Final Report

2nd US-India Network Enabled Research Collaboration Workshop

22nd-23rd March 2012

Westin Arlington Gateway Hotel, Washington DC

Supported By:

Contents

Final Report:

2nd US-India Network Enabled Research Collaboration Workshop

held in Washington, DC on 22nd & 23rd March 2012

James G. Williams - Indiana University
George M McLaughlin – Indiana University

Motivation for the Workshop Series

The motivation for these workshops resulted from the coming together of a number of factors including:

- the desire by the US and India to have a stronger collaborative science engagement;
- the deployment of significant high bandwidth connectivity between India and the US;
- the implementation of the National Knowledge Network in India; and
- existing research programs and potential new ones involving US and Indian researchers that are able to exploit the new connectivity to disruptively enhance these collaborations.

However, the fact that the opportunities for enhanced network-enabled collaboration are now in place does not mean that these will be exploited automatically. Researchers need to be made aware of the possibilities and what they need to do to take advantage of them. Similarly, networking professionals need to understand what the real needs of the research communities are to implement the optimal engineering, protocols and performance attributes in a way that ensures an easy-to-use experience for the researchers.

Identifying champions and using exemplars of network-enabled collaborations that are already happening is a powerful motivator. Presenting and demonstrating what has been done, and planning for what could be done, with the new infrastructure, helps to catalyze uptake. Bringing these elements together in thoughtfully designed workshops was seen as an effective approach.

Acknowledgements

Indiana University gratefully acknowledges the support of the US National Science Foundation via award OISE 0960487, and Indiana University and ERNET India gratefully acknowledge the support of the Indo-US Science and Technology Forum via award 110-2911-WS.

The 1st Workshop

The first workshop was held in New Delhi in December 2010 and attracted 132 registrants from universities, national laboratories and government organizations. A full day of tutorials was held immediately before the workshop. There were 103 registrants from India and 25 from the US, and 4 others. Further details on the first workshop are available from the IU ScholarWORKS repository at:
https://scholarworks.iu.edu/dspace/handle/2022/14422
The final report of this workshop is available at:
http://internationalnetworking.iu.edu/sites/internationalnetworking.iu.edu/files/us-indiaworkshopreportfinal.pdf.

This first workshop provided the opportunity for the cyberinfrastructure service providers and the research users from a number of disciplines to come together to jointly develop a way forward that will lead to significantly enhanced opportunities for collaboration between India and the US.

The 2nd Workshop

This **2nd US-India Network-enabled Research Collaboration Workshop** built on the results and follow-up from the first workshop by identifying how problems and opportunities of network-enabled research collaboration presented in Delhi have been addressed; illustrating what new collaboration opportunities (and problems) have presented themselves and how they might be addressed; and posing the question - if network-enabled research collaboration dialogue has value, how do we proceed forward in a sustainable manner?

Participants included researchers from a number of disciplines; network professionals, government officials and representative from funding agencies from both the US and India, and a range of network-enabled collaborative research activities were presented and demonstrated.

This workshop was intentionally smaller and more focused with 17 presenters, panelist and moderators from the US, and 12 from India putting forward elements of their collaborations, illustrating what worked, identifying inhibitors and proposing ways forward. In addition to these 29, another 17 scientists, network engineers and representatives from NSF, the State Department and NITRD participated in a robust discussion with the presenters and panelists that resulted in identifying 4 main themes and findings (each with a number of sub categories) together with 10 potential action items aimed at addressing the findings.

To illustrate what has been achieved so far, this Interim report concludes with six specific exemplars of collaborations catalyzed and furthered by these workshops.

The presentations and description of the workshop is available at:

http://internationalnetworking.iu.edu/us-india-workshop-2

The final report will be available from the IU ScholarWORKS repository.

Themes and Findings

1. Network Infrastructure and advanced communications services

a. There is already good capacity between the US and India via the Pacific and Indian Oceans (2.5Gbps) and 1Gbps connections to selected NKN connected sites within India. But, planning for increased global capacity should begin immediately.
b. The exchange points in the US, Japan, Singapore and Mumbai are well set up to handle large data flows.
c. The 1Gbps connections within India may be limiting for some high throughput applications.
d. Deployment of a relevant network performance set of tools that allow ready identification of network/throughput failures, bottlenecks, etc should be implemented as a high priority.
e. US-India network-enabled collaboration would be facilitated by the deployment of an appropriate federated identity and security environment.
f. Rock-solid dependable networking for videoconferencing is a high priority for many.
g. In India, the adoption of a low cost audio-visual environment for teaching and collaboration, such as A-VIEW, developed by Amrita, has promise. However this needs to interwork with other AV systems that collaborators have invested in such as Cisco and Lifesize.

See action item #1

2. Network-enabled access to and location of global research facilities and datastores

a. The new network capacity allows access by Indian researchers to facilities that are enormously costly to build, operate and maintain (e.g. remote access to and control of the Synchrotron Beamlines at the Berkeley Laboratories by protein crystallographers in India
b. Conversely, siting certain facilities in India could result in massive improvements in certain areas of science (e.g. installing a Laser Interferometer Gravitational-Wave Observatory facility as part of the LIGO program).
c. During the Medical session, it was noted that there are advantages in data being collected, located and analyzed in India by both US and Indian investigators.
d. The Open Source Drug Discovery program was recognized as an excellent example of Indo-US collaboration facilitated globally by R&E networks and illustrated well India's leading role in crowdsourcing.

See action item #2

3. Adapting to the changing science architecture, explosion of availability of data, and making network-enabled collaboration intuitive to the user

a. Across all areas of science, the volume and potential value of data is increasing. The value will only be realized if the data can be stored, curated, accessed, mined, manipulated, visualized, archived and shared in ways that are easy to use and intuitive to the various disciplines to which that data is relevant.
b. Data needs to be moved between instruments, data stores, analysis systems, visualization facilities and users in ways that are as unrestricted as practicable. The ESNet presentation provided example of the methodology used to gain an understanding of the science requirements that the networks must support.
c. Cloud computing and cloud services are becoming commonplace. Benchmarking of cloud infrastructure is of increasing relevance, as are the related networking implications.
d. High capacity networks and advanced communications systems are key to moving these massive volumes of data, but the complexities of the networking environment and protocols should be as transparent as possible to the users. (you don't need to be a vehicle mechanic to drive a car)
e. Scientists need to be made aware of the opportunities and (any existing) limitations of the networking environment, and should raise any networking-related inhibitors.
f. Networking engineers need to gain a sufficient understanding of the science requirement to develop the optimum networking environment.
g. Training on the use of infrastructure for e-science and student exchange programs seen as positive, particularly where different cultures are involved. PRAGMA was introduces as an example of a framework for persistent collaborations that rely on both physical and human networks.
h. International collaborations take time to mature, face-to-face meetings essential in addition to virtual collaboration

See action item #3

4. Awareness-raising of the potential for Indo-US Network-enabled Collaboration

More exemplars of the kind presented during the workshop will help promote the opportunities of network-enabled collaboration.

See action item #4

Potential Immediate Actions

1. Network Infrastructure and advanced communications services

a. Evaluate and identify capacity bottlenecks that are inhibiting or limiting high-end research programs at the India side and make recommendations accordingly.
(action: NKN and ERNET)

b. Evaluate and identify any current or anticipated capacity bottlenecks on the international circuits between India and the US and determine how these may be addressed.
(action: NKN, ERNET, TransPac, Internet2, GlobalNOC)

c. Develop and schedule a number of in-depth technical meetings to consider harmonized implementation of:

- ***i.*** Network tools (performance measurement, dynamic circuits);
- ***ii.*** Security;
- ***iii.*** Network Research (OpenFlow, SDN networking);
- ***iv.*** Above the net services (federate identity, collaboration environments).

(action: Internet2, NKN, ERNET, ESNET, GlobalNOC)

d. Evaluate the acceptability and practicality of a wide deployment of AVIEW on NKN/ERNET as a low-cost audiovisual collaboration and learning environment capable of interworking with other brands of AV equipment that collaborators have a heavy investment in.
(action: NKN and ERNET)

2. Network-enabled access to and siting of global research facilities and datastores

a) Bring to the attention of the NSF, IUSSTF and other relevant organizations, the advantages of locating nodes of particular science facilities (eg LIGO) or data repositories (Geo, Medical) in India.
(action: Williams, US PI)

b) Within India, raise the awareness across relevant science disciplines of the potential for network-enabled remote access to and control of specialized high-end science facilities located in the US that may significantly enhance US-India collaborations
(action: Mohan Ram, Indian PI)

3. The changing science architecture, explosion of availability of science data and making network-enabled collaboration intuitive to the user

a) Develop proposals, taking into account the ESNet methodology, for determining the high-end science requirements that the networks need to support.
(action: Williams, Mohan Ram and ESNet)

b) Propose to the NSF, IUSSTF and other agencies the formation of a small number of disciplined-based working groups where the disciplines are, or will be, reliant on high capacity networks to gain disruptive enhancement to the ways their US-India collaborations evolve (and that are not already well served). These working groups will provide both focus and exemplars for further exploitation.
(action: Williams and Mohan Ram)

c) Bring to the attention of the science agencies, the effectiveness of the PRAGMA initiative and the approaches that might be taken to enhance Indo-US network-enabled research collaboration and increase Indian involvement in regional research and education activities.
(action: Williams and Mohan Ram)

4. Awareness-raising of the potential for Indo-US Network-enabled Collaboration

Publicize the exemplars, reports and actions from both the first and second workshops and identify other areas that may be ripe for exploitation.
(action: US and Indian PIs and workshop conveners)

Specific activities and collaborations catalyzed and furthered by the workshops

1. Indo-US rapid detection and analysis of astronomical transient events

The *Inter-University Centre for Astronomy and Astrophysics (IUCAA)*, Pune, India and the Caltech Astronomy Department are working together on developing tools and technology for detection of astronomical transients i.e. objects that vary in brightness over short time periods (eg supernova, new asteroid, binary flares).

Transient astronomy requires fast data processing as some of the events evolve within a few seconds and immediate follow-up observations are required to understand the nature of those events.

Since observations are carried out at different locations depending upon the available facilities in the observatories and the geographical location of the observatory, fast communication and data transfer for processing becomes the major bottleneck affecting the effectiveness of the entire process.

It is critical that analysis of these objects happens close to real-time as follow-up observations based on results from the analysis are requires. Some of the methods are being developed by in India and some follow-up facilities are also in India (the origin of the transient detection currently happens in the US). In the near future the number of transients will increase necessitating larger bandwidth and robust networks.

These examples take advantage of India's geographical position as well as the expertise and infrastructure. These will drive use of the network for "near real time" response to these global collaborations.

The new high capacity research networking now available between these two institutions opens up new opportunities for advancing collaboration leading to a better understanding of transient event.

At the time of the first workshop, some of the surveys have already started delivering data, and the institutions were ready to make use of the new infrastructure. At the APAN Delhi meeting held in August of 2011 an impressive demonstration of the high capacity data transfers (600Mbps, end-to-end, limited only by the available bandwidth) was made during the plenary session at the APAN Delhi meeting, with Indian Government Ministers Shri Sachin Pilot and Shri Milind Deora on the podium to witness the demonstration.

Further development of the collaboration was presented and demonstrated during the 2nd workshop and this time involving the LIGO activity. Further enhancement of the collaborative activists is underway with tentative plans to make further demonstrations and presentations alongside the Joint APAN/Internet2 Meeting in Hawaii in January 2013.

2. Bioinformatics Analysis of Cyanobacteria for Carbon Sequestering and Biofuel Production

Following the first workshop, a proposal from IUPUI (Matthew Palakal), involving the establishment of a pilot collaborative research program on carbon sequestration with the Indian Institute of Technology (IIT), Bombay (Pramod Wangikar) was successful. The successful collaborative proposal was a direct result of the first workshop.

The study involved two specific tasks: Prof. Wangikar, an expert in cyanobacteria, provided operon data from his laboratory and Prof. Palakal's (PI) lab conducted bioinformatics analysis and genome-wide studies for species selection. Robust high performing network connectivity between the two sites was critical elements of the data and analysis sharing.

This project has two long-term objectives: (i) fundamental understanding of the regulatory mechanisms of cyanobacteria—so that these mechanisms can be reengineered to produce biofuel directly rather than as a byproduct of the biomass and, more broadly, (ii) the ability to reengineer the bacteria for production of useful products. As part of meeting the above stated objectives, this work looks at the development of a suite of algorithms and software tools customized for the large-scale bioinformatics analysis of the cyanobacteria genome.

Cyanobacteria can be developed as an excellent microbial cell factory that can harvest solar energy and convert atmospheric CO_2 to useful products such as biofuels. If the outcome and the long-term objectives of this work prove to be viable, they will have enormous, global-level impact on the production of clean energy, especially in burning clean coal. The project will also have a broad impact on education and training: this project will educate young professionals—professionals who need to understand how to use cutting-edge technology to produce clean energy

This example focuses on an area of energy, important to both countries, and drawing on expertise from each.

3. Remote Access to and Control of Berkley Synchrotron Beamlines by Homi Bhabha National Institute (HBNI), Anushaktinagar, Mumbai

Third generation synchrotrons are mega-facilities that drive expansions of research-frontiers in a variety of science, arts and engineering disciplines. However, they cost large amounts of money to build and maintain. They also make use of very sophisticated technologies, and therefore only few are available in the world. The ability to remotely access and control the Berkeley facility from India is a great boost to Indian crystallographers and for US-India collaboration.

In the discipline of biology, crystallographers are a very active community that use hard x-rays generated at these synchrotrons. Two unique properties of synchrotrons are particularly significant

from the point of view of crystallographers: 1) the high-brilliance of the almost continuous energy x-ray beam produced, and 2) the time structure of the x-ray beam.

While the former feature allows usage of very small single crystals and experimental 'phase' determination, the latter feature enables structural mapping of biochemical reactions. The accuracy of structures determined using synchrotron data is normally much higher, and approaches the accuracy needed for structure-based drug design. These benefits of synchrotrons to protein crystallographers are now within reach of scientists around the world, thanks to completely automated protein crystallography beam lines, and to development of reliable networks for very fast transfer of commands and data.

The National Knowledge Network in India is linked to international networks, and through this network HBNI has been able to operate remotely various international synchrotron beam lines:

A number of high-resolution data sets have been collected on drug-resistant HIV-1 protease mutant/drug complexes. These structures have been refined to low crystallographic R-factors comparable to those when the diffraction data were collected by onsite operation of the beamline. The molecular models derived are also stereochemically very accurate. There are several protein crystallography beamlines on synchrotrons available in the USA. Providing remote access to these beamlines to protein crystallographers from India would be a good example of net-enabled collaboration in scientific research. Using the remote data collection facility setup at Homi Bhabha National Institute, Anushaktinagar, Mumbai, feasibility of remote operation of the protein crystallography beamline 5.0.2 on ALS, has been established. Actual data collection will be carried out whenever remote access is formalised.

4. NKN Global Classroom Collaboration

The Global Classroom collaboration between Amrita University and SUNY, Buffalo seeks to Architect, Design, Develop and deploy a Cross-functional, Multi-Media, Multi-Channel, High Definition, Interactive, Immersive, Adaptive Classroom experience – a University without walls and borders, that effectively uses the high bandwidth, low latency worldwide network provided by the NKN to enable students attend classes of higher education at UB. Most remote classroom solutions today do not provide a really immersive environment to captivate and motivate the students. The students tend to get tired after short sessions. By providing a really interactive and immersive experience, the collaborators hope to open up a totally new paradigm in remote classrooms and discussion rooms. Initially, the project will develop a system to connect State University of New York at Buffalo [SUNY, UB], through the NKN. This project has the potential to bring much sought after higher education degrees from SUNY, UB to students and working professionals without needing to displace them from their place of study or work. In the future this will even allow remote students and researchers to use labs and facilities at SUNY, UB.

At a technical level, using sophisticated algorithms for Object identification and extraction, including Edge Detection, Face Detection, Gesture analysis, Scene change Detection, Motion Tracking all are done in real time. Multi-channel, Multi-perspective video and audio streams are semantically composited in real-time to create a seamless neighbourliness between the remote teacher and students, effectively creating a media and knowledge environment so that remote students have the same classroom experience as the local students.

Another technical objective is to recreate the Knowledge Environment from the instructor's site to students at remote locations in an immersive manner, i.e. the ability to transport knowledge and information sources as they are interactively and collaboratively navigated, E.g. if the lecturer is navigating media to explain a concept, either through the web or through an interactive book or through various audible means, this interactivity is tracked and presented instantaneously, thereby giving a sense of teamwork between the local and remote sites.

5. Cheminformatics and Open Source Drug Discovery

Indiana University (IU) has an internationally renowned and unique research and education program in cheminformatics, the use of advanced informatics techniques for chemistry, biology and drug discovery.

Open Source Drug Discovery (OSDD) is a CSIR Team India Consortium with Global Partnership with a vision to provide affordable healthcare to the developing world through discovery of novel therapies for neglected tropical diseases like Malaria, Tuberculosis, and Leshmaniasis.

IU and OSDD are engaged in a multi-year educational and research collaboration aimed at:

(i) applying the latest research in cheminformatics techniques to the search for treatments for neglected diseases, and
(ii) bilaterally enhancing cheminformatics learning through engaging of Indian students with research and teaching programs at IU and the development of shared learning resources.

The robust collaboration between the Indian and US collaborators is supplemented by extensive community collaborative science and India's leadership in crowdsourcing. Raising awareness and training in worksflows, collaborative tools and the use of cyberinfrastructure. Public private partnerships as open collaborative endeavours are likely to be effective going forward. OSDD now has more than 5500 members from over 130 countries.

6. Big Data and Cloud Benchmarking

During the second workshop, there was a discussion and interest in holding a cloud benchmarking related workshop in India together with a day of tutorials. Following a subsequent application by Dr

Chaitan Baru, the NSF agreed to support a Big Data Benchmarking Workshop in Pune, India in December 2012 colocated with the IEEE High Performance Computing Conference www.hipc.org.

Information about the first big data benchmarking workshop, held on May 8-9 in San Jose, California is available at, http://clds.sdsc.edu/wbdb2012. Information about the India workshop will be linked from that site as well.

Participants in the second workshop

Presenters, panellists and moderators - US (17 Total, 1 remote)

Peter Arzberger	University California, San Diego	
Chaitan Baru	SDSC, UC San Diego	
Michael Cheetham	IUSSTF	
Eli Dart	ESNet	(remote)
Machi Dilworth	National Science Foundation	
Bharat Jayaraman	University at Buffalo, SUNY	
George Komatsoulis	National Cancer Institute, NIH	
Radha Nand Kumar	NCSA, University of Illinois	
David Lambert	Internet2	
George McLaughlin	Indiana University	
Ashish Mahabal	California Institute of Technology	
Suresh Marru	Indiana University	
Abhik Seal	Indiana University	
Anil Srivastava	Open Health Systems Laboratory	
James Williams	Indiana University	
Roy Williams	CalTech	
Steven Wolff	Internet2	

Presenters, panellists and moderators - India (12 total, 5 remote)

Arun Agarwal	University of Hyderabad	
Anshu Bhardwaj	Open Source Drug Discovery	
Kamal Bijlani	Amrita University	
Prof Samir K Brahmachari	Secretary, Dept of Scientific and Industrial Research	(remote)
Dr P.S.Dhekne	Raja Ramanna Fellow at BARC	(remote)
Dr Debaprya Dutta	Counselor (S&T), Embassy of India	
Madhusoodan Hosur	Indian National Science Academy	(remote)
Rajendra Joshi	C-DAC	(remote)
Ajit Kembhavi	IUCAA	(remote)
Mohan Ram Natarajan	ERNET	
Dipak Singh	ERNET	
Arun K Singh	Deputy Chief of Mission Embassy of India, USA	

Other participants (17)

Heather Boyles	Indiana University
Alice Jackson	Indiana University
Therese Miller	Indiana University
Bill Chang	National Science Foundation
Marjorie Lueck	National Science Foundation
Dane Skow	National Science Foundation
Kevin Thompson	National Science Foundation
Eric Bone	US Department of State
Molly Teas	US Department of State
Grant Miller	NCO, NITRD
Ann Doyle	Internet2
Edward Moynihan	Internet2
Michael Sullivan	Internet2
Harvey Newman	CalTech
Michael Foley	World Bank
Malathi Veeraraghavan	University of Virginia
Paul Love	Private Consultant

2nd US India Workshop Program Schedule

This Schedule available with links to speaker biographies and presentations is available here:
http://internationalnetworks.iu.edu/us-india-workshop-2/agenda

Thursday 22nd March — **Workshop - day 1**

7.30 - 9.00 ***Breakfast***

9.00 - 10.30 ***Welcome***

US Welcome – ***James Williams***, *Indiana University and PI TransPac3 and ACE programs* – 5 mins
India Welcome – ***N. Mohan Ram***, *Director General, ERNET* – 5 mins
Address by ***Dr. Machi F. Dilworth***, *Director, Office of International Science and Engineering, NSF* - 10mins
Address by ***Ambassador Arun K. Singh***, *Deputy Chief of Mission, Embassy of India* - 10mins

Introduction to the Workshop - background, objectives and what we want to achieve from the workshop *James Williams* – 10 mins

Opening Plenary / demonstration – 40 mins

Network-enabled access to globally distributed data repositories
Moderated by ***George McLaughlin***

Astronomy with Cutting-Edge ICT: Making sense of Transients using Geographically dispersed resources
Prof Ajit Kembhavi, *Director IUCAA, Pune, India (remote)* **(Presentation)** ***and*** ***Dr Ashish Mahabal***, *Astronomy Department, Caltech,* **(Presentation),** ***Dr. Roy Williams***, *LIGO, Caltech* **(Presentation)**

Abstract: Massive amounts of data are coming online from new generation sky surveys. Combined with the Geographically diverse archives and processing pipelines that need to run in real-time, these programs are a prime example of needing high bandwidth across continents. The presentation and demonstration will emphasize the critical need for high capacity networks and advanced communication services to exploit US-India collaboration in Astroinformatics and Astrophysics in connection with the current and near-future programs.

10.30 - 11.00 **Morning Tea**

11.00 - 12.30 **Research Collaboration**

*Moderated by **N. Mohan Ram***

How the US and India Network Organizations support Research Collaboration

This session will provide updates since the last workshop on the network infrastructure that is available to researchers today or will be will be available to researchers in the relatively near term (1-2 years) and describe what the organizations that manage the networks do to facilitiate research collaboration.

Current and evolving R&E network infrastructure and research support structures in India

India's National Knowledge Network (NKN) - ***Dr. P.S. Dhekne**, Raja Ramanna Fellow at BARC (remote)* **(Presentation)** – 10 mins

Abstract: The National Knowledge Network (NKN) is a state-of-the-art multi-gigabit pan-India network for providing a unified high-speed network backbone for all knowledge related institutions in the country. The purpose of such a knowledge network goes to the very core of the country's quest for building quality institutions with requisite research facilities and creating a pool of highly trained professionals. The NKN presentation would cover the key highlights of the project, Management overview, Technical overview, NKN connectivity status and key services offered.

*ERNET - **Mr. Dipak Singh**, Director of Network Operations, ERNET **(Presentation)** – 10 mins*

Current and evolving R&E network infrastructure and research support structures in the US and between US and India

TransPAC3 - ***James Williams**, Principal Investigator TP3 project - Indiana University* **(Presentation)** – 10 mins

Internet2 – ***H. David Lambert*** *– CEO Internet2* **(Presentation)** – 10 mins

TEIN3's role in facilitating US-India Collaboration - ***David West*** *– DANTE* **(Presentation)** – 10 mins (remote)

Abstract: The talk will provide a brief update of the TEIN project focusing on the connectivity it provides, project plans, and its current and potential future support for global science applications.

The Energy & Sciences Network (ESNet) support for Network-enabled Research Collaboration- *Eli Dart* (Presentation) - 15 mins

Infrastructure Q/A session – 25 mins

12.45 - 14.00 **Lunch**

14.00 - 14.45 **Network-enabled Access to Distant and High Cost Instruments** - 45 mins
*Moderated by **George McLaughlin***

Remote Access to and Control of Berkley Synchrotron Beamlines by Homi Bhabha National Institute (HBNI), Anushaktinagar, Mumbai
Dr. M V Hosur, *HBNI* **(Presentation)**

Abstract: Third generation synchrotrons are mega-facilities that drive expansions of research-frontiers in a variety of science, arts and engineering disciplines. However, they cost large amounts of money to build and maintain. They also make use of very sophisticated technologies, and therefore only few are available in the world. In the discipline of biology, crystallographers are a very active community to use hard x-rays generated at these synchrotrons. Two unique properties of synchrotrons are particularly significant from the point of view of crystallographers: 1) the high-brilliance of the almost continuous energy x-ray beam produced, and 2) the time structure of the x-ray beam. While the former feature allows usage of very small single crystals and experimental 'phase' determination, the latter feature enables structural mapping of biochemical reactions. The accuracy of structures determined using synchrotron data is normally much higher, and approaches the accuracy needed for structure-based drug design. These benefits of synchrotrons to protein crystallographers are now within reach of scientists around the world, thanks to completely automated protein crystallography beam lines, and to development of reliable networks for very fast transfer of commands and data. The National Knowledge Network in India is linked to international networks, and through this network we have been able to operate remotely following beam lines: 1) BM30 (or FIP) on ESRF in Grenoble, France, 2) BM14 on ESRF in Grenoble, France. A number of high-resolution data sets (1.6 – 2.0 Å) have been collected on drug-resistant HIV-1 protease mutant/drug complexes. These structures have been refined to low crystallographic R-factors comparable to those when the diffraction data were collected by onsite operation of the beamline. The molecular models derived are also stereochemically very accurate. There are several protein crystallography beamlines on

synchrotrons(APS, ALS, NSLS etc.) available in the USA. Providing remote access to these beamlines to protein crystallographers from India would be a good example of net-enabled collaboration in scientific research. Using the remote data collection facility setup at Homi Bhabha National Institute, Anushaktinagar, Mumbai, feasibility of remote operation of the protein crystallography beamline 5.0.2 on ALS, has been established. Actual data collection will be carried out whenever remote access is formalised.

14.45 - 15.30 **Network-enabling of Global Classrooms** – 45 mins
*Moderated by **George McLaughlin***

NKN - A Gateway to a Global Classroom
A collaboration between Amrita University and the University at Buffalo (State University of New York)
Prof Kamal Bijlani *Head of E-Learning Research Lab at Amrita University* **(Presentation)**
Prof Bharat Jayaraman, *Department of Computer Science and Engineering, University at Buffalo, SUNY* **(Presentation)**

15.30 - 16.00 **Afternoon tea**

16.00 - 16.45 **Opportunities for new collaboration -** 45 mins
*Moderated by **James Williams***

A Framework for Persistent Collaborations: PRAGMA Overview, Future, Lessons Learned, and Opportunities for US India Collaborations
Peter Arzberger*, Director of the National Biomedical Computation Resources* **(Presentation)**

Abstract: PRAGMA, a 30 institution, international, grass-roots organization, explores and evaluates practical approaches to how cyberinfrastructure software can be used to enable and enhance scientific collaboration among both small and medium sized groups. Scientific "expeditions" are used to define which software components, available from the PRAGMA partnership and elsewhere, need development and experimental evaluation prior to deployment on larger production infrastructures. Regular face-to-face meetings enables the group as a whole to support new science areas; gain insight to cyber developments in a very timely manner; support and sustain experimental testbeds across multiple administrative domains; create training, education, and network building activities; and provide the persistent interactions that engender the trust needed as the foundation of international scientific and infrastructure development collaborations.

In this presentation we will introduce PRAGMA as an example of a framework for persistent collaborations that rely on both physical and human networks. We will discuss future directions, lessons learned about collaborations, and present opportunities for collaboration

between US and India researchers.

Role of HPC in cyberinfrastructure and some experiences in US-India Collaborations.
***Radha Nandkumar**, Emeritus Director of NCSA's International and Campus Relations*
(Presentation)

Abstract: This presentation will describe opportunities for institutional and individual collaborations in defining the leading edge in high-end computing, information technologies, and cyberinfrastructure. The talk will highlight the role of high end computing in enabling breakthrough science and engineering in general as well as some of the challenges associated with large-scale simulations. An outline of several significant education and outreach activities as well as collaborative international projects will be provided. The presentation will conclude with mention of the impact of these initiatives on society at large.

Fostering Indo-US computational science collaborations
***Suresh Marru**, Indiana University, Program Manager, XSEDE Science Gateways*
(Presentation)

Abstract: A major hindrance in academic collaborations is intellectual property sharing. This talk will dwell upon open source development across multiple collaborating institutions paying attention to Science Gateways, which provide Web-based environments for scientists and students to perform computational experiments online via Web interfaces using Web services and computational workflows. We believe there are important steps that should be taken to go beyond basic open source to address requirements for building open software communities. In addition to licensing and support tools, open communities must have open processes for making design decisions, accepting code contributions, adding new project members, reporting and resolving problems, and making well-packaged and properly licensed software releases. The Apache Software Foundation provides the infrastructure and mentoring experience to help open source communities address these project governance issues. Additionally, Apache has interesting requirements (such as developer diversity) that are designed to emphasize the neutrality of the code base (encouraging competitors to have a safe place to cooperate), help sustain their projects through leadership turnover and funding cycles. I would like to discuss how forums like Apache can help US and Indian counter parts can share code and collaborate without worrying about IP and cross-country funding issues.

Afternoon Q/A - 15 mins

17.00 - 17.15 **Sum-up from first day – *George McLaughlin**, TransPac3*

Friday
23rd March

Workshop - day 2

7.30 - 9.00 ***Breakfast***

9.00 - 10.30 Welcome to day 2 and review of day 1 – ***James Williams*** – 10 minutes

Network-enabling of Medical Research and Drug Discovery Collaboration
Moderated by ***Mohan Ram***

Cheminformatics and Open Source Drug Discovery: a case study in academic collaboration between the U.S. and India – 40 mins
Abhik Seal ***presenting for*** ***Dr. David Wild***, *Indiana University*
Anshu Bhardwaj ***presenting for*** ***Dr. U.C.A. Jaleel***, *OSDD Malabar Christian College*
(Presentation)

Abstract: Indiana University (IU) has an internationally renowned and unique research and education program in cheminformatics, the use of advanced informatics techniques for chemistry, biology and drug discovery. Open Source Drug Discovery (OSDD) is a CSIR Team India Consortium with Global Partnership with a vision to provide affordable healthcare to the developing world through discovery of novel therapies for neglected tropical diseases like Malaria, Tuberculosis, and Leshmaniasis. IU and OSDD are engaged in a multi-year educational and research collaboration aimed at (i) applying the lastest research in cheminformatics techniques to the search for treatments for neglected diseases and (ii) bilaterally enhancing cheminformatics learning through engaging of Indian students with research and teaching programs at IU and the development of shared learning resources. In this talk we will describe the project and highlight the many challenges that have arisen, including infrastructure harmonization, funding, dealing with different time zones, and the need for advanced distance collaboration technologies. We will make recommendations as to steps that can be taken to facilitate stronger mutually beneficial collaboration between the countries.

Protein Structure Modelling on the IUCRG: A BRAF--caBIG® collaboration – 40 mins
Rajendra Joshi, *CDAC/BRAF (remote)* **(Presentation)**

Abstract: The last few decades have witnessed the evolution of biology from what used to be

a purely experimental field, to a high end computational domain, where unrelenting computational power is required to decipher pieces of data generated through high throughput techniques into blocks of information that will help to answer many mysteries of life. To be able to generate knowledge from the oceans of genomic data, enabling technologies like High Performance Computing, Grid Computing and Cloud Computing are the latest weapons in the hands of the modern biologist.

The importance of protein structures can be understood easily from the fact that the function of any protein is directly correlated to its structure. The three dimensional structure of a protein directs its function within a cellular environment. Any mutation in the protein sequence leads to changes in its structure which in turn may render the protein non-functional or even attribute some adverse functions leading to diseases like cancer. Over the decades cancer has become one of the most prevalent diseases with an estimate of reaching over 12 million deaths in 2030 according to World Health Organization. Proteins from almost 1% of the human genome have been identified to be involved in oncogenesis. In the absence of resolved structural data (RCSB database has 73974 resolved protein structures as opposed to 534695 sequence entries in UniProtKB) one has to resort to computational techniques to get the 3D structures of proteins in order to properly understand their functions.

The Bioinformatics Group at the Centre for Development of Advanced Computing (C-DAC) in collaboration with cancer Biomedical Informatics Grid (caBIG®) has developed a grid-enabled web-based automated pipeline for ab initio as well as homology based prediction of protein structures, with an emphasis on cancer related proteins. The pipeline has been deployed on the Bioinformatics Resources & Applications Facility (BRAF) hosted at C-DAC, Pune India. The upstream component of the pipeline retrieves a protein sequence (according to user input) from the gridPIR service of caBIG® that provides a data resource of high quality annotated information on all protein sequences supported by UniProtKB. The retrieved sequence in a FASTA format is then fed to the prediction pipeline. At its core the pipeline consists of two prediction engines, one ab initio based that uses the ROSETTA prediction algorithm and another homology modeling based that uses the MODPIPE program, for determining the 3D structures. The graphical user interface of the pipeline enables the user to choose various control parameters like which secondary structure prediction algorithms to use, number of iterations, number of output structures, uploading NMR constraint files, e-value etc. Once submitted, the jobs get distributed over multiple processors on the Biogene supercomputing system at BRAF, which significantly reduces the prediction time. The resultant output comes in the form of predicted structures in PDB format and parsed energy log files which can be downloaded by the user. All the file transfers are secured over the network by SFTP. JMol has been integrated within the pipeline to provide a visual inspection of the predicted models. Test cases have been run using the pipeline with a few cancer related proteins, downloaded

from The Cancer Genome Atlas (TCGA), where sequence data from various mutated proteins of affected patients are stored and made available in various data formats. Some of these results will be discussed during the presentation.

Indo-US Cooperation in biomedical informatics
George A. Komatsoulis, Ph.D., *interim Director, Center for Biomedical Informatics and Information Technology and CIO at the National Cancer Institute (NCI), NIH* **(Presentation)**

Abstract:

10.30 - 11.00 **Morning Tea**

11.00 - 12.00 **Evolving areas of Network-enabled Collaboration**
Moderated by ***Dipak Singh***

Geosciences, Environmental Networks & Cloud Services, and PRAGMA

A Knowledge R&D Networked Indo-US Collaboration: A case study in Earth Sciences – 20mins
Prof Arun Agarwal, *Dept of Computer & Information Sciences, University of Hyderabad*
(Presentation)

Abstract: Firstly we will cover the role of GEON/PRAGMA projects, initiated primarily by SDSC -UCSD through NSF, in developing CYBERINFRASTRUCTURE in a wide range of Earth Science disciplines in India since 2005. How this "IT head start" helped in the data fusion and visualization of a variety of earth science related data sets. Secondly, we will also highlight significant achievements in producing a new breed of hybrid students in terms of innovative man power development with cross fertilization of different science streams with IT. Thirdly, we provide a review of available data sets that are being generated in India by various organizations and their applications. In conclusion we will make reference of large data sets with a need to build Cloud Cyber-infrastructure a shift from Grid Middleware based Cyber-infrastructure for geosciences.

Big Data and Cloud Benchingmarking - 20 mins
Prof Chaitan Baru, *Director, Center for Large-scale Data Systems Research (CLDS) SDSC*
(Presentation)

Abstract: As science collaborations become data-centric—even moving in the direction of joint analysis of large datasets—there is increasing need for cyberinfrastructure to support data-intensive computing, and an opportunity for collaboration in the area of benchmarking "big data" applications at global-scale. Can we build environments that use distributed

computing and the cloud-based paradigm in which researchers in the US easily access and analyze scientific data from data archives in India, and vice versa? What type of system and network performance is required to sustain such applications? Is there an opportunity for Indo-US collaborations to study performance issues related to such data-intensive applications, and to develop related benchmarks?

In this session, we will discuss a new effort in developing "reference benchmarks" for big data, and "probe benchmarks" for data-intensive clouds, which we refer to as the Cloud Weather Service(TM). The goal of these efforts is to provide clear objective information on hardware, software, and system performance for data-intensive applications on dedicated clusters as well as in cloud-based environments. We will discuss how we might structure new Indo-US collaborations in this area.

Morning Session Q/A - 20 mins.

12.00 - 13.30 **Lunch**

13.30 - 14.45 **Sustainable US-India Network Enabled Research Collaborations - Where to from here?**
*Moderated by **James Williams***

A panel comprising representatives of the US and Indian governments and scientists, taking into account the contributions made during the workshop, will debate and deliberate on ways to significantly further enhance Indo-US network enabled collaboration. In doing so the panel and participants will try to identify key issues, challenges, obstacles, and opportunities needed for the development of action plans, and identify next steps and future deliverables. This session will be followed by a wide-ranging discussion among the participants which will help shape the final workshop recommendations.

14.45 - 14.55 Summing up of workshop - ***James Williams*** **&** ***N. Mohan Ram*** *DG, ERNET India*

14.55 - 15.00 **Closing Remarks - *James Williams***