A Call to Action:
A Roadmap for Expanding Research Infrastructure and Data Services at UCLA

Last updated December 2, 2022. Only minor changes have been made following re-submission of the draft to our Executive Sponsors on November 2, 2022. The submission was accompanied by a request for authorization to begin sharing the draft with campus as outlined in the Research Data Working Group's communication plan.

A Call to Action:
A Roadmap for Expanding Research Infrastructure and Data Services at UCLA

Executive Summary

UCLA has experienced phenomenal success over the past decade such that it is now considered the top ranking public university in the United States. A culture shift, however, is underway that requires immediate action if UCLA hopes to sustain this level of success and retain its place of prominence among R1 institutions. Academia is rapidly moving away from the model of siloed research institutions as popularized in the early 19th century and toward a modern culture of digital scholarship, interdisciplinarity, collaboration, open science, and open data. UCLA is not prepared to support its faculty in participating in this new era, nor in meeting the institutional responsibility to comply with sponsor data-sharing requirements.

This roadmap for enhancement to UCLA's research infrastructure, data resources, and services is the product of a research data working group (RDWG) convened by Roger Wakimoto, Jim Davis, Ginny Steel, Lucy Avetisyan, and Ellen Pollack. The institutional motivation for this effort is fivefold: 1) to better support faculty, innovative research, scholarship, and teaching; 2) to provide open access to knowledge when possible; 3) to meet funder mandates for preservation of and access to research data; 4) to preserve the research of our campus community; and 5) to address increasing cybersecurity compliance and risk as digital technologies become the default for conducting and disseminating research.

Taken overall, this document describes a new approach to research infrastructure and data services at UCLA. Our vision is centered on a full lifecycle approach to open science and scholarship that will position UCLA faculty to continue to be competitive for larger, multi-institutional proposals built on the principles of sharing, collaboration, training, and best practices. Development was guided by a series of principles that respect open data/science, look to the institution to provide core services to help faculty meet their data obligations and develop solutions for data curation and publication, and call for education campaigns and guidelines to train the UCLA community on data best practices. The roadmap as articulated in this proposal addresses those challenges, would generate significant opportunities for faculty and the campus, and represents a direction that validates the research process that developed and launched the DataX Initiative. Implementing this roadmap will establish:

- A logical hierarchy of operational research support that combines specialized and/or discipline-specific resources at the local level with centrally provided baseline research support offerings and services.

- Immediate allocations of resources for new faculty rather than forcing them to negotiate piecemeal technical infrastructure packages.

- Coherent and cohesive information about available research data services and resources, and policy requirements and practices.

- Baseline research data storage, management, computation, and access through infrastructure available at no cost to all faculty.

- Clear pathways to resources and services for faculty with appointments both on campus and in the Health Sciences.

- Clear administrative responsibility for research infrastructure and data services aligned with faculty governance.

- Security, privacy and accessibility integrated with research needs and focused on protecting UC's most valuable data assets, research subjects, and intellectual property.

The RDWG realizes this vision will require a long-overdue and prudent investment in technology and expertise, and that there are many factors to consider before committing to this roadmap. UCLA's current distributed culture has long supported innovative research and exploration, yet there are now emerging and growing data-related needs that are best addressed centrally. Regardless of future potential repository funding that may or may not be available to grant awardees, the shift to data-intensive research requires secure research infrastructure and services with the capacity to seamlessly gather, manipulate, analyze, visualize, share, and publish data. This document builds on the pillars of UCLA's existing research structure and history of faculty innovation while also laying the imperative foundation for the new requirements of academic knowledge production. While investments in research are needed at both the local and central levels, this document focuses on the development of central capacity and capabilities to support data-intensive research while respecting and strengthening our distributed discipline- and domain-specific research environments. Investing in central services and people where appropriate allows UCLA to spend efficiently on technology at scale for needed research infrastructure, consolidate expertise for data-related support, and alleviate some of the administrative burden on faculty, allowing them to focus on their research and scholarship.

This document also establishes a foundational strategy for stable and understandable research infrastructure investments. The RDWG's first assertion is that the campus needs to streamline and enhance the faculty support experience by formally consolidating and clarifying ownership, responsibility, and accountability for initiatives related to research infrastructure and research data services. The second core element of the roadmap is enabling faculty through an investment in shared research infrastructure and data services that will augment locally

provided research support offerings to facilitate new modes of research. Access and use will be at the discretion of individual faculty and the local units; regardless of the mix, the imperative is to ensure that the needs of faculty and researchers in all units are met by a blend of local and central campus infrastructure and services. By investing in much-needed central research data infrastructure, the campus and faculty alike will benefit from better research-related services and infrastructure, campuswide modeling for IT best practices, more understandable and usable security, lower administrative costs, more technical support for units, as well as lowering the costs and time to publication.

The RDWG has met and evaluated data and experiences from a broad cross-section of the research community. In addition, representation from the major IT partners across UCLA joined in the discussion and analysis, including examination of prior infrastructure efforts at UCLA. The RDWG fully acknowledges that this document is a starting point for conversations with the campus about the resources and services needed to ensure UCLA's future success. Planned discussions with the Academic Senate, its subcommittees as appropriate, the board of the faculty-led Institute for Digital Research and Education (IDRE) as a standing subcommittee in the new IT governance structure, our DataX leadership, and researchers in general will ensure that faculty will have direct input on the final roadmap and a voice in shaping investment priorities and process.

A Call to Action:
A Roadmap for Expanding Research Infrastructure and Data Services at UCLA

Section 1: Introduction

UCLA has experienced phenomenal success over the past decade such that it is now considered the top-ranking public university in the United States, 15th overall among the top 1,102 universities in the world, with 40 of our graduate programs in the top ten programs evaluated by the National Research Council (NRC).[1] Moreover, UCLA researchers have a proven track record of innovative, ground-breaking, and impactful research and scholarship, a fact that speaks volumes about the perseverance and brilliance of our exceptional faculty, which includes recipients of academia's highest honors and members of the nation's most prestigious societies.[2] UCLA is also among the leading research institutions, with the university ranking 7th in overall research and development expenditures in the 2020 NSF Higher Education Research Expenditures analysis.[3] More recently, in FY 2021-2022 UCLA researchers were awarded over $1.7 billion dollars in extramural funding, the highest yearly total in the history of the university.

A culture shift, however, is underway that requires immediate action if UCLA hopes to sustain this level of success and retain its place of prominence among R1 institutions. Academia is rapidly moving away from the model of siloed research institutions as popularized in the early 19th century and toward a modern culture of digital scholarship, interdisciplinarity, collaboration, open science, and open data.[4] Moreover, our federal and nonprofit sponsors have introduced mandates that require data sharing in order to accelerate research, encourage multi-institutional collaboration, and ensure that research funded by taxpayers and donors is of optimal benefit to the public. NSF, NIH, NASA, and other federal agencies have established policies and enforcement plans for mandatory data sharing that promote open research. Most recently, the White House Office of Science and Technology Policy (OSTP) has directed federal agencies to establish by December 2025 policies that prohibit federally-funded research from being embargoed by scientific publishers. Rather, taxpayer-funded research must be made available to the public immediately and at no cost. Future funding will depend on institutional compliance with these policies.

[1] https://www.ucla.edu/about/rankings
[2] https://www.ucla.edu/about/awards-and-honors/faculty
[3] See table 20 of the NSF's higher education R&D (HERD) expenditures, ranked by FY 2020 R&D expenditures with longitudinal data from FYs 2010–20. https://ncses.nsf.gov/pubs/nsf22311
[4] https://sparcopen.org/news/2022/nasa-year-of-open-science-set-to-launch-in-2023/

We now exist in a data-centered research world, and this roadmap is the first step toward developing services and resources that will allow UCLA to continue to acquire funding and conduct research. As funder-mandated data preservation and sharing policies have become more standard and complex, so has the academic data repository landscape with its myriad of options that range from the large and small, discipline-specific, generic, institutional, and publisher-hosted. The one commonality is the need for data access, curation, and preservation. For UCLA data researchers to comply with these new research standards and identify appropriate repository options, they need data and technology support, robust local and system-wide services, and dedicated staff with expertise in the development, curation, and preservation of research data. Regardless of any changes that may alter the data repository landscape in the coming years, UCLA faculty will still need data solutions, campus-based support, and infrastructure that facilitates their research.

UCLA is not prepared to support its faculty in participating in this new era of open data and open science, nor in meeting the institutional responsibility to comply with sponsor data-sharing requirements. Our current success has been built on individuals creating their own self-sustaining research environments disconnected from the rest of campus. In an institution famous for its decentralization and its withering-on-the-vine busting bureaucracy efforts, PIs are on their own and the result is 'death by a thousand cuts.' This approach hobbles UCLA researchers in a world that prioritizes collaboration and transparency. Investing in central services enhances distributed efforts, supports compliance, and simplifies collaboration. Providing faculty with access to fundamental research infrastructure and data services that support their ability to generate, store, manipulate, and analyze data is an institutional imperative if UCLA wants to continue to lead in this new era. To ensure excellence in the future, we need to make a standing commitment to invest in research data and infrastructure resources and services, and we need to take action immediately. We have invested in upgrades for our administrative systems to better serve campus needs, and in our learning management system to meet the needs of the student population, and now is the time to make a similar investment in our research environment. The longer we wait to address our research data and infrastructure challenges, the harder it will be to catch up to our peer institutions.

The Research Data Working Group (RDWG) was formed to assess the "as is" and "should be" states of the research data cyberinfrastructure, build a roadmap for enhancement to UCLA's research infrastructure, data resources, and services, and identify the reputational and financial risks to UCLA in the absence of a coordinated response to the proposed roadmap.[5] The institutional motivation for this effort is fivefold: 1) to enable innovative research, scholarship, and teaching; 2) to provide open access to knowledge; 3) to meet funder mandates for preservation of and access to research data; 4) to preserve the research of our campus community; and 5) to address increasing cybersecurity compliance and risk as digital technologies become the default for research.[6] This document is the result of that assessment.

[5] See Appendix 1 for the RDWG charge and Appendix 2 for a complete list of participants of the Research Data Working Group and the five original subcommittees.

[6] In this document, 'research output,' 'research data,' and 'data' are used interchangeably, and are broadly defined to include all digital products that are produced in the course of scholarly work. This includes a

It is intended as a starting point for conversations with all campus stakeholders about necessary research data and infrastructure resources and services.

Research Data Principles

Research and research data are increasingly subject to a complex set of shifting mandates and requirements from the campus, the UC system, and funders. By policy, research data are the property of The Regents of the University of California with UCLA researchers assigned stewardship responsibilities. These shared responsibilities compel the institution to provide faculty with robust, efficient, and sustainable support mechanisms to ensure the longevity of their research data.

The roadmap described herein was created to meet such responsibilities, guided by the following ten principles:[7]

1. In accordance with open-access policies of the Academic Senate of the University of California and the UC Office of the President, UCLA is committed to ensuring that all "published research articles in the broadest sense of the term" authored by any UC employee be made available to the public at no charge through open-access repositories such as eScholarship.[8]

2. Where possible, UCLA expects that this mandate be extended to include other forms of research outputs in accordance with the FAIR guiding principles that call for data to be findable, accessible, interoperable, and reusable.[9] When depositing data, faculty should be encouraged to use repositories that follow the TRUST principles.[10]

---

broad range of output including, but not limited to, structured research data, generated digital assets, multi-media files, and other new forms of digital knowledge production. Similarly, 'research infrastructure,' 'infrastructure,' and 'cyberinfrastructure' are used interchangeably, and are broadly defined as the technical, physical, and human support networks that are needed to support research. Throughout, 'UCLA Health Sciences' and 'Health Sciences' are used as the umbrella term to refer to both the school and the health system. 'UCLA Health' refers only to the hospital system and faculty practice groups.

[7] These principles were adapted from a Research Data Position Paper originally drafted at UCLA in September 2016 and inspired by similar documents from other institutions, most notably the University of Edinburgh and the University of Oxford. The complete paper was included as an appendix in both the 2017 Research Data Support Recommendations and the 2020 Task Force on Research Infrastructure Final Report.

[8] http://osc.universityofcalifornia.edu/open-access-policy/policy-text/presidential/

[9] Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016). The characteristics of FAIR data are described by Wilkinson et al and include requirements for rich metadata and globally unique and persistent identifiers; retrieval with standardized communications protocols; use of "formal, accessible, shared, and broadly applicable language for knowledge representation"; and use of domain-relevant community standards. See also the GO FAIR US website (https://www.gofair.us/about) for information about community efforts to promote and encourage open science and open data.

[10] The TRUST principles are "Transparency, Responsibility, User focus, Sustainability and Technology," and they provide a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders." https://www.nature.com/articles/s41597-020-0486-7

3. Beyond open access, UCLA is also committed to:
    a. Supporting researchers' efforts to comply with funder mandates for data sharing, management, and preservation.
    b. Assisting faculty in meeting expectations for data use and reuse, data sharing, and data deposits into discipline-specific or subject-area repositories.
    c. Developing new digital models for teaching and learning, and encouraging and supporting new forms of digital scholarship.
    d. Treating research data as an institutional asset equivalent to more traditional intellectual property, and providing for its appropriate stewardship.

4. This commitment extends to developing and maintaining the human, physical, and technical infrastructure necessary to support exemplary data practices throughout the research lifecycle, recognizing that:
    a. There is both need for and value in distributed, centralized, and layered approaches for infrastructure.
    b. There is a pressing need for disciplinary and infrastructure expertise on campus.
    c. Any infrastructure effort must include a robust staff and faculty training component, and long-term support mechanisms.
    d. An essential step in this development is an update to our institutional IT plan so that it better reflects the complexity and reality of the current research environment and changes in available technology.

5. Critical to fulfillment of this commitment is the development of a data management, preservation, and curation services that will:
    a. Ensure that UCLA's intellectual assets are appropriately curated and preserved.
    b. Collaborate with researchers to write data management plans, develop metadata and informational data for their research outputs, and coordinate data deposits to domain- and discipline-specific repositories.
    c. Develop outreach efforts on data best practices for researchers across campus.

6. The development and maintenance of infrastructure for storing and preserving UCLA's digital assets is also critical and should be prioritized by the campus.

7. Given that the infrastructure to support digital research is now a fundamental requirement for scholars across disciplines and part of good research practice, UCLA is committed to providing data services and infrastructure to researchers at no cost or at a substantially subsidized rate.

8. Understanding that it is unreasonable – if not impossible – to preserve 100% of the intellectual products of the campus, conversations are needed at the highest levels to prioritize research outputs for curation and preservation.

9. It is imperative to enable and to incentivize faculty to 'do the right thing' when it comes to research data. Critical to use of any institutional data services and support infrastructure

is clearly articulating the benefits for faculty and researchers.

10. Guidelines and governance models for institutional research data need to be developed for all campus data resources. At minimum, the guidelines should address the collection and retention of research data outside of any contractual obligations, and consideration should be given to deposit agreements, terms of use, metadata, ethics, security, privacy, preservation, de-accession, forward migration, and takedown policies.[11]

These principles respect open data/science, look to the institution to provide core services to help faculty meet their data obligations and develop solutions for data curation and publication, and call for education campaigns and guidelines to train the UCLA community on data best practices. The roadmap as articulated in this proposal addresses those challenges and represents a direction for campus that validates the research process that developed and launched the DataX Initiative. The solutions promoted in this document build upon the findings of past UCLA infrastructure planning efforts including, but not limited to, the 2013 North Campus Infrastructure Planning Summit; the first phase of the 2013-2014 Research Informatics Strategic Planning (RISP) Initiative, an inclusive effort that included over seventy faculty and staff in a series of meetings to discuss data requirements, campus infrastructure, and research support; and the 2017 Research Data Support Recommendations. These previous efforts have been stymied by lack of institutional commitment, leadership turnover, and the COVID-19 pandemic.

A Contemporary Vision for Research Data Services and Infrastructure at UCLA

Taken overall, this document describes a new approach to research infrastructure and data services at UCLA. Our vision is centered on a full lifecycle approach to open science and scholarship that will position UCLA faculty to continue to be competitive for larger, multi-institutional proposals built on the principles of sharing, collaboration, training, and best practices. This approach aligns with both the IT Services' 'Strengthening the Core' initiative and the recommendations of the DataX strategy report in its discussion of the requirements for a DataX Support Network.

Implementing this roadmap will establish:

- A logical hierarchy of operational research support that combines specialized and/or discipline-specific resources at the local level with centrally provided baseline research support offerings and services.

- Immediate allocations of resources for new faculty rather than forcing them to negotiate piecemeal technical infrastructure packages.

[11] Sample RDM policies: https://sites.google.com/a/sheffield.ac.uk/rdm_links/repo/policy

- Coherent and cohesive information about available research data services and resources, and policy requirements and practices.

- Baseline research data storage, management, computation, and access through infrastructure available at no cost to all faculty that includes:
    - An understandable, responsive, and protected network environment to generate, store, manage, aggregate, analyze, orchestrate, and use data collaboratively.
    - Tools and mechanisms for analyzing, sharing, and preserving research data.
    - High-performance computing infrastructure that meets a diversity of researcher needs.
    - Research data backup and disaster recovery protections.
    - Data services and support across the research and data lifecycles.
    - A shared data/code repository for both internal and public access, perhaps ordered by security levels.
    - Data and computational infrastructure that addresses security, privacy, and intellectual property protections for a full range of data risk categories and requirements.

- Clear pathways to resources and services for faculty with appointments both on campus and in the Health Sciences.

- Clear administrative responsibility for research infrastructure and data services aligned with faculty governance.

- Security, privacy and accessibility integrated with research needs and focused on protecting UC's most valuable data assets, research subjects, and intellectual property (IP).

The RDWG realizes this vision will require a long-overdue and prudent investment in technology and expertise, and that there are many factors to consider before committing to this roadmap. This roadmap presumes the completion of a number of ongoing technological initiatives including efforts by IT Services to modernize UCLA's identity and access management system, which will reduce friction for researchers campuswide, but particularly for those with appointments both on the campus in the Health Sciences. IT Services is also working to unify the campus network, and the Health Sciences is building a storage system that will support their researchers. The following section addresses the opportunities enabled by a campus investment in research data and infrastructure. Sections 3 and 4 more fully describe the roadmap for this vision, with implementation and possible phasing for the build out of the recommended resources and services discussed in Section 5.

Section 2: Opportunities and Benefits

Despite difficult economic times, an investment in research infrastructure enables a number of opportunities that will both allow UCLA to sustain its reputation as a preeminent research institution in the midst of increasing competition from R1 and aspiring R2 Institutions, and address significant operational and financial risk in the present and near future. Opportunities and benefits for faculty and researchers have been prioritized, followed by items that speak to our institutional concerns of costs, risks, and financial impacts.

1. Enable Publishing

Expanding data-related resources and services at UCLA will facilitate faculty compliance with data-availability requirements that are increasingly becoming conditions of publication. In short, researchers unable to meet publisher data requirements won't be able to publish. This is true both for academic journals and major publishing houses. Authors hoping to publish in *Nature*, for example, "are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications."[12] Similarly, *Science* requires that "all data used in the analysis must be available to any researcher for purposes of reproducing or extending the analysis. Data must be available in the paper or deposited in a community special-purpose repository or a general-purpose repository … "[13] Author services at Taylor & Francis request that journal authors submit with their manuscripts data availability statements that detail the location and conditions of access for the data used in their research, and they provide template language for common scenarios (e.g., data that are in publicly available repositories, under embargo, derived data available upon request).[14]

These data-sharing policies extend to Taylor & Francis' book divisions, with Routledge, a series under the brand, providing the arguments for open research:

- Sharing data publicly improves the robustness of the research process, supporting validation, research transparency, reproducibility and replicability of results.
- Sharing data can lead to re-use and discovery, with greater opportunities for carrying out meta-analyses and the extraction of new knowledge.
- Depositing data in a repository that mints a permanent identifier such as a DOI, allows authors and others to cite the data set, allowing researchers to get appropriate credit for their work.
- Data deposition supports the preservation of data long term.
- Wider public availability of research data supports the translation of research into practice.
- Many funders now make data sharing a requirement … and it's become increasingly commonplace for some subject areas to make data available to everyone.[15]

[12] https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards
[13] https://www.science.org/content/page/science-journals-editorial-policies#research-standards
[14] https://authorservices.taylorandfrancis.com/data-sharing/share-your-data/data-availability-statements/
[15] https://www.routledge.com/our-products/open-access-books/books-data-sharing-policy

In addition to providing information about the open access policies of the Academic Senate of the University of California and the UC Office of the President described above, the UC Office of Scholarly Communication also offers researchers guidance on compliance.[16] Critical to compliance success is the availability of local expertise to support researchers, funding, and robust research infrastructure and data services – all elements of this roadmap document.

2. Enable Innovative Research and Pedagogy

Implementing the roadmap described in this document will not only facilitate researchers' scientific and academic objectives, it will also foster collaboration, broader partnerships, and broader impact. It will speed up time to discovery by reducing the need for individual faculty to 'reinvent the wheel' and allowing them to focus on their research from the moment they join UCLA. The infrastructure and services will enable researchers (faculty and graduate students) to spend more time and effort on innovative research and pedagogy. In short, implementing this roadmap will create "an environment conducive to faculty success."[17] Specific researcher use cases are described in Appendix 3, and the following examples illustrate the opportunities afforded by a modern research environment.

1. Robust data infrastructure and services will foster research and scholarship with the potential for much greater and broader impact. It will also create opportunities for UCLA to assume research leadership based on impactful outcomes and the integration of research into the classroom experience.

2. This roadmap also supports interdisciplinary research by enabling shared research and teaching enclaves that can cross departmental and divisional boundaries. Defining and supporting a baseline of computational research, pedagogical practice, and data discovery enables a data-centered approach to teaching and research. This is a much-needed effort on a campus where cross-departmental and cross-institutional collaborations are equally difficult, and PIs struggle to find support for compute/data aspects of their projects because they are perceived as 'outside the scope' of existing campus structures.

3. Providing researchers with fundamental infrastructure and data services will reduce the time they are required to spend 'managing' the technical aspects of their research. Picture a future where new faculty recruitment doesn't require a major investment in technology and research infrastructure, but rather the partitioning of existing services to easily create data and compute drive shares for specific research needs.

[16] https://osc.universityofcalifornia.edu/
[17] See the 2022 "Joint Senate-Administration Mitigating COVID-19 Impacts on Faculty Working Group Final Report" (https://ucla.app.box.com/s/ddoepdnq8033tl7h2me8v6k6z3hrjrp0)

4. Easily accessible and well-curated datasets would amplify the capabilities of our researchers by reducing the overhead of data-enabled research. Robust data services would also allow faculty to engage with the trends (and mandates) related to research data sharing that follow FAIR, TRUST, and CARE principles.[18]

3. Surpass Peer Institutions

Peer institutions have outpaced UCLA with infrastructure and realignment around research data, with many moving quickly to consolidate research support resources and services and the creation of administrative roles to oversee research data efforts. While we at UCLA have been talking about research data, other campuses have moved forward with investments and an expansion of their support offerings. UCLA fell considerably behind our peers in eight of 13 categories evaluated in the time between two benchmarking studies, the most recent being described in Appendix 4. Most notably we had the lowest score among peers for a research data services program with a website and catalog of services. UCLA also fell behind in support for data sharing, publishing and curation; data acquisition, licensing, storage, and backup; and behind in support for high performance computing.

As a general comparator, the University of Michigan stands out. When the initial White House technology memo was released and federal funders began talking about data management requirements, Michigan officials took action and prepared for this culture shift. Workgroups were formed to plan their response, leadership was reorganized to clarify roles and responsibilities related to data (the Office for VC of OIT), and governance mechanisms were put in place.[19] These efforts led to significant investments in research computing that provided a more robust campus research infrastructure for faculty.[20] In 2014, Michigan's Office of Research funded "an enterprise-wide research data warehouse (RDW) with a user-friendly portal (DataDirect) to enable a wide research audience directaccess to robust data."[21] This RDW and the development of data capabilities led to the establishment of the Michigan Institute for Data Science (MIDAS) in 2015. A "Where's My Data' initiative led to more informed data practices. [22] This collaborative effort has resulted in a cohesive campus vision for computational research with easy to navigate service catalog, coordination of software licenses and easily discoverable

---

[18] FAIR and TRUST were defined earlier in this document. The CARE principles refer to "Collective benefit, Authority to control, Responsibility, and Ethics," and address Indigenous Data Governance. See https://www.gida-global.org/care

[19] (2015: Office of the Vice President for IT & CIO … https://it.umich.edu/vpit-cio)

[20] 2012: $12 million dollar advanced computing center; 2014: Advanced Research Computing (https://arc.umich.edu/); 2019: Running Academic Supercomputing like an Enterprise https://www.nextplatform.com/2020/01/21/the-university-of-michigan-gets-serious-about-supercomputing-roi/
https://michigan.it.umich.edu/news/2019/10/30/armis2-high-performance-computing-cluster-now-available/

[21] https://timeline.it.umich.edu/

[22] https://michigan.it.umich.edu/news/2019/04/29/how-a-simple-question-grew-into-a-vision-for-a-data-informed-future/

procedures and services to ensure that researchers on campus find research to be open, transparent, and easy. (See also Appendix 5 for more peer institution comparison.)

4. Increase Ranking and Reputation and Improve Faculty Recruitment and Retention

As funders and academics compel us to move with the idea that experimentation begets data, data begets publications and publications lead to reuse, the wide adoption of open science and open data are becoming more ingrained with faculty experience. Based on our peer evaluation, investing in critical research infrastructure and data services is key to building experience with and demonstrating the research benefits of open science, and attracting and retaining faculty engaged in data-intensive scholarship. This inevitable shift to collaborative research does not lend itself to incremental change or siloed approaches, and for UCLA to maintain a leading role in the UC system a major investment in central research data infrastructure is required.

Research has been transformed by the increased complexity and scale at which data can now be collected. Use of machine learning and simulation has trivialized research that once took years to complete; they are now an ingrained part of an already competitive research environment. This transformation was also driven by the emergence of fair data, open data, and reproducible computational analysis and is inextricably linked to the cost of data collection, storage, and computation. Methodologies that use complex computational analyses are increasingly the foundation of research and publication. The added infrastructure, technology, policy, compliance, and training requirements for modern research are straining faculty,[23] and adding unexpected training requirements for graduate students that departments and programs are unable to support.

Given these realities, the academic institutions that can distinguish themselves in support of emerging technologies and computational research will clearly have an advantage in attracting and retaining world-class researchers.[24] An investment in campus research infrastructure would also provide a robust on-boarding process for new faculty and allow for campus savings in terms of allocating existing resources rather than creating new technical environments everytime a new faculty is recruited. This is not a massive investment given the success of UCLA in recent years and will address disparities that exist on campus between departments that already provide a wide range of data services with departments that are unable to do so. Indeed the authors of this report believe that if UCLA does not invest in central data services, the sprawling campus environment, as redundant, expensive, and inscrutable as it is, may collapse upon itself.

We need a whole campus approach to research infrastructure and research data services. This committee believes that a coordinated approach can be simultaneously fiscally positive and

[23] Infrastructure shortage may lead to faculty and researchers' dissatisfaction and moving elsewhere (Technol Innov. 2013 Dec 1; 15(3): 269–280. doi: 10.3727/194982413X13790020922068)
[24] 14% faculty did not accept job offer and 7% left UC system due to inadequate facilities (Ref: https://lao.ca.gov/reports/2012/edu/uc-faculty/uc-faculty-121312.aspx)

impactful for the research community. A coordinated, multi-departmental model to collaborate to solve this systemic problem will not only lead to a more financially responsible UCLA, but will lead to a better research environment for all faculty and a more logical approach to IT and data services for staff.

5. Enable and Complement the DataX Initiative

In 2021, UCLA announced the DataX Initiative to improve data science literacy among students and promote data-oriented research across campus. This new data science and scholarship initiative will hire faculty and postdocs, promote interdisciplinary collaboration, design cluster courses and aid with graduate student research over the next three years to help students and faculty learn more about conducting data-intensive, data-informed, or data-critical research.

DataX will be funded with $10 million from UCLA during its first three years. Appointed in mid-2022, DataX Acting Director Safiya Noble has been charged with recruiting outstanding faculty, creating an, "exciting and diverse intellectual and educational community," identifying, "new opportunities in research and education," fostering "new interdisciplinary connections across campus," and bring to life a, "nimble, innovative and well-run entity that fulfills the mission of the DataX Homeworld."[25] The vision for DataX includes – and is dependent on – a "DataX Support Network" that focuses on building capacity in using data across campus and calls for staff "fluent in the language and culture of the disciplines comprising the Division or School that they serve" to "aid in data acquisition, curation, manipulation, analysis, sharing, and management."[26] This broad vision requires the research infrastructures and data services proposed in this roadmap for campus. Campus interest in the DataX initiative reinforces the demand and need for secure, reliable research cyberinfrastructure. Meeting those demands will mitigate the potential risks and consequences for UCLA related to an increase in data-driven research.

6. Modernize UCLA's Research Infrastructure

Implementing all aspects of the roadmap will begin the much-needed process to modernize UCLA's overall research environment. Funding allocations in past years have followed a federated, or siloed, model where redundant services are developed within specific departments or divisions. For some units, this has allowed for the rapid development of resources catered to the discipline-specific needs of their targeted audiences. For others, due to disparities in academic funding models and research interests, they are unable to develop resources at the scale required for their constituents. This decentralized model perpetuates disparities in data literacy education across departments, and groups in certain disciplines have considerable support working with data and digital research, while other units struggle. The end result of this decentralized approach is an uncoordinated and fragmented support network that does not address the needs of the modern academy, fosters inequities in the distribution of computational

[25] Wakimoto, Roger. Email 'Appointment of Safiya U. Noble as Acting Director of DataX' sent on 6/9/22.
[26] https://www3.research.ucla.edu/datax

research, and hinders data discovery and collaboration. While some departments are more than able to cover their user training – like QC Bio – others must rely on tool-based workshops for generic audiences that focus on developing basic digital skills, but stop short of discipline-oriented methods and tools that are sensitive to and aware of disciplinary cultures.

A research and teaching focused institution like UCLA needs both basic data and digital fluency training as well as focused connector courses. By raising the floor, we create new opportunities for research by giving access and training to use technologies in disciplines where these resources are less abundant. Moreover, providing robust centralized solutions minimizes duplicative effort. Enabling research at an institutional level requires the coherent, robust, and transparent environment described in this roadmap document. Baseline infrastructure and data services that are easy to find will streamline innovation and collaboration for all researchers. Departmental or divisional support will be able to take advantage of this infrastructure while catering to unit- and/or discipline-specific needs.

7. Enable Multisite Grant Solicitations

By implementing this roadmap, UCLA faculty are in much stronger positions to participate in or lead large multi-institutional proposals. At the same time as our critical funders are pushing for larger, more diverse, interdisciplinary, and multi-perspective research, they are also calling for compliance with new open data sharing policies. When compared with other UC schools, UCLA may be missing significant economic opportunities when winning sponsored research awards. While UCLA is near the top ranking for awards issued, second only to UCB, UCLA is near the bottom, barely ahead of UCD, in terms of per-award values. One possible explanation for the lower per-award values may be the lack of support or infrastructure that would enable competitive submissions for larger and multi-institution solicitations. This 15-30% difference in award value represents a missed opportunity to include more direct costs for shared infrastructure services and missed opportunities to compete for larger, multi-site solicitations.

NSF AWARDS – FIVE YEAR AVERAGES[27]

Institution	Annual Average	Award Average	Average # Awards
UC Santa Barbara	$52,642,200	$427,708	127
UC San Diego	$74,538,600	$411,669	186
Stanford	$67,260,800	$393,089	175
UC Berkeley	$75,410,000	$376,572	211
UC Irvine	$45,659,200	$311,264	156
UCLA	**$59,998,600**	**$307,474**	**200**
UC Davis	$43,409,250	$302,687	149

[27] See NSF's Budget Internet Information System, https://dellweb.bfa.nsf.gov/starth.asp

In contrast, UCLA ranks second in per-award value in NIH funding in the UC system. It is impossible, however, to make a direct correlation between the award average and shared research infrastructure because not all NIH awards are dependent on data/compute infrastructure.

NIH AWARDS – FIVE YEAR AVERAGES

Institution	Annual Average	Award Average	Average # Awards
UC Lawrence Berkeley Lab	$27,353,910	$868,967	37
UCLA	**$505,164,887**	**$576,769**	**872**
UC Santa Cruz	$43,711,140	$537,299	81
UC Davis	$247,467,915	$495,096	500
Stanford	$533,909,195	$494,904	1,077
UC San Francisco	$664,265,751	$494,122	1,345
UC San Diego	$489,287,281	$489,460	998
UC Irvine	$152,587,848	$448,153	340
UC Berkeley	$139,363,986	$394,581	353
UC Santa Barbara	$19,537,953	$373,150	52
UC Riverside	$31,725,742	$353,151	89
UC Merced	$5,435,407	$263,621	20

8. Address Funder Mandates

Funding agencies have recognized the importance and greater value of research results that are shared. Data management and sharing requirements are in place with all of our major extramural funding agencies including the NIH, NSF, and the NEH. Furthermore, such sharing follows the most recent OSTP directive that requires taxpayer-funded research be made available to the public immediately and at no cost. Future funding depends on institutional compliance with these policies. Implementing this roadmap will provide researchers with data, computation, policy and partnership infrastructure services that are aligned with current and expected funder mandates for data management and sharing.

Robust institutional data services are required to maintain compliance with these funding mandates. These requirements are continuously evolving with directives for ensuring reproducibility of research up to ten years; export controls on data and technology; national security requirements; security, privacy and accessibility requirements on the data and the research; and additional specific requirements for data management, sharing, and preservation.[28] For example, the National Science Foundation requires that investigators share

[28] Proposal & Award Policies & Procedures Guide (PAPPG) Chapter XI.D.4

with other researchers “the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.”[29] Similarly, the National Institutes of Health’s new data management and sharing policy that goes into effect on January 25, 2023, requires sharing of all scientific data “of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications.”[30] Non-compliance “may be taken into account by NIH for future funding decisions for the recipient institution,” introducing a significant risk for the campus.[31] (See Appendix 6 for more information on the history of open-science and open-research policies.).

9. Respond to the UC Policy on Research Data

The new University of California Research Data policy[32] released 8/8/2022 and effective 7/15/2022 is among a constellation of data policies from publishers and funders that have begun to emerge in the support of research transparency and openness. This policy states “The Regents of the University of California owns Research Data generated or collected in the course of University Research” and asserts that “The University of California is committed to disseminating research results as widely as possible. As part of this mission, and in accordance with numerous University policies, the University supports the free and unfettered dissemination of information, knowledge, and discoveries generated by University Researchers.” Finally, the policy establishes that “Principal Investigators are the primary stewards of Research Data,” but the VCR is responsible for the implementation of the policy, with a specific directive to “Provide guidance to promote campus compliance with Research Data obligations arising by law, regulation or agreement.”

Implicit in providing guidance is having a suite of solutions for researchers and infrastructure that supports data management and publishing. Disseminating and preservation are key factors in UCLA’s ability to comply with this policy. It is remarkable that we lack the staff and infrastructure on campus and within the UC system to do the work required to support a $1.7 billion (and growing) research enterprise.

10. Control and Consolidate Costs on IT Spend

A review of campus IT spend excluding the UCLA Health Sciences reveals the distributed − and therefore uncoordinated − nature of our IT environment. Based on purchasing data gathered by

---

NSF Data Management Plan requirements
Requirements by Directorate, Office, Division, Program or other NSF Unit

[29] https://www.nsf.gov/bfa/dias/policy/dmp.jsp

[30] See NIH policy at https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policy/research-covered-under-the-data-management-sharing-policy#research-covered-by-the-2023-data-management-&-sharing-policy

[31] See https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policy/data-management-and-sharing-policy-overview

[32] https://policy.ucop.edu/doc/2500700/ResearchData

Deloitte for the IT Services Transformation effort, in FY21, UCLA spent nearly $7M on what is termed 'enterprise class' storage; 90% was at the departmental level. Similarly, 70% of server spend is also distributed across UCLA.[33] Coordinating and consolidating this distributed IT spend would increase efficiency in three ways: 1) it would better align IT spend with academic and programmatic goals, 2) it would enable lower per TB costs when purchased in bulk, and 3) it would reduce the amount of staff time required to research, purchase, and administer the equipment.[34] Moreover, this distribution of equipment combined with the threat of potential data breaches creates a potential financial impact to local units who likely do not invest sufficiently in cybersecurity or cybersecurity insurance. Creating a UCLA-managed storage platform would distribute financial burdens and risks appropriately allowing the institution to leverage cybersecurity standards and enable local administrators to focus on more immediate research support. A shared research storage service would be leveraged by local IT units for use in their divisions, while gaining the economic benefits of managed services. See Appendix 7 for more information on UCLA's IT spend.

11. Minimize Risk of Breach and Data Loss

IT best practices support data security and privacy, and prevent data loss regardless of the cause (e.g., cyberattack, accidental deletion, equipment failure). These are capabilities that every researcher needs and wants but often finds out of reach whether due to cost, time, or expertise. The question for the institution is how these capabilities are provided so that researchers and the research itself are not burdened. Managed shared research cyberinfrastructure and practice offers the mechanisms to adequately address security and backup redundancy, control costs, and extend consistent IT best practices for all researchers, consistent in a way that is impossible in a distributed environment.

Providing managed shared research cyberinfrastructure is an insurance policy against the financial impacts of data losses due to infrastructure failure or data breaches. The average cost of a breached record is $146 per record.[35] The costs increase significantly per breached protected health information (PHI) or sensitive record from $429 in 2019 to $499 in 2020.[36]

Protecting UCLA's intellectual property is an institutional imperative. While campus security efforts focus on risks along the lines of the UCSF ransomware attack, research data sitting on sub-standard equipment and under-managed systems represent a disaster waiting to happen. Beyond enabling IT best practices, UCLA also needs environments suitable for researchers

[33] UCLA accounting object code classifications are not consistently used across the institution.
[34] In FY21, UCLA conducted over 40,000 transitions to acquire and use a myriad of IT services and equipment. If a transaction represents on average, 3 hours/transaction (cumulative across the lifecycle of the transaction) which equates to a cumulative 120K hours of effort which also equates to 54 FTE allocated to to acquiring non-comp IT. Furthermore, if 1 FTE=$120,000, then $6.4M of compensation expense applied to procure $105M IT services and equipment, representing a 6% 'tax' on acquiring IT services and equipment in the current IT digital ecosystem.
[35] Financial costs of data breach.
[36] Ibid.

working with sensitive or otherwise restricted data and a robust disaster recovery plan for research data. Funding managed and shared infrastructure is the most direct and effective insurance against that risk. We should not wait for an extinction-level event to occur before we evolve and bring UCLA's research infrastructure up to modern security, compliance, and backup standards.

12. Protect UCLA Against Fines and Penalties

UCOP has specific requirements and warnings for violations of federal export controls that can be fined up to $1,000,000 per violation, or ten years of imprisonment, or both. In addition, the Secretary of State may assess civil penalties, which may not exceed $500,000 per violation.[37] The civil penalties may be imposed either in addition to, or in lieu of, any other liability or penalty. Decreasing such risks of data loss in an economically feasible way requires increasing cybersecurity capabilities through managed shared research cyberinfrastructure.

Similarly, the consequences for noncompliance with federal export control on data and technology are very serious for both the University and the researcher (including fines up to $1,000,000 and/or imprisonment up to 10 years for individuals).[38] These penalties apply to single violations; multiple violations in the same project can easily result in enormous penalties. This is a significant risk and concern at UCLA, which touts a highly decentralized model of IT support where in contrast, a well-managed shared research cyberinfrastructure could better manage the risk exposure to the University.

Section 3: Improving Faculty Experiences through the Consolidation of Responsibility for Research Data and Infrastructure

The Research Data Working Group's first assertion is that the campus needs to formally consolidate and clarify ownership, responsibility, and accountability for initiatives related to research infrastructure and research data services. One of the key challenges of our currently decentralized research support network is that there is no clarity about ownership of issues or situations with campuswide implications. UCLA currently has a number of units supporting research in some capacity, a situation that creates two equally problematic possible scenarios related to campuswide issues: either this diffusion of responsibility leads to complete inaction (the 'surely someone else is going to take care of this' phenomenon) or multiple units are sucked into the leaderless vacuum and create duplicative and often contradictory solutions.[39]

[37] UCOP Ethics, Compliance and Audit Services
[38] US Dept of Commerce export regulations
[39] A case in point is NIH's January 2023 requirement that all submitted proposals include a Data Management and Sharing Plan. In the absence of clarity about the ownership of a campus response to

A new leadership position – a **Research Data Rockstar (RDR)**, if you will – with a support staff – **the Office of the RDR (ORDR)** – is needed to coordinate these efforts and implement the vision laid out in this document. This permanent position/unit will work with appropriate campus units, UCLA Health Sciences, the UC system and beyond to ensure that campus data services are robust, well understood, and responsive to anticipated needs. The mandate for ORDR will be threefold: 1) to support faculty and data-enabled research; 2) develop faculty-focused educational materials, guidance documents, and data-related training opportunities; and 3) interface with faculty and campus leadership to improve institutional systems that negatively impact data-related research and ensure transparency and accountability. Specific tasks in each of these three areas are described below.

1. **Supporting faculty and data-enabled research.**

    a. **Continued planning for campus research infrastructure**, including such things as services and resources to facilitate business analytics related to research proposal submissions, support data management and sharing, monitor infrastructure use, and develop pricing strategies to direct charge data sharing activities. Also included are discussions about infrastructure required for emerging scholarship areas (e.g., AI, ML) and support the DataX initiative. (See Appendix 8 for a proposed prioritization schema.

    b. **Developing a faculty incentives program** to promote research data sharing and communicate UCLA's embrace of open data and open science.

    c. **Coordinating Research Facilitation** by working with existing research concierge and facilitation efforts (e.g., the Research Administrators' Forum) to establish and build connections between local and central research support.

    d. **Appointing a data-sharing compliance lead** to work with PIs and staff in developing proposed plans and complying with award terms and conditions concerning data management and sharing. (See Section 5.7.c.i below.)

    e. **Representing UCLA** at appropriate data-related national and international conferences and in leadership roles with appropriate organizations (e.g., the Research Data Alliance, the Data Curation Network) to build knowledge and expertise that will inform ongoing research activities.

---

this issue, first nothing happened, then multiple units began drafting guidelines for their researchers: representatives from DGIT (the IT support team for the Health Sciences) and the Library hosted a series of brown-bag sessions on the topic, members of the RDWG began collecting information for dissemination, and the Clinical and Translational Science Institute (CTSI) team began planning a response. After some discussion, the campus response is now being coordinated by Clifford Kravit (DGIT) and Ibraheem Ali (Library) under the auspices of ORCA. Clarity of ownership and responsibility at the onset would have resulted in a timely and coordinated campuswide response without the confusion and duplicative effort.

2. **Developing faculty-focused educational materials, guidance documents, and data-related training opportunities.**

    a. **Creating a cohesive research-related service catalog and web presence**. This task would focus on identifying campus resources available to researchers following the model of the existing IT Services service catalog. This service catalog would be updated as new resources are made available and could be organized around the research and/or research data life cycles. A distinguishing characteristic of exemplars in the peer analysis was the inclusion of an expert network as part of the service catalog.

    b. **Creating a funder and compliance requirements educational campaign** (e.g., NIH DMSP requirements). This would also be accomplished under the auspices of the RDR working with the DSC, User Engagement liaison librarians, and ORA as appropriate. In addition to funder requirements, this item includes privacy, accessibility, and all other forms of legislative requirements.

    c. **Developing guidance on discipline- and domain-specific repositories** to be used by UCLA researchers. A key assumption is that faculty will continue to use a range of repositories depending on their requirements, the prevailing repositories in use by their disciplines and/or research communities, and their publishing needs. The ORDR is intended to support faculty throughout the research process to find the necessary resources and, ultimately, the necessary repository for their data. So, moving forward, faculty will be encouraged to use all of them as appropriate to their situations.

    d. **Developing a faculty outreach and training program** in conjunction with the OCISO on research-focused security, privacy, and compliance issues.

    e. **Developing a Data Literacy Effort** connected with DataX to teach data skills across the campus to develop the next generation of data workforce and ensure fluency with the new technologies that are shaping our new forms of knowledge production. The goal is a coordinated effort that includes campus computing centers and the Library to develop a ‘data forward, data literate’ community.

3. **Interfacing with faculty and campus leadership to improve institutional systems that negatively impact data-related research and ensure transparency and accountability.**

    a. **Reviewing processes** that negatively impact research with an eye toward streamlining (e.g., the delays in IT procurement due to specialized ITS reviews that are not adequately staffed to ensure quick turnaround times). Work with campus offices and the Academic Senate to advocate for changes that will

address challenges encountered by researchers (e.g., campus restrictions that force PIs to engage with outside partners capable of processing flexible transactions).

b. **Working with campus stakeholders to create a long-term plan to restructure compliance-related efforts** so that they are more cohesive and legible for researchers.

c. **Streamlining of processes related to compliance requirements** which should, at minimum, address TPRM, security, accessibility, and privacy. Also coordinate BAAs affecting campus, the Health system, and other groups to standardize systems and software where possible.

d. **Convening an Accountability Board** to ensure thorough review of proposed research data practices and campus use with an eye toward ethical and equitable data use. Key is the involvement of faculty with expertise to guide choices made about infrastructure and think through the complexity of data use from a systems perspective including, but not limited to, environmental considerations related to data that inform UCLA practices, and the right to deletion and protection of materials.

The creation of this Research Data Rockstar position along with support personnel is in keeping with reorganizations taking place at peer institutions. Where this unit sits on campus is up for debate (see Section 5.2 for that discussion), but what is important is that there is a central team that has responsibility to make a broad campus-wide impact. At minimum, the campus needs to identify the appropriate unit to take the lead on all research infrastructure and data resources and services, and communicate that information to the campus community.

Section 4: Enabling Faculty with Research Data and Infrastructure Solutions

Enabling the faculty-focused benefits and opportunities described above requires a serious investment in shared research infrastructure and data services. Data-intensive research requires secure infrastructure and services with the capacity to seamlessly gather, manipulate, analyze, visualize, share, and publish data. Areas of critical need for faculty success have been categorized into three key areas for inclusion in the research data and infrastructure roadmap: 1) the storage platform with attendant research data support services and tools (aka Bruin Data Hub); 2) expansion and streamlining of existing research data services; and 3) expansion or creation of research infrastructure independent of the storage platform. These initiatives, in tandem, are needed for UCLA to maintain and grow as a research institution. Furthermore, this vision will enable the use of data for teaching the next generation of Bruins a level of data fluency that they need in the modern global economy.

1. Development of a Storage Solution/Technology Platform

The RDWG recommends equipping the campus with a technology platform – referred hereafter as the Bruin Data Hub for simplicity's sake – that combines high-performance storage and software to create a hybrid on-premise and cloud solution that is scalable to meet current and future research data needs. As envisioned, this platform will be the home for campuswide research data services; include collaboration tools to facilitate interdisciplinary work; and will provide storage space for members of the UCLA research community, shareable lab directories for PIs, and secure data access from server environments and workstations. Moreover, this solution will enable the ability to store and protect data acquired from instrumentation via backup or snapshot protection.

At the core of this vision is the idea that this service will be made available to researchers through their local IT teams to augment their own unit's research support offerings. During faculty onboarding or as desired, these local IT teams would provision space for researchers, who could in turn manage and further partition that space for their different research projects and collaborators. The local IT units could also create departmental, divisional, or school shares as desired and manage them on behalf of their constituents. Another key feature of the proposed Bruin Data Hub is its adjacency to associated research data services. These services are described below, but are planned to include a research data commons to facilitate campus access to structured datasets, a support stack for researchers working with databases and datasets, and a suite of web hosting and sharing tools. This new storage platform would be enterprise grade and jointly managed and governed by IT Services and its campus IT partners. Furthermore, this enterprise class storage platform could be the same infrastructure as administrative computing further expanding the economies of scale. When fully implemented, every tenure-track faculty member will be entitled to an allocation of storage and will be able to acquire additional amounts of storage as their needs demand within the same namespace or location where their data currently exists, eliminating the diaspora of data across different storage platforms.

Beyond providing infrastructure for our faculty generally, there are two other compelling drivers to implement this solution. The first is DataX. A basic premise of the DataX initiative is that data and information are the lifeblood of research and the capability to access, manage, analyze and present data and information represents core competencies needed in an ever-increasingly connected digital ecosystem. Providing research infrastructure and services for this initiative is in essence the pilot phase of the storage solution/technology platform. The second driver is a pending limitation on the amount of data UCLA can store on Google Drive under the terms of our existing subscriptions. Approximately 10PB of research data must be migrated by June 2024 to stay within these limits. After June 2024, UCLA will need to pay Google an additional $1.8M per year to continue storing this data in its current location. After these two drivers are

addressed, the storage-for-all component would be expanded and rolled out to all campus departments, divisions, and schools.[40]

The elements of this storage solution/technology platform:

1. **The hybrid on-premises/cloud-based research storage** (aka Bruin Data Hub) provides the UCLA research community with a secure, high-performance platform on which to store, process, and share research data. The platform will provide researchers with the ability to match storage to compute and network resources across on-premises and cloud environments. Additional tiering capabilities will provide cost-effective archive and cold storage facilities to meet needs throughout the research data lifecycle. As part of a broader ecosystem, the Bruin Data Hub will provide the core storage support for the research data services described below. (For a more complete description of technical features of the Bruin Data Hub, see Appendix 9.)

   The plans for the Bruin Data Hub are aligned with the goals and design of a new UCLA Health Research Storage Service being developed by the DGIT team that will leverage a highly redundant, software-defined storage platform optimized for large-scale research data.[41] To meet UCLA Health Sciences security and compliance standards, this service will not be designated for non-health users, though it is critical that the two systems are coordinated to the degree that they can seamlessly support researchers with affiliations both on campus and with UCLA Health Sciences.

2. Associated research data services:

   1. **Data Transfer Nodes (DTNs)** will be deployed and configured to facilitate the movement of data into and out of UCLA's storage solution and networks at the highest rates possible. DTNs not only enable researchers to move data quickly, they improve the security posture of research institutions by reducing the number of systems visible to the outside world, providing a smaller number of well-understood resources to defend against attacks.[42]

   2. **Relocation of data currently stored in Google**. Approximately 10PB of research data is being stored on Google Drive in excess of the storage limits recently imposed by Google. UCLA has until June 2024 to move this data to an

[40] See use cases in Appendix 3. The transition to use of a central networked solution may be challenging for existing faculty with established workflows; new researchers should require no adjustment at all as their local IT unit will offer them access as part of their standard onboarding process.
[41] A UCLA Health Sciences active directory account is required to access the proposed service. It will provide a secure, reliable, and stable environment designed to meet basic needs. A free-tier entitlement and cost for additional capacity are TBD. The system supports data-aware scale-out NAS, unlimited individual file sizes, a simple high-performance file system accessible from any Mednet-connected device (including off-campus access via UCLA Health Sciences VPN), snapshot strategy delivers 90-day retention, and compliance with most legally or contractually restricted data such as Personal Health Information (PHI) and Personally Identifiable Information (PII).
[42] See https://www.es.net/science-engagement/technical-and-consulting-services/data-transfer-nodes/

alternate location to avoid negative impacts to all users of our Google Workspace for Education (Gmail, Google Drive, Google's productivity software, etc.) instance. IT Services will work with the research community to move this data from Google to the Bruin Data Hub.

3. **License server**. Providing a consolidated license management infrastructure will enable more efficient and effective use of the myriad software licenses across UCLA, ensure license compliance, and facilitate use of these licenses by those organizations that do not typically participate in using more advanced or capable software for a variety of reasons. The initial phase of license servers will leverage IT Services' existing VM infrastructure and IAM initiatives underway. When completed, the combination of group security policies of the new IAM will simplify granting/revoking access to software licenses. A central license server will also enable better management of the portfolio of software applications across the UCLA digital ecosystem.

4. **Research mosh pit (aka data commons)**. A critical piece of the storage solution/technology platform is the creation of a research data commons that is integrated with the research process to facilitate the sharing and use of active working data amongst researchers, their labs, and their colleagues. The commons would provide storage, access, and transfer mechanisms to enable researchers to share data sets, code, and tools across teams and disciplines, and introduce a level of control over gathered, purchased, internally generated, and/or aggregated data. Researchers could access data either through a network-mounted drive or through a web portal. Building this service on top of the proposed storage solution/technology platform would ensure performance and efficiency, and facilitate long-term data preservation and access. This sort of service is common at peer institutions,[43] and rightly treats research data as an institutional asset.[44]

5. **Database/dataset tools and service**. A central service and support stack for self-service use by researchers working with databases is a critical component of the Bruin Data Hub. Providing this 'database as a service' (DaaS) would give UCLA researchers an easy way to set-up, host, and maintain their relational databases, and allow them to use their university credentials to create databases through an online portal. Implementation would support MySQL, NoSQL, and PostgreSQL, three of the most common databases used by researchers.

6. **CoLab/JupyterHub Environment**. With the increased demand for data capabilities across all disciplines in the arts and sciences, teaching student researchers how to 'compute' and 'code' collaboratively is creating a demand for

[43] See Texas Data Repository, Illinois Data Bank, Penn State DataCommons@psu, Duke The Research Computing group, Stanford's Data Commons, *et al*.
[44] See also https://www.datacommons.org/about

easy-to-use electronic notebook experiences. Jupyter is an increasingly attractive environment for interactive programming and data analysis and visualization. Supporting JupyterHub on the Bruin Data Hub will make it easy to provision and standardize a computing environment for researchers and provide faculty with a useful teaching tool, allowing greater sharing of teaching material and providing a straight-forward path for taking research data into the classroom.

7. **Web hosting/sharing tools**. Providing no- or low-cost web services within the Bruin Data Hub would result in considerable benefit to the campus. It would enable workload efficiencies for staff (freeing up staff expertise for more impactful activities); eliminate redundant services; introduce a level of control for security, accessibility, and branding; and introduce efficiencies for faculty and researchers. Sharing would be enabled on a per-PI or per-project basis, and web-based and command-line utilities would enable management and auditing of shared information and data.

Researchers have also shown interest in applications of virtual machines and containers, and these could be layered onto the storage solution/technology platform and aligned with the research data commons. The final list of tools and services to provide through the storage platform should be confirmed through discussion with faculty researchers.

2. Research Data Services Expansion

This section will focus on institutionally funded ‘core’ research data services. As articulated throughout this report, expanding and streamlining existing research data services is necessary to respond to funder and scholar mandates and expectations. While many of these services have been recently initiated and piloted at UCLA, they will need to be expanded and more fully funded to meet expectations as outlined in this roadmap. For example, the library's UCLA Data Science Center (DSC) is an expansion and revision of the 1970’s era Social Science Data Archive (SSDA) built primarily to steward survey research and regional census. Since moving into the library, the DSC has seen an expansion of services and staff, but these core services will need to be further expanded to meet the growing demand for campus data scholarship requirements. An administrative home for these services in the library makes sense as the established program can be leveraged to serve a larger number of faculty and the library has an existing core outreach service to all of campus.

Based on our peer research, significant community-building and data infrastructural gaps have been identified. Providing instruction and support on producing well-structured and described data and code is key to supporting research with integrity and impact. Improving data and software practices is critical in preserving UCLA’s scholarly record and making research outputs actionable and computable for reuse and advancing knowledge. The current service portfolio at the DSC lacks sufficient resources to expand at the scale and pace campus needs in providing interdisciplinary-oriented transparent and open research best practice research support. The DSC provides foundational data science and coding education through its membership with the

Carpentries, a community-of-practice train-the-trainer educational program. The DSC, through soft funding, has had a significant impact on improving digital skills but needs to systematize and connect with more departments and programs.The Library needs data and data science professionals to amplify this support even though the faculty already expect that level of support. Due to this scaling issue, the DSC focuses mainly on short-term and best-practice consulting with researchers but cannot engage more intensely. UCLA needs to meet those expectations with an infusion of support. This nascent computational research support unit is already stretched due to the scale of the required service.

1. **Digital/Media Asset Management System** (DAMS/MAMS) This research object management software sits on top of our storage solution. It offers a self-deposit option for researchers focusing on the metadata that makes research data discoverable, findable, downloadable, and storable. Researchers can search, retrieve, upload, curate, store, create identifiers for, publish their research and manage any derived objects from this software layer. The DAMS/MAMS enables data asset management, curation, publishing, archiving, and preservation. The DAMS/MAMS builds on and augments UCLA's existing UCLA Dataverse instance, a data publishing platform, within the DSC.

2. **Expansion of consulting services and expertise**. The additional work required to support faculty and address the complexity of ever changing mandates and compliance challenges will require new dedicated staff.

    a. Expansion of the Data Science Center to provide additional training on basic data practices, expertise on methods and computational practices, access to data tools (no-code and code-based), and support for different data types, particularly data with more sensitive protection levels for on-premises or cloud research services. This need for assistance and training on all things related to research data is felt on both campus and in the UCLA Health Sciences. The creation of a new support area within the DSC to focus on data to be used in a publication and to develop software, code, and work with researchers to adopt good computing practices, so their work is more reproducible. Focusing on openness and transparency, the service will inculcate best practices in preparing datasets for publication in UCLA's Dataverse or appropriate discipline-specific repositories.

    b. Hiring generally assignable data scientists across campus will open the possibility for innovation hires and enable the use of and training on discipline-specific data and/or up-and-coming technologies, including but not limited to, Machine Learning, Density Storage, NLP, Computer Vision, Quantum Computing, and the Internet of Things.

3. **Data Services Outreach & Education Program**. Development of a training program for local support (e.g., unit-level faculty-facing staff, postdocs, grad students) that teaches advocates how to promote best practices in their domains (e.g., FAIR Open Science Principles), the repositories chosen for their work (e.g., those adhering to the TRUST

Principles), and sociotechnical issues related to data use.[45] This unit must work with the Office of the Research Data Rockstar, DSC, User Engagement librarians, and campus data professionals in support roles.

4. **Primary Source Data**. The UCLA Library represents one of the major Archives and Special Collections on the West Coast. The post-COVID impact on the library is years of backlog of primary source collections. One of the biggest challenges the Library has is processing these collections. Making them useful to scholars is very labor-intensive, but the potential scholarship is hugely impactful. Staff support is required to process these collections as well as apply computational processing like tabular OCR, translation services, and image tagging to already digitized collections. Accelerating the forward migration of large existing data collections on brittle technology would serve the campus and the larger academic community.

3. Research Infrastructure Expansion

In addition to the development of a new storage and technology platform, supporting research at UCLA will also require a commitment to expanded investment in existing infrastructure and the development of much-needed services that address specific research support gaps. The campus, for example, has a centrally managed and shared HPC cluster that is tailored to its current user population; expanding access requires equipment appropriate to the data-intensive needs of broader research communities. Support for researchers working with P3/P4 level data is an example of a critical gap in our research infrastructure. Enabling those researchers requires an investment in infrastructure that will meet their needs; the alternative is forcing these researchers to identify and implement solutions independently and potentially outside of UCLA, introducing a level of risk for the institution.

1. **Expansion of On-Premises Compute and Storage.** Operated by the Office of Advanced Research Computing, the Hoffman2 cluster is UCLA's centralized compute resource for campus researchers. The cluster includes both a general campus-use section that is freely available to all interested researchers and a section where researchers with significant computing needs can purchase and contribute nodes for their own priority use. Hoffman2 resources for researchers includes a job scheduler, compilers, applications, and software libraries, including discipline- and domain-specific software. The cluster is heavily utilized by researchers. On average in the past six quarters, the cluster enabled the work of 266 research teams and 1,026 active users that consumed over 22 million CPU hours with 8.5 million submitted jobs.

    a. **High-memory compute**: The widespread adoption of computational techniques for big data analysis and the development of new training and simulation algorithms have led to an increasing need for researchers of all disciplines to have access to compute hardware with large amounts of memory. These nodes

[45] See Gregory et al "Understanding data search as a socio-technical practice" 2019. https://journals.sagepub.com/doi/10.1177/0165551519837182

will be integrated into the existing Hoffman2 cluster to both support general users and allow for benchmarking and job scaling by researchers contributing to the Hoffman2 shared cluster. Following an initial investment, funding would be needed on a three-year refresh cycle.

b. **GPU nodes**: Many cutting-edge software packages used by UCLA researchers in support of publications and grants have been optimized for or require GPUs. Modern GPUs are needed to allow for benchmarking, preparing for computational runs on national-class HPC systems, and supporting development of computational research software by the UCLA researcher community. These nodes will also be integrated into the existing Hoffman2 cluster. Following an initial investment, funding would be needed on a three-year refresh cycle.

c. **Compute nodes for general users**: As part of UCLA's commitment to providing equitable access to research computing resources, OARC provides a baseline of complimentary HPC cluster computational time to all eligible UCLA persons, including undergraduate and graduate students, faculty, and other researchers. New nodes are mandatory to accommodate the growth in demand from new sectors of the university community, to replace old nodes as they reach end-of-life and are removed from the cluster, and to minimize energy use and maximize space efficiency. Following an initial investment, funding would be needed on a three-year refresh cycle.

2. **P3/P4 compliant HPC/storage environment.** While the UCLA Health Sciences has secure solutions for protected health information, the campus does not currently offer a robust storage and compute platform that supports research involving sensitive, restricted, or otherwise encumbered data. There are four options for building out such a platform: 1) expanding the existing highly compliant environment housed in the OCISO, 2) building an on-premises solution on top of the storage platform described above, 3) creating a secure cloud environment as part of the cloud research strategy described below, and 4) combining the best elements of the other three options. If on-premises, the granular group-based security integrated with IAM and a multiprotocol storage and encryption platform such as tiCrypt will enable leveraging of enterprise storage for research usage with increasing security requirements. Depending on the sensitivity level needed, users of the new storage platform will use the same name spaces and interfaces familiar to them while layering new security and control as required by HIPAA and NIST specifications. This infrastructure permits orchestration to cloud to further gain economic scalability.

3. **Expand 'Cloud for Research' Offerings.** To complement on-premises options for storage and compute, the RDWG Cloud Subcommittee identified pathways to include cloud in our overall research support strategy. (See also Appendix 10.)

a. **Campus cloud support for research.** To address researcher demand and enable cost savings, UCLA should create managed and coordinated (facilitated) access to and use of public cloud services offered by the three major cloud vendors (AWS, GCP, and Azure) with the goal of creating an impactful user experience that decreases information security risks and extracts the financial benefits negotiated by UCOP.

   Due to the highly distributed nature of research and IT support, we propose a team be created that includes representatives from across UCLA and UCLA Health Sciences with interests and expertise in cloud services. This group would be charged to define, design, and implement research cloud cyberinfrastructure. This starting mission could be focused on a three-year scope with rotating leadership to promote inclusivity across UCLA. Furthermore, an initial topic could be to support the DataX initiatives with cloud-focused services and integration into the larger research data strategy within the proposed Office of the RDR.

b. **Testbeds/Technology Sandbox.** A priority of the cloud infrastructure team will be enabling use of cloud services for prototypes, pilots, and other testbed services that would benefit from flexible and easily configured storage and compute environments. This would facilitate research projects that require iterative development to demonstrate success before locking down a final production solution or proposing an on-premises hardware buildout. Moreover, the major cloud vendors are already positioned to support innovative work related to machine-learning, AI, and big data processing tasks.

c. **Possible solution for P3/P4 data**. As described above, there is a critical need to provide infrastructure solutions for researchers working with restricted data. Further discussions will be needed to define the best technical approach, but a cloud-only solution may be the fastest pathway to a secure and compliant environment.

4. **Coordination, expansion, and promotion of UCLA's core facilities**. Core facilities feature prominently in evolving plans for increased institutional efficiencies and discussions of UCLA's national and international competitiveness. Core facilities including (bio)chemical and physical analysis instrumentation, state-of-the-art clean rooms, laboratory animal facilities, computing, libraries, studios, and performance spaces are essential components of the baseline research infrastructure needed to maintain UCLA's status as an R1 research university. Funding is requested to coordinate, expand, and promote our core facilities as needed by the demands of our research scientists.

5. **Collaboration and research software and tools**. UCLA researchers currently enjoy access to a limited suite of software tools (e.g., Slack, Box, Google Drive, Adobe, Microsoft Office) In terms of improving existing services, the campus needs to better

identify software, databases, and web services available to users whether licensed for campus use or free and open source; simplify the process to purchase software using approved UC or UCLA agreements; and clarify policies related to purchasing and the installation of free software on campus computers. New services related to software and databases are also needed, including mechanisms for testing and evaluating software for future use; expanded offerings that address research-specific needs (e.g., time or project management software, Git); support for researchers needing help with software to solve a specific problem; and clarification on the process to initiate exploration of new campuswide licenses.[46]

6. **Research data backup.** UCLA is assuming a great deal of risk in the absence of a campus strategy for research data backup. “True” backups (onsite) should be provided for all research data, with a minimum of three versions of each file, and backup policies are needed to ensure sufficient backup coverage for files with high churn (i.e., those files with a high number of updates). As designed, the Bruin Data Hub will include backups for all networked data, but solutions are still needed for data stored on local networks and servers, in the cloud, and on endpoint devices. While there is a UC-wide research data backup effort underway at the time of this writing, it is not clear that the results will be satisfactory, so UCLA should follow the example set by UCI and begin separate negotiations for viable backup solutions.

7. **Streamline research proposal development process**. The proposal preparation and development process is complex, deadline-driven, and frequently has a short lead time. Proposals are high-stakes activities as PIs are seeking funding to support their staff and laboratories. PIs typically manage all aspects of the proposal development process, from academic argument to detailed budgets. Depending on their institutional home, they may have access to staff support, but this is inconsistent across campus resulting in divergent standards. If local support exists, PIs are free to focus on the intellectual pieces of the proposal, while staff take responsibility for understanding the totality of proposal requirements, including the required components, page limits, formatting specifications, budget and justification, facilities and other core components, and ensuring the overall proposal requirements are complete. These local support staff may also be responsible for pulling information together from various sources, including biosketches, current and pending support, budgetary data, and ensuring that the final document has met all funder requirements. The addition of the requirement for a data sharing plan and an estimate of its cost, adds an additional challenge to proposal development.

   Additional staffing and expertise at all levels of the process would support the increasing complexity of proposal requirements. Moreover, consolidation of information and changes to the submission process would benefit campus researchers.

   a. **Establish a single source for guidelines, policies, and templates** for developing data management and sharing plans and other funder requirements.

[46] See also the 2017 Research Data Support Recommendations.

This should include resources to support compliance with data management award terms and conditions. The current diffusion of responsibility has spawned duplicative documents with instructions for researchers and research preparers. This resource should also include educational campaigns developed in coordination with all relevant units and the ORDR.[47]

b. **Streamline workflow** for proposal preparation and development by implementing and promoting best practices in schools, divisions, and departments. Researchers and support staff should also be encouraged to make use of available tools and resources[48] and provide feedback to the ORDR on any additional proposal-related needs (e.g., digital versions of internal documentation, a cohesive proposal dashboard).

c. **Examine staffing related to the proposal development process** in order to improve proposal preparation timelines. Additional resources are needed at all levels to support PIs in the proposal development process. Department research administration FTEs are lacking in a number of schools and departments, and additional central staffing is needed to keep pace with the growth of the campus research program and to improve turnaround times for award negotiation and subcontract issuance.

    i. Develop resources to support research data compliance monitoring and support across campus. Appoint a data sharing plan compliance lead in ORDR to work with PIs and staff in developing proposed plans and complying with data management/sharing award terms and conditions.

    ii. Publish a responsibility matrix that clearly identifies the assigned points of contact for various types of data-related agreements.

d. **Establish appropriate mechanisms for communicating data management and sharing plan requirements.** Determine methods for communicating information about data support services and resources to PIs (e.g., information about DMSPs, source of data to be used, data use agreement tracking, repositories to be used, storage requirements/locations, data ownership, compliance requirements). This communication plan should include the sharing of business analytics about funded grant proposals to inform the ongoing development of research data services and infrastructure.

[47] See https://osr.ucsf.edu/ and https://blink.ucsd.edu/sponsor/ocga/team-portfolios.html for UC examples.
[48] Current resources include training programs offered through the Office of Research Administration, ORA Resource Center Tools, the PI Portal, and the Research Administrators Certification Council within the Health Sciences.

Section 5: Funding and Implementation

There is no question that UCLA needs to invest in research infrastructure and data services. The RDWG roadmap centers on two areas of spend: 1) the Research Data Rockstar (RDR) and their support staff, and 2) research data and infrastructure solutions including the development of the Bruin Data Hub with attendant research data support services and tools, expansion and streamlining of existing research data services, and expansion and streamlining of existing research infrastructure. The authors fully acknowledge that the final roadmap and funding will be the result of conversations with faculty and campus leadership. As a point of reference for these discussions, if the roadmap as described in this document is implemented, the six-year costs are slightly over $74 million, with $19 million in capital expenses or one-time initial costs and the remaining $55 million in recurring costs. The table below provides a top-level look at costs for both full implementation and for what the authors believe is the minimum viable product (MVP). (This MVP has a six-year total just shy of $50 million, with roughly $16 million in one-time costs/$34 million in recurring costs.)

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total	MVP
1. CONSOLIDATE RESPONSIBILITY FOR RESEARCH DATA AND INFRASTRUCTURE								
1. Research Data Rockstar	$444,000	$457,320	$471,040	$485,171	$499,726	$514,718	$2,871,974	$2,871,974
2. Research Data Rockstar Support Team	$222,000	$450,660	$464,180	$700,105	$721,108	$742,742	$3,300,795	$1,435,987
3. Submission Compliance Lead	$296,000	$304,880	$314,026	$323,447	$333,151	$343,145	$1,914,649	$1,914,649
4. Space Required	$123,900	$4,017	$4,138	$4,262	$4,389	$4,521	$145,227	$145,227
5. Travel and Conference Attendence	$9,000	$9,000	$9,000	$9,000	$9,000	$9,000	$54,000	$54,000
6. Website development	$243,400						$243,400	$243,400
Work to enable business analytics from funded awards	$244,200	$251,526	$141,312				$637,038	$411,708
SUBTOTAL (YEARS ONE THROUGH SIX)	$1,582,500	$1,477,403	$1,403,695	$1,521,985	$1,567,374	$1,614,126	$9,167,083	$7,076,945
2.1. DEVELOP A STORAGE SOLUTION/TECHNOLOGY PLATFORM (AKA BRUIN DATA HUB)								
1. Storage Solution/Technology Platform (15 PB)	$6,182,000	$944,210	$839,026	$848,447	$858,151	$6,868,145	$16,539,979	$16,539,979
2. Associated Research Data Services								
a. Data Transfer Nodes	$15,000	$1,500	$1,500	$1,500	$1,500	$1,500	$22,500	$22,500
b. Relocation of Data from Google (10 PB)	$355,200	$355,200					$710,400	$0
c. License Server	$88,800	$91,464	$94,208	$97,034	$99,945	$102,944	$574,395	$574,395
d. Research Mosh Pit (Data Commons)	$557,200	$323,416	$340,818	$359,513	$379,615	$401,253	$2,361,816	$2,361,816
e. Database/Dataset Tools and Service	$133,200	$137,196	$188,416	$194,068	$199,891	$205,887	$1,058,658	$1,058,658
f. CoLab/Jupyter Notebooks	$531,400	$440,032	$454,157	$573,798	$483,980	$499,729	$2,983,096	$1,432,842
g. Web Hosting/Sharing Tools	$133,200	$137,196	$188,416	$194,068	$199,891	$205,887	$1,058,658	$1,058,658
SUBTOTAL (YEARS ONE THROUGH SIX)	$7,996,000	$2,430,214	$2,106,542	$2,268,429	$2,222,972	$8,285,345	$25,309,503	$23,048,848
2.2. EXPAND EXISTING RESEARCH DATA SERVICES								
1. Digital/Media Asset Management System	$532,800	$548,784	$742,848	$942,733	$971,015	$1,000,145	$4,738,325	$3,040,591
2. Expansion of Consulting Services and Expertise	$177,600	$182,928	$177,600	$177,600	$177,600	$177,600	$1,070,928	$0
3. Services Outreach and Education Program	$355,200	$365,856	$554,432	$926,265	$954,053	$982,674	$4,138,479	$2,440,746
4. Primary Source Data	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887	$1,148,790	$0
SUBTOTAL (YEARS ONE THROUGH SIX)	$1,243,200	$1,280,496	$1,663,295	$2,240,666	$2,302,558	$2,366,307	$11,096,521	$5,481,337
2.3. EXPAND EXISTING RESEARCH INFRASTRUCTURE								
1. Expansion of On-Premises Compute and Storage	$920,000	$360,000	$420,000	$860,000	$420,000	$360,000	$3,340,000	$3,340,000
2. Expand Cloud for Research Offerings	$749,000	$765,650	$782,800	$800,463	$818,657	$837,397	$4,753,967	$3,589,967
3. P3/P4 Compliant HPC/Storage	$605,200	$573,056	$590,248	$815,155	$839,610	$864,798	$4,288,067	$3,647,632
4. Core Facilities	$350,000	$385,000	$423,500	$465,850	$512,435	$563,679	$2,700,464	$0
5. Research Software Collaboration Tools	$500,000	$550,000	$605,000	$665,500	$732,050	$805,255	$3,857,805	$0
6. Research Data Backup	$475,000	$522,500	$574,750	$632,225	$695,448	$764,992	$3,664,915	$3,664,915
7. Streamline Proposal Development Process	$732,600	$917,378	$1,152,739	$1,041,770	$1,073,023	$1,105,214	$6,022,726	$0
SUBTOTAL (YEARS ONE THROUGH SIX)	$4,331,800	$4,073,584	$4,549,037	$5,280,964	$5,091,223	$5,301,335	$28,627,943	$14,242,514
TOTALS (YEARS ONE THROUGH SIX)	$15,153,500	$9,261,697	$9,722,569	$11,312,044	$11,184,128	$17,567,112	$74,201,050	$49,849,644

Where possible, the budget presumes effort from existing staff with offsets to backfill current workload. The typical costs associated with new hires (e.g., space, computers, workstations) are

only calculated for the staff in the ORDR; for all other new personnel, it is presumed that the hiring unit will absorb those expenses. Appendix 11 details the costs for the full implementation. A list of possible sustainable funding sources is presented in Appendix 12, although an option would be to fund the roadmap in its entirety as a capital expense, allowing the campus time to develop a sustainable funding model for research infrastructure and data services into perpetuity.

In terms of implementation, the most critical elements of the roadmap should be launched as quickly as possible, followed by expanded build-out and the hiring of additional personnel. Generally, staging implementation is the sound approach to controlling costs and building trust in centralized infrastructure and services, but in this case prioritization will be a challenge because UCLA needs so many discrete research support services. If the campus only had one gap area – the availability of a P3/P4 enclave, for instance – it would be easy to hire a third-party company to build the enclave and one person to support it with the intent to add more staff as demand grows. Paring down services is the only way to pare down staff. The discussion below considers options for staging implementation of the proposed roadmap elements and mirrors the MVP as outlined in the budget, though the RDWG would caution that a deferral of these services and infrastructure elements represents a risk for the campus because further deferral will disadvantage our faculty and increase the eventual cost.

1. Consolidation of responsibility for research data and infrastructure

> **Critical**: Hiring RDR and at least one supporting team member, and developing educational material and cohesive website related to research data requirements and available research infrastructure.
> **Phase 2 and beyond**: Hiring additional ORDR team members and creating an office for the unit.

2.1. Develop a storage platform/technology platform (aka Bruin Data Hub)

> **Critical**: Hiring an engineer to oversee the construction of a bare-bones storage platform, meeting our obligation to transfer data from Google, and prototyping the Bruin Data Hub experience with a select group of data-intensive researchers.
> **Phase 2 and beyond**: Layering in the envisioned research data services as prioritized by the faculty and as needed to accommodate new and emerging technologies and platforms..

2.2. Expand existing research data services

> **Critical**: Two immediate hires to begin working with the RDR and faculty on their data-related challenges.
> **Phase 2 and beyond**: Additional hires based on faculty demands for support.

2.3. Expand existing research infrastructure

Critical: Expanding the capabilities of the shared campus compute cluster to accommodate data-intensive research, developing a ‘Cloud for Research’ program, developing options for researchers working with P3/P4 data, and developing solutions for research data backup.
Phase 2 and beyond: Local and central hires to streamline the proposal process.

Section 6: Conclusion and Summary

We cannot ignore the success UCLA has achieved over the past few years, but we also cannot ignore the facts that our academic culture is changing and that we need to invest in research for us to sustain that success and better support faculty. It is also critical for the administration to recognise the significant impact of the rapid expansion of information technology and data generated by, and required for, research and teaching. As data-driven methods, such as machine learning, take their place alongside simulation and data as pillars of science, the same technologies are revolutionizing the humanities. UCLA needs to respond to these changes.

This roadmap is aligned with campus needs and ongoing initiatives, aligned with larger UCLA mission and goals, and, if adopted, will cement these services in a way that is cohesive and coherent. We request that our executive sponsors commit to this roadmap to expand UCLA’s research data and infrastructure roadmap and find funding to make it happen. Having agreed that UCLA must invest in our research and research data infrastructure, campus leadership must now resolve four key questions that center on funding, reporting lines, coordination, and boundaries in regards to research data outputs. Decisions on these questions must be made at the highest levels of campus leadership to set the stage for the much-needed expansion of UCLA’s research infrastructure. The RDWG would argue that this investment in research infrastructure and research data services is an investment that will pay off for UCLA in the long run by supporting faculty, streamlining processes, and providing a foundation for innovative research.

Section 7: Appendices

Appendix 1: RDWG Charge

Charge Letter for UCLA Research Data Working Group

Executive Sponsors

Roger Wakimoto, Jim Davis, Virginia Steel, Lucy Avetisyan, Ellen Pollack

Last updated September 21, 2021

Data are now critical components of research across every discipline and a key driver for academic innovation and reputational capital. Given the importance of research data to the enterprise, it behooves UCLA to develop infrastructure and services to protect this valuable investment and provide faculty with robust, efficient, and sustainable support mechanisms for their research and research data. Moreover, it is critical that the campus treat research data as an institutional asset equivalent to more traditional intellectual property and appropriately provide for its longevity and ensure its reuse.

There are a number of significant drivers that reinforce the immediate need for the creation of a Research Data Working Group. At the UCLA level, campuswide interest in the DataX Initiative focused attention on the need for hardened infrastructure that supports data at every phase of the research lifecycle. UCLA is also engaged in a push for efficiency and consolidation in the face of budget challenges that demand an organized, robust, and supportable campuswide response to persistent institutional problems. Providing data-related solutions to faculty that enable their research and engender collaboration and innovation addresses this mandate. Protecting, publishing, and preserving UCLA's research data will ensure its availability for reuse as well as signaling to funding agencies the institution's commitment to open data and future scholarship. The changes in leadership of the past year have similarly provided fertile ground for actions and initiatives that will transform the research environment and the research experience for faculty. At the UC-wide level, the recent UCSF ransomware incident has prompted ongoing examination of security and the protection of our most valuable asset – our intellectual property. Similarly, a proposed system-wide policy on Research Data and Tangible Research Materials highlights the importance of managing, curating, preserving, and making accessible our research data. Fundamental research data storage through vendors is also in play following the imposition of storage limits on UCLA's Box agreement and Google's announcement that after July 2022 it will no longer provide free and unlimited storage to the education community.

As UCLA's response to these drivers ramp up, the creation of a Research Data Working Group establishes a critical voice to advocate for research data and institutionally developed research infrastructure that protects and manages our intellectual property, and identify strategies to support the research data lifecycle within a system that is accessible, robust, and secure. The goal of the Research Data Working Group is to develop a high-level view of research and research data needs at UCLA, and establish an institutional support model for researchers built

around the necessary technical, physical, and human infrastructure that streamlines and incentivizes data best practices, encourages collaboration and data sharing, and ensures compliance, security, and accountability related to funder requirements. Put simply, the working group is being asked to acknowledge and address how the campus will identify, store, preserve, and disseminate our digital research output.

The working group should consider five key areas as they relate to research data: policies, technical infrastructure, personnel, pedagogy, and emerging research trends. The scope of this effort includes:

- UCLA's governance structure as it relates to research and research data;
- relevant privacy and intellectual property policies at local, UC, and national levels;
- funder, security, and compliance requirements;
- the potential impact, cost, and risk of data loss and prevention;
- data management and collection practices, reuse scenarios, and restrictions across disciplines and data types;
- incentives and rewards for faculty that support open data practices;
- the technical infrastructure and established best practices for data collection, storage, management, transformation, use, and reuse;
- mechanisms to strengthen, expand, coordinate, and scale existing data management support services and initiatives such as the workshops and consultation services offered through the Library's Data Science Center;
- infrastructure specific to data that is sensitive, restricted, or otherwise encumbered by compliance requirements;
- the requisite communication and education campaigns needed to build awareness of the benefits of research data management defined broadly, encourage adoption of best data practices, create trust in the process, and identify how to access research data support and services;
- funding mechanisms for infrastructure development and maintenance;
- the personnel infrastructure necessary to support data and its effective use and reuse;
- the ongoing support mechanisms required for research data management over the full research and data lifecycle;
- UCLA's funding goals and desired research outcomes; and
- the overall cost of ownership for research data (i.e., defining coherent approaches for protecting and managing reusable data).

To maintain UCLA's competitive advantage and stay abreast of peer institutions, a comprehensive institutional plan is necessary to appropriately support data across the research and data lifecycles. Faculty deserve a seamless research environment that supports their needs without introducing undue work, stress, and complications. UCLA, by creating an accessible environment for comprehensive data services, will encourage collaboration, entice funding, and illuminate requirements to anticipate future campus research data needs. Finally, such an environment will better position UCLA for large-scale funding opportunities that revolve around data-driven research at scale.

The working group is specifically being asked to:

1. Define the ideal campus future state based on priorities across the institution.
2. Conduct a gap analysis against UCLA's current research data needs and environment.
3. Work with related committees both within UCLA and at the system-wide level to coordinate focus and scope.
4. Reconcile competing priorities across UCLA units to coalesce on an institutional strategy for supporting research and securing research data across the full research and data lifecycle.
5. Identify mechanisms to strengthen, expand, coordinate, and scale existing data management support services, infrastructure, and initiatives.
6. Establish a series of operational and steering committees to develop implementation plans for infrastructure and initiatives related to research data (e.g., data storage infrastructure campuswide, research data policies, data management, faculty incentives for open data, technical solutions for restricted or otherwise encumbered data).
7. Charge a leadership group with the coordination of these groups with accountability for implementation outcomes.
8. Develop longer-term strategies related to data across campus disciplines and at varying levels of protection levels as defined by IS-3.

Appendix 2: RDWG Roster and Subcommittees

Active Committees

Executive Sponsors

Ellen Pollack, Chief Information Officer, UCLA Health Sciences; Chief Nursing Informatics Officer
Jim Davis, Vice Provost IT and Chief Academic Technology Officer
Lucy Avetisyan, Associate Vice Chancellor and CIO
Roger Wakimoto, Vice Chancellor for Research and Creative Activities
Virginia Steel, Norman and Armena Powell University Librarian

Steering Committee

Lisa M. Snyder, Director, Campus Research Initiatives/Advanced Research Computing Research Technology Group, Office of Advanced Research Computing (co-chair)
Todd Grappone, Associate University Librarian for Digital Initiatives and Information Technology, the UCLA Library (co-chair)

Albert Duntugan, Chief Data Officer, UCLA Health Sciences
Alex Bui, Professor, Radiological Sciences, Bioengineering, and Bioinformatics; and Director of the UCLA Medical & Imaging Informatics (MII) group; and chair of the Academic Senate Committee on Data, Information Technology and Privacy
Bill Holloway, Chief Academic Information Officer, UCLA Health Sciences
Deanna Needell, Professor, Mathematics, and Executive Director of the Institute for Digital Research and Education
George Morris, Research Infrastructure Consultant to the Office of Advanced Research Computing
Howard Miller, CIO, Anderson School of Management
Jackson Jeng, Senior Director, Office for Research Information Systems, Office of Research Administration
Jacob Foster, Associate Professor, Sociology, and co-lead of the DataX Initiative
Marcia Smith, Associate Vice Chancellor UCLA Research Administration
Mary Okino, Deputy Director, Office of the Vice Chancellor for Research and Creative Activities
Michael Van Norman, Chief Technology Officer, IT Services
Nico Wibowo, IT Director, SEASnet, Samueli School of Engineering
Sarah Klein, Pre-Award Coordinator, Semel Institute
Tim Dennis, Director, Data Science Center, the UCLA Library

Technical Infrastructure Subcommittee

Michael Van Norman (chair)

Brian Pape, Manager, Hoffman2 Systems Research Technology Group, Office of Advanced Research Computing
Edson Smith, Data Management Strategist, UCLA Library
George Morris
Lisa M. Snyder
Mike Shane, Senior Director, Platform Services, UCLA Health Sciences
Nico Wibowo, Director, SEASnet, Samueli School of Engineering
Tito Deveyra, Director, Information Technology, Theater, Film and Television (retired)
TV Singh, Chief Computational Scientist, Office of Advanced Research Computing

Data Services Subcommittee

Tim Dennis (chair)

Clifford Kravit, Program Manager, Research Enablement, Health Sciences
Eleanor Koehl, Research Facilitation Coordinator, Office of Advanced Research Computing
Ibraheem Ali, User Engagement Division, UCLA Library
Keith Chen, Professor, Anderson School of Management)
Rose Rocchio, Director, Mobile and Web Strategy, Office of Advanced Research Computing
Todd Grappone
Yoh Kawano, GIS and Visualization Manager, Office of Advanced Research Computing

Cloud Vendors and Services Subcommittee

Bill Holloway (co-chair)
George Morris (co-chair)

Andrew Browning, Manager, Research Web and Data Platforms, Office of Advanced Research Computing
Clifford Kravit
Hector Felix, Director, Enterprise Technology, IT Services
Lisa M. Snyder
Mike Shane
Mike Van Norman
Shao-Ching Huang, Manager, Computational Science Research Technology Group, Office of Advanced Research Computing
Shehzad Sheikh, Director, Enterprise Data and Analytics Platform, UCLA Health Sciences

Research Proposal Process and Data Subcommittee

Sarah Klein, Pre-Award Coordinator, Semel Institute (chair)

Jackson Jeng, Senior Director, Office for Research Information Systems, Office of Research Administration
Lisa M. Snyder
Marcia Smith
Mary Okino
Todd Grappone

Proposed and Pending Subcommittees

The **SECURITY, PRIVACY, POLICY, AND COMPLIANCE SUBCOMMITTEE** will analyze the completeness of existing and the implications of all known and proposed security, privacy, policy, accessibility, and compliance requirements (e.g., HIPAA, PCI, FERPA, CMMC for DoD research) and propose solutions to address those requirements and confirm secure computing solutions and workflows that support data use, risk, manipulation, and publishing. (Currently inactive; will engage at the point of funding and implementation)

Albert Duntugan (chair)

Alex Bui
Ann Pollack, Assistant Vice Chancellor, Research Policy & Compliance
David Shaw, Chief Information Security Officer, IT Services
Edgar Tijerino, Chief Information Security Officer, UCLA Health Sciences
Ernesto Carrasco, Director, Governance, Risk, and Compliance, OCISO
Howard Miller
Kent Wada, Chief Privacy Officer, Office of Advanced Research Computing
Mark C. Krause, AVC and Chief Compliance and Audit Officer
Paula Kim, Managing Campus Counsel
Todd Grappone

The **FACULTY INCENTIVES SUBCOMMITTEE** is intended to work with the Academic Senate, the Academic Personnel office, and faculty representatives to determine and implement incentives for best practices related to research data and open data. (To be convened.)

The **PUBLIC ACCESS TO RESEARCH DATA SUBCOMMITTEE** would be convened to ensure that UCLA data are made openly available through the most appropriate means possible (e.g., data publishing, data sharing, discipline-specific repositories). (To be convened.)

The **RESEARCH AND TEACHING INTERFACE SUBCOMMITTEE** would ensure the coordination and interoperability of research and research data with Canvas, UCLA's primary learning management system, and use of UCLA data in labs and other training programs. (To be convened.)

The **DATA DIVERSITY AND ETHICAL USE SUBCOMMITTEE** would review algorithmic data use to ensure quality and fairness, discussions of bias, ethics, and provenance. (To be convened.)

Appendix 3: Researcher Use Cases

The use cases below were real-world faculty situations presented to various members of the RDWG during the development of this roadmap. These situations informed the proposed research data services and infrastructure.

- A professor new to UCLA that used to be at the University of Michigan has found the infrastructure at UCLA pales in comparison.

- An established and award-winning researcher has a career's worth of data and is committed to sharing, but needs infrastructure and support to make that happen.

- The faculty team-teaching a planned DataX cluster course on 'Data, Justice, and Society' are going to be teaching data science fundamentals and needs infrastructure that will enable collaboration and data sharing for their course.

- An artist renowned for ground-breaking digital work is nearing retirement and needs assistance finding an archival home for this creative output.

- A researcher with appointments both on campus and in the health sciences is assembling a multi-disciplinary team that will need project workspace and help with data collection and manipulation.

- Faculty leading investigations into NLP manipulations on a repository of video and audio histories has students in need data science fundamentals to successfully complete the analysis.

- Two academic units have joined forces on a new major in Data Theory and need infrastructure that supports their teaching goals (see https://datatheory.ucla.edu/).

- A researcher is in the process of purchasing 50 TB of sensitive data and needs an on-campus storage solution that is compute-adjacent.

Appendix 4: Peer Institution Review

Subject Experts from the Data Services Subcommittee:
Rose Rocchio, Yoh Kawano, Ibraheem Ali, Tim Dennis (chair), Clifford Kravit, Eleanor Koehl

Research data services support researchers throughout the entire data life cycle of activities: research planning and developing data management plans; data acquisition, storage, cleaning, and usage; analysis; publication; curation for reuse, data publication, and preservation. At any point in the life cycle, research data services may include consulting, training, and documentation. Research data services also support connections to high-performance computing when increased computational power is needed due to the scale of the data or demands of the analytic method.

Benchmarking Criteria

Research data services program	Data viz support
Cohesive RDS website & service catalog	Data analysis/stats support
Data sharing, publishing, and curation	HPC Support
Reproducible research & data integrity support	Storage
Education and community	Backup services
Data acquisition support	Compliance services
	Data licensing/IP support

Summary of Findings

Tier	Description	Institutions
1	Robust across all benchmarking criteria	UMich, Harvard, Cornell
2	Strong in most benchmarking criteria, very strong in some areas and weaker in others	Chicago, Northwestern, Stanford, UC Berkeley, UPenn, MIT, UVA
3	Mixed assessment	UCLA, UCSD, OSU
4	Weak assessment in all or most areas	UC Irvine

Draft Recommendations
Requirements in this area must be driven at all stages through engagement with campus researchers. Coordination, collaboration, and leadership between OARC, the Library, IDRE, and others (e.g., ORA, DataX) are critical. We assume that services will not be provided by a single campus entity but should be coordinated for ease of finding and use.

Recommendations: To move from Tier 3 to Tier 2, we must:

- Hire 1 FTE (Research Data Management Coordinator (RDO)) to collect and describe research data infrastructure and services
- Develop a plan to foster coordination with all campus partners to provide data services (e.g., Library, ORCA, ORA, ITS, UHIT, QC Bio, DataX).
- Implement the research data life cycle map and underlying service network and publish it on a high profile campus-level domain
- Develop a workflow for referrals between data services providers
- Establish a seminar series/committee to coordinate services across units
- Hire or allocate 1 FTE to support data publishing and curation
- Support data acquisition and licensing by including funds for purchasing data and support for licensing.
- Conduct a gap analysis to move from Tier 2 to Tier 1

Recommendations: To move from Tier 2 to Tier 1, in addition to all of the previous recommendations, we must:

- Extend the research data life cycle map, filling gaps identified in Tier 2 in services and information about various aspects of research data management, support services, funding, and compliance information.
- Develop an interactive resource with guidance for appropriate repository selections (local, system, society, etc.), deposit guidelines (IP, copyright, etc.), and deposit assistance.
- Develop a complete and usable service catalog documenting services across campus support units
- Document data storage and backup services on a searchable and actionable platform with clear paths for provisioning and payment information where relevant.
- Develop a site with clear guidance to make informed security and compliance decisions about IT services to use with research data.
- Expand service and support offerings to include robust options/storage tiers for research data on-premises and cloud resources as they become available.
- Present consolidated event information for data-related workshops across campus and offer more workshops specific to data management.

Appendix 5: Curated List of Peer Exemplars

Institutional Exemplars

School	Research Proposal Lifecycle	Central Research Data Website for Access, Policy, Storage and Security Info	OVCR Office of Research Website / Org Chart	Office of Sponsored Research Website / Org Chart	Library Research Data Services (RDS)	High Performance Computing	Data Science Institute	Business Intelligence & Data Analytics	Faculty / Staff/Student Onboarding
UCSD	Lifecycle Figure	Central Data	Office of Research Affairs Org Chart	Office of Contract and Grant Administration Org Chart	Library RDS Website Library Strategic Plan	Triton, SDSC HPC Systems, SDSC Cloud Compute	Halicioglu Data Science Institute (Includes proposal for UG in Data)	Business Intelligence Research BIA Hub (hubs based on data area)	RA Staff: Onboarding Checklists
UCSF	Lifecycle Page	Central Data Common IT Structure	Office of Research *Office of Ethics and Compliance	Office of Sponsored Research Org Chart	Library Data Science Website	Computing Environments Overview	Bakar Computational Health Sciences Institute	Business Intelligence Website	Data Training TBD
U Washington	Lifecycle Page Detailed Chart	Data Website Facilities & Resources				Research Computing Overview Computing Strategy	eScience Institute		
U Michigan	Lifecycle Figure	Common IT Structure Single Software Directory	Office of the Vice President for Research	Office of Research and Sponsored Programs		Advanced Research Computing	Michigan Institute for Data Science		Faculty Research Onboarding Roadmap

Research Data Support Exemplars and/or Infrastructure Honorable Mentions

TO BE COMPLETED: Add more links in this section.

Offices of Research Data Rockstar equivalents

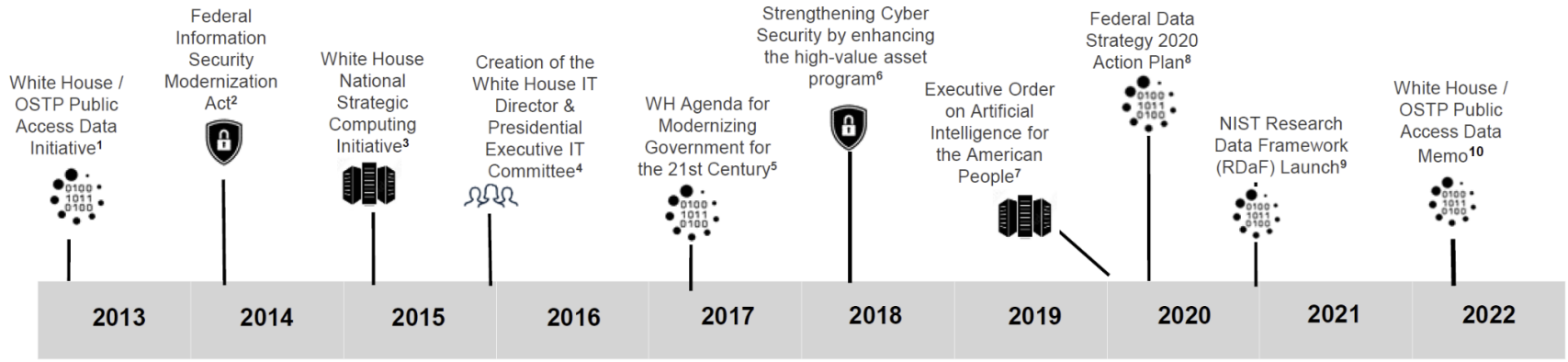
Research storage solutions and technology platforms
UW-Madison: ResearchDrive is a university-wide file storage solution that provides secure, shareable storage space on the campus network for faculty PIs, permanent PIs, and their research group members.

Research Data Services
Harvard: Research Data Management @ Harvard is the landing page for information related to research data services and policies.

Research Infrastructure
Duke: MyRESEARCHpath is maintained in partnership by the Duke Office of Research Initiatives and participating research support offices across the institution.

See also the research data services and infrastructure at Penn State, Virginia, Perdue, and Cornell.

Appendix 6: History of Open-Science and Open-Research Policies



Federal Data-related Policy and Implementation Support

[1] 2013 OSTP Public Access Data Memorandum
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
[2] Federal Information Security Modernization Act (FISMA) of 2104
https://www.cisa.gov/federal-information-security-modernization-act
[3] White House Executive Order for the National Strategic Computing Initiative (NSCI)
https://obamawhitehouse.archives.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative
[4] Presidential Memo Establishing the Director of White House IT and the Executive Committee for Presidential Information Technology
https://obamawhitehouse.archives.gov/the-press-office/2015/03/19/presidential-memorandum-establishing-director-white-house-information-te
[5] President's Management Agenda: Modernizing Government for the 21st Century – Data is the highest asset
https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/03/The-President%E2%80%99s-Management-Agenda.pdf
[6] Plan for Strengthening the Cybersecurity of Federal Agencies by enhancing the High Value Asset Program
https://www.whitehouse.gov/wp-content/uploads/2018/12/M-19-03.pdf

[7] Executive Order on Artificial Intelligence for the American People
https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/
[8] Federal Data Strategy Action Plan
https://web.archive.org/web/20201120105948/https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf
[9] 2020 NIST Data Framework Preliminary Report:
https://www.nist.gov/system/files/documents/2020/12/01/Preliminary%20RaF%20final_12-01-2020.pdf

UCLA Member Organizations Support Material

AAU: Association of American Universities
APLU: Association of Public and Land-Grant Universities
ARL: Association of Research Libraries
CDL: University of California, California Digital Library
COGR: Council on Government Relations
FDP: Federal Demonstration Partnership

2017: AAU-APLU Public Access Working Group Report and Recommendations:
https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Intellectual-Property/Public-Open-Access/AAU-APLU-Public-Access-Working-Group-Report.pdf
2018: APLU/AAU/ARL/CDL (with NSF support) Public Access Workshop Report:
https://www.aplu.org/our-work/2-fostering-research-innovation/public-access/workshop-on-public-access-report-aplu-aau-2019.pdf
2020: APLU/AAU https://www.aplu.org/library/guide-to-accelerate-access-to-public-data/file
2020: APLU/AAU/ARL/CDL Report:
https://www.arl.org/wp-content/uploads/2020/09/2020.09.25-implementing-effective-data-practices.pdf
2020: AAU Report:
https://www.aau.edu/key-issues/research-administration-regulation/renew-investments-vital-scientific-infrastructure

*COGR and FDP non-sponsor specific, data management support guides require member login to access.

Appendix 7: Current UCLA IT spend

Procurement of IT services, in the form of hardware, software and consulting are an integral part of research success. IT spend transaction volume by amount and value are spread across UCLA, with unclear consistency in quality, efficacy and efficiency throughout the acquisition lifecycle. With a variety of options available in hardware, software and consulting, the process of defining and acquiring IT services involved significant investment of time and effort across UCLA. The accounting system does not permit straightforward evaluation of research-related IT spend within the UCLA ecosystem (with the exception of OARC, identified as OIT in the accompanying graphs), so the following discusses total IT spend and estimates as a percentage what might constitute research-related IT spend.

In looking at the non-compensation IT spend there are over 40,000 IT transactions occurring each year, representing 120,000 hours of annual effort (calculated at three hours per transaction) to manage acquisition of IT services, on top of the actual direct purchase costs, amounting to a 6% effort tax. Or stated on an FTE basis, about 54 FTEs needed to support the 40,000 transactions. How much is research related? It's unclear from the data, but one can model differing percentages to understand the overall impact of current procurement strategy. If 30% of the spend was research related, providing shared services could offset this effort burden by nearly an equal amount.

Total

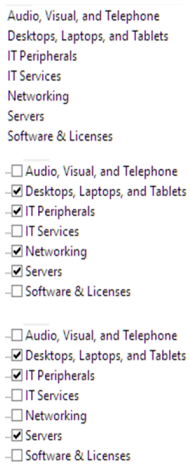
L1 Category	(All)			
Category Level 4	(All)			
Row Labels	Sum of Spend in USD	Average of Spend	Count of Spend in USD3	Sum of Spend in USD4
ITS	$22,101,389.94	$11,523.14	1,918	20.87%
Local	$75,660,435.72	$2,007.17	37,695	71.44%
OIT	$8,150,851.38	$11,352.16	718	7.70%
Grand Total	$105,912,677.04	$2,626.09	40,331	100.00%

Exclude:
-IT services
-Software

L1 Category	(Multiple Items)			
Category Level 4	(All)			
Row Labels	Sum of Spend in USD	Average of Spend	Count of Spend in USD3	Sum of Spend in USD4
ITS	$8,209,561.47	$6,609.95	1,242	19.96%
Local	$32,309,359.42	$1,450.80	22,270	78.57%
OIT	$604,857.99	$4,480.43	135	1.47%
Grand Total	$41,123,778.88	$1,739.07	23,647	100.00%

Exclude:
-IT services
-Software
-networking

L1 Category	(Multiple Items)			
Category Level 4	(All)			
Row Labels	Sum of Spend in USD	Average of Spend	Count of Spend in USD3	Sum of Spend in USD4
ITS	$1,657,535.12	$2,388.38	694	7.61%
Local	$19,522,077.34	$974.54	20,032	89.67%
OIT	$591,244.55	$4,513.32	131	2.72%
Grand Total	$21,770,857.01	$1,043.82	20,857	100.00%



Local and/or distributed IT spend creates technology and data silos. In addition, the overall portfolio of technologies used across the UCLA ecosystem are not interoperable, therefore, as individuals work across differing schools and wish to leverage different services, data and code must be moved or duplicated across systems. The proliferation of software and technology licensing and differing software and hardware versions add additional complexity to conducting research. Shared storage, such as envisioned by the Bruin Data Hub concept would enable the

local IT spend to be directed toward shared services. Distributed IT partners could then focus on consumption of the shared services and ensuring end user support is satisfactory.

Department Name	(All)
Category Level 4	(All)
Line Desc - PO	(All)

IT Groups

Row Labels	Sum of Spend in USD	Count of Spend in USD3	Average of Spend in USD3	Sum of Spend in USD2	
ITS	**$22,101,389.94**	**1,918**	**$11,523.14**	**20.87%**	
Audio, Visual, and Telephone	$6,794,156.76	435	$15,618.75	6.41%	Core IT services
Desktops, Laptops, and Tablets	$78,626.32	44	$1,786.96	0.07%	
IT Peripherals	$891,032.28	512	$1,740.30	0.84%	
IT Services	$1,512,475.72	59	$25,635.18	1.43%	
Networking	$6,552,026.35	548	$11,956.25	6.19%	
Servers	$687,876.52	138	$4,984.61	0.65%	
Software & Licenses	$5,585,195.99	182	$30,687.89	5.27%	
Local	**$75,660,435.72**	**37,695**	**$2,007.17**	**71.44%**	
Audio, Visual, and Telephone	$6,726,859.49	7,056	$953.35	6.35%	Local/specialized IT
Desktops, Laptops, and Tablets	$5,471,183.24	1,543	$3,545.81	5.17%	
IT Peripherals	$10,254,019.62	17,731	$578.31	9.68%	
IT Services	$6,709,660.61	918	$7,309.00	6.34%	
Networking	$12,787,282.08	2,238	$5,713.71	12.07%	
Servers	$3,796,874.48	758	$5,009.07	3.58%	
Software & Licenses	$29,914,556.20	7,451	$4,014.84	28.24%	
OIT	**$8,150,851.38**	**718**	**$11,352.16**	**7.70%**	
Audio, Visual, and Telephone	$165,546.91	77	$2,149.96	0.16%	Specialized core services
Desktops, Laptops, and Tablets	$5,070.20	4	$1,267.55	0.00%	
IT Peripherals	$458,279.38	86	$5,328.83	0.43%	
IT Services	$33,900.78	2	$16,950.39	0.03%	
Networking	$13,613.44	4	$3,403.36	0.01%	
Servers	$127,894.97	41	$3,119.39	0.12%	
Software & Licenses	$7,346,545.70	504	$14,576.48	6.94%	
Grand Total	**$105,912,677.04**	**40,331**	**$2,626.09**	**100.00%**	

Line Desc - PO	(All)

	Column Labels													
	Total Spend			Percentage of tot spend			# of Transactions				Average of Spend in USD			
Row Labels	ITS	Local	OIT	ITS	Local	OIT	ITS	Local	OIT	ITS	Local	OIT		
IT Services	**$323,512.53**	**$675,835.94**	**$33,900.78**	**31.31%**	**65.41%**	**3.28%**	**17**	**173**	**2**	**$19,030.15**	**$3,906.57**	**$16,950.39**		
HARDWARE MAINTENANCE OR SUPPORT	**$323,512.53**	**$675,835.94**	**$33,900.78**	**31.31%**	**65.41%**	**3.28%**	**17**	**173**	**2**	**$19,030.15**	**$3,906.57**	**$16,950.39**		
Servers	**$617,510.07**	**$3,540,779.65**	**$127,126.93**	**14.41%**	**82.62%**	**2.97%**	**40**	**678**	**38**	**$15,437.75**	**$5,222.39**	**$3,345.45**		
ENTERPRISE STORAGE	**$121,468.74**	**$2,196,585.60**	**$25,834.28**	**5.18%**	**93.72%**	**1.10%**	**23**	**484**	**11**	**$5,281.25**	**$4,538.40**	**$2,348.57**		
SERVERS	**$496,041.33**	**$1,344,194.05**	**$101,292.65**	**25.55%**	**69.23%**	**5.22%**	**17**	**194**	**27**	**$29,178.90**	**$6,928.84**	**$3,751.58**		
ABF DATA SYSTEMS INC	$441,811.25			100.00%	0.00%	0.00%	5			$88,362.25				
APPLIED DATA SYSTEM		$329,257.33	$22,858.80	0.00%	93.51%	6.49%		10	3		$32,925.73	$7,619.60		
CDW GOVERNMENT LLC		$79,745.12		0.00%	100.00%	0.00%		2			$39,872.56			
CISCO SYSTEMS INC		$115.28		0.00%	100.00%	0.00%		2			$57.64			
COLFAX INTERNATIONAL		$4,798.20	$22,668.23	0.00%	17.47%	82.53%		4	8		$1,199.55	$2,833.53		
CURVATURE LLC		$936.00		0.00%	100.00%	0.00%		2			$468.00			
DASHER TECHNOLOGIES INC	$801.89	$170,060.07	$47,716.82	0.37%	77.80%	21.83%	1	55	13	$801.89	$3,092.00	$3,670.52		
DATALINK NETWORKS INC		$4,214.58		0.00%	100.00%	0.00%		1			$4,214.58			
DELL MARKETING LP		$352,995.88		0.00%	100.00%	0.00%		23			$15,347.65			
EPLUS TECHNOLOGY INC	$8,020.74	$932.68		89.58%	10.42%	0.00%	3	1		$2,673.58	$932.68			
EXXACT CORP		$31,212.24		0.00%	100.00%	0.00%		2			$15,606.12			
KST DATA INC		$53,544.83	$3,294.31	0.00%	94.20%	5.80%		5	1		$10,708.97	$3,294.31		
NEWEGG BUSINESS INC		$741.99		0.00%	100.00%	0.00%		1			$741.99			
NTH GENERATION COMPUTING INC	$1,000.00	$19,299.84		4.93%	95.07%	0.00%	1	12		$1,000.00	$1,608.32			
ORACLE CORP		$56,330.40		0.00%	100.00%	0.00%		31			$1,817.11			
RELIANT TECHNOLOGY LLC	$42,592.00			100.00%	0.00%	0.00%	4			$10,648.00				
SERVERS DIRECT LLC	$514.65	$31,318.13		1.62%	98.38%	0.00%	2	3		$257.33	$10,439.38			
SHI INTERNATIONAL CORP		$2,951.99		0.00%	100.00%	0.00%		2			$1,476.00			
SILICON MECHANICS INC		$40,136.14	$4,754.49	0.00%	89.41%	10.59%		4	2		$10,034.04	$2,377.25		
SIRIUS COMPUTER SOLUTIONS INC	$1,300.80	$4,137.64		23.92%	76.08%	0.00%	1	2		$1,300.80	$2,068.82			
SUPERMICRO COMPUTER INC		$83.47		0.00%	100.00%	0.00%		1			$83.47			
THINKMATE		$147,604.58		0.00%	100.00%	0.00%		14			$10,543.18			
WIREDZONE COM INC		$1,280.75		0.00%	100.00%	0.00%		5			$256.15			
YANG MING INTERNATIONAL CORP		$12,496.91		0.00%	100.00%	0.00%		12			$1,041.41			
Software & Licenses	**$67,767.27**	**$117,896.07**	**$639.48**	**36.37%**	**63.28%**	**0.34%**	**4**	**161**	**3**	**$16,941.82**	**$732.27**	**$213.16**		
BACKUP OR ARCHIVAL SOFTWARE	**$67,767.27**	**$117,896.07**	**$639.48**	**36.37%**	**63.28%**	**0.34%**	**4**	**161**	**3**	**$16,941.82**	**$732.27**	**$213.16**		

Appendix 8: Prioritization Schema for Proposed Infrastructure

Discussions within the RDWG steering committee about needed resources and services looking to the following list for guidance on prioritization. Similarly, feedback gathered from campus leadership and faculty will be analyzed against this prioritization schema for possible inclusion in the final document and budget that is presented to the campus for funding and implementation.

1. The item addresses funder or institutional mandates that demand immediate attention (e.g., research data management and sharing, security).
2. Other items depend on prior implementation of said service/resource (e.g., dependencies associated with the storage array).
3. There is a glaring gap in our current research support portfolio (e.g., lack of data social services).
4. Implementation will benefit the vast majority of researchers.
5. It is a fiscal imperative (i.e., implementation will save the campus money).
6. Demonstrated growth of researcher demand dictates immediate attention (e.g., a program that supports Jupyter notebooks).
7. There is a current and pressing deadline for action (e.g., deadline to move data from Google).
8. There is an opportunity to develop new areas of expertise or be on the cutting edge of a nascent technology.
9. Implementation will increase reputational capital and/or enable retention and recruitment.
10. Streamlining and/or coordination is required of disparate and duplicative efforts, and/or options to relieve the campus community of administrative burdens.

Appendix 9: Bruin Data Hub Description

The Bruin Data Hub is a service offering that is a mountable storage location for campuswide research data services on Macintosh, Windows and Linux workstations or servers. Overall, the Bruin Data Hub will consist of enterprise-class storage infrastructure, configuration, and support along with additional features and functions to enhance and protect the use and sharing of research data including, but not limited to, a research data commons, a campus-wide license server, and high-speed data transfer nodes. The platform will accommodate the myriad of ways researchers may want to structure their folder hierarchies and the Data Rockstar team will provide consultation on how best to organize data and information.

This network-attached storage will be available to researchers on campus and remote locations where a VPN connection is possible. All researchers will be provided a home directory at no cost along with a starter allocation of shareable lab directory space (also at no cost) for faculty members. This allocation can be expanded at a cost to be determined.[49] The Bruin Data Hub will be configured such that researchers on their own can access and manage or via departmental IT groups if so desired. Use cases for the Bruin Data Hub include the ability to store and protect data acquired from instrumentation and via collaborations internal and external to UCLA, while leveraging group security policies to enable the possibility for a group to manage research data and information for the lab. Volumes of the Bruin Data Hub can be configured for file sharing up to P3 data protection level with researchers anywhere.

To promote such capabilities, access to the Bruin Data Hub will require a valid UCLA ID and password managed via institutional IAM directory services. To simplify access, we expect users to be able to submit a request to provision their Bruin Data Hub lab space via email or web form, along with a list of lab members and UCLA collaborators they wish to grant access to their space. Changes to collaborators or other UCLA staff can be added at any time by contacting the IT helpdesk via email or webform.

As part of Bruin Data Hub, one may request IT to activate the “web sharing” feature of their lab shares. Upon such request approval (to ensure appropriate data security and management policies), IT will create a folder in the Bruin Data Hub space and provide a URL to access data/information placed into the folder. Anything put in the “public_html” folder will be world-accessible via any web-browser and researchers are able to build their own web applications as well. A sample file that is the “landing page” of a Bruin Data Hub web-sharing page will be provided to guide researchers on using the web-publishing features. The web sharing feature will also have an SSO option should researchers want to layer in additional security and access.

[49] The exact amount of storage provided at no charge and subsidized by campus will be determined at a later time and in consultation with faculty.

Appendix 10: Cloud Program for Research

Subject Experts from the Cloud Services Subcommittee:
Bill Holloway, Clifford Kravit, George Morris, Shao-Ching Huang, Andrew Browning

The charge letter of April 2021 to form a research data working group provides a mandate for action needed to meet strategic and business needs related to research data. Cloud-based services are a critical element of this mandate and are addressed through the cloud vendor/services subcommittee with two key deliverables proposed:

> Deliverable #1: A recommendations document that articulates on-premises and cloud options available to researchers for their storage and compute requirements, including consideration of offline storage (see subcommittee #1)
>
> Deliverable #2: An implementation plan with associated costs and resources required to develop the cloud program described in Deliverable #1 including, but not limited to, account creation, administration, documentation and training materials for researchers, a cost analysis for researchers that allows them to understand the true (and long term) costs of their potential cloud use, and benefits accrued as a result of the UC agreements with the major cloud vendors.
> Subcommittee

Analysis

As administrative data storage and processes are shifting to the cloud, it is important to remember that research needs are complex and quite distinct. In contrast to routine cloud services such as website hosting and virtual machines, research computing typically requires unique configurations fine-tuned to address specific research questions and significant administrative effort that are difficult for research faculty to justify. Campus-based support is required to help faculty that are storing and computing large active datasets and using high-performance CPU or GPU compute services in order to manipulate, analyze, and generate secondary data and analysis, in the course of research-focused knowledge production in a cost effective manner.

Furthermore, research use of cloud services is extremely cost sensitive, reinforcing the need for a cloud strategy that incorporates access to both on-premises and cloud-based infrastructure. Even with comparable subsidies, on-premises infrastructure is more cost effective at scale at providing faculty access to storage and computation with the option to buy additional resources at fixed rates. This ensures controlled and predictable costs and allows research to move forward in the absence of extramural funding. To complement on-premises offerings, we call for the development of complementary cloud services that leverage the strengths of cloud for security compliance, rapid scale-up of computations, investigations into new computing architectures, and low-cost deep storage.

‘As Is’ Cloud Spend

Critical data and information on the ‘as is’ state of cloud vendor/services is vital to inform any interim or ‘to be’ states as required by the deliverables, and initial subcommittee efforts were focused on understanding UCLA’s current use of the cloud. The ‘as is’ picture should not only reflect financial metrics but also usage metrics across the UCLA community. According to cloud-spend data provided by Mike Kilpatrick, Technology Program Director for UC/UC Health, ‘The Big Three’ cloud providers (i.e., AWS, GCP, Azure) have enjoyed significant growth in system-wide spend in the past five years. The following table illustrates this growth and projects UC spending on cloud at a conservative 20% annual increase for the next five years.

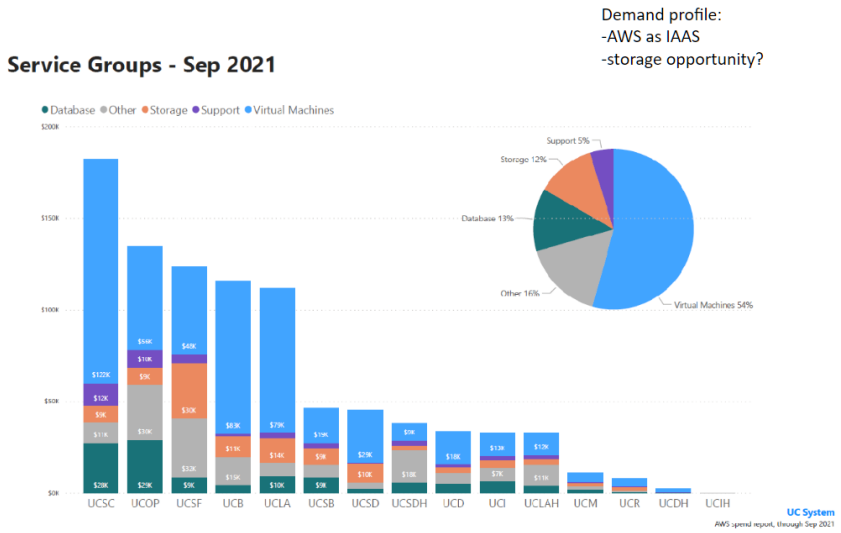
Cloud Growth with 20% forecast

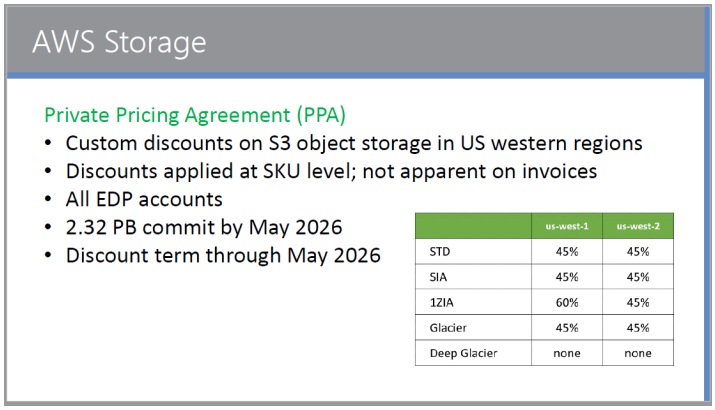
	AWS	YOY Growth	Azure	YOY Growth	GCP	YOY Growth	UC Total	YOY Growth
FY17	$ 1,517,844.64						$ 1,517,844.64	
FY18	$ 6,052,708.41	299%					$ 6,052,708.41	299%
FY19	$ 7,692,842.18	27%	$ 1,299,945.22		$ 17,560.51		$ 9,010,347.91	49%
FY20	$ 9,720,080.03	26%	$ 2,497,352.98	92%	$ 152,357.24	768%	$ 12,369,790.25	37%
FY21	$ 12,648,560.73	30%	$ 5,482,964.59	120%	$ 771,349.82	406%	$ 18,902,875.14	53%
*FY22	$ 15,552,617.09	23%	$ 7,020,823.30	28%	$ 1,113,989.16	44%	$ 23,687,429.55	25%
							$ 28,424,915.46	20%
							$ 34,109,898.55	20%
							$ 40,931,878.26	20%
							$ 49,118,253.91	20%
							$ 58,941,904.70	20%

* Projected through 3 quarters

AWS remains the primary cloud vendor used within the UC system, but growth in all The Big Three is expected to continue across all the cloud service groups. For AWS, virtual machines account for 54% of UC usage (see graph below). Storage, however, remains a minor component of cloud spend (12%). Considering the UC system’s negotiated incentives and attractive rates with AWS (see graph on the following page), integrating S3 storage into the Bruin Data Hub hybrid strategy makes significant economic sense.

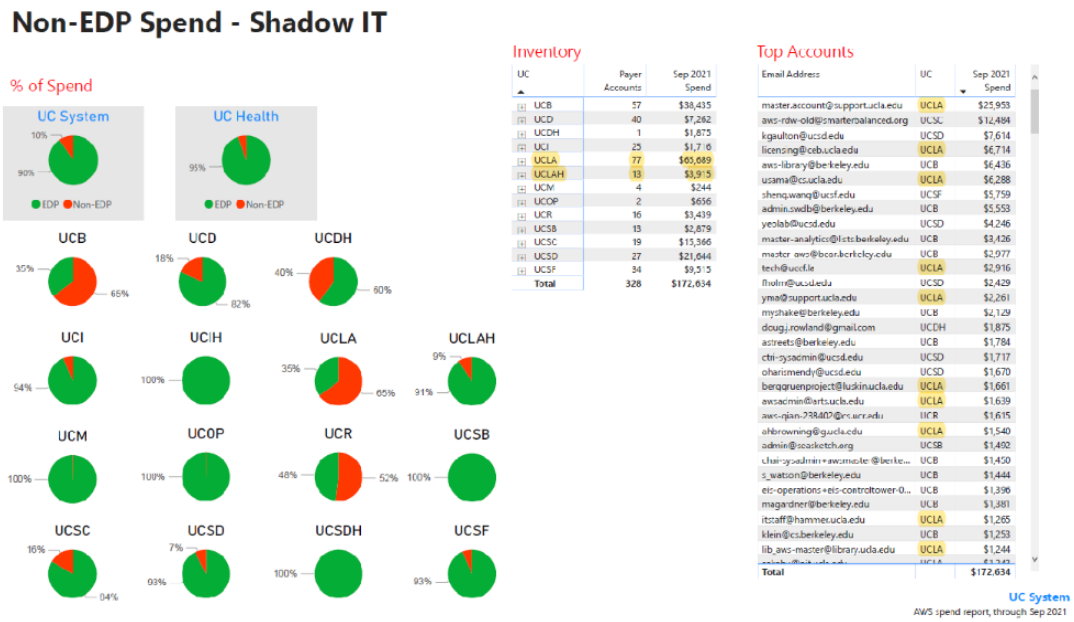
Demand profile:
-AWS as IAAS
-storage opportunity?





EDP versus non-EDP spend

To enjoy the economic benefits of the UC-negotiated agreements, cloud accounts need to fall under our Enterprise Discount Plan (EDP). Leveraging the EDP requires coordination through IT Services, regardless if the account is intended for administrative, departmental, divisional, or research use. UCLA, along with UCB have a high proportion of non-EDP spend, representing missed opportunities to secure the discounts negotiated by UCOP. In the graph below, non-EDP AWS is 65% on the campus vs. 9% at UCLAH.



Recommendations

In the context of research data, there are huge opportunities to incorporate cloud services into research cyberinfrastructure (compute/storage). Current reporting mechanisms, however, make it extremely difficult to distinguish if cloud spend across The Big Three is for research purposes

or how well cloud-based services are configured for security and privacy. In the Health Sciences, DGIT has built AWS ‘templates’ for cloud consumption that have reviewed and approved security features and configurations, but are not widely used yet. The UCLA Health Office of Health Informatics and Analytics (OHIA) has focused on Azure services for clinical data analytics needs across UCLA Health, Faculty Practice Group, and David Geffen School of Medicine.

Therefore, the subcommittee recommends creating managed and coordinated (facilitated) access to and use of public cloud services offered by the three major cloud vendors (e.g., AWS, GCP, and Azure) to create an impactful user experience that decreases information security risks and extracts the financial benefits negotiated by UCOP. Due to the highly distributed nature of research and IT support across campus, we propose a Center of Excellence, or equivalent term, be created and composed of members from units across the campus and the Health Sciences with interests and expertise in cloud services. This group would be charged to formally define a UCLA cloud strategy, and design and implement a research cloud program that realizes that vision. This starting mission could be focused on a three-year scope with rotating leadership to promote inclusivity across UCLA. Furthermore, an initial topic could be to support the DataX initiatives with cloud-focused services and integration into the larger research data strategy within the proposed ORDR.

Appendix 11: Detailed Cost Breakdown for Proposed Roadmap Implementation

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total		MVP
1. CONSOLIDATE RESPONSIBILITY FOR RESEARCH DATA AND INFRASTRUCTURE									
1. Research Data Rockstar	**$444,000**	**$457,320**	**$471,040**	**$485,171**	**$499,726**	**$514,718**	**$2,871,974**		**$2,871,974**
FTE @ $300k plus 48% benefits	$444,000	$457,320	$471,040	$485,171	$499,726	$514,718			$2,871,974
2. Research Data Rockstar Support Team	**$222,000**	**$450,660**	**$464,180**	**$700,105**	**$721,108**	**$742,742**	**$3,300,795**		**$1,435,987**
FTE @ $150k plus 48% benefits	$222,000	$228,660	$235,520	$242,585	$249,863	$257,359			$1,435,987
FTE @ $150k plus 48% benefits		$222,000	$228,660	$235,520	$242,585	$249,863			
FTE @ $150k plus 48% benefits				$222,000	$228,660	$235,520			
3. Data-sharing Compliance Lead	**$296,000**	**$304,880**	**$314,026**	**$323,447**	**$333,151**	**$343,145**	**$1,914,649**		**$1,914,649**
FTE @ $200k plus 48% benefits	$296,000	$304,880	$314,026	$323,447	$333,151	$343,145			$1,914,649
4. Space Required	**$123,900**	**$4,017**	**$4,138**	**$4,262**	**$4,389**	**$4,521**	**$145,227**		**$145,227**
Tenant improvement	$75,000								$75,000
Furniture and computers	$45,000								$45,000
Space use calculation @ $3.25 * 1200 sq ft	$3,900	$4,017	$4,138	$4,262	$4,389	$4,521			$25,227
5. Travel and Conference Attendance	**$9,000**	**$9,000**	**$9,000**	**$9,000**	**$9,000**	**$9,000**	**$54,000**		**$54,000**
Six work trips/conferences per year @ $.5k each	$9,000	$9,000	$9,000	$9,000	$9,000	$9,000			$54,000
6. Website development	**$243,400**						**$243,400**		**$243,400**
Website design and construction	$125,000								$125,000
FTE Contract for web content @ $90k plus 48? benefits	$118,400								$118,400
Work to enable business analytics from funded awards	**$244,200**	**$251,526**	**$141,312**				**$637,038**		**$411,708**
FTE Offset ($150k for .5 plus benefits)	$111,000	$114,330							
FTE Offset ($120k for .75 plus benefits)	$133,200	$137,196	$141,312						$411,708
SUBTOTAL (YEARS ONE THROUGH SIX)	**$1,582,500**	**$1,477,403**	**$1,403,695**	**$1,521,985**	**$1,567,374**	**$1,614,126**	**$9,167,083**		**$7,076,945**

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total		MVP
2.1. DEVELOP A STORAGE SOLUTION/TECHNOLOGY PLATFORM (AKA BRUIN DATA HUB)									
1. Storage Solution/Technology Platform (15 PB)	**$6,182,000**	**$944,210**	**$839,026**	**$848,447**	**$858,151**	**$6,868,145**	**$16,539,979**		**$16,539,979**
Hardware purchase plus incidentals	$5,250,000					$6,000,000			$11,250,000
Maintenance	$525,000	$525,000	$525,000	$525,000	$525,000	$525,000			$3,150,000
FTE Offset ($150k for .5 plus benefits)	$111,000	$114,330							$225,330
FTE @ $200k plus 48% benefits	$296,000	$304,880	$314,026	$323,447	$333,151	$343,145			$1,914,649
2. Associated Research Data Services									
a. Data Transfer Nodes	**$15,000**	**$1,500**	**$1,500**	**$1,500**	**$1,500**	**$1,500**	**$22,500**		**$22,500**
High-end server	$15,000								$15,000
10% Maintenance		$1,500	$1,500	$1,500	$1,500	$1,500			$7,500
b. Relocation of Data from Google (10 PB)	**$355,200**	**$355,200**					**$710,400**		**$0**
FTE @ $120k plus 48% benefits	$177,600	$177,600							
FTE @ $120k plus 48% benefits	$177,600	$177,600							
c. License Server	**$88,800**	**$91,464**	**$94,208**	**$97,034**	**$99,945**	**$102,944**	**$574,395**		**$574,395**
Partial FTE @120k plus 48% benefits	$88,800	$91,464	$94,208	$97,034	$99,945	$102,944			$574,395
d. Research Mosh Pit (Data Commons)	**$557,200**	**$323,416**	**$340,818**	**$359,513**	**$379,615**	**$401,253**	**$2,361,816**		**$2,361,816**
Initial front end configuration	$250,000								$250,000
Licenses/software	$100,000	$110,000	$121,000	$133,100	$146,410	$161,051			$771,561
FTE @ $140k plus 48% benefits	$207,200	$213,416	$219,818	$226,413	$233,205	$240,202			$1,340,255
e. Database/Dataset Tools and Service	**$133,200**	**$137,196**	**$188,416**	**$194,068**	**$199,891**	**$205,887**	**$1,058,658**		**$1,058,658**
Partial FTE @120k plus 48% benefits	$133,200	$137,196	$188,416	$194,068	$199,891	$205,887			$1,058,658
f. CoLab/Jupyter Notebooks	**$531,400**	**$440,032**	**$454,157**	**$573,798**	**$483,980**	**$499,729**	**$2,983,096**		**$1,432,842**
Educational use in cloud	$12,000	$13,200	$14,520	$15,972	$17,569	$19,326			$92,587
Research use on-premises - GPU	$45,000			$45,000					
Research use on-premises - six CPU nodes	$60,000			$60,000					
FTE @ $140k plus 48% benefits	$207,200	$213,416	$219,818	$226,413	$233,205	$240,202			$1,340,255
FTE @ $140k plus 48% benefits	$207,200	$213,416	$219,818	$226,413	$233,205	$240,202			

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total		MVP
g. Web Hosting/Sharing Tools	**$133,200**	**$137,196**	**$188,416**	**$194,068**	**$199,891**	**$205,887**	**$1,058,658**		**$1,058,658**
Partial FTE @120k plus 48% benefits	$133,200	$137,196	$188,416	$194,068	$199,891	$205,887			$1,058,658
SUBTOTAL (YEARS ONE THROUGH SIX)	**$7,996,000**	**$2,430,214**	**$2,106,542**	**$2,268,429**	**$2,222,972**	**$8,285,345**	**$25,309,503**		**$23,048,848**
2.2. EXPAND EXISTING RESEARCH DATA SERVICES									
1. Digital/Media Asset Management System	**$532,800**	**$548,784**	**$742,848**	**$942,733**	**$971,015**	**$1,000,145**	**$4,738,325**		**$3,040,591**
FTE (Data Asset Management)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			$1,148,790
FTE (Data Preservation)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			
FTE (Data Publishing and Curation)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			$1,148,790
FTE (Data Publishing and Curation)			$177,600	$182,928	$188,416	$194,068			$743,012
FTE (Data Services Admin)				$177,600	$182,928	$188,416			
2. Expansion of Consulting Services and Expertise	**$177,600**	**$182,928**	**$177,600**	**$177,600**	**$177,600**	**$177,600**	**$1,070,928**		**$0**
FTE Contract (Data Lifecycle)	$177,600	$182,928							
FTE Contract (Expertise as needed)			$177,600	$177,600	$177,600	$177,600			
3. Services Outreach and Education Program	**$355,200**	**$365,856**	**$554,432**	**$926,265**	**$954,053**	**$982,674**	**$4,138,479**		**$2,440,746**
FTE (Comp Data Science Research)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			$1,148,790
FTE (Comp Data Science Research)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			
FTE (Comp Data Science Research)			$177,600	$182,928	$188,416	$194,068			$743,012
FTE (Comp Data Science Research)				$177,600	$182,928	$188,416			
FTE (Comp Data Science Research)				$177,600	$182,928	$188,416			$548,944
4. Primary Source Data	**$177,600**	**$182,928**	**$188,416**	**$194,068**	**$199,890**	**$205,887**	**$1,148,790**		**$0**
FTE (Data Preservation)	$177,600	$182,928	$188,416	$194,068	$199,890	$205,887			
SUBTOTAL (YEARS ONE THROUGH SIX)	**$1,243,200**	**$1,280,496**	**$1,663,295**	**$2,240,666**	**$2,302,558**	**$2,366,307**	**$11,096,521**		**$5,481,337**
2.3. EXPAND EXISTING RESEARCH INFRASTRUCTURE									
1. Expansion of On-Premises Compute and Storage	**$920,000**	**$360,000**	**$420,000**	**$860,000**	**$420,000**	**$360,000**	**$3,340,000**		**$3,340,000**

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total		MVP
a. High-Memory Compute	$500,000			$500,000					$1,000,000
b. GPU Nodes	$420,000		$420,000		$420,000				$1,260,000
c. Compute Nodes For General Campus Users		$360,000		$360,000		$360,000			$1,080,000
2. Expand Cloud for Research Offerings	**$749,000**	**$765,650**	**$782,800**	**$800,463**	**$818,657**	**$837,397**	**$4,753,967**		**$3,589,967**
a. Campus Cloud Support for Research									
FTE Offset ($150k for 2.5 FTE plus benefits)	$555,000	$571,650	$588,800	$606,463	$624,657	$643,397			$3,589,967
Cloud subsidies	$50,000	$50,000	$50,000	$50,000	$50,000	$50,000			
b. Testbed/Technology Sandbox									
Cloud subsidies	$144,000	$144,000	$144,000	$144,000	$144,000	$144,000			
c. Possible Solution for P3/P4 Data	$0	$0	$0	$0	$0	$0			
3. P3/P4 Compliant HPC/Storage	**$605,200**	**$573,056**	**$590,248**	**$815,155**	**$839,610**	**$864,798**	**$4,288,067**		**$3,647,632**
Third-party vendor to build CMMC environment	$250,000								$250,000
FTE @ $240k plus 48% benefits	$355,200	$365,856	$376,832	$388,137	$399,781	$411,774			$2,297,579
FTE @ $140k plus 48% benefits		$207,200	$213,416	$219,818	$226,413	$233,205			$1,100,053
FTE @ $140k plus 48% benefits				$207,200	$213,416	$219,818			
4. Core Facilities	**$350,000**	**$385,000**	**$423,500**	**$465,850**	**$512,435**	**$563,679**	**$2,700,464**		**$0**
Reserve for core facilities upgrades	$350,000	$385,000	$423,500	$465,850	$512,435	$563,679			
5. Research Software Collaboration Tools	**$500,000**	**$550,000**	**$605,000**	**$665,500**	**$732,050**	**$805,255**	**$3,857,805**		**$0**
Reserve for campus licenses (sitewide)	$375,000	$412,500	$453,750	$499,125	$549,038	$603,941			
Reserve for license compliance (as needed)	$125,000	$137,500	$151,250	$166,375	$183,013	$201,314			
6. Research Data Backup	**$475,000**	**$522,500**	**$574,750**	**$632,225**	**$695,448**	**$764,992**	**$3,664,915**		**$3,664,915**
Endpoint: Crashplan @ $50 x 5000	$250,000	$275,000	$302,500	$332,750	$366,025	$402,628			$1,928,903
Server Backup Solutions	$225,000	$247,500	$272,250	$299,475	$329,423	$362,365			$1,736,012
Local back-up on Bruin Data Hub	$0	$0	$0	$0	$0	$0			
7. Streamline Proposal Development Process	**$732,600**	**$917,378**	**$1,152,739**	**$1,041,770**	**$1,073,023**	**$1,105,214**	**$6,022,726**		**$0**
a. Single source for guidelines									
b. Streamline workflow									
c. Staffing for proposal process									

				Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Six-Year Total		MVP
			FTE (DGSOM; $110k plus benefits)	$162,800	$167,684	$172,715	$177,896	$183,233	$188,730			
			FTE (DGSOM; $110k plus benefits)			$162,800	$167,684	$172,715	$177,896			
			FTE (LifeSciences; $110k plus benefits)	$162,800	$167,684	$172,715	$177,896	$183,233	$188,730			
			FTE (Grant Support-Letters&Sciences)	$162,800	$167,684	$172,715	$177,896	$183,233	$188,730			
			FTE (ORA; $110k plus benefits)		$162,800	$167,684	$172,715	$177,896	$183,233			
			FTE (ORA; $110k plus benefits)			$162,800	$167,684	$172,715	$177,896			
		d.	Communications re: DMSP requirements									
			FTE Offset ($150k for .5 plus benefits)	$111,000	$114,330							
			FTE Offset ($120k for .75 plus benefits)	$133,200	$137,196	$141,312						
	SUBTOTAL (YEARS ONE THROUGH SIX)			**$4,331,800**	**$4,073,584**	**$4,549,037**	**$5,280,964**	**$5,091,223**	**$5,301,335**	**$28,627,943**		**$14,242,514**
	TOTALS (YEARS ONE THROUGH SIX)			**$15,153,500**	**$9,261,697**	**$9,722,569**	**$11,312,044**	**$11,184,128**	**$17,567,112**	**$74,201,050**		**$49,849,644**

Appendix 12: Funding Sustainability Options

Decisions on the ideal funding structure to sustainably support research at UCLA will need to be made at the highest levels of the institution. Critical is the need to modernize and renegotiate our funding structure for centralized research infrastructure and data services to address compliance with growing funder mandates. The RDWG acknowledges that no one funding source is going to provide the budget necessary to implement this plan; more likely, some mix of funding sustainability options will be required. Some suggestions:

1. Establish a defined per-year campus commitment from core funding as a strategic investment in research.
2. Allocate to research support a defined per-year slice of current provisional F&A.
3. Campaign to increase F&A rate to include/support research data and infrastructure initiatives.
4. Increase direct cost allocations in extramural proposals and pool those resources to support research infrastructure and research data services.
5. Redirect existing budget allocations and realign activities within each of the identified service owners (i.e., trimming existing services and reallocating those funds and FTE to support research).
6. Redirect research-related budget from local to central (i.e., local budgets will be cut to partially fund research IT).
7. Expand TIF (or similar ‘tax’ to support research infrastructure).
8. Work with Government Relations to connect our research infrastructure and data services needs to the state’s priorities
9. Redirect to research support savings from improved efficiencies around procurement and consolidation of IT spend.
10. Offer baseline resources and services to all faculty, while recharging premium for above-and-beyond use of resources and services.
11. Develop external partnerships to acquire technology at reduced rate or acquire additional services.
12. Partner with spin-off companies to build research infrastructure.
13. Increase the number of proposals submitted to infrastructure-related grant opportunities.
14. Launch a targeted development campaign to support research data services and infrastructure.