Progress in the Statistical Analysis of the Quality of the FAEIS Data

January 21, 2011

Food & Agricultural Education Information System

http://faeis.usda.gov
mailto:faeis@vt.edu

540-231-4941

- **Mary Marchant**, Ph.D., FAEIS Principal Investigator
  Agriculture and Applied Economics Department, Virginia Tech

- **Timothy Mack,** Ph.D., FAEIS Co-Principal Investigator
  School of Graduate Studies and Research, Indiana University of Pennsylvania

- **Eric Smith**, Ph.D., FAEIS Co-Principal Investigator
  Statistics Department, Virginia Tech

- **Bill Richardson**, FAEIS Project Manager
  Agriculture, Human and Natural Resources Information Technology (AHNR-IT),
  Virginia Tech

- **Eric Vance**, Ph.D., FAEIS Statistical Project Manager and LISA Director
  LISA (Laboratory for Interdisciplinary Statistical Analysis) and Statistics Department,
  Virginia Tech

- **Jolene Hamm,** Ph.D., FAEIS Help Desk Manager
  Agricultural and Extension Education, Virginia Tech

- **Albert Shen**, Ph.D., FAEIS Statistical Graduate Research Assistant
  Statistics Department, Virginia Tech

- **Ashley Bell**, FAEIS Graduate Research Assistant
  Human Nutrition, Foods and Exercise Department, Virginia Tech

- **Lisa Hightower**, FAEIS Graduate Research Assistant
  Agricultural and Extension Education Department, Virginia Tech

Introduction

This is the first in a series of quarterly reports from FAEIS to NIFA, in response to item #9 in the FAEIS RFA which states: "Produce quarterly reports on the progress in addressing transcription errors, outliers and missing values. Include statistical procedures used to correct and process FAEIS data."

Summary

To enhance the usefulness of the FAEIS data that have been collected, we have identified several key areas in which statistical methodology can be applied to improve the quality of the data. In the past quarter we have produced a SAS dataset for the FAEIS data and have verified that the reports generated from this new dataset are identical to those created from Report Builder on the FAEIS website; we have developed SAS algorithms to identify outliers and missing data; and we are currently developing SAS algorithms to identify redundant and miscoded data.

Prior to implementing the statistical quality attributes, we conducted a data quality assurance protocol where each institution's data was examined for the 2003-2009 reporting years and verified for accuracy. Of the 220 institutions who have historically reported to FAEIS since 2003, 170 of those institutions have completed the data quality assurance protocol implemented at FAEIS. At these institutions data was verified by CIP code, enrollment, degrees awarded, gender, ethnicity, disciplines, and degree level. The data quality assurance process at FAEIS has been a tremendous undertaking for the HelpDesk team and it has involved approximately 2000 man-hours and the efforts of four graduate research assistants. Plans for the future include automation of data error detection, comparison with the IPEDS data, and convening a meeting of a statistics expert panel.

1. FAEIS has added significant statistical expertise
(Refers to RFA item 2 in Appendix B, pp.13-14)

In an effort to improve the quality of the FAEIS data and prepare the data for rigorous statistical analysis, the FAEIS team asked Laboratory for Interdisciplinary Statistical Analysis (LISA, http://www.lisa.stat.vt.edu/) at Statistics Department of Virginia Tech (VT) to collaborate and provide leadership on statistical issues from an independent, outside perspective. LISA has a long history of helping VT researchers benefit from the use of statistics. LISA collaborators meet weekly to discuss projects such as FAEIS and to learn from each other.

From the collaboration with LISA, the FAEIS team added two faculty members and a graduate research assistant with statistical expertise:
- Dr. Eric Smith, head of the Statistics Department
- Dr. Eric Vance, director of LISA
- Dr. Albert Shen, statistics graduate student

2. SAS datasets for the FAEIS data have been established and tested

(Refers to RFA items 2, 3, and 4 in Appendix B, pp.13-14)

A significant change in the statistical analyses of the FAEIS data is the creation of SAS datasets. Creating a FAEIS dataset in SAS provides many benefits:

• Flexibility to analyze the data using SAS (or other statistical programs R, JMP, SPSS, ...)
• Visualizing the data
• Detecting outliers, missing data, redundant data, and miscoded data
• Analyzing trends in the data, such as graduation rates
• Consistency of analyses
• Portability of reports

Since October, SAS datasets have been created from the FAEIS Oracle database. In order to verify that the SAS datasets are identical to the FAEIS Oracle data, at least one hundred SAS reports were created and found to be identical to the reports from the Report Builder on the FAEIS website. These comparisons provide us with the confidence that the SAS datasets were successfully created.

3. SAS algorithms have been developed to identify outliers and missing data

*(Refers to RFA items 2 and 5 in Appendix B, pp.13-14)*

As in all voluntary survey data, errors are possible in the FAEIS data. Among the most frequently occurring errors are outliers, missing data, redundant data, and miscoded data. The FAIES team has been trying to identify the erroneous data by manually evaluating the reports (e.g. Enrollments in 2004-2009 by CIP code by Institution) based on the following criteria:

• The change from the previous year's enrollment exceeds ±10%
• Abrupt starts or ends in the enrollment
• Only single year enrollment was entered
• "Holes" or "gaps" (missing data) for certain CIP codes in certain year(s)
• Data were entered into different CIP codes in different years

Since human mistakes might happen while trying to manually identify the erroneous data with the fairly complicated criteria described above, developing SAS algorithms and procedures to detect the erroneous data will be helpful to ensure the data quality. SAS not only provides a reliable method better than the traditional "eyeball" methods, it also has the advantage of providing tools to visualize the data. Two plots that are commonly used to visualize data are the Boxplot and the Strip Plot. Two representative plots are shown below in Figures 1 and 2 for Bachelor Enrollment in Family and Consumer Sciences/Human Sciences – Family and Consumer Economics and Related Studies.

In the Boxplot (Figure 1; top), we should pay attention to two features of the display. The first feature is the underlined outliers, which are labeled as red circles with the values (enrollment) to the right. The second feature is the tall boxes, which indicate large variation of the data between 2004 and 2009. To take a closer look at the variations in the data, the Strip Plot is very helpful.

In the Strip Plot (Figure 1; bottom), the enrollment for each year is plotted as a dot. Clusters of dots will reflect small variation and a small box in the Boxplot. On the other hand, if the dots are widely scattered, there is large variation and a large box is formed in the Boxplot. Another issue to explore from the Strip Plot is the "trend" of enrollment when the variation is large. If there is an obvious trend (increase or decrease) of enrollment with year, the data is more reliable. On the other hand, if there is no obvious trend for the large variation, the data may be questionable and would be flagged for further investigation.
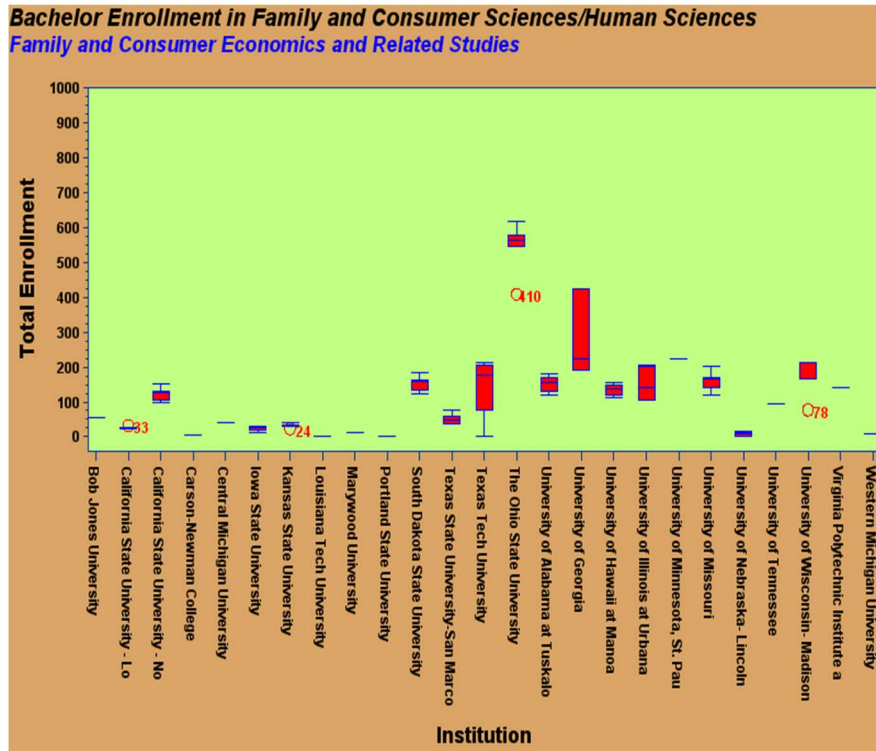
Boxplot of enrollment by institution
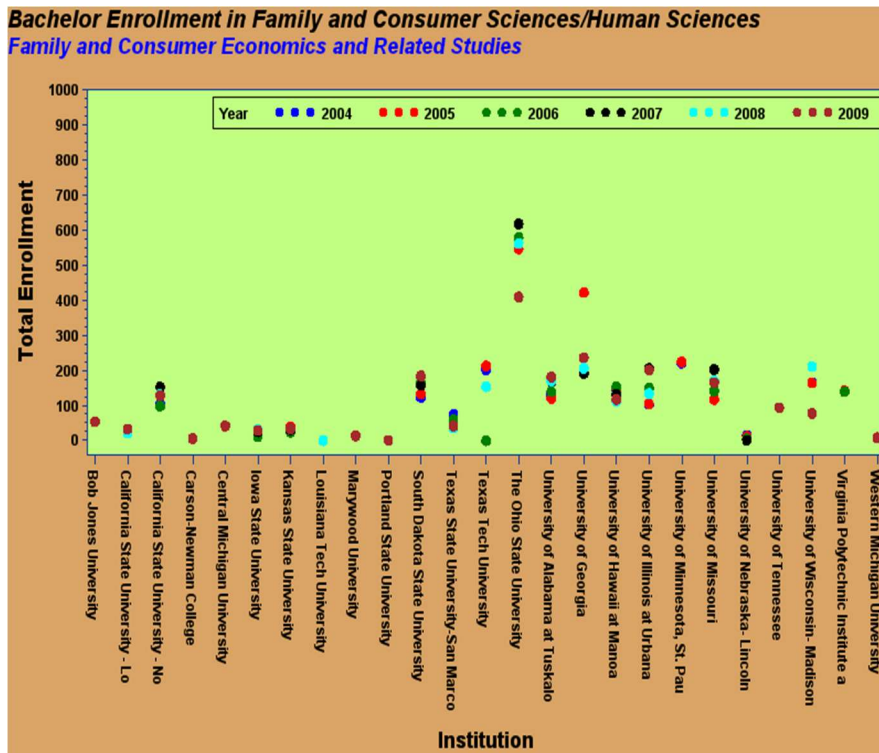
Strip Plot of enrollment by institution

Figure 1. Plots for Identifying Data Quality

In statistical analysis, zero values and missing data have different meanings. SAS can easily identify both zeros and the missing data in the table. Using the above enrollment data in Family and Consumer Economics and Related Studies, the outputs are listed below (Table 1 and Table 2). In this case, they are not large tables and could be easily used to manually identify the zeros, the abrupt starts/ends, single year enrollments, and holes/gaps. However, when the institution list becomes large, the manual identification is inclined to miss errors.

In order to avoid errors associated with manual identification, procedures are being developed to label errors of different types in different colors in the table. For example, cells with zero entries could be colored blue, cells that are holes/gaps could be labeled red, cells that are abrupt starts/ends could be labeled green, etc. Another possible layout is to separate the types of errors into different tables with color labeled cells. For example, we could create one table for holes/gaps, another table for abrupt starts/ends, and third table for single year enrollments. In this way we will be able to minimize the errors caused by looking at a large table. We will test both methods to see which is more effective for manually identifying errors.

Table 1.   Zeros in the Data Entry
Family and Consumer Economics and Related Studies

Institution	2004	2005	2006	2007	2008	2009
Louisiana Tech University	.	.	.	.	0	.
Texas Tech University	202	213	0	.	154	.

Table 2.   Missing Data Entry
Family and Consumer Economics and Related Studies

Institution	2004	2005	2006	2007	2008	2009
Bob Jones University	.	.	.	.	.	54
California State University - Long Beach	26	▌	25	27	22	33
Carson-Newman College	.	.	.	.	.	6
Central Michigan University	.	.	.	.	.	42
Iowa State University	18	▌	12	25	31	29
Marywood University	.	.	.	.	14	14
Portland State University	.	.	.	.	.	1
South Dakota State University	124	133	164	158	▌	185
Texas Tech University	202	213	▌	▌	154	.
The Ohio State University	.	548	580	619	563	410
University of Minnesota, St. Paul	222	225	▌	▌	▌	▌
University of Missouri	.	118	142	203	170	166
University of Nebraska- Lincoln	15	12	5	1	▌	▌
University of Tennessee	.	.	.	.	.	94
University of Wisconsin- Madison	166	165	▌	212	212	78
Virginia Polytechnic Institute and State University	.	142	140	▌	▌	▌
Western Michigan University	.	.	.	.	.	9

4. Data Quality Assurance

(Refers to RFA items 2 and 4 in Appendix B, pp.13-14)

Beginning in the Summer of 2010, the graduate research assistants developed and initiated a detailed data quality assurance protocol, where the GRAs examined data from 2004-2009. GRAs conducted a cross-validation between data found in the nearly 300 institutions in the FAEIS database with each institution's institutional research office data. Once the cross-validation was complete, the HelpDesk team contacted the reporting user to discuss discrepancies in reporting. For each institution, the annual data for the years 2004-2009 is populated and an initial total enrollment comparison between reported data and the institution's institutional research office is conducted by CIP code. Differences in reporting are noted and sent to the data entry user. The data entry user then provides further clarification and correction of these differences and provides data by gender and ethnicity for each CIP code. An example of this would be, the University of Missouri where there are two colleges that report, the College of Agricultural, Food and Natural Resources and the College of Human Environmental Sciences, totaling 26 departments and two schools with 50 majors between the two comprising of hundreds of data points on enrollment and degrees awarded. This data comparison took considerable time and cooperation on behalf of both the institution and FAEIS. Many institutions chose to report data to a greater specificity i.e reporting concentrations of majors and in those cases users were provided an excel file that has been examined for discrepancies such as gaps in data, duplicate data, and changes in CIP code usage. The purpose of this data file was to enable users to verify and correct questionable data.

The data quality assurance process at FAEIS has been a tremendous undertaking for the HelpDesk team and it has involved approximately 2000 man-hours and the efforts of four graduate research assistants. In addition, those relationships between the FAEIS HelpDesk team and the institution's data user have been paramount for enabling reporting. The data quality assurance process requires the HelpDesk team and the institution to have numerous contacts via email, phone, and in some cases in person. It is through these relationships, that have been built over the years, that data quality assurance has had a high response rate.

Currently, all institutions reporting to FAEIS have had cross-validation conducted by the HelpDesk Team. Currently, there are 130 institutions that are complete on data quality assurance. An institution that is considered complete is an institution that has undergone the cross-validation process and has provided verification or changes to data and this information has been corrected in FAEIS. There are approximately 75 institutions that need to respond to FAEIS.

5. Future Work

(Refers to RFA items 1, 2B, and 6 in Appendix B, pp.13-14)

In the subsequent quarters, the following will be enacted:

• SAS algorithms to identify redundant/repeated data entries and miscoded CIP codes –

Redundant or repeated data entries have been found in the FAEIS data, as well as miscoded CIP codes. These types of data errors are not easily identified manually. Redundant data often occurred when the same information was entered multiple times using different FAEIS accounts. Often these data appear to be outliers when compared to other years. A SAS algorithm is being developed to identify the redundant data by searching for multiple accounts and for outliers. Misplaced CIP codes often have the feature of missing data for a certain CIP code in certain years when the data is placed in another similar CIP code. A SAS algorithm is being developed to identify the misplaced CIP codes by matching the CIP codes with missing data.

• Automated identification of invalid/problematic data –

Once the identification of erroneous data with SAS is fully developed and tested, it will be automated with Microsoft Task Manager on a daily or semi-weekly basis. The detected erroneous data will be sent by e-mail to the graduate assistants at the FAEIS help desk for further investigation and follow-up with the corresponding institution representatives. This **will** be a routine data quality assurance procedure.

• Comparisons of FAEIS and IPEDS–

The Integrated Postsecondary Education Data System (IPEDS) from the United States Department of Education has similarities with FAEIS data but with much broader study areas. Systematic testing to compare and contrast the two systems will be done, along with direct comparisons of data collected from groups of institutions. It should be noted that while IPEDS and FAEIS collect similar data for degrees awarded, IPEDS does not collect useful enrollment information.   However, IPEDS will probably be the best source for addressing the "no-universe database problem" (RFA item #6).

Historically, one of the reasons for FAEIS is due to the inadequacy of alternative sources, including IPEDS. In this regard, please see Appendix A, correspondence to NIFA from Dr. Don Boggs, chair of the APLU / Board of Agriculture Assembly's Academic Programs Section (APS). To quote: "FAEIS provides services to the agricultural and natural resource academic community that are simply not available

through any other source, including the Integrated Post-Secondary Education Data System (IPEDS)."

• Statistical Expert Panel meeting –

The FAEIS Statistical Expert Panel meeting is tentatively set to be held in Washington, D.C., in March or April of 2011. Dr. Eric Smith is in charge of the meeting and will be in contact with the invited expert statisticians. One aspect of the expert panel will be to explore the "no-universe database" issue.

RFA Items		Timeline for Deliverables	
3	√	Creation of SAS dataset and report verification	10/2010
2,5	√	Identification of outliers and missing data	11/2010
9	√	Statistical update quarterly report	01/2011
2		Identification of redundant entries and miscoded CIP codes	02/2011
2		Automated identification of problematic data	02/2011
2		Comparison of IPEDS data and FAEIS data	03/2011
1,6		Statistical Expert Panel meeting	03/2011

K·STATE

Kansas State University.

Agriculture-Academic Programs
117 Waters Hall
Manhattan, KS 66506-4015
785-532-6151
Fax: 785-532-6897

Jan. 10, 2010

Jermelina Tupas
USDA – National Institute of Food and Agriculture
1400 Independence Ave. SW
Mail Stop 2215
Washington DC 2025-2215

Dear Dr. Tupas,

In my capacity as chair of the Association of Public and Land-grant Universities' Board on Agriculture Assembly's Academic Programs Section (APS) I write to you in support of the Food and Agricultural Education Information System (FAEIS). FAEIS provides services to the agricultural and natural resources academic community that are simply not available through any other source, including the Integrated Post-Secondary Education Data System (IPEDS). For example, because FAEIS primarily uses data directly reported by colleges or departments of agriculture, it is able to break fields such as animal science into more specific areas than what one may find in IPEDS. Furthermore, because the data comes directly from the academic administrators in charge of these programs and not from an institution's institutional research (IR) office, there is more consistency in the data grouped under a particular heading than what the IR office may provide. This has been demonstrated when comparisons are made against individual institution's IR data. The reason for this is that the program individuals closer to the actual programs are able to provide a better assessment of how to group their majors than people across campus who may not have a feel for the programs involved.

FAEIS provides other advantages as well. For example, FAEIS allows the opportunity for administrators at institutions to easily track enrollment trends in disciplines, and it also allows them to see how they stack up against their peers and to identify areas where they may be ahead or behind. Furthermore, FAEIS is up-to-date on its information versus IPEDS data that may be two years old. FAEIS is also quite easy to use, far easier than IPEDS. Finally, the FAEIS team has been tireless in adding to the information they collect each year and are in the midst of completing a review and correction of any incorrect information in the database. As they have increased the type of data available through FAEIS, the value of the database continues to grow. For example, FAEIS data was heavily employed in the researching and preparation of the latest USDA employment report posted on Purdue's web site, *Employment Opportunities for College Graduates in Food, Renewable Energy, and the Environment, United States, 2010-2015*.

APS strongly endorses FAEIS and encourages USDA to continue its funding.

Sincerely,

Don Boggs

Don Boggs
Associate Dean, Academic Programs
Kansas State University

"Knowledge for Life"

NATIONAL INSTITUTE OF FOOD AND AGRICULTURE; U.S. DEPARTMENT OF AGRICULTURE

Food and Agricultural Education Information Systems (FAEIS)

CONTINUATION APPLICATION GUIDELINES – 2010

INITIAL ANNOUNCEMENT

CATALOG OF FEDERAL DOMESTIC ASSISTANCE: This program is listed in the Catalog of Federal Domestic Assistance under 10.200

DATES: Applications must be received by close of business (COB) (**5:00 p.m. Eastern Time) on the date designated by the NIFA Program Contact**. See the correspondence that accompanies this request for applications.

NIFA often issues requests for applications (RFAs) that contain information divided in eight parts. All information in the eight parts is not necessary for requesting applications for these funds, however, the unnecessary parts are included herein and are denoted by "Reserved," in order to maintain the overall RFA structure for reference to RFAs by other documents.

PART I – FUNDING OPPORTUNITY DESCRIPTION

A continuation award is issued for a specified level of effort for a predetermined period of time with a statement of intention to provide additional support at a future date. The process of applying for the future funding begins with a noncompeting application for an additional funding/budget period within the previously approved project period. Continuation of the funding for the award is provided if performance has been satisfactory, appropriations are available for this purpose, and continued support is in the best interests of the Federal government and the public.

A. Legislative Authority and Background

This project is authorized by section 1417(e) of the National Agricultural Research Extension, and Teaching Policy Act of 1977, as amended (7 U.S.C 3152(3)), which authorizes the Secretary of Agriculture to maintain a national food and agricultural education information system that contains information on enrollment, degrees awarded, faculty, employment placement, and other similar information in the food and agricultural sciences as the Secretary considers appropriate.

B. Purpose and Priorities

FAEIS offers a broad range of higher education data in the fields of agriculture, natural resources, forestry, veterinary medicine and human sciences. Collaboratively it provides information with data from Integrated Postsecondary Education System (IPEDS) from the Department of Education and the NSF (National Science

Foundation) Survey of Earned Doctorates. Working through Memoranda of Understanding (MOU) with the professional associations in agriculture and related sciences, FAEIS will continue to collect data on students and faculty through web-based, on-line surveys. This database contains information on enrollments and degrees awarded by race, gender, academic discipline, institution, state and region. Additional information on placement of graduates as well as faculty in the various disciplines is also included. The database has been expanded to include information on benchmarking and international opportunities for students in these fields. Policy makers often use such data in determining what type of programs to fund. Deans and faculty at land-grant and other colleges and universities use these data to determine what types of academic programs to offer. Additionally grant recipients utilize the data as an assessment of program growth. Counselors and students utilize the information to choose majors and educational institutions. The NIFA Higher Education Programs unit uses these data to respond to congressional and other external requests.

Goals of the FAEIS are:

- Annually collect data on student enrollment, degrees awarded, placement of graduates and faculty in disciplines relating to agriculture, natural resources, human sciences, veterinary medicine and related fields utilizing an on-line database and provide reports to clientele through an on-line reporting system. Maintain longitudinal information for trend analysis.
- Maintain databases for: benchmarking of the Human Science Programs and International Programs involving but not limited to teaching and study abroad opportunities.
- Maintain the website for national Academic Programs Awards and Regional Teaching Workshops and other related sites.

C. Program Area Description

Funding under this authority supports a project for the grantee to administer the FAEIS by performing the following activities:

The recipient of this agreement will perform the following FAEIS activities:

Previous years of Contract

1. Continue the maintenance of the website, **http://faeis.usda.gov**, with appropriate links to include:
   a. Employment Opportunities Report and Technical Addendum (past and future reports)
   b. Teaching Workshops and Awards
   c. REEIS data and other ancillary data residing at NIFA
   d. HEP Programs and summaries of Projects
   e. Other resources relating to Agriculture, Natural Resources, Human Sciences, and Veterinary Medicine including links to participating organizations.

2. Continue to provide information to personnel using FAEIS data to include:
   a. Monthly Newsletter highlighting trends in various discipline relating to data collected
   b. Highlights and Summary reports of data collected by year
   c. Trend Analysis of data collected over time
   d. Institutional Profiles of the institutes participating in FAEIS

3. Increase response rates. Expand the use of and participation in FAEIS by marketing the information to institutions and providing workshops for users and attending conferences where FAEIS can be showcased. Develop marketing materials to be distributed or utilized at conferences by the FAEIS team or Panel of Experts members.

4. Identify institutions not currently participating in FAEIS with programs which would enhance the FAEIS database as well as providing data to those institutions. This could include but is not limited to community colleges and private institutions with related programs. Provide them with information on FAEIS and invite them to participate.

5. Continue to maintain Board on Human Sciences Benchmarking database and provide reporting means for institutions who participate to track their data over time. Continue to work with BOHS on improving and expanding those who participate in the project.

6. Continue to maintain database for institutions participating in international education and exchange of students. Develop a reporting mechanism for extracting data from the database.

7. Conduct an assessment of the USERS of FAEIS data regarding its usefulness.

8. Continue annual assessment of FAEIS data entry personnel to determine and enhance the system based on feedback.

9. Conduct an Annual Meeting of the FAEIS Panel of Experts to review the accomplishments of the year and determine additional items which would benefit the users of FAEIS. This meeting should be augmented by regular briefings either via e-mail or conference/video conferencing with the Panel.

10. Participate in conferences and other workshops or meetings to promote the participation in FAEIS as well as the use of FAEIS data.

11. Expand the composition of the Peer Panel to include additional personnel from currently represented segments as well as unrepresented segments.

12. Produce appropriate marketing and distribution materials to include but not limited to: CDs of data collected and appropriate tables; brochures and banners.

Final Year of Contract

NIFA calls for significant advanced expertise, detailed reporting and communications:

1. Because FAEIS is a national database, it is expected that data management and analyses must be reviewed by an external expert panel. The expert panel will determine the limitations of FAEIS data, proper interpretation and analyses of data from a voluntary data submission process, which is an unreliable data collection source.

2. Data and data analyses must be the products of significant statistical expertise that reflects standards for survey data management, analyses, and interpretation. Transcription errors will be corrected by implementing quality control procedures before to statistical analyses are performed. Methods to accomplish this include:
   a. Proofing data visually in column by row format (Excel or SAS file) by FAEIS employees. In addition, use exploratory data analyses as an additional quality check and test the assumptions.
   b. Systematic testing of data to determine its data accuracy to a "gold-standard" database – which is the IPEDs database.
   c. Outlier tests to highlight abnormal values and eliminate them before other statistics are calculated. Also, remove any redundancy and "orphan records" from the database. Applicant should use appropriate statistical outlier tests can determine if data are wrong and can be removed. Examples include the Shapiro-Wilke test, non-parametric methods and robust statistics (e.g., median and median absolute deviation). Generally, census/survey data are messy and often require multiple imputation methods.

3. SAS software is to be used to conduct data management (Excel and SAS files as output) to ensure ease of data transfer.

4. Data must correctly reflect the real world. All tests and procedures correcting the data must be completed before the FAEIS clienteles are given access to the data.

5. Missing data must be addressed. Examples of analyses techniques to perform this include the (1) Casewise deletion, (2) Pairwise deletion, (3) Mean substitution, (4) Hot-deck imputations, (5) sample weight imputation, and (6) Proxy pattern-mixture analysis or a combination of others.

6. The "no-universe database problem" must be addressed. Each year, the numbers in FAEIS have increased because FAEIS has captured more data – not because the number of students has necessarily increased. Statisticians might call this a trend in the mean/count. In addition, degrees within Classification of Instructional Program (CIP) codes have changed dramatically over the years – including degrees that were not part of the original CIP (1981). For example, the 01 CIP (agriculture) has been changed 35 times since 1978 (Survey of Earned Doctorates – SED, NSF) according to the NSF. This process has by definition increased the numbers by adding new degrees to the CIP. This is also a trend in the mean/count. To adjust for the effect of the population (universe) increasing, the total number of science/engineering majors must be added to the FAEIS dataset.

7. Prepare a complete and up-to-date list of all sources, that is, FAEIS contacts and include in the final report.

8. Develop a final report using a similar format as the National Science Foundation (use SED as example) that shows the improvement of the database and tables with summarized results compared to IPEDs data.

9. Produce quarterly reports (due January 1, April 1, July 1 and October 1) on the progress in addressing transcription errors, outliers and missing values. Include statistical procedures used to correct and process FAEIS data.

10. Conduct a statistically valid, random survey to collect FAEIS clientele data. This survey must clearly define the target population and the random sample must match the target population. The sample size must be large enough and the response rate must exceed 70% (90-99% far better) (NSF=93%, IPEDS >99+ %). The survey must use various methods: mail, Internet, telephone surveys, etc. The survey must be well written, tested and contain no leading questions. The survey personnel selected must be professionals well trained to conduct surveys.

I. TITLE: Improvement and Marketing of the Food and Agricultural Education Information System (FAEIS)

II. AUTHORS:

- M. A. Marchant, Professor, Agricultural and Applied Economics, 313, Hutcheson Hall, Virginia Tech, Blacksburg, VA 24061. Phone: (540) 231-1674; email: Mary.Marchant@vt.edu
- T. P. Mack, Dean of Graduate Studies and Research, Division of Academic Affairs, 129 Stright Hall, Indiana University of Pennsylvania, Indiana, PA 15705. Phone: (724)-357-2244; email: tmack@iup.edu
- E. P. Smith, Professor and Department Head, Department of Statistics, College of Science, Virginia Tech, Blacksburg, VA 24061. Phone: (540) 231-5657; email: epsmith@vt.edu

III. Duration: September 1, 2010 to August 31, 2011 (Year Three of 3-Year Contract)

IV. Purpose:

The Food and Agricultural Education Information System (FAEIS) is a comprehensive web-based survey and database of student and faculty data from the 1862, 1890, 1994 and Non Land-Grant institutions. The purpose of FAEIS is to gather, compile, and distribute a broad range of higher education information related to the food, agricultural, life, human, veterinary, and natural resource sciences. FAEIS provides student enrollment, graduation, and job placement information, as well as faculty count and salary information. FAEIS centralizes these data into a single database. Availability of these data and related analyses is critical to administrators of higher education programs in agriculture, natural resources, forestry, family and consumer sciences, and veterinary science as they plan and deliver educational programs.

For the final year of the three-year contract, we propose to improve the reliability and validity of the data by implementing procedures as outlined in detail below.

We will also continue to market FAEIS so that it has more participants, more users, and provides more value. The current proposal advances the previous work during the past eight years of FAEIS development at Virginia Tech and takes the project to new levels.

At the end of the third year of the contract, the outcome will include improved database quality, expanded FAEIS with enhanced participation, usability and functionality as the primary US database for subject areas that we address.

V. Procedures:

Previous Years of the Contract

Based on the performance criteria outlined in the RFA for the section entitled "Previous Years of the Contract," we propose the following responses:

Performance Criteria 1:

Continue the maintenance of the website, http://faeis.usda.gov, with appropriate links to include:

a. Employment Opportunities Report and Technical Addendum (past and future reports)
b. Teaching Workshops and Awards

c. REEIS data and other ancillary data residing at NIFA
d. HEP Programs and summaries of Projects
e. Other resources relating to Agriculture, Natural Resources, Human Sciences, and Veterinary Medicine including links to participating organizations.

Response 1:

The FAEIS team will continue to maintain all of the web pages on the FAEIS website including:

a. Employment Opportunities Report and Technical Addendum (past and future reports)
a. Project Director Conference website and information
b. Teaching Workshops and Awards
c. REEIS data and other ancillary data residing at NIFA
d. HEP Programs and summaries of Projects
e. Other resources relating to Agriculture, Natural Resources, Human Sciences, and Veterinary Medicine including links to participating organizations.

Where appropriate, the FAEIS webmaster will proactively seek input from suppliers of information that is included on FAEIS web pages. When needs to more effectively communicate the goals of the FAEIS project are identified, the FAEIS team will develop web pages to meet those needs.

Performance Criteria 2:

Continue to provide information to personnel using FAEIS data to include:

a. *Monthly Newsletter highlighting trends in various discipline relating to data collected*
b. *Highlights and Summary reports of data collected by year*
c. *Trend Analysis of data collected over time*
d. *Institutional Profiles of the institutes participating in FAEIS*

Response 2:

The FAEIS Team will continue to develop monthly FAEIS newsletters highlighting FAEIS data and its uses. Newsletters show highlights, summaries and trends of data in order to inform users of the scope and usefulness of the FAEIS data. Profiles of institutions on the FAEIS websites will be enhanced.

Performance Criteria 3:

Increase response rates. Expand the use of and participation in FAEIS by marketing the information to institutions and providing workshops for users and attending conferences where FAEIS can be showcased. Develop marketing materials to be distributed or utilized at conferences by the FAEIS team or Panel of Experts members.

Response 3:

Participation by 1862 and 1890 Land-grant institutions (77 institutions) has been nearly 100% each of the past two years. Participation by 1994 Land-grant institutions has grown each of the past two years. The FAEIS Team

will strive to maintain high levels of participation and increase participation where it is less than 100% through nurturing old relationships and developing new ones with data entry users and institutional administrators.

FAEIS staff will continue to effectively utilize two rollup, floor displays and a FAEIS brochure. These materials will continue to be used at professional meetings and other such functions where we can market FAEIS to potential users. These marketing materials are used at professional association meetings of our sponsoring organizations and other program-related organizations. (Also see performance criterion 10 below.)

Performance Criteria 4:

Identify institutions not currently participating in FAEIS with programs which would enhance the FAEIS database as well as provide data to those institutions. This could include but is not limited to community colleges and private institutions with related programs. Provide them with information on FAEIS and invite them to participate.

Response 4:

The FAEIS Team will continue to contact non-participating institutions, with the goal of reporting the greatest possible undergraduate and graduate enrollments in the program areas of interest to FAEIS. With an already high percentage of participation with 1862 and 1890 Land-grant institutions, the FAEIS Team will concentrate on the 1994 Land-grants and the non-Land-grant institutions with the largest enrollments. We will also seek to engage more non-participating Hispanic serving institutions.

Performance Criteria 5:

Continue to maintain Board on Human Sciences Benchmarking database and provide reporting means for institutions that participate to track their data over time. Continue to work with BOHS on improving and expanding those that participate in the project.

Response 5:

The revised BOHS annual benchmarking survey was published for the second year. Reporting institutions in 2009-10 exceeded institution numbers for the first three years of the survey. The FAEIS Team will continue to work with BOHS on improving the survey and developing a standard report for participants.

Performance Criteria 6:

Continue to maintain database for institutions participating in international education and exchange of students. Develop a reporting mechanism for extracting data from the database.

Response 6:

The International Program Database (IPD) continues with a solid support base among national and international reporting institutions. The database currently has 166 institutions participating, with over 380 projects logged in the system. We will continue to be aggressive in seeking new institutions to add programs to the database. The reporting mechanism will be developed for extracting data from the database.

Performance Criteria 7:

Conduct an assessment of the USERS of FAEIS data regarding its usefulness.

Response 7:

Four years ago we developed a standard set of 10 questions through an electronic survey instrument. This longitudinal survey is administered to over 1200 FAEIS users and data entry personnel each year. This survey will continue for a fifth year because it allows us to analyze trends in level of satisfaction over time. A report of our findings will be presented to the FAEIS Peer Panel and the USDA. We will defer to the guidance of NIFA's National Program Leader for FAEIS in implementing this item, given the survey requested in the "Final Year of the Contract" section (item 10).   We would like to have the survey designed so that information from previous surveys is not discarded.

Performance Criteria 8:

Continue annual assessment of FAEIS data entry personnel to determine and enhance the system based on feedback.

Response 8:

See entry above in Response 7.

Performance Criteria 9:

Conduct an Annual Meeting of the FAEIS Panel of Experts to review the accomplishments of the year and determine additional items which would benefit the users of FAEIS. This meeting should be augmented by regular briefings either via e-mail or conference/video conferencing with the Panel.

Response 9:

We will continue to work with the FAEIS Peer Panel in seeking their advice to insure that FAEIS is responsive to user needs. As in the past, we will also host a semi-annual audio conference of peer panel members, as well as communicate with them regularly through email, to discuss ways of improving FAEIS. The composition of the Peer Panel was expanded in 2009-2010 to increase diversity (see response to performance criterion 11 below). The FAEIS team will work to orient these new members and to maintain relationships with existing sponsoring organizations and their representatives.

Performance Criteria 10:

Participate in conferences and other workshops or meetings to promote the participation in FAEIS as well as the use of FAEIS data.

Response 10:

FAEIS personnel will continue to attend conferences and workshops to market FAEIS. We will continue our increased efforts to submit abstracts for poster and workshop presentations at these conferences. Personnel will also use the FAEIS display and distribute FAEIS brochures at these workshops and conferences. Projected annual meetings that will be attended include the following, many of which are FAEIS' sponsoring professional associations: American Association of State Colleges of Agriculture and Renewable Resources (AASCARR), Association of Public and Land-grant Universities (APLU), Board on Human Sciences (BOHS), First Americans Land Grant Consortium (FALCON), Association of Colleges and Universities (HACU), Society of American Foresters (SAF), Council of Administrators of Family and Consumer Sciences (CAFCS) and North American Colleges of teachers of Agriculture (NACTA).

Performance Criteria 11:

Expand the composition of the Peer Panel to include additional personnel from currently represented segments as well as unrepresented segments.

Response 11:

The composition of the Peer Panel was expanded in 2009-2010 to include a 1994 Land-grant representative from the First Americans Land-grant College Organization and Network (FALCON) and a Hispanic-Serving Institution (HSI) representative from the Hispanic Association of Colleges and Universities (HACU). A possible new segment could be a representative of two-year institutions. The FAEIS Team will work with the Peer Panel and USDA to identify an individual from this segment for membership.

Performance Criteria 12:

Produce appropriate marketing and distribution materials to include but not limited to: CDs of data collected and appropriate tables; brochures and banners.

Response 12:

FAEIS staff will continue to use two rollup, floor displays and our FAEIS brochure developed in 2008. These materials will continue to be used at professional meetings and other such functions where we can market FAEIS to potential users.

Final Year of Contract

Based on the performance criteria outlined in the RFA for the section entitled "Final Year of the Contract," we propose the following responses. Items 1-6 and 10 in this section of the RFA focus on FAEIS data, with responses below:

Data quality assurance

For FAEIS data to be useful it is necessary that the data meet stringent quality criteria. Data quality will be assured through the use of several checks and balances. First, the data will be visually checked to search for transposed values and obvious extreme values. Second, SAS will be used to perform quality checks using 3-sigma checks and outlier tests. Checked and validated data will be used to estimate standard deviations that will be used for the 3-sigma window for new data. Outliers will be checked using standard tests and visual displays, such as box plots and probability plots.

Outliers will be evaluated through a chain of command approach using an approach based on check, verify, validate and adjust. The data analyst and graduate students will check for outliers or odd values using visual or statistical checks. The outliers will be verified by the project manager. Verification of the outliers will be undertaken by the students and data analyst by checking with the provider of the data. Based on the validation step, the observation will either be replaced or not altered. The process will be documented and recorded as part of the Metadata.

SAS software will be used to create programs to check the observations. The SAS code will be written by the data analyst and verified by the project investigator (Smith). The steps in the program will include documentation.

A part of the data assurance program is the education and training of graduate students on the importance of data quality and the steps in ensuring high quality data. Part of the orientation of new student workers will include training on basic and statistical methods for detection of odd observations. A training manual will be produced to help with the process.

Missing data

Data from the reports will sometimes have missing values. Values missing at random will be estimated using standard imputation methods. The approach will be based on pairwise deletion, mean imputation or more complex methods (PROC MI on SAS), depending on the use of the data. SAS code will be documented and information about imputation will be incorporated into the metadata file.

Expert panel

To ensure the quality and utility of the FAEIS data base, an expert panel will be organized. The panel will review the approaches related to data quality and statistical issues related to the data. The panel will consist of four members, selected based on statistical background and knowledge of FAEIS or other similar data bases. We anticipate one meeting with the panel, preferably in the DC area.

Random Survey of FAEIS clientele

FAEIS personnel will collaborate with VT Center for Survey Research to conduct a statistically valid random survey of FAEIS clientele. The survey will be tested using the above expert panel. We will strive to achieve a response rate of over 70% and will use internet with follow-ups using mail and telephone to achieve this response rate. A stratified approach will be used to ensure that the sample is representative of the overall population. We will base the survey on the previous longitudinal survey and adjust the survey to eliminate any misleading questions.

Items 7-9 in this section of the RFA focus on FAEIS contacts and project reporting, with responses below:

Project Contacts and Reports

The FAEIS team will prepare a complete and up-to date list of all FAEIS contacts including their name, title, address, email, phone number, and primary role (data entry, data user, etc) within FAEIS.

The FAEIS team will produce four quarterly reports (January, April, July and September) and one final report in October 2011 to document the progress being made to address database transcription errors, outliers and missing values. The reports will include a description of statistical procedures used to correct and process FAEIS data.

VI. Timetable

Objective	Description	Year One	Year Two	Year Three
1.1	Enhance And Improve The FAEIS Website.	✓	◌	◌
1.2	Develop And Distribute Marketing Materials For FAEIS.	✓	✓	✓
1.3	Participate In Conferences & Workshops.	✓	✓	✓
1.4	Provide FAEIS Data To Users.	✓	✓	✓
1.5	Identify Institutions Not Currently Participating In FAEIS, Provide Them With Information, And Invite Them To Participate.	◌	✓	◌
2.1	Enhance The Board On Human Sciences Benchmarking Database And Provide Reporting Means To Institutions.	✓	◌	✓
2.2	Enhance The International Education Database And Provide Reporting Means To Institutions.	◌	✓	✓
2.3	Enhance The Regional Teaching Workshops website.	✓	◌	◌
3.1	Annual Survey of FAEIS Users Regarding Its Value.	✓	✓	✓
3.2	Conduct a Comprehensive Assessment of FAEIS to Determine Progress during the Three-Year Grant Cycle and Plan for the Future.	✓	◌	◌
3.3	Meet With The FAEIS Peer Panel Annually.	✓	✓	✓
	NEW: Expert Database Management Panel			◌
	NEW: Data Quality Assurance			◌
	NEW: Impute Missing Values			◌

✓ = priority activity, ◌ = activity at a lower level.