

## **PART B. DESCRIPTION OF STATISTICAL METHODOLOGY**

### **B.1. Statistical Design and Estimation**

#### **B.1.1 Survey Population**

The Facilities survey is designed to provide national estimates for U.S. colleges and universities with research expenditures equal to or greater than \$1 million in the prior academic fiscal year (i.e., in FY 2010 for the FY 2011 cycle and FY 2012 for the FY 2013 cycle). The FY 2011 cycle is anticipated to be a census of approximately 495 institutions. The listing of eligible institutions will be derived from the NSF Survey of Higher Education Research and Development. No sampling will be conducted.

NSF is seeking a 95% response rate to this survey. The response rate on the FY 2009 survey was 95%.

#### **B.1.2 Estimation Procedures**

No sampling weights will be required because the survey is a census. However, adjustments will be performed for both unit nonresponse and item nonresponse, with the approach depending on the level of nonresponse and the characteristics of the particular item involved (for item nonresponse).

##### ***Adjustments for Unit Nonresponse***

Since some nonresponse is likely, provisions will be made to compensate for the missing data in the survey estimates. Unit nonresponse (an institution does not respond to the entire survey) occurs when there is no information for a sampled unit most often because of refusal to participate in the survey or failure to contact the respondent.

In the FY 2007 and FY 2009 survey cycles, unit nonresponse for the research space section of the survey (Part 1) was handled by assigning weights to the participating institutions. The nonresponse weight was the ratio of the number of eligible institutions in the survey to the number of responding eligible institutions.

The weights for the academic institutions were adjusted for the known number of academic institutions by: R&D expenditure categories (the quintiles of the distribution), census region, control (public/private), and whether the institution granted Ph.D. degrees. The minimum weights were constrained to be at least 1.0. NCSES anticipates using a similar weighting approach for the research space section for the FY 2011 and FY 2013 cycles of the survey.

The data in Part 2 of the survey (networking and computing) was not weighted in the FY 2007 and FY 2009 cycles because of the greater potential for measurement error within the survey responses. It is believed that substantially greater measurement error may exist in the Part 2 data because the majority of the Part 2 questions change with each survey cycle due to extremely rapid developments in cyberinfrastructure. For example, approximately two-thirds of the Part 2 questions were implemented for the first time in the FY2007 survey cycle. Also, extensive variability in the cyberinfrastructure environments and expertise at different institutions may lead to greater measurement error. NSF anticipates using a similar approach for the Part 2 data for the FY 2011 and FY 2013 survey cycles. NSF plans to consider weighting for unit nonresponse beginning in survey cycle FY 2011.

### ***Adjustments for Item Nonresponse***

Item nonresponse occurs when there is no information for a respondent on an individual item in the questionnaire, most often because of refusal to answer that item or the provision of an invalid response (e.g., one that falls outside of the possible range of values). We will use imputation on selected variables to adjust for item nonresponse for Part 1 variables and will not use imputation for the Part 2 items (see previous section).

The method of imputation will depend on the characteristics of the variable. In some cases logical imputation might be used, with the response to one item being used to infer a response on another item. For example, an institution that indicates on one questionnaire item that it does not have a medical school can be assumed to make equivalent responses elsewhere on the questionnaire, even if the item is left blank. In other cases, statistical imputation might be used, based on a statistical model to predict the expected response of the institution (e.g., based on responses elsewhere in the questionnaire, responses to previous cycles of the survey, or responses of similar institutions).

Flags will be created to indicate all instances in the database of imputed values.

## **B.2. Survey Procedures**

The facilities survey is a mixed-mode mail and web survey, with telephone and email followup.

The president of each institution is mailed a copy of the questionnaire, a cover letter, and a copy of the report from the previous (FY 2009) survey cycle. In addition, the president receives one of two institutional coordinator forms depending on whether the institution planned to retain the previous survey cycle coordinator for the upcoming survey cycle (see below). NSF has found through experience that different sections of the questionnaire are often completed by different offices throughout the institution, so it is important to have an institutional coordinator who can delegate sections of the questionnaire to appropriate individuals, and sometimes prepare a composite response based on their individual responses. The coordinator also acts as the central communication point for NSF and the contractor collecting the data.

During the previous survey cycle each president is asked if they wish to keep that year's institutional coordinator for the next cycle's data collection. If the president indicated that the coordinator would be the same, the president is sent a pre-filled form along with the previously mentioned materials indicating their current cycle coordinator (and providing an opportunity to name a new coordinator if he/she wishes to do so). Simultaneously, the prior cycle's coordinator receives a letter indicating that data collection is beginning and that his or her name has been provided to the president as the past coordinator. At this time the coordinator also receives a copy of all materials.

If the president indicated in the previous cycle's data collection that he/she did not wish to keep the same coordinator for the next cycle's data collection, the president receives a blank form to use to indicate the current cycle's coordinator. To aid a president in selecting a new coordinator, the letter to the president indicates who acted as the coordinator in the previous survey cycle (if the institution responded to the cycle). Simultaneously, the prior cycle's coordinator receives a letter indicating that data collection is beginning, that a letter has been sent to the institution's president requesting a coordinator, and that his or her name has been provided to the president as the past coordinator. At this time the coordinator also receives a copy of all materials.

If no response is received from the president's office within a week, telephone prompts are used to determine the name and contact information for an institutional coordinator. Following designation of the coordinator, the coordinator is notified that he or she has been appointed survey coordinator.

Regular email and/or telephone prompts are used to encourage the institution to respond. Institutions have the option of completing either a paper copy of the questionnaire or providing the data on the web through a designated web site. Based on past experience, we expect over 95% of the institutions to respond using the web. Returned questionnaires are examined for quality and completeness using both visual and computerized edits. In the case of questionnaires completed on the web, computerized edits check for quality and completeness as the data are entered, and prompt the respondents if problems are found. If key items have missing data or other problems appear in the data (e.g., two responses appear to be inconsistent), then respondents are recontacted to resolve the issues.

### **B.3. Methods for Maximizing Response Rates**

NSF is seeking a 95 % response rate in this survey.

A key to achieving this response rate is the tracking of the response status of each included institution, with telephone followup of those institutions that do not respond in a timely manner. The survey responses will be monitored through an automated receipt control system. Approximately three weeks after the initial mailout, the contractor will begin calling nonrespondents to verify that they received the questionnaire and to prompt the individuals to respond. Additional telephone or email prompts will be made as the data collection period continues.

Several other steps will be taken to maximize the response rate. The survey materials will provide a toll-free 800 number that people may call to resolve questions about the survey. Respondents may seek help by email. In addition, standard survey techniques that have proven successful in other academic survey efforts will be employed to achieve a maximum response rate. These techniques include:

- A cover letter signed by the director of NSF.
- Institutional coordinators will be contacted by telephone prior to the intended closeout of the survey. This contact is intended both to offer assistance to respondents and to encourage their speedy response.

- Follow-up telephone calls will be made to nonrespondent institutions as required. These follow-up calls are expected to achieve significant improvements in response rates.

Finally, institutions will be informed in their materials that institution-level survey responses are currently available for the previous survey cycles and institutional responses will also be available for the current FY 2011 (and FY 2013) survey. This data will be available on a publicly accessible database on the NSF website. NSF believes that having publicly available data will maximize responses rates because institutions will be more likely to participate if they believe that the data will be useful to them.

#### **B.4. Tests of Procedures and Methods**

The questionnaire is based on versions of the survey used in previous cycles. As part of survey improvement efforts, after each survey cycle the survey interviewers participate in an extensive debriefing. During the debriefing the interviewers discuss issues such as the questions respondents ask most frequently, the survey questions that posed problems for respondents, any administrative issues that arose, and any other survey improvement issues. In addition, the survey paradata is analyzed after each cycle implementation. The paradata includes: a list of the “other specific responses” for each question, the frequency of error messages for each question, missing data, consistency of question responses, and completed survey return flow. Based on these analyses, survey questions or procedures may be revised.

A total of eight institutions were visited for the purpose of conducting exploratory interviews to determine how respondents interpreted the questionnaire items and the instructions. An additional 6 telephone interviews were conducted to understand the administrative records the institutions kept on high performance computing use and users.

#### **B.5. Individuals Responsible for Study Design and Performance**

The individuals listed below participated in the study design.

Leslie Christovich, NSF	703-292-7782
Fran Featherston, NSF	703-292-4221

John Jankowski, NSF	703-292-7781
Timothy Smith, Westat	240-314-2305
Eric Jodts, Westat	301-610-8844

The contractor for the FY 2011 and FY 2013 data collection is currently unknown. Leslie Christovich at NSF/NCSES will oversee contract.