

MSU INVESTMENT PROPOSAL FOR INSTITUTIONAL PRIORITIES

PROPOSAL OVERVIEW

Title	Institutionalized Support for Management of Research Cyberinfrastructure	Request Date	12/16/2011
Department	Center for Computational Biology	Email	jpmillerusa@gmail.com
Requestor	John P. Miller	Phone	994-7332

STRATEGIC ALIGNMENT

Core Themes and Objectives
(check all that apply)

Educate Students

- Our graduates will have achieved mastery in their major disciplines
- Our graduates will become active citizens and leaders
- Our graduates will have a multicultural and global perspective
- Our graduates will understand the ways that knowledge & art are created and applied in a variety of disciplines
- Our graduates are prepared for careers in their field
- We will provide increased access to our educational programs
- Communities and external stake holders benefit from broadly defined education partnerships with MSU

Create Knowledge and Art

- Students, faculty, and staff will create knowledge and art that is communicated widely

Serve Communities

- We help meet a fundamental need of the citizens of Montana by providing degree programs for our students
- We help meet the educational needs of the citizens of Montana by providing a wide range of educational opportunities to a variety of students
- Our students, faculty, staff, and administrators reach out to engage and serve communities
- Our students, faculty, staff, and administrator reach in to build the university community

Integrate Learning, Discovery, and Engagement

- Each graduate will have had experiences that integrate learning, discovery and engagement
- Outreach activities will educate students and address the needs of the communities we serve
- Students, faculty, and staff will create knowledge and art that addresses societal needs
- MSU is a community that will be characterized by synergy within and across disciplines, roles and functions.



Stewardship

- The public trusts the institution to operate openly and use resources wisely
- The faculty and staff are well-qualified and supported
- MSU will support Native American students, programs, and communities
- MSU will be an inclusive community, supporting and encouraging diversity
- Our publicly provided resources are used efficiently and effectively
- Natural resources are used efficiently and sustainably
- MSU nurtures a culture of resource conservation and ecological literacy among students, faculty and staff
- Our physical infrastructure (e.g., building, equipment, open spaces) will be well-maintained and useful

INSITUTIONAL BENEFIT

Campuses

Bozeman Billings Havre Great Falls FSTS Extension MAES

Cross Depts

Please List: All Departments with funded research or using AgBooks or Faculty Activity Database

TIMEFRAME

Proposed Dates

Start: Summer 2012 End: Continuing base investment

COST AND REQUIREMENTS

Funding Type	One-Time (\$)	Multi-Year (\$)			Base (\$)	FTE
		Year 1	Year 2	Year 3		
Personnel (w/benefits)					130,319	
Materials & Supplies					5,000	
Travel						
Contracted Services						
Capital						
Other Operations						
TOTAL					135,319	

Salary (with benefits)	Position
14,400	1 mo. Summer salary, Director of the RCG
41,046	50% Assoc. Director Cyberinfrastructure
41,046	50% Assoc. Director Software Engineering and Informatics
33,827	50% Sr. Systems Administrator

Operations

5,000 50% Basic yearly operational expenses (site licenses for software, phone service, Kenco building security services, custodial service for non-office space, office supplies.)

135,319 total

Please comment, if necessary, regarding cost and requirements

PROPOSAL SCOPE

The goal of this proposal is to institutionalize the minimal level of base salary funding that is required to support MSU's "research cyber-infrastructure". Specifically, we request 50% salary and benefits for 3 essential technical staff of the Research Computing Group (RCG), 1 month summer salary for the faculty Director of the RCG, and 50% of the typical yearly operating expenses. The remainder of the staff salaries and expenses will continue to be derived through grant-funded contracts with different research units who use RCG services.

Background and Rationale

The mission of the Information Technology Center is to promote, develop, deliver, and facilitate the use of information technology services and resources. The ITC does a spectacular job keeping the computing, communications and network infrastructure up and running on our campus, and provides superlative support for faculty, staff and students in all matters related to the effective, efficient and secure use of those technologies for administration and teaching on our campus. ITC personnel also provide a base level of consulting and troubleshooting expertise to researchers who need to use computers, advanced programs, and the "world wide web" to carry out their research programs.

However, as MSU's research enterprise has grown and evolved over the last few years, researchers' demands for a range of very important, advanced and relatively specialized types of technical services has surpassed ITC's capability for providing those services. Those technical services are generally lumped under the catch-phrase "research cyber-infrastructure" by the agencies that fund research in the US, and include **data acquisition, data storage, data management, data integration, data mining, data visualization, and advanced computation**. The Wikipedia entry for "cyberinfrastructure" defines it as follows: "*In scientific usage, cyber-infrastructure is a technological and sociological solution to the problem of efficiently connecting laboratories, data, computers, and people with the goal of enabling derivation of novel scientific theories and knowledge.*"

The cyberinfrastructure-related services listed above cannot be considered as "luxuries" that researchers could either cobble together or get along without; rather, they have become baseline necessities for all researchers, yet require expertise and knowledge that many researchers simply do not have. As of this last year, every major US funding agency now requires a detailed description of an effective **Data Management Plan** as an essential part of every grant application, along with a description of how the data will be archived and shared with the larger research community. For the somewhat frightening details, see the following link that summarizes the requirements for NIH, NSF, CDC, DOE, DOD, EPA, NASA and USDA: <http://rcg.montana.edu/data-management-plannin/funding-agency-guidelines/>

So if the researchers are smart enough to get grants that require such an advanced cyberinfrastructure in the first place, why can't they just take care of their own cyber-problems? The simple answer is that very few "domain scientists" have the specific, extensive technological expertise required for developing a workable data management plan meeting the funding agencies' requirements, much less the expertise to actually implement and maintain an effective database. (More detail and supporting information for this assertion will be provided below). In other words, as well as a need to "plug in" to computers and the network, researchers also need to "plug into" a group of people who have the necessary expertise to develop and provide solutions to researchers' cyberinfrastructure problems.

For those researchers at MSU that can't do it themselves, how is it currently getting done? The simple answer is as follows: many contract with the **Research Computing Group (RCG)** of the Center for Computational Biology. The RCG provides technical expertise and support for research programs that rely on the use of advanced information technology hardware, software and data management strategies. The RCG is a collaborative effort between the Center for Computational Biology, ITC and the Office of the Vice President for Research, and provides affordable research-specific cyberinfrastructure resources and expertise that lie outside the scope of ITC services. The RCG's website gives a very concise but comprehensive description of the services provided: <http://rcg.montana.edu/> To follow up with the specific example of preparing a data management plan, refer to this link: <http://rcg.montana.edu/data-management-plannin/> Further, the recently-awarded 5-year, \$20M NSF EPSCoR grant has a core, essential cyberinfrastructure component, and a supplemental 3-year, \$3M EPSCoR "Track 2" proposal focused solely on Cyberinfrastructure is currently in preparation. The RCG is playing a principal role in each (doing contract work for MSU and UM researchers!), and the RCG Director is on the leadership committee for both projects, and is taking one of the lead roles in preparing the EPSCoR Track 2 Cyberinfrastructure proposal.

Although the RCG is currently operating very effectively, it is supported completely by “soft money” from several disparate sources, which creates a very fragile operation in several respects.

> First, there are only 4 technical personnel who support the entire workload. Each of these 4 has a different specific realm of expertise, and losing any one of them would severely compromise the RCG’s operation. These 4 are all supported 100% on “soft money”, and fluctuations in the grant funding stream make it impossible for the Director to guarantee continuous funding for more than a year at a time. Though the RCG has managed to operate “in the black” since its inception, it operates right on that thin black line, and even the *perception* of a tenuous funding stream on the part of the technical staff creates stress that is a constant temptation for them to get a “real job.” To be blunt, any of them could immediately move to a higher-paying and more secure job in any of several local high-tech companies. (In fact, this has happened several times over the last 3 years!) Note that this proposal is NOT requesting full salary for these 4 people, but only 50% salary, in keeping with the strategy of maintaining operations as a partially research-funded service center. The current staff have indicated that this level of institutional support would mitigate their concerns about job security, and all of us feel certain that the current and projected levels of contract activity will cover the remainder and even enable a work-driven expansion to meet increasing needs.

> Second, there is currently no institutionalized support for the Director: no teaching relief, nor any base salary component. The current and previous directors have negotiated with the VPR for some level of additional compensation on a year-to-year basis. Considering the significant extra workload associated with directing the RCG, we propose that some support for the Director be provided along with partial central support for the technical staff.

A final logical question is as follows: if these personnel and the services they provide are so essential, why doesn’t ITC provide them? The answer has two parts. First, while the ITC has been (and continues to be) very effective in all aspects of developing and maintaining the core hardware and software infrastructure required for the administration, teaching and some percentage of the research tasks on campus, it has no personnel assigned to support the cyberinfrastructure needs at the level demanded by the growing research enterprise, and could not reassign personnel from other tasks without compromising other central activities. Second, a reprioritization of tasks within ITC could not solve the problem, since there are no personnel with the specific required expertise that could be re-assigned for the tasks currently supported by the RCG.

Detailed Breakdown and Cost of the proposed investment:

The purpose of this proposal is to obtain some base support to insure the long-term sustainability of the RCG. As presented in the budget section, we request 50% salary (and benefits) for 3 core technical personnel, 1 month of summer salary for the Director, and approximately half of the typical yearly operating expenses (e.g., site licenses for software, phone service, Kenco building security services, custodial service for non-office space, office supplies.) The remainder of the staff salaries and expenses will continue to be derived through grant-funded contracts with different research units who use RCG services (see list of examples below). A detailed list of current RCG personnel, along with their responsibilities and the services they provide, are appended to the end of the proposal narrative. The total cost of the investment would be \$135,319.

How MSU would benefit from the investment:

Essential core services will be available to grant-funded units at MSU (Bozeman plus all other campuses).

The investment will effectively subsidize some low-level universal services, and will provide the capacity for the RCG to contract for more advanced services on a recharge basis. Two examples of the types of subsidized assistance that are universal and relatively low-level are as follows: a) consultation for the general planning of data management plans and “boiler-plate” text required for submission of all grant proposals, and b) technical advising on cyberinfrastructure needs and strategies for new researchers who do not yet have grants. An example of a higher-level task to be provided on a contract basis is the development of the “VOEIS” software system, funded through the current EPSCoR grants.

Essential core services will be available to non-research units. Several other academic and administrative units currently have contracts with the RCG for essential services. Two examples are **AG BOOKS** and the **Faculty Activity Database (FAD)**. These services would terminate if the RCG disbanded.

All services will benefit from the “economy and efficiency of scale”. Centralization of these cyberinfrastructure services into a tight, integrated group of 4 technical staff has enabled the distribution of different tasks in a way that optimizes efficiency and effectiveness. Though each one of the RCG staff is extremely highly trained and an excellent

"problem solver", no one staff member has the breadth of expertise to cover all aspects of any single contracts. Having 4 core staff members that cover different programming and service regimes enables the group as a whole to divide and integrate work effectively: each does what he does best, and passes tasks components to other partners (or asks for help) when a problem is encountered. Furthermore, many tasks have been identified that are fairly low-level, and a scheme has been developed through which undergrad helpers from CS and Engineering departments are hired for specific "sub-contracts". For example, one contract involved the development of an extensive web-based user tutorial, for which an undergrad was hired for one semester. That undergrad learned how to use the software, under the supervision of the RCG member who wrote the software, and then worked relatively independently under that staff member's supervision to develop a truly remarkable, effective and engaging set of tutorials. Thus, the student was paid by a research grant sub-contracted through the RCG to help a research group in a very cost-effective manner, and benefited immensely (adding to his training and technical resume!) in the process.

> In summary, the existence of the RCG as a centralized group having diverse expertise enables an efficiency of scale that yields reduced cost, reduced turn-around time for each job, and increased job quality.

BELOW WE PRESENT ADDITIONAL DETAILS OF THE RCG SERVICES AND PERSONNEL:

THE RCG PROVIDES THE FOLLOWING CORE SERVICES:

Data management: Our experienced database designers and administrators develop comprehensive solutions for designing an infrastructure to capture, manage, store, and publish data. We support multiple datastores from the familiar relational-style such as SQL and PostGRES, to document databases like persevere and mongo db. Where appropriate, the RCG provides estimates of the cost of its services along with other competitive services in the private sector.

Data Management Planning: We work with researchers to determine needs related to their data. We also provide data management templates for grant submission in response to recent changes implemented by the National Science Foundation and National Institutes of Health.

System Administration: We administer systems, both physical and virtual, in compliance with all university guidelines as well as industry best practices. The RCG system administration team has more than 20 years combined system administration experience.

IT Consultation and Training: Our consulting team will work with individuals on a variety of complex implementation decisions. We can provide input, guidance and training on data gathering, sharing, storage, and dissemination.

Managed Server Hosting

Virtual Server Hosting The RCG supplies virtual computing servers and data storage to clients at a significantly lower price point than individually purchased hardware. The infrastructure also provides these services much more reliably through the use of server/data back ups, uninterruptible power supplies, and cooling systems.

Physical Server Hosting Managed Server hosting of physical machines in our datacenter is also available if a virtual machine does not meet the requirements for the research project.

System Administration Our system administration services incorporate setup (which includes machine installation, user creation, software configuration, etc), maintenance (modifying users, installing and configuring new software, hardware repair, etc), as well as system and service monitoring and notifications.

Storage and Computation

Large-scale Storage The RCG Storage Cloud platform offers large data storage at a very competitive price. Since the storage system is on-campus, the researchers can keep their data within the network of the University so as to not burden themselves (or the university) with off-campus data bandwidth costs. The cluster is comprised of dual redundant, high-availability dell servers that provide access to a scalable, dynamic storage system using the Gluster open-source software. The RCG Storage cloud currently offers 12 TB (usable) of resilient, highly available storage (this space is expected to more than double over the next six months). Long-term cold-storage backup is currently implemented using our legacy 10TB storage cluster.

Dedicated Computation The RCG supplies computational resources to aid researchers in performing computationally intense operations. We currently have a 16-core server with 128Gb of ram for large computation purposes. In addition, we act as a liaison between MSU and cloud service providers like Amazon and Microsoft via the Pacific Northwest Gigapop and Internet2.

Software Development The RCG has experienced application developers on staff able to provide software infrastructure to aide in research and administration. Our group's development is based on an agile style of software architecting to provide the quickest access to proto-type software for users to identify features and gather feedback early on. This enables us to develop projects in months instead of years and to add features as requested.

Four Examples (out of MANY!!!) of RCG Applications Developed at MSU for Academic, Research and Outreach Infrastructure. A comprehensive list can be provided upon request.

The **Faculty Activity Database (FAD)** is an application that is used annually to collect performance data for all faculty at Montana State University. FAD provides department heads, deans, and administrators with data collected directly from faculty annually. This application has been in production for 4 years, and approximately 80% of faculty use it for their annual review reporting.

The **Protocol Authorization Manager (PAM)** is designed to help the MSU Institutional Biosafety Committee review and oversee research and teaching activities that use recombinant DNA, infectious agents and toxins.

The **Virtual Observatory and Ecological Informatics System (VOEIS)** is a collaborative project between MT and KY and partners from industry and the public sector to develop an integrated sensor and ecological informatics system through the use of modern cyberinfrastructure resources. The RCG is responsible for the development of the data management and computational infrastructure aspects of this project.

College Choices for Adults (CCfA): An outreach program supported by WICHE requires software that allows a person interested in pursuing online education access to information about online curricula. The CCfA software application lets the institutions log in and edit their data, and it allows the WICHE staff to validate and publish the changes made by the institutions. The site is easy to browse and navigate with lots of data for prospective online students.

RCG STAFF

Director, CCB Research Computing Group [Dr. John Miller] Dr. Miller is responsible for all aspects of the CCB and RCG's operations, staffing, interaction with MSU faculty and administration to arrange jobs and contracts, overall workplace guidance related to any problems or issues that arise through the contractual interactions with other researchers and Departments, and HR functions (including hiring and termination processes, performance evaluations, and general personnel issues). He serves as the head of the CyberInfrastructure component of the new EPSCoR grant, and is one of the lead Investigators on the Track 2 proposal. He also works with the accountant to manage the budgets for all RCG contracts and grants. He reports to Dr. Jim Rimpau, CIO and Dr. Tom McCoy, VPR. This role realistically represents a time commitment of about 2 hours/day, corresponding to a 0.20 FTE time commitment. We request only a single month of summer salary, equivalent to approximately 0.11 FTE.

Associate Director, Cyberinfrastructure & Information Systems, Senior Software Engineer [Pol Llovet] Pol is responsible for the design and maintenance of RCG hardware and software systems, and serves as the primary technical contact for researchers interested in RCG services. In this role he supervises the Senior Systems Administrator (Matt Harris). Pol's current responsibilities include service as the lead software developer for the Faculty Activity Database, the BioSafety Forms application, AgBooks, and the development of the NIH-funded Yogo Framework.

Associate Director, Research Software & Informatics, Senior Software Engineer [Sean Cleveland] Sean is responsible for the design and management of RCG software and serves as the primary contact for researchers who have contracted RCG software projects. In this role he supervises another software engineer (Thomas Heetderks). Sean contributes to all RCG software projects but his primary research effort is focused on the development of the VOEIS software (a Yogo Framework application), funded by NSF.

Senior Systems Administrator [Matt Harris] Matt is responsible for the administration and upkeep of the hardware infrastructure of the RCG. This includes the large-scale storage cloud, DNS servers, VM cluster, database clusters, and computational cluster. In addition to this, he is also responsible for managing the hardware associated with grant-funded research. Matt is also responsible for the configuration management and notifications system that stores information about RCG servers and notifies staff when there are problems.

Research Software Engineer [Thomas Heetderks] Thomas is an RCG Software Engineer responsible for applications programming on current software projects. While he is involved in all RCG software, his primary duties are currently focused on the VOEIS software application.

Additional personnel, for which no base funding is requested, are funded through contracts. These include:

Accountant and AgBooks Steering Committee Chair [Adell Westrick]

Administrative Assistant [Martha Sellers]

PROPOSAL SCOPE

A sustainable RCG will provide universal web-based access to the services needed to develop a successful grant proposal. Although most of the users will be faculty, staff and students at MSU Bozeman, we note that the services can be offered to anyone at any of the other MUS sites. For example, one of the largest RCG contracts has come from researchers at U.M. working on their component of the previous EPSCoR Track 2 proposal: there is no unit equivalent to the RCG at UM, and UM researchers contracted with the RCG due to the many benefits summarized above. But even more significant within the context of **broader impacts** is the **provision of cyberinfrastructure support to faculty at the tribal colleges**. For those sites, access to the services provided by the RCG is not simply a matter of cost-effectiveness, but a matter of necessity: the expertise required for the development and implementation of an effective data management plan simply does not exist at those sites. The RCG's ability to carry out remote web-based delivery of cyberinfrastructure services to UM research groups (which include support of data being collected at remote sites around the state, including Flathead Lake) validates their capability of supporting service to researchers at the tribal colleges. In fact, the current RCG director (John Miller) is working directly with Sara Young, and several contacts designated by her, in the development of the new EPSCoR Track 2 Cyberinfrastructure preproposal to MAKE CERTAIN that RCG services will be accessible to researchers, staff and students at the tribal colleges.

Beyond this, to the extent that every research grant to a national funding agency now requires a clearly articulated statement of **broader impacts**, as well as a section detailing a specific **data management plan**, funding for this proposal would establish a service center that would, in essence, leverage the broader impacts of every research grant that goes through MSU.

ADDITIONAL INFORMATION

The Research Computing Group has been fully functional for 3 years in its current operational scheme. Implementation will simply be the continuation of the current scheme, but with a more secure funding base. The only aspect of continuing operations that would require negotiation would be the precise definition of tasks that would be offered for no cost to the faculty and administration.

Note that the Director and all technical personnel have strong links to, and strong confidence in, the ITC. If it would make sense at some future time to re-implement and integrate the RCG as a unit within ITC, the RCG Director and Staff would be amenable.

The Director currently reports to the CIO and VPR. Evaluation depends upon several performance metrics, two major metrics being a) success at maintaining a sustainable level of contracted funding, and b) satisfying the very objectively-specified expectations and "deliverables" of those contracts. Dr. Miller and the RCG staff are currently in the process of developing a more systematic and comprehensive assessment plan that includes a survey instrument to a) assess user needs, b) solicit and analyze constructive feedback and criticism from the contracting units. We have just constructed a "test survey" to be distributed to all research participants in the new EPSCoR-funded Institute of Ecosystems (IoE), which could be provided on request.

SIGNATURES

Signature	Date
John Miller	
Dept Head Priority (please circle one): Very High High Medium Low Very Low	
Dean/Director Priority (please circle one): Very High High Medium Low Very Low	
Executive/VP	Date
Tom McCoy	1/11/12
Executive/VP Priority (please circle one): Very High High Medium Low Very Low	