Computationally Intensive Research Electronic Proposal Guide

Deadline for receipt of complete proposals for CIR is 5:00 PM PDT, July 9, 2008

A **Letter of Intent** must first be submitted and **accepted** by EMSL before a full proposal should be submitted. EMSL recommends the Team Leader review the Call for Proposals of Computationally Intensive Research projects (CRI; formerly known as Computational Grand Challenge Applications) in Environmental Molecular Science Research before submitting their team's proposal to the EMSL User Portal. Please contact the EMSL computing staff (mscf@emsl.pnl.gov or 509-371-6448) with any questions on the call for proposals or this form.

Proposal Submission Requirements. The Team Leader (or delegate) is required to submit the CIR proposal for their team. All proposals are required to be submitted electronically via the EMSL User Portal. No paper proposals will be accepted. All parts of the electronic proposal form must be completed and submitted together at the same time for a proposal to be considered complete. Incomplete proposals will not be accepted or reviewed. All proposals must follow the format specified below. Proposals that do not follow the instructions, omit required information, or deviate from the proposal format will be considered as incomplete proposals, and will not be accepted or reviewed. Once a proposal is submitted, it cannot be modified by the author. If changes need to be made, please contact EMSL computing staff (see first paragraph).

Before entering the EMSL Usage System (EUS), PIs should gather information regarding team members' names, affiliations, and other information (described below) for entry into the form. Part of the EUS includes the capability to attach large sections of your proposal as Adobe PDF files (preferred) or as Microsoft Word files. **The complete proposal must be submitted via the EMSL User Portal at http://eus.emsl.pnl.gov/Portal.** An acknowledgment for receipt of a complete proposal will be sent by e-mail to the Team Leader. Each research participant entered into the EUS will receive an EMSL user ID number if they don't already have one from a previous proposal.

The following information will be needed for each participant on the proposal:

- Prefix (Dr or Mr or Ms)
- First Name
- Full Middle Name (use NMN if no middle name)
- Last Name
- Suffix (if necessary)
- Primary Citizenship
- Profession (*i.e.*, Professional, Student, Postdoc, Research Scientist, Faculty/staff, Retired)
- Type of Institution
- Institution Name
- Business Address
- Country
- Telephone number
- Facsimile number
- Email address
- Funding agencies
- Curriculum Vitae (for PI and up to six principal participants)

The EUS is designed as a central web-based utility for establishing and tracking collaborations with EMSL and is entered by logging into the EMSL User Portal. As you go through the EUS, the first webpage will have general information and list open or saved proposals. For a new CIR proposal, click on the "Create New Proposal" button in the upper right. The actual proposal submission section is divided into three parts:

- **Participants** – needs to be repeated for each team member
- **Details** – research details are attached here including computer time needed
- **Logistics** – questions about funding and materials.

The first participant should be the Primary Author who is also referred to as the Team Leader or Principal Investigator. The Principal Investigator cannot be a Post-doc or graduate student. The second section includes details of the research, a place to attach the bulk of the proposal, and anticipated computational needs. The third section contains questions about funding, materials to be used, and any additional comments. CIR projects are generally non-proprietary in nature with results to be published in the open literature. Users engaged in proprietary research at EMSL are obligated to pay full cost recovery for their use of EMSL.

If you need to save your work and return later to complete it, then use the "Save" button at the top of the page. [To return to the saved proposal if you have closed your browser, go to http://eus.emsl.pnl.gov/Portal and log in again. Select from the list of saved proposals the proposal you want to return to.]

A step-by-step description of what data needs to be entered into the EUS starts below. Please review the requirements and have information and documents ready **before you begin**. The Adobe PDF or Microsoft Word files should be prepared ahead of time for quick attachment. **Please follow the maximum page limits for each section. All nine sections listed below under "Attachments:" must be included for the proposal to be complete.**

Before you enter the EMSL Usage Proposal system, you may need to agree to the Terms and Conditions for Using EMSL. The URL for the EUS is **http://eus.emsl.pnl.gov/Portal.** After reading the terms and conditions, and you agree with the terms and conditions, select "I Agree" to enter the EUS.

Participants

Research Team Leader:
Please enter the information listed below for the Team Leader first. A separate window will open asking you to "create new user" or select "returning EMSL user." If a new user, enter required information (see below); if a returning EMSL user enter User ID or if EMSL user ID is unknown, click on "Forget your ID?" and enter name to pre-fill returning EMSL user information. You will also be asked if the team member plans on visiting the EMSL. Please respond accordingly. Once you have added the primary author (team leader), select the "Add Another Participant" to enter information for each additional team member. Postdoctoral and graduate students only need to be included if they will need computer accounts. After all team members have been added, then select the "Continue" button at the bottom of the page.

Additional Team Members:
If a previous EMSL User:
Enter EMSL User ID Number or click on "Forget your ID?" and enter user name to pre-fill the information. Check the information shown for accuracy and make the necessary corrections.

If new EMSL User:
Enter the following information for each team member.

Proposal Participant

Prefix:

First Name:

Full Middle Name: (no initials; if no middle name, use "NMN"

Last Name:

Suffix: (if necessary)

Profession (*i.e.*, Professional, Student, Grad Student, Postdoc, Research Scientist, Faculty/staff):
[If a student under the Office of Fellowship Program (OFP) or the
DOE Science Undergraduate Laboratory Internship (SULI), select "Yes" }

Telephone number:

Facsimile number:

Email address:

Institution Information

Country:

Type of Institution:
{Depending on the type of institution, you may get a pop-up window}

Institution Name: (select from pop-up window if one appears)

Department: (if applicable)

Business Address:

Details

Proposal Description

Proposal Title:
The first field is the title for the proposal. This should be short and concise.

Proposal Abstract [Maximum 2 paragraphs]:
This should be about 500 words in length, briefly describing what will be done, include anticipated results. This can be cut and pasted from a previously prepared Word document. Do not duplicate in the main body text.

Proposed Research (.doc or .pdf):

This is where the bulk of the proposal will be entered. This can be either a single document or several smaller documents. They will be attached to the EUS proposal. A description of each section follows. The maximum

length for each text section is shown in brackets. Requested benchmark information on scalable performance for parallel codes must be reported in terms of speedup factor versus number of processors and be based upon real elapsed wall clock times. Specifications of requested computer resources must be given in terms of node-hours (each node has two quad-core processors).

1. Project Definition [Maximum 2 pages]:
Define project and objectives including background and relevance to the environmental problems and research needs facing the U.S. Department of Energy and the nation. What is the "Computationally Intensive" aspect of the project? Into which EMSL Science Theme(s) does the research fit?

2. Proposed Work [Maximum 8 pages]:
Describe the proposed work in enough detail to justify resource allocation. Describe the computational methods/approach that will be used. Give expected outcomes/results from the project and impacts on DOE environmental mission goals and/or national environmental research needs. Include a project work plan (in six month periods) for the duration of the proposed project (maximum of 3 years). We expect more detail at the beginning of the project and less detail at the end.

3. Software [Maximum 3 pages]:
Describe the software to be used in enough detail to demonstrate efficient scalable utilization of EMSL's computing resources for parallel calculations starting at 16 processors and going higher. If the code is not parallel (or does not efficiently scale), estimate how much computer time will be required to develop an efficient, scalable parallel implementation of the code. If the code is parallel, provide benchmark information in enough detail to demonstrate the performance and scalability. Benchmark information must be reported in terms of speedup factor versus number of nodes and be based upon real elapsed wall clock times. If software development/porting efforts are planned, identify the length of time projected for that work and the amount of computer resources needed. Identify any available software (*e.g.*, NWChem, ECCE, parallel tools and libraries) that your project proposes to use. EMSL does not guarantee the ability to provide any external third-party software for a project or user. Users will be required to provide a valid executed software license agreement (or provide suitable public domain documentation) before being able to install software on any EMSL computer system.

4. Proposed EMSL Computational Resources [Maximum 2 pages]:
Estimate the EMSL computational resources (in node-hours) required for the proposed work for each six-month period in the work plan. For Chinook there will be 2 quad-core processors per node. Be as specific as possible (*e.g.*, 64 nodes x 250 wall clock hours = 16,000 node-hours) and state the basis for your estimate (*e.g.*, runs of previous jobs on MPP2 or other parallel computer system). Remember your code may not be able to use all 8 cores on a node effectively. We plan to have benchmarks available for NWChem on Chinook. For other code, use your best estimate. If the basis of your estimate is from results on a computer system that differs from the EMSL resources, use the ratio of the theoretical MFLOPS of the processors to convert to node-hours based on 6 or 8 cores/node. If development/porting of software is planned, also include the computational resources required for it. *The requested computational resources specified here must sum to match those given in the* **Logistics** *section of the proposal.*

5. Proposed EMSL Computing Storage Resources [Maximum 1 pages]:
Describe the type and amount of storage needed for the proposed project on a yearly and cumulative basis. Distinguish between short-term and long-term disk storage. Identify what, if any, types of data in the NWfs

database your project will make freely accessible to other EMSL researchers and/or publicly available. *The storage resources specified here must be consistent with those given in the* **Logistics** *section of the proposal.*

6. Access to Other Computational Resources [Maximum 1 pages]:
Identify and quantify the other computational resources the team will have for accomplishing the proposed project. Note that calculations that do not scale efficiently with large numbers number of processors or that can be done effectively with less than 16 processors are expected to be run primarily on smaller computers not part of EMSL's supercomputing resources. Teams that are planning to acquire and use EMSL's ECCE software should identify the types of desktop workstations used (including versions of the operating system). Participants at PNNL can have access to the Graphics and Visualization Lab computers and workstations.

7. Requested Support Resources [Maximum 2 pages]:
Identify and describe any computing support resources that are needed for the proposed project (*e.g.*, specific features/functionality not presently in NWChem or ECCE, guidance in using the parallel tools and libraries to develop efficiently scalable parallel software, etc.). Note that porting and parallelizing of purely sequential code is not likely to be supported.

8. References:
Include the appropriate reference citations for the proposal.

Curriculum Vitae for Team Leader and Team Members [Maximum 2 pages each].

Include curriculum vitae for the Team Leader plus up to **five** of the team members. Not all team members need to have a vitae. The vitae should include recent publications relevant to the proposed research.

Preferred Start Date:
Choose a date (month, day, and year) when you expect to begin work. The earliest date is October 1, 2008 since this is when we plan to have computational resources available.

Preferred End Date:
Choose a date (month, day, and year) when you expect to complete the work. For a 3-year proposal this would be September 30, 2011.

Primary Research Area:
Choose the research area most closely associated with your proposal. These areas are most in line with the DOE's research directives. If none of these listed are appropriate, select "Other" and specify it in the next field.

- ☐ Biological and Life Sciences (excludes medical applications)
- ☐ Chemistry (excludes materials chemistry)
- ☐ Earth Sciences
- ☐ Engineering
- ☐ Environmental Sciences
- ☐ Materials Sciences (incl. condensed matter phys. & materials chem..)
- ☐ Medical Applications
- ☐ Optics
- ☐ Physics (excludes condensed matter physics)

- ☐ Polymers
- ☐ Other - Specify

What type of access are your requesting?

☑ **Standard** ☐ **Rapid**

Select standard.

What type of proposal are you submitting?

☑ **General** ☐ **Partner**

Select general, partner proposals require prior interaction with EMSL staff in a letter of inquiry.

What is the proprietary status of your proposal?

☐ **Proprietary** ☑ **Non-Proprietary**

The proposed research will likely be considered non-proprietary, unless otherwise specified. CIR projects are generally non-proprietary in nature with results to be published in the open literature. Users engaged in proprietary research at EMSL are obligated to pay full cost recovery for their use of EMSL. Since EMSL's computers are fully subscribed with non-proprietary projects, proprietary research on EMSL computers is not an option.

Which call are you responding to?

Choose from the list "Grand Challenge: Computation"

Is this proposal associated with a National Science Foundation Supplemental Funding Request?

☐ **Yes** ☐ **No**

Select yes if it is, otherwise select no. Pretty straight forward.

Will you desire the assistance of EMSL staff in obtaining and interpreting the results?

☐ **Yes** ☐ **No**

Select yes if you anticipate the assistance of EMSL staff, otherwise select no. We are here to help.

PNNL Staff Contact:

If you have discussed this proposal with a PNNL Staff member, please enter his/her name.

EMSL Resources

Primary EMSL Facility:

☑ **Molecular Science Computing - Vorpagel, Erich R**.

Since you will be primarily using the HP Linux cluster, select the MSCF.

EMSL Facilities List:

For each facility within EMSL, select the facility (*e.g.,* Molecular Science Computing") from the list and then select from the **Resource List:** an MSCF Resource. First select the "Computing: Chinook (HP 2310-Node Linux Cluster"

Estimated Hours of Usage:

Click the mouse on this field and enter the total number of node-hours requested for the first year. (Allocation awards are expected to be on the order of 100,000-2,000,000 node-hours for a one-year period). For example: "750000" then, click the "Add Resource" button and the selected resource request will appear in the "**Selected Resources:**" field. These amounts should correspond to the time estimates in Section 4 (Proposed EMSL Computational Resources) of the main proposal. Repeat this for any other instrument you will need.

If you make a mistake, just click on the wrong entry in the "**Selected Resources:**" list then click the "Remove Resource" button.

Next select the "Computing: Data Storage Archive (NWfs)" in the "**Resource List:"** Enter the estimated amount of Disk Storage space in Gigabytes for the first year. The resource request will appear in the field below along with the computing hours. These amounts should correspond to the time estimates in Section 5 (Proposed Computing Storage Resources) of the main proposal.

When you have completed entering all the data, select the "Continue" button at the bottom of the page.

Logistics

Finally, we ask a few questions regarding funding agencies and whether you will need additional equipment or if you intend to bring any equipment or materials to the EMSL as part of this proposal.

Proposal Funding

Funding Agency (Agencies):
Use the pick list provided in the EUS to select all the agencies from which team members are funded. If "other", then please specify. Multiple selections can be made by holding down the Ctrl key while clicking with mouse.

Work Package #:
If the Primary Author (Team Leader) is a PNNL employee, include a work package number for reasons described in the EUS.

Materials & Equipment

This section probably will all be answered with a No response. But we need to ask for your response. It is possible that you will be including experimental components in the proposal that require one or more of the following:

Will your research involve the use of human blood, tissues, DNA, cells, cell lines, or human biological samples in any form?

☐ **Yes** ☐ **No**

Does this work involve the use of live animals?

☐ **Yes** ☐ **No**

Will you be bringing or sending any *chemicals* to the EMSL facility?

☐ **Yes** ☐ **No**

Will you be bringing or sending any *samples* to the EMSL facility?

☐ **Yes** ☐ **No**

If you intend to bring any equipment to EMSL as part of this proposed research, please list it in the **User Equipment:** field. If you are bringing computers that will need to connect to the PNNL network, please list them as well.

Comments/Additional Needs

If you have any additional needs or comments regarding the proposal or the process, please enter them in the **Comment:** field.

If you need to save your work and return later to complete it, then use the "Save" button at the top of the page. [To return to the saved proposal if you have closed your browser, go to http://eus.emsl.pnl.gov/EUS (as if you were going to submit another proposal), enter your User ID (or click on "Forgot your ID?") and select from the list of saved proposals.]

When you are finished, select the "Submit Proposal" button at the bottom or top of the page. **Remember, once the proposal is submitted, you can no longer modify it.** Modifications can be made by the Capability Steward or EMSL Host if necessary. Just send an e-mail to mscf@emsl.pnl.gov with any modifications.

Congratulations! You are now finished.