UC Irvine Office of Research, Office of Information Technology and The Libraries

Faculty Assessment of the State of Research Computing (FASRC) at University of California, Irvine

May 22, 2013

Contents

Faculty Assessment of the State of Research Computing (FASRC) at UCI

May 22, 2013

Executive Summary

This report provides a faculty-led assessment on the state of research computing at UC Irvine. At the request of the UCI Vice Chancellor for Research, Associate Vice Chancellor for Information Technology, and the Campus Librarian, a faculty committee was formed which also included staff from the Libraries, and Offices of Research and Information Technology. During the summer and fall of 2012, the committee met with and surveyed campus faculty on their opinions and concerns about research computing services and support at UCI. Along with an extensive online survey (with 125 respondents), focus group meetings and a capstone 'Townhall' meeting were convened in which a statistically meaningful faculty response was captured. This report summarizes these results along with individual comments and analyses done by the committee members.

Informed by this information, the committee is making a set of recommendations to the campus administration on the continuing and expanding need for research computing services (interpreted in a very broad form). Among these recommendations (in no particular order) are:

- UCI needs to develop a research data storage system and management services for long term and safe storage of large, important raw data sets and a large number of smaller, but even more critical, processed data sets. This need crosses all boundaries of research computing interest and is the most critical concern brought out by this project. This need is so strong that, unlike all other identified needs, faculty members are willing to contribute grant funds to the construction of such a system.
- UCI needs more research computing support staff in OIT and the Libraries:
    - System administration and programming support for the 'condo' research computing clusters which support faculty consortia, e.g., the HPC and Green Planet clusters.
    - Short-term, project-based programming services for research projects in the areas of database, web page, and system development.
    - Management, preservation, and organization of research project data especially in relation to increasingly critical data management and preservation demands of funding agencies.
- UCI needs to develop a much faster network for the movement of research data both across campus and between UCI and other locations. High capacity, high speed network equipment is needed to bridge campus sites that generate, store, and process the large data upon which all areas of research computing depend. An example of within-campus transfer would be from the Genomics High-Throughput Facility in the School of Medicine to the HPC compute cluster in

Engineering Gateway; examples of off-campus transfer include CERN, and the National Center for Atmospheric Research.

- UCI needs better communication to faculty regarding what research computing resources are available to them on campus, the accessibility of resources, and who to contact for resources.

The above recommendations are not meant to determine a specific plan for how to meet these needs nor a priority order for any of the specific recommendations.

Background Information

In Spring 2012, the Vice Chancellor of Research, John Hemminger, the Associate Vice Chancellor of Information Technology, Dana Roode, and the University Librarian, Lorelei Tanji, asked their staffs to assess the state of research computing at UCI. The results of the assessment were to be used to determine the medium (1-3 years) and long term direction and priorities for research computing services and support from the sponsoring units as well as to provide top level campus management direction in support of campus research goals.

The project was to be led by a faculty-led committee that was selected to have representatives from all areas of campus that required research computing defined in its broadest sense, i.e., any faculty research activity that depended critically on all but the most common forms of information technology (e-mail, word processing, etc.) to succeed. With this broad definition, faculty demographics were not limited to the traditional 'computationally' intensive users from the traditional research areas (physical science, engineering, and some areas in the biological sciences). As such, faculty representatives were chosen by the organizers from almost all research areas on campus. Appendix A shows the list of faculty and the staff representatives from the sponsoring organizations.

The project was modeled on a similar assessment at UCSD in 2009-10, the "Blueprint for the Digital University, A Report of the UCSD Research Cyberinfrastructure Design Team" http://research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf. In November 2011, a 'Summit' meeting between UC Vice Chancellors and CIOs held at UCLA also prompted concern about the state of research computing at UCI and the need for a faculty perspective on what were the current issues and problems as well as our long-term needs for research computing services and support.

Following the UCSD model, a faculty committee was created by the sponsoring groups in coordination with School Deans, School-based IT Directors, and the Research Computing Support group within OIT. The resulting committee is being lead by three faculty 'editors' —Profs. Robin Bush of Biological Sciences, Antoinette LaFarge of Arts, and Doug Tobias of Physical Sciences—who edited this report and created the set of recommendations to the sponsoring organizations. The committee met three times to discuss the project plan, review the survey results, and discuss this report along with many online conversations. A website containing project documents resides at http://sites.uci.edu/fasrc.

The main initial effort of the committee has been to elicit comments from UCI faculty members on their opinions, concerns, and needs in research computing. Three main tools were used: a campus-wide online survey using the EEE-based survey tool; a series of focus group meetings involving in-person discussions between faculty and the committee members; and a 'Townhall' meeting where the sponsoring unit leaders heard directly from faculty.

The survey was open to comment from October 2012 until February of 2013 and received 125 respondents (of 1200 invited faculty). The focus groups were held from November 2012 to January 2013 in which approximately 25 faculty members attended. Lastly, the Townhall meeting occurred on January 7, 2013.

Report Summary

The second meeting of the FASRC Committee occurred after approximately 110 faculty had taken the survey and Tony Soeller and Allen Schiano of OIT had partially analyzed the responses for trend and strength of response (see Appendix C for the 27 survey questions).

One important result was that the faculty had a broad interpretation of "research computing," in particular with respect to the Arts and Humanities. These researchers do not use computing clusters, but it is clear that they are not simply using computers for e-mail, word processing, calendaring, etc. Examples of non-cluster use include video in the performing arts, digital archiving in the visual arts, and the assemblage of knowledge as web-based content such as the Vietnamese American Oral History Project (http://sites.uci.edu/vaohp/).

We found that the number of faculty members with strong research computing needs has increased at a faster rate than the numbers of support personnel provided by OIT, the Libraries, and their individual schools. While Appendix B shows data confirming this trend in one School, the committee feels this trend is a campus-wide phenomenon. The committee's conclusion is that we are in need of more research computing support personnel, especially in those areas highlighted in this report.

Our survey response of 125 faculty members represents more than 10% of the campus and as such likely represents an accurate portrayal of campus faculty opinions. The results of the focus group meetings and Townhall meeting brought out very similar comments (see Appendix E) albeit with a smaller sample.

The basic findings of this report fall into two categories: there is a need for more support personnel for research computing, and a need to better store and manage highly critical research data. Our findings are very similar to those in the 2009 UCSD report. However, there were some differences that highlight the progress that UCI had already made in several key areas:

- Co-location services for research IT equipment: UCI has had such a service for researchers for many years. Faculty now do not consider there being a need for expansion of further development of the service. Co-location services are meeting the needs of current faculty.

- Cluster computing services: UCI has two cluster computers, HPC and Green Planet, that already are meeting the needs of many faculty groups. The need to further develop these successful services by adding more computational power is being addressed through both central funding and faculty research grants. However, both the present and future operation of these clusters is negatively impacted by an insufficient number of support staff. Expanding the human resources associated with our computing clusters is critical to the Schools of Physical Sciences, Biological Sciences, Medicine, and Engineering.
- High speed networking: Only a few faculty members mentioned problems with the performance of the current UCInet for their research needs. However, the data storage safety concerns addressed by the faculty may be an indicator of this need but in a form that does not appear immediate. More on this below.

We now outline the major results of the project and the recommendations of the committee:

The Need for a Redundant, Long Term, Research Data Storage and Management Services

Our survey clearly revealed that the UCI research community would benefit from more access to redundant, long-term data storage and data management services. Many faculty members do not systematically protect their data while others use off-site storage services that are not protected by agreement with UCI. Such services are only subject to 'consumer' level assurances about reliability and safety. In the survey several faculty members reported problems accessing media on old storage devices for which reading devices no longer exist, and difficulties in making conversions from old to new media.

Additionally, Federal funding agencies' requirements that sponsored data are to be accessible from outside of the University for extended periods of time, as well as researchers' recurring need to reuse legacy data, are increasingly critical. The February 23, 2013 White House directive (http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf) to Federal departments and agencies which mandates public access to federally funded research output, including data, compels the university to meet this need.

Over 50% of faculty members who responded to the survey said they would financially support a long term storage system using funds from their grants or departmental/school funds. This commitment strongly indicates both the need and priority of a storage service to faculty. However, it should be borne in mind that most faculty in Humanities and the Arts still do not have access to large outside research grants, even though their use of research computing has ramped up along with that of faculty in other schools; this creates a situation where the real and increasing needs of some faculty may not be able to be met with a simple cost-sharing model.

Our assessment is that long-term research data storage, and associated data management, is the single most critical research computing need not being met on campus. The FASRC committee believes that a well-run data storage service would allow many faculty groups to coordinate data storage using a centralized system, foster research collaboration, and provide access to archived research data. Faculty

expressed a need for having a secure place to archive their data, if not centrally, elsewhere on or off campus.

As a major component of the University's scholarly product, research data must not only be stored securely but preserved and curated in trusted repositories so that the data remain accessible to the research community after a project is completed. Such accessibility enables secondary analysis of research data originally collected by University faculty and researchers.

The Need for More Research Computing Support Staff

A large component of the concerns voiced by faculty and the committee relate to the lack of support staff in many specific areas. There are unmet needs within units that are only important to those units, e.g., staff to operate devices and facilities, support for specific applications, etc. But there are some areas that cut across wide swaths of campus, one of which is the fact that faculty and graduate students often provide computing support for their colleagues which impacts time they could devote to working on their own research. Were commonly needed services assigned to trained IT staff we could achieve an improvement in efficiency akin to that which occurred when OIT constructed 'condo' research clusters and assumed cluster administration previously handled by the faculty.

Some of the most pressing needs by researchers involved assistance with transient tasks, such as the initial creation of a database or web site, with writing scripts, with data conversion and with task automation. These needs could be met in part by IT staff who act as 'hired guns', a model that OIT has successfully implemented in the areas of GIS, bioinformatics, parallel programming, and scientific visualization.

We see the need for the following additional staff:

- System administrators and application programming staff for the HPC and Green Planet clusters. These clusters are critical to a significant portion of faculty in several schools. We do not presently have enough staff to maintain the current level of expected support while the clusters continue to expand. The faculty members were very vocal about this need.
- Additional programming staff as 'hired guns', i.e., limited duration support for any one research effort or project, in support of database, web, data conversion, and research applications. This need extends across campus, from the sciences to the arts.
- Additional staff with expertise in data management planning, data preservation, data meta-data organization. This need comes from several parallel concerns: 1) the increasing demands of funding agencies to have funded research data shared with others; 2) the funding agencies' requirements to have fully configured data management plans; and, 3) the difficulty that researchers have with organizing, preserving, and creating 'meta-data' catalogs for large datasets.
- The creation of a long-term research data storage service as outlined below will also require additional system administration staff with expertise in large scale (PetaByte and larger), distributed file and management systems.

The Need for Higher Speed and Capacity Networking

Unlike the 2009 UCSD assessment, overall, the UCI faculty members do not see a pressing need for a massive increase in the speed of the campus computing network. Faster networking is however a critical need of a small set of faculty, particularly those who transfer large amounts of data to and from the campus to remote sites.

Faster networking will, however, be needed should we implement the redundant, long term storage system mentioned previously. Moving large amounts of data from multiple sites to a centralized site or moving data between several distributed sites (to improve data redundancy) requires a network with the ability to move Petabytes (the equivalent of a thousand modern disk drives) of data each day. While the current UCInet is not at 'capacity' at present, we would be pushing capacity with added load of a massive storage system. The committee suggests developing an enhanced capacity between around 10 selected buildings with heavy research usage. Particular targets are buildings that house 'big data' creation engines, e.g., genomics sequencers, or where large processed data is known to exist, e.g., archives of media, such as high definition digital video. These buildings would be the starting points for an augmented research network. NSF and other funding agencies have calls for proposals for such networks that UCI will pursue.

The Need to Better Communicate Availability of Research Computing Resources

In the faculty survey and focus group meetings several faculty members expressed concerns that they did not know what research computing resources on campus were available to them, or that they did not know who to contact, or where to look, for information about resources. In some cases, faculty know about research computing resources but they did not know how to access those resources. New faculty, in particular, experience these problems.

The committee recommends that a better mechanism be developed to communicate to faculty what are the research computing resources on campus and how to access those resources. Traditional emails and web sites are not sufficient for communicating with faculty about current services available to them. Coordination between campus service providers and departmental and school support groups to provide venues for one-on-one and group presentations (e.g., departmental meetings) were deemed to be better ways to convey this information to faculty.

Conclusion

The overwhelming consensus on the part of all participants in this assessment, whether from OIT, the sciences, the arts or the library, was that UCI currently suffers from a startling lack of human support for research computing.

We cannot even begin to meet the growing demand for data management and archiving by federal funding agencies without additional storage capacity, but hardware alone is useless in the absence of the highly trained IT personnel required to manage such systems.

The current level of support for cluster computing is barely meeting the needs of faculty in the sciences; in the near future needs for increased networking speed will impact campus research, particularly among those who collaborate internationally.

As a result of this assessment, faculty in the sciences, who to date have been most active in campus planning for research computing, now see much common need with their colleagues in the arts and humanities, and can better guide future growth with the entire campus in mind.

Everyone involved in this project gained insight into ways the UCI Libraries can help build research computing at UCI in the future.

Appendix A: FASRC Committee Members

Faculty

- Robin Bush, School Biological Sciences
- Donald Dabdub, School of Engineering
- David Goldberg, School of Humanities
- Antoinette LaFarge, School of the Arts
- Tony Long, School of Biological Sciences
- Christa Lopes, School of Information & Computer Sciences
- James Meeker, School of Social Ecology
- Maria Pantelia, School of Humanities
- Suzanne Sandmeyer, School of Medicine
- Doug Tobias, School of Physical Sciences
- Steve White, School of Medicine
- Ted Wright, School of Social Sciences

Staff

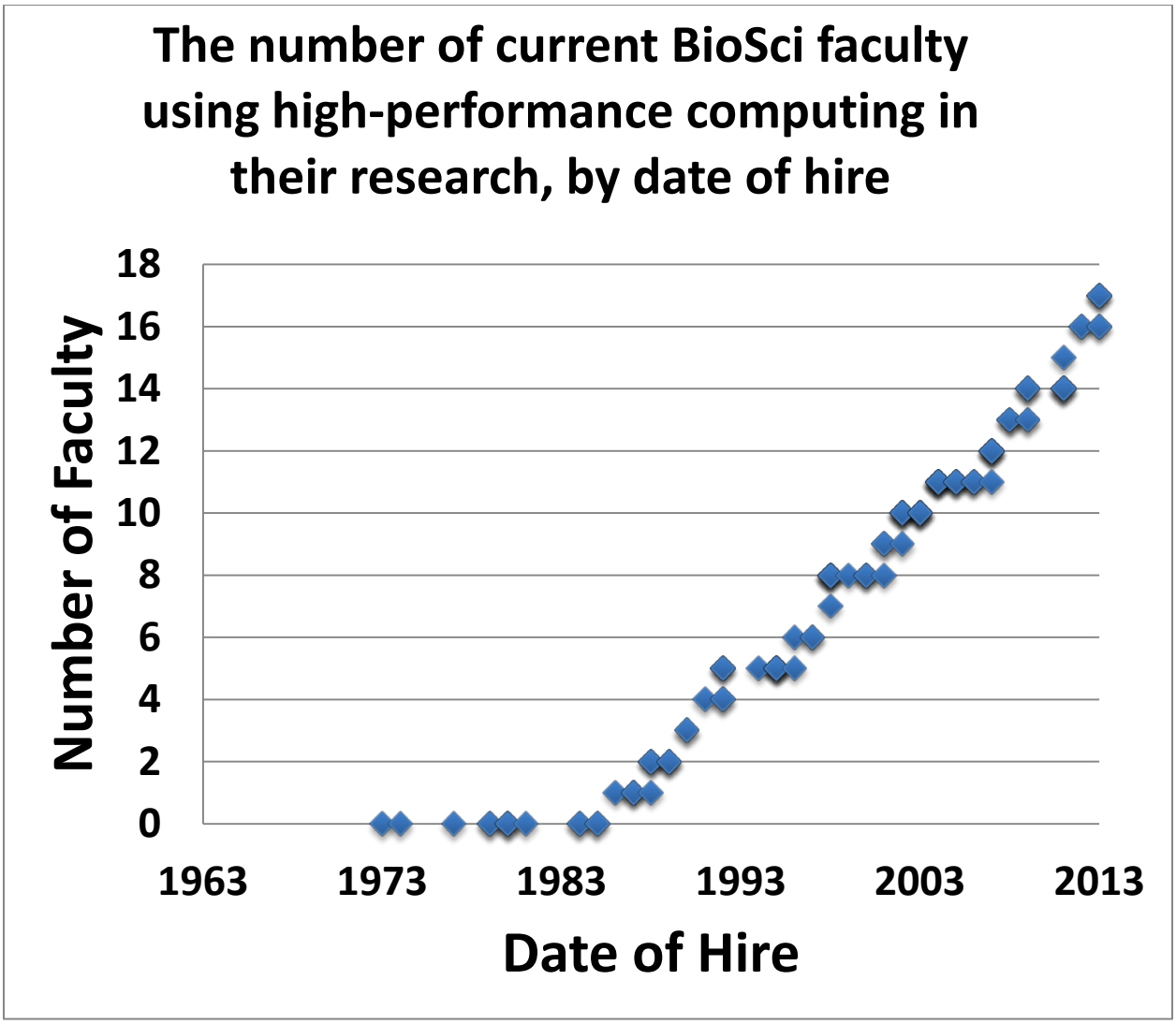
- Lisa Dahm, School of Medicine
- Jill Kay, Office of Research
- Allen Schiano, Office of Information Technology
- Tony Soeller, Office of Information Technology
- Dan Tsang, UCI Libraries

Appendix B: Growth of Faculty with Research Computing Needs

Robin Bush and Doug Tobias of the committee have attempted to accurately determine the numbers of faculty in their Departments and Schools that have significant needs for research computing support and services. Inspecting the faculty data in the School of Biological Sciences, Prof Bush was able to generate the following spreadsheet.

The specific metric used by Prof Bush was faculty involvement in the use of research computing clusters in their research efforts. During the period from 2000 to the present the number of FTE staff in both OIT and Biological Sciences assigned to research computing support (and research computing clusters) has been approximately constant. With all aspects of research computing being much broader than this specific service, the discrepancy between demands and supply for service is even greater.

The use of high performance research computing clusters in the biological sciences for genomic based research and other areas is well documented across academia and was repeated by faculty members contacted by this project. However, similar statements were made by all members of the committee. As such, the need for additional staffing as outlined in this report is strongly linked to the increases in faculty needs as both new faculty come to UCI as well as faculty research thrusts change.


The number of current BioSci faculty using high-performance computing in their research, by date of hire
Number of Faculty
18
16
14
12
10
8
6
4
2
0
1963
1973
1983
1993
2003
2013
Date of Hire

Appendix C: Faculty Survey Questions

Following are questions presented in the *Faculty Research Computing Assessment Survey* in their entirety. The survey was open to faculty from August 28, 2012 to February 28, 2013. One hundred and twenty five members of the faculty participated in the survey.

Survey Introduction

The Office of Information Technology, the Office of Research, and the UCI Libraries are interested in your responses to this survey. Your answers will assist a faculty-led workgroup to complete an assessment of UCI's Information Technology (IT) computing and networking infrastructure and services in support of your research needs. Your answers and comments will be analyzed by the workgroup and the results will be shared with the campus in the aggregate.

In the survey, we use the term "Research Computing" to broadly reflect the use of computing, storage, network and associated equipment and software in computation, or in data processing in all its forms, to advance research. Other terms such as "cyber-infrastructure", "e-Science" or "e-Research" could be applied to the use of information technology in research across disciplines. Research computing encompasses computers, research-oriented software, network connectivity, storage, laboratory and field analytical instruments, visualization and other related technologies, as well as data management. The application of research computing is applied to research performed in all academic disciplines on campus not just the traditional hard sciences.

Instructions: Please complete all the questions that are relevant to your research efforts. We realize there are many questions due to the breadth of the subject. Hence, any information that you can provide will be helpful even if you do not complete all questions.

Questions about you

1. What is your name?

2. What is your Department?

3. Your representation

Are you representing:

- o Yourself
- o Yourself and others (e.g., your research group, faculty in your department, or a research institute)
- o Other

If you are representing more than yourself, how many people?

If "Other," identify.

Your Research

4. Describe the areas of research where you use (or would use) research computing services and/or support.

Research Computing Hardware and Software

5. Describe the configuration of hardware (computers, storage, and instrumentation) currently being used to achieve your research outcomes?

6. What key research computing software do you currently use?

7. What research computing resources are used away from the UCI campus?

Research Data

8. Data management for research typically involves the collection, cleaning, merging, analysis, storage, and sharing of research data. In the following support areas for data management, indicate those areas where you could use more assistance and support from the campus:

- ☐ Developing a data management plan
- ☐ Documenting the research process
- ☐ Developing better ways to catalogue datasets (such as rapid identification of files, descriptive metadata, etc)
- ☐ Sharing data with project colleagues
- ☐ Storing (or backing up) data temporarily and reliably
- ☐ Long term archiving and preserving of data
- ☐ Securing sensitive data
- ☐ Depositing research data in a data repository for subsequent sharing with others
- ☐ Publishing data for sharing with researchers and/or the public
- ☐ Migrating legacy datasets to current technology
- ☐ Finding data produced by other researchers
- ☐ Ensuring professional credit/citation for research data that is created

9. What type of research data do you create, collect, or work with?

- ☐ Non-digital text
- ☐ Digital text or digital copies of texts
- ☐ Hand-written notes/sketches/figures
- ☐ Non-digital images
- ☐ Digital images or digital copies of images
- ☐ Artistic products

- ☐ Audio recordings
- ☐ Video recordings
- ☐ Survey data
- ☐ Socioeconomic data
- ☐ Health data
- ☐ Spatial or GIS data
- ☐ Time-series data
- ☐ Spreadsheets
- ☐ Digital databases
- ☐ Computer code
- ☐ Data from lab and field experiments
- ☐ Restricted human subject data with personal identifiers
- ☐ Biological/organic/inorganic samples or specimens
- ☐ Digital gene sequences or similar digital renditions of biological/organic/inorganic samples or specimens

Other?

10. The federal government has recently embarked on requiring sponsored research to include an explicit data management plan. To support faculty in these efforts the California Digital Library has rolled out the "DMPTool" [https://dmp.cdlib.org/].

If you have created a data management plan, how have you done so, and specifically how have you used tools such as DMPTool? Also, how have these tools met your needs?

11. What for-fee or licensed data do you need or currently purchase for your research?

12. If you use sensitive data what are the issues you are having with the data (e.g., security, obtaining, usage, contractual obligations on its use)?

<u>Your Unmet Research Computing Needs</u>

13. If you have been hesitant to use research computing resources, what is precluding you from using them?

14. What are the research computing hardware, software or instrumentation resources that could help you further your research goals?

15. What augmented/new infrastructure resources do you need to achieve your research goals? Topics may include, laboratory or field hardware, laboratory space, computer laboratory support hardware (e.g., air conditioning, electrical power, network connectivity), etc.

16. What additional support personnel for research computing tasks (e.g., programming, data analysis, data management, system administration) do you need to further your research goals?

17. What additional research computing training do you or your research group need?

18. What additional assistance do you need with purchasing research computing equipment or what recommendations do you need regarding the most optimum equipment to purchase for your research?

Specific Research Computing Service Improvements

In recent years, several suggestions and ideas have been proposed to improve the state of research computing at UCI. These follow similar assessments done at other universities (for example, see the UCSD's "Blueprint for the Digital University" http://research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf).

19. Based on your needs, what is your opinion regarding the following campus research computing improvements.

Strongly agree *Agree* *Disagree* *Strongly disagree* *No opinion*

a. UCI should provide more data center co-location space for computers and data systems for my research needs.
b. UCI should provide a centralized data storage system for my research data.
c. UCI should have more data management services.
d. UCI should improve its network to provide 10 Gbit, or better, performance to many sites on campus.
e. UCI has several "shared computing clusters" which are high performance tightly-networked clusters of computers that perform both serial and parallel computations. Examples are Green Planet, HPC/MPC, and BDUC. UCI should expand these computing services to researchers.
f. UCI should provide more support personnel for research computing.
g. UCI should provide more specialized research computing training.

20. Based on your choice in #19b, if you agree that UCI should provide a centralized data storage system, what should its qualities be?

Strongly agree *Agree* *Disagree* *Strongly disagree* *No opinion*

a. It should have very large capacity.
b. It should be a very high speed system.
c. It should have the ability to store sensitive data with advanced security requirements (FISMA, HIPPA, etc.).
d. It should have long term storage capabilities.

21. Based on your choice in #19c, if you agree that UCI should have more data management services, what should they provide?

Strongly agree *Agree* *Disagree* *Strongly disagree* *No opinion*

a. More data curation services
b. More data discovery and integration services
c. More data analysis and visualization services

22. Based on your choice in #19f, if you agree that UCI should provide more support personnel for research computing, what should be the services that they provide?

Strongly agree *Agree* *Disagree* *Strongly disagree* *No opinion*

a. More software development support
b. More code porting and optimization support
c. More web programming support
d. More database design and operation support
e. More statistical analysis support
f. More topic driven (GIS, mathematical software, CFD software, etc.) support
g. More network-based applications and services (streaming media, telematics, shared events or presentations)

23. Assuming the following services were going to be cost-shared between the campus, other faculty, and you, what specific percentage of your research area/lab/office for this service would you pay? (For example, if total costs for a service were $1 per user, what percentage of that $1 would you pay?)

0% *25%* *50%* *75%* *100%*

a. More data center co-location space for computers and data systems for my research needs
b. Centralized data storage system for my research data
c. More data management services
d. Provide 10 Gbit, or better, performance to many sites on campus
e. Expanded shared computing clusters
f. More support personnel for research computing
g. More specialized research computing training

<u>Other Suggestions</u>

24. What suggestions do you have for improving the services and support for research computing at UCI?

Further Outreach to You

Aside from this survey, we will be soliciting your opinions in other forms as well.

25. Would you be willing to talk about your answers to this survey with our faculty-led workgroup?

- o Yes
- o No

26. Would you be willing to attend an informal focus group with other faculty discussing the state of research computing at UCI?

- o Yes
- o No

27. Would you be interested in attending a "forum" where presentations on current research computing services and support by OIT, UCI Libraries, and other faculty support groups would be made? A faculty "townhall" type meeting on the state of research computing would also occur at the "forum."

- o Yes
- o No

Appendix D: Staff Analysis of Faculty Survey Data

Tony Soeller and Allen Schiano of OIT analyzed the responses from the survey and summarized them as follows for the committee. This information was present at the TownHall Meeting on January 7, 2013.

Question 4: Describe the areas of research where you use (or would use) research computing services and/or support.

Analysis : There is a wide variety of computing uses from this cohort – not limited to 'traditional computational' users

- *not many ways to group*
- *many said 'all' their work relies on IT services*
- *some made comments about support needs or who they work with*

Question 5. Describe the configuration of hardware (computers, storage, and instrumentation) currently being used to achieve your research outcomes?

Analysis: Very varied desktop environment – as many PCs, as Macs, as Linux systems

- *Very few off-campus computing resource users*
- *Some usage in Dropbox and Google for file sharing, backup, etc.*
- *About 10-20% percent use of campus clusters*

Question 6. What key research computing software do you currently use?

Analysis : Software needs quite varied

- *More usage of acquired software over created software (about 10% locally programmed)*
- *Acquired Software listing:*
    1) *Mathematical (Matlab, Mathematica, etc.) 28*
    2) *Statistical (SAS, SPSS, R, STATA) 23*
    3) *Common Software tools (Word etc) 18*
    4) *Locally programmed 13*
    5) *Chemistry 9*
    6) *Arts (Final Cut, Adobe) 9*
    7) *Engineering 7*
    8) *Open Source 7*
    9) *Biological 5*
    10) *Specialized to hardware 5*
    11) *Database 5*
    12) *GIS 5*
    13) *Visualization (IDL) 3*
    14) *Big Data 1*

Question 7. What research computing resources are used away from the UCI campus?

Analysis: Some confusion about the question

- *Earlier questions pointed to minimal computing resources off campus, now the majority site off campus resources*
- *About 20% report 'None'*
- *Some DropBox and Google cloud use*
- *Lots of VPN usage, most to Library resources*
- *A few remote computational services (SDSC, NERSC, Amazon Cloud)*

Question 8. Data management for research typically involves the collection, cleaning, merging, analysis, storage, and sharing of research data. In the following support areas for data management, indicate those areas where you could use more assistance and support from the campus:

Analysis:

- *Biggest results related to making data secure and reliable in the short and long run*
- *Sharing data also important*
- *Creating a data management plan important*
- *Moving old data to new formats is a problem, but not for all*

Question 9. What type of research data do you create, collect, or work with?

Answer: Answers: The most popular data types used are (in order of usage):

- *Images*
- *Text Files*
- *Programs*
- *Spreadsheets*

Question 11. What for-fee or licensed data do you need or currently purchase for your research?

Analysis:

- *Most people answered 'none'*
- *Many people did not understand the question. Some answered with commercial software titles*
- *No two answers the same*

Question 10. The federal government has recently embarked on requiring sponsored research to include an explicit data management plan. To support faculty in these efforts the California Digital Library has rolled out the "DMPTool" [https://dmp.cdlib.org/].

If you have created a data management plan, how have you done so, and specifically how have you used tools such as DMPTool? Also, how have these tools met your needs?

Analysis:

- *Only about half of the respondents answered the question*
- *A large proportion of the respondents had not created a plan or said 'Not Applicable'*
- *Some have used DMPTool successfully*
- *Many want to know about DMPTool*
- *Lots of confusion about the need for plan*
-

Question 12. If you use sensitive data what are the issues you are having with the data (e.g., security, obtaining, usage, contractual obligations on its use)?

Answer:

- *About 20% of respondents had meaningful answers*
- *Most have dealt with issues already*
- *Some problem with storage needs*

Question 19. Based on your needs, what is your opinion regarding the following campus research computing improvements.

a. UCI should provide more data center co-location space for computers and data systems for my research needs.

#	%	Answer Option
12/99	12%	Strongly agree

22/99	22%	Agree
8/99	8%	Disagree
2/99	2%	Strongly disagree
55/99	56%	No opinion
99/110	90%	# of responses to this question

b. UCI should provide a centralized data storage system for my research data.

#	%	Answer Option
30/101	30%	Strongly agree
34/101	34%	Agree
8/101	8%	Disagree
3/101	3%	Strongly disagree
26/101	26%	No opinion
101/110	92%	# of responses to this question

c. UCI should have more data management services.

#	%	Answer Option
9/98	9%	Strongly agree
38/98	39%	Agree
4/98	4%	Disagree
1/98	1%	Strongly disagree
46/98	47%	No opinion
98/110	89%	# of responses to this question

d. UCI should improve its network to provide 10 Gbit, or better, performance to many sites on campus.

#	%	Answer Option
26/101	26%	Strongly agree
42/101	42%	Agree
4/101	4%	Disagree
1/101	1%	Strongly disagree
28/101	28%	No opinion
101/110	92%	# of responses to this question

e. UCI has several "shared computing clusters" which are high performance tightly-networked clusters of computers that perform both serial and parallel computations. Examples are Green Planet, HPC/MPC, and BDUC. UCI should expand these computing services to researchers.

#	%	Answer Option

22/101	22%	Strongly agree
23/101	23%	Agree
7/101	7%	Disagree
0/101	0%	Strongly disagree
49/101	49%	No opinion
101/110	92%	# of responses to this question

f. UCI should provide more support personnel for research computing.

#	%	Answer Option
30/102	29%	Strongly agree
30/102	29%	Agree
6/102	6%	Disagree
1/102	1%	Strongly disagree
35/102	34%	No opinion
102/110	93%	# of responses to this question

g. UCI should provide more specialized research computing training.

#	%	Answer Option
19/102	19%	Strongly agree
35/102	34%	Agree
5/102	5%	Disagree
1/102	1%	Strongly disagree
42/102	41%	No opinion
102/110	93%	# of responses to this question

Analysis:

- *The order of agreement (Strongly Agree+Agree): 68% Network, 64% Storage, Personnel 58%, Training 53%, Data Management 48%, Clusters 45%, 34% Data Center*
- *Strength of Agreement (Strongly/Agree): 100% Personnel, 96% Clusters, 88% Storage, 62% Network, 56% Training, 55% Data Center, 23% Data Management*
- *Network: most use it and want to improve (#1) it but not in need of critical improvements (#5)*
- *Storage: Many need improvements (#2) and it's critical for some (#3)*
- *Personnel: Many need improvements (#3) and it's critical for some (#1)*
- *Training: Some see improvements needed (#5) and it's critical for some (#6)*
- *Data Management: Some see improvements needed (#6) but few think it's critical (#8)*
- *Clusters: A minority see a need for improvement (#7) but they think it's critically needed (#2)*
- *Data Center: A minority see a need for improvement (#8) but few think it's critical (#7)*

Question 20. Based on your choice in #19b, if you agree that UCI should provide a centralized data storage system, what should its qualities be?

It should have very large capacity.

#	%	Answer Option
35/81	43%	Strongly agree
24/81	30%	Agree
0/81	0%	Disagree
0/81	0%	Strongly disagree
22/81	27%	No opinion
81/110	74%	# of responses to this question

It should be a very high speed system.

#	%	Answer Option
20/80	25%	Strongly agree
34/80	43%	Agree
3/80	4%	Disagree
0/80	0%	Strongly disagree
23/80	29%	No opinion
80/110	73%	# of responses to this question

It should have the ability to store sensitive data with advanced security requirements (FISMA, HIPPA, etc.).

#	%	Answer Option
17/79	22%	Strongly agree
16/79	20%	Agree
4/79	5%	Disagree
2/79	3%	Strongly disagree
40/79	51%	No opinion
79/110	72%	# of responses to this question

It should have long term storage capabilities.

#	%	Answer Option
29/81	36%	Strongly agree
33/81	41%	Agree
2/81	2%	Disagree
1/81	1%	Strongly disagree
16/81	20%	No opinion
81/110	74%	# of responses to this question

Analysis:

- *The order of agreement (Strongly Agree+Agree): 77% Long Term, 73% Large Capacity, 68% Fast, 42% Sensitive*
- *Strength of Agreement (Strongly/Agree): 143% Large Capacity, 110% Sensitive, 88% Long Term, 58% Fast*
- *Long Term: Most want it to improve (#1) and it is important to many (#3)*
- *Large Capacity: Most want it to improve (#2) and most think it's critical (#1)*
- *Fast: Most want it to improve (#3) and many think it is critical (#4)*
- *Sensitive: Minority want it to improve (#4) but most of those think it's critical (#2)*
- *A quarter of the survey takers did not respond to this question*

Question 21. Based on your choice in #19c, if you agree that UCI should have more data management services, what should they provide?

More data curation services

#	%	Answer Option
7/69	10%	Strongly agree
24/69	35%	Agree
2/69	3%	Disagree
0/69	0%	Strongly disagree
36/69	52%	No opinion
69/110	63%	*# of responses to this question*

More data discovery and integration services

#	%	Answer Option
4/69	6%	Strongly agree
21/69	30%	Agree
4/69	6%	Disagree
0/69	0%	Strongly disagree
40/69	58%	No opinion
69/110	63%	*# of responses to this question*

More data analysis and visualization services

#	%	Answer Option
12/72	17%	Strongly agree
22/72	31%	Agree
2/72	3%	Disagree

0/72	0%	Strongly disagree
36/72	50%	No opinion
72/110	65%	*# of responses to this question*

Analysis

- *The order of agreement (Strongly Agree+Agree): 58% Visualization, 45% Data Curation, 42% Data Discovery*
- *Strength of Agreement (Strongly/Agree): 55% Visualization, 28% Data Curation, 17% Data Discovery*
- *Visualization: Most want it to improve (#1) and it is important to a majority of them (#1)*
- *Data Curation: A minority want to see it improve (#2) but it's critical only to a few (#2)*
- *Data Discovery: A minority want to see it improve (#3) but it's critical only to a few (#3)*
- *A third of the survey takers did not respond to this question*

Question 22. Based on your choice in #19f, if you agree that UCI should provide more support personnel for research computing, what should be the services that they provide?

More software development support

#	%	Answer Option
16/77	21%	Strongly agree
22/77	29%	Agree
5/77	6%	Disagree
1/77	1%	Strongly disagree
33/77	43%	No opinion
77/110	70%	*# of responses to this question*

More code porting and optimization support

#	%	Answer Option
9/72	13%	Strongly agree
14/72	19%	Agree
5/72	7%	Disagree
1/72	1%	Strongly disagree
43/72	60%	No opinion
72/110	65%	*# of responses to this question*

More web programming support

#	%	Answer Option

12/76	16%	Strongly agree
28/76	37%	Agree
3/76	4%	Disagree
0/76	0%	Strongly disagree
33/76	43%	No opinion
76/110	69%	*# of responses to this question*

More database design and operation support

#	%	Answer Option
6/78	8%	Strongly agree
29/78	37%	Agree
5/78	6%	Disagree
0/78	0%	Strongly disagree
38/78	49%	No opinion
78/110	71%	*# of responses to this question*

More statistical analysis support

#	%	Answer Option
10/74	14%	Strongly agree
19/74	26%	Agree
4/74	5%	Disagree
4/74	5%	Strongly disagree
37/74	50%	No opinion
74/110	67%	*# of responses to this question*

More topic driven (GIS, mathematical software, CFD software, etc.) support

#	%	Answer Option
13/73	18%	Strongly agree
22/73	30%	Agree
1/73	1%	Disagree
2/73	3%	Strongly disagree
35/73	48%	No opinion
73/110	66%	*# of responses to this question*

More network-based applications and services (streaming media, telematics, shared events or presentations)

#	%	Answer Option
11/75	15%	Strongly agree
25/75	33%	Agree

7/75	9%	Disagree
1/75	1%	Strongly disagree
31/75	41%	No opinion
75/110	68%	*# of responses to this question*

Analysis

- *The order of agreement (Strongly Agree+Agree): 53% Web programming, 50% Programming, 48% Topic Driven, 48% Network, 45% Database, 40% Statistical, 32% Optimization*
- *Strength of Agreement (Strongly/Agree): 73% Programming, 68% Optimization, 60% Topic Driven, 54% Data Statistical, 45% Network, 43% web programming, 22% Database*
- *Web programming: Important to most (#1) but critically important to a minority (#6)*
- *Programming: Important to many (#2) and critically important to the majority (#1)*
- *Topic Driven: Important to many (#3) and critically important to the majority (#3)*
- *Network: Important to many (#4) but critically important to a minority (#5)*
- *Database: Important to some (#5) but critically important to only a few (#7)*
- *Statistical: Important to some (#6) but critically important to a minority (#6)*
- *Optimization: Important to only a minority (#7) but critically important to that few (#2)*
- *A third of the survey takers did not respond to this question*

Question 23. Assuming the following services were going to be cost-shared between the campus, other faculty, and you, what specific percentage of your research area/lab/office for this service would you pay? (For example, if total costs for a service were $1 per user, what percentage of that $1 would you pay?)

More data center co-location space for computers and data systems for my research needs

#	%	Answer Option
47/71	66%	0%
16/71	23%	25%
5/71	7%	50%
2/71	3%	75%
1/71	1%	100%
71/110	65%	*# of responses to this question*

Centralized data storage system for my research data

#	%	Answer Option
27/76	36%	0%
23/76	30%	25%

15/76	20%	50%
5/76	7%	75%
6/76	8%	100%
76/110	69%	*# of responses to this question*

More data management services

#	%	Answer Option
50/73	68%	0%
11/73	15%	25%
10/73	14%	50%
1/73	1%	75%
1/73	1%	100%
73/110	66%	*# of responses to this question*

Provide 10 Gbit, or better, performance to many sites on campus

#	%	Answer Option
46/76	61%	0%
18/76	24%	25%
6/76	8%	50%
5/76	7%	75%
1/76	1%	100%
76/110	69%	*# of responses to this question*

Expanded shared computing clusters

#	%	Answer Option
49/73	67%	0%
13/73	18%	25%
7/73	10%	50%
2/73	3%	75%
2/73	3%	100%
73/110	66%	*# of responses to this question*

More support personnel for research computing

#	%	Answer Option
39/77	51%	0%
20/77	26%	25%
11/77	14%	50%
3/77	4%	75%
4/77	5%	100%

77/110	70%	*# of responses to this question*

More specialized research computing training

#	%	Answer Option
47/73	64%	0%
15/73	21%	25%
5/73	7%	50%
3/73	4%	75%
3/73	4%	100%
73/110	66%	*# of responses to this question*

Analysis

- *Order of Whose Would Pay (25 and 50% recharge): 50% Storage, 40% Personnel, 32% Network, 30% Data Center, 29% Data Management, 28% Clusters, 28% Training*
- *Should be Free (0% recharge): 68% Data Management, 67% Clusters, 66% Data Center, 64% Training, 61% Network, 51% Personnel, 36% Storage*
- *Most will pay for up to 50% of their storage costs*
- *A large minority would pay up to 50% for personnel costs*
- *With the exception of storage costs, most think all these costs should be mostly borne by the campus*

Question 24. What suggestions do you have for improving the services and support for research computing at UCI?

Compilation of Comments:

- *Build a centralized, efficient and well managed campus storage system (5 responses)*
- *There is no support in Arts and Humanities*
- *Why recharge at all/should be paid with overhead (3 responses)*
- *We need bioinformatics support (3 responses)*
- *None/I'm happy (4 responses)*
- *Need 'hired gun' programmers (2 responses)*
- *Need for support for local clusters*
- *We need more support staff (2 responses)*
- *We need more campus wide site licenses for software (2 responses)*
- *I have a non research computing need*
- *Need more telepresence/telematics support*
- *Use/pay for more cloud services (2 responses)*
- *Make more databases available*

- *We need more centralized services*
- *We need fewer centralized services*
- *Local school support groups are great/we need to support them more (3 responses)*
- *Higher network performance*

Appendix E: Comments from the FASRC Townhall Meeting

Town Hall Meeting, January 7, 2013

A 'townhall' type meeting was held on this day in the CalIT2 Auditorium. E-mail invitations went to all faculty as well as personal requests to the campus School IT Directors to encourage attendance. There were 67 attendees who signed up for the event to accommodate a lunchtime meal. An unknown number of attendees came to either the morning or afternoon sessions.

The morning portion of the event consisted of a presentation by UCI Library staff on the use of data management tools for research, specifically the DMPTool which is a UC wide collaboration with the California Digital Library consortium.

In the afternoon, presentation of the faculty survey results was followed by a roundtable discussion chaired by the Directors of the Sponsoring organizations, namely Dana Roode of OIT, Lorelie Tanji of the Libraries, and John Hemminger of the Office of Research. An audio recording of the survey presentation and the Townhall discussion was recorded with 'UCIReplay'. The recordings can be accessed at

http://replay.uci.edu/uci-only/winter2013/FASRC_Town_Hall_-_Flash_%28Large%29_-_20130107_12.29.08PM.html

for the morning session and at

http://replay.uci.edu/uci-only/winter2013/FASRC_Town_Hall_%28afternoon%29_-_Flash_%28Large%29_-_20130107_01.50.52PM.html

for the afternoon session.