

SRS HERDS ICR
Response to OMB Questions
July 20, 2009

1. Page 5, under A2, item 3: We believe that this language was revised by the COMPETES Act. Please confirm and update as needed.

Yes, the language was revised as follows:

Paragraphs (1) and (2) of section 4(j) of the National Science Foundation Act of 1950 (42 U.S.C. 1863(j)(1) and (2)) are amended by striking ` , for submission to' and `for submission to', respectively, and inserting `and'.

We have corrected the supporting statement (attached).

2. Page 8, A4, 2nd paragraph – Does “character of work” refer to defense/non-defense or just basic/applied/development? It might be useful to collect defense/non-defense without getting into the classified/unclassified split.

This refers to basic/applied/development, the breakout which is currently requested for total R&D expenditures. The survey also requests federally funded R&D expenditures to be reported by agency, including the Department of Defense (DOD). Therefore NSF does collect data on DOD-funded academic R&D expenditures. NSF does not plan to request the classification of all R&D projects as defense-related or not, although this may be a classification that is added in future years as we continue to evaluate the best taxonomy (or taxonomies) to be used for academic R&D.

3. Page 8, A5, This is obviously not germane to this collection, but how are the SBIR/STTR expenditures tracked?

The data are submitted by federal agencies to the Small Business Administration.

4. Are we consistently wrong in terms of the discrepancy between what Federal agencies report funding and what performers report expending?

If the question is whether the differences between agency and performer reports of R&D are consistently in the same direction, i.e. agency reports are always larger, the answer is no. If the question is whether within a given sector, the differences between agency and performer reports are consistent over time, the answer is yes. For the higher education sector the amount of federally funded R&D expenditures reported by universities is *larger* than the amount federal agencies report obligating to universities for all years. For the business sector the amount of federally funded R&D reported by businesses is substantially *smaller* than the amount federal agencies report obligating to the business sector for all years.

5. Page 10, A10

a. Please clarify the rationale for releasing only aggregate totals for lines D(1) and d(2).

The decision to keep the unreimbursed indirect costs line confidential was made in 1978 when the survey began requesting this breakout of institution funds in order to improve the data quality of the overall category. From the beginning the survey had requested the full costs of R&D to be reported, including both recovered and unreimbursed indirect costs associated with the R&D projects. However, in discussions with respondents NSF discovered that many institutions were not calculating the unreimbursed indirect costs portion at all, or were not doing it consistently.

Therefore for the FY 1978 survey, the category of institution funds was subdivided to include a breakout for unreimbursed indirect costs. The instructions for this subcategory included details on how to calculate the unreimbursed amount for those institutions who were not currently tracking these costs. Confidentiality was promised because many institutional respondents expressed hesitance at releasing information on the unreimbursed indirect costs and cost sharing portion of their R&D expenditures total. The main concerns were that (1) since many institutions do not "book" such expenses in their accounting systems, they were concerned about releasing such estimates that could not be tracked back on a project-by-project basis, and (2) the information would be used to justify lowering indirect cost reimbursement on grants, or to judge public institutions by how well they recovered indirect costs on R&D projects. Respondents felt that both uses would be inappropriate and misleading, because of the variety of types of projects and sponsors represented within the total. Because certain agencies cap their indirect cost reimbursement well below a normal institutional negotiated rate, some amount of unreimbursed costs is necessary and expected.

We asked about retaining the confidentiality of these sub-items on the redesigned HERD survey during our recent site visits and cognitive tests. The majority of respondents preferred keeping the confidentiality for the reasons stated above.

b. How are the "amount of unreimbursed indirect costs" verified?

There is currently no method of verifying this amount. Current university financial reports do not provide the amount of unreimbursed indirect costs at a project level. The survey instructions provide detail on how this amount should be calculated, but NSF is not able to access institutional accounting systems in order to verify that the calculations are done correctly.

6. Burden

a. Please break down the portion of the anticipated burden of the pilot into the one-time reprogramming burden and recurring burden.

Unfortunately NSF does not have a valid method of estimating that breakdown at this time. The estimate of 80 hours per institution was an extrapolation from the 22 hours currently estimated, based on the number and scope of new questions added. Our site visit discussions indicate that for many of the new questions the majority of the burden would be initial reprogramming, but this is highly variable depending on

whether schools are able to set up automated programs to collect the data or must extract the information manually.

As part of the pilot test respondent follow-up, we will request a breakout of one-time vs. recurring burden hours in order to better estimate the burden for the full scale rollout in FY 2010. It is expected that the recurring burden will decline as more institutions are able to set up automated systems to capture the new questions over the next several years.

b. Page 12, A12, what is the basis for the assumption that “the survey population will continue to grow by approximately 25 institutions per year?”

There has been an average increase of 22 institutions each year since the population criteria was revised in FY 2004. This average was simply rounded up to 25 for ease of use and to be overly cautious in developing the burden estimates.

c. Please show in the table on page 11 years beyond FY 2002.

We have not collected burden hours on the survey since FY 2002 because as noted in the supporting statement, the burden reported by respondents in FY 2002 included all of the current survey questions. We plan to collect updated burden hours on the pilot test of the redesigned survey in order to form an estimate for the FY 2010 and beyond surveys.

d. Why is no burden indicated for the screening phase described in B1? We would like to see this broken out as a separate item.

This has been added to Table A-12.2. The screening consists of a general email to a contact within the university accounting or sponsored programs office which asks if an institution is above or below the \$150,000 threshold in R&D spending for their previous fiscal year. A copy of the screening email is attached for reference. Because no precise data need to be extracted and reported to answer this question, we estimate the burden to be under one hour for each institution.

e. Please indicate the anticipated change in burden under the pilot by the types of institutions shown in Table A-12.1. Given the extraordinary increase in anticipated burden, please differentiate among the proposed additional sets of questions (i.e., the list of bullets on page 4, B4) by relative priority (and to who) as well as anticipated burden (i.e., which are the most/least burdensome to add).

Although it is not possible to break down estimates for each of these types of institutions (doctorate, master's, bachelor's) with any degree of validity, NSF anticipates the same relative burden distribution. The doctorate-granting institutions will likely continue to have the largest burden since they generally have many more R&D accounts to compile and report. Master's and bachelor's institution respondents will generally need less time to compile the survey data. The FFRDCS will not have any change in burden because they are not included in the pilot study.

Regarding the relative priority of the additional questions, NSF considers all of the questions to be high priority to certain key stakeholders. The list of additional questions, in rough priority order for NSF, is annotated below with information on the key stakeholders requesting this information as well as the relative burden based on what was learned from site visits:

- Total R&D expanded to include R&D expenditures in non-S&E fields as well as clinical trial expenditures:
Requested by survey respondents
Lowers existing burden (inclusion of non-S&E fields will be easier for respondents because they will no longer have to separate those accounts from their R&D totals; clinical trials are often already coded as R&D as well)
- Total R&D expenditures funded by nonprofit institutions:
Requested by National Science Board members and numerous data users throughout survey history
Low burden (code already exists on accounts at majority of institutions)
- Detail by field (both S&E and non-S&E) for R&D expenditures from each source of funding (federal, state/local, institution, industry, nonprofit, and other):
Requested by National Science Board members and data users throughout survey history
Low to medium burden (information is readily available based on site visits, but will require more data entry on survey)
- Test module on intellectual property and commercialization:
Requested by data users during redesign workshops, the Kauffman Foundation, and some items requested by National Science Board. These are key data elements to begin to gain a better understanding of innovation.
Medium burden (most of the questions ask for counts already collected by university technology transfer offices)
- Total R&D within interdisciplinary R&D centers:
Some kind of interdisciplinary R&D measurement requested by National Science Board members, as much of science is now conducted in interdisciplinary settings.
Low burden (question utilizes the location code on R&D accounts)
- Total amount of R&D expenditures funded from all types of foreign sources:
Requested by data users during redesign workshops
Low burden (code already exists on accounts at majority of institutions)
- Total R&D expenditures by direct cost categories (salaries, software, equipment, etc.):
Requested by BEA
Low burden (cost categories are existing account codes)
- Headcount of R&D personnel (principal investigators and other staff):
Requested by data users during redesign workshops and by NSF for international comparability
Medium-high burden (site visits and cognitive tests found varying degrees of sophistication in tracking R&D personnel at academic institutions)
- Counts of proposals submitted to funding organizations during fiscal year:
Requested by data users during redesign workshops
Low burden (sponsored programs offices already keep track of this metric)
- Counts and dollar amounts of R&D awards during fiscal year:
Requested by data users during redesign workshop
Low burden (sponsored programs offices already keep track of this metric)

- f. Please confirm whether the planned “actual burden hour reports” collected from pilot participants will allow separation of one-time versus recurring burden increases.**

As part of the pilot test debriefing, we will request a breakout of one-time vs. recurring burden hours in order to better estimate the burden for the full scale rollout in FY 2010.

7. Page 13, A14, What is the basis for the assumption of 3% inflation rate?

The basis for this assumed inflation rate was simply a rough average of the annual rate over the previous five years. However, if OMB prefers we use a different assumed inflation rate, we will revise the table accordingly.

8. Page 18, B4, will the expanded detail split out defense/non-defense work?

The survey has collected R&D expenditures by field and federal agency (including DOD) since FY 2003. The new survey will continue to collect this information. However, the survey does not include a question on defense versus non-defense work.

9. Where are all of the materials such as advance and cover letters? Only the Supporting Statement and questionnaires were submitted to OMB.

Copies of the advance and cover letters (which are email contacts) for the regular FY 2009 survey are attached. Draft versions of the advance and cover letters for the pilot are also attached.

10. Please provide the report of the 17 site visits approved by OMB in November 2008 and conducted from December 2008 to February 2009.

Attached.

11. Please provide more information about how the pilot sample will be identified and whether it is designed to be representative.

The pilot sample was selected using a systematic probability proportional to 2007 science and engineering (S&E) R&D expenditures from a frame constructed based on the 2007 Survey of R&D Expenditures at Universities and Colleges. Characteristics of universities that were taken into consideration in the design of the pilot sample were:

- type of ownership (public or private),
- presence of a medical school,
- minority serving institution
- status (Historically Black College or University or High Hispanic Enrollment),
- amount of 2007 S&E R&D expenditures (small, medium, or large),
- 2007 Non-S&E R&D expenditures, and
- simulated burden.

The purpose of simulated burden was to take into account a variable that may not be skewed or skewed as highly as R&D expenditures. The use of a probability sample design is intended to allow for the unbiased estimation of population characteristics such as R&D expenditures and their associated sample variances and be representative of R&D performing higher education institutions. While the sampling variance of the estimates is

expected to be high because of the small sample size (particularly for smaller subsets of universities and detailed questionnaire items), we still expect to perform statistical comparisons of the pilot sample results with the ongoing survey results.

12. Please clarify which of the proposed new content areas are directly responsive to BEA's expressed interests, as well as which of their requests are not met by the proposed redesign. Is it SRS's understanding that BEA requires an annual census to meet each of the identified information needs?

Below is the list of new content areas which BEA requested that will be covered by the proposed redesign:

- Cost detail for compensation, materials and supplies, overhead (indirect costs)
- Expenditures on R&D equipment and software
- Detail on type of funding transaction: contract, grant or pass-through
- University revenues from sale of R&D

Below is the list of new content areas that will NOT be included on the proposed redesign and explanations for why they are not included:

- R&D expenditures that are not separately budgeted
Institutions currently have no way of reliably tracking these intermingled departmental research funds, but NSF is continuing to investigate other methods for estimating this amount.
- Cost detail for depreciation on R&D capital
Due to the guidance of both respondents and a consultant with expertise in university accounting practices, the redesigned survey will continue to request expenditures rather than expenses (which would allow depreciation to be reported separately). However, depreciation is included in the associated indirect costs reported on the survey, but it is not able to be separately reported.
- Patent costs
NSF has decided to simplify the focus of the IP module to income received from intellectual property rather than the costs incurred to file patents and perform other R&D output-related tasks. This detail may be added in future years depending on the success of the currently proposed metrics on intellectual property. NSF is committed to the development of an intellectual property module, as an important element in better understanding innovation.
- Expenditures on R&D structures
It is not necessary to gather the information in this survey because it would duplicate information already collected in the NSF Survey of Science and Engineering Research Facilities (completion costs for new construction and repair/renovation of existing structures)
- Detail on type of transaction (contract, grant, pass-through) by type of entity (domestic/foreign)
During the site visit and cognitive test findings, the collection of the type of transaction (contract vs. grant) was found to be problematic for some institutions, but as noted above will be included on the pilot. Further breakouts for this question by foreign or domestic funding source was viewed by NSF to be unduly burdensome and also prone to measurement error. The pilot test will provide feedback on the availability and reliability of the codes for foreign funding as well as transaction type. If both categories are easily

reported NSF will consider adding the foreign breakout to the contracts vs. grants question.

- Expenditures used to create software
No codes currently exist within institutions to track this level of spending detail.

13. Please elaborate on the process by which SRS and its advisors will evaluate the pilot results.

On a microdata level, each pilot institution's response will be compared to its historical response(s) for items on the pilot survey that correspond to items on the ongoing survey. The comparison will determine whether the response is consistent with prior responses given a one-year differential in response. Items that are new to the pilot survey will be evaluated as being plausible assuming the other items as given. If a response to the pilot is found to be inconsistent, the pilot institution will be contacted and SRS will attempt to resolve the inconsistencies.

SRS will compute estimates of important items such as total S&E R&D expenditures and federal S&E R&D expenditures at the total U.S. level and by important subclasses such as public or private universities, medical schools, minority serving institutions, etc. We will also compute estimates of the sampling variability of the magnitude estimates. We will then make statistical comparisons of the estimated totals from the pilot with those from the ongoing survey to see if the pilot results are inline with the ongoing survey results.

Regarding methodological testing, respondent debriefings will be conducted by phone with each of the 40 participant institutions. The topics covered will assess the following:

- Success rates for retrieving requested survey information
- Ranking of survey items based on retrievability
- Ranking of survey items based on level of effort required
- Comparisons of level of effort required with effort actually used
- Assessment of respondent motivation and associated factors
- Perceived level of respondent frustration
- Respondent understanding of survey concepts and definitions
- Respondent satisfaction with final survey answers submitted
- Satisfaction with web survey edit messages
- Overall satisfaction with web survey
- Satisfaction with other survey tools (Excel spreadsheets, data printouts, etc.)

Help desk records will also be collected to document questions respondents asked during the field period. Similar records will also be recorded by the survey manager and project staff whenever respondent contacts occur.

Attachments:

Revised 83-I

Revised 83-I Supporting Statement

Kauffman Foundation Letter of Support (new Attachment 5 for Supporting Statement)

FY 2009 survey population screening email

FY 2009 survey advance and cover letters

Summary of Winter 2008-09 Cognitive Site Visits