NREL/TP-320-6696 • UC Category: 900 • DE94011823

Using Research Metrics to Evaluate the International Atomic Energy Agency Guidelines on Quality Assurance for R&D

Mark Bodnarczuk
National Renewable Energy Laboratory

Prepared for the 21st ASQC National Energy and Environmental Quality Division Conference, Tucson, Arizona, September 18–21, 1994

National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, Colorado 80401-3393
A national laboratory of the U.S. Department of Energy
Operated by the Midwest Research Institute
for the U.S. Department of Energy
under Contract No. DE-AC02-83CH10093

June 1994

NOTICE

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

Printed in the United States of America
Available from:
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

Price: Microfiche A01
Printed Copy A03

Codes are used for pricing all publications. The code is determined by the number of pages in the publication. Information pertaining to the pricing codes can be found in the current issue of the following publications which are generally available in most libraries: *Energy Research Abstracts (ERA)*; *Government Reports Announcements and Index (GRA and I)*; *Scientific and Technical Abstract Reports (STAR)*; and publication NTIS-PR-360 available from NTIS at the above address.

Printed on recycled paper

Using Research Metrics to Evaluate the International Atomic Energy Agency Guidelines on Quality Assurance for R&D[1]

Mark Bodnarczuk
National Renewable Energy Laboratory
1617 Cole Boulevard
Golden, Colorado 80401

Abstract

The objective of the International Atomic Energy Agency (IAEA) Guidelines on Quality Assurance for R&D is to provide guidance for developing quality assurance (QA) programs for R&D work on items, services, and processes important to safety, and to support the siting, design, construction, commissioning, operation, and decommissioning of nuclear facilities. The standard approach to writing papers describing new quality guidelines documents is to present a descriptive overview of the contents of the document. I will depart from this approach. Instead, I will first discuss a conceptual framework of metrics for evaluating and improving basic and applied experimental science as well as the associated role that quality management should play in understanding and implementing these metrics. I will conclude by evaluating how well the IAEA document addresses the metrics from this conceptual framework and the broader principles of quality management.

I. Framing the Question

In a previous paper,[2] I discussed the contrast between the successful application of quality management (QM) in industrial R&D laboratories[3] and the more difficult case of applying it to basic experimental science performed at national laboratories. I claimed that if one could successfully define metrics for evaluating and improving basic experimental science in disciplines like high-energy physics and experimental astrophysics (which are the most difficult cases), then defining such constructs for more applied science would be much less problematic. In what follows, I extend this work by giving a number of examples of metrics for *applied* experimental science and discussing how the principles of QM relate to the metrics for both basic and applied

---

[1] This document is the International Atomic Energy Agency's (IAEA's) Proposed Nuclear Safety Standard (NUSS) Safety Guide for Quality Assurance for Research and Development (50-SG-QA16).

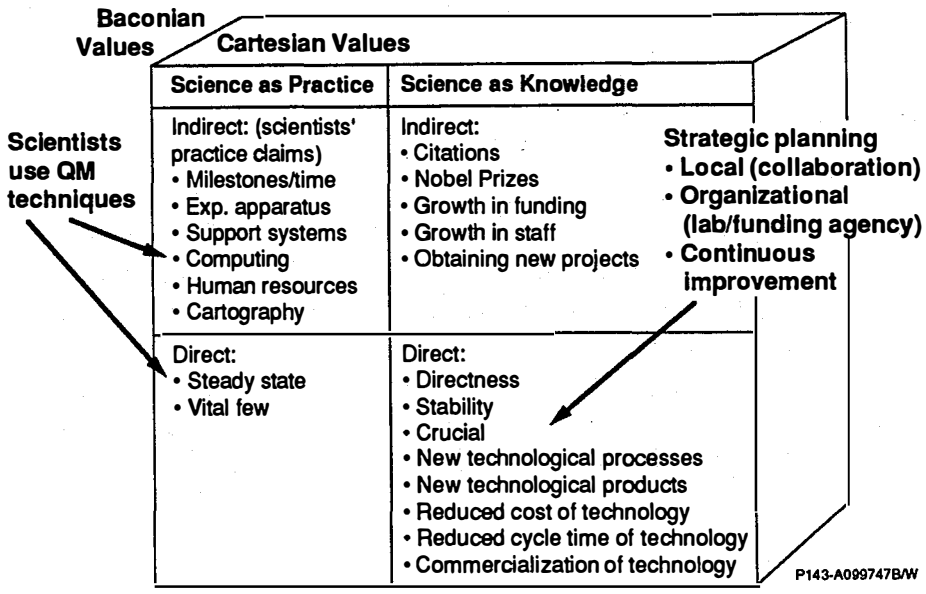
[2] See Mark Bodnarczuk, "Defining Metrics for Evaluating and Improving Basic Experimental Science," in *The Proceedings of the ASQC 48th Annual Quality Congress*, May 22-25, 1994.

[3] See Jerry D. Holmes and David J. McClaskey, "Improving Research Using Total Quality Management" in *The Proceedings of the Juran Institute - IMPRO92*, November 11-13, 1992, Session 3B: Quality in Research and Development, p. 13 ff.

experimental science. Using this taxonomy, I go on to evaluate how well the IAEA document addresses these metrics and the more general principles of QM.

II. A Role for QM in Experimental Science

What is the purpose of science? The figure below defines a framework within which this question can be discussed by interrelating Baconian and Cartesian values about science. On one hand, those holding to Baconian values tend to view science as the servant of the taxpayers who finance it. The knowledge produced should be utilitarian, e.g. applied science that improves the quality of life through technological advances.[4] Alvin Weinberg's model of internal and external criteria for evaluating which types of science should be funded exemplifies Baconian values. His internal criteria attempt to answer two questions. First, is the field ready for exploitation? Second, are the scientists in the field actually competent? Weinberg claimed that these decisions could be made only by scientists. He identified three external criteria that could be decided by nonscientists: technological merit, social merit, and scientific merit. The criterion of scientific merit assessed the degree to which the knowledge produced by the discipline requesting funding contributed to its neighboring disciplines.[5]



[4] Francis Bacon, *The Works of Francis Bacon*, James Spedding, Robert Leslie Ellis, Douglas Denon Heath (eds.), (Boston: Taggard and Thompson, 1893).

[5] See Alvin Weinberg, "Criteria for Scientific Choice," in *Minerva*, vol. 1, 1963, pp. 159-171.

On the other hand, those holding to Cartesian values tend to view basic experimental science as the "feedstock" of future utilitarian (applied) science. In addition, they might claim that all cultured societies should support scientific knowledge for its own sake (like art or music), even if there are no immediate known applications of the knowledge produced. Although I have shown them diagrammatically as orthogonal, there is actually a kind of dialectic tension between Baconian and Cartesian values and the outcome of this dialogue is crucial to developing a science policy. Having established this macrosocial framework, I will focus in the remainder of the paper on the microsocial aspects of scientific practice, beginning with the individual scientific proposal.

When scientists propose research programs they normally concretize them into an experimental proposal, and these proposals becomes an important part of the information that the funding agency uses to decide whether to support the research. In such experimental proposals, scientists make two types of claims. First, they make *knowledge claims* that are calculated hypotheses of the experimental path that must be taken, and the expected results and new knowledge to be obtained by performing the experiment. Second, they make *practice claims* that define the planning and managing elements needed to obtain that new knowledge, including estimates of how long it will take to travel the experimental path, the types of experimental equipment that will be needed, how much on-line and off-line computing resources will be needed, how much engineering and technician support will be needed, and other costs and resources. Experimental proposals are often submitted to funding agencies two or more years prior to performing the experimental work, and because of the exploratory nature of science the *knowledge* claims are like moving targets that cannot be well defined. Consequently, the *practice* claims needed to obtain that knowledge are often "place holders" that cannot be nailed down. But the *knowledge* claims become increasingly refined and fine-tuned as they move toward the performance of the current year's research; and, as a result scientists ought to be able to articulate and evaluate the *practice* claims more precisely. But nailing down practice claims (e.g., managing the research) has been problematic for reasons I will explain below.

Metrics for Science-as-Knowledge

I will use the notion of a "metric" to indicate a standard of measurement, but I want to make the distinction between a direct metric (a close logical, causal, or consequential relationship to the things being measured) and an indirect metric (a collateral or circumstantial relationship to the things being measured). Direct metrics when obtainable are preferred because of their more veridical characterization of the thing being measured. For example, when evaluating whether an experiment has produced "good" physics, one could use Irvine and Martin's indirect metric of counting the number of citations that a particular publication receives; but this is based on the inference that if a scientific publication contains "good" physics, then it will be cited more frequently than one that does not.[6] Using the number of Nobel Prizes received by Americans as a

[6] See J. Irvine and B. R. Martin, "Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators," in *Social Studies of Science*, 1985, vol. 15, p. 300, and also Steven Yearly, *Science, Technology, & Social Change*, (London: Unwin Hyman, 1988), p. 88 ff.

metric of the effectiveness of U.S. science policy, or using a laboratory's growth in funding and staff as an indication of the quality of its research are also based on the inference that these are the consequence of quality research. The fact is that increases in funding or staff may be due largely to political factors. Although both scientists and non-scientists use them, indirect metrics provide *no direct* indication of the quality of the content of the science.

Unlike indirect metrics, direct metrics for science-as-knowledge *do* provide a *direct* indication of the quality of the content of the science. For *basic* experimental science, I have identified the direct metrics of directness, stability, and crucial experiments (see Figure).[7] *Direct* measurements bring experimental reasoning another rung up the ladder of causal explanation (for example, measuring a background previously only calculated).[8] *Stable* measurements mean varying a feature of the experimental setup (including changes in the test substance, apparatus, arrangement, or data analysis) and obtaining the same basic result.[9] Stable measurements are desirable because each variation introduced into the experimental design makes it more difficult to postulate an alternative causal story that will satisfy all the observations; this is because the effect is nested within ever more complex loops of experimental demonstration.[10] A final type of direct metric for basic experimental science-as-knowledge is the *crucial* experiment - an experiment at the "crossroads" - that yields new effects, new experimental or theoretical directions, more direct and stable measurements, or data that enable scientists to choose between competing theories.[11] It is important to note that even when scientists are involved in an entire program of related research, they almost never repeat the same experiment. Detailed case studies show that experiments are performed in series or "strings." The follow-up experiments that constitute these experimental "strings" are almost always attempts to *improve* the directness and stability of the measurement using improved techniques for detection, data acquisition/monitoring, or data analysis (see figure).[12]

Direct metrics for *applied* experimental science-as-knowledge are more tangible and include things like the development of new technological processes, development of a new technological products, reduced cost of the technologies, reduced cycle time of technologies, improved performance of technologies as evidenced in increased reliability, availability, maintainability, safety, or demonstrations of the knowledge's usefulness through commercialization. Examples of metrics for applied experimental science are legion, and I will

---

[7] I have adopted Peter Galison's notion of directness and stability here. See Peter Galison, *How Experiments End*, (Chicago: The University of Chicago Press, 1987), pp. 259-260. Also see Mark Bodnarczuk, "Defining Metrics for Evaluating and Improving Basic Experimental Science," in *The Proceedings of the ASQC 48th Annual Quality Congress*, May 22-25, 1994, pp. 7-9.

[8] Galison, pp. 259-260.

[9] Galison, p. 234 ff. Latour describes roughly the same notion in terms of the "trials of strength" that various effects and substances endure at the hands of experimenters in Bruno Latour, *Science in Action*, (Cambridge, MA: Harvard University Press, 1987), p. 74 ff.

[10] Galison, p. 260.

[11] See Ian Hacking, *Representing and Intervening*, (New York: Cambridge University Press, 1983), p. 249.

[12] I discuss such case studies in Mark Bodnarczuk, *The Social Structure of Experimental Strings at Fermilab; A Physics and Detector-Driven Model*, Fermilab-Pub-91/63, March 1990, p. 14 ff.

give just a few. Note that milestones and associated metrics are so interrelated that they cannot be clearly separated, with the milestone being expressed as a goal or outcome and the metric expressed as the unit of measure that quantifies the goal or outcome. One example might be applied research in wind energy aimed at wind characterization at wind turbine sites and characterizing aerodynamic phenomena experienced and created by wind turbines during operation. Examples of milestones and associated metrics might be things like (1) performing a complete analysis of U.S. and European aerodynamic data with the goal of defining methods of gaining a 5% improvement in energy capture, (2) developing and validating a blade-fatigue model to allow design of a blade with 30% greater lifetime than current standards, or (3) developing and implementing a wind resource monitoring strategy at remote villages and using the data gathered as the basis for at least 20 sales of wind-hybrid systems in the next five years.

Another example might be applied research in solar industrial technologies with the goals of developing new industrial applications of solar technologies and introducing those technologies into a broad spectrum of the industrial sector. Examples of milestones and associated metrics might be things like (1) negotiating and establishing cost-shared projects with three electric utilities to encourage use of solar by industrial and commercial firms, the installation of 20 solar systems, annual energy savings of 24,000 gigajoules, and the creation of 30 primary and secondary jobs, (2) establishing the technical and economic feasibility of advanced ceramics production, with industry sharing 30% of the project costs and resources, or (3) selecting three new solar advanced processes concepts for initial exploration, recording one record of invention from these studies, and obtaining two expressions of interest from industrial partners for further collaborative investigations.

Metrics for Science-as-Practice

There are many stages of experimental practice that could be discussed, but I will focus on that aspect of experimental practice in which the experimental design, procurements, installation, and overall configuration of the apparatus have settled down enough to actually perform the experiment. At this point, the scientist must concern himself with bringing the operation of the experimental apparatus into a "steady state," where all possible operational parameters of the apparatus are understood and functioning as designed. At this stage, the majority of processes that occur in the organizational infrastructure of the laboratory in which an experiment is embedded exert little or no causal efficacy on the outcome of an experiment. In addition, scientists construct organizational and experimental "protective belts" around their work to protect it from all but the most devastating laboratory perturbations - usually drastic reductions in resources and funding. So how does one isolate the "vital few" activities, and support organizational interfaces that can actually affect the output of an experiment?

The vital few are those events that exert *direct causal effects* on the stability of the apparatus and computing because only those can affect the outcome of the experiment itself, for example, background and noise in the apparatus. Scientists control potential perturbations to the stability of the experiment by doing things like (1) visually or computationally monitoring the

apparatus to ensure proper operation of components like power supplies, gases, and calibration of equipment; (2) ensuring that the proper materials, targets, and chemicals are being used; (3) ensuring that data rates are appropriate and that the data acquisition and software systems are functioning as designed and intended; (4) ensuring that the cartographic function of experimental science is carried out in scientific notebooks and that correct data are recorded to magnetic tape, disk, or other media; and (5) ensuring that the appropriate measurement uncertainty analysis is performed on experimental results that are reported in the literature. Only these types of things can seriously affect the production of scientific knowledge. They are a minority of laboratory activities. They are the only activities on which scientists are even interested in implementing management controls (see figure).

Our final component is indirect metrics for science-as-practice which are an important component of the practice claims that scientists make in their proposals. This includes estimates of how long it will take to travel the experimental path, the types of experimental equipment that will be needed, how much on-line and off-line computing resources will be needed, how much engineering and technician support will be needed, time frames, milestones, human resources, and other costs. As I have described elsewhere,[13] although many scientists tend to devalue these planning and managing aspects of science, I see no fundamental reason why many of the continuous improvement methodologies developed by Juran cannot be tailored and applied to these practice claims. Scientists and laboratory managers could form cross-functional teams (composed of scientists, engineers, and technicians) that would more closely examine why scientists miss the milestones that they have defined for themselves. Following a standard quality method like storyboarding, the team could analyze the symptoms of the missed milestones, formulate theories about why they were missed, test these theories, and identify root causes. Was the problem attributable to a limitation that nature imposed on the experiment? Was it impossible to push the technologies involved any further? Was the problem an inevitable part of the pedagogic process of obtaining knowledge by actually *doing* an experiment and could not be avoided? Or was it a systemic problem where a spokesperson had no authority to make collaborators come through on their commitments? Was it a lack of planning or managing on the part of the principal investigator, or the inability to stop introducing new parameters and changes into an experimental design that should have been fixed? Was the problem due to the lack of supervision of a graduate student by his or her senior professor or laboratory manager? Problems imposed by nature may be unavoidable, but the practice of science would certainly be improved if most of these other problems were solved.

[13] See Mark Bodnarczuk, "Defining Metrics for Evaluating and Improving Basic Experimental Science," in *The Proceedings of the ASQC 48th Annual Quality Congress*, May 22-25, 1994, pp. 12-13.

The Role of Quality Management

When viewing the two components of science-as-knowledge, I see little or no role for QM with regard to the indirect metrics for knowledge (see figure). However, the direct metrics for science-as-knowledge clearly lend themselves to some QM elements, for example, strategic planning. In terms of strategic planning at the local level, individual collaborations of researchers must be sure that their research is out on the front of the scientific and technological power-curve, obtaining more direct and stable measurements than other scientists working in related areas. Strategic planning at the organizational level requires that laboratory management construct an institution-wide research "menu" that is world class - a research "menu" that is constituted by increasingly more direct and stable measurements and crucial experiments that may some day lead to the kind of applied science that supports the development of new technological processes and products (see figure). As I mentioned previously,[14] this is what constitutes continuous improvement in science - this is the foundation of scientific and technological progress.

The application of the tools of QM to the science-as-practice side of the equation is much more self evident. In terms of direct metrics for science-as-practice, once scientists identify the "vital few" activities that can actually affect their experiment, they can use many of the traditional tools of QM to ensure that the goals and performance objectives of the R&D work are achieved. I discussed them above and will not reiterate them here (see figure). The application of QM tools to the indirect metrics for science-as-practice is also more clear cut in terms of managing the resources needed to meet milestones, use computing power, use human resources, and myriad other aspects of the practice of science.

III. Evaluating the IAEA Guidelines on Quality Assurance for R&D

In this section, I will evaluate the IAEA Guidelines on Quality Assurance for R&D using the distinction between science-as-practice and science-as-knowledge and the associated metrics. I believe this is one way to determine whether the document is user friendly to the scientists that will have to use it, and whether it will be value-added when properly implemented. I will begin by defining an historical context within which the document should be viewed. Over the past decade, the nuclear industry in general (and the IAEA in particular) has recognized that its traditional perceptions of QA were not contributing to plant safety and reliability as meaningfully as they could or should.[15] During this time, the IAEA's hierarchy of documents consisted of a Safety Code that described QA requirements and a number of Safety Guides that provided additional guidance on implementing the Safety Code for various activities. This hierarchy is similar to the

---

[14] See Mark Bodnarczuk, "Defining Metrics for Evaluating and Improving Basic Experimental Science," in *The Proceedings of the ASQC 48th Annual Quality Congress*, May 22-25, 1994, pp. 8-9.

[15] See Frank Hawkins and Nestor Pieroni, "Quality Assurance at Nuclear Power Plants: Basing Programmes on Performance" in *IAEA Bulletin: Quarterly Journal of the International Atomic Energy Agency*, Vol. 33, No. 4, 1991, Vienna Austria, p. 29.

current structure of DOE's nuclear safety policy.[16] In 1990, the IAEA Nuclear Safety Standards Advisory Group (NUSSAG) began a program designed to enhance nuclear safety by revising and improving its QA code and safety guides. The goal was to update IAEA's QA documents so that they depicted contemporary principles and techniques for managing, achieving, and assessing quality. Some of the products that emerged from the process were a new performance-based Safety Code (50-C-QA; Code on the Safety of Nuclear Power Plants: Quality Assurance) and a revised Safety Guide (50-SG-QA1; Establishing and Implementing a Quality Assurance Programme) that provided guidance for implementation. The transformation of IAEA's QA requirements into a performance-based approach to quality was similar to the process that led to the development of DOE 5700.6C, DOE-ER-STD-6001-92, and 10CFR830.120.[17] In 1992, NUSSAG determined that a new Safety Guide for R&D quality should be developed and asked me to write Draft 1 of the document. What emerged subsequently from a Vienna-based committee composed of Neil Redman, Jim Mullan, Nestor Pieroni, Chengkai Chen, John Hillairet, and Kent Truss, and myself was Draft 2 of the IAEA Guidelines on Quality Assurance for R&D (Safety Guide 50-SG-QA16). Draft 3 of the document is currently being reviewed by the 135 countries that constitute the member states of the IAEA. Comments on Draft 3 will be received and reviewed before the end of 1994, and the final document should be issued shortly thereafter.

Safety Guide 50-SG-QA16 will be used to provide guidance for developing QA programs for R&D work on items, services, and processes that are important to safety and support the siting, design, construction, commissioning, operation, and decommissioning of nuclear facilities in the 135 members states of the IAEA. Although nuclear facilities may vary from one member state to another, 50-SG-QA16 defines a nuclear facility as any facility used for power generation, spent fuel storage, waste storage and treatment, fuel reprocessing and plutonium processing, fuel fabrication, research, radioactive isotope production, and handling and storage and disposal of radioactive materials. Safety Guide 50-SG-QA16 is designed to supplement Safety Code 50-C-QA and Safety Guide 50-SG-QA1 by providing more specific guidance for applying the Code's basic requirements to R&D activities. Structurally, the document is divided into three sections: managing R&D work, performing R&D work, and assessing R&D work. As with DOE 5700.6C, this is roughly a Plan, Do, Check, Act cycle. Safety Guide 50-SG-QA16 differs from DOE 5700.6C in some minor ways. For example, Criterion 3 of Safety Guide 50-SG-QA16 is entitled "Non-Conformance Control and Corrective Action," rather than the DOE 5700.6C counterpart, "Quality Improvement."

---

[16] See DOE Notice 1321.138 (Departmental Directives System: Interim Improvement Notice 2), February 16, 1993.

[17] For a brief history of the development of DOE 5700.6C and DOE-ER-STD-6001-92, see Mark Bodnarczuk, "DOE 5700.6C, 10CFR830.120, DOE-ER-STD 6001-92, and Covey-Based TQM: A Historical Perspective on Current Issues in Research Environments," in *The Proceedings of the Twenty-First ASQC National Energy and Environmental Quality Division Conference*, September, 1994. For a comparison between ASME NQA-1 and 10CFR830.120, see Mark Bodnarczuk, "The Application of 10CFR830.120 in a Basic Research Environment," in *The Proceedings of the Eighteenth Annual ASQC National Energy Division Conference*, October, 1991.

Creating an R&D Environment

There are a number of things that must be present in an organization to create the kind of environment in which R&D will flourish. One of the most important is to understand the contrasts and differences between basic and applied research and engineering development. In order to address these and other related issues, the section on managing R&D work opens with a comparison of the objective and direct products of basic and applied *research* (the R - new knowledge or analytic studies that may or may not have any known application to technological processes or products) and the objective and direct products of engineering *development* (the D-development of prototype devices, new software, new testing methods, new or improved technologies, or new industry standards). Among other differences, the document points out that whereas engineering design activities have well-known specifications for many of the materials, tests, inspections, and methods, R&D work can begin with some procedures, but often quickly deviate from them as a legitimate component of the conduct of research; entirely new procedures are often developed through the process of trial and error (see sections 201-205).[18]

Safety Guide 50-SG-QA16 discusses other things that are crucial to creating and sustaining an environment in which R&D will flourish. Example include fostering and encouraging creativity, intellectual stimulation, innovation, and collaboration; requiring that good work practices are the only acceptable way to perform or support R&D work; leading by example and demonstrating personal commitment to continuous improvement; empowering personnel at all levels in the organization and recognizing them for excellence in performance; ensuring that sufficient resources are available and priorities set for their deployment; and avoiding the temptation to overload scientists with administrative tasks (see section 206). In what follows, I will evaluate Safety Guide 50-SG-QA16 against the conceptual framework of interrelated metrics discussed in the first part of the paper and shown in the figure.

Direct Metrics for Science-as-Knowledge

Safety Guide 50-SG-QA16 does not discuss any of the *indirect* metrics for science-as-knowledge that I identified (citations, science policy matters, Nobel Prizes, etc.), but it does discuss the more crucial aspects of *direct* metrics for science-as-knowledge. For basic experimental science, I have identified the direct metrics of directness, stability, and crucial experiments. Applied experimental science includes things like the development of a new technological process, development of a new technological product, reduced cost of the technology, reduced cycle time of the technology, improved performance of a given technology as evidenced in increased reliability, availability, maintainability, safety, or demonstrations of the knowledge's usefulness by commercialization (see figure).

I have already discussed how scientists must perform strategic planning at the local level as a part of defining direct metrics for science-as-knowledge. This means that the research

[18] The references to section numbers are the section numbers of Safety Guide 50-SG-QA16.

collaboration proposing the R&D must be sure that their research is out on the front of the scientific and technological power-curve relative to work performed by other scientists working in related areas. Also, Safety Guide 50-SG-QA16 requires the responsible principal investigator or scientist to prepare an R&D plan. The R&D plan should provide a brief historical overview of the work, including publications that describe previous experiments, theories, feedback from users of previous R&D work, or technological developments that have led to the work described in the R&D plan (section 304). The R&D plan should also describe any relationships, interrelationships, or dependencies that the R&D work has to other known projects or research programs. When it is known that similar work will be performed elsewhere, this should be stated with a brief explanation of how (and if) the work will be coordinated (section 308). The R&D plan should also describe the purpose of the R&D, as well as identify criteria that can be used for assessing the success or failure of the R&D (section 303). Unfortunately, Safety Guide 50-SG-QA16 does not discuss the kind of institution-wide strategic planning that is a crucial component of QM in general and direct metrics for science-as-knowledge in particular.

Other areas discussed in Safety Guide 50-SG-QA16 that are relevant to metrics for science-as-knowledge are issues involving data analysis and reporting the final results of R&D work. When analyzing data for acceptability, researchers should define (1) the methods for the analysis of all data and results used and received, (2) the assumptions and methodologies used for the analysis of data so that competent experts can evaluate how the data were interpreted, and (3) the methods used to identify and minimize measurement uncertainty (section 321). The final report issued by the responsible principal investigator or scientist should describe such items as (1) the results contained, their range of application and validation, (2) the relationship of the results to previous publications, experiments, theories, or technological developments, (3) a description of the apparatus and the operations/data gathering activities, (4) a description of significant problems that occurred during operation/data gathering activities, (5) a description of data analysis issues similar to those listed in section 321, and (6) a summary of the work, including any conclusions and/or recommendations. Finally, under the auspices of independent assessments, Safety Guide 50-SG-QA16 describes how independent assessments (peer reviews) should be used to evaluate the success criteria defined in the R&D plan (section 404).

Direct Metrics for Science-as-Practice

I have defined *direct* metrics for science-as-practice as bringing the operation of the apparatus into a "steady state," where all possible operational parameters of the apparatus are understood and are functioning as designed. As I noted, these are the "vital few" activities embedded within the organizational and operational structure of laboratory life. Safety Guide 50-SG-QA16 addresses these issues in a number of ways. To begin with, the R&D plan developed by the responsible principal investigator or scientist should contain a description of the relevant components of the experimental equipment/apparatus and their configuration, along with a discussion of how any unusual or potentially problematic techniques, special tools, and methodologies will be handled (section 305). In terms of the actual conduct of the R&D, during

the commissioning of the equipment/apparatus or prototype, the calibration and performance requirements of test, measurement, and diagnostic equipment/apparatus (TMDE) should be defined to a level of detail adequate to ensure the R&D goals are achieved. During the experimental operations/data gathering stage of the R&D work, the responsible principal investigator or scientist should ensure that calibration and performance requirements for TMDE are maintained and understood (section 318), and that the systems and subsystems of the experimental equipment/apparatus are functioning as intended (section 319).

Indirect Metrics for Science-as-Practice

As might be expected, the majority of the content of Safety Guide 50-SG-QA16 focuses on what I have defined as *indirect* metrics for science-as-practice, e.g., estimates of how long it will take to travel the experimental path, the types of experimental equipment that will be needed, how much on-line and off-line computing resources will be needed, how much engineering and technician support will be needed, time frames, milestones, human resources, and other resources and costs. Due to space constraints and my goal of evaluating the document against criteria used in scientific practice, I will only include highlights of this portion of the document because they are familiar to most quality professionals. In terms of indirect metrics for science-as-practice, Safety Guide 50-SG-QA16 provides guidance on the following aspects:

- The structure of the R&D organization, including roles, responsibilities, authorities, and interfaces for principal investigators/scientists and the senior managers who are responsible for the organization within which the R&D work is performed (sections 207-210, 216).
- Training and qualification that takes into account the collaborative nature of R&D and the widely divergent levels of education, skills, and experience of the scientists, operators, designers, engineers, welders, technicians, and craftsmen that constitute the workforce needed to perform today's R&D activities (sections 211-215).
- A graded approach based on things like the intended end-use of the knowledge, data, technological process, or technological product, particularly in terms of their impact on nuclear safety (section 217).
- The establishment of a non conformance control and corrective action process that is tailored for R&D work (sections 218-221).
- The establishment of a document control and records management system that is tailored for R&D work (sections 222-225).
- The planning and preparation of R&D work, including the preparation of an R&D plan (sections 301-302, 306-307, 309-314).
- The conduct of the R&D work (sections 315-317, 320) and the reporting of final data results (section 323).
- The performance of management self-assessments (sections 401-402) and independent assessments (403-406).

IV. Conclusion

The struggle to develop quality management concepts that "map" onto the cultural and work practices found in basic and applied experimental science has been (for better or for worse) an attempt to differentiate basic and applied research from the culture found in the non-research components of the nuclear industry. I believe that Safety Guide 50-SG-QA16 successfully addresses those matters of quality and safety that are crucial to the nuclear industry, and at the same time addresses the issues that are crucial to developing a value-added approach to managing nuclear-related R&D work. In closing, let me highlight four areas that the document does *not* adequately address. First, Safety Guide 50-SG-QA16 does address a number of strategic planning elements, but only on a local level, e.g., at the level of the experimental collaboration proposing the work. Strategic planning on an institution-wide level (including the development of a vision, mission, values, critical success factors, and metrics) are not discussed in the document.

Second, the more modern notions of quality improvement are not articulated in the document, not even in the criterion that comes closest to fulfilling some of those functions - Non Conformance Control and Corrective Action. Although the elements listed in this criterion are a necessary component of quality improvement (especially with nuclear safety considerations), the Safety Guide 50-SG-QA16 encourages the notion of quality improvement without articulating the principles and practices of quality improvement as espoused by Juran, Deming, or even the Malcolm Baldrige National Quality Award Criteria. Third, the document does not actually address the kind of cultural and human factors that are so crucial to the cultural transformation that is prerequisite to implementing QM - for example, Covey-based principles and values. Finally, the document does not really address the issue of customer focus. But if an organization would adopt a graded approach to implementing the document that incorporates the elements needed to correct these few omissions, I believe that Safety Guide 50-SG-QA16 would not only be user friendly and value-added to nuclear-related R&D; it would also help to lead the implementing organization down the road toward implementing TQM for R&D. Of course, the most crucial component of implementing TQM is to *believe* it and then to *do* it, and this requires us to change our values, beliefs, and the way we do business.