

OIT Grant Preparation Support

INTRODUCTION	2
YOUR BUDGET	2
STAFFING.....	2
DATA ACQUISITION	2
PRIMARY DATA.....	2
<i>How Much Data to Collect</i>	3
<i>Manual Data Entry</i>	3
<i>High Speed Scanners</i>	3
<i>Web Surveys or Forms.....</i>	3
<i>Instrumentation & Data Logging.....</i>	4
SECONDARY DATA.....	4
DATA MANAGEMENT SERVICES.....	4
RELATIONAL STRUCTURES VS. DATA WAREHOUSE	5
UT MICROARRAY DATABASE	5
ARCHIVES.....	5
DATA ANALYSIS	5
STATISTICS	5
TEXT ANALYSIS	6
DATA MINING	6
IMAGE ANALYSIS	6
MATHEMATICS	6
THEMATIC MAPPING	6
COMMUNICATION AND COLLABORATION.....	6
SOFTWARE	7
NETWORK.....	7
HARDWARE.....	7
ANALYSIS.UTK.EDU.....	7
UNIX	8
LINUX CLUSTER FOR HIGH PERFORMANCE COMPUTING	8

Introduction

Applying for a grant is a challenging task. The staff of the Office of Information Technology (OIT) can help in a variety of ways. Our technology and services can help you plan the best strategies for data acquisition, management, analysis, visualization and presentation.

We can also refer you to other support units on campus that provide related services. Whether you are looking for the answer to a quick question or looking to build a more collaborative relationship, OIT can help.

This document provides a brief overview of these services. To get started or for more information, call the OIT HelpDesk at 974-9900 and ask for Grant Planning Support or complete the OIT HelpDesk Contact Form at <http://remedy.utk.edu/contact/>.

Your Budget

One of the biggest challenges in preparing a grant is, of course, your budget. In some of the sections below we provide some cost estimates. However, even if you have used OIT services before, it is best to meet with one of our consultants to discuss the cost of our services. You may have used our services on a smaller scale for free and assume they will still be free on a larger scale. That may not be the case. We may also be planning changes that will be in effect by the time your grant is awarded.

It is also important to realize that software that you use for free for your own research may not be free when working for an outside agency (see the section on Software below).

We recommend at least a brief meeting with one of our consultants as early in your planning as possible. Minor changes made during the early planning stage can have a big impact on your budget requirements.

Staffing

If your grant needs dedicated staff for any of the services that we offer, we may be able to help you write job descriptions and advertisements. We may also be able to help you select interview questions and even assist in the hiring process. Once staffed, we may be able to help you manage the work of your research staff.

Office space for research associates or post-docs is always a challenge. That is an area in which we are most unlikely to be of assistance.

Data Acquisition

Primary data

Our consultants can help you decide how much data to collect and choose the method that offers the best balance of speed, accuracy, ease-of-use and cost.

How Much Data to Collect

If your project involves collecting your own data, the grant providing agency will want to ensure that you will collect only enough data to verify your hypotheses. Determining this optimal sample size is called *power analysis*. This typically takes an hour or two to run a power analysis and the cost is usually centrally funded. The most important step concerns the design of the study itself. For example, a study looking into how a group taking a blood pressure drug compares to a control group may take five to ten times as many subjects as one that compares the responses of the same group on and off the drug. We can help discuss variations like this, choose a powerful design and then determine the optimal sample size.

Manual Data Entry

Manual data entry is the slowest and most error prone way to record data, but there are times when it is the only option. OIT does not offer manual data entry services. However if you have such clerical assistance we can help make the most of it. We have software and services to help you create data entry forms that set minimum and maximum values validation for continuous measures (e.g. no age>120), lists of valid values for categorical measures (e.g. no gender=femail) and complex cross-validation rules (e.g. no pregnant males).

While these methods greatly increase the accuracy of manual data entry, they cannot stop all errors. The method that can is the dual-entry approach. With this method, you actually enter the data twice and then generate reports that cross validate the two datasets to find the errors.

High Speed Scanners

One of the best ways to avoid costly and error prone manual entry is to use high speed scanners. Our scanners can read the data from both sides of a form in about a second, with a very low error rate. You can choose from a wide range of off-the-shelf forms for economy or create custom forms that put survey items on the scan form itself giving it a professional appearance. You can even create multi-page booklets and separate the pages later for scanning.

We have both optical mark recognition scanners and image scanners to read a wide range of form types. However, it is critically important to design and test the scanning form in advance. Our scanners work with specific types of forms.

We can also refer you to outside data scanning companies that can prepare custom forms and handle extremely large volumes of data. You can read more about scanning at its costs at <http://oit.utk.edu/research/scanning>.

Web Surveys or Forms

If collecting data with a web survey or form is possible, we have a server in production to perform that task. Our web survey server lets you design your survey in a web browser, activate it, run frequencies as it continues to collect data and then download your data, all without our intervention. This data is backed up on tape each night and stored in a

separate building. The server and its associated routers also have long-term backup power in the case of an outage.

We can help you export the data in any form you need. The survey or form items become variable labels and the response options become value labels in an SPSS, SAS or Excel file. You can save both numeric data and qualitative text responses up to about eight pages each (32K).

Once we provide you with an account and about 30 minutes of training, you're ready to use it. It can send out email invitations and then send reminders to only those people who did not respond.

For budgeting purposes, consider a survey that is given one time with 40 items and containing no more than five logical skips (e.g. if someone says they don't do something, skip further questions about that topic). Once you have finalized the wording, a time consuming process in itself, figure on 10 hours of time to put this on the web. If someone on your research team does the work, our consultants can assist usually at no charge. If you want us to do it for you, we can do so only if you schedule this work with us at least six weeks in advance. The rate is \$70/hour. More complex surveys require detailed information to estimate costs.

Instrumentation & Data Logging

One of the cheapest, fastest and most accurate methods of data acquisition is to have a computer monitor and optionally control your experiment and record the data as it is created. Our data acquisition consultants can help you choose equipment that will connect your computer to your laboratory instruments and then program the computer to interact with it and collect data. We can even show you how to monitor and control it remotely using a web browser.

Secondary data

If you can answer your research questions without having to collect your own data, you will save a great deal of time and effort. There are many data repositories that store data that researchers have already collected, along with programs to read and manage it. We work closely with UT Library's Data Librarian to help you see if data already exists to meet your needs. We can then show you how to download it, read it using the included programs (or write one if none are included) and then translate it into any format you need.

Data Management Services

You can expect to spend between 70% and 90% percent of your data analysis time on cleaning and managing your data. The more carefully you plan your data structures, the less time you will have to spend on this effort.

Relational Database vs. Data Set or Data Warehouse

If your project requires many simultaneous people doing data entry at the same time, you may need to use a relational database. This is not a service that we routinely offer so we will need advanced notice to plan one.

A relational data structure may also help with speed of entry, ensuring accuracy and saving disk space. However, you cannot apply most analytic methods directly to data in relational form. We can help you plan the steps to “denormalize” or transform it into the data warehouse or dataset form. We can help you choose those data structures and even begin writing programs that you will need to convert from one form to another before your data is even collected.

UT Microarray Database

If your research uses microarray technology we can help you plan how to store the data in our local installation of the Stanford Microarray Database application. The database provides you with a standard approach to storage plus a widely accepted way to share the data with other researchers if you wish. For more information, see

<http://genome.ws.utk.edu>.

Archives

Our staff can help you evaluate and implement the latest in storage technologies. Recent developments in storage can enable you to create archives of your work that are accessible locally or across the Internet.

Data Analysis

Your grant proposal will of course need to include a description of your data analysis plan. We can help you develop one using the various tools below and even help you write that part of the grant.

Statistics

We can help you chose from a wide range of statistical methods that granting agencies find acceptable. These include a wide range of measures of association, group comparison methods, analysis of variance and covariance as well as more advanced multivariate approaches.

For budgeting purposes we offer the following scenario. If you had one set of data already entered, with up to 100 variables and 100,000 observations, that you collected at one point in time, stored in a standard format (columns are variables, rows are observations) we could usually run basic descriptive statistics plus up to five analysis in 10 hours. If we are acting as your team’s consultants that would be centrally funded (i.e. free). If instead you wanted us to do the work for you, it would cost \$70/hour and we would need to plan it to occur during our less busy parts of the semester. Anything more complicated we would need to discuss before providing cost estimates.

Text Analysis

We can help you plan the analysis of text using several different approaches. If you use the more manual qualitative approach, we can recommend software to speed your work.

If your problem is amenable to a more computerized content analysis approach, we can also help you choose off-the-shelf categorization dictionaries prepared by other researchers which measure scales like anxiety, communication complexity and even truthfulness. If you need to build a custom categorization dictionary, we can help you plan computational, linguistic and statistical methods to speed your work.

We also support the automated Latent Semantic Analysis approach. We have site licenses for software that uses a linguistic approach, although we do not fully support it at this time.

Data Mining

If your project has massive amounts of data that you need to automatically scan for useful relationships, we can help you plan the stages. There are two main data mining approaches, the SEMMA approach preferred by SAS Institute and the industry standard CRISP-DM approach. We can help you choose between these approaches and develop a plan to apply artificial intelligence, machine learning or statistical methods to your problem.

Image Analysis

If your grant involves the use of images from satellites, microscopes, cameras or the like we can help you plan the steps needed to extract critical measurements from them. Measurements that we can help you extract include line lengths, angles, x-y point coordinates, area, percent of total or item counts. We can help you import images from almost any source such as digital cameras, flatbed or medical scanners or microscopes.

Mathematics

If your grant requires the use of mathematics such as the calculus or differential equations, we can help you solve those problems.

Thematic Mapping

If your grant includes the collection of data with a geographic component, we can help you plan how to display it. The software we support can do almost any type of thematic map including those with bars, pies and shading. We can also help you determine when to use our services and when to use those of UT's Cartographic Lab. At this time we are not able to support full Geographic Information Systems such as ArcGIS.

Communication and Collaboration

Our staff can help you evaluate and implement the latest in communication and collaboration technologies. Tools such as blogs, wikis and various social software packages may help your team work more efficiently together.

Software

Your grant will likely require you to use software to solve your research problem. UT has licenses for a wide range of software. If your grant work will result in only an academic publication, then you can use these licenses. However if your grant requires you to provide a report directly to a government agency or an outside company (profit or non-profit), you will need to budget for the cost of a commercial software license in your grant. Commercial licenses can be quite expensive. You may also have to run that software on computers owned by the grant providing agency.

OIT provides a wide range of research software. Some of it is centrally funded, i.e. free for you to use. Also some of it is fully supported by OIT staff. For details on what software is available for use on your computer, see <http://oit.utk.edu/software>. For details about which software we support see <http://oit.utk.edu/research/software>.

Network

OIT offers Gigabit Ethernet connectivity on the Knoxville campus to over 100 buildings, and DSL, T-1, or 100 Megabit Ethernet to over 20 more buildings. Most of these buildings are also serviced by Nomad, our wireless network. We offer high-speed connectivity to other universities and research organizations on Internet 2 at around 400 Mbps. If your grant needs access to higher bandwidth connections, please see the additional information at <http://www.ns.utk.edu/researchers.html>.

Hardware

Computers are available in OIT's public computer labs. For details, see <http://oit.utk.edu/labs/>. You can also use a free UNIX account to do your analyses. See <http://oit.utk.edu/accounts>.

Analysis.utk.edu

Our Analysis.utk.edu Terminal Server system lets you run research software as if it were installed on your own computer. The software actually runs on a server and displays an image of it on your computer. It is very easy to use since you can open and save files on your own computer's disk drive and print to your own printer. You can use it from any Windows, Macintosh, Linux, or Solaris computer. All the software on Analysis.utk.edu is for university work only. For outside consulting, both for-profit and non-profit, contact the vendor for information regarding commercial licensing.

The server itself has four 3.2 GHz Intel processors and is running Windows Server 2003. You can use 100 megabytes of disk space for free. We can help you determine how much disk space you will need and if it exceeds that limit, what it will cost.

The software currently available on Analysis.utk.edu and instruction on how to use the server is listed in <http://oit.utk.edu/research/HowToUseAnalysis.utk.edu.pdf>. For descriptions of each package, the locations of any tutorials or online manuals as well as how we can help you use each package, see our web page: <http://oit.utk.edu/research/software>.

In addition to limits on each piece of software, the total number of connections that the server currently allows is 20. We have not yet had that many log in simultaneously. We also monitor the use of each package and people rarely have to wait for a copy to become available.

UNIX

Our Sun Solaris system is called unix.cas.utk.edu. It consists of two systems. One system has four 750 MHz 64-bit Sparc processors and another has four 900 Mhz 64-bit Sparc processors. Each has 8 gigabytes of random access memory. You can use 50 megabytes of disk space for free. Each additional 100 megabytes of disk space costs a one-time fee of \$25.

The software includes C, C++, CSW, FORTRAN, Frame, GAMS, GCC, GCG, Imagine, IMSL, Java, Lindo, Maple, MATLAB (with many toolboxes), NetBeans, Pascal, PVWave, S-PLUS, SAS, StarOffice, SPSS, SunStudio, TeX.

The software packages in our *full support* category are IMSL, Maple, MATLAB, SAS, SPSS and TeX. The rest are in our *minimal support* category. We are considering removing IMSL and PVWave due to low usage.

Linux Cluster for High Performance Computing

Our Linux cluster is named Newton.usg.utk.edu. Its operating system is Red Hat Enterprise Linux 4. The cluster consists of 82 EMT 64-bit, 3.2 GHz Intel Xeon CPUs in a 32 node configuration. They are connected with two 24-port InfiniBand switches. Each processor has 2 gigabytes of random access memory. The high performance computing software is Platform Computing's Platform Rocks. It includes C, C++ and FORTRAN compilers. Software is run in batch execution mode using the LSF facility.