An Earth Data File is Worth a Thousand Pictures:
going beyond multimedia for education

Bruce Caron, PhD
The New Media Studio [www.newmediastudio.org]

This essay explores how incorporating data files into "multimedia" educational applications changes the usefulness of these applications. Giving students direct access to data files through a dedicated "rich-client" (which sounds better than a "thick" client) opens up the possibility of independent discovery, and the beginnings of data literacy.

In the beginning was Macromedia Director™. And it was, well, pretty good, considering. This was in the mid 1980s, and the whole idea of multimedia was emerging as fast as the hardware and content formats could grow. While other multimedia authoring programs (notably Hypercard) would come and go, Director™ developed into the most widely used authoring software for educational (and commercial) multimedia CD-ROMs. John Thompson (pictured on the left; source: http://inventors.about.com/library/inventors/bl_Lingo_programming.htm), the chief scientist at Macromedia, developed the Lingo code that brought interactivity to the software. This meant that pedagogical feedbacks were possible, and the user (the student) could be given a number of pathways through the educational content. As the Internet developed, Macromedia has been migrating the capabilities of Director™ from CD-ROM delivery to this new media. First came Shockwave™, which allowed the user to play a "shocked" Director™ file. And later came Flash™, based on a new vector file format. But the idea was the same: create an authoring platform that can deliver an interactive, multimedia-rich user experience.

At roughly the same time (from the mid 1980s), earth science researchers were using their computers to create images and models and build data resources of data from newly launched earth orbiting remote sensors. The growth of research software, such as MATLAB, IDL, and ARCinfo, enabled a new generation of graduate students to explore data and models in ways their professors had never dreamed a generation before. A new cadre of scientists emerged, and a new science that was data-driven. These scientists were computer and software—and, of course, model and data literate. They would discover new knowledge about the earth system. But their tools, while powerful, grew enormously complex, being built to handle a wide range of data access, visualization and analysis methods.

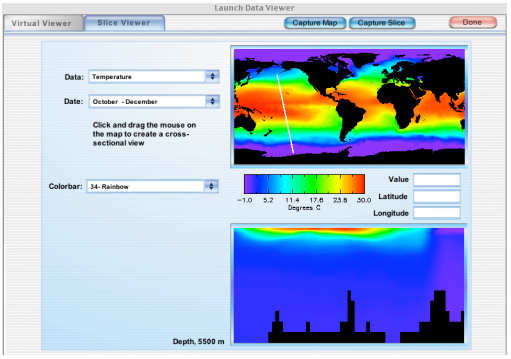
The output of the science, in terms of graphics, has been enormous. NASA's Visible Earth website [http://visibleearth.nasa.gov/], for example, boasts more than 6000 images. And NASA alone has thousands of other WebPages that offer tens of thousands of images. So when it comes to creating a multimedia educational application that contains pictures of the subject under study, there is, at last, a wealth of images available. However, if the goal is to help the student understand not only the content, but also the science that created it, new problems emerge.

In the mid 1990s, under a NASA cooperative agreement, I managed a team that was developing middle-school courseware that included NASA data resources. [NOTE: These are still available: http://www.planetearthscience.com./products.html]. These products were developed using Director™, and were designed to guide the student through the scientific process of data access, visualization, and analysis. For this purpose, the NASA data files were brought into IDL™, visualized, and then images—dozens, hundreds, thousands—were output. These images were then

integrated into the product to simulate the various choices that a scientist would make in using the data. But even with thousands of images, the software could only simulate a small subset of an actual data use case: any decision tree more than a few steps long and with a few interactively available parameters, would create the need for not just thousands but potentially millions of images. In another project designed to simulate a local land-use decision model, we determined that several hundred million graphic images would be needed. It appeared that we had hit the limit on where interactive multimedia could go.

We had been exploring the notion of building a data access and visualization tool for education using Java™, but the task seemed far too ambitious. Stepping back from the problem and looking at the possibility of extending Director™ (rather than starting from scratch), we hit on the possibility of developing a plug-in to this program. And since we were already using IDL™, I wondered if it might be possible to create a "plug-out" of IDL™ that could connect to our plug-in to Director.™ The goal was to have an IDL™ data window (or more than one) inside a Director™ authored application. This window would offer the students the same data analysis capabilities of IDL, but would be preprogrammed for the task at hand.

This data window would give the student a research-level tool in an interactive educational user interface. And the result would be an application where the students can discover data in much the same way as the scientist does. And since a single data set can be re-visualized any number of times, there is no need to pre-script what that choices were for the student. The actual tool would replace the simulation of the tool. [At Left: a data page from our latest tool. The data displays are from actual data sets and the students can get data values using their mouse.] To produce this application we would not need thousands of (pretty) pictures for each data file, but only the one file. And since the data would be on the desktop, the response time would fall in the under one-second range (instead of the multiple second—or even multiple minute—response time typical for thin-clients).



With support from the NSF (DUE #0121550), and from NASA and the ESIP Federation, the Data Discovery Toolkit technology was created. This is now being updated, upgraded, and enlarged to bring new data access, visualization, and analysis capabilities into Director™ and soon into Flash™. The code resources for this technology are freely available on the Data Discovery Toolkit and Foundry NSDL site [http://www.newmediastudio.org/DataDiscovery/]. This is the next generation of multimedia, where media might include images, sound, and video, and also data and links to models.

However, the promise of putting real data into the hands of students can only be fulfilled through an active partnership with teachers and curriculum builders. Giving students an ability to actively discover data in a real tool also means that students will need to learn about real data use and misuse. Data literacy becomes both possible and necessary for the students to learn about the earth through the data they can now manipulate. A data file is certainly worth a thousand pictures, but only to those who have learned to pull information properly from these data.