This summer I worked on collecting and organizing geologic data. The data came from multiple map sources and boring logs. It is essential in two projects that concurrently develop cyber-infrastructure for geological data, and for regional data in San Diego area. The first project is GEON (The Geosciences Network), which creates tools for joint analysis and visualization of distributed heterogeneous geological data. The second project is RWBC (Regional Workbench Consortium), which attempts to create interactive and participatory methods for social learning about sustainable development in the San Diego-Tijuana city-region. The RWBC is promoting multidisciplinary research and service learning aimed at understanding how problems of environment and development interrelate across local, regional and global scales. The web pages of the two projects are http://www.geongrid.org/ and http://www.regionalworkbench.org/.

The GEON project focuses on national-scale and regional-scale datasets, while RWBC assembles local datasets. Geologists involved in the RWBC have been trying to find ways to connect the projects together by integrating local geologic data into larger-scale analysis conducted within GEON. The focus of my project has been on researching the integration of these datasets. Specifically, I assembled and cataloged local geologic datasets and created a query-able database of boring records. The databases and online query interfaces I used in the course of the project are compatible with the tools used in GEON, which should make integrating local geologic data into GEON seamless.

On the first step of the project, I focused on collecting and organizing geologic datasets utilizing the expertise of Jerome Jaminet. Jerome Jaminet is a Senior Project Scientist at TRC, a vice-president of the San Diego Association of Environmental Professionals, and a member of RWBC. He provided me with a collection of boring logs for various parts of the San Diego county. Then, with Jerome Jaminet and Ilya Zaslavsky, a schema was developed for a database containing the boring logs, which later would be served online. I implemented this database in Microsoft Access using 2 tables that are linked together by a company id. The first table is designed to be general information about the written boring log. This includes contact information (name, address, etc.) of the company, the name of the .pdf file of the scanned document, and the x,y,z coordinates of where the boring happened. The second table is specific information to the actual boring log. In this table is information about the soil, the depth levels, geologic stratum, depth ranges, and fine detailed information that was included in the boring log. The main challenge of developing such a database was designing a query system capable of querying depths (ranges, maximum and minimum values) for different strata across a collection of boring holes. Another challenge was adjusting local geologic designations to the naming conventions and terminology adopted in GEON.

The next step was developing a spatial database on geologic maps. These data were provided by Professor Eric Frost, of the SDSU Geology Department, who is also a member of the GEON project. With the help of Christopher Paolini, the system administrator of the College of Engineering at SDSU, I was granted a super-user account on the Attila server and given 50GB of space on the disk array. Then, with the help of John Kaiser of the SDSU Geography Department, I scanned 12 large-format geologic maps at different resolutions, covering San
Diego County. Since the scanning computer is isolated from the campus network, I copied these maps on zip disks and transferred them into a machine equipped with ESRI’s geographic information systems software, including ArcGIS and ArcIMS. The ArcGIS software was used to georeference the images to real-world coordinates, and ArcIMS was used to serve these maps online. This part of the project is still ongoing.

Since geologic data served by this project and geologic data in GEON are served from the same types of servers, it is possible to integrate local and regional/national geologic maps and explore irregularities caused by resolution mismatches. For example, along the edges where more detailed local and less detailed national datasets “meet” there are lots of inconsistencies caused by the different resolutions of the data. Also merging multiple projections of the same general mapped area together to make different maps fit together nicely results in a slight skewing, which would result in a slight precision error in calculating an exact location. There is a major discrepancy in reconciling the maps of the United States/Tijuana border created in Mexico with the maps created in the United States. These inconsistencies are usually more pronounced in mountainous areas with complex geology. However, deciding how to resolve such inconsistencies automatically in the course of querying across datasets of different resolutions was beyond the scope of my current effort.

This project allowed me to interact with a broad group of people, each having their own unique skills and helping me in different ways. I am very grateful to all the efforts that were donated to help me succeed with this project by all the individuals I talked to. Through Michael Bayliss, the administrative coordinator for electrical and computer engineering, I was able to obtain space in the graduate area of the engineering to work on this project. This proved to be incredibly helpful because I was able to use a wire in that room to connect directly into the Attilla server.

I also had a very diverse experience working with both practicing geologists and computer scientists. It was interesting seeing how the different people approached the different obstacles that arose. I, like the computer scientists, tended to approach most of the issues from a logic standpoint, however, the practicing geologists approached them thinking how the results would be best be used in their work. Talking and meeting with Jerome Jaminet showed how computer scientists could develop something that would be completely useless to the practicing geologists. In this project Jerome Jaminet really cared about trying to implement the water level, and a digital copy of all the boring logs with the creation of the boring log database.

Working on this project helped me understand better the intricacies and challenges of data integration techniques. It was difficult trying to combine different boring logs into a single working database because there were written in many different formats. It seemed like there was a general consensus for setting up a boring log, but there was no standard in the units of the measurements that were taken. For example, some of the logs used blows per feet and others used blows per six inches. Another problem was combining the resolutions of the various maps together into one entity. Prior to this project, I never questioned maps as being hard to make. I naturally assumed that making / combining maps would be relatively simple since the Earth does not drastically change over time. Now I can see that trying to make a perfect world map can be quite challenging.
By doing this internship I hope to have benefited both organizations (GEON and RWBC) as well as the practicing geologists with the data I have created. I hope in the future that this will provide the means of a way to create the 3-D model as I utilize the technology that is shared between GEON and RWBC.