

Laboratory Directed Research and Development Program Berkeley Lab FY 2001 Coversheet

Project Title: Collaborative Cross-scale Bioinformatics Tool Development

Investigator(s): Adam Arkin (PBD), Inna Dubchak (NERSC), Eddy Rubin (LSD)

Division:

Funds Requested (FY 2001):

Proposed Project Duration:

New Proposal

**Out Year Funds Requested:
(for multiyear projects only)**

Continuation

Prop No.
Budget No.
PAO OFFICE USE ONLY

Long-Term Funding (amount, source, likelihood):

DARPA, NIH

Collaborating Divisions or Institutions:

PBD, LSD, JGI and UCB

Summary

Purpose /Goals:

The purpose of this LDRD is the creation of a strategic partnership for bioinformatics. This partnership will primarily be a service organization staffed by programmers, electronic publishing experts, and bioinformatics staff trained in the use of laboratory developed and third party bioinformatic tools. The partnership will be responsible for development of innovative bioinformatic techniques developed at the Laboratory and UCB into industrial strength tools that may be served to the biological community at large. This involves code polishing, algorithmic development for scalability and used on high-end parallel computers, development of user interfaces. In addition, the partnership will collect (where permission is given) tools or interfaces to tools developed by other individuals in order to create central lab access to often used programs. The partnership will provide analytical help, training, and workshops on new and existent technologies. Finally, as the partnership solidifies, partnership scientists will engage in their own bioinformatic research. An initial tool set for development is presented

Approach/Methods:

The partnership will be formed by recruitment of a head responsible for overseeing day-to-day operations of the partnership. A steering committee will be formed in areas of DNA sequence, DNA/RNA structure, protein sequence/structure, molecular profiling, imaging, pathway, and modeling analyses and will advise the partnership on which areas should be further developed, apprise it of new tools and technologies and be responsible for review of the partnership on a periodic basis. The steering committee will be composed of laboratory and UCB computational and experimental biologists whose tools will be fed into the partnership "code pipeline." An external advisory committee of biologists and computer scientists will also be formed with broad oversight responsibilities. They will review the partnership twice a year initially, falling off to once a year after the three-year start-up. The partnership will run a web-site with online access to tools, news of new literature, provide expert help on a contract basis, run a seminar series, and training and scientific workshops. There are a number of "focusing problems" of immediate interest to the laboratory that will initially be pursued by the partnership.

Relationship to other Berkeley Lab projects sponsored by DOE or other agencies:

Are there human subject data, cells, or tissues and/or animal use on this project? If yes, fill in the Human/Vertebrate Animal Use form. Yes No

(See instructions)

On an attachment (3 pages, maximum), please provide a brief description of the project:

Purpose / Goals; Approach / Methods; potential results or significance and, if multi-investigator or multi-divisional, proposed organization.

**BERKELEY LAB FY 2001 BUDGET REQUEST
LABORATORY DIRECTED RESEARCH AND DEVELOPMENT PROGRAM**

LABOR COSTS	FTE OR %	EXPENSE
SCIENTIFIC LABOR		
Salaries & Wages (<i>List last name & classification of all key personnel to be supported</i>)		
.....	_____	_____
.....	_____	_____
.....	_____	_____
.....	_____	_____
Payroll Burden (if not included above).....	_____	_____
Subtotal Direct		_____
Scientific Division Burden	_____	_____
SUBTOTAL SCIENTIFIC LABOR		_____
SUPPORT LABOR (TECHNICAL)		
Salaries & Wages (<i>List last name & classification of all key personnel to be supported</i>)		
.....	_____	_____
.....	_____	_____
Payroll Burden (if not included above).....	_____	_____
Subtotal Direct		_____
Support Division Burden	_____	_____
SUBTOTAL SUPPORT LABOR		_____
SUBTOTAL LABOR COST		

PURCHASES	%	EXPENSE
CONSULTING SERVICES / SUBCONTRACTS		
Purchase Cost		_____
Procurement Burden	4.6%	_____
SUBTOTAL CONSULTING SERVICES / SUBCONTRACTS		_____
MATERIALS / STORES / CAPITAL EQUIPMENT		
Purchase Cost (<i>e.g., supplies and materials</i>)		_____
Capital Equipment (please describe):		_____
Procurement Burden	4.6%	_____
Material Handling Burden	4.8%	_____
SUBTOTAL MATERIALS COST		_____
SUBTOTAL PURCHASES COST		

OTHER COSTS		EXPENSE
Travel: <input type="checkbox"/> Domestic <input type="checkbox"/> Foreign		_____
Recharges		_____
Indirect Charges		_____
Other		_____
SUBTOTAL OTHER COSTS		_____
TOTAL LDRD BUDGET REQUEST		

NOTE: Human Subject and/or Vertebrate Animal Use Questionnaire is required at proposal stage *only* for those LDRD proposals that do have human subject and/or animal use, including the use of human-derived data, cell lines, or DNA as well as animal tissues. Information for NEPA/CEQA/EH&S Review form(s) will be required of all proposals *after* preliminary selection (see check list below).

Status	Form Requirements
Proposal Stage	Cover sheet (1 page limit) and Budget page required of all proposals. Human Subject and/or Animal Use Questionnaire required only if project does have human subject and/or animal use.
Preliminary Selection (prior to start of fiscal year)	NEPA/CEQA and EH&S Review form.

Proposal #: _____ Type (e.g., WFO, CRADA, FTP, LDRD, DOE M&O, Gift): _____

Project Title: _____

1. Identify LBNL, UCB, DOE facility, and any other offsite locations where work will be performed:

Yes	No	Uncertain	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2. For LBNL locations: Does the project involve room or building modifications?
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	3. Would the project disturb hazardous substances or contaminants that preexist in the environment, such that there would be a release not covered by agency permits?
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	4. For work at non-DOE locations only: Would the project disturb any of the following environmentally sensitive resources?

- Property (e.g., sites, buildings, structures, objects) of historic, archeological or architectural significance designated by Federal, state or local governments or property eligible for listing on the National Register of Historic places;
- Federally-listed threatened or endangered species or their habitat, Federally-proposed or candidate endangered species or their habitat; or state-listed endangered or threatened species or their habitat;
- Wetlands regulated under the Clean Water Act and floodplains;
- Federally- and state-designated wilderness areas, national parks, national natural landmarks, wild and scenic rivers, state and Federal wildlife refuges, and marine sanctuaries;
- Prime agricultural lands;
- Special sources of water, such as sole-source aquifers, wellhead protection areas, or other water sources vital to the region;
- Tundra, coral reefs, or rain forests.

<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5. Would the project result in hazardous emissions, wastes, or effluents outside permit limits?
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	6. Is new or additional safety documentation needed for the work?

Division	Principal Investigator	Date
----------	------------------------	------

To be completed prior to beginning work:

Proposed classification:

LBNL NEPA/CEQA Program	Date	DOE NCO	Date
------------------------	------	---------	------

1.0 INTRODUCTION AND CONNECTIONS TO OTHER PROJECTS AT LBNL CONNECTIONS TO OTHER

It is clear that there is a convergence of activities around LBNL that require the formation of a bioinformatics infrastructure to promote collaboration and development of bioinformatics tools. With the PSF producing 30 MB/day of sequence at the JGI, structural genomics initiatives current and planned producing protein structures at the ALS and miscellaneous genomics projects and individual high-throughput biological experiments producing genetic, molecular profiling, physiological and image data, a small industry in databasing and computational analysis has grown up around these projects to mine their data for the biological insight that is ultimately their goal. The scientists engaged in the computational projects are developing tools that are directed towards solving a particular biological problem at hand or are innovative research-grade programs that are rarely polished into industrial-grade tools for use by the LBNL and other general biological communities. Further, many of these tools, even were they in polished form, are of sufficient complexity that the average bench biologist would be hard pressed to properly use the tools, let alone know which one is appropriate for their problem.

Here we propose the formation of a bioinformatics partnership with three primary purposes: 1) Training of and consulting with laboratory scientists in use of new and existing tools to aid in their research problems. This effort includes dissemination and documentation of tools, creation of tutorials and the development of courses and workshops in bioinformatics theory and application; 2) Extension of existing tools for client laboratories with specialized needs or identification of LBNL and UCB computational biology laboratories that may have expertise in the problem to be solved; 3) Polishing of tools developed by LBNL/JGI/UCB researchers for robust and rapid performance and serving to the general community; the adaptation of complex bioinformatic tools for uses with high speed supercomputers. When the partnership begins to function well as a service entity, partnership staff will also be free to pursue their own research agendas.

We have an unusual opportunity to take advantage of the existing critical mass of computational biology at the laboratory and to create a focal point for the professional dissemination of these efforts for Laboratory and general use and to provide a central identifiable pathway for polishing and dissemination of laboratory product. This will also lead to a centralized location at which the computational biology community at the Laboratory can (at least conceptually) convene. That is, it provides a central resource to increase communication among the groups. Further, the partnership will also provide a location for the import of external computational expertise, the creation and implementation in the theory and application of computational methods to biology, and will be a natural administrator for a directed seminar series and for the hosting of yearly workshops to bring in outside "clients" and colleagues.

The nature of this partnership is such that it will be greatly strengthened by coordinating efforts at all three of the closely placed institutions, the JGI, LBNL, and UCB. Thus, contributing scientists should be drawn from all of these communities, as should the primary scientific clients. This is naturally a cross-divisional undertaking. We propose an administrative structure in section 4 below, but this partnership is of sufficient complexity and criticality that the finalized structure will only be arrived at the end of this proposal period. Follow on funding should come from DOE, NIH and DARPA. Matching funding from campus should also be considered.

It is critical to the Laboratory that such a research pipeline be developed. The Laboratory is host or partner to a large number of genomic and post-genomic efforts including the *Drosophila* genome project, human, mouse and other genomic efforts at JGI, and the structural genomic programs at ALS. Further, the future programs in biology at the DOE, DARPA and NIH involve funding for complex quantitative biological projects such as the Microbial Cell (DOE, \$12M dollar line item in next years budget). The partnership will position the Laboratory and its personnel as major location for computational biological analysis that takes advantage of the existing infrastructural and scientific strengths of LBNL. Further, this partnership will prevent duplication of effort among these projects, ensure that the bioinformatics research at and around the Laboratory is finished into distributable products (thus raising the profile of bioinformatics), and that these tools can be utilized by all the projects in need of their analysis. Finally, the staff of the partnership will provide the necessary technical skill for biologists untrained in bioinformatics to use these tools, will provide customization of the tools for particular problems, will direct new bioinformatic problems driven by the biology users group to the appropriate contribution bioinformatic contributing scientist, and provide an intellectual home-base for contributing scientists and biological users to interact and trade problems and solutions.

2.0 Mission areas and activities. As the primary goal of this partnership is not research but service the mission areas are all based on collecting and honing technology of use to a biological client base. With a suitable armory of tools the partnership may then be involved both with training the community in the use of the tools and the application of these tools on a contractual basis. As the partnership team becomes more expert at application and training in the tools they will become a natural partnership for courses, seminars and workshops in bioinformatics in the Berkeley communities. We have identified eight areas that must be developed in order to make this partnership as useful and strong as possible.

2.1 Central Collection of Tools. In order for the partnership to accomplish its goal to be a bioinformatics service the staff must collect and familiarize themselves with the wide range of biological data analytical tools available from Laboratory researcher and in the general literature. Among the areas of bioinformatics identified as central to the Laboratory mission are genome assembly and finishing, genome annotation, regulatory site identification and sequence analysis, cross-genome/phylogenetic/polymorphism analysis, RNA identification/structure/function, protein structure/function, molecular profiling (microarray, proteomic, metabolomic) analysis, molecular (e.g protein, RNA) engineering, cellular simulation and image analysis. For every tool collected, a "tool page" will be created on the web site that gives credit to the originator of the

idea and to the finishers of the tool, a description of the tool and its input and output data, and pointer to documentation, tutorials and test data sets for the tool. For every tool type, benchmarking test for speed and function will be developed. In addition, common user and programmer interfaces to the tools will be developed as will standards for data output format. Results from analyses will be archived.

2.1.1 Negotiating for use of published tools and training on these tools. The partnership will initially recruit a list of “contributing scientists” from the LBNL/UCB/JGI community. These scientists will provide the unique Laboratory bioinformatics resources that will differentiate us from other bioinformatics sites. (see section 3.0 for a list of contributors and the procedures for becoming a contributor). However, these tools will not cover all the areas of interested to biological users of the facility and thus external tools will have to be collected. For the cases, in which the support for the external tool does not meet partnership standards, the partnership will negotiate for use of this tool by the partnership. partnership staff will bring the code up to partnership standards; produce proper documentation and a training procedure for use of the tool.

2.1.2 Coding up from public literature. When external tools are identified that have been published or otherwise made free but where well-tested, efficient and clear code are not available, the partnership staff will take on the job of creating an implementation of the tool and an interface according to partnership standards. The decision to spend time on such an effort will be ultimately up to the partnership Head and the Steering committee or may be performed on contract to a client laboratory. As with contributor and negotiated tools these will have a tool page, documentation, tutorial and test data set.

2.1.3 Polishing Laboratory developed tools. Contributing scientists will submit their research tools for polishing to the partnership. These submissions enter a job queue. In first approximation, tools will be “finished” on a first come, first serve basis. However, partnership personnel can prioritize specific tools based on partnership needs. Further, tools can be finished on contract basis by partnership staff. Finishing a tool entails cleaning up the code, creating user-interfaces, creating standardized documentation and tutorials for it, benchmarking it against standard data-sets for its tool type and comparison to other tools. In some cases, course materials will be produced for use in classes and workshops offered by the partnership.

2.1.4 High-end algorithms. In some cases, such as assembly of full genome sequence from shot-gun data, primer design for microarrays, and comparative genomic analysis, the use of high speed supercomputers or clever combinatorial algorithms will be necessary. In this case, the problem may be handed off to interested NERSC staff and other contributing scientists for adaptation of the code for use on partnership computers.

2.1.5 Collection of experimental validated results from tools. One of the most important goals of the partnership will be to keep track of which tools bench biologists have used to successfully predict or confirm experimental results. The ability of a tool to yield validated biological results is the main indicator of quality. An example of such data would be data from the CASP structure prediction competition. These results will be archived by the partnership.

2.1.6 Search programs. All tools, manuals, tutorials, and test-data will be entered into a searchable database for easy navigation to the section of the site of interest. Further, a decision support tool will be created to aid the biologist in the choice of tool for their particular problem.

2.2 Documentation, Training, Dissemination. There are three central general services of the partnership. These are documentation of tools, training in their applied use and general theory, and dissemination both of new tools built by contributing scientists and news of tools developed by others off-site. A “Tools” newsletter will be developed in order to disseminate information about new tools, their usage, comparison with similar programs and announcements of successes.

2.2.1 General fundamental tutorials. One of the products of partnership staff will be to compile and write tutorials on the basic science behind the major classes of bioinformatic tools. This effort is important both in the training of partnership staff and in presenting the partnership as a useful, professional service to the external community. Such tutorials, collected in one location will serve as a draw to the partnership site and set up the partnership as an educational as well as a service facility.

2.2.2 Specific tutorials and manuals. For every tool collected by the partnership, staff will produce a technical manual and specific tutorials for the tool. The manual will comprise the basic usage of the tool, format of the input and output files and a brief description of the function of the tool. The manual will also contain a specific technical description of the algorithms and specific code implementation. All code will be documented and referenced to the manual. The tutorial will describe the basic theory and usage of the tool and go through a number of worked examples. The data sets used in the examples will be made available on the site. Comparisons with other tools will be made where appropriate. Tutorials will be served online.

2.2.3 Training. The partnership staff will create a set of courses each on a set of related tools. The classes will be dedicated to teaching basic theory and application of partnership tools. These classes will be offered on a regular schedule for free for Laboratory personnel and by subscription for external researchers.

2.2.4 Help desk. The partnership will maintain a help-desk function for troubleshooting and explanation of the use of partnership tools. This function is designed for quick questions and bug reports. The Help Desk will be available to the Tri-Institutional community for free. Calls will be handled on a first come, first serve basis via the web if possible. Users may request a telephone follow-up.

2.2.5 Workshops. Once a year, the partnership will organize a Bioinformatics workshop designed to bring in cutting edge researchers to the LBNL community. It will serve as a small, possibly international venue for scientists (chosen by the steering and advisory committee; see below) to present their work, provide non-binding commentary on the partnership and its tools in meeting with the Advisory and Steering committees, and interact with partnership contributing scientists in a relatively relaxed venue. The format of the workshops will be determined by partnership staff but will likely be three day affairs with morning and early afternoon talks, afternoon poster-sessions and evening free-interaction time.

2.2.6 Maintaining Links, News, Central Journal Access. In addition to its other duties, partnership staff will maintain web pages that explain and link to external bioinformatic resources, reports on important news in the bioinformatic community, provide central access to relevant journals for which the Laboratory has an online subscription, and will provide direct pointers to articles about tools at the partnership site. The partnership will also send representatives to the major bioinformatics meetings internationally. The staff will report on the cutting edge research presented at these conferences.

2.3 Collaboratory locus. In order for the partnership to remain an integral part of the Bioinformatics community at LBNL, a suitably localized space should be provided for personnel along with suitable interaction space so that partnership associates and contributing scientists can meet and discuss their work in a collegial manner. This will promote the community to be coherent and interactive. Further, the web site will have private message boards and working document and code sections that the partnership community can access to help in development of the partnership.

2.4 Contract Work. The primary way that the partnership staff will be supported in the long run is by contract work. Individual biology laboratories in need of high-end computational biology experts can contract with the partnership to apply computational tools to their problem. This may entail application or extension of current tools, creation of custom databases, training of client personnel, or direction of clients to specific contributing scientists when the problem is beyond the scope of current tools and their simple extension. A charge-out system such as that in other service divisions will be developed.

2.5 Research. One of the functions of the partnership will be to perform its own research in bioinformatics and bioinformatic algorithms. Once the service pipeline is in working order, we foresee partnership personnel having time to pursue their own research interests. These interests may be funded by grants from outside sources and partially subsidized by contract work with partnership users. Though all partnership members will have significant percentage efforts in the service aspects some will be able to maintain small research groups. Post-doctoral fellows, for example, may be recruited by the partnership to provide a small research core and to obtain training in applied bioinformatics.

3.0 Participants and Tool Development. The Laboratory, JGI and UCB already have a vibrant computational biology community that, however, does not have a central facility in which to meet or organize and develop their tools. However, a large number of these recognize the advantage of the bioinformatics partnership as proposed. Accompanying letters indicate the willingness of the bioinformatic community at the Laboratory and UCB to donate tools and expertise to the Center. The biological community also has written letters of support. *The partnership will start in year1 with this strong base of scientists and their tools. It is this large collection of top-quality programs that will make this partnership initially functional and propel it future success.* The partnership's web site will provide an application to become a contributing scientist. Ultimately, contributing scientists will be allowed from anywhere in the world. To start, the Center will chose a few focusing problems in order to best serve the immediate needs of the community. These problems may include supporting and maintaining local databases such as the ASDB: alternative splicing database, bringing tools like Vista (Dubchak), RNAGENE (Holbrook) and Guacomole (Rokhsar) into general usage, and, in collaboration with high-end bioinformaticists, developing a shot-gun sequencing genome assembly program needed by the JGI. This latter project is of sufficient complexity that a contributing scientist would have to be the lead developer.

4.0 Organization. The organization of the partnership will necessarily be dynamic during the start-up period. However, there must be a central head who is strongly advised on the partnerships aims and progress. In the first cut, the Head (or co-heads) are responsible for coordinating five sub-departments. (See Figure 1.) There are three central technical departments: bioinformatics, coding, and database. These three are separated based on the primary expertise of the individuals who staff those departments. The bioinformatics department is responsible for knowing the theory and application of partnership tools, for helping clients use these tools or for using these tools themselves for client's needs, for training and tutorial development, and identifying areas in need of development and/or better code or algorithms. The Coding core is responsible for the finishing of contributor code, the coding of public domain algorithms, development of advanced algorithms for efficient execution of calculations on supercomputers, and for creation of tool technical manuals. The Database core is staffed by database design and use experts able to design custom databases for clients, develop code for interface to existing databases, and provide documentation on database access routines and point to new database technologies of use to bioinformatics.

Two administrative departments support these three technical departments: central administration and electronic media. Central Administration is responsible for maintenance of the partnership computer servers and website, as well as, planning of courses and class schedules, and implementation of the yearly workshop. Electronic media is staffed by technical writers and web-site designers who aid in finishing of manuals, tutorials, class materials, publications, grants, and the web-site. These two departments free the technical core to pursue the central scientific jobs efficiently.

Each department will have a head responsible for directly reporting progress and problems to the head. In addition, the partnership will assemble a Steering committee made up of contributing scientists as well as biologists and computer scientists from the tri-institutional community. The Steering committee will provide continual guidance on what areas should be partnership priority and will provide technical advice and partnership review at multiple times throughout the year. An external advisory committee of experimental and computational biologists and computer scientists will also be formed as a review committee that will meet yearly at the partnership workshop to provide a written review of the partnership's success and progress. The review will be based on informal interactions with the partnership throughout the year, the web site, and an evening of presentations by the partnership leadership (Head and department heads) and contract clients.

The first year of this LDRD will be engaged in recruiting the partnership staff and assembly of the Advisory and Steering committees. We project the need for one or two heads, two-four coding staff, two-four database staff, three to seven

bioinformatics staff, two electronic media staff and three central administrators. The role of the heads will be to: 1) Oversee the broad mission planning for the division including identifying new bioinformatic areas for development, areas of client need and the decide on the priorities of the partnership, 2) Recruit personnel to the division appropriate for these missions, 3) Actively promote the partnership to the local and external communities, 4) Attend to budgetary concerns, 5) Report to the Advisory and Steering committees on progress of the partnership, and 6) Report to Laboratory leadership on programmatic progress, needs, and problems. Two heads may be necessary in order to best represent the wide range of activities encompassed by bioinformatic research. The coding staff is responsible for analyzing, polishing and extending tool codes supported by the partnership. Because of the complexity of some of these tools and because of their number, two programmers is the absolute minimal set with which the partnership could function. Two database staff members are necessary because a large number of bioinformatic tools use databases of information as their starting point. The partnership will need professionals who have sophisticated knowledge of current databases and are able to create efficient scalable database designs for use with partnership tools. Two is the absolute minimum given the number of projects the partnership should be able to handle. The three bioinformatics staff is the minimal number to cover the a few focusing described in section 3.0. The two electronic media staff members are dedicated to electronic/web-based dissemination and technical writing for the manuals, documentation and tutorials. Three central administrators are necessary for system/web administration, course/workshop administration, and general support for the partnership.

The competitiveness of this field means that there will probably be a rapid turn-over in key personnel. This may be a strength of the partnership in the following way. The partnership will attract bright young researchers who want to shore up their foundations in applied computational biology. As such the partnership can act as a training program for these people and act as a leaping off point for other academic and industrial positions. The training aspect of the partnership will allow it to apply for specific funding in this area. Further, it will allow the hiring of new staff with particular expertise in the evolving “focusing problem” set under development by the staff.

5.0 Timeline. This program is divided into a number of milestones:

Task 1: Recruitment and situation of personnel on site

Task 2: Formation of the Steering and Advisory Committee

Task 3: Collection and polishing of tools from the initial pool of contributing scientists

Task 4: Collection and polishing of publicly available tools

Task 5: Creation of documentation, tutorial, and class materials

Task 6: Design and creation of web site.

Task 7: Setting up of services: contract systems, help desk functionality, classes and workshops

Task 8: Annual review

The schedule for the tasks is outlined in the Gantt chart in Figure 2. Research at the facility will progress as the staff has the time and funds. The partnership will create formal ties to the campus genomics training grant programs and the Department of Bioinformatics/Computational Biology within Bioengineering. The Steering committee will naturally be populated with bioinformatics specialists from PBD, LSD, JGI, the Drosophila Genome project and NERSC. The initial search committee is made up from the PI's of this LDRD who will then take on the roles of contributing scientists and/or be members of the Steering committee.

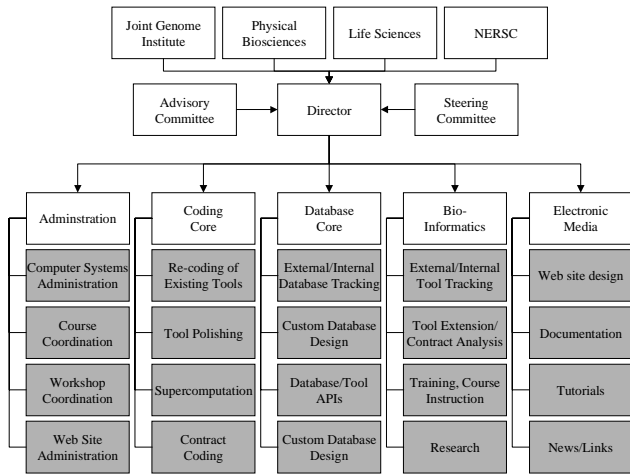


Figure 1. A rough organizational chart for the partnership

	Year 1				Year 2				Year 3			
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Task 1 Recruitment	[Blue bar]											
Directors	[Blue bar]											
Technical Personnel	[Blue bar]											
Administrative Personnel	[Blue bar]											
Task 2 Formation of Advisory committees.	[Red bar]											
Steering Committee	[Red bar]											
External Advisory Committee	[Red bar]											
Task 3. Contributed Tool Finishing	[Green bar]											
Task 4 . Public Tool Finishing	[Yellow bar]											
Task 5 Documentation	[Black bar]											
Manuals and Tool Documentation	[Black bar]											
Online Tutorials	[Black bar]											
Course Notes	[Black bar]											
Workshop Proceedings	[Black bar]											
Task 6 Web site	[Purple bar]											
Task 7 Services	[Light Purple bar]											
Help Desk	[Light Purple bar]											
Contract Work	[Light Purple bar]											
Classes	[Light Purple bar]											
Workshops	[Light Purple bar]											
Task 8 Annual Review	[Grey bar]											

Figure 2. The time-line for partnership Tasks