Unidata 2013
University Science Survey Results

March 2013

Unidata Program Center
UCAR Office of Programs
P.O. Box 3000
Boulder, CO 80307-3000
www.unidata.ucar.edu

Mohan Ramamurthy, Director

1 Introduction

In early 2013, the Unidata Program Center surveyed members of the Unidata academic community to form a picture of how Unidata software and data services are being used by working scientists and educators.

The survey was conducted via electronic mail and telephone contacts with 21 professors who are members of the Unidata community associated with university departments that grant advanced degrees in the atmospheric sciences or related disciplines. Of the professors contacted, 13 were able to participate.

The survey itself consisted of several open-ended questions:

1. How many PhD candidates in your department are using Unidata software or data services as part of their doctoral theses? How many Masters candidates? How many postdocs?

2. What types of scientific projects are Unidata software and data services aiding in your institution?

3. How is the process of "doing science" changing in your institution? How can Unidata help meet scientists' changing needs for geoscience data?

4. How is the process of teaching science changing? How can Unidata help meet the needs of geoscience educators?

Responses to many of these questions were more anecdotal than rigorous, and no effort was made to enforce consistency between respondents. This document contains a brief summary of the responses to the survey, stripped of identifying data.

In addition to the above questions, we asked respondents to suggest any scientific projects at their institutions in which Unidata technologies have played an important supporting role, as a basis for further research by UPC staff. These suggestions are addressed not addressed in this document.

2 Use of Unidata Software and Data Services

Respondents reported that in their departments, a total of more than 500 masters and PhD students made use of Unidata software and data services in their academic work. (Not all respondents reported use by post-doctoral scientists; those that did reported a total of 34 using Unidata software and services in their research.)

Of these, the large majority (roughly 80% of students and postdocs) make use of data stored in Unidata’s netCDF data format. The netCDF community is large and vibrant, and many software tools for working with netCDF files exist. Several respondents reported that “nearly everyone” in their department uses data in netCDF format, adding that many may be using netCDF without realizing the format’s connection to Unidata.

Use of data supplied via Unidata technologies is also nearly ubiquitous in the respondents' departments, with 11 of 13 reporting local use of the Local Data Manager (LDM) and Internet Data Distribution (IDD) network, and 5 of 13 reporting use of either the THREDDS Data Server (TDS) or RAMADDA to access data.

Respondents described less extensive use of Unidata-supplied visualization software (IDV, GEMPAK, and McIDAS). Of these, 10 of the 13 respondents reported use of GEMPAK for either teaching or research in their departments. The IDV was mentioned by 7 of 13 respondents, most often in the context of teaching as opposed to research. Four respondents reported use of McIDAS in their departments.

3 How Unidata Technologies are Used

Survey respondents described a wide variety of activities in which Unidata technologies are used in their departments. These activities, whether in the context of research or education, generally fell into the following broad categories:

- Weather analysis and forecasting
- Real-time weather observation
- Model analysis
- Climate research

4 How are Scientific Activities Changing?

The following are some representative excerpts from the answers to the question "How is the process of "doing science" changing in your institution?"

Large data sets were mentioned frequently:

One huge challenge is how to visualize very large datasets. A model simulation with 10000x20000 grid cells cannot be visualized using IDV very easily, nor will it work with GEMPAK. Resolution continues to increase. We need help storing, manipulating, analyzing, and visualizing these data.

***

Someone analyzing a year of model surface output needs to access roughly 10^10 floating point variables - roughly 100 GB of data. Go for a decade and multiple levels as many modern research projects require and you're talking 10 TB of data. The big issues of science then become:

1. *Where to store that much data?*

2. *Do we move the data to the user or the computation resources to the data? Bandwidth is the constraint nowadays, not cpu cycles or mass storage.*

3. *If we move the data how? The Internet is too slow and no one is yet set up to mail hard drives like we used to mail magnetic tape.*

4. *If we move the computing resources to the data how do we allocate them?*

*****

First thought that comes to mind ... we are increasing being overwhelmed with data and dealing with data. Simple, yet powerful, ways to manage datasets continues to be extremely vital to 'doing science'. Unifying data formats is one way to address this issue along with designing display software that can seamlessly deal with a variety of data formats is another.

*****

I think the need to analyze and interpret ever larger datasets, whether they be observations or models or reanalyses, is one way that things are changing. As a result, students need more skill with data analysis and programming than ever, but it seems that many of the students entering the program are not as adept with this as they probably should be.

*****

We are analyzing and manipulating larger and larger data sets generated by both instruments and models. In the end, we have to condense, consolidate, analyze, and understand this information to the point where we can present it in papers in a few key figures (with possible reasonably sized electronic supplements).

The theme of “data friction,” discussed in Unidata’s strategic plan, also came up:

The distinction between the subdisciplines of meteorology is becoming increasingly blurred, as well as distinctions between the atmospheric and related sciences. Students deal with a diverse range of datasets and increasingly we have students from other disciplines looking to use atmospheric datasets. Ease of access is increasingly important, as is reducing data friction.

*****

In Met research a lot of time is lost in learning how to achieve data access and manipulation, with friction in both command line approaches and GUI interfaces. In data processing, scientists increasingly are migrating to software applications with intuitive user interfaces and extended tool boxes (e.g., Matlab). Some might claim that we are becoming lazy, but I think rather that it is impatience to get work done and results out faster.

*****

I’m also noting an increasing trend of collaborative funded projects where multiple faculty in our department are involved. This at times involves challenges as one person may be most familiar with one type of data set and/or visualization tool, while others use different ones. Each may be unaware of datasets and/or visualization/analysis tools that their colleagues may typically use. Often it is the grad *students who must do the "cross-pollination" of learning each other's preferred data sets and analysis/display packages.*

Among the other challenges noted by the respondents was a dearth of IT support:

[...] a substantial degradation in our computing support infrastructure. The data flood makes it difficult to keep track of what we have available, and interdisciplinary brings with it incompatibilities, inconsistencies and "data friction" in general. The advent of relatively cheap storage actually exacerbates the problem. The infrastructure degradation means students, faculty and researchers have to take on more sysadmin tasks, administering our own machines and dealing with our own installs and upgrades.

5 How is Geoscience Education Changing?

The following are some representative excerpts from the answers to the question "How is the process of teaching science changing?"

Advancing technology was a recurring theme here:

The hardest part is keeping up with the students and the way they work with technology!! We need to integrate those innovations, but not be too 'blinded' by them too. Sometimes the new innovations or current fads in technology, are really just fads and not useful tools for teaching. So, we need to be careful when designing new software that may incorporate these 'bells and whistles' that they really do enhance teaching and help to visualize weather systems and weather events.

*****

The era of the meteorological computer lab may be coming to an end. Students are bringing iPads to class to take notes. They look at meteorological analyses and data on everything from their phones to desktop computers. They were born in the cloud and the idea of having to go to a lab to analyze the weather seems arcane to them.

*****

The ability to plot data on hand-held devices, tablets, ipads, etc. is starting to arrive. Interactive material that interfaces with mobile technology could be helpful in the classroom - now, a smart phone can double as a "clicker".

*****

As is becoming widely known, the students today seem almost "wired-by-birth". They expect to interact with data in a visual manner, and to do so in a very fast and intuitive manner. While traditional means of science education must not be cast away due to current trends, programs such as Unidata must lead the students, not merely attempt to follow them.

The “expanding data volume” theme was also reprised in answers to this question:

We now teach big-data processing and artificial intelligence system building to our undergraduates. That requires timeseries of array or volume data, not just Unidata-style snapshots and case studies. In the traditional (synoptic and mesoscale weather) classes we're making more use of in-house WRF output.