

REPORT OF THE COMMITTEE OF VISITORS, 2008-2010

Surface Earth Processes Section (SEP) of the Earth Sciences Division (EAR)

INTRODUCTION

The Committee of Visitors (COV) for the Geobiology and Low Temperature Geochemistry (GG), Geomorphology and Land-Use Dynamics (GLD), Hydrological Sciences (HS) and Sedimentary Geology and Paleobiology (SGP) programs reviewed proposal e-jackets, analyzed program data, and spoke with Program Officers (POs) and administrators on June 6-8 2011 at NSF headquarters. The 2011 COV members were: George Hornberger, Chair (as a member of the Advisory Committee for the Geosciences Directorate, AC-GEO) (Vanderbilt University), David Hyndman (Michigan State University), Patricia Kelley (University of North Carolina at Wilmington), Ben Odhiambo Kisila (University of Mary Washington), Richard L. Reynolds (U.S. Geological Survey), Peter D. Roopnarine (California Academy of Sciences), Alan T. Stone (Johns Hopkins University), and Annette Summers Engel (Louisiana State University). The charge to the COV was to review actions taken by the Geobiology and Low Temperature Geochemistry (GG), Geomorphology and Land-Use Dynamics (GLD), Hydrological Sciences (HS), and Sedimentary Geology and Paleobiology (SGP) programs during the last three fiscal years (2008-2010) with a focus on (1) integrity and efficiency of proposal processes used to solicit, review, recommend, and document proposal action, considering intellectual merit, broader impact, transformative value; and (2) the relation between program portfolio and program goals with respect to a) the research supported, b) the integration of research and education, c) the balance in support of sub-disciplines, d) the award size and duration, e) the support of high risk-high reward research, f) the balance among single investigator, collaborative, and interdisciplinary proposals, g) the diversity of investigator and institution types and

underrepresented groups supported; and (3) relevance of supported research and program portfolio to national priorities and NSF strategic goals.

GENERAL FINDINGS AND RECOMMENDATIONS

In general terms, the COV was very favorably impressed with the performance of SEP personnel over the period covered by our charge and by the qualifications and dedication of the current Program Officers (POs) and staff. In terms of the main elements of our charge, we find (1) that the review process is handled very well, with proper consideration given to all aspects – intellectual merit, broader impacts, and transformative components of proposed research; and (2) that the program portfolio and program goals are well balanced and of high quality. Below, we will discuss observations, findings, and recommendations.

- (1) The COV was impressed by the ability of the POs to carry out their work in the face of an almost overwhelming proposal load. POs in SEP handle a large number of proposals within their core programs, they actively secure meaningful collaborations within EAR, across NSF, and with other agencies, and they serve key roles in proposal actions for cross-Division and cross-Foundation initiatives. All of these tasks are critically important and none can be sacrificed. Given staffing levels, however, there is a serious workload issue. This is not a new problem and we understand that some steps have been taken and ideas put forward about how to redress the most serious of the workload issues since the last COV reported. We also understand some of the constraints on the system in terms of acting to alleviate the problems. We have no ready solution to suggest, but we believe that it is our responsibility to point out once again that a breaking point must exist where overloads become so burdensome that work cannot be done effectively, even if no

serious consequences have occurred to date. We recommend that NSF continue to find ways to deal with the problem.

- (2) The two-tier system of review used in SEP – ad hoc mail review plus panel review – works well and we recommend that it be continued. We recognize that retaining the two-tier system constrains potential responses to deal with the workload issue (e.g., a no-deadline solicitation would be impractical). Furthermore, we think that having two calls for proposals per year is important, especially for early-career investigators and those new to the NSF world, and we recommend that this practice be continued.
- (3) Success of SEP POs at leveraging through co-funding within EAR, within GEO, across the Foundation, and with other agencies, is notable. We recommend that such initiatives continue. (Again, we recognize that this does nothing to address the workload issue.) We also recommend that POs across EAR be alert for possible collaborations between SEP and the Deep Earth Processes (DEP) section of EAR.
- (4) Awards made in the core programs, in general, continue to be of quite modest size and of relatively short duration (<3 years). We recognize that the desire to fund the highest quality, deserving proposals given limited budgets creates a Morton's fork for the programs. There are many more qualified proposals than there is available funding. We further realize that principal investigators (PIs) in the community may perceive and decide that the modest size and short duration projects have the best chance for funding, so they write proposals to fit into the size and duration restrictions. This may condition the programs to have portfolios with modest size and short duration applications. Nevertheless, as the previous COV pointed out, there is a limit, in terms of both size and duration, at which viability of a project is threatened. Based on the portfolio of the programs during this COV review period, this size-duration-viability threshold may have been

reached for some programs, especially those requiring severe cuts to awarded budgets. Although we do not have a ready solution to propose, we recommend that NSF POs carefully monitor the size-duration-viability issue within their programs in managing their portfolios.

- (5) Even with size and duration restrictions, success rates remain quite low for regular competition grants. We believe that the EAGER program is very important in terms of stimulating submission of proposals for high-risk research. EAGER grants should stimulate work on preliminary ideas, some of which will blossom and ultimately yield high rewards. It may be useful for SEP to track work that progresses from EAGER through other NSF funding to identify examples where such innovation occurs.
- (6) We were very impressed with the effectiveness of the collaboration among programs within SEP and encourage this to continue. We note, however, that proposals reviewed by several panels, especially when reviewed across the Foundation, appear to enjoy a lower success rate than those reviewed by a single panel within SEP. We could not determine whether this is an indication that inter- and multidisciplinary proposals are penalized in some way, but we recommend that the results be assessed. If bias against such proposals is suspected, then ways should be considered to overcome it. Some SEP programs are aware of this problem and try to assemble panels of an inter- and multidisciplinary nature. Also, for example, we were told that DEP recognized this apparent bias as a problem in the past and now arrange to have panels from two programs hold their meetings on overlapping days so panels can meet jointly to evaluate the inter- and multidisciplinary proposals. They evidently found that this practice has made a marked change in the panel dynamics for rating such projects.
- (7) Within the core programs, we noted that successful proposals have (statistically significantly) fewer ad hoc reviews than do declined proposals. Although the

differences may be argued to be relatively small in most cases, we note that for four programs in SEP for three years (12 observations in total), the data showed that successful proposals on average had fewer reviews than the proposals that were declined. We think that this bears examination by programs and we recommend that a more robust review and careful analysis be done. If there are no logical and consistent explanations for the differences observed, we recommend that programs strive to secure a uniform number of ad hoc reviews per proposal. This should strengthen the two-tiered system of review. We gather that the GLD program has had recent success in securing four reviews per proposal, so there may be lessons that can be shared with other programs that may prove useful. We think that the use of modern databases should assist in increasing the efficiency of getting reviews and note that if the NSF does not have appropriate software, they should upgrade. Review requests that ask for immediate response (accept, decline) may prove to be best practice.

- (8) Core grants tend to be on the order of \$100K/per year for 2 to 2.5 years, whereas grants through broad initiatives (e.g., the Critical Zone Observatories—CZOs) tend to be on the order of \$1M/per year for up to 5 years. We note that there appear to be few opportunities to secure funding for projects that may fall between these two types of grants. We think that SEP POs should be alert for opportunities that might require cross-program cooperation to fund such projects, if they find a need within the community.
- (9) The strategy of encouraging SEP investigators to participate in proposals to broad initiatives (e.g., Water Sustainability and Climate) is sound. By the nature of some of the recent initiatives, participation of investigators from some of the programs seems to be easier than for those from other programs. We think that accessibility of investigators from all SEP programs to broad initiatives should be a consideration in a balanced portfolio for EAR. For example, we are encouraged by

the broad call through the Frontiers of Earth System Dynamics (FESD) initiative, which clearly speaks to investigators across all of GEO.

(10) Issues raised in the report from the previous COV have generally been dealt with effectively. In particular, we note the following.

- a. Dwell time for proposals has been controlled. Exceedances of the six-month target are not egregious in general terms.
- b. Steps have been taken to clarify the use of the Broader Impacts criterion for reviewers and panelists.
- c. The provision of contiguous space for SEP has had a salutary effect on instilling a collaborative and collegial atmosphere for the programs housed on the sixth floor.
- d. Issues with proper reporting of ad hoc versus panel reviews still remain in some programs. (See specific comments section of this report.)

(11) Participation and funding of minority PIs appears to be relatively static across recent years. Some (as yet undiscovered?) new and innovative measures may be in order. We recognize that strong efforts have been made within SEP during the period covered by our charge and commend these. We recommend that aggressive measures continue to be taken and new ideas explored. As highlighted in O'Connell and Holmes (June 2011, *GSA Today* 21: 52-53), a continued multifaceted approach seems to be the best way forward. Summer research experiences and paired programs with minority serving institutions have proven effective and thus should continue to be encouraged. More attention should also continue to be on the overall geosciences pipeline (though this is a much broader issue).

(12) For the period of our review, there was a good balance between permanent POs and “rotators” – those on temporary appointments through the Intergovernmental

Personnel Act (IPA). The vitality and new ideas brought by rotators is impressive. We encourage SEP to maintain a balance of permanent and IPA POs.

(13) Participation in international conferences is very important for the POs to maintain connection with the community and to learn where the frontiers of the field are headed. We recommend that such participation continue to be valued and supported.

(14) We note that FY09 was anomalous within the three-year period covered by our review. It is noteworthy that ARRA (American Recovery and Reinvestment Act) funds provided significant stimulus in FY09, particularly through enhanced funding of postdoctoral fellows and early career awards, but also in many other ways such as funding high quality projects that would not have been awarded in more typical years because of budget limitations.

(15) We recommend that future COVs be provided with the following statistics as part of the reports from each core program.

- a. Success rates by type of proposal (e.g., RAPID, workshop, standard grant), as well as the lumped rates.
- b. Success rates of co-reviewed proposals.
- c. Geographic distribution of funded proposals.
- d. Resubmission success rates (when it is possible to identify such proposals).
- e. Co-funding and co-review details.
- f. Longitudinal data (for ~ a decade) for items in which gauging temporal progress is important, for example minority participation.

Specific Responses to the Charge to the COV

I. Questions about the quality and effectiveness of the program's use of merit review process.

Question 1. Are the review methods (for example ad hoc, panel, site visits) appropriate?

[GG] Both ad hoc and panel reviews are essential for a thorough process. For every proposal, a panelist writes a review. The consistency on this is useful. The number of ad hoc reviewers per proposal, however, varied from panel to panel. This program adequately and clearly handled the proposals, marked panel reviews so that they could be distinguished from ad hoc reviews in the proposal records. The Review Analysis documentation generated by the POs was commendable.

[GLD] GLD reviews were well handled, and the overall review practices were considered by the COV to be excellent, especially for the ad hoc review solicitation. Reviewers for GLD have been responsive and have taken their charge seriously, providing reliable, high value reviews. GLD also attempted to keep records of the gender and minority status of reviewers, but it was noted that minority status was difficult to track. If this record keeping is seen as important, perhaps reviewers could be prompted voluntarily to add gender and minority information.

[HS] The review process in HS was very thorough and effective. Despite an extremely high workload, the POs have done a remarkable job of getting ad hoc reviews, running effective panels, and providing input to the PIs about their proposals. The two-tier system of review is very effective for HS. The ad hoc reviews provide in-depth expertise in areas that may not be well represented by panel members. The panel then synthesizes these data and sorts out potential issues with ad hoc reviews. The overall

response (as reported in the Review Analysis) is generally more extensive for proposals likely to be funded or resubmitted with success than with those proposals that ranked low, which seems appropriate.

[SGP] The two-tier system of review was highly appropriate. Both methods were considered essential; ad hoc reviews provide comments from experts in the proposal topic area and panel members (who may not be experts in the proposal field) have a broader perspective of the entire competition, allowing interpretation of the ad hoc reviews and ranking of proposals. The COV noted that proposals that did not receive a full complement of ad hoc reviews, and for which panel members served as the ad hoc reviewers, were prone to be rated toward extremes (i.e., either reviewed more positively or more negatively) more so than the proposals that received a full suite of ad hoc reviews that were independent of panelists. The latter situation should be avoided if at all possible. In some cases, reasons for decisions that go against the recommendations of panel and reviewers were not explained in the e-jacket materials (e.g., Review Analysis); in such cases, a thorough documentation of the justification for the decision was needed.

Question 2. Are both merit criteria addressed

a) In individual reviews? b) In panel summaries? c) In Program Officer analyses?

[GG] Reviewers received directives to evaluate both merit criteria. Most reviewers (~80%) discussed both criteria in some detail. There were a small number of instances, however, where reviewers did not strictly adhere to the distinctions between Intellectual Merit and Broader Impacts. Also in rare instances, Broader Impacts was treated somewhat superficially, or reviewers used their own definition of Broader Impacts. In rare instances reviewer responses amounted to endorsements of proposals, rather than

thoughtful critiques. Fortunately, POs and panels identified reviews with shortcomings, and adjusted their conclusions accordingly. The POs noted that in the future problematic reviewers may be bypassed.

All panel summaries that we examined contained extensive written descriptions for both Intellectual Merit and Broader Impacts. Proposal analyses provided by the POs included posted numerical scores for each panelist for both merit criteria.

[GLD] In all sampled cases, both merit criteria were excellently addressed in ad hoc reviews, panel reviews, and panel summaries. The PO reviews were succinct, evaluated the different criteria, and addressed differences in reviews, all with excellent clarity.

[HS] Intellectual merit and broader impacts both play a significant role in the process, starting with comments by most ad hoc reviewers, continuing to the panel response, and the PO's comments to PIs. The COV extensively discussed the broader impacts criteria. In HS, there appeared to be an increasing importance given for BI by proposers and the review process. Broader impacts in the portfolio include a wide range of emphasis, from inclusion of K-12 education, to components that are relevant for policy makers.

[SGP] Based on the proposal materials evaluated, the BI criterion was not consistently applied by all ad hoc reviewers. Some reviewers ignored this criterion or, more commonly, misunderstood the criterion and thus misapplied it. Some comments under the BI criterion addressed aspects more appropriately considered as IM. Panel summaries did a better job than ad hoc reviews in appropriately applying the BI criterion. POs generally applied both criteria. Some ad hoc reviewers and panels were critical of proposals that included "traditional" BI elements, such as training of graduate

students. We note that, in this funding climate, it may be difficult to implement more extensive BI elements (e.g., museum exhibits, workshops), especially depending on the resources available to PIs from non-NSF sources. Perhaps a more appropriate criterion is whether the BI statement simply represents "lip service," or if any adequate plan for implementation exists.

It should be noted that at present (and not within the COV review period), SGP panelists are asked to give two scores for each proposal, one for each criterion. These are weighted 25% BI and 75% IM. Weighting ensures that both criteria are considered in the PO decision.

Question 3. Do the individual reviewers provide substantive comments to explain their assessment of proposals?

[GG] The POs have accrued a great deal of experience selecting reviewers and panelists, and have used their knowledge to good effect. There are, of course, some reviewers who return less insightful reviews.

[GLD] Nearly all of the ad hoc reviews for proposals examined by the COV contained substantive comments.

[HS] Reviewers in HS tend to provide extensive and substantial comments. Even the shorter ones tend to get right to the key points.

[SGP] On average, those reviewers who do respond to review requests are conscientious, although the numbers who do so are alarmingly low. Probably 80% of the reviews examined are substantive. However, POs need to be discerning in assessing the utility of each review. Non-substantive endorsements are not useful.

The program could provide instructions and more detailed guidelines to assist reviewers in writing more substantive reviews, such as asking reviewers to comment on such aspects as significance of proposed work, adequacy of proposed methods, etc.

Perhaps the GG panel template questions would be useful to reviewers.

Question 4. Do the panel summaries provide the rationale for the panel consensus (or reasons consensus was not reached)?

[GG] In the vast majority of instances, panel summaries were substantive and defended.

[GLD] The panel summaries provide thorough rationale for the recommended ranking.

[HS] In general panels provide adequate rationale for their decisions. In cases where the ad hoc reviews and the panel recommendations were at odds, the panel summary should be clearer in the reasoning. This was not always the case in the FY2008 summaries, but the process for the Fall 2009 competition has more extensive evaluation.

[SGP] Some inconsistencies among panel summaries were noted. Most are excellent and provide clear rationale, but some merely summarize the proposal without critiquing it. This problem was apparent mostly for one particular 'rotating' PO's panel summaries, which may have been due to any number of reasons, including the steep learning curve and overwhelming workload for that panel at that time.

Question 5. Does the documentation in the jacket provide the rationale for the award/decline decision?

[GG] For the GG program, the template for the review analysis is superb. Overall, analyses are carefully documented and consistent. Each analysis provides extensive

and specific documentation to the most relevant strengths and weaknesses within a proposal, regarding the investigators (including prior support), review quality, panel summaries, and the PO's own commentary and justification for funding decisions. Reviews that are unduly negative (i.e. unfounded, with insufficient documentation) or overly positive (i.e. endorsements without detailed explanation) are flagged and appropriately discussed.

[GLD] The PO comments in the review analysis have a clear, concise documentation of the rationale for the decisions.

[HS] Rationales for decisions are generally clear. Retaining the scores and text from the primary lead and readers in the summary would be helpful, which was not always done in the 2008/2009 jackets that were provided. The new PO is providing an extensive evaluation for each decision.

[SGP] Some inconsistencies in documentation of the decision were noted. The problem was exacerbated when more than one panel reviewed the proposal. Typically, one panel provided a very positive review, whereas the other panel was much more critical. (This situation may put interdisciplinary proposals at a disadvantage, as noted in our general comments.) Inadequacies in PO review analysis pertain primarily to one particular 'rotating' PO.

Question 6. Does the documentation to PI provide the rationale for the award/decline decision?

[GG] The rationale for each award/decline decision is carefully documented.

[GLD] PIs received clear and constructive comments in the "review analysis" with appropriate redactions.

[HS] Documentation to the PI generally included rationale, with content taken from ad-hoc and panel members, but more synthesis would be helpful than what was found in some documentation from earlier years.

[SGP] Yes, documentation of decisions generally is adequate.

7. Additional comments on the quality and effectiveness of the program's use of merit review process.

[GG and GLD] The GG and GLD POs did an excellent job of providing constructive criticism both to awarded and declined projects.

[HS] The HS program effectively used the merit review process. Additional efforts to delineate ad-hoc versus panel reviews would be helpful. Review summaries should carefully reconcile these where they are different.

[SGP] Co-reviewed trans- or interdisciplinary research proposals are being penalized by review by multiple disciplinary panels. The COV suggests that the program have reviews of these proposals by a panel comprised of people in cross-disciplinary areas.

II. Questions concerning the selection of reviewers.

1. Did the program make use of reviewers having appropriate expertise and/or qualifications?

[GG] It is clear that the POs expend a great deal of effort to solicit reviews from researchers with appropriate experience and qualifications. As can be expected, there were cases where only three ad hoc reviews were obtained. In these cases, the choice of ad hoc reviewers is very important. There were a few proposals where perhaps one of the ad hoc reviewers was outside the immediate subject area. When there are four

or five of these reviews, this is not much of a problem. Labeling reviews from panel members is useful because their expertise might not be tied as closely to that of the PIs' as are the ad hoc reviews. Instead, a review from a panel member, and the subsequent discussion, can be useful for placing the proposal's potential contributions into a broader context.

[GLD] Reviewers with a wide range of expertise were asked to provide comments. The reviewers provided insightful and thorough advice, with a few exceptions. The panel and PO clearly noted reviews that were deficient in content or detail.

[HS] The POs are getting reviews from experts in the field, but the number of reviews received can be quite variable from one proposal to another. The perceived issue of fairness of deciding to fund or decline proposals with a small number of reviews versus those with a high number of reviews should be addressed. A few cases were identified where the expertise was likely too focused on one aspect of the proposal and the panel may not have fully considered the proposal as a whole.

[SGP] We observed a disparity in handling the review process between POs, with a tendency for one of the POs to select reviewers with inappropriate expertise.

Reviewers without the appropriate expertise in some cases appeared to rubber-stamp proposals or to be inappropriately critical of proposal details. In an unusually large number of cases, panelists were pressed into service as ad hoc reviewers, sometimes having an inordinate effect upon the review decision despite lacking the requisite expertise. This situation was most severe for proposals reviewed by the Fall 2009 panel. Only four ad hoc reviews were requested for jackets reviewed by this panel, with an average return of 1.4 reviews per proposal. In some cases, decisions were made with no ad hoc reviews at all. We find this practice detrimental to the decision-making

process. It appears that some POs may need help in tracking down expertise within the large review community.

2. Did the program recognize conflicts of interest when appropriate?

[GG] Clearly the POs are very careful in this area, much more so than in many funding agencies, and other programs within NSF. In the proposal set that the COV examined, there were one or two gray area examples, ad hoc reviewers who might not be intellectually independent of each other. A good practice would be to avoid soliciting reviews for the same proposal from people who are spouses or close collaborators.

[GLD and HS] NSF is very careful about resolving potential COIs. Any COIs were carefully noted and panelists left the room as per the NSF-wide policy. POs are extremely sensitive to COIs. They have done a highly effective job and are proactive to identify COIs. We did not recognize any COIs at any step in the review process.

[SGP] Among the selection of proposals reviewed, we found rare instances of COI inconsistencies; for instance, the COV noted instances where a collaborator or postdoctoral advisor's review was not recognized as a COI, even though the post-doctoral relationship was mentioned as a COI by the reviewer and the PI listed this collaborator as a COI.

Additional comments on the review process.

Obtaining ad hoc reviews is an important but time-consuming process. COV members wondered if there is a way to streamline the process and simultaneously obtain a more consistent number of ad hoc reviews. In some cases, decisions were made based on a small number of ad hoc reviews, with panelists serving as additional reviewers (and also contributing to the panel score - a case of "double-dipping"). In other cases, proposals received a much larger number of reviews than "normal" (sometimes including two

panel summaries). This generally resulted in a lower ranking. Possible ways of streamlining the reviewer selection process might include having the PO vet an initial list produced by an assistant staffer (e.g., program intern?), or developing a system in which fewer reviewers need to be selected (e.g., have an online system similar to that used by many journals, with an electronic invitation that a potential reviewer can either decline or accept; tracking of reviewer past performance and number of reviews requested; in cases of failure to respond or of decline to review, then additional reviewers could be requested).

III. Questions concerning the management of the program under review.

1. Management of the program.

[GG] The POs in GG are obviously very careful and strategic in making decisions about which proposals to fund at what funding level, at identifying other cooperative sources of funding, and leveraging Section funds with these other funding sources. Within the current funding climate, the fact that the mortgage in GG has been kept below 25 % for the COV years being evaluated reflects good management, and ensures that funds will be available when opportunities arise in the future. Moreover, the number of awards that have received <20%, and even 0%, budget reductions (see comments below) is also testimony to the strong commitment by GG to attempt to provide sufficient funds for proposed work.

[GLD] The POs have done an excellent job of managing the program. The COV notes that the new PO is continuing this good practice.

[HS] The POs are doing an excellent job of reviewing a huge number of proposals. Management rose to the challenge, especially in 2010 when the proposal load doubled from 2009. Going forward, with a growing pool of proposals, the HS POs should be

given some relief, especially with the significant amount of effort involved with getting the WSC program up and running.

[SGP] Program management was uneven for the period of our review. Coupled with the significant under-funding of this program, problematic past management practices by one particular 'rotating' PO have resulted in a heavily mortgaged program, reversing several years of a declining mortgage. Current POs are struggling to decrease the mortgage by issuing few continuing grants, funding fewer proposals of shorter duration, and making significant cuts to new award budgets. This obviously has an impact on the community, but also taxes the POs because they have to discuss frustrating funding outcomes for highly competitive proposals with PIs.

2. Responsiveness of the program to emerging research and education opportunities.

[GG] GG has quite successfully fostered and supported rapidly evolving lines of inquiry. The program has done a good job keeping abreast of new developments and new opportunities through sponsoring a number of workshops, attending topical conferences, etc.

[GLD] The program is ahead-of-the-curve in terms of the portfolio. The POs have linked, with vision and effectiveness, with emerging interdisciplinary activities and programs such as NCED, Frontiers in Earth System Dynamics, and CSDMS. The Program has provided a balanced mix of SGER and RAPID awards. The program is in an outstanding position to build on the research priorities in interacting landscapes and climate, including coevolution of ecosystems and landscapes, reconstruction of landscape dynamics across time, and future landscapes. Educational opportunities are well considered in the program. GLD supported or helped support 10 workshops during

FY08-FY10 that covered major issues of broad interest to the Earth-surface-processes and other Earth-Sciences communities.

[HS] The HS program has aggressively responded to emerging opportunities, and connects very well with other programs. The very nature of hydrological sciences involves extensive collaboration and cooperation, which has defined the culture of the program. The previous PO was exceptional in getting cooperation with others, and the current POs are continuing and extending this capacity.

[SGP] The current POs are providing strong leadership to the research community regarding emerging research and educational opportunities. For many years, the SGP permanent PO has attempted to mobilize the community and organize researchers to act like a community rather than solely as individuals. These efforts have included support of workshops, town hall meetings, recent NRC reports on Understanding Earth's Deep Past: Lessons for Our Climate Future, Understanding Climate's Influence on Human Evolution, etc. As a result, some promising initiatives are under development (e.g., DETELON).

3. Program planning and prioritization process (internal and external) that guided the development of the program.

[GG] Overall, there is a sense that GG POs have been very resourceful. Nothing short of heroic efforts on the part of the permanent PO and rotating POs have gone into leveraging GG funds and finding joint funding sources for individual projects, including ushering proposals through co-panel/program reviews and consideration, as well as shopping proposals through other agencies. This has expanded what GG has been able to accomplish, predominately within FY09 and FY10. The program also strategized how some proposals could be awarded at different times of the year, even alerting PIs

SEP COV REPORT – JUNE 2011

to this issue (i.e. Fall panel success rates are lower than Spring success rates). Because there is a tradeoff between the number of proposals that can be funded and the funding level that each proposal can receive, the POs have also kept sight on realistic funding levels that are needed to conduct the proposed research so that draconian cuts are not required. It is clear that the POs try to fund as much innovative science as possible and to maintain a diverse and balanced portfolio that represents the GG community.

[GLD] The current strength of the GLD portfolio reflects a well-conceived and implemented process of program planning. We see the results of past planning and observe excellence in current planning.

[HS] The Eagleson Blue Book report (NRC 1991, Opportunities in the Hydrologic Sciences, Nat. Acad. Press) was very helpful as an early guide for the HS program. Workshops, including those promoted by CUAHSI, have promoted collaboration in the community. POs are doing an excellent job of promoting collaboration with other scientific communities.

[SGP] The COV noted problems in portfolio development. The current POs expressed concern about the unbalanced state of the SGP portfolio within the paleobiology discipline. The problem is most apparent for FY 2010 awards, in which proposals in the vertebrate paleontology subdiscipline were funded significantly out of proportion to its representation in the discipline and the proposal pool, with corresponding declines in funding rates of other subdisciplines. The COV examined the list of proposals taken to Fall 2009 panel. At this panel, success rates were 53.3% for vertebrate paleontology proposals and 17.6% for invertebrate paleontology, a statistically significant difference (Fisher's Exact test, $p = 0.01$). The disciplinary imbalances in the program and problems with program management have made a historically underfunded program even more

difficult to manage. The current SGP POs inherited this problem and are working valiantly to address imbalances and deal with the budgetary constraints of a heavily mortgaged program.

4. Responsiveness of program to previous COV comments and recommendations.

[GG] Following upon the comments from the previous COV report, GG has effectively employed two rotators. They have been especially thorough in record keeping and written explanations of proposal decisions. Dwell times do not appear to be a significant issue, and most proposals are receiving three or more ad hoc reviews. GG decisions appear to follow from ad hoc reviews and panel deliberations, and not relying unduly on prior PI track record. The previous COV reported noted that funds awarded to each grant were too low, lower than for the other three sections. During the past three years, funds awarded to each grant only rose when ARRA funds were available.

[GLD] NSF responded as much as possible to the previous COV recommendations regarding the GLD Program. The program now has a permanent PO, and, to every possible extent, GLD is responding to needs and opportunities to advance science in areas of geomorphology and land-use dynamics, with strong linkages to related topics.

[HS] The new PO was hired after the retirement of the previous PO. The new PO should be commended for rapidly coming up to speed and moving the community forward. The previous COV recommended that hydrologic sciences be given more help, a recommendation that remains appropriate today.

[SGP] The program has responded to previous COV comments. Questions about consideration of BI led to the current approach to a formal 25% weighting of panel BI scores. The program had been criticized by the 2005-07 COV for “panel instability,” noting the service of some panelists for as little as one panel without justification.

During the currently reviewed COV period, the POs provided sound reasons for this practice, including “trying out” panelists before longer panel terms, helping new PIs and underrepresented PIs become familiar with the review process, etc.

IV. Questions about portfolio

1. Does the program portfolio have an appropriate balance of awards across disciplines and sub disciplines of the activity?

[GG] The GG program requires that proposals have some type of geochemical context or contribution. This is clearly described in the 6-year-old program description and RFP, and distinguishes this program from purely biological or ecological sections within NSF or other agencies. Generally, within the geochemistry constraint imposed by the GG program, its portfolio exhibits remarkable diversity. Regarding funded projects during the COV review period, some subject areas are represented more than others (e.g., inorganic geochemistry, general biogeochemistry, and geomicrobiology). This is predominately due to larger numbers of submitted proposals in these areas. But, for any given COV year evaluated, projects were awarded that spanned a range of topics. Efforts to increase funding in emerging areas such as environmental genomics are being explored. The POs discussed ways to encourage more submissions in these emerging areas, including revising the current RFP, sponsoring workshops, posting announcements, etc.

[GLD] The portfolio is well diversified, and it includes the funding of emerging areas.

[HS] The PO's have done an excellent job of managing a program with great diversity.

[SGP] We observed a growing imbalance of the portfolio during the period covered in our review. Current awards do not reflect the diversity of submissions, as mentioned earlier.

2. Are awards appropriate in size and duration for the scope of the projects?

[GG] The POs are creatively working within overall SEP and program budget constraints. In general, award amounts are small (FY08-10 average, \$113K) and awards are short in duration (FY08-10 average, 2.5 years). Evidence of budget constraints is reflected by the percentage of budget reductions within the program for funded projects (Table 1). During the COV review period, 43% of the budgets for funded projects in FY08 had to be reduced by more than 20%, but considerably less in budgets for FY09 (with and without ARRA). For FY10, almost half of budgets had some reduction. These data could also indicate the PIs submitting proposals are requesting realistic budgets to conduct their research.

Table 1: Percentage of funded projects without and with budget reductions for GG.

	FY09			
	FY 08	FY 09	w/out ARRA	FY10
0%	39	68	85	51
up to 20%	17	5	8	37
> 20%	43	25	8	12

[GLD] The funding amounts are low, as in other SEP areas, and the duration of proposed research is commonly very short relative to the environmental forcings that are being examined in some projects. In general, award amounts are small (FY08-10

average, \$110K) and awards are short in duration (FY08-10 average, 2.5 years). GLD POs indicated that they have a goal of moving the duration closer to three years for regular grants. The three-year duration, while welcome, could also fall short of the needs of some projects engaged in measuring environmental changes.

[HS] The level of funding for core HS proposals is generally modest (FY08-10 average, \$120K), the durations relatively short (FY08-10 average, 2.7 years), and the success rate despite very high quality proposals very low. The low rate of success was especially true in 2010 when the number of proposals spiked. PIs probably know, from discussions with POs and reviewing funding portfolios for the program, that shorter duration, less expensive proposals are more likely to be funded. But, science and education (e.g., training PhD students) would greatly benefit from longer term proposals with a higher rate of funding. There is a limit to small size and short duration of a grant to be effective. Furthermore, PIs have to spend too much time writing proposals with these levels of funding.

[SGP] This program is dramatically under-funded. In general, award amounts are small (FY08-10 average, \$158K) and awards are short in duration (FY08-10 average, 2.2 years). This situation calls into question how SGP research that requires infrastructure (e.g., continental drilling, geochronology, paleoclimate, geoinformatics) could be supported. Perhaps this factor has had an effect within the community because there are generally low numbers of proposals submitted that fall into these categories. Moreover, the limitation in funding and funding duration make student, post-doc, and research staff support problematic. Even when proposals had co-funding, budgets were cut, sometimes drastically. Data clearly show this to be an acute problem during the COV period (Table 2).

Table 2: Number of proposals with budget reductions for COV period by panel. Spring 09 panel excludes ARRA awards.

Number of awarded proposals falling into budget reduction categories (by percentage)						
	Total awards	<10%	10-20%	20-50%	>50%	Total awards with >20% reduction
F07	30	13	8	5	4	30.0%
S08	18	8	8	1	1	11.1%
F08	21	13	5	3	0	14.3%
S09	6	2	3	0	1	16.7%
F09	32	4	11	14	3	53.1%
S10	13	5	3	5	0	38.5%

3. Does the program portfolio include awards for projects that are innovative or potentially transformative? Does the program support shifts in existing fields?

[GG] The number of proposals submitted has increased by ~10% over the full COV period for this program, with more proposals being submitted in the Fall than in the Spring submission rounds. During any year, there are 120 to 140 active proposals, and ~14 continuing grants.

This program has done a very consistent job of building the foundations of important new lines of inquiry. This is not to say that work has not been imaginative or transformative. Funds devoted to methods development and provocative ideas in environmental genomics are starting to pay off. Linkages between biological activity and important geochemical phenomena have been identified. From FY09 and FY10 funded proposals, innovative and transformative projects were identified. About 25-45%

of the projects were identified by the POs as being novel and potentially transformative (including ARRA-supported projects but excluding EAGERs, SGERs, RAPIDs, and Workshops). This is an exciting trend and demonstrates the inherent nature of the program to include interdisciplinary research (e.g., Geo-biology). Seven ‘risky’ proposals were identified by the GG POs in FY10.f 7 when this became part of the reporting process.

[GLD] The GLD program has made awards that, taken together, are balanced and innovative. One example is the science community-driven needs for research in the emerging area of ecogeomorphology. Another example is GLD support and commitment to the Community Surface Dynamics Modeling System (CSDMS) that continues to gain traction and is yielding important results.

[HS] The research in the portfolio includes significant transformative elements, and has been supportive of significant shifts in the field with emerging topics being supported including hydroecology and hydrogeophysics.

[SGP] It is very difficult to do transformative science cheaply. Because of limitations of funds and award duration imposed by the limited budget of the program, POs indicated a trend for proposals being more scientifically conservative during the COV period compared to in the past. Nonetheless, some shifts are occurring, for instance in the funding of rare, innovative projects involving drilling, geochronology, stratigraphy, paleoecology, and early life.

4. Does the program portfolio include inter- and multi-disciplinary projects?

[GG] The GG section is appreciative of investigators who have personally invested effort in building bridges and establishing linkages between disciplines. The program has also been effective at supporting imaginative and productive collaborations among

investigators with different skill sets. The number of PO-identified ‘interdisciplinary’ projects that were funded has increased during the COV evaluation period.

The COV observed that these interdisciplinary projects have a complicated review process, with one to two other panel reviews per proposal (in rare instances, three panels reviewed proposals). During the COV period, the POs sought co-reviews for interdisciplinary projects at differing rates (e.g., 13 co-reviewed proposals for Fall ‘07 submissions versus 6 co-reviewed proposals for Spring ‘10). The success rates for these proposals was low, with 0% success rate for FY 09 proposals for Fall and Spring panels, to as many as 33% success rate for Spring ‘10 proposals. For other reviewed panels, the rates were 16-25%.

[GLD] The GLD Program portfolio contains a large number of interdisciplinary and multidisciplinary projects. Answering important questions in Earth-surface dynamics is commonly based on concepts that must account for interactions among atmosphere, biosphere, geosphere, hydrosphere, and human influences. This kind of approach then requires blending traditional and innovative field, chronologic, and other laboratory methods, imaging, and modeling. These broad approaches are well represented in GLD-funded projects.

[HS] The HS portfolio is interdisciplinary, and the level of interdisciplinary has increased with time based on the materials reviewed by the COV. Proposals in areas related to climate change, ecohydrology, and biogeochemistry, for example, represent a significant portion of the portfolio.

[SGP] Inter- and multi-disciplinary projects are indicated by co-funding with other programs (paleoclimate, BIO Venture, EHR, OISE, DEB, DEM), but in diminishing numbers. In 2007-2008, there were eight such co-funded proposals, but in 2009-2010 only three.

5. Does the program portfolio have an appropriate geographical distribution of principal investigator?

Data provided show an appropriate distribution of proposals received and reviewers solicited. Data were inadequate to consider the geographic distribution of funded PIs.

6. Does the program portfolio have an appropriate balance of awards to different types of institutions?

[GG] Research-intensive institutions have been more aggressive in hiring investigators with interests in biogeochemistry. The success of these investigators in receiving support is reflected in the portfolio. Based on the documentation seen by the COV for this three year period, EPSCoR funding rates are lower than the overall average. It is difficult to draw conclusions from this observation, given the low numbers of proposals received from EPSCoR states (some EPSCoR states had no submissions to this program during the COV review period). Some universities have been slower than others at moving into the geobiological aspects of the program. When proposals have not been successful, POs have provided summary comments to PIs that will point the way to improved proposals with a better chance of success. For EPSCoR states with no submissions, however, some effort to reach out to potential PIs may need to be considered.

[GLD, HS, and SGP] The statistics provided indicate that the balance of awards to different types of institutions is being appropriately considered by POs. Awards to research intensive, PhD-granting institutions far outnumber those to other types of institutions, but this is not unexpected based on rates of submission.

7. Does the program portfolio have an appropriate balance of awards to new investigators?

[GG] There is a 28 % success rate for proposals submitted by new PIs. This is a good percentage, which is similar to the funding rate for all proposals submitted to the program. The funding rate for new PI's has been similar to the overall rate for each of the last 3 years (Table 3). It also reflects efforts to introduce new PIs to the system (e.g., service on panels) and constructive summary comments to PIs about making improvements to proposals from the panel summary and PO Review Analysis notes.

[GLD] The success rate for first-time investigators has been good, but was a bit low in 2009 (Table 3). The new PO shows commitment to new PI success rates.

[HS] The funding rate was low for new PI's in 2008. The success rate for first time investigators has been reasonable in the last two years, and is being carefully considered by the new PO (Table 3).

Table 3: Success rate percentages for overall versus New Investigator awards.

		2008	2009	2010
	Total	20	37	27
GG	New PI	23	42	25
	Total	13	32	24
GLD	New PI	14	22	20
	Total	26	39	14
HS	New PI	16	32	10
	Total	23	31	21
SGP	New PI	16	27	8

[SGP] Funding rates for new PIs in 2010 were inordinately low, but in the other two years were not vastly different from success rates for the overall pool (Table 3).

8. Does the program portfolio include projects that integrate research and education?

[GG] CAREER proposals have strong commitments in this area. It is clear that, in order to receive funding, CAREER proposals must have a carefully planned and presented Broader Impacts section.

[GLD] The CAREER grants in the portfolio combine well integrated aspects of education and research. Graduate and undergrad education components are well integrated into the broader research portfolio.

[HS] The CAREER grants in the program closely integrate education and research. Graduate and undergraduate education are also well incorporated in the broader research portfolio.

[SGP] The portfolio includes awards with well integrated research and education components. POs are committed to funding proposals with well integrated BI, as indicated by the weight assigned panelist BI scores. The Panama Canal PIRE is an example where research and education are integrated.

9. Does the program prepare and engage a diverse STEM workforce motivated to participate at the science frontiers?

Overall the SEP POs, panels, and PIs are committed to preparing and engaging a diverse STEM workforce. The data summarized below for HS and SGP indicate the levels of funding achieved in SEP.

[HS] The HS program has similar success rates for many of the diversity categories, in comparison with the total success rates. Percentages for women appear to have improved to roughly the same as the full pool of proposals.

Table 4a: Percentage of success rates for different groups in HS.

	2008	2009	2010	total
Total	26%	39%	14%	24%
Women	18%	52%	13%	24%
Minority Serving Institutions	7%	40%	10%	16%
EPSCoR	24%	29%	18%	22%
Minority-PI	27%	31%	9%	17%

[SGP] An indication of the program’s commitment to preparing a diverse workforce is the extent to which awards are made to minority-serving institutions (MSI) and EPSCoR institutions (Table 4b). Success rates for MSI vary from year to year, but overall the rate exceeds that of the general pool. EPSCoR rates are close to that for the general pool.

Table 4b: Percentage of success rates for different groups in SGP.

	2008	2009	2010	Total
Total	22.6	30.5	21.1	24.0
Women	22.4	32.7	16.8	22.7
Minorities	8.3	12.5	18.8	13.9
MSI	37.5	36.4	12.5	29.7
EPSCoR	22.4	23.4	19.3	21.3

10. Does the program portfolio have appropriate participation of underrepresented groups?

Success rates for proposals involving women are comparable to those for the overall pool (e.g., Table 4). The number of minority PI's is small, and their success rates tend to be lower than those for the overall pool, although the POs make efforts to encourage submissions by including minority panelists, seeking minority reviews, etc.

11. Is the program relevant to national priorities, agency mission, relevant fields and other constituent needs? Include citations of relevant external reports.

[GG] The program is responsive in this area, and looks for ways to be both scientifically grounded, innovative, and progressive. It should be noted that research funding at other agencies are in flux (e.g., EPA, USGS), with NSF being the remaining funding source in subjects critical to the nation. For example, EPA funding of hydrology and low temperature geochemistry used to be significant, but is barely discernible now.

[GLD] The GLD Program is clearly relevant in all of these ways. The report, Landscapes on the Edge, is one outstanding example of the linkage between GLD and national priorities that further encompass relevant scientific research directions, as well as the needs of education and public communities. In the overall GLD portfolio, there is solid representation in "Hazards" and "Human" Categories, as well as in other topics (such as "Coastal") that address societal challenges.

[HS] The HS program is highly relevant to national priorities especially as we look forward to altered responses due to changes in climate and land use. As human populations grow, demands for sustainable clean water supplies grow. Competition for these supplies is exacerbated due to the need to feed the growing population and provide energy with biofuels.

[SGP] One indication of SGP relevance is their sponsorship of the recent NRC reports on Understanding Climate's Influence on Human Evolution and Understanding Earths' Deep Past: Lessons for our Climate Future. The program is also encouraging the community to come together to address national priorities such as biodiversity conservation and sustainability through such activities as the DETELON initiative and the Conservation Paleobiology workshop. Another example is the Bighorn Basin Coring Project, which seeks to understand the relationship between climate and biotic and geological change during a time of global greenhouse conditions.

12. Does the program enhance research infrastructure and promote data access to support researcher's and educator's capabilities?

[GG] CZOs are examples of this, which has been defining for the program, and a support structure for future achievements. Bringing minority institutions and investigators into this enterprise could be very beneficial.

[GLD] CSDMS, CZO program support, and NCED are examples where the program has enhanced infrastructure for the community.

[HS] Examples where the program has enhanced infrastructure include CUAHSI (including the "spin-offs" of the Hydrologic Information System, the Hydrologic Measurement Facility, and two Synthesis grants) and the CZOs.

[SGP] Support for research infrastructure and data access is exemplified by Paleobiology Database, Macrostrat, GeoStratSys, NEOTOMA, Morphobank and other database funding; workshops on such topics as Continental Drilling, DETELON, Variable Atmospheric Laboratory; Deep Time NCAR Paleoclimate Liaison; National Center for Earth Surface Dynamics. Support for student attendance at meetings (including 3 international meetings) supports educators' capabilities.

13. Does the program innovate for society? Does it make investments that lead to results and resources that are useful to society and build the capacity of the nation's citizenry for addressing societal challenges through science and engineering?

Biological processes are being incorporated into geosciences research, enhancing the capabilities of the Directorate; the effects of climatic variability and land uses on landscapes and human communities are being addressed; critical issues at the interface between humans and the environment form a significant part of the SEP portfolio, including for example, responding to the Deepwater Horizon oil spill.

14 Additional comments on the quality of the projects or the balance of the portfolio:

Funding success rates across SEP generally fall in the range of 20-25% considering all proposals, including those for workshops, RAPID, and EAGER grants. Success rates for the “standard” program competitions are lower yet. These rates are quite low for an incredible quality of research proposals, many of which have clear, direct, and critical relevance to society. As we note, the programs within SEP attract proposals from a very broad spectrum, including many that are multi- or interdisciplinary in nature. The management of programs that have modest success rates at best and a diverse array of high quality proposals was a challenge during the period covered by this COV. We anticipate that the challenges will remain substantial for the programs in the future as well.



For the SEP-EAR COV
George M. Hornberger
Chair