies in the Senate and in the administration would need to discuss this idea as part of the process of writing a charter for DataX.

Today's large-scale proliferation of "big data" presents incredible opportunities for public health. Large data collection efforts within health systems promise to advance our understanding of how genetic traits and environmental factors influence common disease and condition status. In addition to the standard collection of the electronic health record and genomic data, many studies now use wearable devices that allow frequent measurement of phenotypic information. These biosensors produce digital data at an unprecedented scale of size and complexity. These data pose a significant big data challenge; developing tools and techniques to access, organize, and glean discoveries from them is an urgent unmet need. **—Dept of Computational Medicine WP**

A second mechanism for staffing Institute courses is through **teaching replacement buyouts**, involving Institute Fellows across the three cohorts. These buyouts (sometimes described as "internal rate") would supply replacement funding to cover a course in the faculty member's home department at an entry-level salary. This is quite similar to the mechanism used by the existing cluster program. The frequency with which Fellows can participate in these teaching buyouts in response to Institute educational priorities should be described in a standard memorandum of understanding between the Institute and the home department, to make the process as predictable, painless, and equitable as possible.

Incubator and "Doorway" Grants: Each quarter, the Institute would solicit proposals from interdisciplinary groups for an **Incubator Residency at the Institute. These internal awards would provide teaching replacement buyouts for 6 ladder faculty (who might bring along postdocs and graduate students) to spend the quarter in residence at the Institute.** These Incubator Residencies would be similar in cost to existing transdisciplinary seed grant programs, and would allow faculty to launch new DataX projects (which we anticipate would grow into Institute Working Groups). Faculty participating in a residency would also be obligated to lead a reading group or informal seminar related to their project. Such Incubator Residencies would be available to Institute Faculty Fellows in good standing. They have a close analogy in the practice of buying out UCLA faculty from certain departments at internal rate to participate in IPAM Long Programs.

Much of IDRE's strategic plan aligns with the components of the DataX initiative aimed at enabling and promoting innovative research in data science as well as in pipelining faculty and research groups to move up the ladder in data science competency. Another important aspect of the IDRE strategic plan is its recognition that data science and computational science both rely on advanced computing infrastructure, algorithms and software, and that multidisciplinary teams of experts are needed to tackle complex data science problems.
—Institute for Digital Research and Education WP

We also propose to offer **"Doorway" residencies. These internal fellowships would allow faculty who want to transition to a deeper use of data in their research to spend a quarter at the Institute to jump-start the process.** Such buyouts, at the internal rate, will only be available to faculty who are not current Fellows. Indeed, we view the "Doorway" residency as a critical step in the pipeline that connects faculty whose research is largely supported by the DataX Support Network to those who engage with cutting-edge data science at the Institute. Faculty who complete successful residencies would be prime candidates for affiliating with the Institute as Fellows. This model is similar to a residency program at the UW eScience institute, but adds the critical teaching buyout (in analogy to buyout programs at UCLA's CCPR) to give faculty time to focus on their new projects.

We propose twenty-four 1-quarter replacement buyouts each year and six 1-quarter GSR's. These would typically be allocated to three Incubator Residences (6 faculty buyouts + 2 quarter GSRs each) and six "Doorway" Residencies (2 faculty buyouts per quarter).

We have proposed internal-rate teaching buyouts in several contexts, and we would like to indicate how we think these would work out cumulatively. In the case of DataX cluster courses, these buyouts would provide a net increase in the number of courses being taught at UCLA, with the buyouts envisaged as funding a course in the faculty member's home department. We would anticipate that departments will see students subsequently attracted to their offerings when their faculty participate in teaching DataX

cluster courses. The replacement buyouts for Incubator Grants and Doorway Grants would be neutral in terms of total course offerings, but would entail potentially replacing a course taught by a senior-level faculty member by someone at a more junior level. In the case of Incubator Grants, the benefit to the department and the university would come from the formation of a new interdisciplinary working group, with a research program that would offer opportunities to postdocs, graduate students and, when appropriate, undergraduates. In the case of Doorway Grants, this would provide a faculty member with an opportunity to refocus their research in a more data-intensive way, which would have benefits for their students and their department. At any given moment, the impact of replacement buyouts would be small given the campus-wide scale of the Institute's activities.

Institute Postdoctoral Fellows: We propose an **eventual cadre of 30 Institute Postdoctoral Fellows. These will be divided evenly across the three cohorts in their primary affiliation.** These would be 3-year positions, rolled out over the first 9 years by hiring 6 the first year, 6 the second year, then 6, 8, 8, 8, 10, 10, 10, resulting in totals of 6, 12, 18, 20, 22, 24, 26, 28, 30. Once again, after considerable reflection on what we learned in preparing this report, we concluded that this is the appropriate scale for the activities of the Institute. DataX is an area of prolific intellectual ferment, and having a constant infusion of exuberant energy, new ideas and diverse perspectives will be crucial lifeblood for the Institute. The scale is also comparable to the activity at some peer institutions. The postdocs will have responsibilities for teaching and research. They will provide a crucial layer in Working Groups, connecting together the vertically integrated research teams of faculty, postdocs, graduate students and undergraduates. These Working Group teams would provide one source for undergraduate capstone projects in DataX. This gradual rollout would allow time for the program to become widely known and would grow apace with the expansion of research and education at DataX at UCLA. Depending on the teaching load, these appointments may need to be formally constituted as "Adjunct Assistant Professors" (in parallel to similar positions in mathematics). Institute faculty will be able to buy out some of the teaching responsibilities with grant money, helping to further integrate postdocs with Institute research.

TA's and GSR's for Instruction: Instruction in DataX will largely be project-based. Students will need ongoing support in writing code, using databases, cleaning, analyzing, and visualizing data, etc. As student enrollments grow, the Institute would work to establish a reasonably predictable number of graduate TA's necessary for its instructional mission. **TA's would be hired from departments across campus. This would allow departments to expand the number of graduate students that they can support, and to attract graduate students with DataX interests and capabilities.** Note that these TA's would be responsible for helping students with course projects. The DataX Support Network staff would be available to help students with research projects (even if these are undertaken for course credit as independent studies, honors theses, etc.). This parallels the distinction currently drawn around course TA's and stats consulting or other support provided by OARC Research Technology Groups.

The Institute would also offer GSR positions for students to curate instructional databases, create Jupyter notebooks to be used in courses, etc., When students undertake this work, their contributions would be clearly documented, thereby enhancing the instructional side of their resumes. We envision these GSR positions being offered to support instructional development beyond the Institute educational ecology; in essence, the Institute would "sponsor" GSRs to develop materials for DataX courses in departments across campus.

Datasets: Funds are needed for obtaining, curating, and securing datasets of cross-cutting interest so that they are widely available on campus for research. In some cases, this would involve negotiating

agreements to purchase external datasets. In other cases, it would involve developing and making accessible many of the **amazing datasets held by UCLA, which are, in some cases, underutilized.** We consider it especially important to **make UCLA a leader in enabling access to datasets relevant to research on racial equity and racial justice,** with a view toward furthering UCLA's *Rising to the Challenge* initiative and UCLA's commitment to becoming a Hispanic Serving Institution.

*ATLAS is a UCLA Health wide community biobanking initiative that will recruit 150,000 UCLA Health patients to create California's largest and most diverse genomic resource for translational precision medicine...Because we have bio-banked samples, other "-omics" modalities, such as methylation and genome or exome sequencing will be conducted as opportunities arise, further enriching the resource. Notably, having this infrastructure in place has allowed us to rapidly respond to COVID-19 and participate in multiple international efforts focused on understanding the host (human) response to the virus.***—Institute for Precision Health WP**

Industrial Joint Appointments and Industrial Residencies: We want to encourage a flow back and forth between the Institute and those working in industries relevant to DataX. A major (and increasingly popular) modality for such cross-fertilization is a joint academic-industry appointment. **It is essential that UCLA create policies for joint academic-industry appointments that encourage them where appropriate while taking due account of the needs of academic departments and students.** We are cognizant that this is a difficult issue with a long history, and that its resolution falls outside the scope of this report. The market for faculty in the DataX arena is highly competitive, with a substantial disparity between academic and industrial salary scales. This suggests taking a fresh look at how one might structure such appointments in the DataX context; the Academic Senate (particularly its Council on Academic Personnel) can play a critical role in establishing a reasonable and equitable policy. Taking action is important; colleagues at peer institutions noted that departments can face mass exodus when such appointments are not permitted. On a shorter time scale, company and lab employees might spend a quarter in residence at the Institute as Industry and Lab Partners, or Faculty Fellows might spend a quarter embedded in a company or lab on leave. We anticipate that these would primarily be funded by the industrial sponsor.

Fund-Raising: The Institute constitutes a fund-raising opportunity of the first magnitude. Peer institutions have attracted major gifts around their data science efforts. Institute leadership should work closely with Development to cultivate and secure large gifts from private and corporate philanthropy.

Funding from off-campus and private-sector partners is not zero-sum. By offering more opportunities for collaboration and philanthropy, the likelihood of securing such funding is enhanced. Peer institutions have succeeded in obtaining substantial funding from foundations and in securing very large gifts for their Data Science programs. The fundraising that DataX will do will be broadly inclusive, matching the broad intellectual footprint we have proposed.

Resources for DataX Support Network

A number of existing entities already engage in DataX support activities. Grant funding should be used to expand support activities wherever possible, e.g, by hiring additional support personnel whose time is paid for by various grants. However, **it is essential that assistance be available freely and at scale to**

those who need it—both faculty and students working on research projects. This point was underlined by colleagues at peer institutions.

Incentives are needed to expand the DataX Support Network to meet the anticipated rising demand. **We propose that divisions and departments be able to apply for 1-1 matching funds for such expansion, with the percentage of central support declining 50%, 40%, 30%, 20%, 10%, 0% over 6 years.**

Funding Model

Given the uncertainties associated with the new activity-based budgeting model, it is difficult to predict exactly what can be covered by the income associated with the Institute's courses and return of overhead on grants run through the Institute. **At least initially, many of the Institute's activities will need to be supported by a dedicated Institute budget.**

It would be natural to support these activities, in part, with funds generated by the indirect costs on external grants, since the Institute and Support Network contribute directly to the research capacity of the university. **A small percentage of the face value of annual grant funding, taken out of overhead, would go a long way toward covering the costs. The amount could be phased in gradually as DataX activity ramps up.** These funds would be used across campus independent of their origin. There is a need to nurture activity in areas where interest and expertise in DataX is still in its early stages (and grant funding is often scarce). We therefore anticipate that this would to some extent result in a South-North transfer.

EQUITY, DIVERSITY, AND INCLUSION
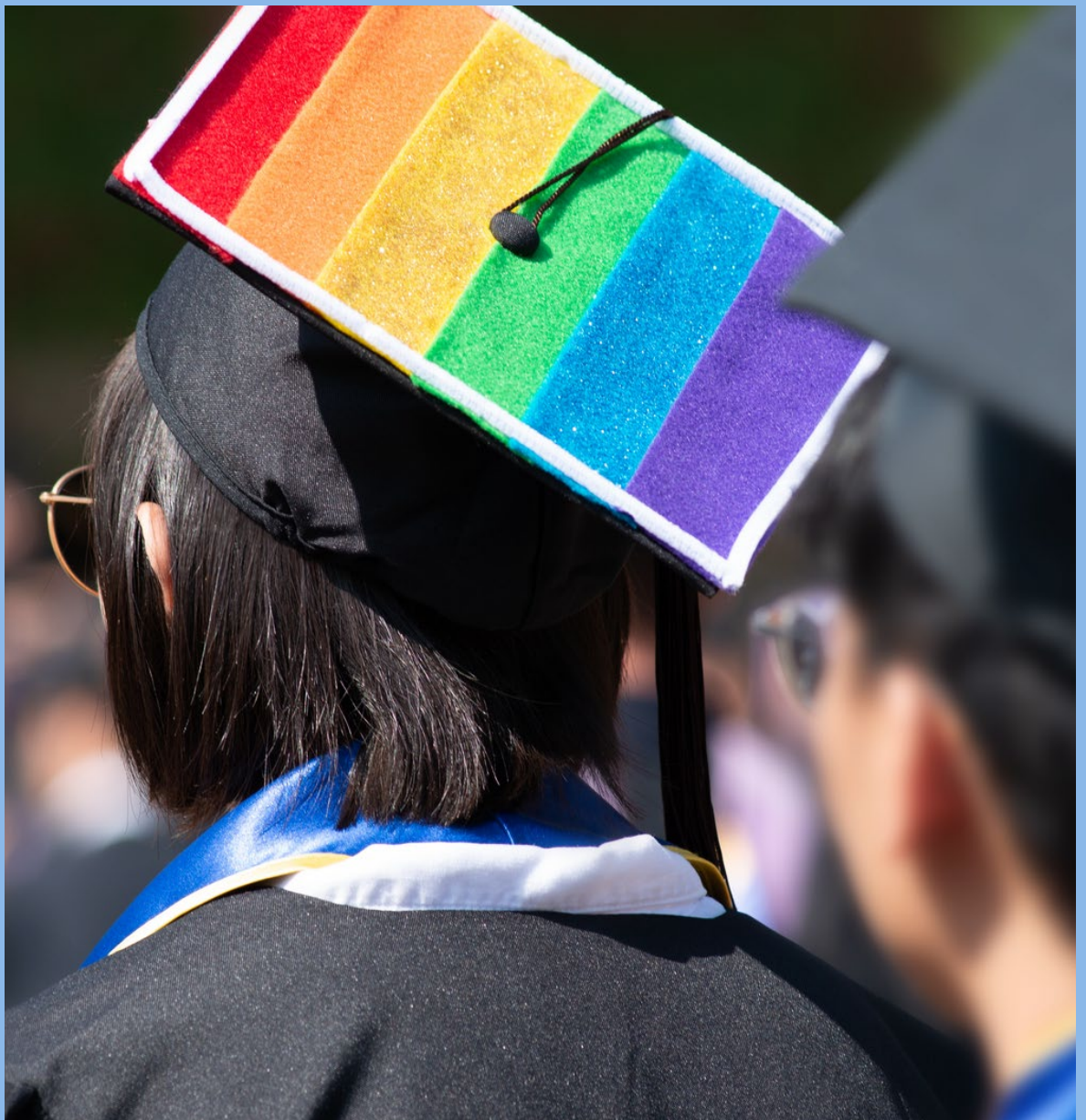
Equity, Diversity, and Inclusion

We have made a commitment to equity, diversity, and inclusion central to the proposed DataX strategy. Here we collect the key tactics by which that commitment will produce long-term and material change.

First, from a broad strategic perspective, it is essential that **DataX align and partner with other major campus initiatives dedicated to equity, diversity, and inclusion.** DataX leadership should work closely with efforts pursued under the *Rising to the Challenge* initiative, and UCLA's commitment to become a Hispanic-Serving Institution. In particular, DataX should partner with these efforts to recruit and retain faculty.

The breadth and scope of DataX scholarship means that research in DataX can and must address themes of equity, diversity, and inclusion across the three cohorts. Research in *fundamental data science* on topics like privacy, bias in machine learning, algorithmic fairness, interpretability, and explainability has direct implications for equity, and will benefit from the inclusion of a diverse cohort of researchers with a range of backgrounds and life experiences. There are countless *innovative applications* of data science that speak to equity, diversity, and inclusion, from data-intensive work on minority health disparities and COVID or LGBTQ suicide risk to scholarship on incarceration or homelessness. Finally, work on the *social, ethical, and policy implications of data* and its uses provides an obvious home for rigorous, critical scholarship that centers issues of equity, diversity, and inclusion. The DataX initiative should make a commitment to recruiting Fellows whose scholarship involves such themes; to supporting working groups that embrace the equity, diversity, and inclusion dimensions of DataX; and to ensuring that discussions of social, ethical, and policy implications take place in conjunction with the substantive topic of a working group. More modestly, our proposed policy of requiring that postdoctoral scholars have two mentors drawn from two different cohorts means that many postdocs working on fundamentals or innovative applications will be able to receive mentorship from scholars working on the social, ethical, and policy implications of DataX, weaving equity, diversity, and inclusion deeply into their training. Finally, the initiative should commit to acquiring and supporting the collection of data relevant to underrepresented groups and marginalized communities, with participation from members of those groups and communities.

Several aspects of the proposed educational expansion support our aspiration of an inclusive initiative that produces the nation's most diverse cohorts of DataX scholars. Recognizing longstanding disparities in the funding and provision of K12 education, the Initiative insists on and will support the development of courses that do not depend on assumptions about prior mathematical and computational training. The strategy proposed for the cluster courses--which emphasizes themes that speak to our students, like climate and sustainability, social justice, or environmental justice--will likewise help to reach a broad population of students, some of whom (e.g., first generation students) may not have prior exposure to the idea of data science or data-intensive scholarship. Finally, the Institute should ensure that all parts of the DataX educational ecology include training on the social, ethical, and policy aspects of data and its use; in particular, the cluster courses should make this a major theme.

The DataX Initiative should partner with and support existing, funded projects and centers at the intersection of DataX and equity, diversity, and inclusion, with the ultimate goal of cultivating a robust funding stream to support such scholarship at UCLA. For example, we can imagine robust partnerships with Kelly Lytle Hernández and the Million Dollar Hoods team; with Safiya Noble, Sarah Roberts, and their Center for Critical Internet Inquiry; or with the recent Mellon-funded Social Justice

Curriculum project. We can also envisage the development of mentoring networks to cultivate and attract recipients of MPS-Ascend Postdoctoral Research Fellowships, and believe that the Institute can play a critical role in supporting pre-tenure faculty in pursuing awards like LEAPS (Launching Early-Career Academic Pathways in the Mathematical and Physical Sciences) or CAREER; some of these opportunities (Ascend, LEAPS) have explicit goals of broadening participation.

These latter efforts dovetail with **commitments on the faculty hiring side that will support the recruitment of a diverse DataX faculty.** We advocate for making at least 2 early hires of faculty whose research focus and education interests center on racial equity and racial justice. We also encourage close collaboration between Institute hiring and UCLA's emerging Mentor-Professor program, which recruits faculty with a special interest in mentoring students from underrepresented backgrounds. We also encourage close collaboration with hiring pursued under *Rising to the Challenge* and UCLA's Hispanic Serving Institution commitment, so that new DataX faculty from underrepresented backgrounds can participate in multiple cohorts, e.g., a cohort with overlapping substantive interests through *Rising to the Challenge* and a cohort with overlapping DataX interests.

Finally, our insistence on the cultivation of opportunities for **community engagement and service provides many opportunities to address inequities in our home city and home state.** Research pursued under the aegis of the Homeworld/Institute and hands-on education pursued through internships and formal instruction can provide material support to marginalized communities in Los Angeles and beyond.

EVALUATION & METRICS

Evaluation and Metrics

Given the scope and transformational ambition of the DataX initiative, it is essential that our campus investment be *data-driven*. UCLA must do more than roll out new structures, new educational initiatives, and new outreach efforts. We must also develop a strategy for evaluating these activities, so that effective interventions are supported and ineffective ones are improved or abandoned. This strategy should begin with a set of core goals, and only then devise appropriate metrics. Such an approach avoids a major pitfall of evaluation: Confusing what is important with what is easy to measure.

For the DataX Homeworld and eventual DataX Institute, we see the key goals as (i) *building a robust, multifaceted and diverse campus community at the cutting-edge of DataX* and (ii) *developing new research initiatives* that cut across disciplines and the three DataX cohorts.

For the new hiring recommended by the report—Institute faculty and postdoctoral fellows—we see the key goals as (i) *recruiting outstanding scholars across the disciplines;* (ii) *ensuring equity, diversity, and inclusion in these new hires*, with a particular focus on including and supporting scholars from currently underrepresented groups; and (iii) *promoting fulfilling and successful career trajectories as appropriate for the career stage and fostering a culture within the Institute that values and rewards a diversity of intellectual approaches and interdisciplinary work.*

For the DataX Support Network, we see the key goals as (i) *increasing the number and diversity of faculty and students involved in DataX scholarship* and (ii) *robust development of campus interest and capacity in DataX* (in terms of concrete research skills).

For the educational initiatives, we see the key goals as (i) *developing DataX courses, certificates, and degree programs that engage a diverse group of students with a range of interests and technical backgrounds* and (ii) *ensuring that the DataX educational ecology is easy to navigate and legible within and beyond campus*. Closely related to these goals, we also prioritize the (iii) *involvement of UCLA students in DataX research on campus, and in DataX internships and community engagement beyond campus.*

For outreach and community engagement, we see the key goals as (i) **establishing UCLA's reputation as an internationally recognized leader in DataX**; (ii) **promoting the robust use of UCLA data, code, and educational resources**; and (iii) **building partnerships with local industry and community organizations of value to UCLA students and researchers.**

For equity, diversity, and inclusion, we see the key goals as (i) **ensuring that UCLA's DataX faculty and staff reflect the diversity of the state that we serve;** (ii) **creating DataX educational opportunities that attract and retain students from all backgrounds, and particularly promote the participation of underrepresented groups;** and (iii) **ensuring that DataX research and education promote UCLA's goals and commitments to equity and social justice, as well as embracing critical scholarship on the social, ethical, and policy impacts of data science and related fields.**

It is a high priority to develop appropriate metrics that **capture the Initiative's progress toward these goals and to foster a culture of continuous improvement** across the Initiative. Over time, learning from experience, instruments should be developed within the DataX initiative that will be the most helpful in the

design of processes and procedures across the Initiative. It is also important to embrace qualitative as well as quantitative metrics; surveys and other data-gathering instruments should incorporate established design principles, such as including open-ended questions. We believe such questions can reveal both the best outcomes (e.g., transformational research or educational experiences) and the most urgent critiques.

We recommend that **periodically (at 5 year intervals), the entire DataX initiative be evaluated as to whether it continues to serve a campus need.** It is difficult to foresee how the focus of research and education might change over the course of decades, and an existential evaluation on this time scale makes sense.

TRANSITION TO IMPLEMENTATION
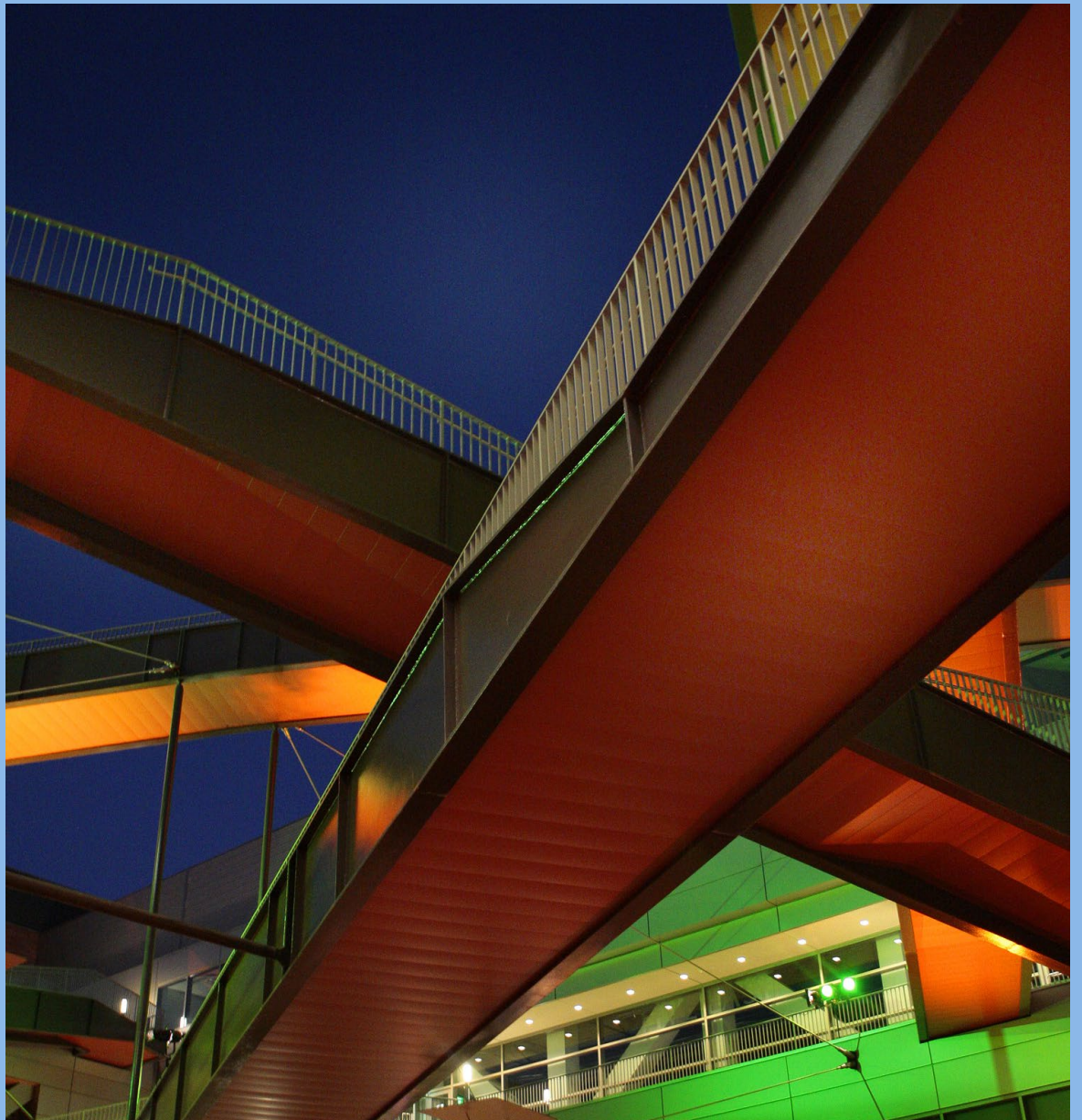
Transition to Implementation

The fundamental principle governing implementation is to move forward with creating the DataX Institute and DataX Support Network as expeditiously as possible. The present fiscal constraints do not diminish the importance of acting quickly. We have taken these constraints into account by recommending a structure in which resources will be phased in over a period of a few years. Because it may take some time to secure the resources required for the DataX Institute and DataX Support Network, we recommend establishing a **transitional entity** (the **DataX Homeworld**) that will allow the campus to initiate a range of DataX activities in research, hiring, education, outreach etc. These activities will allow the campus to begin its transformational investment in DataX while fundraising and laying the groundwork for a DataX Institute and Support Network.

We recommend that the campus initiate a search for an Institute Director only after the Institute and Support Network are up and running. During the startup phase, leadership for the Initiative and the DataX Homeworld should be drawn internally from UCLA.

It is essential that the leadership of the Homeworld and eventual Institute be as even-handed as possible between fields—both among the three cohorts and within fundamental data science itself. Getting the "initial conditions" right is absolutely essential to realizing the vision set forth in this document. Some peer institutions have had to "reboot" their data science initiatives because the initial setup was not compatible with long-term goals.

There are some critical first steps that should be initiated as soon as possible.

First, a systematic taxonomy of existing and planned DataX research, teaching, support, and outreach should be created. While we gleaned much of this information from our interviews and the submitted white papers, implementation requires detailed awareness of the campus landscape, and systematic data collection is needed to fill in the blanks and capture the unknown unknowns. Such information will also provide the raw material for the research and teaching portals described above; since constructing these portals as high-quality, sustainable resources will take considerable time, it is best to start early.

Second, and in keeping with UCLA's principles of shared governance, the Initiative should engage in close consultation with the Academic Senate leadership and appropriate committees about a range of issues raised by this strategy. We want to clarify that in writing this report, we took on the task of **proposing a strategy for research and education in Data Science and allied fields at UCLA,** after extensive bottom-up consultation. We take it as a given that there will be a process of iteration to settle on the precise details of how the programs and structures we recommend would be adapted to satisfy the rules and procedures that govern UCLA. We expect that an important piece of this process will unfold through working in partnership with the relevant Academic Senate committees in an organic process as the entities we envisage are animated and built out.

Taking personnel issues as an example, the Senate can play an essential role in crafting policy for the joint hires proposed above, ensuring a fair, equitable review process that acknowledges the complexity of their roles and the difficulty of evaluating intrinsically interdisciplinary work. The Senate is also the appropriate venue for working out the fraught issue of joint UCLA-industry appointments, which are a significant challenge in securing top data science talent. The Academic Senate would also play an

important role in the creation of new courses and degree programs developed as part of this initiative. Specifically, the Graduate Council and/or Undergraduate Council will need to approve new and updated courses, minors, majors, or degree programs. The Council on Planning and Budget will provide feedback on any degree proposals. The Committee on Rules and Jurisdiction will review any changes to bylaws or regulations before approval by the Legislative Assembly. The Council on Research may engage in an advisory capacity with issues of how research will be structured and supported.

Third, we recommend a systematic consultation with UCLA's undergraduate and graduate student populations, e.g., through meeting with student leaders and through student-directed town halls. Their feedback will guide the design and implementation of activities informed by this report, particularly those in education.

Fourth, the Administration should identify activities that can be pursued without the full resources required by the Institute and Support Network, and—balancing the urgency of DataX against the present fiscal constraints—launch those activities under the aegis of the DataX Homeworld.

Fifth, we recommend the early convening of an education working group under the auspices of the DataX Homeworld. This working group should refine the educational principles outlined in the education section; begin to develop a shared understanding across campus of the core aspects of DataX curricula at varying technical levels; and cultivate experiments in pedagogy and course design.

Sixth, we recommend that the DataX Homeworld, upon its establishment, work to articulate key goals for the initial phase of the Initiative and develop internal metrics that can be used to provide continuous feedback on the benefits and limitations of its programmatic efforts.

Finally, the Administration should develop a coherent fundraising strategy around DataX, coordinating efforts to secure traditional philanthropic donations with pursuit of federal and foundation grants and industry support. Peer institutions have raised hundreds of millions of dollars to pursue campus transformation around data science, AI, and computation more broadly, and UCLA must have similar ambitions.

UCLA

REPORT OF THE DATAX STRATEGY TASK FORCE
A TEMPLATE FOR TRANSFORMATION

PREPARED BY JACOB G. FOSTER AND MARK L. GREEN
SPECIAL ASSISTANTS TO THE VCRCA FOR STRATEGIC INITIATIVES IN DATA SCIENCE

Report Designer: Adela Hernandez // Content Management Specialist