Faculty Assessment of the State of Research Computing Town Hall Meeting

- Agenda
  - 11:00  Meeting Overview, Timeline and Logistics
  - 11:05  Review of FASRC Project and Introduction of FACRC Committee
  - 11:15  Data Management Services Presentation by UCI Libraries
  - 12:00  Light Lunch in CalIT2 Lobby Area
  - 12:30  Preliminary Review of FASRC Faculty Online Survey Results and Discussion with FASRC Committee
  - 1:00   Introductory Remarks by Project Sponsors
  - 1:10   Discussion with Audience Members, Project Sponsors and FASRC Faculty Members
  - 2:00   Conclusion
Faculty Assessment of the State of Research Computing

- Faculty Committee
  - Robin Bush, Biology
  - Donald Dabdub, Engineering
  - David Goldberg, Humanities
  - Antoinette LaFarge, Arts
  - Tony Long, Biology
  - Christa Lopes, ICS
  - James Meeker, Social Ecology
  - Maria Pantelia, Classics
  - Suzanne Sandmeyer, Medicine
  - Doug Tobias, Physical Sciences
  - Stephen White, Medicine
  - Ted Wright, Social Sciences

- Project Tasks
  - Committee Meetings
  - Faculty Survey
  - Focus Groups
  - Town Hall
  - Report to Sponsors
Analysis of FASRC Faculty Survey

- Respondents’ Representation
  - 121 respondents
  - 63% represent themselves only
  - 37% represent 302 people (self and others in research group, and other affiliated faculty)
    - Average is 7
    - Range is 2 to 15
<table>
<thead>
<tr>
<th>Department</th>
<th>Responses</th>
<th>School/Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemistry</td>
<td>18</td>
<td>PS</td>
</tr>
<tr>
<td>Ecology &amp; Evolutionary Biology</td>
<td>9</td>
<td>Bio</td>
</tr>
<tr>
<td>Cognitive Sciences</td>
<td>7</td>
<td>SS</td>
</tr>
<tr>
<td>Drama</td>
<td>7</td>
<td>Arts</td>
</tr>
<tr>
<td>Music</td>
<td>5</td>
<td>Arts</td>
</tr>
<tr>
<td>Civil &amp; Environmental Engineering</td>
<td>4</td>
<td>Eng</td>
</tr>
<tr>
<td>Informatics</td>
<td>4</td>
<td>ICS</td>
</tr>
<tr>
<td>Developmental and Cell Biology</td>
<td>3</td>
<td>Bio</td>
</tr>
<tr>
<td>Education</td>
<td>3</td>
<td>Edu</td>
</tr>
<tr>
<td>Merage School of Business</td>
<td>3</td>
<td>MSOB</td>
</tr>
<tr>
<td>Pharmaceutical Sciences and Chemistry</td>
<td>3</td>
<td>Program</td>
</tr>
<tr>
<td>Physics and Astronomy</td>
<td>3</td>
<td>PS</td>
</tr>
<tr>
<td>Chemical Engineering and Materials Science</td>
<td>3</td>
<td>Eng</td>
</tr>
<tr>
<td>EECS</td>
<td>3</td>
<td>Eng</td>
</tr>
<tr>
<td>Anthropology</td>
<td>2</td>
<td>SS</td>
</tr>
<tr>
<td>English</td>
<td>2</td>
<td>Hum</td>
</tr>
<tr>
<td>Logic and Philosophy of Science</td>
<td>2</td>
<td>SS</td>
</tr>
<tr>
<td>Mathematics</td>
<td>2</td>
<td>PS</td>
</tr>
<tr>
<td>Molecular Biology &amp; Biochemistry</td>
<td>2</td>
<td>Bio</td>
</tr>
<tr>
<td>Physiology and Biophysics</td>
<td>2</td>
<td>Bio</td>
</tr>
<tr>
<td>Planning, Policy and Design</td>
<td>2</td>
<td>SE</td>
</tr>
<tr>
<td>Political Science</td>
<td>2</td>
<td>SS</td>
</tr>
<tr>
<td>Program in Public Health</td>
<td>2</td>
<td>Program</td>
</tr>
<tr>
<td>Sociology</td>
<td>2</td>
<td>SS</td>
</tr>
</tbody>
</table>
# School / Program Representation

<table>
<thead>
<tr>
<th>Department</th>
<th>Responses</th>
<th>School/Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine</td>
<td>2</td>
<td>SOM</td>
</tr>
<tr>
<td>AfAm</td>
<td>1</td>
<td>Hum</td>
</tr>
<tr>
<td>Biological Chemistry</td>
<td>1</td>
<td>SOM</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td>1</td>
<td>Eng</td>
</tr>
<tr>
<td>BME</td>
<td>1</td>
<td>Eng</td>
</tr>
<tr>
<td>Criminology, Law and Society</td>
<td>1</td>
<td>SE</td>
</tr>
<tr>
<td>Dance</td>
<td>1</td>
<td>Arts</td>
</tr>
<tr>
<td>Earth System Science</td>
<td>1</td>
<td>PS</td>
</tr>
<tr>
<td>East Asian</td>
<td>1</td>
<td>Hum</td>
</tr>
<tr>
<td>Economics</td>
<td>1</td>
<td>SS</td>
</tr>
<tr>
<td>Film &amp; Media Studies, English, Informatics</td>
<td>1</td>
<td>Art</td>
</tr>
<tr>
<td>History</td>
<td>1</td>
<td>Hum</td>
</tr>
<tr>
<td>MAE, ChEMS</td>
<td>1</td>
<td>Eng</td>
</tr>
<tr>
<td>Merage, Marketing</td>
<td>1</td>
<td>MSOB</td>
</tr>
<tr>
<td>Microbiology and Molecular Genetics</td>
<td>1</td>
<td>SOM</td>
</tr>
<tr>
<td>Neurobiology and Behavior</td>
<td>1</td>
<td>Bio</td>
</tr>
<tr>
<td>Nursing Science</td>
<td>1</td>
<td>Program</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>1</td>
<td>Program</td>
</tr>
<tr>
<td>Political Science</td>
<td>1</td>
<td>SS</td>
</tr>
<tr>
<td>Psychology &amp; Social Behavior</td>
<td>1</td>
<td>SE</td>
</tr>
<tr>
<td>Radiological Sciences</td>
<td>1</td>
<td>SOM</td>
</tr>
<tr>
<td>School of Law</td>
<td>1</td>
<td>Law</td>
</tr>
<tr>
<td>Statistics</td>
<td>1</td>
<td>ICS</td>
</tr>
</tbody>
</table>
FASRC Survey Topics

- Hardware Used
- Software Used
- Off-campus IT Resources
- Data
  - Data Management Support Needed
  - Types of Data Used
  - Data Management Plan
  - Purchased Data
  - Issues with Sensitive Data
- Research Computing Resources
  - Obstacles to Using RC Resources
  - What RC Resources Would Help
  - Augmented / New Infrastructure Resources
  - Additional RC Support Personnel
  - Additional RC Training
  - RC Equipment purchasing assistance
Opinions
- RC Improvements
  - Data Center Co-location space
  - Centralized Data Storage System
  - Data Management Services
  - Improve Network
  - Expand computing clusters
  - Provide more support personnel
  - Specialized RC Training

Specific Storage Needs
- Very large capacity
- Very high speed
- Ability to store sensitive data with advanced security requirements
- Long-term storage capabilities

Data Management
- More Data Curation services
- More Data Discovery and integration services
- More data analysis and visualization services
Opinions (cont.)

- Support Personnel
  - More software development support
  - More code porting and optimization support
  - More web programming support
  - More database design and operation support
  - More statistical analysis support
  - More topic-driven support
  - More network-based applications and services

Percentage you would pay for service improvements

- More data center co-location space
- Centralized data storage system
- More Data Management services
- Provide 10 Gbit performance to many sites on campus
- Expanded shared-computing clusters
- More support personnel for research computing
Our comments regarding responses

- A wide variety of computing uses from this cohort – not limited to ‘traditional computational’ users
- not many ways to group respondents
- many said ‘all’ their work relies on IT services
- some made comments about support needs or who they work with
Our comments regarding responses

- Highly 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
Our comments regarding responses

- Software needs are quite varied
- More usage of acquired software over created software (about 10% locally programmed)

Acquired software listing

- Mathematical (Matlab, Mathematica, etc.) 29
- Statistical (SAS, SPSS, R, STATA) 26
- Common Software tools (Word etc) 19
- Locally programmed 13
- Chemistry 11
- Arts (Final Cut, Adobe) 10
- Engineering 7
- Open Source 7
- Biological 6
- Specialized to hardware 6
- Database 5
- GIS 5
- Visualization (IDL) 3
- Big Data 1
Survey Question Example: Data Management

Question: 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

- Our comments regarding responses
  - 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, but not very critical
Our comments regarding responses

- 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 a plan
Our comments regarding responses

- Some researchers appear not to be aware of the research computing resources that are available to them.
  - Or the impression that not many resources are available.
  - Or they do not know who to contact for assistance.
  - Or they do not know enough about computing to ask what would help.
  - Some do not know how to access resources.
  - Or the “language” of research computing.
  - Or lack of documentation and access.
Our comments regarding responses

- Again, some researchers apparently were not aware of the resources available on campus.
- Many diverse interests. Some recurring themes:
  - Storage and backup systems; archiving bulky data; sharing of data; networked data storage
  - Many specific needs
  - More clusters/nodes
  - Making Webfiles easier; storage up to 20 Gb+
What Additional Support Personnel

- Our comments regarding responses
  - System administrators is a common need.
  - Also database managers to design and manage databases, and find databases.
  - Programmers for programs as well as scripts.
  - Some specifics:
    - Information visualization specialists
    - Lab maintenance and support staff with knowledge of audio and music technologies and systems
    - Help with genomic analyses / bioinformatics
    - Programmers for text analysis and acquiring data from online.
Opinions Regarding Campus Research Computing Improvements

- Our comments regarding responses
  - Most Widespread Need for New Storage Services, followed by Network Improvements
  - Strongest Need for Additional Personnel, followed by Improved Cluster Support
  - Network: All use it and want to improve it but not in critical need for need for improvement
  - Storage: Many need improvements and it’s critical for some
  - Personnel: Many see need for improvements and it’s critical for some
  - Data Management: Some see need for improvements but few think it’s critical at this time
  - Clusters: A minority see a need for improvement but they think it’s critically needed for them
Our comments regarding responses

- Most widespread need is for long term storage followed by large capacity
- Of those that see this need as critical, most see large capacity followed by the ability to deal with sensitive data as important
- Most think it has to be fast
Our comments regarding responses

- Most people wanted improved Data Visualization services followed by Data Curation.
- Of those interested in Visualization improvements were seen as important.
- Of those interested in Data Curation improvements it is not currently critical but will likely be in the future.
- Only a minority want to see Data Discovery improved and to those it’s not critical.
Our comments regarding responses

- The most widespread need was for programming mostly for web and then database programming
- Web programming critical to some
- Programming on general was important to most and critical to all
What Percentage You Would Pay for Service Improvements

- Our comments regarding responses
  - A majority would pay for *some* (up to 50%) of the costs for storage improvements
  - A large minority would pay for increased support personnel costs
  - A large majority think none of the suggested improvements should cost them (Data Management, 67% Clusters, 66% Data Center, 64% Training, 61% Network)
Suggestions for Improving Services and Support for Research Computing

- Summarized faculty opinions
  - Build a centralized, efficient and well managed campus storage system
  - “Why recharge at all?/Should be paid with overhead”
  - We need bioinformatics support
  - “None/I’m happy“
  - Need ‘hired gun’ programmers
  - Need for support for compute clusters
  - We need more support staff
  - We need more campus wide software site licenses
  - Higher network performance
Go to http://sites.uci.edu/fasrc

Content

- Project Mission & Plan
- List of Committee Members
- Meeting Notes
  - Committee Meeting
  - Focus Group Meetings
- Files
  - Preliminary Analysis of Survey Responses
  - This PowerPoint
  - (Upon completion, the formal FASRC Report to Sponsors)