UC VCR-CIO 2015 Summit

Anchor Principles and Key Conclusions

Version 12
September 22, 2015

Background

The modern research enterprise continues to evolve dramatically: Science and digital scholarship are becoming data-driven, research now occurs in increasingly collaborative environments, researchers must be both domain and data experts, and data as a language enabling research and scholarship is the new normal.

This new environment is driving change and presents new challenges and opportunities, from ethics to data access to human analytical capacity. Clearly, this evolving environment requires the University of California to consider and plan for its collective future, and a thoughtful research cyberinfrastructure strategy is required to ensure UC addresses these challenges and that every opportunity is leveraged.

The costs of not addressing this collective UC need are significant*.* The grand, complex challenges facing humankind can only be resolved with robust, coordinated, and collaboratively utilized cyberinfrastructures and related services and support.

VCR and CIO Cyberinfrastructure Conference – March 2015 at UCLA

On March 23, 2015, UC's Vice Chancellors of Research (VCRs) and Chief Information Officers (CIOs) sponsored a cyberinfrastructure visioning conference, which was held at UCLA. The goal was to prioritize and recommend a UC cyberinfrastructure plan of action for the next five years. The conference was a day-long event and featured panels discussing emerging digital scholarship and research trends and the associated cyberinfrastructure requirements these opportunities will demand. Over 140 UCR faculty, research support staff, and VCRs and CIOs participated.

The conference featured five panels as follows:

- Physical Sciences, Life Sciences and Engineering
- Libraries, Arts and Architecture, Theater Film and Television
- Management, Law and Public Affairs
- Social Sciences, Humanities, and Education
- Health Sciences

Other panels / presentations were provided by the National Science Foundation, CIOs (current technology initiatives), and faculty discussing "blue sky" possibilities.

Conference Themes

During the conference, several consistent themes emerged across the presentations and panel sessions. Importantly, as UC addresses these themes in the months and years ahead, it is essential these cyberinfrastructure, services, and support offerings be inventoried and be made transparently available to faculty, whether these services and infrastructures are supplied by a campus, the UC system, or cloud providers.

The seven themes below received particular attention from conference participants.

- Cyberinfrastructure "Concierge" Service (digital technology resource guidance)
- Collaboration Tools, Portals, and Services
- Storage Vision and Eco-System
- Data Management, Curation, Metadata / Interoperability
- Data Access – UC and Beyond
- Skills Development, Training, "Boot Camps"
- Polices and Ethical Considerations

Vision Document

In many ways, the 2015 VCR and CIO Cyberinfrastructure Conference was a call to action. Based on conference themes and observations, this cyberinfrastructure vision document has been created to offer prioritized recommendations and a series of action plans for each recommendation. This plan has been reviewed and vetted by UC's VCRs, CIOs, Librarians, and the over twenty UC faculty members who served as conference panelists.

UC's cyberinfrastructure vision provides a roadmap that will enable UC to optimally support the future success of its research enterprise. Clearly, data driven science, digital scholarship, and the associated (and enabling) cyberinfrastructures this vision document discusses are core to the University of California's collective ability to address the grand challenges facing California, the nation, and the entire world.

Defining Cyberinfrastructure

*G*iven the intensity of expectation around "cyber-enabled" research, the following terms are defined to distinguish key aspects of "cyberinfrastructure" and to facilitate discussion around UC's pressing need to take action. Cyberinfrastructure is itself an area of research that is developed and deployed institutionally as broad-area, shared infrastructure. Cyber-*enabled* research is the more expansive researcher-driven disciplinary and trans-disciplinary research that cyberinfrastructure facilitates.

a. **Cyber facilities** – the physical compute, storage, data center and network facilities and the operational standards, software and code that comprise the computational, storage and network system layers of cyberinfrastructure. Facilities include sophisticated routers, servers, fiber, cabling, data centers, power and cooling, etc.

b. **Cyber collaboration infrastructure** – tools, capabilities and processes that are layered on the cyber facilities
    i. *collaboration tools* for multiple research groups to work together with analytics, modeling, simulation and visualization capabilities
    ii. *software-based processes* for data management, data modeling, curation, preservation, and aggregation for accessing, reusing and building broadly used research data assets, as well as protecting and securing them
    iii. *cyber environments* for readily promoting, accessing, using and collaboratively building software applications, i.e., research software stores
    iv. *networked tools and mechanisms* for discovering and accessing expertise, both formally and informally and in directed team-based projects, to spark innovation, discovery and trial
    v. *network-based channels* for conducting team-based R & D securely, tech transfer that manages IP, processes that manage regulated data, etc. not only within higher education, but also with commercial and industry partners, recognizing that data are valuable intellectual property and technology transfer assets

c. **Platforms** – platforms combine *cyber facilities* (what we can think of now as basic needs) and *cyber collaboration infrastructure* (new, enabling tools and processes) to create integrated cyberinfrastructure facilities and services that, in aggregate, offer new functions, often taking into account the full research data life cycle or the end-to-end process of collaboration. An institutional research cyberinfrastructure platform might, for example, integrate network, computation, data, workflow and security facilities and services to facilitate the ability of researchers at different locations and institutions to progressively analyze data sets. Mobility services might be added to facilitate distributed human-centered data input. Different database structures might be integrated

to facilitate different data analysis and integration needs. A HIPAA compliant platform might make it possible to do health sciences research involving patient data. Discipline-specific platforms could be built separately or over general-purpose platforms.

d. **Sociotechnical infrastructure** – this term, in increasing use in higher education, refers to the *technical expertise*, *guidance*, *workflow*, *procedures*, *interfaces* and other human-technology interventions (such as the concierge service described later in this document) that facilitate the use of cybertechnologies by humans in the research environment. The importance of this type of service was stressed at the conference and must be developed in concert with the facilities and infrastructure that accompany it.

Summit Trends

A decade ago at the 2005 UC VCR-CIO Summit, the emphasis was on the *cyber facilities* needed to provide capacity and capability for high-performance computation-based research. The National Science Foundation (NSF) focus was on the national research network infrastructure, computation resource availability through the Teragrid, the build-out and aggregation of campus computational facilities, and the advent of Petascale facilities. The Top 500 competition had just become a metric of cyber research capability and leadership. Today's roles for research data, data management and managed storage were in early discussion.

By the 2011 Summit, the tenor of the discussion had shifted from cyber facilities to a direct focus on the researcher-defined, front-end research capabilities that comprise *cyber collaboration infrastructure*. Cyber facilities were not strongly referenced by frontline researchers, although IT infrastructure providers and the relevant infrastructure programs within funding agencies continued to strongly emphasize them. Data management and analytics were becoming a stronger focus, while the focus on facilities, especially physical facilities, had shifted to a focus on the tools and services that would more directly meet these research data needs. The 2011 Summit can be characterized as the moment when the importance of cyber collaboration infrastructure really took hold, and the questions of what tools were needed and how to invest in them were raised.

The 2015 Summit did not reveal significant differences in researcher perspectives on the importance of data, analytics, modeling and important tools. It did reveal a much more extensive cross-disciplinary research interest, an increased diversity of targeted uses, and an expectation of precision in findings, predictions and insights. All disciplinary areas now depend on data and analytics in some way. The 2015 Summit featured widely cross-disciplinary breakout sessions, and all disciplines noted the importance of infrastructure and expertise to support research data management, preservations and analytics (without using cyberinfrastructure terminology). Facilities such as compute, storage and transit were presumed to be essential but are not always present at the necessary levels. The term "informatics" was used frequently. The expected precision of solutions and team-based

informatics amplifies the dependence on agile and flexible research tools that facilitate shared, team-based research. This in turn generates further need for more a purpose-built integrated, end-to-end collaboration infrastructure, which we refer to here as *platforms*. The institutional role, and the need for platforms that no single researcher or research group can individually provide was underscored, along with the role of people and the importance of sociotechnical infrastructure.

The NSF's long-term vision for cyberinfrastructure stresses that the complexity of research analytics is increasing. Solving the grand-challenge problems of society has become an increasingly important priority but provides IT challenges. There is an unprecedented growth in data, both facilitated by technology and also in response to the improved ability to apply meaningful and timely analysis and action. This growth is expected to continue increasing dramatically for many years. Many of the grand challenges require approaches to "big data" and strategies to deal with data from new technologies (mobile devices or social media). More importantly, meaningful solutions demand interoperable expertise, capabilities and resources. Partnerships are required. As more data become available, interoperability and standards become important, as well as rational access, analytics, and archiving strategies. All universities have similar shared challenges: to reduce costs, create policies, address data management and curation requirements, etc. The successful universities will be those that leverage their unique strengths *and* an appropriately open environment of integrated, federated and/or shared resources, expertise and true partnerships.

These trends are reflected in other agencies and initiatives. The President's Council of Advisors on Science and Technology (PCAST) has recognized the role of digitization in the national economy. In response, the White House has established the Advanced Manufacturing Office in the Department of Commerce and the National Network of Manufacturing Innovation Institutes. To date, three of these institutes are directly related to information technology. The National Institute of Standards and Technology has reoriented many of its programs around "smart" technologies: Smart Buildings, Smart Grid, Smart Health, Smart Transportation, Smart Manufacturing, etc. The Department of Defense has oriented programs around a strong cyberinfrastructure emphasis on accelerated product and parts design, manufacture and management. The Department of Energy has oriented cyberinfrastructure initiatives around energy reduction, renewables, and environment, as well as the science around energy. The National Academy of Engineering has promoted the U.S. Grand Challenge problems that have led universities throughout the country to re-orient their educational programs in direct response.

Many new physical technologies such as 3D printing, materials development, etc., depend on IT and cyberinfrastructure. The concept of the Internet of Things is motivating the extensive connectivity of devices to the Internet, and the Industrial Internet Consortium is encouraging use of networked data. The Federal Communications Commission has supported network neutrality to preserve democratized access to Internet capacity and data. The Office of Science Technology Policy has strongly advocated the open publishing of data. The National Institutes of Health is investing resources to explore the opportunities and identify the challenges

associated with building a large research cohort with complex data elements (clinical, imaging, genomic) as part of the newly announced Precision Medicine Initiative. Germany, the United Kingdom, India, China, and Korea also all have large government-driven cyberinfrastructure initiatives.

An External View of UC

UC campuses are individually recognized as world-class research universities. Each campus supports a wide range of research and each campus claims particular areas of research leadership. When UC's research areas, grants, patents, scholarship recognitions, etc., are considered as a whole, the University is unrivaled as an institution. In general, though, UC research and cyberinfrastructure capabilities are operationally separated by campus, with little inter-campus visibility, access or interaction. Both in research and in cyberinfrastructure, UC is perceived as ten individual campuses, not as a system. Indeed, UC has a history of competing as individual campuses rather than aggregating strengths as a system or cluster of campuses when responding to state and national initiatives and funding opportunities.

Positioning UC Action

UC recognizes that cyberinfrastructure requires research and development in its own right. Research on cyberinfrastructure needs to be aligned and in lockstep with the frontline domain-specific research. Today's platform infrastructure research will become tomorrow's platform tools, used pervasively by our researchers to pursue innovative and next-generation research problems and needs. We need to recognize the transformational nature of cyberinfrastructure technology, the role of cyberinfrastructure research in facilitating frontline research, and the need to create a pipeline from cyberinfrastructure research to application.

The 2015 Summit generated a spectrum of topics worthy of consideration. However, seven of these received particularly strong, cross-disciplinary attention, as measured by how often they surfaced in the disciplinary sessions and summit panel sessions. They can be grouped into seven priority areas for UC action:

- Cyberinfrastructure "concierge" service
- Collaboration tools, portals, and services
- Storage vision and ecosystem
- Data management, curation, metadata / interoperability
- Data access – UC and beyond
- Skills development, training, "boot camps"
- Policies and ethical considerations

We can further organize them into the following four themes: (1) the *need for cyber collaboration tools*, (2) *strong cross-discipline desire for skills and access to expertise*, (3) *data as*

research assets to be managed, curated, and preserved; and (4) bringing it all together into a *platform "ecosystem"* that reflects *associated policy and ethical considerations.*

Cyber collaboration tools

1. *Enabling a broader base of researchers.* Easier-to-use, self-guided and more highly abstracted transformative tools and services that embody informatics expertise will enable a broader base of researchers to conduct novel research without having to develop or invest in the same expertise**.** In addition, new models for research informatics support will support researchers who may be in silos or who lack resources to establish independent infrastructure and support systems. Such models may also realize cost savings. Emerging technologies and access to standardized approaches to data management will be accessible to all faculty, including those in fields where such capabilities have traditionally been underdeveloped. Finally, widely available training for students, research staff and faculty in applying new technologies to research will help develop cyberinfrastructure skills into standard research techniques.

Cross-discipline desire for skills and expertise

2. *Cross-disciplinary collaboration.* Collaboration and partnerships across departments, schools and fields of study will increase our ability to solve complex research problems. Innovative approaches for generating, collecting, and analyzing data to bridge disciplinary languages, dictionaries, and areas of interest will provide vast opportunities for cross-disciplinary researchers to share ideas, data, tools, and algorithms and to approach research and global problems with a shared context.

Data as research assets

3. *Data ownership and big data*. Big data has three attributes: volume (scale), variety (its many forms, e.g., structured/unstructured, text, multimedia), and velocity (dynamism/real-time qualities). The ability to more readily collect, access and analyze data beyond the walls of the institution, and to store and analyze large amounts of disparate data (or big data) generated both locally and distally, will increase opportunities for new kinds of research, analysis and decision-making. Real-time dynamic data and analysis will transform traditional research approaches and methodologies by accelerating the generate-analyze-apply-learn cycle. Systems will use networked, information-based technologies to integrate intelligence in real time across entire enterprises and will use data-driven modeling, simulations and Key Performance Indicators to communicate optimal actions and results in real time. There are significant policy, regulatory, security, privacy and ethical issues to be managed.
4. *Multi-use data.* The line between operational, business and research data is

blurring. Data is quickly becoming dual-purpose or multi-use as organizations integrate potential research data collection more seamlessly into business workflow and operations. Policy and governance will be critical to efficiently and effectively manage data in organizations with potentially multi-purpose data innovative approaches to human subjects protection and compliance issues. Business operations will have to consider how to support business and research simultaneously.

5. *Data visualization.* Of increasing importance for managing large data sets, data visualization involves the graphic display of data too complex for manual processing or assessment; the resultant imagery is typically the end result of an algorithmic process or generated from large-scale data sets. It encompasses a broad range of analytic tools and techniques that include statistical visualization, GIS, and 3D modeling, all of which share the common goal of organizing data into a coherent visual display that can be readily interpreted and understood.

Platform "ecosystem"

6. *A federated but connected and interoperable infrastructure of platforms*. This will be key to helping the campuses enhance capacity and capability individually and across the system. Such infrastructure will extend the tools and capabilities that form the institutional "nervous system" (distributed resources, capabilities, expertise, policy and ethics) through which data can be moved and methodologies accessed. Organized for campus leverage, this federated infrastructure will cultivate individual researcher capability. Mobile information and communication technologies will play a major role. Policy will be an important driver, and initiatives need to reflect the ethical values that the UC wished to project.

<u>Recommended Actions</u>

[Please note that several Action Items contain timelines while others do not. Timelines for all action items will be created once the Steering Committee has reviewed the recommendations / actions items and has prioritized them.]

ACTION 1: Build the policies necessary for a federated approach to shared services.

UC is a wellspring of innovation, new ideas, and creative approaches, which occur in parallel across campuses. There is significant untapped potential, however, to link best practices and to federate services as an *additional* mechanism to services that are shared simply by extending capability provisioned on one campus to others.

A federated research cyberinfrastructure by definition involves distributed resources and capabilities in the form of staff, facilities, services, investments and individualized campus models that have been structured and configured to align to particular researcher-defined areas of emphasis, and to research partnerships, as

well as the campus' mission, culture and location. A federated approach to frontline faculty research at each campus, across our campuses, and across the full diversity of research partnerships will better align cyberinfrastructure capabilities, including research and development on research cyberinfrastructure itself. Federation builds from a starting position that each campus has and needs to build local capacity and capability to its strengths and needs. Federation then addresses the ways that these individual campus capabilities are not only supported but also significantly enhanced through shared visibility and appropriately shared and/or integrated capabilities through cross campus infrastructure, platform and sociotechnical collaborations that create win-win situations. Federation also addresses the ways that individual campus research and capabilities can be formed into a collective strength.

Currently, UC policies are not organized to facilitate federation or collaboration. Specifically, policies, practices, and incentives often encourage the creation rather than the dissolution of silos. While exceptionally difficult, UC should tackle and promote the development of "federated services" that result in "deploying once for the benefit of many campuses." The following actions are essential to develop and promote federated services.

- Establish as an organizing principle a systemwide "research cyberinfrastructure federation" of services, platforms, technical expertise, and accessible, reusable research data. Federated services need to be distinguished from centrally shared services with respect to approach, resources and operations. A federated service is the true value-driven coordination of services drawing upon the strengths and diversity of the distributed approaches. This is very different from centralization, which implies centralized provisioning and then extended access. Federation and centralization are not mutually exclusive, just different. Although a "federation" is challenging to the currently fully decentralized business and financial structures of the UC system, it is highly valuable. Precisely because of the broad nature of individual campus research strengths, UC is well positioned to build and demonstrate the power of federation. UC federated services would allow individual campuses to retain their interests and strengths, and to build on them and draw on crossover strengths where there are multi-campus benefits. Federation should be used to create interoperability opportunities that take advantage of the infrastructure and expertise at each campus for the purposes of accelerating, enhancing and promoting the development of each campus's unique research strengths.

- To make this work, determine the appropriate infrastructure (such as network connectivity), transparency, and incentives necessary to facilitate federated resource sharing between campuses. Federated resources must not be determined solely in a top-down, system-level manner, but must be allowed to emerge from individual or collaborative campus efforts and identified and selected for federation. Bottom-up structures are often more agile, approach

new technologies sooner and address a broader range of disciplinary and cross-disciplinary research activities. Top-down transparency, organization and facilitation can be combined with campus-level development, expertise, and emerging skills to maximize impact.

- Break down policy barriers to collaboration with specific timelines and the following deliverables:

    - *Inventory of services, systems, and support.* Strategies are needed to communicate the existence of shared services and to facilitate inter-campus use of such devices, systems, tools, and services.

    - *Institutional support for sharing services across the UC system.* The barriers to entry for sharing and utilizing common tools across campuses must be eliminated or greatly reduced. These barriers include financial, cultural, incentive, policy, and organizational constraints.

    - *Federated services strategy.* Importantly, not all campuses must utilize a particular service, nor it is necessary for all shared services to be provided by UCOP or a particular campus or center. Rather, UC's strategy should recognize that intercampus collaborations of two or several campuses or research centers might generate significant efficiencies and benefits. (This does not preclude such services being identified as shared service opportunities at a later time.)

    - *Common approach to data access, security, etc.* UC does not have a common (campus, discipline, health sciences) approach to data access, security, availability, etc. UC should develop and support a suite of transparent policies, procedures, and incentives that are easy to understand / utilize and that promote the wide availability of data and resources within UC. Issues that must be addressed include compliance (e.g., HIPAA), security, bio-ethical topics, and clinician / researcher relationships.

    - *Ethical considerations.* As access to data increases, UC must ensure appropriate policies and standards for privacy, confidentiality, data ownership, public / private partnerships, etc., are considered and adopted.

    - *External (non-UC) data.* UC must investigate policies and practices relating to data security, access, privacy, etc., that will facilitate the acquisition of data from organizations, firms, and other groups outside UC.

- Create a **UC Cyberinfrastructure Institute** tasked to define, build, stage and orchestrate federated *and* centralized operations and policy.

    - Federation needs to be viewed as an operation in its own right that facilitates and sustains value-driven federation-oriented policy, infrastructure activities and interoperability collaborations, which together produce measurably increased campus and collective research capability and capacity. In sharp contrast to centralization, federation involves sustaining an evolutionary development lifecycle that will generally consist of (1) identification of a high potential federated capability, (2) an inventory and visible exposure of campus capabilities, e.g., websites and workshops, (3) a detailed review of federated potential, consideration of approaches and funding, policy and capacity needs/barriers, (4) a highly visible pilot orchestrated with a small subset of campuses to champion, demonstrate and shape an approach, (5) resolution of funding, policy, infrastructure or capability barriers, (6) scaling from the successful pilot, moving to operational requirements and scaling to critical mass interest and (7) adjusting and sunsetting a capability when requirements, technologies and value changes.

    - To execute on this development pattern, a working group for each potential federated capability needs to be identified. Each working group must be supported with increasing involvement and project management. This will ensure the demonstration of value and review on the merits of capability, and will avoid the loss of capabilities because of lack of support, resources or commitment at any one step. Federated capabilities that survive the pilot process need to be able move into a managed operational start-up and scale-up mode with identification of appropriate federated value, investment in resources, and resolution of policy barriers. The VCR-CIO Summit identified a first slate of candidate federation capabilities. The descriptions for each of the following recommended actions provide proposed agendas for the associated working groups.

    - A **Federation Governance Board** (FGB) should be established and staffed as the initial federation operating entity. As a start-up itself, the FGB will be responsible for prioritizing federated capabilities, commissioning working groups and supporting and orchestrating the activities of each working group. The FGB must include a funded project management position, since it will need to coordinate and manage resources from the beginning. This need will only grow as the first federation capabilities move into the pilot steps. As capabilities become operational and others enter the development process, the FGB will need to become an operating entity. The FGB should eventually form a UC business entity, a

UC Cyberinfrastructure Institute, responsible for federated operations. The FGB should comprise two VCRs, two CFOs, two CIOs, two librarians and several key faculty members from multiple campuses. The FGB will interact with campuses through existing senate and administrative structures, as well as create events, such as workshops, to define, shape and build operational direction and interest and to build the infrastructure needed to facilitate capability.

ACTION 2: Make research data an institutional asset.

It is important to acknowledging the role of research data as valuable University intellectual property, and to develop and implement a set of guidelines for its management. Further, it is important to develop new – and integrate existing – tools and services based on these guidelines, bringing together local campus data management initiatives and system-level tools where appropriate. The libraries' critical role in building research data into a University research asset emerged strongly in the Summit — issues relating to data management (short and long term), data quality, curation, retention practices, and metadata structures that enable interoperability, etc., are foundational to optimizing UC's effectiveness and cementing UC's reputation as a leader. UC must leverage expertise within its libraries and partner with technology organizations to address this important need.

a. ***Create a Working Group to guide development***. The Working Group will include three librarians from different campuses across the system, including one representing CDL, and two to three research-focused technologists and/or data-intensive faculty members from different campuses and who are broadly knowledgeable about their local campus research products. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the FGB/Institute. The group shall consult with a minimum of ten faculty members (drawn from multiple campuses) whose research produces a range of data types from representative communities or domains (e.g., data types common to multiple campuses or particularly associated with UC research). The working group will be designated for one year and tasked with the following deliverables:

    - ***Write a canonical set of data guidelines, based on community standards, funder mandates and UC policies (by February 1, 2016).*** These guidelines should not be particular to any campus or domain, but apply broadly to data produced across the system. These guidelines will necessarily be basic, to encompass the wide array of data, and will be driven by practical concerns, including sharing mandates and technology requirements.

    - ***Complete a survey of existing data tools and services in the UC system (by April 1, 2016).*** This survey will expose the current data infrastructure

landscape at the campuses and CDL, and should highlight common goals and services, competing goals and services, and gaps. Significant work has already been done in this area, and this survey will help avoid duplication of effort. Where appropriate, this survey will include information on what is being done at local campuses.

- ***Produce a practical online "data guidebook" for researchers (by August 31, 2016)***. Based on the data guidelines and survey, this guidebook should contain a concise set of directions for data producers, indicating the "UC approved" data services available to them, with clear ties to funder compliance. This document will also include relevant information about local campus processes and services, as outlined in the survey.

- ***Produce a "Data Management at UC" manual that explains in depth the current state of data services (by August 31, 2016).*** This document should be a deep-dive into all of the relevant facts of data management in the UC system, including policy and compliance issues, technology and infrastructure options, and the role of libraries in research. The document will have two main purposes: 1) To provide the foundation for the "Data Guidebook," giving clear and transparent explanations for all decisions and recommendations; 2) to serve as a living document that leads the UC into the immediate and longer future, giving an initial set of guideposts for future data asset management. This could include goals for future funding opportunities, shared development, and new policies.

- **Create an ongoing process to actively monitor and maintain the data services landscape.** The "Data Guidebook" and "Data Management at UC Manual" will need to be updated and maintained on a continuous basis. The working group (or a future group created after the initial year) will be in charge of ongoing updates managed by the institute, with the actual work to be done by key stakeholders.

ACTION 3: Scale discipline-similar requirements.

Not all research areas have large, concentrated discipline-specific data needs that are accommodated by formal structures, such as centers. There is a huge diversity of research and scholarship programs working with smaller and equally valuable data assets. These programs may lack the ability to scale data resources. Institutional and cross-institutional discipline-specific data resources should be leveraged to allow smaller data assets to take advantage of shared resources and scale.

a. ***Create a Working Group to guide planning development.*** The Working Group will include the follow membership three to six faculty members from multiple disciplines with data as defined within this initiative and three to six support staff knowledgeable about data repositories, metadata, and collaboration tools. A

working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the FGB/Institute. The working group will be designated for one year and tasked with the following deliverables:

- ***Data Assets – Vision for discovery and asset description.*** The Working Group will develop, vet, and gain consensus on a plan and suite of data asset descriptors that will define and describe data assets as they are discovered and documented for collaborative use. These data asset descriptors will enable the development of a centralized catalog of data assets. The descriptors will include the tools, services, and systems that are utilized to deliver and maintain the data asset.

- ***Vision for a Data Asset Resource Catalog.*** The Working Group will develop, vet, and gain consensus on the specifications / vision for an online system that will enable UC researchers and digital scholars who are stewards of relatively smaller, individual data resources to register their assets with a centralized catalog/service listing data repositories.

- ***Creation of a Data Asset Resource Catalog.*** The Working Group will serve as an oversight / advisory group to the technical team that will create / acquire (e.g., vendor or cloud service) the online digital asset repository system.

- ***Phase II – Data / research collaboration tool.*** In Phase II, the Working Group will develop, vet, and gain consensus on the specifications / vision for an online system that enables researchers and digital scholars to share information about their data assets and to establish connections for collaboration as part of inter-campus teams. This effort will be tightly linked to other initiatives aimed at creating online data / research collaboration platforms and may include the development of a UC Researcher Profile tool.

- ***Other considerations.*** To support its overall efforts, the Working Group will create a timeline for completing the tasks noted in this document and will also note and escalate any policy issues / considerations that are discovered. Additionally, the Working Group will create a vision for supporting and maintaining this service over time.

ACTION 4: Position health, patient and clinical data.

The five UC medical centers and many health science programs and their attendant health, patient and clinical data present unparalleled data assets for research. The UC ReX and Big Cogito pilot are examples. Key challenges will be standardization of terminology across UC, and the development of appropriate policies and data

governance that allow UC to simultaneously work as one collaborative system in certain situations while promoting healthy competitive innovation and excellence as individual campuses.

a. ***Create a Working Group to guide development***. A relatively small Working Group will be identified and charged with executing the steps listed below. The Working Group will include three School of Medicine CIOs, and two to three research-focused data-intensive faculty members from the Schools of Medicine and who are broadly knowledgeable about their local campus research products. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the FGB/Institute. The group shall consult with a minimum of ten faculty members (drawn from multiple campuses) whose research produces a range of data types from "omics" to "sensing" to patient-reported data to clinical data. The working group will be designated for one year and tasked with the following deliverables:

    - Define a HIPAA compliant approach and infrastructure to advance research collaboration.

    - Identify data workflows, interfaces, and standards to allow for precision medicine within the electronic medical record.

    - Determine a model that provides easy access to de-identified clinical data to faculty outside of the School of Medicine or outside of health sciences.

    - Examine challenges around specific types of data, such as imaging or whole genome as it relates to storage and high performance computing, and report recommendations.

    - Highlight data visualization needs for clinical trials research or research around medical decision making or quality improvement.

    - Engage the lay public – patients and the community – in hypothesis-generating activities around clinical and medical questions.

ACTION 5: Develop systemwide and campus "concierge" services.

"Concierge" (digital technology resource guidance) and related sociotechnical services will bring federated expertise and capabilities together to help guide faculty to the appropriate cyberinfrastructure services to meet their research needs. This was a strong theme at the Summit, with the aim of reducing faculty search time and bringing cloud, national, UC wide and local campus cyberinfrastructure capabilities together. UC needs to sponsor and create "ask an expert" services and provide "how to do things or get things done" guidance.

a. ***Create a Working Group to guide development***. The Working Group will include four to six digital technology staff or management from at least four campuses and who are broadly knowledgeable both their local campus and cloud resources. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as or designate a liaison to the FGB/Institute. The group shall consult with a minimum of ten faculty members (drawn from multiple campuses) whose research requires technology from multiple campus or off-campus resources (e.g., cloud computing, server colocation, local storage, library curation, instructional technology). The working group will be designated for one year and tasked with the following deliverables:

   - ***Identify a lead Digital Technology Resource Advisor ("concierge") for each campus (by January 1, 2016).*** The Working Group may develop additional guidelines for selection and will work with campuses to provide a nominee. These should be high-level professionals or middle management with a clear understanding of their campus's digital technology resources, excellent connections across campus technology providers, a broad understanding of available cloud services and their appropriateness in research applications, demonstrated understanding of the research process, and strong communication skills. These personnel will participate in the systemwide Digital Technology Resource Advisory (DTRA) team and will act as leads for teams on each campus.

   - ***Develop systemwide team charge (by February 1, 2016)***. The systemwide DTRA team will maintain the systemwide digital technology resource index, make references across the system for needed resources, and share best practices and case studies to ensure the highest level of service within each concierge group across the system. The charge should include modes of communication and frequency of meeting for the systemwide team. The systemwide team will be a long-term commitment with a regular communications schedule. The Working Group will refine and flesh out this charge and submit it for Oversight Committee approval.

   - ***Develop guidelines for services and resources to implement at campus and systemwide levels.*** In collaboration with the FGB/Institute and the working groups responsible for federation/shared services, research data management, cyber-platform interconnects, software stores, expertise, and others as appropriate, determine where such services and resources will be located and how they will be supported on an ongoing basis.

   - ***Develop campus-level plans for funding and implementation of digital technology concierge services at each campus (by August 31, 2016)***. In partnership with each campus's VCR and other appropriate stewards, the Working Group will guide each campus in developing an appropriate local

plan to staff a funded team that will provide high-level digital technology resource advice directly to faculty. Campuses are expected to commit funds and human resources to support this important service.

ACTION 6: Build cyber platform interconnects.

UC needs to agree on standards and build the necessary campus network interconnects, scheduler technologies and cloud service management technologies to make it possible for federated facilities and tools to interoperate. This will enable UC to take advantage of cross-system and commercial cloud technologies to assemble services for particular research needs. It may also realize efficiencies.

a. ***Create a Working Group to guide planning development.*** The Working Group will include three to six faculty members whose research might benefit from service federation and/or who are currently utilizing tools that are or would benefit from federation, and three to six support staff who are knowledgeable about various research technologies and the interconnects / middleware available to interconnect these tools. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the FGB/Institute. The working group will be designated for one year and tasked with the following deliverables:

   - ***UC Information Technology Architecture Group (ITAG).*** The Working Group will explore various partnerships as it creates its project plan, including leveraging the UC ITAG group that provides inter-campus architecture / middleware support for UC's operational and analytics / decision support systems.

   - ***Federated repository design.*** The Working Group will engage in the following activities that will yield a prioritized roadmap for platforms that might benefit from the UC sponsored / developed / support interconnects (These efforts will be tightly linked to other initiatives aimed at creating and facilitating cyberinfrastructure federation or the creation of cyberinfrastructure shared services.):

     - Develop, vet, and gain consensus on specifications for a catalog of platforms (systems, tools, other assets, and cloud resources) that are priority candidates for federation.
     - Ensure this catalog also describes the method (or methods) that are most commonly used to interconnect these platforms.
     - Record the disciplines that are (and will be) positively impacted by federating these tools, prioritize opportunities for interconnecting the platforms based on the positive impact to UC's research enterprise, and produce a roadmap with timeline and milestones.

- ***Creation of interconnect resources.*** The Working Group will serve as an oversight / advisory group to the technical team that develops / acquires and deploys the various interconnect services.

- ***Other considerations.*** To support its overall efforts, the Working Group will create a timeline for completing its tasks and will note and escalate any policy issues / considerations that are discovered.

ACTION 7: Build a software store*.*

UC must create a software brokerage infrastructure and appropriate policy for sharing/promoting/buying software applications across the UC system. Similarly, the UC federation should be set up to facilitate a technology channel for data and software with respect to internal and external partnerships. Collectively, UC research is a major producer of software and this asset can be leveraged within the system to enhance research achievements for all.

a. ***Create a Working Group to guide development***. Commission a Working Group by November 2015 comprised of representative members from the following areas: software license managers for academic software, software IP and licensing, UC Research Technology Group (RTG) member experienced with research software, Educational Technology Leadership Group (ETLG) members experienced with educational software, librarians experienced with curation, and a finance person experienced with sales and service of software models. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the overall FGB/Institute. The working group will be designated for one year and tasked with the following deliverables:

   - ***Create an inventory of use cases and models.*** Inventory use cases and categories of software sharing/transaction potential as well as software sharing systems and models across UC campuses and post by February 2016. Include:
     - Internal and external to UC, contributing and using
     - No cost, at cost, buying, supported, unsupported
     - Open source, experimental, level of validation, certification
   - ***Develop criteria and evaluate.*** In parallel, inventory, establish evaluation criteria, and evaluate structures and operating models for national software exchanges and post by Spring 2016. Examples include:
     - A number of national disciplinary institutes have software sharing and download frameworks – Hubone at Purdue
     - The Digital Manufacturing and Design Innovation Institute (DMDII) out of Chicago has partnered with GE on a national software store
     - The Smart Manufacturing Leadership Coalition is building a national software store with an integrated deployment

infrastructure
    - UC and IMS have partnered and prototyped a federated store for software sharing called CASA – Community App Sharing Architecture
- ***Make recommendations to the UC IT Leadership Council (ITLC)***. Review the evaluation matrix with the RTG and ETLG and recommend to the ITLC one or more structures to be considered for pilots, as well as how to structure them, by March 2015.

ACTION 8: Support and build on UC's expertise.

Finally, it is essential to develop platform tools that bring researchers and their work into a more visible, discoverable state to facilitate shared expertise and to increase the potential for collaboration. For example, how does one researcher find another researcher doing something similar with cyberinfrastructure, especially across disciplines? We need to invest in the professional development of research IT staff across the UC system, and build a collaborative cadre of such staff across the system. By staff, we include the full range of domain experts who choose non-faculty career paths supporting researchers, as well as IT experts in infrastructure technologies who keep research operations running. Professional development includes the soft (interpersonal) and hard (technical) skills needed so that research IT professionals can move comfortably from helping to address local problems to participating in cross-campus and multi-campus collaborations.

a. ***Create a Working Group to guide development***. The Working Group will include 3-5 digital technology representatives and 2-3 Library representatives from at least four campuses with broad knowledge of both their local campus and cloud resources. A working group lead responsible for guidance and deliverable management will be designated. The lead will serve as (or designate) a liaison to the overall FGB/Institute. The working group will be designated for one year and tasked with the following deliverables:

    - ***Survey and identify current offerings and best practices (by February 1, 2016)***. The Working Group should communicate with all campuses to survey current offerings in the areas of faculty profiles and research catalogs; other tools that enable the sharing, discoverability, and research collaboration for data, expertise and tools within the campus research community, and formal IT staff training opportunities. They will also study current collaboration and training models at other higher education institutions, EDUCAUSE and other organizations. They should work in cooperation with the federation/shared services, research data management, cyberplatform, and "concierge" working groups during this discovery phase. An online report of findings on collaborative offerings, training offerings, and observed best practices should be produced.

- ***Develop training and internship recommendations (by April 1, 2016).*** Based on the results of the survey, the Working Group should identify recommendations for cross-campus and centralized technology and soft skills training that the Institute should provide or coordinate. They should also recommend a structure for cross-campus internships that facilitate the sharing of new technology competencies across the system. These recommendations should be presented in a report to the FGB/Institute.

- ***Produce an online "Guidebook for Building UC's Technology Expertise" (by June 1, 2016).*** Based on the survey findings, the recommendations in the report, and subsequent analysis by the Working Group, and in collaboration with the cyberplatform working group as appropriate, the Working Group should publish a guide providing best practices for developing UC's technology expertise across the system.