4B.3 ACCESSING THE IPCC AR4 DATA: A DAY IN THE LIFE OF THE GFDL DATA PORTAL

Kevin M. O'Brien*[1], S. Nikonov[2], R. Schweitzer[3], S. Hankin[4], V. Balaji[2], A. Manke[4], J. Li[5], and J. Callahan[1]

[1] University of Washington/JISAO, [2]Princeton University,[3]Weathertop Consulting,[4]NOAA/PMEL,[5]Macrostaff

1. Introduction

After running models and generating copious amounts of data for their contribution to the Intergovernmental Panel on Climate Change (IPCC) AR4 assessment report, the Geophysical Fluid Dynamics Laboratory (GFDL) had a problem. What was the best way to offer data access to the IPCC working group researchers, while simplifying the complex hierarchy of data that naturally evolved as part of running the experiments? Part of the solution was to partner with Pacific Marine Environment Laboratory (PMEL) to create a data portal which would harness the Live Access Server (LAS) technology to aid in data analysis and visualization. The LAS is a highly configurable web server that is designed to provide flexible access to geo-referenced scientific data, and allows browsing, visualizing and comparisons of data on the fly.

In this presentation, we will be discussing the Live Access server and the role it plays in the data portal at the GFDL. As mentioned, the data portal at GFDL is being used to disseminate their model data contributions in support of the IPCC AR4 assessment report. We will take a look at some of the typical research questions an IPCC WG2 scientist would ask and demonstrate how such questions can be answered using the Live Access Server. In addition, we will show how we used the Live Access Server technology in conjunction with the Curator database project to ease the complexity of the Data Portal data hierarchy, both in terms of end user access as well as for potential installers of LAS.

2. Background

As one of the modeling centers that participated in the IPCC AR4 experiment, GFDL was responsible for running several ocean and atmosphere coupled experiments. GFDL itself contributed two sets of 10 different scenario runs, one each using the new family of coupled AOGCMs that is referred to as the CM2.x family. The results of these experiments were both sent to the PCMDI/IPCC archive as well as hosted on GFDL's own data portal. Though the IPCC AR4 model outputs are but one part of the data available through the GFDL data portal, they are the most rapidly growing segment.

* *Corresponding author address:* Kevin M. O'Brien, University of Washington, Joint Institute for the Study of the Atmosphere and Ocean, Seattle, WA, 98115; email: Kevin.M.O'BrienATnoaa.gov

The challenge presented to the GFDL data portal was how to meet the evolving demands for access to the data portal holdings. This includes making available datasets containing specific requests of particular variables and time ranges. In order to give the Data portal users increased options, GFDL turned to the Live Access Server, which is developed at PMEL. Integrating LAS into the data portal gives users more options in terms of browsing, discovering and analyzing data.

One of the challenges facing LAS, from a user perspective, was how to simplify the incredibly complicated hierarchy of experiments that confronted the user. From a technical standpoint, other challenges arose as well. For example, because the volume of experiments was so high, finding a way to automate LAS configuration in way that would support the complex data structure was also paramount.

The solution that presented itself was to integrate the curator metadata database with the configuration stop of LAS.

3. Curator Database and LAS

3.1 Configuring LAS

Typically, to populate an LAS server with datasets, the LAS maintainer will point a utility at the desired dataset (for example, a NetCDF file, or several NetCDF files aggregated through an OPeNDAP service such as TDS) which will generate XML configuration information for that particular dataset.

When the complexity of data holdings rises to such a level as it does for the GFDL data portal, the task of installing these datasets into an LAS can become a time consuming task. In addition, though LAS is designed to guide users through a hierarchical path of data discovery, too many levels of data hierarchy can ultimately frustrate and confuse the user.

The challenge was to find a way we could automate both the generation of dataset configuration for the LAS installer, as well as ease the process of data discovery for the end user.

3.2 Curator Database

The running of the CM2.x family of models at GFDL begins with the Flexible Modeling System (FMS) Runtime Environment (FRE). The details of FRE will be quite lacking here, as that is not the focus of this presentation. For this presentation, it is enough to mention that the brain behind the FRE system is called the curator database. The curator database contains all the metadata required to assemble and run and experiment as well as the information need to locate the results of the experiment on the external data areas of GFDL. It is the centralized metadata storage for the entire modeling process.

3.3 Curator and LAS

The obvious answer to the problem of configuring LAS to efficiently serve the GFDL data portal holdings lay within the guts of curator. After all, curator was the brains behind the whole FRE operation and knew everything that was necessary to successfully configure an LAS.

LASurator (LAS + curator) was born when the generic LAS utility addXML was extended to gather all of it's necessary configuration information from the curator database. Not to be confused with lacerate, "LASurating" was an efficient, automated and simple way to generate XML configuration information for all of the IPCC AR4 datasets that GFDL had exposed to the public via their NOMADS system.

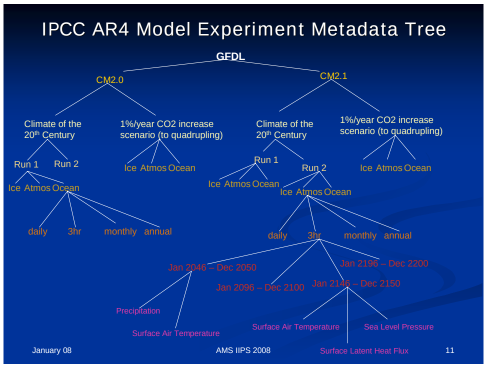

Figure 1. *Snapshot of AR4 Experiment tree*

Figure 1 represents a simplified snapshot of just how complicated the GFDL AR4 experiment tree is. Each one of those nodes actually has many more branches than depicted in the figure above. By combining curator and LAS, it is no longer of any concern how complicated the data hierarchy becomes. As long as the information is maintained and correct in the curator database, it will require simply one command to generate the configuration information needed to populate the LAS.

4. The User Experience

4.1 Fewer degrees of separation

One of the strengths of the Live Access Server is it's ease of customization. It is trivial to create custom categories in LAS that represent the data hierarchy. However, when the hierarchies become as complex as the GFDL data portal, the default configuration can lead to user frustration.

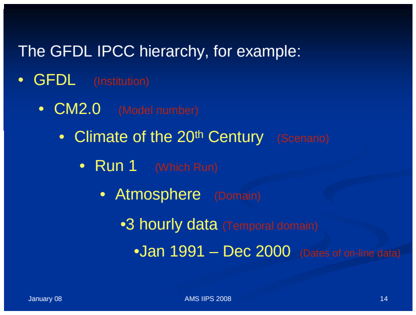

Figure 2. *An example hierarchy*

Figure 2 represents an example hierarchy of one data set in the GFDL data portal. Each line represents a "click" of the mouse a user would have to make in order to reach their final destination. Although not extremely laborious, having to click seven times on the mouse repeatedly for different experiments can raise the level of frustration almost as much as waiting in line at the DMV. One of our goals was to reduce those degrees of separation.

LAS already uses the velocity templating language to render pages. In order to achieve our goal of less separation, we turned to velocity templates. We developed a generic velocity template that, when combined with the information from the curator database, is able to clearly and simply represent all of the data sets under one experiment. The result of applying the generic template to an experiment in the data portal can be seen in figure 3.

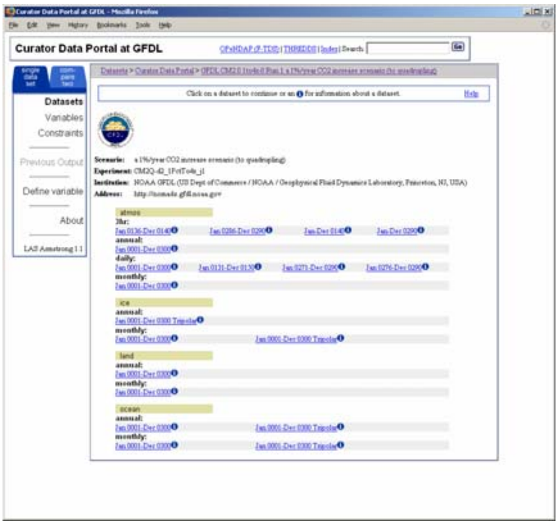

Figure 3. *Redesigned access to a data portal dataset*

What once would have taken seven taps on the mouse button, can now be done in two. In addition, the temporal ranges of each data set that are available are clearly displayed for the user. Remember that often there are datasets that contain custom variables and/or time ranges. With the developed template, this information is right at the users fingertips, and one more click on the information button provides the user with a simple method of discovering the experiment metadata.

4.2 Applying LAS to a specific problem

Now that the entire IPCC AR4 data that is available through the GFDL data portal has been configured into LAS, the next step is to actually use the data for scientific purposes. This will allow the user to utilize the various products that LAS is able to deliver in approaching real world science problems.

An example of such a scenario would be someone who is interested in the evolution of air temperature around the Hudson Bay region. One reason to investigate such questions is concern for how polar bears might be affected by warming temperatures.

The easiest way to describe the abilities that LAS has in investigation such a problem is to show a variety of figures generated from the data portal LAS. What follow is a series of six figures, each with an explanation, actual products drawn from the Live Access Server. All of the following data is taken from the CM2.1 model running the Climate change of the 20th Century scenario. That scenario contains five different runs.

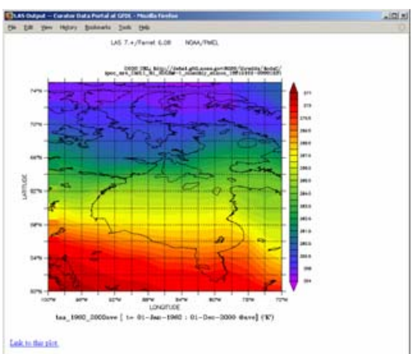

Figure 4. *The averaged surface air temperature from 1960 – 2000 over the Hudson Bay region. The data is from Run 1 of the Cm2.1 Climate of the 20th Century scenario*

Figure 4 shows the surface air temperature averaged over the Hudson Bay region. However, because there are five different runs within this scenario, it is desirable to compare the results from the different runs. Figure 5 shows the difference in the averaged surface air temperature for two separate experiments.

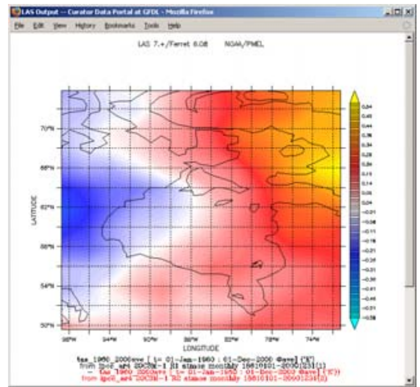

Figure 5. *Difference in averaged surface air temperature from 1960 – 2000 between Run 1 and Run 2 of the CM2.1 Climate Change of the 20th Century scenario's*

Another visualization available through LAS is something called the Slidesorter. The Slidesorter allows the user to look at and compare several different plots on the same screen. This is especially beneficial when comparing results from the different runs of the CM2.1 Climate of the 20th Century scenarios, as in Figure 6

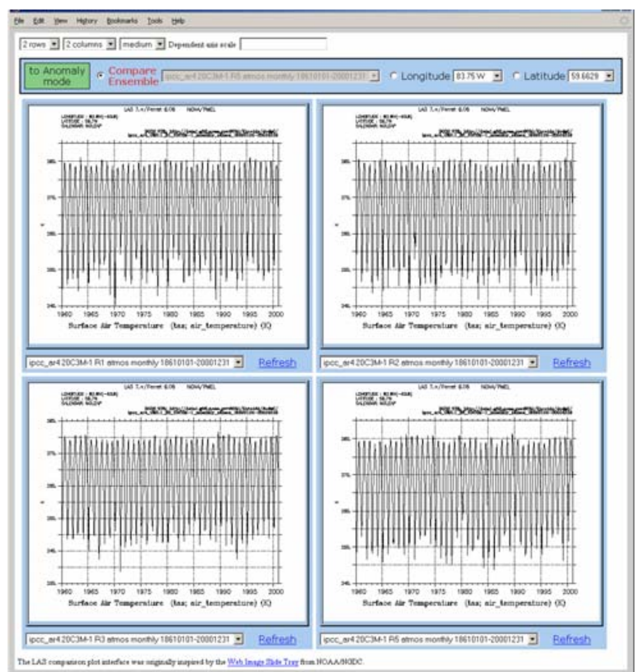

Figure 6. *Comparing time series of surface air temperature between four different runs (R1, R2, R3, R5) of the CM2.1 Climate of the 20th Century scenario.*

The Slidesorter also utilizes an "anomaly mode", which allows for the user to assess change from a baseline experiment, as in Figure 7.

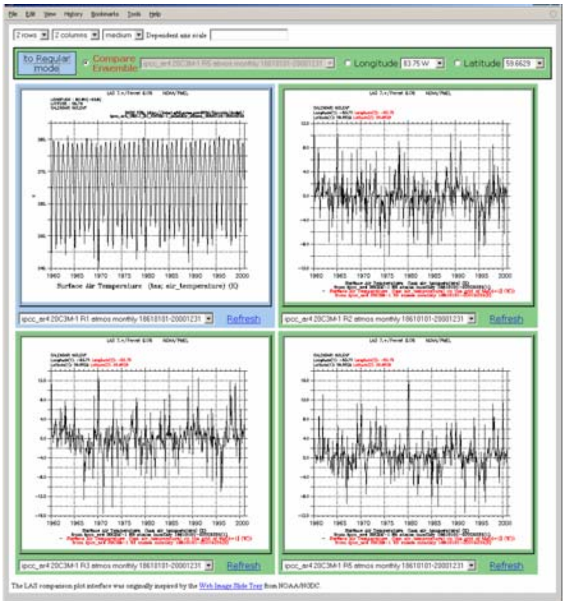

Figure 7. *LAS Slidesorter in "anomaly mode". CM2.1 Climate of the 20th Century Run 1 is set as the baseline experiment. Runs 2, 3 and 5 are then compared to the baseline. A time series of surface air temperature in Hudson bay is compared.*

Additionally, the data portal LAS allows the user to apply different averaging periods to the monthly data available on the data portal (see Figure 8). The different types of averaging include single month averages (for example, every February), three month averages (centered on the month of user choice) and five month averages (again, using the center month of the users choice).

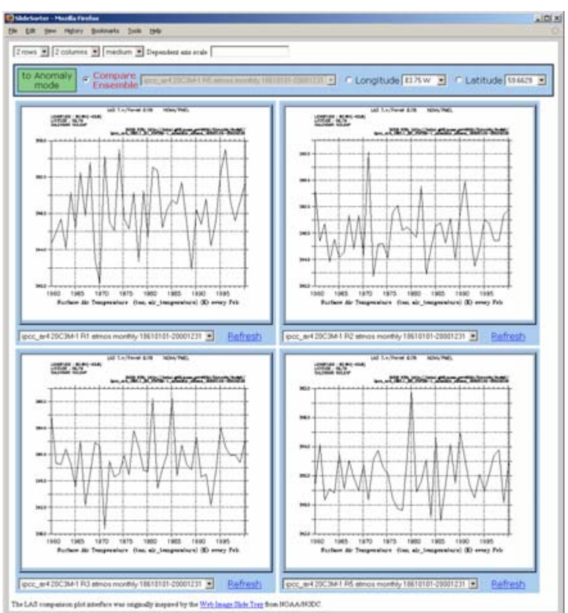

Figure 8. *Surface air temperature from Runs 1,2,3,5 of the CM2.1 Climate of the 20th Century scenario. The time series has been averaged on the fly to show only the February data for the selected time range.*

As previously, with the Slidesorter anomaly mode, it can be very useful to set one experiment as the baseline and compare the single month, February averages, as in Figure 9.

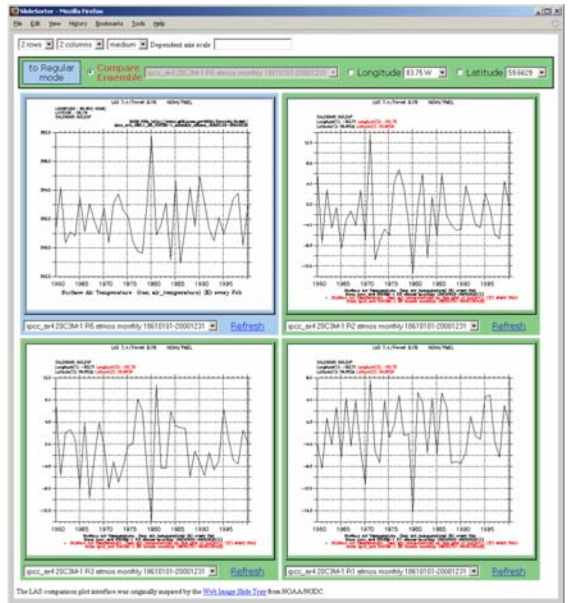

Figure 9. *Surface air temperature from Runs of the CM2.1 Climate of the 20th Century scenario. Run 1 is used as the baseline, and Runs 2,3 and 5 are then compared to the baseline.*

5. Conclusion

Combining the technical forces of GFDL's curator database with PMEL's Live Access Server has resulted in the ability to serve the IPCC AR4 data from GFDL in user friendly, efficient manner. Being able to serve the data in that way ensures that the user is able to harness the capabilities to browse, subset, visualize and analyze the IPCC data in a comfortable and stress free environment. The innovations applied to LAS were designed to ease data interaction for the consumer, as well as provide them with the ability to perform scientific analysis from their own desktop.

As the user community continues to evolve, and the needs of the community evolve with it, LAS and curator will also evolve to meet and even stay ahead of such needs.

6. Acknowledgments

This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA17RJ1232, Contribution #1465, as well as NOAA/PMEL Contribution #3177.