Discovering Data in the Classroom: The SDSC “Discover Data” Portal

Jeff Sale

The San Diego Supercomputer Center’s Education Group is developing a web-based portal for K-12 and undergraduate educators to integrate real-world data into their curriculum. It is called the “Discover Data” portal. The portal includes standards-based step-by-step lessons with a focus on freely available data archives from a range of scientific disciplines. Many of these archives may be challenging to use in their current form. The lessons are designed to help teachers find their way through them.

The portal is developed with the Moodle learning management system. Moodle is a “free, open source software package designed using sound pedagogical principles, to

The portal is developed with the Moodle learning management system. Moodle is a “free, open source software package designed using sound pedagogical principles, to

continued on page 8

Figure 1: The Discover Data portal at http://education.sdsc.edu/discoverdata/

2006 BioQUEST Workshop
June 10-18, Beloit College

Exploring Complex Data Sets

The 2006 BioQUEST Summer Workshop, “Exploring Complex Data Sets,” invites participants to construct new curricular materials that are data-rich, problem based, collaborative, and mathematically and computationally intensive. This nine-day workshop will feature several distinct mini-workshops in a collaborative setting:

- Phylogenetics
- Bioinformatics
- Investigative Case Based Learning
- Multivariate Statistics
- Data Mining and Exploratory Data Analysis
- Image Analysis and GIS
- Realtime Data Acquisition and the Collaborative Classroom

We hope that these nine days of sharing, borrowing, transferring and translating ideas will encourage meaning-

continued on next page
ful cross-disciplinary connections. Certainly the mix of presenters and participants should provide ample opportunities for the first step towards fruitful and lasting collaborations.

A long time activist and author in the BioQUEST Curriculum Consortium, Bill Sterner, Director of the Instructional Laboratory of the Department of Computer Science of the University of Chicago, wrote in 1988 that the substantive changes in biological research will require a very different sort of undergraduate education (see Fig.1). Similarly, the NRC Bio 2010 and MAA Math & Bio 2010: Linking Undergraduate Disciplines reports have highlighted the need for both undergraduate biology and mathematics educators to parallel changes in the research environment by connecting biology and math much more deeply in their curricula. So as educators, how do we bring into the classroom the “Data Deluge” that is rapidly becoming part of most contemporary biology research?

In order to address these challenges, we will explore the use of graph theoretical, multivariate statistical, and quantitative geometric approaches in the context of undergraduate biology education. Workshop leaders have been recruited who have expertise in reforming undergraduate science and mathematics education.

**Workshops**

Investigative Case-Based Learning with Margaret Waterman, Southeast Missouri State University and Ethel Stanley, BioQUEST. For more about their LifeLines OnLine project see http://www.bioquest.org/icb

Evolutionary Bioinformatics with Sam Donovan, University of Pittsburgh, and Anton Weisstein, Truman State University. For more information about their BEDROCK Bioinformatics Education Dissemination: Reaching Out Connecting and Knitting Together see http://www.bioquest.org/BEDROCK/

Multivariate Statistics with Danny Kaplan, Macalester College. Danny has developed a course: “Linking Introductory Calculus and Statistics With Multivariate Modeling,” that he has presented at the AMS-MAA-MER Special Session on Mathematics and Education Reform, and with David Bressoud presented a poster on “Quantitative Methods for Public Policy” in an MAA Poster Session www.macalester.edu/qm4pp

Physioinformatics with Jeff Sale, San Diego Supercomputer Center. Currently, Jeff is Staff Scientist with the Education Center on Computational Science and Engineering, where he overseas work on the incorporation of high-performance computing including modeling, simulations, visualization and data-intensive computing into the undergraduate curriculum. He has extensive experience in education courseware development.

**Keynote Speakers**

Complexity, Fractals, and Data Analysis, Larry Liebovitch, Director, Center for Complex Systems and Brain Sciences, Florida Atlantic University

Phylogenetic Analysis of Complex Networks Due to Hybridization, Endosymbiosis, & Horizontal Gene Transfer, Peter Lockhart, Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, New Zealand
As a research assistant professor at the Center for Computational Science at Boston University, I have spent the past four years developing resources for faculty interested in introducing modeling and simulation into their classroom instruction and laboratories. BioQUEST has been a partner in these activities. As members of the former Education, Outreach and Training Partnership for Advanced Computational Infrastructure (EOT-PACI) and current Engaging People in Cyberinfrastructure (EPIC), both NSF funded initiatives, we have collaborated on numerous professional development projects. These projects have focused on supporting faculty and students to take advantage of the computational infrastructures available in the country. Through these efforts we have introduced to faculty Biology Workbench and Biology Student Workbench, developed exercises that utilize the numerous molecular databases available via the web and created simulation tools such as Gepasi, Stella and JDesigner. However the environment of professional development is different than the classroom. In the spring semester of 2005, I had the privilege of spending four months at Beloit College, co-teaching Cellular, Developmental Biology with John Jungck and working with the BioQUEST staff to host and provide professional development training at Emory University. BioQUEST and Beloit College created an exciting physical, philosophical and pedagogical context for translating materials development into classroom learning.

I planned to explore the implementation of resources that I developed for modeling the cell cycle. These materials are preliminary components to my book “Computational and Modeling Basics for Cell Biologists” expected to be released next year.

Why were BioQUEST and Beloit an exciting context? As we know, BioQUEST has organized within the biology education reform movement. The Consortium brings together faculty, students, and researchers from a variety of backgrounds to co-create new ways of teaching biology that empower both faculty and students as learners and scientists. As such, BioQUEST has provided me with a philosophical and pedagogical refuge in the area of biology. As a newbie to the classroom environment, BioQUEST seemed an ideal framework to use to further develop my teaching skills in the context of reform and Beloit a good location for interacting with a number of BioQUEST members during the semester.

The Challenges

The professional challenge that I saw and chose for my work with BioQUEST was to make full use of the modeling materials that I had developed. Would they translate to the classroom? Would the students learn something? Would I still remember enough biology that could be related to the materials that students would be learning? A different challenge that I saw as well, was: How would John Jungck and I work together? Would he accept my way of teaching? How would we co-teach?

Challenge 1: Use of materials in the classroom

Within the Cellular, Developmental Biology Course, the module I taught was one of four sections. The two hundred-level, research-based course was taken by twenty-two students, freshmen to seniors. Text-based materials consisted of excerpts from the Cell Cycle by Murray and Hunt and two chapters from my book in progress, “Cell Cycle” and “Introduction to Modeling.” I also provided students with a list of existing models online. The lab exercises and research projects made use of Stella, a non-domain specific, dynamics simulation tool. The materials we used are based on already published models. In particular, we used an oscillator model published by Goldbeter in 1991.

This model is a nice introduction because it has a small number of variables, associated rate equations and differential equations. The biology includes key concepts (feedback, active and inactive states of enzymes, and changes in concentrations) without excruciating detail. Students were provided with the data (components, rate constants, initial parameters and rate equations) to re-create the minimal model (a model with the fewest variables required to re-recreate the observed phenomena) of Goldbeter (1991). They were then asked to modify the model to include one of seven additional aspects of the cell cycle (size, kinetic...
models, regulation, etc). To modify the models, students reviewed published cell-cycle models and biological reviews. From these they developed their own models using Stella, and reported their projects as scientific articles.

There were seven self-selected groups of students with a group size of three or four. Group composition varied including a mix of freshmen, sophomores, juniors and seniors. The groups chose their topic from the list of seven possibilities. The seven features of the cell cycle were chosen with two minimum requirements 1) they were considered important aspects of cell cycle mechanics and regulation, 2) they have been examined in publicly available mathematical models. A list of resources was provided that included references to existing mathematical models. Students used resources on the handout and found additional resources (web sites, papers, presentations) via the web. The students were required to submit a proposed research project based on their topic. The proposal included a discussion of their hypothesis or reason for examining a particular factor, the primary literature they would use as references, and background information.

Some of the challenges that arose during the course and using the materials were the following: understanding rates of change and how to create rate equations in Stella, understanding how reaction rates relate to changes in mass, and dealing with the absence of lab data linked to models. Models are never correct, they are approximations. However, in the absence of a standard, experimental data or tables of previous results, students were unable to evaluate for themselves what worked and did not work in their models. The lab data provides a constraint for the possible results one should obtain from the model. It provides a concrete value to compare against simulation results. In the absence of these values, one must evaluate the overall patterns and behaviors of the variables and rates of change. The focus on patterns and behaviors includes a greater attention to the relationship between variables (e.g., Is active MPF high when cyclin concentrations are low? Is this consistent with the model and biological system?). This is not a detailed, reductionist evaluation that can be more familiar to the students. Evaluating patterns and behaviors requires the students to design their own standard for success of the model. Given that students were not told that they needed to design their own measure of success for the model and that students were still learning to create their own research projects, it is not surprising to me that more concrete standards would have helped the students better understand and investigate their models.

The entire course was an immersion in the activity of being a scientist. The students generated a hypothesis, proposed a method of investigation and pursued that investigation including the constant revision of the project due to success and failures of an approach. This can be an exciting, awkward, and frustrating process for students. The students became more comfortable during the course in working on projects that they defined, crafted and refined. Through the research process students begin to learn to search for being correct within the context of their research. This is different than being “right” as measured to an abstract standard in an answer book or from the instructor. Modeling accentuates the same iterative research process. One proposes a set of relationships, begins to design a model, finds that it does not work and needs revision. The modeling project was the second project in the course. It may have been better placed as one of the latter projects. As a latter project, students would have been more comfortable with the process of doing research. The students in this second project were still invested in getting the right answer. This often prevented them from being able to discuss what was happening in their model.

Challenge 2: Performing co-teaching
I have worked for 14 years in supplemental education arenas with cultural, community and group building methodologies. This work has led me and others to an nth P in BioQUEST language. (I believe that many participants in
BioQUEST recall the 3P’s Problem posing, Problem solving, and Peer persuasion. And, many have identified additional P’s to the list. The P for me has been performance, as in theatrical performance. Performance-based learning is a cultural approach, in which we understand learning to be a socially created activity. The social-cultural context (environment) of learning that is created by participants is inseparable from the learning that occurs. When seeing what we (students and faculty) do as performance, we have the opportunity to relate to each challenge or conversation as an opportunity for a new performance—an opportunity to do something out of our ordinary character or to accept someone else doing something out of character. This is an opportunity for growth and learning. As we learn, we do things that were not previously attributed to character, our identity or ourselves.

Performance is a way of understanding what we do as human beings. It is also a metaphor for creating the learning environment. Performance in classrooms, including science classrooms is essentially improvisational. Improvisation, theatrically, refers to building a group that listens to one another, accepts what each member contributes and uses what was said and done to add to the collective narrative and support of the group. In our classroom, our performance was of scientific conversations. This brings us to how would John and I worked together.

John and I were committed to creating an environment in which the students were related to as scientists and that built a community of learners in the context of the classroom. We did not know this at the beginning of our work together, but discovered it as we learned to teach together. My initial concerns were focused more on my own teaching. Would I have the room to explore my own teaching method in the classroom? What would John and the students do with what I gave in the class? Framing my work with the class in this way set up an awkward and untenable dynamic. My fellow participants (John and students) were placed in a position to be evaluators of my teaching rather than colleagues or collaborators in producing the learning environment. Given John and I come from very different backgrounds— we are from different generations, scientific trainings, genders, races and the list could go on— the potential for conflict was everywhere. However, John and I improvised a very successful collaboration that built on our strengths and commitment to relating to students as capable of doing science. As someone who looks at what people, including myself, do as performance, I chose to create a different performance for myself. I chose to stop looking at what I had to prove to the students or John about what I knew. Instead I worked with John to learn how he, through apparent chaos (both in practice and theory), supported students to perform as scientists. An important direction for my performance and the class came from a conversation with John about how the class was doing.

John: I’m concerned about the class.
Raquell: Why?
John: They’re not engaged.
Raquell: How do you know?
John: Because, we’re not having scientific conversations.

I took the direction and challenge from John to have scientific conversations in the classroom. As an improviser and performer, the challenge was to work with John and the students to use everything that was in the classroom to create scientific conversations.

In practice, my performance involved asking questions. John led many of the lectures in class. John provides a broad range of examples and concepts that students might relate to from their different backgrounds and interests. John would introduce concepts that I did not know or was not sure how they related to the biological topic we were studying. Rather than perform observer or critic, I performed as conversational collaborator. I would ask John questions. Questions that clarified what he was saying for me and would continue to support his performance as presenter. If I had a different opinion or understanding of the scientific topic, we explored it publicly. We had scientific conversations in the classroom. I do not have a quantitative evaluation of the affect of this on the students. I do know from student comments that they found our conver-

Figure 2: Raquell Holmes, the author, and John Jungck
A recent NSF report (Donoghue and Harris 2005) has identified computational visualization as one of the key techniques of scientific analysis for the twenty first century. Appropriately enough, two of the Excel-based modules developed for the Excel Simulations and Tools for Exploratory, Experiential Mathematics (ESTEEM) (http://www.bioquest.org/esteem/) project at Beloit College--BioGrapher and WDisplay--emphasize data visualization in unique and useful ways.

**BioGrapher** (Jungck and Viswanathan 2003) is an Excel front-end for the AT&T GraphViz graphical visualization package available in the public domain (http://www.graphviz.org/). While a number of excellent stand-alone applications that incorporate GraphViz for different platforms and operating systems are available, **BioGrapher** is unique in that, it keeps with the ESTEEM philosophy of using the standard and ubiquitous Excel environment as the front end and graphical user interface. It implements the visualization of graphs (nodes and edges) that are specified as standard adjacency matrices using the Excel spreadsheet paradigm of rows and columns, with connections between nodes represented by a non-zero (non-empty) cell for the appropriate row and column. A complete Visual Basic for Applications (VBA) Excel interface for the Windows versions (4.5) of GraphViz has been programmed and implemented as an additional menu bar menu (http://home.so-net.net.tw/oodtsen/wingraphviz/). This additional menu has enough functionality for the user to invoke many graphical drawing and manipulation routines, as shown in Figure 1. This figure shows a visual and graphical representation of part of the yeast genome protein expression data analyzed by Rives and Galitski (Rives and Galitski 2003), and clearly illustrates the existence of clusters for families of proteins in a fashion that is much easier to understand and interpret than the corresponding adjacency matrix or even a standard gray-scale connection matrix diagram.

The second module--**WDisplay**--is a server-side Excel spreadsheet with Macros (VBA) that can simultaneously, and in real-time, display data (both in numeric and graphical form) acquired from client computers communicating with each other over a wireless network. The module as-

---

**Figure 1:** A screen shot from **BioGrapher** showing a visual and graphical representation of part of the yeast genome protein expression data.
sumes use of a standard Wi-Fi network and and uses standard Windows remote file sharing protocols. Each of the client computers runs a client version of the spreadsheet (WClient) that provides an interface and options for acquisition of electrical transducer data (voltage, resistance, capacitance, etc.) as a function of time using a readily available serial-port equipped Radio Shack voltmeter (Viswanathan et al. 1996). Thus, the client version of the Excel spreadsheet can serve as a stand-alone single channel electronic “strip chart” recorder, while the server-side spreadsheet (implemented on a dedicated server computer that can communicate with up to ten client computers) can display ten channels of data in real time, superimposed on the same graph in Excel. The aggregated data and graphs can be projected using a LCD projector for real-time, interactive laboratory or classroom discussion. The figure below shows typical room temperature data (measured as the resistance of thermistors connected to voltmeters) acquired from four portable laptop computers set up in a student laboratory Wireless Room Area Network (WRAN) and transmitted to a fifth server laptop computer.

In summary, both modules provide user-friendly front ends for the visualization of different types of data that are relevant in the biological context, namely graphical visualization of connections and relationships, as well as real time monitoring and comparison of experimental parameters. Both modules would work well in a laboratory or classroom environment.

References


Figure 2: A screen shot of WDisplay displaying room temperature data acquired from four laptop computers and transmitted to a fifth.
help educators create effective online learning communities” (from the Moodle website). Two sets of lessons from the Discover Data portal will be highlighted at the upcoming BioQUEST Summer Workshop June 10-18, 2006.

The first set of lessons integrates a complex, spatiotemporal dataset from the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program. The CalCOFI program derives from a unique partnership of the California Department of Fish and Game, the NOAA Fisheries Service and the Scripps Institution of Oceanography. It began in 1949 to help explain the collapse of the sardine populations off California. Today it continues in an effort to study the marine environment off the coast of California and the management of its living resources.

While the data is readily available on the web, it is in a form that might be somewhat daunting for an educator to attempt to incorporate into their curriculum. In an easy to follow step-by-step manner, the series of lessons simplifies this dataset and walks the student through a sequence of data representations in a gradual progression from simple one-dimensional time series graphs to elegant three-dimensional volumetric visualizations. Through this process, students gain insight in three main areas:

- How plankton population variation may be correlated with environmental phenomena such as the El Nino Southern Oscillation,
- How to interactively explore a very complex dataset with a variety of visualization tools
- How to represent complex data in multiple ways

Students begin by using the popular spreadsheet program Microsoft Excel to generate one-dimensional time series from one of several data sampling stations. They then have an opportunity to interactively explore the entire dataset one station at a time with an interactive clickable map. Next, students explore the entire dataset at one time by creating a three-dimensional bar chart using Excel, where they begin to understand the benefits and disadvantages of this complex mode of representation (Figure 2).

Finally, students learn advanced two- and three-dimensional visualization techniques with a free powerful software program called EasyViz developed by the National Center for Supercomputing Applications (NCSA).

The second set of lessons to be presented seeks to help students better understand phylogenetics and the origin of new species. One of the most powerful counters to the argument that evolution does not convincingly explain the origin of new species is the observation of a rare but fascinating and still controversial phenomenon known as “ring species.”

Ring species occur when a single species becomes geographically distributed in a circular pattern over a large area. Neighboring populations of the species vary slightly but can still interbreed. However, at the opposite ends of the pattern that link to form a circle, natural variation has produced so much difference between the populations that they function as though they were two separate, non-interbreeding species.

Students learn how researchers use electrophoretic separations of selected enzymes and DNA sequence pattern matching to better understand why some researchers argue that the two subspecies *Ensatina klauberi* and *E. eschscholtzi* are in fact different species. This leads to the core question: should the species *E. eschscholtzi* be split into two or more species, or be considered a single species, and if so, where does one specify this distinction? Students first play the role of a researcher who has col-
lected a set of data from field observations of salamanders found in California, distributed throughout the coastal mountain ranges and Sierra Nevada surrounding the San Joaquin valley. Using an interactive map to plot their data, students can create a compelling graph of the considerable variation within a single species to infer the possible emergence of a new species.

Students then use the Student Biology Workbench to process actual genetic sequence data obtained from the National Center for Biotechnology Information to generate cladograms of the species in question. They also explore similar data from other species of salamander that are known to be part of a single species in order to compare the results with *E. eschscholtzii*. Through this process, students learn about the challenges researchers face in their attempts to demonstrate the complex process of evolution and the origin of new species.

In addition to these series of lessons, the Discover Data portal includes lessons based on complex datasets from many other scientific disciplines including astronomy, earth science, muscle physiology, biochemistry, geography, and more.

The portal is being developed by Jeff Sale, educational programmer at SDSC, in collaboration with a select group of teachers from SDSC’s TeacherTECH program. These teachers are exploring different approaches to using the lessons within the social constructivist paradigm of the Moodle LMS. The portal will be officially available to the public as of July 2006. However, teachers may currently submit a request for an account to assist with the evaluation of its content (http://education.sdsc.edu/edportal/moodle/).

Jeff will be presenting these lessons at the upcoming BioQUEST Workshop “Exploring Complex Datasets” in June 2006, along with other interactive exploratory tools for earth science education, including a real-time interactive 3D visualization tool for recent seismic activity in California currently being used in undergraduate geoscience courses.

An added feature of this module is the ability to run it on the GeoWall, an immersive three-dimensional stereoscopic visualization system that allows users to discern subtle details in a complex dataset that would be difficult to identify using just two-dimensional representations.


Figure 4: A 3-D visualization “shakemap” of seismic activity in California.
sation/presentations interesting and helpful. The students benefited from topological, organismal, and quantitative approach of John’s combined with the qualitative, mechanistic, cellular view of my own. I think significantly, the students joined in the performance by asking questions of us and each other. Every student spoke at some point, gave an opinion, asked a question. As a group they participated in the scientific dialogue.

In the language of performance, we--faculty and students--created an ensemble in which the students were as responsible for their learning as the faculty. We created an environment of scientific dialogue. We developed the skills to present our ideas; ask questions; revise our opinions and experiments; challenge our abilities and collaborate on the advancement of all of these. In the process of research, you investigate an unknown. It is impossible to know if you are right; there is no right answer to check yourself against. You can only do experiments. You define, describe results in the context of the experimental design and then investigate the potential ramifications of these findings in the context of current bodies of knowledge. It is commonly accepted that for students to develop these skills they need to be involved in the process of research. Performance provides a context in which the students are related to as scientists even as they are becoming scientists. The focus is on the investigation, the conversation and environment in which we develop the skills to push the boundaries of who we are and what we know.

Performance, including improvisation, is a methodology that enables us to create with the unknown: what students say and think; what reagents will work or not. We create without having to know how the class dynamics and logistics will turn out. As scientists, we use skills that we have developed over years to create understanding without knowing what the results will be. The challenge that we face as educators is how to not use what we have learned (knowledge) to stifle creativity and learning itself. Performance gives us as educators the opportunity to relate and respond in new ways to the classroom and our content. The students and I had never worked together to learn how to model the cell cycle. We had to create a way of doing that together. We had to create our learning ensemble, a group that could make offers to one another and together build a scientific scene from what was given. Similarly, John and I improvised with a focus on inclusion, community and science. In the process, we created an environment that was inclusive of us, our students and scientific research. It was a wonderful semester of professional growth for all involved.

From the description of DNA structure in 1953 to the recent discovery of “Hobbits” in Flores, the field of biology has undergone a revolution. At the same time, textbooks for “introductory” biology have rapidly grown from 200 pages to well over 1000 pages. As the amount of information has grown, biology education has evolved to include PBL, case studies, computer simulations, open-ended laboratory projects, and many other innovative methods. The importance of biology over the last half century is undeniable.

Traveling this last half century of change, the Association of College & University Biology Educators (ACUBE) has supported undergraduate and graduate biology educators who wish to keep their educational and research scholarship current, fresh and innovative. It has supported those who strive to translate the changes in biology, technology, and our understanding of learning into the classroom. Members of ACUBE share their ideas, concerns, and course innovations; present their work at the annual meeting; and publish their work in Bioscene, ACUBE’s peer reviewed journal. The ACUBE publication archive including the bioscene Journal of College Biology teach-
Looking back
http://www.micrographia.com/tutoria/michasic/michapt01/michb0100.htm
http://www.micrographia.com/index.htm
Great site for looking at history, tutorials, references

Looking forward
http://www.molec.com/products_options.html#microscope
To contribute to the newsletter or to subscribe to BioQUEST Notes, please contact Robin Greenler, Editor (greenlrr@beloit.edu). For general information about the activities of the BioQUEST Curriculum Consortium, contact us at:

Biology Department
Beloit College
700 College St.
Beloit, WI 53511
(608) 363-2012
bioquest@beloit.edu
http://BioQUEST.org

Major support for BioQUEST development and activity has been provided by the Howard Hughes Medical Institute, the Annenberg/CPB Project and The National Science Foundation, with additional support from Apple Computer Inc., The Foundation for Microbiology, and The Center for Biology Education at the University of Wisconsin-Madison. Any opinions, conclusions or recommendations in these materials are those of the author(s) and do not necessarily reflect the views of the funding sources.