Problem Statement: AvidaLab

1. Problem

Avida (Ofria & Wilke, 2004) is a software platform for digital evolution, used for studying evolutionary processes relevant to both natural and artificial systems. Avida is free, open-source software written in C++, and can be easily extended to add new features and system capabilities. Avida is the most widely used digital evolution software in the world and its users encompass a wide range of backgrounds and expertise, including biology, computer science, engineering, and philosophy. Avida places a population of self-replicating computer programs, called *digital organisms* or *Avidians*, in a user-defined computational environment. Avida is not a simulation of natural evolution, but rather a separate instance of evolution in its own right: the digital organisms compete, replicate, and mutate, thus satisfying the fundamental requirements of evolution (Dennett, 2002).

Users may configure nearly all aspects of Avida to accommodate the needs of an experiment. System configurations include the size of the world, the length of the experiment, the presence and behavior of resources, mutation rates, organism age limit, and many other factors. In addition, users can record many different statistics during the experimental run, recorded in plain text files. This information is then analyzed after the experiment has ended. The user can also add new statistics to go with added system functions. Experiments using Avida are normally conducted using multiple runs of the same experimental setup (typically 50-100 runs for each experimental setup), with each run initialized with a different seed value. In this way, the individual runs will produce different outcomes, since Avida's evolution engine uses probabilistic methods to determine elements such as mutations and placement of offspring in the environment. Depending on the complexity of the experiment and the length of the run, Avida experiments can last from several hours to a number of days.

Because of the extensibility of the platform and the flexibility of system configurations, Avida experiments produce vast amounts of data that need to be processed. Avida contains some limited analysis tools, but individual researchers have to write most of their own analysis tools outside of Avida itself. The computer scientists and engineers who use Avida do not generally find this situation particularly burdensome, except that writing such custom tools is time-consuming. The situation is a larger problem for researchers who are not trained or experienced programmers, such as the biologists who use Avida. Learning the tools and programming skills required to write software for basic analysis is time consuming, and presents a significant barrier for non-programmer users of Avida. Exacerbating this issue is the fact that most of the software tools used to produce analysis programs are heavily command-line based. Most non-programmer users are unaccustomed to working with command line interfaces, and the added complexity of learning a new interface scheme along with programming skills makes the task even more daunting.

Another difficulty in Avida experiment analysis is that, in the past, most existing Avida data analysis software was written using Matlab, an expensive propriety software product. More recently, Avida researchers have turned to Python and other freely available software to write analysis tools. Although this solution eliminates the need for high cost tools such as Matlab, it still fails to address the lack of tools available to non-programmers, or the potentially

complex tool chains and configurations that may be required in order to use non-propriety software.

2. Objectives

The objectives of the AvidaLab project are to provide:

- Tools for analysis of Avida data, including importing data sets, calculating common statistics.
- Data visualization tools for analyzed data, such as line plots and box and whisker plots.
- A user-friendly environment that is accessible for users who are not programmers.
- An integrated software toolkit that is easy to install, configure, and manage.

3. Functional Requirements

- Users must be able to import archived/zipped files, with file sizes ranging from 50 MB to 250 MB (typically in tgz format).
- Archive files must be prepared for use.
- Users must be able to select files for processing, from a single file to a whole set of files in a treatment (up to 100 files).
- Users will be able to select columns of data from files to process.
- Users will be able to select preprocessing features, such as computing log values, before statistics or visualization.
- The user will be able to compute statistics and statistical tests on the aggregated data, including:
  - Mean
  - Median
  - Standard deviation
  - Mann-Whitney U test
  - Kruskal-Wallis one-way analysis of variance
- The user will be able create visualizations of data, including:
  - Line plots
  - Histograms
  - Box and whisker plots
  - Scatterplots
- The user will be able to export plots in a variety of file formats, including .eps, .pdf, .jpg, and .png.

4. Nonfunctional Requirements

a. Usability.
   AvidaLab will provide an intuitive user interface. A user manual will be provided with easy access through the system, and troubleshooting information will be available through the system.
b. Reliability.
   The system will provide mechanisms for the user to correct input errors without failure of the system. The system will provide validation for inputs from the user, and will provide

user-friendly error messages in the event of user input error. AvidaLab will produce correct output results.

c. Supportability.
   Support for AvidaLab will be provided through the following mechanisms:
   - System documentation.
   - Online help-ticket system.
   - E-mail.

d. Implementation.
   - AvidaLab will be implemented with non-proprietary software tools.
   - Project documents will be produced in PDF format.

e. Interface.
   - AvidaLab will use a graphical user interface.
   - The system will run on standard computers (desktops/workstations, laptops).

f. Packaging.
   - AvidaLab will be available for download online.
   - The system will be installed on the user's computer and will not require use of remote resources.
   - Installation and configuration of the system will be simple and require a minimum of software needed outside of the system itself.

g. Legal.
   - AvidaLab will be free software.
   - The GNU General Public License (GPL) will govern AvidaLab, unless superseded by the license of one or more of the tools used to build the system.

5. **Target Environment**

AvidaLab must be cross-platform compatible, so that it runs on a variety of operating systems, including Windows, Mac OS X, and Linux.

6. **References**

Dennett, D. (2002). The new replicators. In M. Pagel (Ed.), *Encyclopedia of Evolution*. New York, NY: Oxford University Press, E83-E92.

Ofria, C., and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life , 10*, 191-229.