

Proceedings of the First Meeting on Computing Resources for Biomedical Research

April 30-May 1, 2001

An ad-hoc meeting on Computing Resources for Biomedical Research was held in Philadelphia at the Fox Chase Cancer Center on April 30 and May 1, 2001. The stated goals of the meeting were as follows:

- Establish personal contacts and bring together those responsible for informatics and other research computing activities at biomedical research institutions
- Identify and document common problems and interests
- Seek opportunities for partnership and/or consortium activities
- Identify common issues that should be brought to attention of home institutions, government and other agencies.

Representatives from the 10 exempt Cancer Centers, plus a number of AIRI (www.airi.org) members were present. A list of attendees is attached.

During the meeting, it was decided that group would hold future meetings under the name Biomedical Research Information Technology Exchange, or BRITE.

Genesis of the Meeting

The meeting came about as the result of a survey commissioned in early 2000 by Dr. Robert Robbins, Chief Information Officer of the Fred Hutchinson Cancer Research Center. The survey examined how major research institutions support the Information Technology (IT) needs of their research groups. Ten organizations involved in life sciences research on a scale similar to Fred Hutchinson were surveyed. The survey's goal was to identify the state of IT support for scientific research – including common trends, and strategic directions for the future – to guide cost-effective development of high level technical and infrastructure support. The focus of the survey was on support for research informatics rather than issues surrounding administrative computing (i.e. payroll, human resources) or patient management. Issues not related to research IT were specifically *not included* unless they had bearing on the research informatics infrastructure.

Results of the Fred Hutchinson Survey

The survey revealed substantial differences in the organization of groups providing research support. At some institutions the research computing and research informatics groups have responsibility for bioinformatics, general scientific computing, and desktop support. Others include components of server support, network infrastructure, and in some cases, even phone services. There was substantial variability in the level of funding and available manpower among institutions. All the survey participants indicated that the level of resources available to them do not support current or expected demands in this area.

Dr. Robbins then discussed that there was a correlation between the overall percent of institutional R01 grant funding and the level of decentralization of control over information technology and corresponding support services. Conference participants agreed that this result has implications for funding mechanisms from the various national agencies supporting research, particularly if these agencies seek to insure adequate access to state-of-the-art technologies for supported research activities. The group felt that this issue should be

Proceedings of the First Meeting on Computing Resources for Biomedical Research

examined further by these agencies, and that this recommendation should be conveyed to the Center Directors.

Review of Stanford meeting of the Microarray Gene Expression Database Group

An overview of the Third Annual Meeting on Microarray Gene Expression Database Group (MGED, <http://www.mged.org>), held in March 2001 at Stanford, was presented from slides on the web site, with additions from Frank Manion and Dr. Michael Ochs from Fox Chase Cancer Center. The following issues resulting from that meeting were discussed:

1. The MGED group is developing standards for the various stages of experimental design, array data collection, within and across experiment data normalization, data distribution, and data analysis.
2. At the Stanford meeting, Dr. Roger Bumgarner from the University of Washington presented data demonstrating non-linear fluorescence response from cy3 and cy5 dyes on various spotted arrays, and discussed the resulting serious implications for experimental design, data normalization efforts, and data analysis.
3. The MGED group is planning on making recommendations in the near future to the funding agencies and major journals on requirements for data submission to approved international repositories.
4. Although consensus has not yet been reached, the MGED group appears to be converging toward favoring a recommendation for the storage of the raw image data from the confocal array scanners in the national and international repositories. The rationale is to have sufficient data to allow future image processing and expression data mining technologies to process past experimental data for knowledge discovery. This in-turn has important implications for the management of such datasets at our institutions, and for the size and cost of the systems needed for supporting genomics research.
5. There are a substantial number of technical and scientific issues as well as international data standards groups which the MGED group is trying to integrate with, including the Gene Ontology Consortium.
6. From a technical perspective, anyone building in-house data systems for the support of microarray work will need to include the information specified in the Minimal Information About Microarray Experiments (MIAME) in their data model and functional systems. The system will also need to be able to deal effectively with, at a minimum, exporting all data in a form that can be converted to the MicroArray Markup Language (MAML) format.

All those doing microarray work of any sort, both scientific and technical, will be well served by keeping up with the progress of the MGED. (Note: the MGED meeting was also summarized in the April 21 copy of *Nature*, on page 851).

American Association of Cancer Institutes (AACI)

There was discussion about this group and its relationship to the current effort by the AACI to examine informatics at the cancer centers. Dr. Joyce Niland from City of Hope Cancer Center presented an overview of the AACI initiative. There are four AACI subcommittees focused on 1) data representation and interchange standards; 2) clinical informatics infrastructure; 3) bioinformatics infrastructure; and 4) consumer health information. The subcommittees are

Proceedings of the First Meeting on Computing Resources for Biomedical Research

developing a report due mid-Summer 2001 to the NCI. It is unknown at this time whether the AACI subcommittees will continue to meet after that time. The group felt that the AACI effort is mostly focused on issues directly involved with clinical research and clinical trials support, although there are also subcommittees on infrastructure. The group believes that in addition to the AACI initiative, there continues to be a strong need for an ongoing forum to discuss the scientific, technical, and operational issues – as well as the resulting financial and organizational implications – in the area of research informatics within the Cancer Center and the closely allied research community setting. It was felt that this group should work with the AACI where possible, particularly in those areas where there is substantial overlap, and in bringing critical recommendations of common interest to the attention of the NCI.

National Cancer Institute Center for Bioinformatics

Dr. Ken Buetow, from the National Cancer Institute, discussed the status of the NCI Center for Bioinformatics. Funding for various informatics programs, including the NCI Director's Challenge grants and the Mouse Models of Human Cancer grants, were discussed. Initiatives from the Center for Bioinformatics, including the development of common ontologies and vocabulary standards, development of web consumable forms for support of clinical trials, clinical trials screening, monitoring and outcomes analysis were discussed. The Center is recommending that all software and systems developed for bioinformatics and clinical research be built with Open Source principles and licensing.

Issues Surrounding Microsoft Software Licensing

Changes in licensing policy are resulting in a large and increasing recurring cost burden for a number of institutions. Several institutions have had their academic discounts for Microsoft products discontinued within the past year. The loss of this discount will likely result in new recurring operating costs in the six-figure range at the institutions already affected. We believe there are indications that others may lose their discounts in the future. Microsoft appears to be moving in the direction of being more restrictive in their definitions of an academic institution, increasingly relying on strict interpretation of a "four-year degree granting" clause.

Substantial discussion ensued about the various possibilities for reducing our exposure. The current technical status of the Sun StarOffice and the public OpenOffice projects were discussed. It was felt that there are currently problems with using these tools in a production environment. The burden of tracking current and past MS Office document formats, to retain the ability to exchange such documents for collaboration, was viewed as problematic. A further critical shortcoming of these tools specific to a research environment is the lack of significant integration with common reference management products such as EndNote from ISI. Additionally, there is little or no support for the creation of portable document format (PDF) that is increasingly required for electronic grant submission, although there are probably workarounds for this issue.

It was felt that we all should carefully review MS license agreements to see what tactics could be taken in developing a group purchasing or service consortium to achieve volume or other discounts. Several participants agreed to do this and report back to the group. A research purchasing consortium with enough leverage might be able to influence software vendors to invest effort to develop the important integrated components, such as the reference manager, for other non-proprietary office suites. It was felt that the possibility of assistance in obtaining GAO pricing in conjunction with the funding agencies should be explored through our

Proceedings of the First Meeting on Computing Resources for Biomedical Research

administrations. The group believes that the entire issue should be brought to the attention of the funding agencies and the government.

Other Data Resources and Software

Following along the lines of the Microsoft discussion, the group considered other ways in which partnering would enable institutions to access an expanded set of genomics data from groups such as Celera. There was a sense among the group that a number of institutions would be very interested in obtaining access to the Celera SNP and mouse genome data were it made available in a reduced cost fashion. New York State has apparently worked out a group purchasing arrangement for state funded researchers. Jerry Latter from CSHL agreed to gather information on this arrangement and disseminate it to the group. The discussion then turned to other genomic data, software and hardware resources that it might be possible to leverage via group consortium agreements. Relational database licenses were identified as one of these. Although most of the institutions represented at the meeting are running a variety of commercial database products, it was felt that this might be a fruitful area for further investigation. Frank Manion of Fox Chase agreed to begin investigations in this area and report back to the group.

The discussion concluded with strategies that may be effective at enticing vendors to sell or resell to us through a consortium or group purchasing agreement at heavily discounted prices. The group felt that two tactics might be effective and should be considered. The first is structuring in some hook or gimmick to allow the vendor to draw a firewall around the special pricing, such as may have been done in the New York agreement with Celera. The second factor is to structure routine provision of support through the consortium, to the participant institutions, in such a manner as to allow the vendor to pass on their savings from a single point of contact for a large purchase.

Informatics support on grant submissions

There was substantial discussion on involving informatics personnel in the grant review process, both within the research centers and funding agencies. Many in the group felt that grant applications routinely underestimate the amount of effort, complexity and cost of even simple informatics. The demands of larger grants with more complex informatics are often severely under budgeted. From anecdotal evidence presented by several attendees, it appears that the level and sophistication of review by the funding agencies of grant submissions with an informatics component is or will be increasing. This implies that our institutions should insure that grants and contracts which require informatics support be reviewed by faculty or staff knowledgeable in the field prior to submission or risk problems from the study sections.

The Structure of Funding for Informatics Facilities

The group felt that a lack of ongoing support for critical operational components of core informatics groups is a serious problem that must be addressed. It was noted that money for hardware or software is often more readily available through the current grant mechanisms than for ongoing support issues such as personnel and software upgrades. This has the consequence of increasing demands on staff, and limiting the ability of informatics shared facilities to achieve necessary critical mass.

Proceedings of the First Meeting on Computing Resources for Biomedical Research

Recommendations to Center Directors and the AACI

The group felt that a number of recommendations should be communicated to the Center Directors of our home institutions and to the AACI informatics committee for further action, including possible recommendations to the NCI and other funding bodies. These include:

Issues of Relevance to Grant Funding:

- The NCI Centers program should consider developing a series of Core Grant supplements to promote the development of both shared infrastructure and novel applications of informatics and information technology.
- The NCI should work to establish a series of best informatics practice and infrastructure guidelines, along with Core funding supplements specifically targeted to bring all groups to a minimum level of proficiency. This series of supplements should concentrate on improving the groups with the lowest expertise rather than on improving the best current groups. This is similar to grant approaches developed as part of the NSF Long Term Ecological Review (LTER) program.
- To further promote the establishment of minimal standards and capabilities within an institution, we recommend that the NCI consider establishing a baseline amount for information infrastructure and informatics support awarded as a fixed percentage of a Core Grant award.
- Review guidelines from NCI should be amended to require routine informatics reviewers on all study sections to ensure that appropriate informatics technologies are being proposed, that appropriate informatics means exist to accomplish the goals, and that the level of budgeting for these resources is appropriate to the grant. We believe that such review will have the additional benefit of more effectively leveraging already developed and innovative informatics resources at both the local and national level.
- There is evidence that the current R01 funding mechanism itself discourages the appropriate utilization of shared informatics resources and computing support groups. If this hypothesis is true, the implication is that R01s and similar vehicles reduce the ability of research centers to effectively leverage in-house informatics and information technology expertise in general. Further, it implies a reduction of the ability of these groups to deal with a rapidly changing field at an appropriate level of resolution and in the appropriate scale. We believe the NCI should study this issue. We also recommend that in addition to the aforementioned grant supplements, the Center Directors, together with the NCI, should consider establishing and funding novel new mechanisms, such as targeted informatics consortia and information technology magnet centers as mechanisms to counteract the effect.
- The NCI should study additional ways to recognize and promote important new information technology advances of relevance to the Cancer Centers. Important new informatics technologies routinely emerge from other fields, such as computer science, industry, and other key academic information technology groups such as EduCause and the AAMC. To promote innovation in Cancer Center informatics, as well as to represent the specific needs of our community, our informatics people should be routine participants in initiatives developed within these other communities. The importance of this participation should be recognized and encouraged.

Proceedings of the First Meeting on Computing Resources for Biomedical Research

Intellectual Property

- Important, innovative software developments are, in general, routinely emerging from the open source development community and in particular from our colleagues developing software for research purposes. We believe that a free interchange of software source codes developed at our institutions is in keeping with an academic and research tradition, supports the concept of software peer-review, and promotes software quality and staff intellectual development. Ultimately, it will enhance our clinical, prevention, and basic research programs. The group believes that it is important for the Centers and the funding agencies to adopt a clear policy of favoring and supporting open source software development to the greatest extent possible. It probably is reasonable to retain technology transfer agreements and licensing for large profitable efforts, but the group specifically recommends against licensing every algorithm, computer program or small project.

Consortium Activity

- The group agreed that Cancer Center Directors should be made aware of the value of a standing group to discuss common issues and concerns across a variety of areas of informatics. We would like to continue these meeting as an NCI sponsored Informatics and Information Technology directors meeting, similar to the Cancer Center administrators meeting. We are asking the Center Directors to recommend this to Dr. Brian Kimes.
- There are a large number of software, hardware, and data resources, such as the Celera mouse data, commonly in use, or of great value to all of our institutions. We recommend that the institutions work together to develop consortium and/or group discount agreements to develop better pricing for the research centers. Additionally, the NCI and the Centers should seek recognition of standalone cancer centers and research institutions, regardless of degree granting status, as full-fledged “educational” institutions for the purposes of discount agreements. EduCause is the primary forum, advocate, and policy advisor to the software and hardware vendor community in the academic area. A number of our institutions are also EduCause members. We believe that at a minimum the Centers should consider drafting a letter to EduCause advocating for recognition of all not-for-profit, peer-reviewed basic research institutions as being “educational” in nature. Additionally, we recommend that the Center Directors explore the possibility of obtaining GAO pricing for software, hardware, and data resources required for the fulfillment of grant and contract awards.

Specific Technology Recommendations

- The group recommends that the NCI give priority to informatics initiatives that fund the development of systems based on *object oriented software design*, “pluggable” *component* and *framework architectures*, and on the development of software systems for informatics that result in shared, *open source* resources. Emphasis should be given to multi-site consortia applications over single-site applications, and these could be constructed in a manner that targets different areas of interest to multiple Centers, e.g. integrated microarray data management and analysis. Such an approach would increase the informatics critical mass and leverage a better return on investment of research support.

Proceedings of the First Meeting on Computing Resources for Biomedical Research

Plans and future directions

As noted above, we feel that there is a high degree of value in having regular meetings of a professional group of peers within the Cancer Center and research informatics area. Consensus indicated that a semiannual meeting should be held. The group believes that in conjunction with the AACI and NCI informatics groups, we could play an important and vital role in developing informatics consortium; advise cancer centers on required standards for infrastructure, e.g. define minimum network speeds necessary to adequately share microarray data for clinical and basic research; provide common planning for network infrastructure, etc. Possible additional activities for this group would be participation in the definition of process for peer review of CCSG supplements; policy recommendations on chargeback rates for informatics and computing support cores and the development of core review criteria.

A steering group, including Jerry Latter (Cold Spring Harbor), Kim Johnson (Duke University Medical Center), Bob Robbins (Fred Hutchinson), and Frank Manion (Fox Chase), was selected to coordinate plans for the next meeting, to be held in the Fall of 2001 at one of their four facilities. The steering group will work to identify appropriate individuals to be contacted at each of the 60 Cancer Centers, and will send out letters to determine interest and size of next meeting.

After discussions about what the group name should be it was decided to operate in the interim under the name Biomedical Research Information Technology Exchange, or BRITE for short. Tony Popowicz from Rockefeller University agreed to set up a web-based discussion group and document exchange web site for group participants at <http://cs.rockefeller.edu/brite/>.

As a possible agenda topic for the next meeting, Dr. Robbins broached the issue of informatics lying on the critical path for the life sciences in the twenty first century. The key question is at what point does funding for informatics become crucial to the ability of a life science research institution to successfully conduct research. Dr. Robbins agreed to research this question with hard data from other fields and circulate a short white paper to the participants at some point during the next several months.

A tentative agenda for next meeting was discussed. Although a topic largely operational in nature, it was felt that the implications of the HIPAA legislation for research, research informatics, and informatics infrastructure should be a major theme for the next meeting, given the relatively short timeframe until the implementation deadlines. Specifically, the group will focus on a review of the final HIPAA regulations and their impact on research applications, collaborative data networks, systems, databases, middleware systems and policies. Other topics that will be considered for inclusion on the agenda are:

- Feedback from the first meeting
- Reports on possible purchasing consortium and software licensing cooperatives
- Discussions with NCI representatives on infrastructure and grant issues,
- Development of common service level agreements for network and technology support,
- Common security standards and specifications,
- Common software environments that can be shared.
- CORBA is emerging as a very important *framework system* (term used in a formal sense) for software development and implementation of distributed data resources within the life sciences community. The use of CORBA based information systems and the resulting impact on projects, staffing, and infrastructure is of interest to a number of attendees.

Proceedings of the First Meeting on Computing Resources for Biomedical Research

We intend that this meeting serve as a highly valuable, on-going forum for exchange to all participant institutions to inform the development of reasonable informatics capabilities and requirements. We hope that all participants find this group useful for planning purposes; in developing grant vehicles such as the CCSG “core” grants; or for any other research or grant endeavor with complex informatics and infrastructure requirements. It is our intent that all participants at Centers and research institutions gain perspective and understanding of the multifaceted issues impacting the use of information technology as part of a research environment in the 21st century.

Attendees
First Meeting on Computing Resources for Biomedical Research
April 30 and May 1, 2001

Michael Bookman, M.D.
Director, Medical Information
Management
Fox Chase Cancer Center

Ken Buetow, Ph.D.
Chief, Laboratory of
Population Genetics
National Cancer Institute

James Gergel
Cold Spring Harbor Lab

Dr. Elizabeth A. Greene
Biocomputing Facility
Manager
Fred Hutchinson Cancer
Research Center

Ms. Patricia Harsche
Vice President, Strategic
Planning
Fox Chase Cancer Center

Mr. Ron Hood
Manager, Telecom/Network
Fred Hutchinson Cancer
Research Center

Kim Johnson
Director, Cancer Center
Computing
Duke Comprehensive Cancer
Center

Lisa Johnson, MHA
Strategic Planning
Fox Chase Cancer Center

Jeff Krischer, Ph.D.
H. Lee Moffitt Cancer Center
& Research Institute

Gerald Latter
Director of Information
Technology
Cold Spring Harbor Lab

Frank Manion
Director, Research
Information Technology
Fox Chase Cancer Center

Linda Marks
Memorial Sloan-Kettering
Cancer Center

Tony Murillo
Director of Informatics
Huntsman Cancer Center

Joyce Niland, Ph.D
Chair, Information Sciences
Director, Biostatistics
City of Hope National Medical
Center

Michael Ochs, Ph.D.
Manager, Bioinformatics
Facility
Fox Chase Cancer Center

Tony Popowicz, Ph.D.
Assistant Director, Computing
Rockefeller University

Mr. James Quirk
Senior VP for Research
Resources & Government
Affairs
Memorial Sloan-Kettering
Cancer Center

Robert Robbins, Ph.D.
CIO & VP MIS
Fred Hutchinson Cancer
Research Center

Eric Ross, Ph.D.
Director, Population Studies
Facility
Fox Chase Cancer Center

Joanne Ruh
Roswell Park Cancer Center

John Speakman
Project Manager, Biostatistics
Memorial Sloan-Kettering
Cancer Center

Donn Young, Ph.D.
Director, Biostatistics
Arthur James Cancer Center