Western Consortium of Idaho, Nevada, and New Mexico

To: Track 2 External Advisory Committee: Kelvin Droegemeier, Janice Gobert, Alan Hevner, Henry Neeman, Cherri Pancake, Jordan Powers, Robert Sherwood, David Tarboton

From: **Track 2 Leadership Team**: Gayle Dana, Bill Michener, Mary Jo Daniel, Peter Goodwin, Rick Schumaker, Karl Benedict, Nancy Glenn, Fred Harris

Re: Responses to Findings and Recommendations from the External Advisory Committee site visit of the NSF EPSCoR Tri-State Cyberinfrastructure Project (Track 2) on February 25, 2010, in Las Vegas, NV

Date: May 4, 2010

The External Advisory Committee (EAC) conducted a site visit and meeting at the Desert Research Institute in Las Vegas, Nevada on February 25, 2010 in order to review the progress to date on the Tri-State Consortium, NSF EPSCoR project, "Cyberinfrastructure Development for the Western Consortium of Idaho, Nevada, and New Mexico." The EAC submitted a report of their visit and findings and recommendations for the project.

The Leadership Team of the Tri-State CI project very much appreciates the valuable feedback and recommendations of the EAC. We have responded to each recommendation with strategies on how we plan to incorporate these recommendations. The responses are embedded after each recommendation in the EAC's February 25, 2010 Report. If after reviewing our responses, the EAC has any concerns about our approaches, please do not hesitate to contact Gayle Dana (775-674-7538; Gayle.Dana@dri.edu).

External Advisory Committee (EAC)

NSF EPSCoR Tri-State Cyberinfrastructure Project (Track 2)
Findings and Recommendations
February 25, 2010

Introduction

The External Advisory Committee (EAC) for the NSF EPSCoR Tri-State Cyberinfrastructure Project (Track 2) met with the project team on Thursday, February 25, 2010. The EPSCoR states of Idaho, Nevada, and New Mexico received a collaborative award of $6 million from the NSF EPSCoR program for the project "Cyberinfrastructure Development for the Western Consortium of Idaho, Nevada, and New Mexico." The funding is for three years and commenced at the end of 2009. Thus, the project is in its beginning stages and is in the process of defining its activities in order to achieve project goals in the research area of climate change science. The team gave the EAC presentations on the current status and plans for the project. The three high priority objectives of the project are:

1. Increase connectivity and bandwidth within the consortium states
2. Enhance data and model interoperability within climate change science
3. Utilize CI to integrate research with education

The EAC engaged the project team with many questions and suggestions. Extensive discussion on project opportunities to achieve the above objectives ensued.

This report summarizes the findings and recommendations of the EAC. We begin with general observations on the overall project and then break out observations, questions, and recommendations on the three project goals.

General Project Findings and Recommendations

The EAC is impressed with the capabilities of the states' project teams to eventually succeed in this project. The presented project materials address the opportunities of the project with an achievable project management plan, although the particulars of the science activities were a little vague. The team members in the meeting were very open to the suggestions and ideas proposed by the EAC. We came away excited by the potential of the project to make important impacts to CI for climate change research and education in the tri-state region.

A key issue to a full understanding of the Track 2 project is the relationship of its CI goals to the research goals of the ongoing Track 1 projects. Much discussion dealt with how STEM research and education will impact, and will be impacted by, the development of CI across the three states. We agree with the NSF recommendation that the project must keep its focus on the science to be supported by the CI.

In line with this focus on climate change science, the EAC is not necessarily looking for research contributions from the CI designs and implementations. However, the CI must support and encourage innovative research and education ideas, for climate change projects in particular and STEM in general, both within the three states and beyond to national and international research groups.

Recommendations:

1. Build initial momentum by framing the work in terms of real STEM research and education problems and needs. Use case studies and examples as a foundation for describing your project's concrete contributions.

   Response: The model and data interoperability group is going to prepare an Innovation Working Group (IWG) proposal to support three workshops with targeted tri-state stakeholders: climate science researchers, modelers, and educators. The IWG program is funded by the three states to support working groups that emphasize the collaborative development and testing of important ideas and theories, cutting-edge analysis of recent or existing data and information, the use of sound science policy and management decisions, and investigation of social issues. The proposed workshops will provide the needed case studies for the initial development of the CI capabilities that will support the science and education goals of the three state's Track 1 projects. These case studies will define the user-centered capabilities that are required for the three state projects. These capabilities will be built upon the existing and planned standards-based systems already under development as part of both the Track 1 and Track 2 projects.

   In addition, all of the outreach components in the Cyberlearning Component use STEM research and education problems as case studies. For example, the cyberlearning material development for MS/HS students involves the use of climate data from the Track 1 science projects.

2. To provide future sustainability of the CI projects commenced under this project, it is important for the team to produce clear and convincing deliverables that demonstrate success to the stakeholder communities of researchers, educators, government, industry, etc. Engage potential champions from each of the stakeholder groups to participate and to advocate for future funding.

   Response: As part of both the proposal process and the project's strategic planning process in fall 2009, we developed deliverables for each component. We will work to engage potential champions from each of the stakeholder groups as we make progress on producing the deliverables.

   Also, the capabilities developed as an outcome of the case study workshops described in the response to 1) above will be evolved in an iterative fashion with the relevant stakeholders so that they meet both the immediate goals of the current EPSCoR projects, as well as establish an ongoing dialog between the CI team and targeted climate scientists and educators. It is expected that this ongoing interaction will produce new collaborations that will result in new projects and continued support for the maintenance and continued evolution of the capabilities developed as part of the EPSCoR projects.

In addition, the Cyberlearning team is currently pursuing additional funding through NSF and leveraging partner organizations (e.g. New Mexico SuperComputing Center).

3. The project team should be fully aware of and, as feasible, participate in international and national groups developing standards, frameworks, toolsets, and other capabilities in the areas of data/model transmission, storage, and interoperability. The project leaders must effectively select and adapt the most effective standards to the CI design and implementation.

   Response: Individual model and data interoperability group members are involved in national and international standards and interoperability groups and programs such as: EDAC (UNM): Open Geospatial Consortium, Global Earth Observing System of Systems (GEOSS), Federation of Earth Science Information Partners (ESIP Federation). Individuals will continue these involvements as well as integrate into their work explicit information sharing about and, as appropriate, outreach to relevant interoperability organizations, including detailed discussions about the standards developed or recommended by these organizations.

   The Cyberlearning team is leveraging international and national groups through each members own expertise/research focus in their field. For example, we are leveraging CUAHSI initiatives and OpenTopography interoperability in our cyberlearning material development.

Increase Connectivity and Bandwidth: Findings and Recommendations

The connectivity objective will provide substantial improvements to high performance networking capacity in Idaho and Nevada and will establish cyberinfrastructure “gateways” (video conferencing facilities) in New Mexico. One of the challenges in evaluating these activities is that, due to space constraints, the materials provided haven’t been able to include a substantial degree of specificity. The EAC’s key recommendation is that significant detail be collected and analyzed on specific objectives, milestones, equipment to be purchased, and especially the list of specific research and education projects that have been becalmed for lack of the intended resources and what these projects will be able to achieve once these resources are deployed (specific needs and problems solved).

Response: Data collection and assessment of needs was done several years ago, in preparation for the writing of our Track 2 proposal, on a state-by-state basis. This assessment included bandwidth needs, and scientific research needs. In Nevada it was done by NevadaNet (as explained below), in New Mexico it was done by NMCAC (as described below), and in Idaho it was done by the state EPSCoR committee in collaboration with the Higher Education institutions. Regarding the milestones for the project, each objective (with the equipment purchased to meet it) has a data collection plan in place as part of the proposal assessment.

We also recommend that the project team explain more fully how the tri-state's interstate connectivity will be improved by these activities and whether this kind of tri-state collaboration is currently limited or already well underway. For example, does the current connectivity between Nevada and New Mexico limit extant or emerging research and education endeavors? State-specific questions and recommendations are presented below.

Response: Without the bandwidth increases provided by the Track 2 funding, the collaboration among states in the Track 1 and Track 2 projects would be severely limited: in Idaho key researchers would still have slow networks; in Nevada the data pipe into the state would be full; and in New Mexico, as described below, the collaboration within the state would be limited. On the networking side, this will impact the interoperability component of Track 2 and the data portal components of the Track 1 projects. These then cascade to impact the researchers collaboration and data sharing and proceed all the way to the education components, which will use the collected and modeled data in their training and outreach.

Idaho: In addition to the connectivity into each participating institution, it would be useful to know the connectivity within these institutions.

Idaho Response: Idaho State University (ISU) partners with the Utah Education Network, which gives it access to Internet2 and the National Lambda Rail (NLR) and membership in the Front Range GigaPoP for 310Mbps connectivity. The University of Idaho (UI) is linked to Internet2 via the Pacific Northwest GigaPoP (through Seattle) with an OC-48 (2.4 Gbps). In the fall of 2009 UI Information Technology Services (ITS) invested in hardware to provide connection at 10 Gbps to the Idaho Regional Optical Network (IRON).

Boise State University (BSU) recently connected to Internet2 and NLR through the Idaho Regional Optical Network (IRON). IRON Associates and organizations eligible for Associate status have access to high-speed (10 Mbps to 10 Gbps) IP bandwidth and L2VPPN connectivity to the commodity Internet, Internet2, and NLR. Switches will make connections from the engineering building via the college to the BSU campus (at 1 Gbps). From campus it currently goes to the outer world by 400 Mbps as one connection; the other route is 1Gbps from campus to IRON, and from IRON to I2NLR is 256Mbps.

INSIDE Idaho tested network capabilities and confirmed transfer > 600Mb/s on the UI Moscow Campus. AT UI, connections to the core range from 10 to 100 MB/s, with growing number of 1Gbps connections. Approx 80% are at 100 Mbps. Within campus the infrastructure will easily support up to 10Gbps as needed.

The Kimberly Research Station in southern Idaho is connected via frame-relay through Boise's T1s at 768 Kbps, burstable to 1.5 Mb. This slow connection is not sufficient to allow efficient research data transfer. Hagerman research facility uses 2 T1 lines to provide a 3Mbps connection through a commercial provider.

At ISU, prior to the blades being installed in the core, connectivity to the buildings (that are the subject of the Track 2 upgrade) was 100Mbps uplinks Intra and Inter-building. The upgrade now allows uplink connectivity to those buildings at 1Gbps.

Connectivity leaving campus is 155Mbps to our Qwest Commodity Internet provider and 155Mbps to ISU's UEN I2 provider (310Mbps total). Additional bandwidth to Commodity Internet is being investigated as lower prices become available resulting from the Idaho Education Network initiatives.

No bandwidth shaping or limits are placed on Inter-building or Intra-building connections. All upgraded network access ports (workstations & servers) have full access to wire speeds afforded by the new 1Gbps ports and uplinks. Bandwidth shaping only takes place at the Internet borders.

Idaho: The Track 2 proposal rightly highlights the need to support "key university researchers' labs and desktops," so it's crucial to establish precisely how will this be accomplished once connectivity into the institution is available.

Idaho Response: ISU has installed the network backbone improvements and the core network switches in order to upgrade four buildings. Physical Sciences is complete (Geosciences); the remaining three buildings (Biology, Colonial (Engineering), and Graveley (GIS Center, Anthropology)) are next.

ISU has installed the network backbone improvements (Gig Blades) into two WS-C6509 core network switches in order to provide 1Gbps uplinks to four buildings. Fiber optic work is in process to increase the backbone connectivity from the core at 100Mbps to 1Gbps into those buildings (the blades installed in the core will allow additional buildings to be connected as funding becomes available).

1Gbps links have been established between the (2) WS-C6509 core network switches to improve transport speeds between upper and lower campus.

The Geosciences offices and labs (Physical Sciences Building) have been upgraded with new 1Gbps access switch ports to increase connectivity from 10Mbps to 1Gbps for desktops and servers.

The Physical Sciences / Chemistry building (lower campus) is waiting for additional fiber optic work to be completed to connect the building (uplink) to the new 1Gig blades in the core.

The remaining three buildings (Biology (upper campus), Colonial/Engineering (lower campus), and Graveley/GIS Center/Anthropology (lower Campus) will be completed after the Physical Sciences Building. All four buildings need fiber optic work to complete the connections.

We will, under the scope of this project, upgrade select access ports from 10Mbps to 100Mbps as identified by Geosciences.

The new network capabilities will be leveraged in the ISU GIS Training and Research Center to improve scientific collaboration and the exchange of large geospatial datasets with rangeland scientists at other Universities and institutions. The increased speed availed by this network improvement will greatly improve data sharing and the real-time exchange of ideas and information especially as increasing amounts of data are transitioning to cloud services.

In the ISU Department of Geosciences, new gigabit network capacities are essential for transferring large topographic and remote sensing datasets (both traditional and LiDAR) between the community server and the desktop users. These desktop users included graduate and Post-Doctoral researchers, faculty in various departments and buildings as well as students in computer labs. The previous system inhibited sharing of large files and thus multiple versions of the same data were present on many local machines, preventing rapid sharing of GIS progress.

Idaho: For example, will researchers need to pay higher monthly charges for higher bandwidth, and if so, how will they pay (e.g., funded by grants, by departmental internal funds, or by upcharges waived as institutional commitment)?

Idaho Response: The connectivity improvements being implemented under the Track 2 award are based on strategy to invest in sustainable improvements. Universities are committed to maintaining the hardware upgrades resulting from the awards. Connections to IRON and related investments are being integrated into the institutional budget.

Some at UI Kimberly and Hagerman will probably need to pay more. The directors both facilities are aware that if any additional charges are necessary, beyond their current budget, they will be responsible for identifying appropriate funds. They are aware the Track 2 funds are available only for a short period of time and that it will most likely be used to build infrastructure that will bring a higher speed connection to their facilities. That is why we are taking a very careful approach to the Request for Proposal (RFP) for connectivity solutions. The RFP will allow a broad spectrum of solutions ranging from little investment required upfront to significant investment required upfront, and ranging from very affordable and expandable amounts of bandwidth to very expensive amounts of bandwidth in the beginning. We will be asking for flexibility in monthly payments depending on how much money can be made available to purchase infrastructure.

At ISU the improvements are hardware solutions with a one-time cost. The switches have a useful life expectancy of 3-5 years after which they will need to be replaced with whatever technology is current at that time. The switches would be replaced by institutional funds when the time comes to replace/upgrade.

Idaho: What is the expected level of use of these capabilities?

Idaho Response: Researchers who will benefit from the improved connections are involved in a variety of activities related to the theme of the Tri-State Consortium. The improvements range in the nature of their specificity to particular faculty. At

BSU, investments are directed to serve the needs of Dr. V. Sridhar to access and utilize remote computing facilities at INL. At Hagerman, they benefit the research group at the aquaculture research station. At Kimberly, they benefit the Hydroclimatology and remote sensing work led by RII Track 1 leader Rick Allen. At ISU they benefit multiple faculty (many involved in RII-supported research) in several departments.

The minimum amount of new bandwidth at Hagerman and Kimberly will be 10Mbps or more. Kimberly currently has 768Kbps burstable to 1.5Mbps and Hagerman has 3Mbps. Again, UI will "prefer" solutions that are scalable well beyond 10Mbps, to 100Mbps and beyond. But as mentioned above, if the scalable solutions are not sustainable from the facilities perspective, they will most likely not be selected during the RFP.

Idaho: What are the leadership projects that have been waiting for this capability?

Idaho Response: At BSU, Dr. V. Sridhar and his group have been heavily accessing the Idaho National Laboratory High Performance Computing (HPC) from BSU through Internet connections. They have been testing the simple version of newly parallelized VIC model. For the last six weeks, they have been working on this process and their test runs are successful. But it is the parallel upgrade for only surface hydrology part of the hydrology model at the moment. The Track 2 funds will help to improve the efficiency of their connections for faster up/downloads.

At UI Hagerman research is moving into a 'next level' stage of molecular biology that involves measuring gene expression across the entire genome of fish subjected to various treatments, such as diet, water temperature, etc. This involves an element of 'cloud' computing in that the expensive software we bought runs on servers operated by the software developers. The challenge with this program is that the pyrosequencing done on fish from their research studies generates a tremendous amount of data. The sequencing is done in Moscow. They need to have the capability to send and receive data, plus do bioinformatic computing on the cloud servers.

Also, UI partners, CRITFC, are co-PIs with Idaho Dept. Fish & Game on a new project that involves genotyping all of the salmon and steelhead returning to Idaho hatcheries, an average of 16, 500 fish per year. They will do this at Hagerman. Genotyping is done using SNPs, single nucleotide polymorphisms. SNPs are sequences of genes that differ among populations and individuals. To get a good differentiation between stocks and individuals will require either 96 or 192 SNPs per fish (one or two plates). This $5.5M project will utilize cyberinfrastructure and will require the Hagerman Station to have the capacity to move tremendous amounts of data, plus access this data for analysis.

Researchers in Hagerman are going to be more involved in classroom teaching as well as other short courses, some with Columbia River Inter-Tribal Fish Commission (CRITFC). This will require a higher quality of video than they now have. Their video equipment is fine but our bandwidth is inadequate. Among other things, they will be starting a collaborative project with the Washington State University College of

Veterinary Medicine in which we will offer a rotation in aquatic animal health for 4th year vet students. This will begin in spring semester, 2011, and will involve video teaching as well.

Idaho: Also, given that some of the funding will be spent on a subscription, how will that be sustained after the project period?

Idaho Response: As stated earlier, sustainability is a key factor in the implementation plan. In many cases, the highest cost is associated with the initial investment. The connections and associated fees will be paid by the university as a part of the ongoing IT network service structure.

At UI, for example, they are forcing the facility directors to look to the future sustainability of the potential solutions before we collectively make a decision. Using the grant to heavily subsidize the increased monthly fees is not a sustainable approach. Also, both directors plan to pursue additional funding and one of them is able to increase the amount of money his facility spends on bandwidth with currently available funds. Our implementation approach is to have multiple (and sustainable) solutions to select from in the RFP.

Nevada: What is the list of key projects that expect, either individually or collectively, to need 10 Gbps connectivity rather than the current 2 Gbps? What opportunity costs have been associated with the lesser level of connectivity that will be resolved by this improvement?

Nevada Response: All of the connectivity upgrades proposed in the Track 2 proposal were prioritized by NevadaNet. NevadaNet, a division of the Nevada System of Higher Education (NSHE), is a statewide backbone network that provides data connectivity and runs the statewide K-20 education network and videoconferencing into and within the state. NevadaNet provides prioritization for all proposals that include connectivity upgrades.

The Connectivity between Nevada and the outside world before this proposal was a 2Gbps Internet 1 (I1 or Commodity Internet) pipe in the South and a 2Gbps shared I1 and Internet 2 (I2) pipe in the north. The I1 pipe in the south had reached more than 85% saturation and was estimated (based upon conservative growth rates over the past few years) to reach 100% saturation before the Track 1 proposal had finished (next 5 years). The I1/I2 pipe in the north (to CENIC) had also reached 85% saturation and was also projected to reach 100% saturation at approximately the same time.

In the North the outside I1 and I2 connectivity is contracted through CENIC. The smallest pipe they offer is a 10Gbps connection. We have fiber that could handle 10, but hardware on each end that could only handle 2. Therefore, this proposal will change the hardware on each end of the fiber in the north to use the bandwidth currently being paid for. Without this connectivity improvement, large research projects such as NEES (and any collaboration with researchers outside of Nevada) would encounter difficulties in sharing their data outside of (and into) the state of Nevada in a timely fashion.

In Nevada, once this external connectivity is improved we will then work on the connectivity within the state. This will take several tracks. In the scope of this proposal we are upgrading network lines in the north (Reno to Elko, Reno to Ely) and then purchasing enhanced video conferencing hardware to increase connectivity within the state.

Nevada: What level of "last mile" connectivity is extant on the relevant campuses so that they can exploit the new capacity?

Nevada Response: The Networking connectivity between NevadaNet and the campuses is 10Gbs to UNR, UNLV, and DRI-South; 2Gbs to DRI-North; 2Gbs between the North and South; and at least 1 Gbps to all the other campuses in the state.

On Campus, DRI has 10Gbs main core routers, and 1Gbs to all the desktops (some have 10Gbs). UNR has a 10Gbs backbone with 1Gbs to most buildings (Geography is still 100 Mbs but should be upgraded soon). UNLV has a similar structure to UNR.

Nevada: What kind of new videoconferencing capability will be deployed and how will it be used? In particular, the committee would like to see some examples of projects that have expressed a need for this capability and what they will be able to accomplish that currently isn't possible.

Nevada Response: NevadaNet handles all compressed video conferencing across the various campuses in Nevada. The new equipment proposed will allow better quality video and more connections across the state. Current use includes distance education at both Universities and community colleges across the state as well as meetings at a variety of levels (faculty can schedule ad-hoc meetings, research groups schedule bandwidth for meetings, some academic departmental meetings are held via video, colloquia are broadcast on compressed video for several of the campuses, all the way up to board of regents meetings) for members across the state. As budgets are shrinking due to large shortfalls in the state budget, more emphasis is being placed on distance education and video meetings for research and administration.

New Mexico: "Gateways" at the various types of institutions (e.g. research, minority serving, tribal) aren't well-defined in the proposal. What will they be used for, and how?

New Mexico Response: The 20 Gateways installed to date are currently being used for:

- Statewide collaboration between colleges and universities as high-definition videoconferencing for meetings and distance education where students at multiple campuses are connected together taking a common course
- High-definition 3D stereo visualization as a decision theater in which complex problems (water systems, smart grids, traffic flow, etc.) are modeled in

advance of real-time changes in conditions to understand possible options in addressing an issue

- Scientific rendering/display/analysis of output (large scale datasets) for molecular dynamics, nanoscience, molecular genomics/computational biology [gene expression microarray data], particle physics, climate modeling, computational electromagnetics, computational mechanics, etc.

- Upper Hondo Water Availability and Decision Support Model for Lincoln County

- Providing computational services for the residents of Espanola area; teaching regular engineering courses that are part of the program at NNMC

- Dual credit High School programs to seven Hobbs area high schools

- Broadcast of online courses for Renewable Energy Curriculum, Electrical Applications, Energy Career Explorations; collaboratives with Louisiana Energy Services, Waste Isolation Pilot Plant, Enrichment Technology US, Waste Control Services; distance Learning for Nuclear Energy Education Program

<u>New Mexico</u>: Who at these various institutions has expressed a need for these gateways, and why do they need them?

New Mexico Response: In addition to the above, Gateways have identified the following future uses of the Gateway equipment:

- Engage business and industry to access and utilize the 3D capabilities for tele-health, data visualization, high definition 3D training, and professional development.

- Outreach/potential partnering with the NM Film industry (lab use to edit or send edits to studios), teleconferencing in support of NSF, DOE, and DoD collaborative science, teleconferencing in support of NSF, DOE, and DoD collaborative science, High School outreach and involvement in supercomputing, advanced visualization, and math/science (Cafe' Scientifique, SCC.)

- Articulation Conferences in Degree programs (have integrated Gateway capabilities into more than 40 degree programs of study at LCC), community use for local Health Agencies with remote locations, local events hosted by LCC, gaming competitions, distance learning with 15 Las Vegas area high schools, city to city conference meetings, workshops for community based organizations, qualification testing and recertification exams for local firefighters and Homeland Security.

- Professional development workshops via teleconference for continuing education in EMT Basic, EMT Intermediate I & II, EMS First Responder, CPR, Certified Nursing assistant, Hospitality and Tourism Management, Construction Trades, and Computer an Network Security Certification; develop programs in simulation and modeling, gaming, and digital video arts and entertainment at ENMU – Ruidoso.

New Mexico: What opportunity costs have been accrued without them, and what new abilities will be enabled by them?

New Mexico Response: In terms of opportunity costs, both ideas below are functions of having a large rural state and not having an infrastructure, such as the gateway infrastructure, to support statewide needs.

- Numerous programs have started at any one of the state schools only to have the program cancelled because the individual school did not have the critical mass at their institution to sustain the program. In 2008, the state legislature and the Governor signed into law the ability for an institution of Higher Education to allow their students to participate in distance education classes offered by other New Mexico institutions of higher education and to award the student the appropriate credit at their home institution. Not only has the state lost unsustainable programs at the Regional levels, but professors have gone underutilized in some cases and, in the worst cases, the professors have left the state to offer their expertise in more urban settings. The gateways provide the capability to address the costs lost to this phenomenon.

- The single largest rural economic development issue in the state of New Mexico is the lack of adequate broadband services to the rural parts of the state. Individuals in those communities have found it necessary to move to urban centers or they are unable to benefit from broadband access and, consequently, unable to contribute to the needs of their communities. The gateways and the communities they serve are largely Hispanic or tribal rural communities. The gateways serve as a major driver of expanded broadband for those communities. The opportunity cost of the brain drain in those communities because of the lack of access to broadband has been and continues to have major economic implications to those communities.

New Mexico: What software will be used to integrate the gateway components? What does integration mean in this context? A key issue for the EAC here is understanding the value proposition for this acquisition, especially in the differences between the needs at the research institutions and the needs at the primarily undergraduate, minority serving, and tribal institutions.

New Mexico Response: The software is to be determined by the NMCAC Education Committee – Budgeted at $99,960

NMCAC financial investment for 20 Gateway sites as of April 1, 2010:

Equipment	$ 839,380.00
40 Desktops	$ 50,040.00
QA & training	$ 63,820.00
Coordinator	$ 28,045.00
Software Budget	$ 99,960.00

Total NMCAC match contribution - $1,081,245.00

Enhance Data and Model Interoperability: Findings and Recommendations

The EAC felt that the issues of data interoperability and model interoperability are different enough to deserve separate sets of findings and recommendations.

Data Interoperability

Your group clearly has a good grasp of the state-of-the-art in this area. The RGIS software stack appears to use one of the best-in-class products at each layer and heavily leverages the open-source efforts of the last decade. The system's support for a wide range of the most common export formats is another strength; once data is uploaded and metadata created, data extraction and usage should be very smooth.

The project has a well-developed evaluation plan that shows considerable thought and depth. Further, your team is highly motivated and includes all the essential types of expertise. The group is also able to leverage a number of past projects and a strong experience base.

A key weakness is the fact that the success of the project's data interoperability efforts will ultimately depend on tri-state scientists' willingness not only to upload data but also to create high-quality metadata that will encourage others to utilize the data. From what we heard, the current scientific culture in your region, as elsewhere, is that metadata is not a well-understood or valued concept. The bottom line is that scientists don't want to have anything to do with metadata creation, so the burden is on data repository developers to make the creation process easy and to motivate it by providing some kind of key service that the user will not have access to unless he/she creates metadata. It puts your project at risk to be so susceptible to something that is outside your control.

The EAC suggests that you develop some "carrots" that will encourage scientists to upload data. You could begin with an exercise to identify what you think the value-add is for the user (e.g. long-term preservation and the ability to download in other formats, thereby saving oneself the trouble of doing the format conversions). Then be realistic about how to simplify and streamline the metadata creation. Consider issues such as: what are the minimal fields that are really required; how can tools simplify metadata entry by providing smart defaults or "remembering" values previously entered; and the availability of very clear examples alongside each field that will encourage users to provide high-quality metadata. Use this information as the basis for developing the metadata creation interfaces.

> **Response:** A previously planned instance of GeoNetwork Open Source has now been installed and subjected to initial testing at the Earth Data Analysis Center at the University of New Mexico. This system is planned for direct integration with the data discovery and delivery system under development as the foundation for the NM data portal. This integrated system will be available to all three states as an Open Source platform that supports the full suite of interoperability standards identified as key to the success of the CI work of all tree states. GeoNetwork will provides the ability to develop custom metadata templates that may be used to streamline and make more efficient the metadata creation process for partner scientists. The integration with the data discovery and delivery capabilities of the data portal will provide near-real time availability of uploaded data/metadata both for use by researchers and their collaborators.
>
> This rapid availability of science data provided by project researchers through the multiple standards-based interfaces provided by the data portal will support flexible use of those data - even by the researchers that created them - nearly instantly. The rapid publication of added data will also facilitate collaboration with other researchers through shared access to research data in near real time, without the researchers needing to do much more that upload their data and related metadata when they become available. It is anticipated that these capabilities will serve as "carrots" to provide an immediate reward to researchers that contribute to the shared data and metadata repository created by the EPSCoR programs.

A second weakness that the EAC notes is that some of the metrics for evaluating interoperability are off-base. Specifically, some of the metrics are measures of complexity, rather than value, such as:

- # of modules in the tool's design document
- # of functions/methods in the tool's API
- # of tables in the database schema document
- # of web services posted for data

The very presence of these metrics raises a red flag, in the sense that it implies complexity is a target — when, in fact, simplicity should be the target.

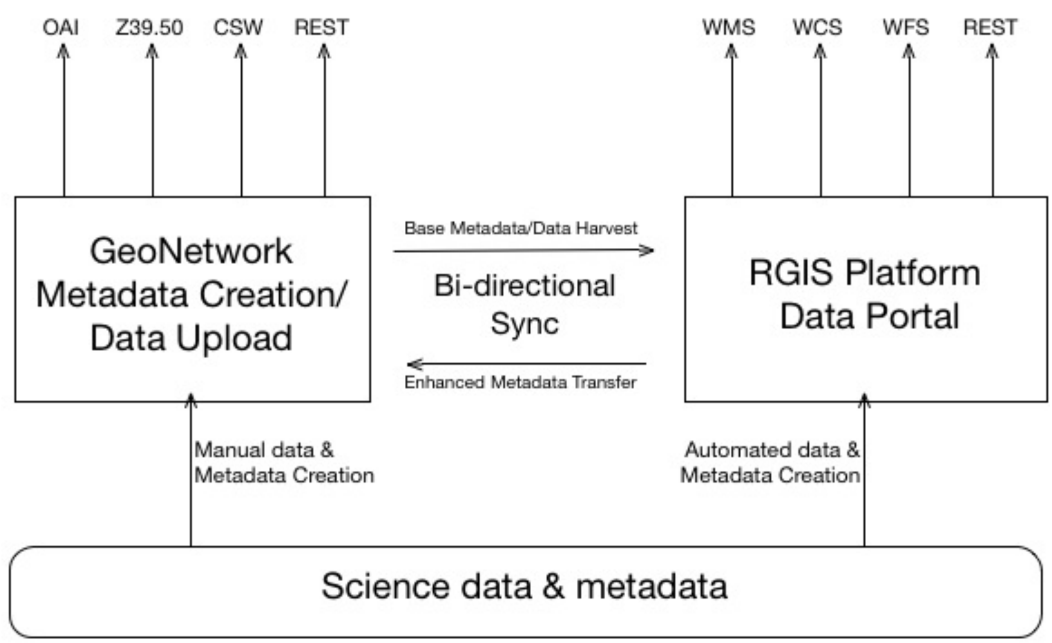
We suggest that you replace any metric that appears to reward complexity by one that focuses instead on software quality and efficiency, such as:

- # of software requirements met
- # of uploaded datasets for which metadata quality meets or exceeds our documented expectations (has the side benefit of requiring that your expectations for metadata quality be defined clearly)

Response: The data and model interoperability team will work with our Track 2 project external evaluator to refine the project metrics, with an explicit focus on the linkage between proposed capabilities and specifications and the metrics that measure progress towards their completion.

The EAC also noted that the project team lists many data and metadata standards, but has not articulated clearly how it will approach the specific organization or integration of these capabilities into a comprehensive and useful system. We recommend that this be done.

Response: The integration of the identified standards will be accomplished by each state using existing and planned capabilities within each state - but a specific instance of component integration (as being implemented in New Mexico) that is available to all three states as an integrated Open Source platform includes both the RGIS OGC and web services platform for the data portal and GeoNetwork Open Source as the metadata creation and catalog services platform. These two systems will be integrated through metadata synchronization, and data transfer from GeoNetwork to RGIS. RGIS will enhance base metadata added through the GeoNetwork interface with links to the available OGC and data download services hosted by the RGIS platform, and synchronize these metadata records with GeoNetwork for the catalog services hosted by GeoNetwork. This synchronization model will allow for manual data and metadata creation through GeoNetwork and automated data and metadata upload through the web service interfaces published by RGIS. This model is illustrated in the following figure.



Finally, although the project has a solid timeline of incremental development, we note that there is no real plan or mechanism for creating "success stories" along the way. The project could be missing a big opportunity to use incremental successes as a basis for motivating more participation by regional scientists and, perhaps, to increase chances of funding for any desired second phase of the effort. We suggest that you consider a plan for building incremental successes in the form of "case studies" that can serve two purposes: (i) demonstrate the value of your project to NSF and the participating institutions, and (ii) motivate wider participation in the project. For example, make a point of working with a couple of key scientists and making sure that it is simple and painless for them to upload their data into the repository, to create the associated metadata, and then to download the data in a different format and put the data to practical use (e.g. creating a visualization or integrating multiple data sets using Google Earth). Then help them develop a "testimonial" describing how easy the process was and how valuable the result was. Similarly, help a cyberlearning target group to access the data, to apply the data to a useful classroom activity, and to develop a descriptive testimonial. This type of case study can be extremely useful in a wide range of contexts where you want to demonstrate the value of your project or entice new users.

> **Response:** The data and model interoperability group is already performing initial work with targeted scientists in each jurisdiction, and in each case we hope to have early, demonstrable capabilities that can yield the results described above. For example, the New Mexico team is working with the lead Acequia researcher in importing the historic flow and meteorology data that are related to that particular research component. These data will be integrated into New Mexico's data portal as an early product that will be available through all of the interoperable service interfaces outlined in the diagram above.

Model Interoperability

In evaluating the model interoperability component, the EAC struggled with the specificity presented of the planned activities: it was too vague for us to evaluate the specific problem that the project would address in a tangible way. The users and their specific needs need to be identified. Users must see what is in it for them and understand what the proposed system will allow them to do that they cannot do now.

We recommend developing a plan for focus group activities to elicit information and identification of specific needs. Do not underestimate the importance of doing this correctly. The requirements of, and use for, the technological solutions need to be better articulated. The EAC suggests describing the existing physical problems and the approaches to their solution. Describe how a sequence of model runs and analyses might comprise an end-to-end solution, identify critical points where improvement is needed, and then focus the CI solutions developed on these real problems. Specifically, where the team lists existing systems there is a need to articulate why and to what extent the existing approaches fall short. For example, Kepler is a model interoperability system that has been developed at NSF-supported supercomputer centers, so its use would be seen to leverage NSF investment. However, there are also commercial packages (e.g. Matlab) that have similar capabilities (at least superficially). Thus, there is a need to articulate what is new and innovative about any solution considered by the project team that is different from one of these existing systems.

> **Response:** The case study development workshops described above are expected to provide needed science needs and requirements around which the model framework work may be performed. They will define specific science workflows that would benefit from an integrated system, and yield a concrete set of initial capabilities that are currently not well met by existing systems while also bounding the problem to be addressed in the development of the model interoperability components of the tri-state system.

There was also concern that model linkages proposed are over-ambitious. There is a need to pick small, doable examples that demonstrate capabilities to solve a meaningful problem quickly.

Many of the problems (model coupling, interoperability, data models) being addressed are also the subject of broad research and development around the world. The EPSCoR team needs to ensure that they properly understand, are linked with, and can leverage other efforts so that their contributions are state-of-the-art, effective, and cost-efficient. In other words, do not limit interactions to the three states or embark on re-inventing a wheel.

> **Response:** In addition to the above outlined connections with existing model interoperability activities (see page 2), the team will identify key projects and activities where additional linkages may be created. One significant step towards this goal is the participation of several team members in the IEMSS conference in July 2010, with one of the team members (Dan Ames) serving as a co-organizer for a workshop focused on open source software models in the development of integrated modeling systems. This workshop will bring together the developers of over 20 integrated environmental modeling systems with a goal of providing a broad overview of the development models used and capabilities developed.

Utilize CI to Integrate Research with Education: Findings and Recommendations

The project team has undertaken four major activities under the auspices of their cyberlearning component, namely: 1) support CI training in computation; 2) develop and disseminate educational materials for middle schools and high schools; 3) develop and support extracurricular CI activities; and 4) develop and deliver industry CI days. We address these in two categories: those to support graduate students, post-docs, faculty, and industry members (1 & 4 above) and those to support middle and high school learning (2 & 3 above).

With regard to support for graduate students, post-docs, faculty, and industry members, there is a wealth of workshops, short courses, etc. that are mentioned in the proposal. The EAC also noted the potential usefulness of http://shodor.org/cserd and the OU Supercomputing Center for Education & Research (http://www.oscer.ou.edu/education.php) to the project.

> **Response:** Thank you for the pointers to specific materials and we will explore these. We are specifically talking to Shodor about a NSF CI-TEAM proposal.

In terms of the activities to support middle and high school learning, there were several questions raised with respect to these materials:

1) What pedagogical model is being employed in the development of these materials?

   Response: The pedagogical model will change depending on the state's needs (UNLV versus ID-MOSS) and the audience/data used. For example, the models we are using include: inquiry-based models (5-e) and lecture-based (structural) models.

2) Regarding activity 1, the group mentioned the National Center for Learning and Teaching Nanoscale Science and Engineering framework (Shin et al. 2008). This has relevance in that they are dealing with the topic of nanotechnology, and one of the principal issues here is scale, namely size. In the proposed project, the topic is climate change and some of the challenging pedagogical concepts are physical scales (size, e.g. size and concentration of oxygen in water/air, etc.) and temporal scales (climate processes occurring over different time scales and research has shown that people have difficulty understanding time scales). Related questions are how to ensure that the data vis-à-vis time scales will be appropriately decided upon for the grade levels and how the driving questions/pedagogical tasks will engage students in deep learning regarding issues of time scale

   Response: One of the 'big ideas' of climate change that we are exploring for curriculum development is the notion of temporal scale. Size scale exists but time scale is harder to communicate and is needed in outreach. How do you accurately represent the scale of time? Visualization should be able to help with this. Consider: "Weather is not climate" – taking students through today's weather and change scales (minutes, months, years). A year scale is too short for climate change so we need to make the jump to a much larger scale, for example, consider:

   Lunch (period), Grade 10 (year long), Parent's Life, Lifetime of our country, How long ago were Dinosaurs on Earth?, etc.

3) What are the driving questions that will be used to guide students in their inquiry? Driving questions must have scientific worth, high interest to the grade level in question, etc.

   Response: We have ideas of some of the 'big questions' that we will use to guide students' inquiry (e.g. timescale); however, we are still identifying these 'big questions'. We met with a consortium of scientists on April 7, 2010 (Tri-State Meeting, Incline Village) to engage in this conversation and the process is still on going.

4) What are the bases for the formative data to be collected? For example, how will the group package the data so that they are usable for middle and high school students? And how will the dependent measures address efficacy of the materials for middle and high school students?

   Response: We have the users of the materials integrated into the design and development process to ensure usability of the materials. For example: MOSS, GUTS, SCC, and middle and high school teachers are involved. We are developing an efficacy checklist to be completed by the teachers involved in the process during the formative evaluation stage (before it is delivered).

5) Similarly, how will the issue of the regional climate system model's (atmosphere, hydro, land surface) deficiencies and problems be taught to the students using the yet-to-be-built Regional Climate Model (RCM) or its predictions? It's crucial that the students understand numerical modeling's limitations and that output offered today as the foundation of their study may be undermined tomorrow by realizations of: misunderstandings of physical/chemical processes and linkages; errors in forcing datasets and parameterizations; and bugs in software and coding. These are particular vulnerabilities in numerical prediction. The new RCM will be an incomplete approximation of the earth system and its diverse facets/forcings (atmospheric, oceanographic, hydrologic, chemical, biologic, cryospheric, anthropologic, solar, etc.), and this needs candid explanation to the students. There is a danger of misapplication or misinterpretation of this limited tool by young students who are not yet scientifically astute (middle schools and high school students, and undergrads as well).

   Response: The computer modeling and simulations developed through SCC and GUTS provide a foundation to address questions about limitations and applications of models. In the process of experimenting with existing models and building their own models and simulations, students necessarily confront issues about assumptions, simplifications, and the resulting limitations of models to correctly represent and predict reality. This forms the basis for understanding limitations in models--such as those that will be in the RCM. The materials developed through this project will tackle these issues directly.

6) How will these materials be used in classroom implementations versus more informal settings, such as in extracurricular activities (e.g. after-school programs)?

   Response: We are considering: SCC is both in and out of the formal classroom; GUTS is after school; NV and NM are working with teachers in the 'traditional' classroom; and MOSS is outdoor classroom. Thus, the materials will be used in all instances depending on the state program. The source materials will be similar but used in multiple ways as they are conveyed in different environments.

7) Will other dependent measures also be used in addition to content measures, will process skills be measured (NSES inquiry skills), and will attitudinal measures be used (they may yield pre-post gains for students)?

   Response: No, we are not focusing on implementation and associated dependent measures. Our focus is development and delivery based on the scale of funding and what we proposed in the proposal.

8) Will you be developing your own portal or making use of one already in existence, e.g. WISE/TELS?

   Response: We will make use of WISE/TELs as well as local portals (idahostem.org; climatechange.education.unlv.edu; cs.nmt.edu/~epscor; tri-state EPSCoR websites). We will ensure cross linkages between our portals in order to maximize dissemination.

9) Will your decisions about topic areas to be addressed be driven by scientists' data sets or will you, for continuity across the curriculum, develop units for climate that are not necessarily being addressed by the larger project?

   Response: Please refer to our response to Question 3 – we believe it is imperative to be driven by the 'big ideas'.

Prepared and Reviewed by the External Advisory Committee Members listed below:

Janice D. Gobert, Ph.D.

Alan R. Hevner, Ph.D.

Henry Neeman, Ph.D.

Cherri M. Pancake, Ph.D.

Jordan G. Powers, Ph.D.

Robert D. Sherwood, Ph.D.

David G. Tarboton, Sc.D.