Fellows' Report on Metrics for Scientific Excellence at Los Alamos
May 7, 2006
LA-UR-06-3393

Summary

Maintaining the Laboratory's credibility in meeting its national security mission demands metrics for scientific excellence, and strong incentives to support them. Our analysis of metrics led us to four principles:

1. Metrics should adapt the government-wide COSEPUP criteria of quality, relevance, and leadership to the complex mission and programs of our national laboratory.

2. Quality, relevance, and leadership should be evaluated via objective metrics such as publications, citations, and honors wherever possible.

3. Because objective metrics alone may give a distorted view of our scientific process and our ability to solve complex, interdisciplinary, national security problems, we also require expert review, focused on (a) scientific process, (b) intellectual leadership, and (c) outcomes. A sound scientific process will include a healthy interaction between theory, simulation, and experiment.

4. The integrity and independence of expert review is critical and must be improved.

We propose a collaboration between LANS managers and senior scientific staff to develop and implement metrics for scientific excellence.

Introduction

The Los Alamos Fellows are committed to excellence in science and technology. Excellence and technical credibility have been at the heart of the Laboratory's identity, making our national security mission possible. We are heartened by the dedication to excellence expressed by the LANS leadership team, and want to help them identify appropriate tools to manage for excellence. Metrics are important because, when evaluated carefully and meaningfully, they change behavior at all levels. Metrics are the tools that allow the Director to require accountability from management and staff. It is an axiom of management that one can manage what can be measured. Although quantifying excellence can be elusive, metrics of excellence are essential.

The Los Alamos Fellows have therefore assembled a task force to recommend metrics for scientific excellence, pursuant to a commission from Laboratory Director Robert Kuckuck dated March 13, 2006 (Appendix 1). Membership of the task force and the schedule of their meetings with LANL and LANS leadership are listed in Appendix 2.

To be consistent with the overriding theme of cross-laboratory integration set out by LANS, the evaluation of science and technology excellence should be ubiquitous, applying to all organizations with a science and technology mission. There should be no special circumstances that preclude the evaluation of scientific excellence. Metrics *cannot* be uniform, but must be tailored to an organization, because no single set of metrics could be appropriate for the breadth of the Laboratory, with its range of fundamental and applied, classified and unclassified R&D. Our recommendations are a toolkit from which tools can be drawn as appropriate.

What are metrics trying to measure?

If metrics could be boiled down to a single word, the word would be "impact". The Laboratory's value depends on how it makes a difference to the nation, the world, and science. Impact has so many facets that it must be broken into tractable components.

In 1999, the National Academies commissioned a Committee on Science, Engineering, and Public Policy (COSEPUP) to determine criteria for excellence in federally funded research.[1] These criteria have been adapted for the Laboratory Directed Research and Development program (LDRD), and are generally applicable to science and technology. These criteria are quality, relevance, and leadership, which COSEPUP recommends be evaluated by expert review.[2] They translate into these particulars for Los Alamos:

[1] http://www.nap.edu/catalog/10106.html

[2] COSEPUP defines expert review as "more than traditional peer review by scholars in the field. It also includes the users of the research, whether they are in industry, nongovernment organizations, or public health organizations or are other members of the public who can evaluate the relevance of the research to agency goals."

Quality of science and technology. In the context of peers nationally and internationally, the quality of R&D should be at the upper end, consistent with the purview of a National Laboratory.

Relevance to the mission and objectives of the Laboratory. Relevance may include connections between the science of a division and its mission, the flow of ideas from fundamental research to applications in national security programs, and integration rather than duplication across the Laboratory.

Leadership. An organization or person in a position of leadership is looked to for advice and direction by peers and sponsors. Leadership results in changes in R&D directions that range far beyond the Laboratory. Leadership and mentoring within the Laboratory nurtures our science and technology and makes it possible for the institution to be a leader.

The Task Force recommends that quality, relevance, and leadership be evaluated by both objective measures of activity, which can be evaluated without much interpretation, and by analysis of scientific process and outcomes, which requires the Laboratory to establish an unbiased process of expert review. Both processes are necessary, the first because of its objectivity, the second for its deeper look and predictive value.

Objective metrics of activity

In priority order, we find value in the following objective metrics:

1. Peer-reviewed publications and citations. These are clearly the most important metric of good scientific work. They train the staff in rigor and teach them to defend their work. Citations demonstrate that our work is in the main stream of scientific discourse. Prestigious publications (*e.g.* Science or Nature) should be given particular weight. The criteria for acceptable publication activity will vary from field to field, and from program to program.
2. Objective metrics for classified research need to be developed. Possible metrics include publication in Defense Research Review and intelligence community publications, archival classified reports, participation in classified conferences and seminars, and peer review reports from external committees such as the UC S&T panel and the JASONS.
3. Awards, such as professional society awards, National Academies memberships, Lawrence Awards, and awards at the pinnacles of scientific achievement.
4. Success in recruiting, measured by the fraction of postdoctoral fellows and strategic hires that accept our offers; and their subsequent success.[3]
5. Program development success in peer-reviewed venues.[4]
6. Journal editorships, with weight given to more prestigious journals.

[3] Measured by the prestige of their subsequent positions and their scientific and technical accomplishments.

[4] Both in fundamental research venues that feature a peer-reviewed proposal process (*e.g*. NASA, DOE/OS, and NIH) and in peer-reviewed venues for applied R&D such as the National Reconnaissance Office's Director's Innovation Initiative, DARPA, DOE/NN, DNDO, etc.

7. Intellectual atmosphere, measured by factors such as regular colloquia and seminars, and scientific/technical visitors, especially those on extended visits for interactions with staff.
8. Scientific and technical society fellowships and officerships.
9. Patents, patent disclosures, and licenses, although we note that the absolute rates of these, Laboratory-wide, are as dependent on our patent lawyers as on our scientists.
10. Scientific meeting participation, with greater weight given to conference and session organization and invited papers.
11. Service on external advisory panels and committees.

These metrics should be applied differently for different organizations. For example, publication-associated criteria should be different in an unclassified research environment like T division than in an organization closely associated with classified programs like X division. Nonetheless, there are unclassified aspects of applied programs that can be published, *e.g*. methodologies like instrumentation and computational advances. Encouragement of publication will build the reputation of staff across the Laboratory. Integration across the Laboratory is demonstrated when applied research as well as basic research produces peer-reviewed publications. We note also that most of these metrics are secondary, having to do with someone's opinion of our process or output – a reviewer or referee – even if calculating the metric is objective. However, these review and refereeing processes are visibly independent of the Laboratory.

Objective metrics require central tracking at a Laboratory-wide level. For example, we commend the recently announced AWARDS database. For useful, uniform, and timely tracking of publications and citations, the Laboratory should maintain a central database, executed by some combination of S-7 (who approve publications for external release) and the Technical Library (who catalog and disseminate results). We do not see the possibility of uniform metrics if they are managed on a group-by-group basis. Similarly, metrics on recruiting success and post-LANL career accomplishments could be managed by HR. Our objective metrics cannot be evaluated without the appropriate data.

Expert review of scientific process and outcomes

Expert review should focus on [three] aspects: (1) scientific process, (2) intellectual leadership, and (3) outcomes, *i.e.,* is the Laboratory contributing to the nation's security by solving the nation's problems?"

The scientific process is the means by which we investigate physical processes in a rigorous and reproducible fashion. It includes ensuring that theory and simulation are verified by experiment, and that experiments are informed by theory, on a regular and systematic basis. This interaction between experiment, theory, and simulation is necessary to assure the success of many scientific and technological efforts. The metric of verifiable scientific process applies also to areas where experimentation is not appropriate. For example, data mining for counterterrorism may have no connection to any laboratory experiment, but may be verifiable by independent analyses of the databases or by results obtained in the real world. Similarly, advances in theoretical and simulation methods may not need experimental tests; *e.g.,* an improved algorithm that speeds up a calculation by a quantifiable factor and agrees with previous calculations, or a theoretical approach that presents a new way of arriving at already known results but provides additional insights, are instances of verifiable scientific process.

An important sign of a healthy scientific process is vigorous peer engagement and review within the organization, both formal and informal.

Expert reviewers should emphasize primary metrics of the scientific process, based on *evidence,* as opposed to secondary metrics based on the *opinions* of experts and peers. Examples of the evidence that can produce primary metrics are: (1) the Columbia/Challenger disasters; (2) an underground nuclear test; (3) the result of an experiment designed to test a theoretical prediction; (4) successful test and operation of an instrument developed for space exploration. The Laboratory's reputation as a premier scientific institution is based largely on primary metrics, just as the high points of NASA's reputation come from successful missions in space.[5] In the absence of nuclear testing, a set of new primary metrics for the Laboratory needs to be clearly identified.

Additionally, Laboratory management should be assessed on their intellectual leadership, particularly their success in translating national mission needs into science and technology challenges, and conversely in integrating the staff's technical innovation into large-scale mission solutions. This kind of integrative leadership is essential for Los Alamos to be a National Laboratory rather than a job shop.

A final objective of expert review should be to analyze and anticipate outcomes, *i.e.* the ways in which the solutions that we devise have provided the nation new options to solve national security and other problems. Because of the heterogeneous nature of outcomes, it is difficult to imagine its evaluation by an objective metric. We therefore recommend that outcomes be analyzed by divisions and programs, and these analyses be assessed by expert reviewers. These analyses should consider the degree to which scientific excellence has influenced success in meeting schedules and budgets. Evaluation of outcomes will ensure that a big laboratory does big things.

Integrity of expert review

Given the finite amount of time that can be spent on the metrics process, the most effective path for expert review is face-to-face interaction with an expert panel. Both presentations to the panel, and supporting written self-assessments, should stress the verifiable scientific process. However, face-to-face interactions, for understandable human reasons, pose a risk of loss of independence that must be countered by careful management of the review process.

---

[5] While experiments are essential to the scientific process, we need to be careful not to define them as events that should not fail. This would move them towards "admiral's test" demonstrations, which yield no new insight into the world. Some risk is essential for scientific breakthroughs to occur, and risky experiments can yield insight from apparent failures as well as successes.

In the past, expert review at LANL has come largely from the division review committees (DRCs), which evaluate both impacts and outcomes, and also other factors that might better be addressed by objective measures. We are concerned about the independence of the DRC process from management. We therefore recommend that the senior leadership of the Lab must strongly and openly ensure that expert reviews (DRCs or their successors) are free of institutional bias against truthful-but-unfavorable conclusions. This is essential to maintaining the technical credibility of the Laboratory. The ethical leadership taken by the Director and senior management is essential to the integrity of expert review. Another way to enhance independence is to partially distance the nomination process from the organization to be reviewed, *e.g.* by a Livermore division nominating 1/3 of the review panelists for its Los Alamos counterpart, and vice versa. We advocate not that the partner lab supply the panelists, but that they nominate them.

We recommend that some fraction of expert review at Los Alamos be conducted by the national "gold standard", the National Academies. For National Academy reviews, independence is guaranteed because the Academy sets up the process and selects panel members, decoupling the Laboratory from selecting sympathetic reviewers. The Academies might be involved on a trial basis to evaluate two or three divisions, including some with the strongest scientific reputations.

Next steps:

Our recommendations are based on our career-long experience in science, technology, and their review. It may be useful to analyze case studies of these metrics to determine how they would have captured the merit of Laboratory initiatives that were eventually successful – if the metrics had been applied at an intermediate stage of the research. Candidates for case studies could include proton radiography, space gamma-ray and nuclear detection for nuclear explosion detection, and fundamental actinide chemistry in support of the Rocky Flats cleanup.

We re-emphasize the commitment of the Fellows to excellence in science and technology at the Laboratory. We would like to help maintain that excellence. While it may not be appropriate to task the Fellows as an integral part of the regular peer review process, *e.g.*, annual division reviews, we would be happy to review particular aspects of the Los Alamos scientific process, or review the review process itself as it develops.

Our objective is to help define excellence, not to dictate the way to achieve it. The right metrics will themselves drive the institution towards excellence. The push towards excellence can be reinforced by vigorous and effective postdoctoral and LDRD programs, and real improvements in Laboratory operations that let our scientists do science.

APPENDIX 1

Director's commission

Los Alamos
NATIONAL LABORATORY
EST.1943

memorandum
Office of the Director

To/MS: Rusty Gray, Fellows Coordinator, G755
Bill Priedhorsky, Deputy Fellows Coordinator, B241
From/MS: Robert W. Kuckuck, A100 Bob
Phone/Fax: 7-5101/Fax 7-2997
Symbol: DIR-06-083
Date: March 13, 2006

SUBJECT: Metrics for Excellence

I know and appreciate that the Laboratory Fellows are concerned with the vitality and excellence of science and technology at Los Alamos. In order for the incoming LANS team to at least preserve and hopefully even enhance the Laboratory's excellence, they would benefit from having metrics. There will certainly be metrics of excellence for the new contract, and it is important that they be the right ones. I would, therefore, appreciate your advice on the critical components for metrics of Laboratory science and technology excellence.

It would be very helpful for the Fellows to address the following question: "What metrics of scientific and technological excellence are appropriate for the particular circumstances of LANL?"

Metrics of excellence for this Laboratory should be applicable to the mix of applied and fundamental research at LANL, to science and technology that is both classified and unclassified, and should consider the extensive body of work on metrics for scientific excellence that has been developed at the national level.

It would be helpful to receive your thoughts in the form of a short white paper, accompanied by a short PowerPoint briefing, which you could deliver to me by April 15. I would like to be able to meet with you to discuss your suggestions.

APPENDIX 2

Task force membership:

Robert F. Benjamin, P-23, retired
Gregory H. Canavan, P-DO
David W. Forslund, CCS-DO, retired
Hans Frauenfelder, T-10
Philip Jeffrey Hay, T-12
Mikkel B. Johnson, P-25, retired
Norman A. Kurnit, P-24, retired
Terence E. Mitchell, MST-8, retired
Christopher Morris, P-25
Bill Priedhorsky, ISR-DO (chair)

Task force interviews:

February 16:	Tom Bowles, Chief Science Officer
February 23:	David Watkins, LDRD Program Manager
February 24:	Terry Wallace, ADSR
March 2:	Sue Seestrom, ADWP
April 7:	Doug Beason, ADTR