FEBRUARY 2022

What Is Safeguards Effectiveness?

Valeri **Bytchkov** and Mark **Hibbs**

A PRODUCT OF
The Future of IAEA Safeguards: Rebuilding the Vienna Spirit through Russian-U.S. Expert Dialogue

20 YEARS OF NTI
BUILDING A SAFER WORLD

Center for Energy and Security Studies
www.ceness-russia.org

Verification as an Important Component of International Treaties

There are a number of international treaties and agreements in the field of nuclear arms control and non-proliferation. Their viability depends on mutual trust between the parties involved, reinforced by mechanisms for verification. For example, in U.S.-Russian nuclear arms control agreements, there is mutual verification of each state's compliance with obligations under agreement. The following factors may contribute to an agreement's viability: (a) the capability to detect a significant violation of the agreement, (b) the timeliness of such detection, and (c) the response to the violation. In general terms, the viability of such an agreement may rest upon two components: a verification component consisting of detection and reporting of findings, and a "political" component consisting of the parties responding to findings, as appropriate, through enforcement.

There are several international agreements aimed at ensuring that nuclear energy will be used peacefully, the most important of which is the 1970 Treaty on the Non-Proliferation of Nuclear Weapons (NPT). The International Atomic Energy Agency (IAEA) safeguards system, which was developed beginning in the 1960s, was adopted as the verification mechanism for the NPT. The obligations of the non-nuclear weapon states that are parties to the NPT have been translated from Articles II and III of the NPT to the text of Paragraphs 1 and 2 of the comprehensive safeguards agreements (CSA) that these states are required to conclude with the IAEA. The purpose of this system is to verify compliance of the states with their obligations under safeguards agreements. If the IAEA detects non-compliance in a state, the IAEA reports this conclusion to the United Nations Security Council (UNSC). If in such a case the state does not cooperate in resolving the non-compliance, enforcement[1] measures may be applied by IAEA member states through the authority of the UNSC; enforcement is the prerogative of the international community and the UNSC, not the IAEA Secretariat.

The nuclear non-proliferation regime includes legally binding obligations of states under the NPT and nuclear-weapons-free zone (NWFZ) treaties, as well as numerous states' initiatives, such as proliferation-resistant nuclear fuel cycles, multilateral export-control arrangements, and international nuclear fuel banks. The effectiveness of the non-proliferation regime may ultimately be measured in terms of the

[1] The term "enforcement" was introduced together with the term "safeguards" in 1945 by the leaders of Canada, the United Kingdom, and the United States in their joint declaration concerning possible measures "to prevent the use of atomic energy for destructive purposes" ("Joint Declaration by the Heads of Government of the United States, the United Kingdom, and Canada," November 15, 1945; U.S. Department of State Historical Office, Bureau of Public Affairs, Washington, D.C., Documents on Disarmament 1945–1959), http://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/publications/documents_on_disarmament/1945-1956/DoD_1945-1959_VOL_I.pdf.

The focus on effectiveness will become more critical in coming years, in part because the demand for nuclear verification can be expected to increase with the expansion of nuclear power into more NPT non-nuclear-weapon states.

number of states that have acquired nuclear weapons since the establishment of the regime. The *effectiveness* of the non-proliferation regime depends, first of all, on the viability of the legally binding arrangements, most of all on the viability of the NPT and NWFZ treaties. In practice, it also depends on the ability of the IAEA to detect non-compliance—in other words, on the effectiveness of IAEA verification. Beyond verification, the regime's effectiveness depends on the international community to appropriately respond to a finding of a state's non-compliance with its safeguards obligations. This paper addresses the topic of the effectiveness of IAEA verification. The defining and reporting of non-compliance by the IAEA Board of Governors, although very important, is not considered in this paper. The term "safeguards effectiveness" means, for the purpose of this paper, the effectiveness of IAEA verification—that is, the ability of the IAEA to detect non-compliance.

Effectiveness of IAEA Safeguards

There is no official or formal definition of "safeguards effectiveness" used by the IAEA. Most generally, effectiveness of IAEA safeguards might be defined as the IAEA's ability to detect non-compliance of a state with its obligations under its safeguards agreement concluded with the IAEA. In the course of performing routine verification activities, IAEA inspectors may detect indications of possible non-compliance. In such a case, further IAEA investigations and consultations with the state would be carried out in order to conclude whether or not a formal non-compliance finding by the IAEA is warranted. In its annual safeguards implementation report, the IAEA makes a statement about states' compliance with their safeguards obligations that rests upon the implied effectiveness of verification procedures that were implemented and, consequently, upon the ability of the IAEA Secretariat to detect indications of non-compliance.

In order to ensure that its verification is effective, the secretariat has established, pursuant to safeguards objectives defined in safeguards agreements, its technical verification objectives; it has also has developed verification procedures to meet the objectives. The *IAEA Safeguards Glossary, 2001 Edition* says that for the purposes of an evaluation, "effectiveness" is the "extent to which the IAEA's implementation of safeguards is able to achieve the safeguards objectives."[2] For states with CSAs, the glossary says the safeguards objective is to ensure that safeguards are applied to "all nuclear material in the state for the exclusive purpose of verifying that such material is not diverted to nuclear weapons or other nuclear explosive devices."[3]

[2] *IAEA Safeguards Glossary, 2001 Edition* (International Nuclear Verification Series No. 3), Paragraph 12.23, "Safeguards effectiveness evaluation," https://www.iaea.org/sites/default/files/iaea_safeguards_glossary.pdf.

[3] *IAEA Safeguards Glossary, 2001 Edition* (International Nuclear Verification Series No. 3), Paragraph 2.1, "Objectives of IAEA safeguards," https://www.iaea.org/sites/default/files/iaea_safeguards_glossary.pdf.

This statement is consistent with the language of the CSA itself, in Paragraph 2.

For many years before 1991, when a clandestine nuclear weapons program in Iraq was uncovered, the IAEA used "inspection goal attainment" as its key benchmark for safeguards effectiveness, evaluated against universal safeguards criteria established by the IAEA for prescribed inspection activities at specific types of nuclear facilities in all states. Before 1991, the inspection goals for Iraq were 100 percent attained, implying that IAEA safeguards were considered effective.

Following the exposure of Iraq's clandestine nuclear weapons program, during the 1990s the IAEA began making a transition from "traditional" safeguards, focused on declared materials and activities in a state, to safeguards for the "state-as-a-whole," in part to detect nuclear materials and activities that might not declared by a state. In this process, the IAEA has reconsidered "effectiveness," and it introduced measures to strengthen safeguards, including in the pursuit of greater effectiveness. The IAEA Secretariat and member states have studied, discussed, and debated how to best define, evaluate, and improve safeguards effectiveness, with the intention of establishing a new safeguards evaluation methodology and new benchmarks. This process has generated considerable literature about safeguards effectiveness, but no firm consensus has yet emerged about what safeguards effectiveness is and how it can be best evaluated. Benchmarks for evaluating the effectiveness of strengthened safeguards, to complement or replace the IAEA's prescriptive and universally applied criteria for traditional safeguards, have yet to be established.

With the advent of the implementation of the state-level concept (SLC) of safeguards at the beginning of this century, the IAEA has been departing from the use of the criteria and hence from established benchmarks for measuring safeguards effectiveness. Instead, and increasingly, the IAEA selects an array of specific safeguards procedures for each state, chosen on the basis of a complex analysis of each state's nuclear fuel cycle and related technical capabilities.

For implementation under strengthened safeguards, the IAEA has also enlarged the scope of its generic and technical safeguards objectives. In response to the revelations of Iraq's clandestine nuclear activities, the IAEA introduced an important and challenging new generic objective: "to detect undeclared nuclear material and activities." The introduction of the new objective required a new implementation concept; this was first designated as "integrated safeguards," and then as the SLC for safeguards implementation. Under this method, state-level approaches (SLAs) for individual states are established to implement the new enlarged objectives. Whereas under "traditional" safeguards, "diversion" had meant initially the undeclared removal of safeguarded nuclear materials from a safeguarded facility,[4] under the SLC, the focus has shifted toward Paragraphs 1 and 2 of the CSA, referring to any "diversion of nuclear material to manufacture of nuclear weapons or other explosive devices." In other words, diversion is considered as a state-level process that may involve not only declared nuclear materials and facilities but also undeclared nuclear materials and activities.

Since 2005, under the SLC there are three generic objectives for CSA implementation, which are the foundation for considerations of effectiveness: (a) to verify whether the declared nuclear materials have has been adequately accounted for,[5] (b) to detect undeclared production or processing of nuclear materials at declared facilities, and (c) to detect undeclared nuclear materials and activities in the state as a whole. Following from these generic objectives, for each state with a CSA, a safeguards plan is defined and implemented. The basis of the safeguards plan is an evaluation of the state's

[4] IAEA, *The Agency's Safeguards* (INFCIRC/26), March 30, 1961, https://www.iaea.org/sites/default/files/publications/documents/infcircs/1961/infcirc26.pdf.

[5] This objective was initially described as: "To detect diversion of declared nuclear material."

nuclear fuel cycle and related capabilities, culminating in an acquisition path analysis for the state that identifies, characterizes, and prioritizes plausible routes for the state to acquire weapons-usable nuclear materials. Using this state-specific analysis as a departure point, the IAEA develops for each state detailed technical objectives, corresponding to indicators of diversion that are derived from the analysis of the state's nuclear capabilities; the indicators are selected to facilitate the discovery of possible indications of nuclear materials diversion. On the basis of these technical objectives, the IAEA develops and implements verification procedures for each state that, collectively, should have a high probability to detect a diversion in that state or attain the state's technical safeguards objectives.

For any given state with a CSA, the SLA should represent the IAEA's optimum verification tool to verify compliance with the state's obligations under its safeguards agreement. The question is: How effective is this tool? For a state with a CSA, the effectiveness of its SLA might depend on the following key attributes:

- The completeness of the set of indicators of diversion—in other words, the completeness of the set of detailed technical objectives selected.
- The appropriateness of the verification procedures designed to attain the detailed technical objectives.
- The robustness of the process of optimizing safeguards in the state, given cost and resource restraints.
- The overall results of SLA implementation—that is, the attainment of the technical objectives and the performance of follow-up activities to address any anomalies and indicators detected.

Under traditional safeguards, for any state with a CSA, the set of indicators of diversion could not be complete because consideration of the presence of undeclared nuclear materials and activities was not included in the set of traditional safeguards objectives.

© Katy Laffan

Measuring radiation from a container storing nuclear materials.

Why Effectiveness Evaluation Matters

Under the SLC, the IAEA is turning away from decades of nearly exclusive focus on routine accountancy procedures underpinned by the assumption that states could be trusted to declare all their nuclear materials and activities; it is also modifying or abandoning strictly non-discriminatory yardsticks that considered effectiveness according to whether universal inspection goals were met. In the current SLC context, measuring safeguards effectiveness is a far more complex undertaking because each state is subject to a unique analysis relying on more information and more kinds of information, which need to be assessed and contextualized. Compared with criteria-based safeguards, state-level safeguards implementation includes a more qualitative element and may therefore imply greater potential risk. In order to conclude whether an SLA represents an effective and efficient verification tool for a given state, the IAEA should be able to evaluate the attributes in the above four bullet points. Presently, it does not have a method to do so and is not in a position, therefore, to perform such an evaluation. However, such an evaluation is important to help ensure the effectiveness of IAEA safeguards implementation.

The focus on effectiveness will become more critical in coming years, in part because the demand for nuclear verification can be expected to increase with the expansion of nuclear power into more NPT non-nuclear-weapon states. Also, technology for nuclear power generation and for the nuclear fuel cycle is evolving toward industrial-scale deployment of more advanced and in some cases more sensitive technologies.

Consensus about metrics for evaluating safeguards effectiveness will require broad agreement about the definitions of such terms as "safeguards effectiveness," "safeguards objectives," "technical objectives," and "diversion of nuclear material." Both Russian and American experts have contributed to the extensive literature on these topics.[6]

The quest for better understanding, guidance, and, eventually, procedures to enhance and evaluate effectiveness needs to be more highly prioritized by the IAEA Secretariat and member states. Given the likely stress on the IAEA safeguards system in coming years, in view of growing demand, cost pressures, and the analytical requirements of the SLC, insufficient attention to the evaluation of safeguards effectiveness might contribute to a safeguards failure event that could lead to serious proliferation consequences and precipitate loss of confidence in IAEA safeguards and multilateral nuclear verification.

[6] See, for example, Valery Bytchkov, Vladimir Ryzhikov, and Yuri Stepin, "A Method to Evaluate the Effectiveness of IAEA Safeguards" (Center for Energy and Security Studies, April 2021) http://ceness-russia.org/data/doc/21-04-12_A_Method_to_Evaluate_the_Effectiveness_of_IAEA_Safeguards_ENG_Final.pdf. Bytchkov et al. propose that the IAEA create a library of detailed technical objectives and corresponding verification procedures and that each verification procedure be assigned a probability of detecting the corresponding indication of diversion, established based on expert judgment. This method would permit assessment of the four SLA attributes listed above in the text. A performance indicator, suggested in the paper, could serve as a measure of SLA effectiveness. A paper presented by Dunbar Lockwood, Mark Goodman, and J. Steven Adams, "The Role of Performance Targets in Safeguards" at the IAEA Symposium on International Safeguards (Vienna, November 5–8. 2018, https://media.superevent.com/documents/20181023/5af1cd731929cbff2feb51cc771efbb2/book-of-abstracts.pdf) argues for the establishment of SLC performance targets and discusses ways to define the technical effectiveness of verification efforts and also the challenge of developing metrics for effectiveness in detecting undeclared activities.

20 YEARS OF NTI BUILDING A SAFER WORLD

Center for Energy and Security Studies
www.ceness-russia.org

Nuclear Threat Initiative
1776 Eye Street NW, Suite 600
Washington, D.C. 20006
www.nti.org

@NTI_WMD nti_wmd Facebook.com/nti.org

Center for Energy and Security Studies
(Центр энергетики и безопасности)
Mosfilmovskaya Str., 42, Bldg. 1
Moscow, Russia 119285
www.ceness-russia.org