AGENDA

1. (1:00-1:05): Approval of October 22, 2010 Meeting Summary (Kathleen Komar)

2. (1:05-1:15): Governance Proposal (Kathleen Komar)
   Review and Endorsement

3. (1:15-1:45): Research Informatics and Computational Data Initiative (Warren Mori, Jim Davis, Art Toga)
   Review, Input, and Endorsement

4. (1:45-2:15): IT Strategic Plan: Digital Citizen Concept (Jackie Reynolds)
   Review, Input, and Endorsement

5. (2:15-2:40): E-Dossier Update (Tom Rice)
   Information

   Information
   a. IT Investment Strategies Subcommittee
   b. Faculty Digital Presence
   c. Student Systems Interface

7. (2:55-3:00): Next meeting and adjournment (Kathleen Komar)
Information Technology Planning Board
Meeting Summary
Friday, October 22, 2010
3517 Ackerman Union

ITPB Attendees: Jim Davis, Deborah Estrin, Jerry Kang, Patricia Keating, Kathleen Komar, Jonathan Kuo, Vincent Riggs (for Mike Kusunoki), Michelle Lew, John Mamer, Sam Morabito, Warren Mori, Susan Parker (for Gary Strong), Janice Reiff (also for Andrew Leuchter), John Riley, Tim Stowell, Christopher Waterman

Stephanie Hokama, recorder

Absent: Andrew Leuchter, Alan Robinson, Gary Strong, Arthur Toga

Invited Guests: Marsha Smith, Lisa Snyder

Guests: Ross Bollens, Michelle Chen, Bill Labate, Larry Loeker, Tom Phelan, Rose Rocchio, Mike Schilling, Andrew Wissmiller

Chair Kathleen Komar called the meeting to order at 2:03 PM.

Agenda Item #1: Approval of the Meeting Summary and Introduction of New Members (Komar)

The summary from the June 9, 2010 meeting was approved. The agenda format will include links to all the documents needed for the meetings, but no hard copies will be brought for distribution. The agenda will be annotated to indicate updates, information, future voting topics, etc., and will be sent to members at least three business days prior to the meeting.

Kathy Komar then introduced the new members:

Andrew Leuchter, Professor, Psychiatry – Academic Senate Vice Chair
Replaces Ann Karagozian

New Three Year Appointments, 2010-13
Deborah Estrin, Professor, Computer Science
Christopher Lee, Professor, Chemistry and Biochemistry
John Mamer, Professor, Anderson School
John Riley, Professor, Economics

New Undergraduate Representative
Jonathan Kuo

Chair Komar referred to a recently submitted annual report when asked of ITPB’s significant accomplishments in the last year. The board addressed both emergent and long term issues such as the video furnace, TIF recommendations and oversight of the IT Strategic Plan. In the coming
year, the strategic plan will continue to be an agenda item, as well as consideration of problematic budget issues that affect IT.

**Agenda Item #2: Governance Meeting (Komar)**

During the summer, the chairs and co-chairs of ITPB, CITI and CSG met to discuss the function of each group and how they intersect. Representatives from each committee would attend the others’ meetings.

ITPB – IT Planning Board is the primary oversight body for UCLA. This board has an overview of IT throughout campus. It is a combination administration/Academic Senate appointed group. The board considers the strategies, infrastructure, technology renewal, research administration, oversight of implementation, and the balance between central and distributed projects. The ITPB also is the governing board of planning and policy, any IT issue that goes to project status, and any topic that is systemwide.

CITI – Committee on IT Infrastructure is the group that considers funding decisions about projects.

CSG – Common Systems Group provides oversight/input on technical issues. They also have input on what goods/services should be central/dispersed.

**Agenda Item #3: UC Online Instruction Pilot (Davis)**

In January 2010, UCOP discussed the establishment of an eleventh campus (online instruction). The matter was referred to the campuses for consideration and input. The ITPB’s position was that while it supported cyberlearning, it found the emphasis on revenue to be problematic. The entire program was reformed and presented to a systemwide steering committee. An RFP will be distributed from EVC Scott Waugh to the faculty.

The topic of revenue has been put on a back burner. There is more faculty involvement and review of the effectiveness from a student viewpoint. These classes will be fully or largely online across all campuses. There would be a focus on undergraduate education, particularly in the lower division, higher enrollment classes. For example, the Calculus series 1-3 and Chemistry 1-3 have approximately 39,000 enrollees per year.

This program will be run according to faculty guidelines. It is estimated that it will cost $65-75K per course, although no dollar amount has been set. There is no intention of circumventing senate approval. UC must consider that Illinois and Columbia failed in this endeavor. However, it opens the door for UC to prepare for other future cyberlearning issues. There are many questions regarding the student/professor interaction, scalability and language issues. In the case of foreign students, they may receive lectures in English, but speak with TAs in their own languages. ITPB needs to hear from successful programs such as Brian Copenhaver who taught Philosophy online and Vijay Dhir whose department (Engineering) has given successful online courses.
Funding for online courses is a question that needs to be addressed. UC says there are no funds for TA hiring; where will the money for more TAs come from? The RFP process will produce more concrete topics for discussion, so this should remain on the ITPB agenda.

Agenda Item #4: New Developments with CCLE/Moodle and Strategic Issues with Educational Repositories (Lew)

Moodle was begun in 2007 and has approximately 20,000 users per quarter. At the September CSG meeting, CCLE repositories were discussed. Version 1.9 of Moodle is currently being used, with v. 2.0 expected next year. This newer version will be able to link to repositories, enabling faculty to use others’ repositories. There are a lot of developing standards, and ITPB should be kept up-to-date. This item should be on the agenda of future meetings.

Agenda Item #5: Research Informatics and Computational Data Initiative (Mori, Toga, Snyder, Davis, Smith)

Jim Davis gave a historical background on this initiative, noting that ITPB has spent two years on this topic. Research data is an important asset that needs to be managed. EVC Waugh is using the term “digital presence” which came out of the Chancellor’s retreat. NSF and investigators applying for ROIs are concerned about data management.

The campus has a variety of data management needs, such as the humanities (IDRE HASIS), PII in medical research, and simulation data (extremely large data sets). 12 proposals will be developed by Marsha Smith. The framework of proposals will set the stage for proposal endorsement. The concept of these proposals will include internal calls for the campus, projects and workshops, medical informatics data assets, etc. These projects will inform the campus vision, utilizing projects to test and evaluate approaches, as well as define impact. This will provide more information for data stewardship as a part of a broader campus vision.

Agenda Item #6: Reports and Updates

A. Video Furnace Dispute (Amy Blum)

Amy Blum, Campus Counsel, gave an update on the AIME threat to bring suit against UCLA for violation of copyright. There has been no filing, but UCLA has received two more cease and desist notices. PBS is claiming copyright infringement; ironically, universities and university professors provide much of their programming content. Direct Cinema, although not selling a product, are licensing materials for 10 years. Despite the fact that no one has asked for the materials to be returned, they claim we are beyond the scope of the license, therefore in violation of the agreement.

It was pointed out that UCLA should not waste time and money on a three year gain. We should develop a longer strategy that includes lobbying Congress and utilizing the library to negotiate terms. This topic should be discussed in future meetings to determine what faculty require for teaching.
B. FACTECH (Jerry Kang)

A lunch event was held in the School of Law that was likened to “speed dating with technology.” Many people attended and were exposed to various technology devices. It was felt that to obtain buy-in from staff and faculty, a change in dynamic was achievable with these quick “how-to-do” lessons.

Agenda Item #7: Next Meeting and Adjournment (Komar)

The next meeting will be held on Friday, December 10, 1:00-3:00 PM in 3517 Ackerman Union.

The meeting was adjourned at 3:55 PM.
Building UCLA’s Digital Future

An Approach and Action Plan
for Building
Research Informatics, Computational Data and Data Management
Capability

An Information Technology Planning Board Initiative

December 8, 2010

Sponsored by
The Office of Information Technology, The Institute for Digital Research and Education, and The Institute for Informatics.
Building UCLA’s Digital Future

The Office of Information Technology (OIT), in partnership with the Institute for Digital Research and Education (IDRE) and the Informatics Institute (I2), is spearheading a one-year planning process to engage and focus the UCLA community to develop solutions on critical processing, storage, and management issues for research and education data that face the campus as it looks towards its digital future. Grounded in the belief that digital data is an important institutional asset that must be valued, preserved, and made accessible, this effort will include holding planning sessions that center on digital stewardship and cyberinfrastructure issues, and supporting internal informatics\(^1\) and computational data development projects that will inform long-term institutional improvements.\(^2\)

These two activities will augment work already being done to build out UCLA’s cyberinfrastructure and develop a campus vision and action plan for supporting UCLA’s digital research and education efforts through UCLA’s IT governance and leadership.\(^3\) This plan is further driven by the following objectives:

1. To respond to a renewed and growing national commitment to science, technology, and innovation. Addressing computational and data management challenges will enable UCLA researchers to successfully compete in critical growth areas including DNA sequencing, smart therapeutics, smart manufacturing, smart energy management, solar energy, nanotechnology, personalized medicine, regenerative medicine, transformative educational technology, and advanced simulation and modeling, including moving toward Exascale level computation.\(^4\)

2. To harness and elevate UCLA success stories in research and education areas of national significance and to inform and propel an integrated vision and plan for UCLA’s digital future. This is particularly important in order to spur collaboration and to allow researchers to build on existing knowledge – both of which are essential for innovation.

3. To respond with specific strategies to accommodate the exponential growth in digital data that challenges the university at every level (technical, policy, and organizational) and to protect the university’s core competitive asset – research data.\(^5\)

---

\(^1\) Informatics includes both medical and non-medical data. It broadly defines information that can be stored in digital form and accessed electronically: numeric data, text, publications, sensor streams, video, audio, algorithms, software, models and simulations, images, etc.

\(^2\) As used herein, digital stewardship is broadly defined as the careful and responsible management of digital assets. The term has also been variously defined as “the management of digital objects over the long term through careful digital asset management practices” (Harvard Library OIS definition), and the “useful accessibility, annotation, curation, and preservation of research data” (National Academies Press, 2009, *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*).

\(^3\) Using the definition promoted in Developing a *Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies* (2009), “Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.”

\(^4\) On September 21, 2009, President Barack Obama released his *Strategy for American Innovation*. The strategy outlines the Administration’s plans to foster innovation for sustainable growth and the creation of high-quality jobs.

\(^5\) As late as February 2010 the NSF’s Blue Ribbon Task Force on *Sustainable Digital Preservation and Access* noted that the remarkable growth of data-intensive research in all knowledge domains is outpacing our current capacities and abilities to make decisions about long-term preservation.
4. To position UCLA to respond to increasingly specific requirements of national granting agencies (most notably NSF and NIH) to make data sharable and to incorporate data management plans in all future proposals.

5. To respond to the growing governmental demand for transparency in how public funds are used and in how data created through the use of public funds are used, stored, and made available to others. This demand is the impetus for UCLA to place a priority on expanding its digital presence, defined as the integrated approach to making research data – and the results stemming from that data – easily accessible, regardless of its origin, processing, or final destination.

Meeting these stated objectives and integrating the results of the proposed planning sessions and development projects into meaningful action will require participation and leadership at the campus, division, and departmental levels. To be sure, this process will create multiple challenges, and the discussion may require a reassessment of our assumptions and closely held beliefs about research and the responsibilities of the various campus stakeholders.

While accepting that a university as complex as UCLA will spawn more than one approach to the data challenges ahead, this proposal also posits that it is counter-productive for dozens of isolated approaches to be launched in light of our very real budgetary and regulatory situation. We therefore offer this proposal as a means to propel the campus toward a unified vision that can accommodate the diverse needs of campus researchers.

It is of particular importance that UCLA capitalize on the existing work that has been done in this area to inform a campus vision and plan based upon the real challenges that researchers and educators already face regarding their data requirements – from data ingestion, through computation and analysis, to the publication of results and longer-term preservation. The program will be structured so that anyone can participate and will:

1. Use the data management expertise and innovation that already exist and are being utilized on campus in a variety of research and educational efforts including within the Library, the Graduate School of Education and Information and Sciences, Digital Humanities, Social Sciences, and the Social Science Data Archive.

2. Take advantage of the multitude of internal and external reviews reports and recommendations that have been generated on this topic and incorporate them into the planned workshops and projects. (See Appendix 1)

3. Use the internal data inventories that have already been completed to inform the planned workshops and projects. For example, an internal task force recently completed a detailed inventory of north campus data that identified both the strengths and challenges inherent in the current UCLA approach to digital research. (See Appendix 2 for the project team’s final summary report).

The proposed program funding will be used for three basic line items: 1) staff support including a program manager on a one-year contract to manage and coordinate the planning efforts, 2) a series of working sessions on digital stewardship and cyberinfrastructure, and 3) internal informatics and computational data development projects that will inform long-term institutional improvements.

**Working Sessions**

The working sessions will be used to build an understanding of the issues the campus faces in its role as a steward of digital assets. These working sessions will be defined, planned and
implemented through the executive sponsors, a sponsorship group that includes the Library, Medical Enterprise, and the Office of the Vice Chancellor of Research (see Sponsorship and Planning Section below). Each workshop would be in the form of a structured brainstorming session, drawing upon existing work done at UCLA as well as a wealth of external reports on this topic listed as Appendix 1. The sessions will involve all of the campus stakeholders necessary to have a productive discussion about the data stewardship and research infrastructure needs for these specific UCLA communities and recommend actionable outcomes. The sessions will use an external facilitator and call in external experts, as needed.

The primary goals of the sessions are to coalesce a number of independent efforts in medical and non-medical informatics into an integrated effort, build a unified campus vision for supporting UCLA’s digital research and education efforts, create consensus for future expansion of our cyberinfrastructure, and develop an action plan for data stewardship to amplify UCLA’s digital presence.

**Informatics and Computational Data Development Projects**

The second element of the proposal is to provide funding for a series of informatics and computational data development projects that emphasize campus infrastructure and/or common services such as storage, networking, or data centers/repositories. The Institute for Digital Research and Education and the Informatics Institute will use their annual request for proposals process to solicit proposals using criteria endorsed by the IDRE-I2 Executive Committees. Proposal topics could include, but are not limited to, internal projects, workshops, and symposiums in areas of informatics, high-end computation, simulation, imaging, modeling and visualization, data curation, and data management. The committees will provide a peer review of the proposals, organize them into a tangible approach, and make proposal funding recommendations to the executive sponsors of this initiative. At its October 2010 meeting, the IDRE-I2 Executive Committees endorsed the following principles. Successful proposals must:

1. Support innovative research now in informatics and computational data research that will inform the broader campus vision to build UCLA’s digital future.
2. By example, demonstrate that our own research and educational efforts inform and guide a unifying and cohesive vision for UCLA’s digital future.
3. By example, demonstrate that these same efforts can encourage agreement about what data stewardship means for our future – stewardship that encourages people toward common paths to harness, protect and make available our digital assets.
4. As the specifics of the unified campus data vision develop, use actual research projects to test and evaluate specific approaches, strategies, and methodologies.

Successful proposals must also include as a deliverable a summative discussion with campus stakeholders and/or a white paper that describes how the project results can inform campus-wide projects that will build out UCLA’s next cyberinfrastructure and digital stewardship approaches.

---

6 Four reports were drawn upon heavily in the development of this proposal and could be used to outline the major issues that would inform working session agendas. The reports are: the February 2010 NSF Blue Ribbon Task Force Final Report: *Sustainable Digital Preservation and Access*, the February 2009 Educause report: *Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies*, the March 2007 report by the NSF Cyberinfrastructure Council: *Cyberinfrastructure Vision for 21st Century Discovery*, and the 2008 report by UCLA’s Christine Borgman, *Supporting the “Scholarship” in e-Scholarship*. 
Sponsorship and Planning Structure

UCLA’s established IT governance structure (below) should be fully engaged to sponsor, inform, guide, and make decisions about UCLA’s digital future. To augment this structure we are suggesting that the campus form an executive sponsor group representing a partnership between the Office of Information Technology, the Office of the Vice Chancellor for Research, the Library, and the Medical Enterprise. This group will have overall responsibility for project funding recommendations, guiding the planning process and its components through governance and decision-making groups at UCLA, and planning and project oversight. At the conclusion of the planning process the sponsors will produce a final report to the campus leadership that integrates and brings forward the discoveries and recommendations of the working sessions and sponsored projects. To excerpt and paraphrase from the NSF’s Sustainable Digital Data Preservation and Access Network (DataNet) project, that closely parallels UCLA’s internal efforts, the final report should outline a strategy for UCLA’s next cyberinfrastructure and digital stewardship approaches that includes the following:

1. “A vision and rationale that meet critical data needs, and create important new opportunities and capabilities for discovery, innovation, and learning.
2. A sustainable economic model, rooted in a compelling value proposition, with clear incentives to act.
3. An organizational structure that provides for a comprehensive range of expertise and cyberinfrastructure capabilities and digital preservation to ensure active participation and effective use by a wide diversity of individuals.
4. A plan to provide resources and services for the full data management life cycle, where digital data are routinely deposited in well-documented forms, are regularly and easily consulted and analyzed, are openly accessible while suitably protected, and are reliably preserved.
5. A plan to develop new tools and capabilities for learning that integrate research and education at all levels.
6. Activities that provide for active community input and participation in all phases and all aspects of our post-planning efforts, and include a vigorous and comprehensive assessment and evaluation program.”
Planning and Project Budget

We are asking for new planning funds of $610,000 to augment existing resources and matching funds through the various entities participating in the planning effort. The category amounts below are estimates and fully fungible based upon further planning input and priorities.

<table>
<thead>
<tr>
<th>Project or Issue</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OIT staff support and a contract position for program manager</strong> (includes salary, benefits, office space and support, and general overhead)</td>
<td>$160,000</td>
</tr>
<tr>
<td><strong>Working sessions and related activities</strong> to build an understanding of the issues the campus faces in its role as a steward of digital assets. These working sessions will be defined, planned and implemented through the Executive Sponsors, a sponsorship group that includes the Library and Medical Enterprise. Each workshop would be in the form of a structured brainstorming session run by an external facilitator that involves all of the campus stakeholders necessary to have a productive discussion about the data stewardship and research infrastructure needs for these specific UCLA communities.</td>
<td>$100,000</td>
</tr>
<tr>
<td><strong>Informatics and computational development projects</strong></td>
<td>$350,000</td>
</tr>
<tr>
<td>The funding allocation for applied projects and workshops/symposiums will be dependent upon the chosen proposals. It is anticipated that grants will be awarded for a mix of small, medium, and large projects as described below.</td>
<td></td>
</tr>
<tr>
<td><strong>Small Projects</strong>: With the addition of a small amount of resource ($5,000-$15,000), the project or activity will make use of, or expand upon existing cyberinfrastructure to remove an obstacle or solve a problem with regard to data workflows (accessing data, storing data, moving data, ingesting data, etc.).</td>
<td></td>
</tr>
<tr>
<td><strong>Medium Projects</strong>: For grants from $15,001-$50,000, the project proposal must describe research where the results are highly dependent upon new ways of thinking about end-to-end solutions for data ingestion, computation, visualization, short- and long-term storage, archiving and preservation, and data accessibility.</td>
<td></td>
</tr>
<tr>
<td><strong>Large Projects</strong>: For grants from $50,001-$100,000, there must be a direct link between the project to build out components of UCLA’s cyberinfrastructure resources and services whose use is broader than the particular research area of interest. In addition, the project must show potential to enable the research to become a center of excellence under the OIT-IDRE-I2 partnership umbrella and to explicitly position a researcher or research group to be competitive for future external funding.</td>
<td>$350,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$610,000</td>
</tr>
</tbody>
</table>
# Appendix 1 – Literature Review 2000-2010

Arranged chronologically, internal UCLA reports highlighted in gray.

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Report Title, Contributor, and Description</th>
<th>Website Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Borgman, Christine, Professor and Chair, UCLA</td>
<td>Scholarship in the Digital Age: Information, Infrastructure, and the Internet, (Presentation).</td>
<td><a href="http://www.itpb.ucla.edu/documents/">http://www.itpb.ucla.edu/documents/</a></td>
</tr>
<tr>
<td>2008</td>
<td>Palm, Jona, Director, Head of Department of Preservation</td>
<td>The Digital Black Hole, Riksarkivet National Archives, Stockholm, Sweden.</td>
<td><a href="http://www.tape-online.net/docs/Palm_Black_Hole.pdf">http://www.tape-online.net/docs/Palm_Black_Hole.pdf</a></td>
</tr>
<tr>
<td>April, 2008</td>
<td>ITPB</td>
<td>ITPB Subcommittee on Research and Education Data Management</td>
<td><a href="http://www.itpb.ucla.edu">www.itpb.ucla.edu</a></td>
</tr>
<tr>
<td>Year</td>
<td>Author</td>
<td>Report Title, Contributor, and Description</td>
<td>Website Address</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>September, 2009</td>
<td>Davis, Davis, AVP and Chair, Morabito, Jim Davis, AVC and Chair, Morabito, Sam, VC and co-Chair VC and co-Chair</td>
<td>UCLA IT Strategic Plan. Working Document.</td>
<td><a href="http://evc.ucla.edu/reports">http://evc.ucla.edu/reports</a></td>
</tr>
<tr>
<td>September, 2009</td>
<td>Obama, Barrack</td>
<td>President’s Strategy for American Innovation</td>
<td><a href="http://www.whitehouse.gov/sites/default/files/microsites/ostp/innovation-whitepaper.pdf">http://www.whitehouse.gov/sites/default/files/microsites/ostp/innovation-whitepaper.pdf</a></td>
</tr>
<tr>
<td>July 2010</td>
<td>Orszag, Peter R. and Holdren, John P., Director, OMB, Director, Office of Science Technology Policy</td>
<td>M-10-30 Science and Technology Priorities for the FY 2012 Budget</td>
<td><a href="http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-30.pdf">http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-30.pdf</a></td>
</tr>
</tbody>
</table>
Appendix 2 – Data Asset Inventory

North-Campus Data Asset Inventory

FINAL SUMMARY REPORT
Last updated by Lisa M. Snyder on October 18, 2010
Generated following the group meeting of September 2, and modified based on feedback from the DAI consultants on a draft summary report.

DAI Consultants/Project Team Members: Lisa M. Snyder, Heather Kelsey, Zoe Borovsky, Libbie Stephenson, Todd Grappone, Bruce McCrimmon, Stephen Davison

Project history
The inventory of north-campus data was launched as one element in a broader effort to build an understanding of the UCLA digital research environment in order to make informed decisions about long-term infrastructure and research-technology-related initiatives. The project team scoured publicly available websites, faculty rosters, grant data, and other planning documents to populate a Data Asset Inventory spreadsheet that is a composite of the Data Audit Framework used at the University of Glasgow and UCLA-specific data fields. For the purposes of the DAI, north-campus was defined as the departments and centers associated with Humanities Division; Social Sciences Division; School of the Arts and Architecture; Theater, Film & Television; the Graduate School of Education and Information Studies; and the School of Public Affairs. Most entries on the DAI are related to stand-alone research projects, and do not include data generated in the course of typical academic production (e.g. the paper research leading to the publication of a monograph) or course materials. Active work on the DAI has been stopped. The information on the spreadsheet is incomplete, but enough to make generalized points about the state of north-campus research support and data management.

DAI spreadsheet analysis
Findings from the researcher requirements report conducted for an Oxford-based project on “Supporting Data Management Infrastructure in the Humanities”7 are generalizable to the broader humanities community. The work to-date on the UCLA DAI reinforces a number of Wilson and Patrick’s conclusions:

- “Humanities research is hugely diverse, and makes use of an enormous range of types of sources.
- The intellectual value of humanities datasets tends not to depreciate over time.
  [Discussion omitted]
- There appears to be a growing trend towards structuring data within the humanities.
  [Discussion omitted]
- Methods of organizing data also vary considerably, as does the extent to which researchers succeed in creating and maintaining a well-functioning system.
- Good information management is time consuming, and academics often find themselves with insufficient time to keep on top of it.

---

• Most researchers are willing in principle to share their data with others, but in practice choose not to do so, for a variety of reasons.\(^8\)

• Data storage is generally on personally owned machines and backing-up is generally also to personal devices on an ad hoc basis. Knowledge of centrally provided services is limited and they are seldom used.

• There is a significant amount of confusion over the ownership of research data. This is exacerbated by complex situations in which multiple people or organizations may have different claims on the same resource.

• Most researchers are unaware of formal policies relating to data management, although those who have applied for funding may know something of the requirements of a specific funding body.”

The DAI consultants agreed that these general conclusions about Humanities digital research were applicable to the UCLA environment. To these, an analysis of the DAI and discussion with the DAI consultants revealed the following additional UCLA-centric issues:

1. Demand for research IT infrastructure and data solutions (storage, management, archiving) on campus is extensive and growing.

2. There are a handful of extremely well-maintained data archives (e.g., DLP, TFT Film Archive, ISR, and the Ethnomusicology Archive) that serve a percentage of north-campus researchers, but the vast majority has no IT/data support for their research.

3. The ability of existing archival/repository facilities and/or IT units to assume responsibility for more materials or additional file formats has been negatively impacted by the ongoing budget crisis, as evidenced by recent cuts and staff reductions in the social sciences and other campus units.

4. There is growing evidence that UCLA researchers are looking to outside vendors and other institutions for services and support because solutions available on campus do not meet their needs or are more expensive than outside options.

5. Because of the diverse research interests on campus, there is no single solution for data that will be appropriate for all disciplines.

6. External pressure for data management is not as great on north campus as with the medical and computing communities; the exception to this is Social Sciences where faculty are likely to get NSF funding.

7. Raw data is virtually inaccessible, and almost zero outside of Social Sciences.

8. There is strong support for instruction, but little coordinated support for research.

9. Researchers have varying levels of staff and technical expertise available to them.

10. There is no single opinion as to what IT services (and by extension, data management) should be centralized vs. distributed.

11. There is no clear boundary/definition between the various IT/data responsibilities of departments, divisions, schools, and campus entities.

12. The internal and external resources that faculty tap into for research support are often dictated by disciplinary/departmental norms.

13. UCLA’s online presence includes stale or unfinished web content that should be updated or pulled.

14. Digital data production includes the work of an unusually high number of high-profile professionals (artists, architects, filmmakers) where there might be IP issues beyond the academic norm.

\(^8\) Items in bold were identified by the DAI consultants as being very relevant to the UCLA research environment.
Possible next steps on DAI
Suggested tasks for DAI team (Lisa/Bruce/Heather)

1. Analysis of the DAI
   a. Calculation of the percentages of faculty within a department seeking external funding, internal grants, and no activity
   b. Categorization of DAI projects by type (what general comments could be made about the projects and their likely stewardship?)
   c. Patterns of IT-enabled research.

2. Expansion of the DAI
   a. Fill in grant titles from RNET and federal funding spreadsheets
   b. Search for gift/donation information
   c. Continue personal interviews (not likely to be fruitful)
   d. Expand to include the entire campus
   e. Attendance at CSG meeting and other IT support events

3. Further research for IDRE white paper
   a. UCLA needs analysis
   b. Other campus’ IT infrastructure and research support environments
   c. Generalizable workflow and project phases

4. Lobbying efforts
   a. Inclusion of DAI data in DAT (AS)
   b. Expansion of RNET fields (OCGA)
   c. Changes to Goldenrod (OCGA)
   d. Expanded Research/Grant support in OVCR and in divisions/schools
   e. Development of UC-wide policy on data management
   f. Creation of task force for IP vis-à-vis digital content/data
   g. Plan for networking connectivity across north campus
   h. Solutions for deep data storage, web hosting, and repository services

Case study options for north campus OIT planning
There are a number of possible options, the challenge is to define a case study that is clearly within OIT’s domain, or one that would involve campus-wide action and resources and have a demonstrable broad and positive impact on the UCLA community.

1. Coordinate a workshop/brainstorming session that involves all of the necessary stakeholders to discuss the ideal research environment for the UCLA north-campus community.
   a. Include presentation by outside expert on IT infrastructure strategies
   b. Include outside facilitator (i.e., someone without a campus bias)
   c. Develop workshop materials that are thought provoking and speak to the needs of the campus

2. Develop a mechanism for making the DAI data available to the campus (privacy issues on some content)
   a. Integration of DAI with Data Center surveys
   b. Integration of DAI with expanded CAP database
   c. Integration with Library planning efforts (UC3)

3. Test of commercially available and open-source repository systems
4. Exploration of north-campus networking with a focus on the new YRL Research Commons.
5. See list of possible lobbying efforts from above.
Appendix 3: IDRE-I2 RFP – Draft Call Letter

Overview: UCLA aims to create a strategy for its digital future to surface the challenges and develop an actionable plan related to developing and sustaining UCLA’s cyberinfrastructure and the resources and services required to manage, sustain, preserve and make accessible its research and instructional data.

Research and education are increasingly digital and increasingly data-intensive. Digital data is cyclic in that the output of research can provide new hypotheses and insights to drive innovation. As summarized in the NSF’s DataNet program, our major challenge is “how to develop the new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams.”

The Institute for Digital Research and Instruction (IDRE) and Institute for Informatics (I2) are requesting proposals that will inform this strategy. Executive sponsorship and resources for this initiative have been provided by the Chancellor’s Office, the Office of Information Technology (OIT), the Medical Enterprise, the Office of the Vice Chancellor of Research, and the Library. The IDRE-I2 Executive Committee will review all proposals, and based upon available funding and merit, will recommend proposals to the executive sponsors of this initiative. Funding will be made available for projects and workshops and symposia that inform a particular challenge related to digital data management. The proposal process is open to all campus communities and individuals.

Funding: Three levels of funding are available: Small projects (up to $15,000), Medium projects (up to $50,000), and Large projects (up to $100,000).

Proposal Due Date: January 31, 2011

Requirements: Activities would include models for the full data management life cycle, facilitate research where digital data workflows or indexing schemes can impact research in significant ways, and/or develop new tools and capabilities to integrate research and education. The following questions will be considered within the context of the proposal:

- What is the impact on the research and the researcher now?
- How will the results of the project generalize to other researchers or research areas?
- Is the project a part of an overall design strategy that will require future funding? If so explain.
- How does the design strategy generalize or inform a broader campus-wide strategy?
- How can your project be leveraged with other infrastructure projects?
- Does this project necessitate a partnership between organizations to integrate data services in a cohesive way, especially for multidisciplinary work.

Successful proposals must also include as a deliverable a summative discussion with campus stakeholders and/or a white paper that describes how the project results can inform campus-wide projects that will build out UCLA’s next cyberinfrastructure and digital stewardship approaches, policies, and services.

For each category there are also specific criteria as follows:
Small Projects: With the addition of a small amount of resource ($5,000-$15,000), the project or activity will make use of, or expand upon existing cyberinfrastructure to remove an obstacle or solve a problem with regard to data workflows (accessing data, storing data, moving data, ingesting data, etc.).

Medium Projects: For grants from $15,001-$50,000, the project proposal must describe research where the results are highly dependent upon new ways of thinking about end-to-end solutions for data ingestion, computation, visualization, short and long term storage, archiving and preservation and data accessibility.

Large Projects: For grants from $50,001-$100,000, there must be a direct link between the project to build out components of UCLA’s cyberinfrastructure resources and services whose use is broader than the particular research area of interest. In addition the project must show potential to enable the research to become a center of excellence under the OIT-IDRE-I2 partnership umbrella and to explicitly position a researcher or research group to be competitive for future external funding.
Informatics and Computational Data Development Proposal

GOALS
- Facilitate innovative and impactful use of UCLA research
- Foster and facilitate innovative research
- Build search engine prominence
- Facilitate discovery through data accessibility
- Engender broad public reach

Digital (Web) Presence
- Public, Press, Gov’t & Inst. Marketing
- School, Dept & Individual Program
- Research Collaboration Hubs
- Data Hubs
- Library & Publication Linked Data Sets
- Education Portfolios, Collaboration
- Enterprise Info., Training, Resources

Content Management, Survey, Voting, On-Line Meetings and Management, Webinars, Shared Documents, Comment, Registration Validation, Workflow, Publishing, Portfolio Management, Templates
Developing an Electronic Dossier System for UCLA Faculty

Presentation to: Information Technology Planning Board

December 10, 2010
Problems with Current System

• In most units on campus, faculty dossiers are compiled manually and often re-done from scratch year after year

• A single paper copy and publication box with hard copy of articles is shipped from place to place

• Review of paper copy typically must occur in chair’s office; cannot be reviewed on one’s own computer

• Data cannot be used for reporting or for generating campus statistics
Activities Up Until Now

• Conducted visits to the four UC campuses with electronic systems: Davis, Riverside, San Diego, San Francisco

• Established and received recommendations from joint “Committee on Developing Electronic Dossier and Review Systems at UCLA”

  Conclusion: “…the campus should begin the process of developing an electronic dossier system immediately” [including] “both an electronic dossier and review system”

• Conducted interviews with key administrators, faculty, and staff

• Presentations and endorsements from EVC-Provost, CAP, Deans, and Restructuring Steering Committee

• Recommendation by CITI for 9-month planning grant, with expectation that this will lead to the implementation phase
Conclusions from Other Campus Systems

• Systems need to be tailored to each campus’ needs

• No existing system (inside UC or outside product) met the needs of any campus; each tailored it to their personnel and IT system

• Fully interactive systems (e.g., not unsearchable pdf’s) is essential

• Faculty input and buy-in is essential

• Implementation time has varied from 2-8 years, total cost approximates $1.5 - $2 million
Proposed System

- Electronic generation of dossiers
- Ability for faculty to generate other CVs (e.g., NIH grants)
- Hot links for faculty publications
- Preparation of reviews (deans, chairs, outside letters) online
- Ability for all reviewing agencies to conduct electronic review on own computers
- Ability for campus and faculty to track accomplishments for fundraising/grants
- Potential to have all recruitment activities done online
Proposed Timeline Estimate

Planning: Feb – Oct ‘11
Development: Nov ‘11 – Aug ‘12
Pilots and Demonstrations: Sept ‘12 – June ‘13
Implementation and Improvement: July ‘13 – June ‘14
Proposed Planning Team

The Design Committee Provides Overall Governance.

- Senate Leadership
- Deans
- Chairs
- Other Key Constituents

Design Committee
Provide high-level design requirements and performance expectations

The Full-Time Project Leader Manages the Overall Project and Makes Day-to-Day Decisions.

- Overall Project Leader (Full-Time)
  - Overall project planning
  - Project communication
  - Project coordination
  - Planning and budgeting

- Functional Lead (Full-Time)
  - User requirements definition and validation
  - User interface

- Technical Lead (Full-Time)
  - Technical requirements definition
  - Technical architecture

Cross-Functional Planning Committee
Participate in user design and review sessions (committee meeting participation only)

The Cross-Functional Planning Committee Provides Support and Periodic Validation
Building UCLA’s Digital Future

An Approach and Action Plan
for Building
Research Informatics, Computational Data and Data Management Capability

An Information Technology Planning Board Initiative

December 8, 2010

Sponsored by
The Office of Information Technology, The Institute for Digital Research and Education, and The Institute for Informatics.
Building UCLA’s Digital Future

The Office of Information Technology (OIT), in partnership with the Institute for Digital Research and Education (IDRE) and the Informatics Institute (I2), is spearheading a one-year planning process to engage and focus the UCLA community to develop solutions on critical processing, storage, and management issues for research and education data that face the campus as it looks towards its digital future. Grounded in the belief that digital data is an important institutional asset that must be valued, preserved, and made accessible, this effort will include holding planning sessions that center on digital stewardship and cyberinfrastructure issues, and supporting internal informatics\(^1\) and computational data development projects that will inform long-term institutional improvements.\(^2\)

These two activities will augment work already being done to build out UCLA’s cyberinfrastructure and develop a campus vision and action plan for supporting UCLA’s digital research and education efforts through UCLA’s IT governance and leadership.\(^3\) This plan is further driven by the following objectives:

1. To respond to a renewed and growing national commitment to science, technology, and innovation. Addressing computational and data management challenges will enable UCLA researchers to successfully compete in critical growth areas including DNA sequencing, smart therapeutics, smart manufacturing, smart energy management, solar energy, nanotechnology, personalized medicine, regenerative medicine, transformative educational technology, and advanced simulation and modeling, including moving toward Exascale level computation.\(^4\)

2. To harness and elevate UCLA success stories in research and education areas of national significance and to inform and propel an integrated vision and plan for UCLA’s digital future. This is particularly important in order to spur collaboration and to allow researchers to build on existing knowledge – both of which are essential for innovation.

3. To respond with specific strategies to accommodate the exponential growth in digital data that challenges the university at every level (technical, policy, and organizational) and to protect the university’s core competitive asset – research data.\(^5\)

---

\(^1\) Informatics includes both medical and non-medical data. It broadly defines information that can be stored in digital form and accessed electronically: numeric data, text, publications, sensor streams, video, audio, algorithms, software, models and simulations, images, etc.

\(^2\) As used herein, digital stewardship is broadly defined as the careful and responsible management of digital assets. The term has also been variously defined as “the management of digital objects over the long term through careful digital asset management practices” (Harvard Library OIS definition), and the “useful accessibility, annotation, curation, and preservation of research data” (National Academies Press, 2009, *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*).

\(^3\) Using the definition promoted in Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies (2009), “Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.”

\(^4\) On September 21, 2009, President Barack Obama released his *Strategy for American Innovation*. The strategy outlines the Administration’s plans to foster innovation for sustainable growth and the creation of high-quality jobs.

\(^5\) As late as February 2010 the NSF’s Blue Ribbon Task Force on Sustainable Digital Preservation and Access noted that the remarkable growth of data-intensive research in all knowledge domains is outpacing our current capacities and abilities to make decisions about long-term preservation.
4. To position UCLA to respond to increasingly specific requirements of national granting agencies (most notably NSF and NIH) to make data sharable and to incorporate data management plans in all future proposals.

5. To respond to the growing governmental demand for transparency in how public funds are used and in how data created through the use of public funds are used, stored, and made available to others. This demand is the impetus for UCLA to place a priority on expanding its digital presence, defined as the integrated approach to making research data – and the results stemming from that data – easily accessible, regardless of its origin, processing, or final destination.

Meeting these stated objectives and integrating the results of the proposed planning sessions and development projects into meaningful action will require participation and leadership at the campus, division, and departmental levels. To be sure, this process will create multiple challenges, and the discussion may require a reassessment of our assumptions and closely held beliefs about research and the responsibilities of the various campus stakeholders.

While accepting that a university as complex as UCLA will spawn more than one approach to the data challenges ahead, this proposal also posits that it is counter-productive for dozens of isolated approaches to be launched in light of our very real budgetary and regulatory situation. We therefore offer this proposal as a means to propel the campus toward a unified vision that can accommodate the diverse needs of campus researchers.

It is of particular importance that UCLA capitalize on the existing work that has been done in this area to inform a campus vision and plan based upon the real challenges that researchers and educators already face regarding their data requirements – from data ingestion, through computation and analysis, to the publication of results and longer-term preservation. The program will be structured so that anyone can participate and will:

1. Use the data management expertise and innovation that already exist and are being utilized on campus in a variety of research and educational efforts including within the Library, the Graduate School of Education and Information and Sciences, Digital Humanities, Social Sciences, and the Social Science Data Archive.

2. Take advantage of the multitude of internal and external reviews reports and recommendations that have been generated on this topic and incorporate them into the planned workshops and projects. (See Appendix 1)

3. Use the internal data inventories that have already been completed to inform the planned workshops and projects. For example, an internal task force recently completed a detailed inventory of north campus data that identified both the strengths and challenges inherent in the current UCLA approach to digital research. (See Appendix 2 for the project team’s final summary report).

The proposed program funding will be used for three basic line items: 1) staff support including a program manager on a one-year contract to manage and coordinate the planning efforts, 2) a series of working sessions on digital stewardship and cyberinfrastructure, and 3) internal informatics and computational data development projects that will inform long-term institutional improvements.

**Working Sessions**

The working sessions will be used to build an understanding of the issues the campus faces in its role as a steward of digital assets. These working sessions will be defined, planned and
implemented through the executive sponsors, a sponsorship group that includes the Library, Medical Enterprise, and the Office of the Vice Chancellor of Research (see Sponsorship and Planning Section below). Each workshop would be in the form of a structured brainstorming session, drawing upon existing work done at UCLA as well as a wealth of external reports on this topic listed as Appendix 1.6 The sessions will involve all of the campus stakeholders necessary to have a productive discussion about the data stewardship and research infrastructure needs for these specific UCLA communities and recommend actionable outcomes. The sessions will use an external facilitator and call in external experts, as needed.

The primary goals of the sessions are to coalesce a number of independent efforts in medical and non-medical informatics into an integrated effort, build a unified campus vision for supporting UCLA’s digital research and education efforts, create consensus for future expansion of our cyberinfrastructure, and develop an action plan for data stewardship to amplify UCLA’s digital presence.

Informatics and Computational Data Development Projects

The second element of the proposal is to provide funding for a series of informatics and computational data development projects that emphasize campus infrastructure and/or common services such as storage, networking, or data centers/repositories. The Institute for Digital Research and Education and the Informatics Institute will use their annual request for proposals process to solicit proposals using criteria endorsed by the IDRE-I2 Executive Committees. Proposal topics could include, but are not limited to, internal projects, workshops, and symposiums in areas of informatics, high-end computation, simulation, imaging, modeling and visualization, data curation, and data management. The committees will provide a peer review of the proposals, organize them into a tangible approach, and make proposal funding recommendations to the executive sponsors of this initiative. At its October 2010 meeting, the IDRE-I2 Executive Committees endorsed the following principles. Successful proposals must:

1. Support innovative research now in informatics and computational data research that will inform the broader campus vision to build UCLA’s digital future.
2. By example, demonstrate that our own research and educational efforts inform and guide a unifying and cohesive vision for UCLA’s digital future.
3. By example, demonstrate that these same efforts can encourage agreement about what data stewardship means for our future – stewardship that encourages people toward common paths to harness, protect and make available our digital assets.
4. As the specifics of the unified campus data vision develop, use actual research projects to test and evaluate specific approaches, strategies, and methodologies.

Successful proposals must also include as a deliverable a summative discussion with campus stakeholders and/or a white paper that describes how the project results can inform campus-wide projects that will build out UCLA’s next cyberinfrastructure and digital stewardship approaches.

----

6 Four reports were drawn upon heavily in the development of this proposal and could be used to outline the major issues that would inform working session agendas. The reports are: the February 2010 NSF Blue Ribbon Task Force Final Report: Sustainable Digital Preservation and Access, the February 2009 Educause report: Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies, the March 2007 report by the NSF Cyberinfrastructure Council: Cyberinfrastructure Vision for 21st Century Discovery, and the 2008 report by UCLA’s Christine Borgman, Supporting the “Scholarship” in e-Scholarship.
Sponsorship and Planning Structure

UCLA’s established IT governance structure (below) should be fully engaged to sponsor, inform, guide, and make decisions about UCLA’s digital future. To augment this structure we are suggesting that the campus form an executive sponsor group representing a partnership between the Office of Information Technology, the Office of the Vice Chancellor for Research, the Library, and the Medical Enterprise. This group will have overall responsibility for project funding recommendations, guiding the planning process and its components through governance and decision-making groups at UCLA, and planning and project oversight.

At the conclusion of the planning process the sponsors will produce a final report to the campus leadership that integrates and brings forward the discoveries and recommendations of the working sessions and sponsored projects. To excerpt and paraphrase from the NSF’s *Sustainable Digital Data Preservation and Access Network (DataNet)* project, that closely parallels UCLA’s internal efforts, the final report should outline a strategy for UCLA’s next cyberinfrastructure and digital stewardship approaches that includes the following:

1. “A vision and rationale that meet critical data needs, and create important new opportunities and capabilities for discovery, innovation, and learning.

2. A sustainable economic model, rooted in a compelling value proposition, with clear incentives to act.

3. An organizational structure that provides for a comprehensive range of expertise and cyberinfrastructure capabilities and digital preservation to ensure active participation and effective use by a wide diversity of individuals.

4. A plan to provide resources and services for the full data management life cycle, where digital data are routinely deposited in well-documented forms, are regularly and easily consulted and analyzed, are openly accessible while suitably protected, and are reliably preserved.

5. A plan to develop new tools and capabilities for learning that integrate research and education at all levels.

6. Activities that provide for active community input and participation in all phases and all aspects of our post-planning efforts, and include a vigorous and comprehensive assessment and evaluation program.”
Planning and Project Budget

We are asking for new planning funds of $610,000 to augment existing resources and matching funds through the various entities participating in the planning effort. The category amounts below are estimates and fully fungible based upon further planning input and priorities.

<table>
<thead>
<tr>
<th>Project or Issue</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OIT staff support and a contract position for program manager</strong> for one year (includes salary, benefits, office space and support, and general overhead)</td>
<td>$160,000</td>
</tr>
<tr>
<td><strong>Working sessions and related activities</strong> to build an understanding of the issues the campus faces in its role as a steward of digital assets. These working sessions will be defined, planned and implemented through the Executive Sponsors, a sponsorship group that includes the Library and Medical Enterprise. Each workshop would be in the form of a structured brainstorming session run by an external facilitator that involves all of the campus stakeholders necessary to have a productive discussion about the data stewardship and research infrastructure needs for these specific UCLA communities.</td>
<td>$100,000</td>
</tr>
</tbody>
</table>

**Informatics and computational development projects**

The funding allocation for applied projects and workshops/symposiums will be dependent upon the chosen proposals. It is anticipated that grants will be awarded for a mix of small, medium, and large projects as described below.

**Small Projects**: With the addition of a small amount of resource ($5,000-$15,000), the project or activity will make use of, or expand upon existing cyberinfrastructure to remove an obstacle or solve a problem with regard to data workflows (accessing data, storing data, moving data, ingesting data, etc.).

**Medium Projects**: For grants from $15,001-$50,000, the project proposal must describe research where the results are highly dependent upon new ways of thinking about end-to-end solutions for data ingestion, computation, visualization, short- and long-term storage, archiving and preservation, and data accessibility.

**Large Projects**: For grants from $50,001-$100,000, there must be a direct link between the project to build out components of UCLA’s cyberinfrastructure resources and services whose use is broader than the particular research area of interest. In addition, the project must show potential to enable the research to become a center of excellence under the OIT-IDRE-I2 partnership umbrella and to explicitly position a researcher or research group to be competitive for future external funding.

$350,000

**Total** $610,000
<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Report Title, Contributor, and Description</th>
<th>Website Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>Borgman, Christine, Professor and Chair, UCLA</td>
<td>Scholarship in the Digital Age: Information, Infrastructure, and the Internet, (Presentation).</td>
<td><a href="http://www.itpb.ucla.edu/documents/">http://www.itpb.ucla.edu/documents/</a></td>
</tr>
<tr>
<td>January 2007</td>
<td>Peterson, Michael President, Porter, Jeff, Chair, Zasman, Gary,Co-Chair, Jojica, Peter, Co-Chair</td>
<td>100 Year Archive Requirements Survey, Storage Networking Industry Association (SNIA) Data Management Forum.</td>
<td><a href="http://www.snia.org/forums/dmf/knowledge/100Yr">http://www.snia.org/forums/dmf/knowledge/100Yr</a> ATF_Archive-Requirements-Survey_20070619.pdf</td>
</tr>
<tr>
<td>2008</td>
<td>Palm, Jona, Director, Head of Department of Preservation</td>
<td>The Digital Black Hole, Riksarkivet National Archives, Stockholm, Sweden.</td>
<td><a href="http://www.tape-online.net/docs/Palm_Black_Hole.pdf">http://www.tape-online.net/docs/Palm_Black_Hole.pdf</a></td>
</tr>
<tr>
<td>April, 2008</td>
<td>ITPB</td>
<td>ITPB Subcommittee on Research and Education Data Management</td>
<td><a href="http://www.itpb.ucla.edu">www.itpb.ucla.edu</a></td>
</tr>
<tr>
<td>Year</td>
<td>Author</td>
<td>Report Title, Contributor, and Description</td>
<td>Website Address</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>-------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>September, 2009</td>
<td>Davis, Davis, AVP and Chair, Morabito, Jim Davis, AVC and Chair, Morabito, Sam, VC and co-Chair VC and co-Chair</td>
<td>UCLA IT Strategic Plan. Working Document.</td>
<td><a href="http://evc.ucla.edu/reports">http://evc.ucla.edu/reports</a></td>
</tr>
<tr>
<td>September, 2009</td>
<td>Obama, Barrack</td>
<td>President’s Strategy for American Innovation</td>
<td><a href="http://www.whitehouse.gov/sites/default/files/microsites/ostp/innovation-whitepaper.pdf">http://www.whitehouse.gov/sites/default/files/microsites/ostp/innovation-whitepaper.pdf</a></td>
</tr>
</tbody>
</table>
Appendix 2 – Data Asset Inventory

North-Campus Data Asset Inventory

FINAL SUMMARY REPORT
Last updated by Lisa M. Snyder on October 18, 2010
Generated following the group meeting of September 2, and modified based on feedback from the DAI consultants on a draft summary report.
DAI Consultants/Project Team Members: Lisa M. Snyder, Heather Kelsey, Zoe Borovsky, Libbie Stephenson, Todd Grappone, Bruce McCrimmon, Stephen Davison

Project history
The inventory of north-campus data was launched as one element in a broader effort to build an understanding of the UCLA digital research environment in order to make informed decisions about long-term infrastructure and research-technology-related initiatives. The project team scoured publicly available websites, faculty rosters, grant data, and other planning documents to populate a Data Asset Inventory spreadsheet that is a composite of the Data Audit Framework used at the University of Glasgow and UCLA-specific data fields. For the purposes of the DAI, north-campus was defined as the departments and centers associated with Humanities Division; Social Sciences Division; School of the Arts and Architecture; Theater, Film & Television; the Graduate School of Education and Information Studies; and the School of Public Affairs. Most entries on the DAI are related to stand-alone research projects, and do not include data generated in the course of typical academic production (e.g. the paper research leading to the publication of a monograph) or course materials. Active work on the DAI has been stopped. The information on the spreadsheet is incomplete, but enough to make generalized points about the state of north-campus research support and data management.

DAI spreadsheet analysis
Findings from the researcher requirements report conducted for an Oxford-based project on “Supporting Data Management Infrastructure in the Humanities”7 are generalizable to the broader humanities community. The work to-date on the UCLA DAI reinforces a number of Wilson and Patrick’s conclusions:

- “Humanities research is hugely diverse, and makes use of an enormous range of types of sources.
- The intellectual value of humanities datasets tends not to depreciate over time.
- [Discussion omitted]
- There appears to be a growing trend towards structuring data within the humanities.
- [Discussion omitted]
- Methods of organizing data also vary considerably, as does the extent to which researchers succeed in creating and maintaining a well-functioning system.
- Good information management is time consuming, and academics often find themselves with insufficient time to keep on top of it.

Most researchers are willing in principle to share their data with others, but in practice choose not to do so, for a variety of reasons.\(^8\)

Data storage is generally on personally owned machines and backing-up is generally also to personal devices on an ad hoc basis. Knowledge of centrally provided services is limited and they are seldom used.

There is a significant amount of confusion over the ownership of research data. This is exacerbated by complex situations in which multiple people or organizations may have different claims on the same resource.

Most researchers are unaware of formal policies relating to data management, although those who have applied for funding may know something of the requirements of a specific funding body.”

The DAI consultants agreed that these general conclusions about Humanities digital research were applicable to the UCLA environment. To these, an analysis of the DAI and discussion with the DAI consultants revealed the following additional UCLA-centric issues:

1. Demand for research IT infrastructure and data solutions (storage, management, archiving) on campus is extensive and growing.
2. There are a handful of extremely well-maintained data archives (e.g., DLP, TFT Film Archive, ISR, and the Ethnomusicology Archive) that serve a percentage of north-campus researchers, but the vast majority has no IT/data support for their research.
3. The ability of existing archival/repository facilities and/or IT units to assume responsibility for more materials or additional file formats has been negatively impacted by the ongoing budget crisis, as evidenced by recent cuts and staff reductions in the social sciences and other campus units.
4. There is growing evidence that UCLA researchers are looking to outside vendors and other institutions for services and support because solutions available on campus do not meet their needs or are more expensive than outside options.
5. Because of the diverse research interests on campus, there is no single solution for data that will be appropriate for all disciplines.
6. External pressure for data management is not as great on north campus as with the medical and computing communities; the exception to this is Social Sciences where faculty are likely to get NSF funding.
7. Raw data is virtually inaccessible, and almost zero outside of Social Sciences.
8. There is strong support for instruction, but little coordinated support for research.
9. Researchers have varying levels of staff and technical expertise available to them.
10. There is no single opinion as to what IT services (and by extension, data management) should be centralized vs. distributed.
11. There is no clear boundary/definition between the various IT/data responsibilities of departments, divisions, schools, and campus entities.
12. The internal and external resources that faculty tap into for research support are often dictated by disciplinary/departmental norms.
13. UCLA’s online presence includes stale or unfinished web content that should be updated or pulled.
14. Digital data production includes the work of an unusually high number of high-profile professionals (artists, architects, filmmakers) where there might be IP issues beyond the academic norm.

\(^8\) Items in bold were identified by the DAI consultants as being very relevant to the UCLA research environment.
Possible next steps on DAI
Suggested tasks for DAI team (Lisa/Bruce/Heather)

1. Analysis of the DAI
   a. Calculation of the percentages of faculty within a department seeking external funding, internal grants, and no activity
   b. Categorization of DAI projects by type (what general comments could be made about the projects and their likely stewardship?)
   c. Patterns of IT-enabled research.

2. Expansion of the DAI
   a. Fill in grant titles from RNET and federal funding spreadsheets
   b. Search for gift/donation information
   c. Continue personal interviews (not likely to be fruitful)
   d. Expand to include the entire campus
   e. Attendance at CSG meeting and other IT support events

3. Further research for IDRE white paper
   a. UCLA needs analysis
   b. Other campus’ IT infrastructure and research support environments
   c. Generalizable workflow and project phases

4. Lobbying efforts
   a. Inclusion of DAI data in DAT (AS)
   b. Expansion of RNET fields (OCGA)
   c. Changes to Goldenrod (OCGA)
   d. Expanded Research/Grant support in OVCR and in divisions/schools
   e. Development of UC-wide policy on data management
   f. Creation of task force for IP vis-à-vis digital content/data
   g. Plan for networking connectivity across north campus
   h. Solutions for deep data storage, web hosting, and repository services

Case study options for north campus OIT planning
There are a number of possible options, the challenge is to define a case study that is clearly within OIT’s domain, or one that would involve campus-wide action and resources and have a demonstrable broad and positive impact on the UCLA community.

1. Coordinate a workshop/brainstorming session that involves all of the necessary stakeholders to discuss the ideal research environment for the UCLA north-campus community.
   a. Include presentation by outside expert on IT infrastructure strategies
   b. Include outside facilitator (i.e., someone without a campus bias)
   c. Develop workshop materials that are thought provoking and speak to the needs of the campus

2. Develop a mechanism for making the DAI data available to the campus (privacy issues on some content)
   a. Integration of DAI with Data Center surveys
   b. Integration of DAI with expanded CAP database
   c. Integration with Library planning efforts (UC3)

3. Test of commercially available and open-source repository systems

4. Exploration of north-campus networking with a focus on the new YRL Research Commons.
5. See list of possible lobbying efforts from above.
Appendix 3: IDRE-I2 RFP – Draft Call Letter

**Overview:** UCLA aims to create a strategy for its digital future to surface the challenges and develop an actionable plan related to developing and sustaining UCLA’s cyberinfrastructure and the resources and services required to manage, sustain, preserve and make accessible its research and instructional data.

Research and education are increasingly digital and increasingly data-intensive. Digital data is cyclic in that the output of research can provide new hypotheses and insights to drive innovation. As summarized in the NSF’s DataNet program, our major challenge is “how to develop the new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams.”

The Institute for Digital Research and Instruction (IDRE) and Institute for Informatics (I2) are requesting proposals that will inform this strategy. Executive sponsorship and resources for this initiative have been provided by the Chancellor’s Office, the Office of Information Technology (OIT), the Medical Enterprise, the Office of the Vice Chancellor of Research, and the Library. The IDRE-I2 Executive Committee will review all proposals, and based upon available funding and merit, will recommend proposals to the executive sponsors of this initiative. Funding will be made available for projects and workshops and symposia that inform a particular challenge related to digital data management. The proposal process is open to all campus communities and individuals.

**Funding:** Three levels of funding are available: Small projects (up to $15,000), Medium projects (up to $50,000), and Large projects (up to $100,000).

**Proposal Due Date:** January 31, 2011

**Requirements:** Activities would include models for the full data management life cycle, facilitate research where digital data workflows or indexing schemes can impact research in significant ways, and/or develop new tools and capabilities to integrate research and education. The following questions will be considered within the context of the proposal:

- What is the impact on the research and the researcher now?
- How will the results of the project generalize to other researchers or research areas?
- Is the project a part of an overall design strategy that will require future funding? If so explain.
- How does the design strategy generalize or inform a broader campus-wide strategy?
- How can your project be leveraged with other infrastructure projects?
- Does this project necessitate a partnership between organizations to integrate data services in a cohesive way, especially for multidisciplinary work.

Successful proposals must also include as a deliverable a summative discussion with campus stakeholders and/or a white paper that describes how the project results can inform campus-wide projects that will build out UCLA’s next cyberinfrastructure and digital stewardship approaches, policies, and services.

For each category there are also specific criteria as follows:
**Small Projects:** With the addition of a small amount of resource ($5,000-$15,000), the project or activity will make use of, or expand upon existing cyberinfrastructure to remove an obstacle or solve a problem with regard to data workflows (accessing data, storing data, moving data, ingesting data, etc.).

**Medium Projects:** For grants from $15,001-$50,000, the project proposal must describe research where the results are highly dependent upon new ways of thinking about end-to-end solutions for data ingestion, computation, visualization, short and long term storage, archiving and preservation and data accessibility.

**Large Projects:** For grants from $50,001-$100,000, there must be a direct link between the project to build out components of UCLA’s cyberinfrastructure resources and services whose use is broader than the particular research area of interest. In addition the project must show potential to enable the research to become a center of excellence under the OIT-IDRE-I2 partnership umbrella and to explicitly position a researcher or research group to be competitive for future external funding.