SUPPORTING STATEMENT

for

FY 2011 and FY 2013 Science and Engineering Research

Facilities

3145-0101

PART A: JUSTIFICATION

A.1.   Circumstances Necessitating Collection of Information

The National Science Foundation's (NSF) Survey of Science and Engineering Research Facilities is a congressionally mandated biennial survey that has been conducted since 1986.   The Facilities survey has two parts. Part 1 collects data on "bricks and mortar" research facilities; Part 2 collects data on components of cyberinfrastructure (currently networking and high performance computing).

Background

Academic research facilities in the fields of science and engineering (S&E) are an important national resource. Extensive hearings were held during the 99th Congress in the House and Senate committees on science and technology to examine the research facilities needs of universities and colleges. Both committees found "sufficient evidence to suggest the presence of a serious and growing problem…" and expressed concern that the Federal government did not have in place an ongoing analytical system to document the current status of and needs for research facilities by major field of science and engineering. Such systematic information was needed to understand current and future facilities needs and pressures and to formulate sound solutions over time.

In recognition of the need for objective information about research facilities, Congress directed the National Science Foundation (NSF), in the Authorization Act of November 22, 1985 (P.L. 99-159, Section 108

> The National Science Foundation is authorized to design, establish, and maintain a data collection and analysis capability in the Foundation for the purpose of identifying and assessing the research facilities needs of universities. The needs of universities, by field of science and engineering, for construction and modernization of research laboratories, including fixed equipment and major research equipment, shall be documented. University expenditures for the construction and modernization of research facilities, the sources of funds, and other appropriate data shall be collected and analyzed. The Foundation, in conjunction with other appropriate Federal agencies, shall conduct the necessary survey every 2 years and report the results to the Congress.

In response to this directive, NSF has now conducted twelve biennial surveys. The National Institutes of Health (NIH) cosponsored the surveys with NSF beginning with the initial survey. At the request of NIH, the survey also collected data from nonprofit, biomedical research organizations and hospitals that received NIH research funds.

Since 1986, NSF has modified the survey several times with the goals of improving data quality and reducing respondent burden. The original survey was limited and was conducted using NSF's Quick Response Survey questionnaire capability in order to meet the Congressional timeframes. The surveys from 1988 to 1996 built upon the 1986 Quick Response Survey instrument but expanded the level of detail in the data collected and expanded the size of the universe surveyed.

During the FY 1998 cycle, a Large Facilities Follow-up survey was added to the Facilities survey. The Follow-up survey collected data on the construction of buildings costing at least $25 million and housing at least some science and engineering research facilities as required in OMB circular A-21, revision dated June 1998. The data collected mostly concerned the costs of construction.

During the next survey cycle, data collection was expanded from a sample of institutions to a census of institutions in order to make the data collected more useful to Federal and State agencies and to the higher education community. A census of institutions allowed for more precise estimates of the status of S&E facilities, and allowed for the analysis of subgroups of institutions.

In the FY 2001 survey cycle NSF reduced the number of survey items from 10 items to two. The two remaining survey items requested data on: 1) the amount of net assignable square feet used for instruction and research by field of science; and, 2) the adequacy of the amount of S&E research space available at each institution. NSF also eliminated the Follow-up survey. The significant reduction in the survey was done in full consultation with the Office of Management and Budget, relevant Congressional staff, the National Institutes of Health and other interested parties.

The purpose of implementing the abbreviated survey was for NSF to focus completely on redesigning the questionnaire for the FY 2003 survey cycle. The goal of the redesign was to improve data quality and relevance and to reduce respondent burden. Until the abbreviated FY 2001 survey, the questionnaire had essentially remained the same over the years; a substantial detailed review and redesign of the survey had not been undertaken for 14 years.

The redesigned survey was initially implemented during with the FY 2003 survey cycle. Beginning in FY 2003, the redesigned survey included a new section focusing on computing and networking capacity. Computing and networking, along with other components of cyberinfrastructure, have become an essential part of the infrastructure for science and engineering research and will continue to be critical in the future.

Beginning with the FY 2011 survey, NIH will no longer be a survey cosponsor. NIH was legislatively mandated to cooperate with NSF in the production of a report that included data relevant to their National Center for Research Resources' Construction/Renovation program and the Animal Facility Improvement program. These two programs support construction and renovation awards at academic and biomedical research facilities. At the request of NIH, the survey has collected data from nonprofit, biomedical research organizations and hospitals that receive NIH research funds. However, NIH is no longer required to participate in the NSF report and the construction programs are no longer funded by Congress. While NIH is still a user of the survey's data on academic institutions, given general budget constraints and the lack of funding for the construction program, NIH no longer cosponsors the survey. Consequently, biomedical research institutions will no longer be included in the survey population and several survey questions specific to NIH's programs will be deleted.

General survey description

The NSF National Center for Science and Engineering Statistics (NCSES) is responsible for developing and maintaining the research facilities survey system. NCSES now requests approval to conduct the next cycles of the Survey of Science and Engineering Research Facilities (hereafter Facilities survey) for fiscal years 2011 and 2013. The following is a broad overview of the survey.

As required by the authorizing legislation, multiple survey questions ask respondents to report data according to fields of science and engineering. The Facilities survey incorporates science and engineering fields from the Classification of Instructional Programs (CIP) developed and used by National Center for Educational Statistics (NCES). Because colleges and universities must report information to NCES using this field classification system, many institutions use these categories in their institutional databases, including space databases. Therefore the Facilities survey uses this same classification system so that institutions can report on the science and engineering fields with minimum burden. Approximately every decade NCES updates CIP based on changes in the study programs

offered at higher education institutions. The Facilities survey updates the fields on the survey following each NCES update. The FY 2011 survey is updated to reflect the changes made to the FY 2010 CIP.

Questions 1 through 5 request data on the amount of space used for science and engineering research. The questions request specific information about institutional research facilities (e.g., how much research space is located in medical schools) because different kinds of space have different costs and requirements for preparing and using the space.

Question 6 asks about the condition of the research facilities because of a recurring concern about aging research infrastructure in the U.S. Also, this question may provide data on whether existing research space is appropriate for the rapidly changing needs of scientific research.

Questions 7 and 8 ask about the costs of repairs and renovations in the current and immediately previous fiscal year. Similar to Question 6, these questions also relate to the issues of a potentially aging infrastructure and the changing needs of science, but from a financial perspective.

Question 9 and an accompanying project worksheet ask about new construction in the current and immediately previous fiscal year. New construction allows institutions to increase the total amount of research space, to replace inadequate facilities, and to adjust to the changing needs of science. The worksheet requests information on individual projects such as costs per square feet of new construction.

Question 10 asks about the sources of project funding for repairs/renovations and new construction to determine the roles various sectors play in funding research facilities.

Questions 11 through 14 ask about planned new construction and planned repairs and renovations over the next two fiscal years. These provide data on the extent to which institutions are addressing their research space needs.

Questions 15 through 18 ask about deferred repairs/renovation and new construction. These provide another measure of institutions' anticipated need for space, while also examining the ability of institutions to meet their needs.

Part 2 of the survey requests information on cyberinfrastructure components and currently focuses on high performance computing, data storage, and networking capacity. Cyberinfrastructure is

an important component of the academic institutional research infrastructure and had been largely unmeasured until this section was added to the survey. Researchers widely agree that the increased power of computers, combined with the use of networking to share resources and facilitate communication, is significantly changing the nature of scientific research. In some cases, these capabilities may also reduce the need for "bricks and mortar" research facilities, now and in the future, because cyberinfrastructure can provide the ability to share remote facilities or to conduct computer simulation.

Questions 1through 5 and question 9 in this section focus on external network connections and access. The first 5 questions ask respondents to report their total institutional bandwidth, their high performance bandwidth and their sources of bandwidth. Networking, especially high performance networking, has become essential to a wide variety of scientific activities. For example, some areas of scientific inquiry increasingly rely on networks to access massive databases, collect data, or access remote instrumentation. In addition, networking has become an essential communication mechanism. Question 9 requests information on wireless network coverage at institutions. Some institutions use wireless communications as a substitute for hard wire connections, while others use wireless to collect data (weather sensors or from patients readings in hospitals) or for specialized applications such as robotics.

Question 6 asks about the speed of desktop port connections. Desktop port connections are often called the "last mile" of networking because of their importance to overall institutional networking capacity. Institutions can have connections to fast networks outside their institution and they may have fast internal networks connecting desktops across campus, but it is equally important to total networking capacity that they also have fast connections to actual desktop computers.

Question 7 asks about the presence of dark fiber at an institution. As networking fiber that has been laid but is not currently being used, dark fiber is an important measure of an institution's ability to expand networking capacity.

Question 8 asks about the speed of an institution's internal network. The institution's internal network is the between the desktop ports where the research is performed and external bandwidth connections to providers.

Question 10 is an open ended opportunity for respondents to provide any clarifications or additional information about their networking capabilities.

Questions 11 through 18 request information on the availability of centrally administered high performance computing (hereafter HPC) both to the internal campus community and to potential users external to the institution. The extent of centralized HPC has just begun to grow at many campuses over the last few years. Previously, HPC was only available to those researchers who received grants to fund such a capability. These resources were not usually shared. However, with increased concern about security, cost, and serving the entire campus community, more and more institutions are making this a centralized service available to all faculty, students, staff, and researchers. Further, with the expansion of HPC to numerous fields of study and the increasing necessity of HPC to address large scientific issues, centralized HPC will likely continue to grow. Questions 19 through 21 ask about the data storage capacity for HPC that is so important to making full use of HPC capacity.

Finally, Question 22 allows respondents to elaborate on any of their responses on HPC in an open ended question format.

A.2.   Purposes and Use of the Data

Data from this survey serve several audiences.   At the national level the data provide Congress and federal agencies with a broad, quantitative picture of the inventory and condition of existing S&E research space at research-performing academic institutions. The survey also provides data on the current and future capital expenditures for research facilities by universities and colleges; sources of funding for research facilities at these institutions; plans for future repair/renovation and new construction of science and engineering research facilities; and on computing and networking capacity. Congress uses the Facilities survey data to assist them in estimating appropriations and programmatic decisions. For example, Congress used the Facilities survey data to assist in determining funding levels for NIH's Construction/Renovation Program under the American Recovery and Reinvestment Act of 2009 (ARRA). NIH also used the information to help them design the size and type of procurements using the ARRA funds. NSF used the information to assist the programs in designing procurements for cyberinfrastructure under ARRA.

While no longer a survey sponsor, NIH still uses the data to understand the magnitude and condition of biomedical research facilities at higher education institutions. Further, NIH has previously used survey data to prepare reports on the Government Performance and Results Act.

At the state and local level, state legislatures, state agencies, and individual universities and colleges use the data to assist them in making budgetary decisions and in planning future activities. Private sector architectural, engineering, and construction firms use the data for several purposes including forecasting the demand for their products and services. Finally, professional associations such as the Association of American Medical Colleges use the data to assist them in assessing needs related to their associated fields of science.

In an effort to increase the usefulness of the data, beginning with the FY 2003 data, most survey data became publicly available with institutional identifiers (see A.11. on sensitive questions for exceptions). The facilities survey data is included in NCSES's public-use database that allows institutions (and others) to access the data for analytical purposes. We believe that this change increased the number of data users, the quality of the survey responses, and the ways in which the data are used.

A.3.   Use of Improved Information Technology

The survey is a mixed mode survey: respondents have the option of completing a paper version of the survey or a web version. The large majority of institutions (94% in FY 2009 cycle) respond to the survey via the web.

The web version of the survey is a .NET application. There is a real-time monitoring system, which allows NCSES to monitor data, response status, and comments from respondents. From the perspective of the respondents, the web version is more convenient and simplifies the survey (e.g., by automating skip patterns, and automatically calculating totals). NCSES also benefits from use of the web version by receiving improved data quality. Data quality is improved using several techniques such as data checks, skip pattern checks, and consistency checks. For example, data checks are used to calculate totals and inform respondents if totals do not add to 100 (in the situation of percents). For inconsistent data, respondents will receive messages while completing their survey if their responses do not appear consistent with their responses to other relevant questions.

A.4.   Efforts to Identify Duplication

NSF monitors the availability and release of higher education facilities data. While some sources provide data on amount of construction or costs of construction, no organization provides data on higher education research facilities, individual fields of science, or science and engineering in the aggregate. For example, at the national level, data on the book value of higher education physical plant assets is collected periodically on the NCES Integrated Postsecondary Education Data System. However, these data are only collected in the aggregate for an institution; the data are not available for research space, and the data are not available by S&E field.   The Survey of Scientific and Engineering Research Facilities is the only survey to collect data from research-performing higher education institutions on research space and fields of S&E.   The data do not duplicate statistical data from any other sources.

Currently, no other organization collects data on the fundamental electronic/digital infrastructure (cyberinfrastructure) underlying an institution's capacity to electronically support their faculty, researchers, students, and staff. A limited number of organizations collect data on information technology systems or devices that make use of cyberinfrastructure. For example, The Campus Computing Project collects data on the presence and pervasiveness of mobile apps and laptops, among other information technology information. An institution's ability to support this and other information technology however, is dependent on the general underlying cyberinfrastructure.

A.5.   Collection of Data from Small Businesses

Data are not collected from small business.

A.6.   Consequences of Less Frequent Data Collection

The Congressional mandate requires that data collection occur every two years, which ensures that the data are current. This also lessens burden when compared with an annual survey. Conducting the survey less frequently would adversely affect the relevance of the data both for policymakers and users of the data. The Foundation has consulted with an expert panel and numerous individual institutions (see section A.8. on public comment and consultations outside the agency), and a two-year period largely corresponds with the frequency that many institutions update their records and with their planning processes.

A.7.   **Special Circumstances**

No special circumstances.

A.8.   **Public Comment and Consultations Outside the Agency**

Following implementation of the FY 2009 facilities survey, NSF consulted with survey respondents and cyberinfrastructure experts outside the agency.

In the spring of 2011, NSF conducted seven site visits to conduct early scoping interviews on the survey's bandwidth and HPC questions. In selecting institutions for the visits, NSF considered several variables including: institutional total bandwidth, source of bandwidth, access to high performance networks, and the amount of R&D expenditures. In addition, institutions may have been selected if they provided high performance computing to their campus and if they made their HPC available to external users.

The first Federal Register notice for public comment was published on May 16, 2011 at 76 FR 28244 and no comments were received.

Interviews were conducted to understand the current networking and HPC environments at academic institutions, to determine how respondents interpreted various jargon associated with networking and HPC, and to elicit their opinions on emerging issues. Participants typically held positions such as Vice Provost of Information Technology, Chief Information Officer, Vice Provost for Infrastructure, Director of Network Engineering and Technology, or similar. The following institutions participated in the on-site interviews.

George Mason University
University of Maryland, Baltimore County
University of Maryland, College Park
Georgetown University
Johns Hopkins University
George Washington University
Morgan State University

Phone interviews were made to 5 institutions indicating on their FY 2009 survey that they had HPC. These were exploratory one hour interviews during which the interviewees described their HPC resources, their HPC internal and external users, and how the institution tracked individual user usage. The following institutions participated in the phone calls:

University of Minnesota
Cornell University
State University of New York, Buffalo
Louisiana State University
Purdue University

NSF obtained the consulting services of Dr. Curt Hillegas, Director of Research Computing, Academic Services, Office of Information Technology, at Princeton University. He also sits on the Steering Committee for the EDUCAUSE Campus Cyberinfrastructure working group. Dr. Hillegas engaged in several activities to assist NSF's assessment of both high performance computing and networking capacity including bandwidth measurement, high performance networking allocation, networking consortia, HPC measures, HPC architectures and storage, and emerging cyberinfrastructure issues. Dr. Hillegas reviewed FY 2009 survey questions and draft FY 2011 survey questions, reviewed the survey responses of specific institutions with which he was familiar, and conducted analyses of previous survey data.

Finally, NSF conducted analyses of the FY 2007 and FY 2009 data for data anomalies and potentially inconsistent data. Responses to open ended questions were also reviewed.

Based on information from the site visits, information from expert review of survey questions and consultation, and analyses of previous survey results and paradata, NSF revised the survey questions for the FY 2011 survey cycle.

A.9.   Payments or Gifts to Respondents

Respondents will not be paid or receive gifts.

A.10.   Assurance of Confidentiality

Data on individuals are not collected. Historically, the survey promised respondents that all individual institutional data would be confidential and that responses would not be used in any manner which would identify any individual institution's responses. Only aggregated data for statistical summaries would be presented. However, over the years, respondents and other organizations requested the data in a less aggregated and more accessible form. For example, colleges and universities wished to compare their institution to other institutions they define as their peers. State agencies and legislatures are interested in comparing institutions within their state to institutions in other states that they define as relevant.   Professional associations wish to compare institutional data. And finally, private sector firms wish to know which institutions are planning significant repair/renovation or construction work.

Based on the idea that the data should be available to all interested parties and users and that making the institutional level data more accessible will provide institutions with a reason to respond to the survey, the NSF made the data publicly available on an institution by institution basis with a few key exceptions (see section A.11 on sensitive questions) for the first time during the FY 2003 survey cycle. Institutions are clearly informed of NSF's confidentially policy at the front of the survey and the confidentiality of specific questions at the location of each specific confidential question.

A.11.   Sensitive Questions

This is a voluntary survey. Institutions are not required to participate in the survey and respondents may decline to answer any question in the survey. While NSF decided to make the large majority of the survey data publicly available, the expert panel and the cognitive interviews during the survey redesign in 2001 identified several survey items as particularly sensitive. The release of data on those questions could make institutions less inclined to respond to the survey. Also, during pretests of the redesigned survey respondents expressed concern about the confidentiality of certain items. To address this issue, the following survey items remain confidential: condition of research space (question 6) and the amount of research animal space (questions 3 and 9f).

The beginning of the paper and web surveys includes a section informing survey respondents of the confidentiality of the above survey items. The section identifies the relevant questions and informs

respondents that these responses will not be disclosed in identifiable form. In addition, each survey item contains a reminder of the confidentiality of that item.

A.12.   Estimated Response Burden

Burden hours

The Facilities survey is a biennial survey of approximately 495 academic institutions. For academic institutions the hour burden for any particular institution will be affected by two major factors--the size of the institution (in terms of number of S&E departments and/or the number and size of research facilities) and the status of the institution's computerized central records system. Previous surveys showed that many institutions, as part of their compliance with OMB Circular A-21, have created centralized databases, including a measure of space devoted to research.

Because of these factors, the completion time varies among institutions. The time to complete the research space section of the survey (Part 1) ranges from 10 to 85 hours with an average of 40 hours (based on pretests). The time to complete the computing and networking section of the survey (Part 2) averages 1 hour. Therefore, in total, the time per institution to complete the survey is expected to average approximately 41 hours. Assuming a 95% response rate (based on the FY 2009 response rate), this would result in an estimated total burden of 19,270 hours in FY 2011 and a similar burden in FY 2013. [(.95 response rate x 495 institutions) x 41 hours = 19,270]

Costs to responding institutions

Costs that will be incurred by responding institutions involve the completion time of the survey and the time of an institutional coordinator associated with managing the distribution of survey forms and transmittal to the contractor. The institutional coordinator makes the initial contacts with the relevant individual offices at the institution informing them of the survey, distributes the forms, follows up on the return of the forms or data to the coordinator's office, and transmits the forms or data to the contractor.

At an estimated cost of $35 an hour for institution staff, the average cost to each academic institution is $1,435 for FY 2011 and a similar cost for FY 2013. Assuming a 95% response rate, the

total respondent costs for the FY 2011 survey are estimated to be $674,450 [(.95 response rate x 495 institutions) x $1435].

These costs are expected to be similar for the FY 2013 cycle of the survey.

A.13.   Annualized Cost to Respondents

There are no capital or startup costs to the respondents for the Facilities survey.

A.14.   Annualized Cost to the Federal Government

The estimated total cost of the FY 2007 Facilities survey contract to the Federal Government is as follows:

Facilities Survey Manager	$ 152,635
Contract to conduct the survey	$1,253,000
Total estimate	$1,405,635

The contract will include all direct and indirect costs of data collection, analysis, reporting, and the production of public and proprietary data sets. These costs are expected to be similar for the FY 2013 survey cycle.

A.15.   Reasons for Program Changes

No program changes.

A.16.   Publication Plans and Project Schedule

The FY 2011 survey implementation is planned to begin in October of 2011. Data collection will take place over a six-month period in order to provide institutions adequate time to identify a coordinator and to gather the appropriate records and data. Data collection is expected to end the spring of 2012. The data collection for the FY 2013 cycle will begin in October of 2013 and will end in the spring of 2014.

The contractor will submit draft data tabulations within three months following completion of data collection.   These will be reviewed and revised as needed by NSF before production of final data tabulations. NSF will be responsible for electronic dissemination of the tables to NSF offices, participating academic institutions, professional associations, and others who may request information about the survey.

The contractor will also provide data for a public use data file that is accessible via the NSF website and includes electronic data files for use on personal computers.   The data files, including documentation, will contain institutional data and will be delivered within 3 months of data collection. In addition, the contractor will provide a methodology report detailing all survey activities, materials, and procedures.

An NSF-produced Infobrief on the results of the FY 2009 survey is scheduled for spring/summer of 2011. The Infobrief will include a presentation of major findings, charts and graphs relevant to the survey, and relevant statistical tables. It will also refer the reader to additional information about survey limitations and other technical information.

NSF anticipates a similar schedule for the FY 2011 cycle of the survey.

A.17.   Approval for Not Displaying the Expiration Date for OMB Approval

Approval not requested.

A.18.   Exceptions to the Certification Statement

There are no expectations.