

# Internet Data Distribution – extending real-time data sharing throughout the Americas

T. Yoksas<sup>1</sup>, W. Gambi de Almeida<sup>2</sup>, D. Garrana Coelho<sup>3</sup>, V. Castro Leon<sup>4</sup>, and T. Spangler<sup>5</sup>

<sup>1</sup>Unidata Program Center/UCAR, CO, USA

<sup>2</sup>Centro de Previsão de Tempo e Estudos Climáticos/INPE, São Paulo, Brazil

<sup>3</sup>Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>4</sup>Universidad de Costa Rica, Costa Rica

<sup>5</sup>COMET/UCAR

Received: 15 September 2005 – Revised: 5 March 2006 – Accepted: 5 March 2006 – Published: 6 June 2006

**Abstract.** The Unidata Program Center (Unidata) of the University Corporation of Atmospheric Research (UCAR) is involved in three international collaborations whose goals are extension of real-time data delivery-to and sharing-of locally held datasets-by educational institutions throughout the Americas. These efforts are based on the use of Unidata's Internet Data Distribution (IDD) system which is built on top of its proven Local Data Manager Version 6 (LDM-6) technology. The Unidata IDD is an event-driven network of cooperating Unidata LDM servers that distributes discipline-neutral data products in near real-time over wide-area networks. The IDD, a collaboration of over 150 mostly North American institutions of higher education, has been the primary source of real-time atmospheric science data for the US university community for over a decade. In addition to providing a highly reliable mechanism for delivering real-time data, the IDD allows users to easily share locally held datasets.

discipline-neutral data products in near real-time over wide-area networks.

The IDD was developed in the early 1990s in response to challenges related to weather-data ingest via satellite broadcast (e.g., local sources of terrestrial interference, data outages caused by solar occultation, weather-related outages due to signal degradation, and the difficulty in locating satellite reception systems near departmental computing resources) and to provide access to datasets that were not commonly available. Starting with a modest goal of internet delivery of data available in the US National Weather Service's Family of Services satellite broadcast, the IDD has grown to become the leading Internet2 advanced-application.

The Unidata IDD has expanded from a US-centric delivery system to one that includes 13 countries on 5 continents. Additionally, the LDM is being used as the data distribution engine in systems akin to the Unidata IDD: by private industry; by several US government agencies including the National Weather Service and NASA; and by the national weather services of South Korea and Spain.

A collaboration among Unidata, Brazil's Centro de Previsão de Tempo e Estudos Climáticos (CPTEC, a division of the Instituto Nacional de Pesquisas Espaciais), the Universidade Federal do Rio de Janeiro (UFRJ), and the Universidade de São Paulo (USP) has resulted in the creation of a Brazilian peer of the North American IDD, the IDD-Brasil.

Unidata, the Universidad de Costa Rica (UCR), and the University of Puerto Rico at Mayaguez (UPRM) are working to extend IDD data sharing throughout Central America and the Caribbean.

Collaboration between Unidata and the World Meteorological Organization (WMO) Regional Meteorological Training Center (RMTTC) in Barbados, the Caribbean Institute for Meteorology and Hydrology (CIMH) was launched in November, 2004 to investigate the expansion of IDD data sharing throughout Caribbean RMTTC member countries.

## 1 Introduction

The mission of the Unidata program of the University Corporation for Atmospheric Research (UCAR) is to provide universities with innovative applications of current computing and networking technologies to access and use atmospheric and related data for education and research.

The development in the United States of the NSFnet and its successors provided a substrate on top of which a multi-way communications system could be built. The Unidata Local Data Manager (LDM) software has evolved to be the vehicle that enabled the multi-way sharing of data in the Unidata community through a project known as the Internet Data Distribution (IDD) system. The IDD is an event-driven network of cooperating Unidata LDM servers that distributes

*Correspondence to:* T. Yoksas  
(yoksas@unidata.ucar.edu)

These last two efforts are aimed at the creation of a Central American peer of the North American IDD, the IDD-Caribe.

The two year old IDD-Brasil and incipient IDD-Caribe are now delivering, in near real-time, the full set of Global Telecommunications System (GTS) observations and US National Centers for Environmental Prediction (NCEP) and UK Met Office global model output available in the US NOAAPORT data stream, and all GOES-East imager channels to top-level relay nodes hosted at prominent national universities and research institutes. The most active of these efforts, the IDD-Brasil, is relaying data to a rapidly growing community of university users in Brazil, Portugal, Argentina, and, most recently, Chile. The IDD-Brasil is also being used to share in near real-time Brazilian mesonet data that are not available through the GTS or other sources. Local, regional, national, and global model output produced by one or more of the CPTEC, the UFRJ, and the USP are also being made available in the IDD-Brasil.

The international expansion of the IDD is, in part, an outgrowth of the MeteoForum pilot project (Spangler and Fulker, 2001; Laing et al., 2002; Yoksas et al., 2004) being conducted by the Unidata and COMET programs of UCAR. MeteoForum is designed to assist WMO RMTCs to improve service in the areas of hydro-meteorology, agriculture, and disaster management. By accessing comprehensive collections of training materials and real-time and historical data, the RMTCs will be able to enrich the education they offer to hydro-meteorological professionals in their region.

## 2 Unidata's role

The existing North American Unidata community has grown into the vital entity that it is through a variety of efforts spearheaded by the NSF-sponsored UCAR Office of Programs Unidata project whose goals are:

- Establishment of an easy-to-use data distribution network that provides access to a wide variety of real-time, atmospheric science-related datasets
- Provision of data analysis tools and high quality support services for use of those tools
- Encouraging of active participation through community-based governance and guidance

The international expansion of the IDD began in earnest as the first phase of the MeteoForum pilot project. Unidata's role in MeteoForum is particularly well aligned with its primary mission, the provision of: *Data, tools, and community leadership for enhanced Earth-system education and research*

Unidata conducts the following activities in MeteoForum:

- Facilitate data access to a broad spectrum of observations and forecasts

- Coordinate a data-relay network that collects and distributes data in real-time at no cost to educators and researchers
- Build a community where data, tools, and best practices in education and research are shared
- Support faculty at research and educational institutions in the use of Unidata systems

## 3 International data relay inauguration

Since Unidata's forte is interacting with the university community, it has concentrated its MeteoForum efforts in the extension of real-time hydro-meteorological data delivery to WMO Region III and IV RMTC-associated universities:

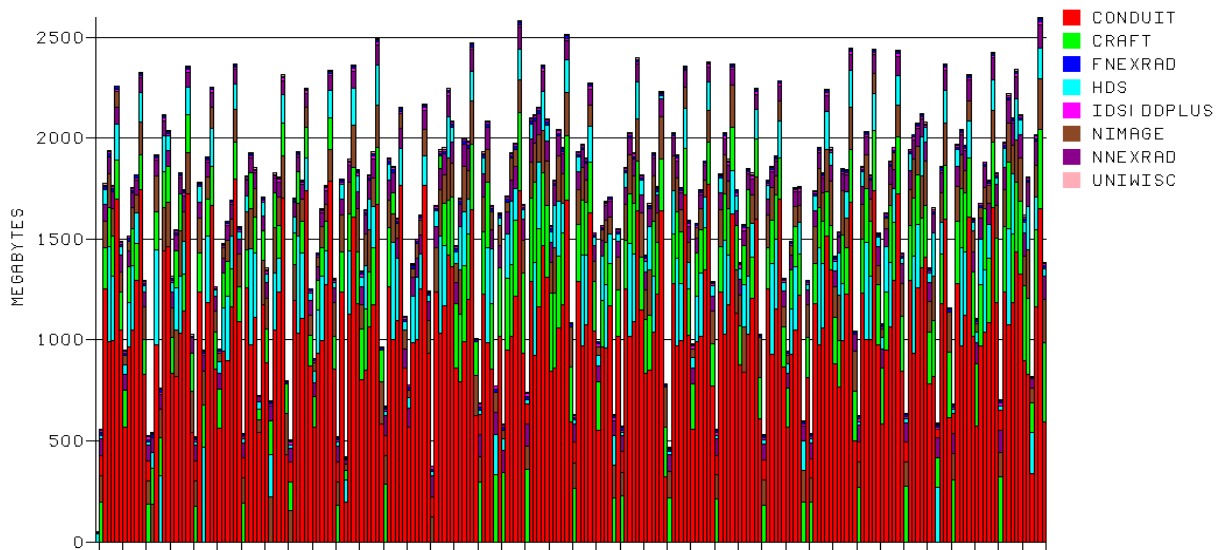
|            |  |
|------------|--|
| Argentina  | Universidad de Buenos Aires (UBA)      |
| Barbados   | University of the West Indies (UWI)    |
| Brazil     | Universidade Federal do Pará (UFPA)    |
| Costa Rica | Universidad de Costa Rica (UCR)        |
| Venezuela  | Universidad Central de Venezuela (UCV) |

In the fall of 2001, Brazilian participation in the IDD was simultaneously inaugurated at the UFPA and the UFRJ with the installation of the Unidata LDM-5 (Davis and Rew, 1994). Data relay tests reinforced previous observations (Chiswell, 1999) that the delivery engine behind the IDD, the LDM-5, was inefficient when relaying data between electronically distant machines. Counter-intuitively, relaying data to a sequence of intermediate hosts actually improved the end-to-end performance of the IDD when using LDM-5.

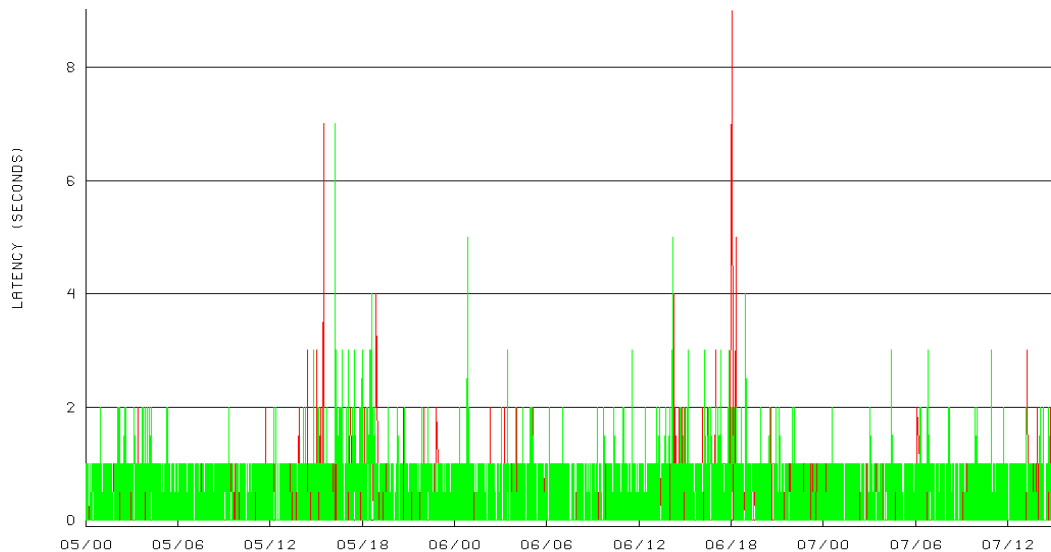
Lessons learned in the UFRJ/UFPA data relay tests were combined with independent results at the Hong Kong University of Science and Technology, and at the University of Melbourne (Melbourne, Victoria, Australia) in a redesign of the LDM. This effort resulted in a next-generation LDM, the LDM-6 (Unidata internal communication), that is able to relay *substantial* volumes of data to both local and remote sites with little to no latency. (Latency is the time difference from when a product is first injected in the IDD and the time the product is received). Stress testing at Unidata in summer, 2005 demonstrated that the LDM-6 is capable of relaying well over 5 terabytes of data per day without introduction of product latencies off of Unidata's IDD relay cluster.

## 4 Network infrastructure is crucial

An element critical to success real-time data delivery is the availability of good-quality, high-speed networking at IDD participating institutions. High speed Internet2 interconnects between Brazil and the US were made possible by connections provided by Brazil's Rede Nacional de Ensino e Pesquisa and the Academic Network at São Paulo through the NSF-funded America Pathways project of Florida International University and Global Crossing.



**Fig. 1.** Time series of ingest data volume as a function of time and color coded by data type.



**Fig. 2.** GTS global observation latencies measured during UFRJ IDD stress testing.

Figure 1 shows hourly data volumes received by the UFRJ IDD node over a ten day period at the end of December 2003. The time series shows that over 1.5 GB of data was received each hour during the test, and peak ingest rates exceed 2.5 GB per hour.

The GTS global observations time series depicted in Fig. 2 shows that the great majority of data products were received with latencies of one second or less.

In fall, 2003 co-author Waldenio Gambi de Almeida, a member of the data section of the CPTEC (<http://www.cptec.inpe.br/>), learned of the data relay experiment being conducted by Unidata and UFRJ while attending the fall series of Unidata training workshops in Boulder, Colorado. Since

the IDD offered a means by which CPTEC could receive a reliable stream of real-time hydro-meteorological data, and since it also provided a mechanism that allowed for sharing of locally held datasets, CPTEC joined in the data relay experimentation and offered to act as a top-level IDD relay node and perform outreach to institutions like the USP.

The ability of the UFRJ and CPTEC to actively participate in the peer expansion of the Unidata IDD was a direct result of their excellent network connections to Internet2.

The data relay collaboration among Unidata, the UFRJ, CPTEC, and USP was named the IDD-Brasil.

## 5 Inauguration of the IDD-Brasil

One of the first parts in the establishment of the IDD-Brasil was the creation of *principles of participation*:

- For the most part, participants must be educational institutions
- Participants must acquire and maintain appropriate computer hardware and Internet access
- Real-time data must be relayed free-of-charge
- Cost of participation is sharing of locally-held datasets with fellow participants
- Top-level relays must take ownership of the expansion and support processes

Efforts aimed at broadening participation in the IDD-Brasil have been ongoing since its inception. Co-authors Waldenio Gambi de Almeida (CPTEC) and David Garrana Coelho (UFRJ) have been promoting the benefits of participating in the IDD-Brasil and in the usefulness of Unidata display and analysis systems through discussions with a variety of Brazilian universities. This effort has been very successful: currently, Brazil ranks second only to the US in IDD participation.

In summer, 2005 CPTEC held the first Brazilian training workshop on the use of the LDM-6/IDD-Brasil, and in the application of Unidata's distribution of the GEMPAK analysis and visualization package in support of weather forecasting. CPTEC has plans to continue these courses at their training facilities in Cachoeira Paulista and eventually to extend the courses to cover use of the next-generation visualization package from Unidata, the Integrated Data Viewer.

The first non-Brazilian university to connect to the IDD-Brasil was the Universidad de Avarez (Portugal); the second was the Universidad de Buenos Aires (Argentina); and the third was the Universidad de Chile (Chile). More foreign participants are expected to join the IDD-Brasil as outreach efforts continue.

## 6 Central American data sharing expansion

In parallel with the establishment of the IDD-Brasil, the University of Puerto Rico at Mayaguez was approached to help relay real-time atmospheric science data via the Internet to Caribbean and Central American universities.

When Internet delivery of real-time data is not possible and when a site is located within the NOAAPORT broadcast footprint, Unidata recommends the installation of satellite-based data reception system. In February 2004, Unidata worked with the Universidad de Costa Rica to install a Unidata-designed and built NOAAPORT satellite ingestion

system on the UCR campus in San Jose. With the installation, the UCR was able to ingest real-time global observations and NCEP model output for use in education and research for the first time. The UCR has agreed in principle that, as its Internet connectivity improves, it will assume a leading role in extending access to its real-time meteorology data to Central American universities that also have sufficient Internet connections.

In November, 2004 Unidata began working with the Caribbean Institute of Meteorology and Hydrology, to test IDD-delivery of real-time data into Barbados. These tests have shown that data delivery is possible, but not spectacular given the limited, 256 Kbps network connection that the CIMH currently has to the global Internet.

Ultimately, the success of the incipient data distribution/sharing efforts among the UPRM, UCR, CIMH, and Unidata university community, named the IDD-Caribe, will depend entirely on the quality of network connections available at participating sites.

## 7 Building communities through data sharing

The same approach used to create the vibrant North American Unidata community is now being extended to research and education institutions in WMO Regions III and IV. The establishment of the IDD-Brasil and IDD-Caribe represent a first step in this community building process.

The second step in the community building process is encouraging the sharing of locally-held datasets by all participating sites with colleagues throughout the extended community. The need for full and open international exchange of environmental data has been articulated in many documents. In order to understand the myriad complexities of the earth system and the way the different elements interact, it is crucial to collect, share and analyze environmental data from all parts of the world.

The US Unidata community recognizes the importance of access to environmental data from other countries, particularly those in the southern hemisphere, to enhance education and research capabilities toward studying global problems like climate change, ozone depletion, and the El Niño-Southern Oscillation. A critical requirement for such research studies is the acquisition and assimilation of a complete spectrum of global meteorological, oceanographic and hydrological observations. The data relay infrastructure that has been installed in Brazil, Costa Rica, Puerto Rico and Barbados represents the beginnings of a hemispheric-wide network that can act as conduit for multi-way sharing of locally held data sets among peers in an expanded Unidata university community.

## 8 Conclusions

Unidata plans to continue playing an active role in establishing a robust IDD relay network for South America in collaboration with the UFRJ, CPTEC and the USP. Similarly, it intends to continue working to establish a data sharing network for Central America and the Caribbean. Participation in the 2004 El Niño conference of the Red de Universidades del Pacífico Sur (the South Pacific University Network) in Santiago, Chile was used to inform leading Latin American universities of the availability of Internet-delivered, real-time hydro-meteorological data through the IDD and IDD-Brasil networks. The Universidad de Chile has recently joined the IDD as a result of this outreach effort. The Universidad de Simon Bolivar (Venezuela) is scheduled to join the IDD as a result of this outreach effort in fall, 2005.

It is envisioned that the data sharing capabilities available in the North American IDD, the IDD-Brasil, and the IDD-Caribe will help foster new collaborations among prominent universities, national meteorological agencies, and WMO Regional Meteorological Training Centers throughout North, Central, and South America.

*Acknowledgements.* The Unidata Program Center is funded by the US National Science Foundation. COMET is funded by the US National Weather Service.

Edited by: E. Cutrim, M. Ramamurthy, S. Nativi, and L. Miller

Reviewed by: anonymous referees

## References

- Chiswell, S. R. and Miller, L.: Toward Expanding Model Data Access Through Unidata's Internet Data Distribution System, Preprints, 15th International Conference on IIPS for Meteorology, Oceanography, and Hydrology, 441, 1999.
- Davis, G. P. and Rew, R. K.: The Unidata LDM: Programs and Protocols for Flexible Processing of Data Products, Preprints, 10th International Conference on IIPS for Meteorology, Oceanography, and Hydrology, 131–136, 1994.
- Laing, A. G., Miller, L., Spangler, T. C., and Fulker, D.: MeteoForum: A Pilot Project for Advanced International Meteorological Training Centers, Preprints, 18th International Conference on IIPS for Meteorology, Oceanography, and Hydrology, 326–327, 2002.
- Spangler, T. C. and Fulker, D.: MeteoForum – an International Network of Meteorological Training Centers for the 21st Century, A Pilot Proposal developed jointly by the COMET® and Unidata Programs of the University Corporation for Atmospheric Research, Boulder, CO, USA, March 2001.
- Yoksas, T., Coelho, D. G., Spangler, T. C., Cutrim, E., 2004: MeteoForum – Initial Successes in Data Sharing Leading to the Creation of the IDD-Brazil, Preprints, 20th International Conference on IIPS, Seattle, WA.