Global Advanced Research Networks

2006 Annual Report (April)

Paul O'Nolan, IRRI

Audience: CGIAR ICTKM Community

2.0 Executive Summary

A year ago we expected advanced research network (ARN) connectivity between IRRI and CIP, the first two centers connected to such networks in April 2005. CIP was connected to Internet2 in September 2005 but it wasn't until April 2006 that IRRI and CIP finally had an Access Grid video conference that used ARN connectivity from end to end.

A year ago it was not clear that other centers would tap the project for funds and the possibility of needing to return funds was mooted if demand for advanced research network connectivity didn't pick up. CIAT was the only other likely candidate at the time. In the past year CIAT, then CIMMYT and ICARDA have applied for funds from the project and have established ,or are in the process of establishing, ARN connectivity. Within a short time it's expected that 5 CGIAR Centers will have and will be using such connectivity regularly.

The project has met its major objectives. However, the difficulty of establishing ARN connectivity in developing countries has affected the timetable, so that the project is late, and some funds remain unspent ($6,579).

A three month no-cost extension to the project is requested. A budgeted $2,500 remains to bring a speaker from Japan to an event at IRRI. In addition, permission to contribute remaining funds to the construction of a portable Access Grid node at IRRI is requested (estimated cost $13,000).

3.0 Progress Report

3.1 Financial Management and Reporting

The following disbursements have been made:

From project funds

1. $20,000	CIP for local loop connection; details below
2. $20,000	CIAT for CLARANET link
3. $20,000	CIMMYT for membership of academic research network
4. $20,000	ICARDA for infrastructure for GEANT2 network link
5. $15,000	IRRI for grid cluster for bionformatics
6. $ 5,000	Indirect costs
3. $15,000	Contribution to IRRI grid cluster
4. $ 2,250	ASTI for Access Grid node consulting and support
$105,021	Total

By IRRI

1.	$ 8,198	AG audio system
2.	$ 8,000	AG projection system
3.	$ 5,500	AG video system
4.	*$ 5,700*	*IT infrastructure*
5.	$ 1,000	IT infrastructure (in kind; existing)
6.	$ 7,536	Communications infrastructure
	$35,934	Total

3.2 Project Performance

3.2.1 Progress vis a viz milestones in work plan or indicators in M&E plan

3.2.2 Results

Construction, testing and operation of a cluster for Grid Computing

The project contributed $15,000 to the acquisition by IRRI a high performance computing cluster (HPC) used by IRRI's Biometrics and Bioinformatics unit (subsequently merged with its counterpart at CIMMYT to form the IRRI CIMMYT Crops Research Informatics Laboratory; CRIL).

Hardware configuration: 1 master node + 7 slave nodes in Umali data center, 1 mirror master node in Cantrell data center (to insure against research data loss; funded by the project)

Each node has dual 2.0 Ghz Opteron CPUs with 4 Gb of RAM. Each master node has 1Tb of RAID5 storage.

Applications of this technology:

- High throughput analysis of the rice genome for use in SNP[1] detection
- High throughput statistical analysis of large micro-array experiments
- Sequence analysis of candidate regions in rice for specific agronomic trait projects

Significance

1. The rice genome analysis is to prepare DNA sequence information for the Perlegen project[2] to identify "whole genome" genetic variation at the DNA base pair level ("single nucleotide polymorphisms" or SNP for short) in 20-

[1] Single nucleotide polymorphism
[2] http://www.irri.org/media/press/press.asp?id=113

25 representative Oryza sativa varieties, as an essential first step to surveying the genetic diversity of the entire genetic resources collection. Such a survey is necessary to undertake association genetics to correlate specific genotype (DNA sequences) with specific desired agronomic traits for poverty alleviation. Such large scale surveys were not possible before the sequencing of the rice genome. The computational preparation of a ~430 megabase genome is very computationally intensive, hence the use of the HPC.

2. The rice genome contains tens of thousands of predicted genes. Microarray experiments designed to detect gene activity at the subcellular level (in the RNA) each generate many thousands of data points that require heavy statistical and classification analysis to remove noise, detect significant signals and interpret the data meaningfully. Again, HPC usage turns such computations into hours or days, what might taken weeks beforehand, for such large data sets

The cluster has been used to do the “masking” (dealing with transposable element and other redundancies in the rice genome) that is, to identify unique regions in the rice genome for use in the experimental design by IRRI research partner Perlegen. Perlegen’s analysis will result in a SNP database across 21 rice genomes. These data will be used to design a set of tagged SNPs which uniquely identify haplotype regions (regions with very low genetic recombination rates). Everything within a haplotype co-segregates, accordingly these tagged SNPS can be used to follow candidate genes: if there’s a candidate gene within the block the tagged SNP can be used as a surrogate for the association with the desired trait. The presence of particular tags in a plant’s genome will be tested for using micro-array experiments. This will allow plant breeders to undertake whole genome scans and identify which candidate genes are associated with traits of importance. Having identified these they can work to introgress desired traits to existing breeding lines. IRRI currently expects outputs from Perlegen in May 2006. After this, IRRI’s use of the HPC will increase further, dramatically.

The equipment is operational and in active use helping IRRI do information intensive science that couldn’t be done without it.

In principle, IRRI, CIP, ICRISAT and ILRI are connected via Platform LSF Cluster Grid, and CIP has enabled LDAP authentication using Active Directory (to facilitate secure, wider CGIAR participation in the use of the machines). However, things have been fairly quiet on this front over the past year. As IRRI’s need for additional computational power increases after May 2006 efforts to tap resources on the CGIAR’s network will grow. At the moment the CGIAR is in the position of the IT being a bit ahead of demand but this is expected to change. In the long term outsourcing all computational processing over the Internet is likely to be an option, but this is not a viable proposition just yet except on a very large scale (e.g., Dreamworks outsourcing animation rendering to HP).

Access Grid node construction and commissioning

Access Grid

IRRI's connection to APAN was broken during mid 2005 as the national termination point moved to ASTI, our partner in the Access Grid project, and as PHNET, the former termination point, relocated. IRRI's 2Mb link to PHNET was left dangling. A new higher capacity link directly to ASTI had been planned and commissioned but it took one year to complete. The new link became operational in Sept. 2005, shortly after CIP established it's own ARN link. IRRI was without an ARN link for some months.

Once the new link was installed IRRI was able to connect to Australian unversities with it's completed Access Grid installation without any difficulty. We then began trying to connect with CIP. It took several months of trying before the centers were finally able to have a successful Access Grid video conference. A litany of technical problems needed to be resolved at CIP. The reasons for the delay included

- Problems with the national infrastructure in Peru
- Driver problems with the systems used by CIP
- CIP lacking any partner like ASTI with more experience of the technology
- CIP not having identical or even comparable equipment to IRRI (it used a much cheaper setup but this was costly in terms of time)
- Difficulty of coordinating testing with a 13 hour time zone difference (only a couple of tests a week at most were attempted; these always involved CIP staff staying late)

Success was finally achieved in April 2006. A note from the IRRI Bulletin of 10 April 2006 follows

IRRI and CIP establish first CGIAR Access Grid Link

On the 7th of April, last Friday, after a successful test the previous day, IRRI and CIP made a little CGIAR history by becoming the first two CGIAR Centers to have a video conference using the Access Grid. They were joined in the conference by the Advanced Science and Technology Institute (ASTI) at the U.P. Diliman, Quezon City, IRRI's partner in using Access Grid technology (ASTI and IRRI are the only Access Grid nodes in the country). The Access Grid is a technology for enabling multiple locations, even dozens, to particpate simultaneously in a video conference.

CIP obtained advanced research network connectivity last September when it established a link to the local Agrarian University, already connected to Internet2. IRRI and CIP staff have tried on and off since that date to establish a link but

there were invariably problems at CIP's end and it took months to iron out all the difficulties. Thanks to perserverence Rolando Navarro Jarra of CIP and Lino (Roy) Suarez of IT Services finally succeeded in establishing a working connection and the CGIAR entered a new era. ITS staff were joined by colleagues from IRRI's Crop Research Informatics Laboratory (CRIL). Graham McLaren and Thomas Metz, who have been keenly awaiting this moment showed up to have a look. In a note to IT colleagues in the CGIAR Paul O'Nolan wrote

"Next, since Rolando still didn't show any signs of wanting to go to bed (I guess we were all pretty excited) I grabbed Gene Hettel from publications upstairs and suggested he might want to see and perhaps record a little history. He brought IRRI's videoographer and cameraman Joe Ibabao. Poor Rolando was completely outnumbered and a bit disappointed that not one of us spoke Spanish! So we fetched Gene's colleague, writer editor Bill Hardy, from upstairs. Not just is he fluent in Spanish he worked at CIP for 4 years. We had a room full of people most of whom had no idea that this was on the cards at all.

Rolando entertained us by rehearsing some of the fun moments when he and Roy had to write messages saying "I CAN'T HEAR YOU" to each other on whiteboards and we congratulated him on his dogged perseverence. Zeppy Vilan, who was online with us from ASTI in Quezon City, disappeared for a while (lunch I think, not a technical problem), but he's visible in the attached photo which shows two views from IRRI's AG rooom and two of Rolando at CIP and one of Zeppy. We only used one projector here but could have used two if we'd needed to conference in additional sites or project a powerpoint presentation or something like that."

IRRI and CIP are collaborating on an ICTKM project on Enterprise Security and Business Continuity (ESBC). We expect and plan that consultants from the Philippines (from SGV & Co.) will be able to "work abroad but stay in the Philippines" when they work with CIP staff in the coming months. It remains to be seen if they will stay in Los Banos on Peruvian time however (there's a 13 hour time difference). CIMMYT will be the next CGIAR Center with an Access Grid connection. CIAT and ICARDA are in the process of establishing advanced research network links required to support Access Grid connections. The ICTKM program of the CGIAR is funding the Advanced Research Networks project and the ESBC project. Both projects are led by IRRI IT Head, Paul O'Nolan.

First Access Grid video conference in the CGIAR: IRRI and CIP; 7 April, 2006

Promotional material

A brochure summarizing the importance of ARN connectivity was prepared and circulated to CGIAR IT staff in April 2005. Although it was a modest production it seemed effective in finally persuading some centers, notably CIMMYT, to take the plunge and apply for project funds.

IRRI's Access Grid Room with Access Grid hardware shown in back room (also used for rear projection)

The Access Grid System at IRRI

Report on CIP ARN Link

The following report was received from CIP in July 2005

CGIAR ICT-KM ARN project

The established goals of the CGIAR ICT-KM ARN project have been fully achieved, and the groundwork is in place for significant further advances in 2005.

- **CIP has completed the installation of an Internet2 connection in partnership with the Agrarian University and the Peruvian university research networks consortium (RAAP).**

  The connection is currently in test, and will be available for PC users within the CIP Lima HQ network during July 2006. After IRRI, CIP is No2 in CGIAR to incorporate within the global Internet2 network.

- **A prototype Access Grid video conferencing workstation has been set up and is currently being tested in ITU.**

  Pending the arrival of imported components, a production system is planned for implementation in August 2006. This facility will enable cooperation between IRRI and CIP Research Informatics Units, within the GCP SP4 HPC and SP5 projects, and this will be announced at the September 2005 GCP annual meeting in Rome.

It appeared that July 2006 was a typographical error. It actually took until September before CIP was connected to Internet2. However, the connection was essentially for demonstration purposes only. No domain name resolution service (DNS) was available on the link so it’s utility was rather limited at first.

Report on CIMMYT ARN Link

A letter of agreement was signed in April 2005 and funds were disbursed to CIMMYT.

A few months later CIMMYT reported:

From: Lopez, Carlos Gabriel (CIMMYT) [mailto:c.g.lopez@cgiar.org]
Sent: Tuesday, 2005 November 01 7:55 AM
To: O'Nolan, Paul (IRRI)
Cc: Van Weerdenburg, Martin (CIMMYT)
Subject: RE: ARN status

Dear Paul,

There are two major tasks to be carried out for the implementation of the ARN project: the incorporation to CUDI, and the implementation of the E1- 2 Mb link. This is to report you the status of both activities:

1. After a long proceeding, last week CIMMYT has been accepted as an affiliate member of CUDI: The consortium of Universities and Advanced Research Institutions in Mexico that promote Internet II. This is a mandatory prerequisite to implement Internet II in Mexico. Today we have paid the annual membership.

2. Engineering studies are been conducted at HQ by Avantel and TELMEX the only two Internet II connectivity providers in Mexico. We are in the process of getting the better deal from them to make a decision on the best technical and economical proposal.

Regards,
Carlos

More recently CIMMYT reported as follows:

From: Lopez, Carlos Gabriel (CIMMYT)
Sent: Tuesday, 2006 March 21 3:54 AM
To: O'Nolan, Paul (IRRI)
Subject: RE: ARN report

Dear Paul,

The Internet 2 link is in the final deployment step. We have not announced this to the CIMMYT community yet because we are replacing our current ATT Internet link with UNINET. This is part of a major effort to reduce ICT costs by renegotiating with our ISPs for getting better prices. The ATT Internet link provides us the set of IP numbers that we need to publish to CUDI (Consortium of Mexican Universities to Promote Internet 2) so that we can implement NAT tables with Internet/Internet 2 numbers. We are finalizing the contract with UNINET and once we have the IP numbers from them, we will announce the availability of Internet 2 and therefore we can make that ARN VC. I hope this to happen in the next three weeks.

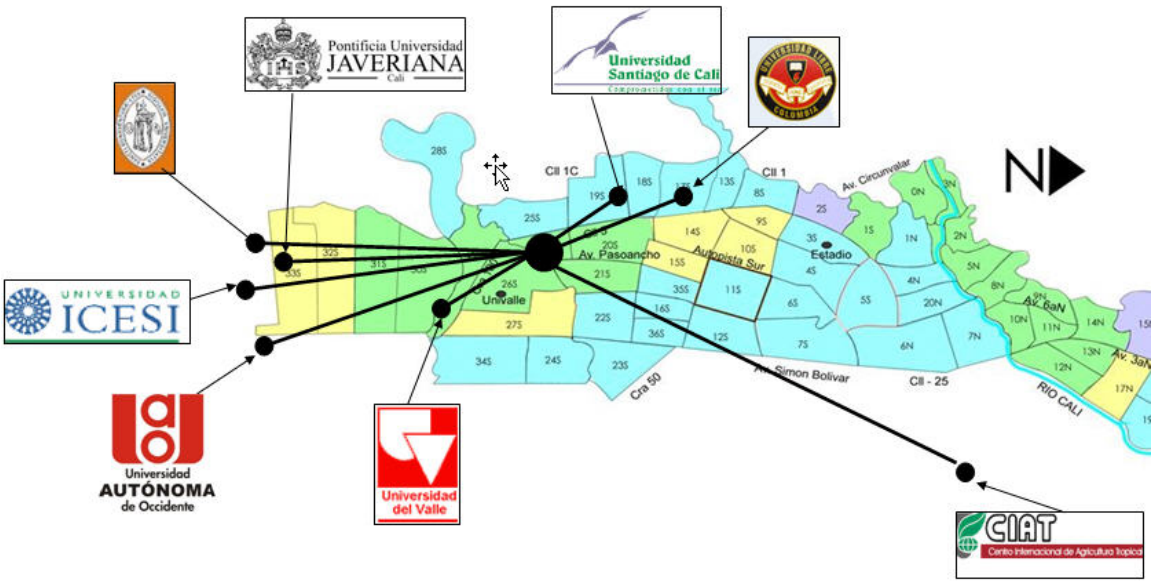
Regards,
Carlos

IRRI has yet to have a video conference or data exchange with CIMMYT via ARN but CIMMYT has used ISABEL software and ARN connectivity successfully to interact with CIP and CIAT and others. Given reports that CIMMYT will implement Access Grid technology in future we can likely expect some 3 way conferences between IRRI, CIMMYT and CIP before the year is over. Of course each institute will have greater need to conference with external partners but it's notable that Access Grid is emerging as a killer application for ARN bandwidth and that this project has stimulated the centers to make the investments needed.

Report on CIAT ARN Link

A letter of agreement was signed in April 2005 and funds were disbursed to CIAT

Later Carlos Meneses, CIAT's CIO reported that the government launch of the new network for high speed networking took place in the first week of February 2006.



RUAV2: Red Universitaria de Alta Velocidad: Colombia's national research and education network, commissioned in February, 2006.

Later, on March 20th CIAT reported:

From: Meneses, Carlos (CIAT)
Sent: Monday, 2006 March 20 9:18 PM
To: O'Nolan, Paul (IRRI)
Subject: RE: ARN report

Dear Paul,

Regarding ARN, I'm going to send the documentation support in a few days.

Now we already finished with the installation: Fiber optical, IPV6, small switch and memory for the routers. Additional last friday we already received from

CLARA the IP numbers, on wednesday we are going to have video conference test.

Sorry for my delay, but it's no easy to deploy this kind of project, even more when the public universities and the government are involved on it.

CarlosM

CIAT has connected with CIP, CIMMYT and others using ISABEL software video conferencing software. No plans for an Access Grid node have been made to date.

Report on ICARDA ARN Link

The following report was received from ICARDA at the end of March 2006.

Completed Activities

1. Agreement with Aleppo University on the link between ICARDA and Aleppo University. As part of this agreement, ICARDA will give access to the researchers and staff at the College of Agriculture access to ICARDA library to enable them to access various research documents and search through the library.
2. Site survey was carried out to determine how best to link ICARDA and Aleppo University. It was decided to use an E1 channel currently available on our microwave link to Aleppo City and a special cable link with suitable high speed modems from the exchange at which the microwave link terminates to the location of the Computer Center in Aleppo University.
3. The equipment required were identified and the specifications were developed. Offers were solicited from vendors and the order was made. Partial delivery has already been made by the vendor and it is expected that the rest of the equipment will be soon delivered.
4. The order for the communication lines was raised to the Syrian Telecommunications Establishment and agreement was reached on how to handle.

Pending Activities

5. Delivery of the rest of the equipment

6. Installation of the equipment at both ends.

7. Installation and testing of communication lines

8. Testing of the entire link

9. Testing the ARN connection through Aleppo University

No target completion date has been given to date. As the university link is already operational the key to getting ARN connectivity at ICARDA appears to be obtaining the equipment.

3.2.3 Project Design and Implementation

3.2.4 Reflection and Learning

If, as seems likely, the number of Centers with advanced research network connectivity increases to five in 20006 the year may be seen in future as a tipping point in getting the CGIAR connected to such networks, as 1992 and 1995 in getting email and Internet connectivity on the desks of CGIAR staff.

The adoption was something that likely could not have been done any faster. In retrospect, it seems that some external expectations may not have taken into account the realities of working in the countries where the CGIAR operates. If even Microsoft, with effectively unlimited resources, can't deliver new products on time, I hope the CGIAR's donors may understand that patience is also required with projects of this nature.

Some interesting synergies are emerging from the ICTKM projects in supporting each other and shared agendas of the centers:

- The ARN project will support the ESBC project, as noted in the IRRI Bulletin. The investment in the grid cluster made by the ARN project was largely to add resilience to it in the interest of data protection.
- The VRC project's CGXCHANGE portal has been used by the ESBC project
- CIMMYT's decision to get ARN connectivity and expected adoption of Access Grid technology is seen as critical to the success of the joint IRRI CIMMYT Crop Research Informatics Laboratory

CIP has become a recognized leader in Peru and has had a deputation from the other institutes with ARN connectivity visit to learn about Access Grid technology. While CIP was having trouble with Access Grid software it adopted and quickly exploited another less complicated open source video conferencing solution called ISABEL and it used this regularly. We can recommend in future that centers who start out working with Access Grid software start with this rather than resort to it after they have failed to make Access Grid work first time.

4.0 Annexes

Appendix 1

Agriculture Research Network and Agroinformatics: An Inevitable Partnership

International Rice Research Institute
Consultative Group on International Agricultural Research

The communication problem

There are a number of communication challenges facing researchers in the CGIAR today:

- The need to exchange large volumes of data
- The need to transfer data rapidly, especially in the context of grid computing
- The need to backup data offsite
- The need to communicate one on one and in groups in real time (typically using teleconferencing or videoconferencing)

The Internet supports all of these needs, but always with limitations, invariably related to limited bandwidth and congestion. While capacity on the Internet will continue to increase, and costs decrease, it is not a suitable media for the heavy demands of leading-edge research in Future Harvest Centres

The solution

Research networks exist today with much higher capacity than the Internet, and the CGIAR can meet its needs by participating.

Governments, universities and research institutions in many countries have developed high speed National Research and Education Networks (NRENs). Some examples are:

- Internet2 (http://www.internet2.edu) is a research and development consortium led by 207 US universities working in partnership with industry and government to develop and deploy advanced network applications and technologies
- The GEANT project (http://www.geant.net/), on the other hand, is a collaboration between 26 NRENs representing 30 European countries, the European Commission, and DANTE (which operates the Pan European NRENs - http://www.dante.net/). Its principal purpose is to develop the GÉANT network - a multi-gigabit pan-European data communications network, reserved specifically for research and education use. It reaches out to the Asian Pacific countries through TIEN2 (http://www.tein2.net/) for the benefit of developing countries in Asia. It also reaches out to Latin America with its ALICE Project (http://www.dante.net/server/show/conWebDoc.156)
- The Asia Pacific Advanced Network (APAN: http://www.APAN.net/) connects research and academic institutions in countries in Asia and the Pacific region
- Proyecto CLARA (Cooperación Latinoamericana de Redes Avanzadas) serves Latin America (see http://alice.dante.net/server/show/conWebDoc.1165).
- CANARIE (http://www.canarie.ca/about/index.html) in Canada
- CUDI (http://www.cudi.edu.mx/) in Mexico

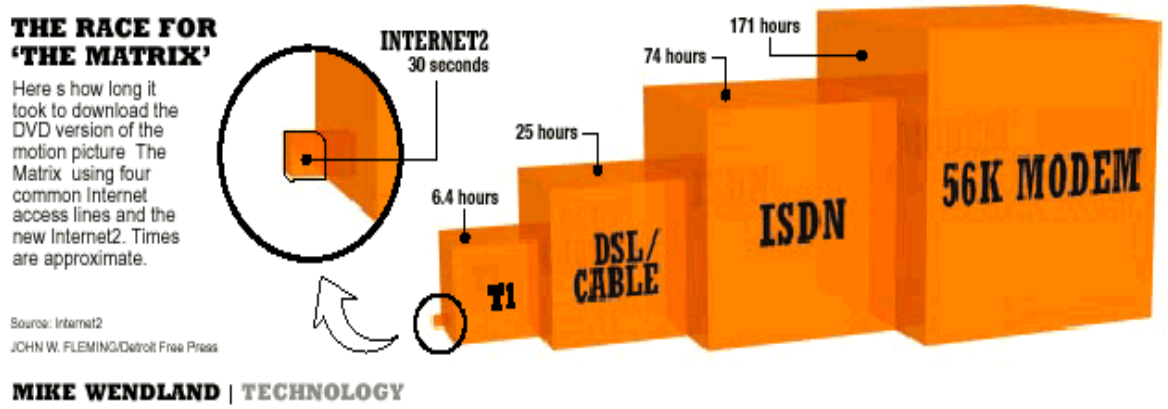
These networks are interconnected, making it possible to work globally.

Generally, NRENs have the following objectives *(Meeting the Innovation Challenge, The R&E Networks of Canada, Phil Baker, President/CEO, October 28, 2004):*

- link researchers, educators and learners globally, driving new knowledge exchange;
- connect and support leading research projects/activities and major scientific, data and computational resources;

- enable new ways to access learning resources regardless of distance and time;
- foster and support public/private sector collaboration on next generation applications, services and technologies; and
- help establish innovative research and education collaborations across disciplines and across regions.

The image below shows the greatly improved transmission speed in comparison with the standard Internet.



NRENs are funded and supported by their respective governments. Activities of these NRENs support and sometimes lead government initiatives to strengthen specific sectors of society, including agriculture. NRENs also serve as catalysts and implementers of government policies benefiting their respective constituents. It is appropriate for the CGIAR to integrate its activities with those of the other members of NRENs. A possible network scenario is shown below: