

**AN ASSESSMENT OF  
SELECTED DIVISIONS OF THE  
INFORMATION TECHNOLOGY LABORATORY  
AT THE NATIONAL INSTITUTE OF STANDARDS  
AND TECHNOLOGY**

**FISCAL YEAR 2021**

Panel on Assessment of Selected Divisions of the  
National Institute of Standards and Technology (NIST) Information Technology Laboratory

Committee on NIST Technical Programs

Laboratory Assessments Board

Division on Engineering and Physical Sciences

A Consensus Study Report of

*The National Academies of*  
**SCIENCES • ENGINEERING • MEDICINE**

THE NATIONAL ACADEMIES PRESS

*Washington, DC*

[www.nap.edu](http://www.nap.edu)

**THE NATIONAL ACADEMIES PRESS**

**500 Fifth Street, NW**

**Washington, DC 20001**

This activity was supported by Contract No. SB13411CQ0017 with the National Institute of Standards and Technology. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-09256-2

International Standard Book Number-10: 0-309-09256-6

Digital Object Identifier: <https://doi.org/10.17226/26354>

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; <http://www.nap.edu>.

Copyright 2021 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2021. *An Assessment of Selected Divisions of the Information Technology Laboratory at the National Institute of Standards and Technology: Fiscal Year 2021*. Washington, DC: The National Academies Press.  
<https://doi.org/10.17226/26354>.

*The National Academies of*  
**SCIENCES • ENGINEERING • MEDICINE**

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at [www.nationalacademies.org](http://www.nationalacademies.org).

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

**Consensus Study Reports** published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

**Proceedings** published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit [www.nationalacademies.org/about/whatwedo](http://www.nationalacademies.org/about/whatwedo).

**PANEL ON ASSESSMENT OF SELECTED DIVISIONS OF THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST) INFORMATION TECHNOLOGY LABORATORY**

PHILIP M. NECHES, NAE,<sup>1</sup> Entrepreneurs Roundtable Accelerator, *Chair*  
RUZENA K. BAJCSY, NAE/NAM,<sup>2</sup> University of California, Berkeley  
SUZANNE BAKKEN, Columbia University  
PETER A. BELING, Virginia Polytechnic Institute and State University  
JAMES O. BERGER, NAS,<sup>3</sup> Duke University  
DUANE BLACKBURN, MITRE Corporation  
PHILLIP COLELLA, NAS, Lawrence Berkeley National Laboratory  
DORIN COMANICIU, NAE, Siemens Healthcare  
JAMES DARNELL, Blackbourne Holding Corporation  
MARK EDWARD DEAN, NAE, University of Tennessee  
DON EUGENE DETMER, NAM, University of Virginia  
JUAN E. GILBERT, University of Florida  
GERARD J. HOLZMANN, NAE, Nimble Research  
PETER M. KOGGE, University of Notre Dame  
MING C. LIN, University of Maryland  
CHRISTOPHER MANNING, Stanford University  
MARI OSTENDORF, NAE, University of Washington  
JOSEPH MARC OVERHAGE, NAM, Anthem, Inc.  
STEPHANIE A. SCHUCKERS, Clarkson University  
PETER TU, General Electric Corp.  
STEPHEN B. VARDEMAN, Iowa State University  
ALYSON G. WILSON, North Carolina State University  
WING HUNG WONG, NAS, Stanford University  
MARY ELLEN ZURKO, MIT Lincoln Laboratory

***Staff***

AZEB GETACHEW, Senior Program Assistant  
EVA LABRE, Administrative Coordinator  
JAMES MCGEE, Director

---

<sup>1</sup> Member, National Academy of Engineering.

<sup>2</sup> Member, National Academy of Medicine.

<sup>3</sup> Member, National Academy of Sciences.

## COMMITTEE ON NIST TECHNICAL PROGRAMS

ELSA REICHMANIS, NAE, Lehigh University, *Chair*  
MICHAEL I. BASKES, NAE, Mississippi State University  
LEWIS BRANSCOMB, NAS/NAE/NAM, University of California, San Diego  
MARTIN E. GLICKSMAN, NAE, Florida Institute of Technology  
JENNIE S. HWANG, NAE, H-Technologies Group  
CHRISTOPHER W. MACOSKO, NAE, University of Minnesota  
C. KUMAR PATEL, NAS/NAE, Pranalytica, Inc.  
BHAKTA B. RATH, NAE, Naval Research Laboratory  
ALICE WHITE, Boston University

### *Staff*

AZEB GETACHEW, Senior Program Assistant  
EVA LABRE, Administrative Coordinator  
JAMES MCGEE, Director

## LABORATORY ASSESSMENTS BOARD

JENNIE S. HWANG, NAE, H-Technologies Group, *Chair*  
MICHAEL I. BASKES, NAE, Los Alamos National Laboratory  
KEVIN MACK GATES II, Strategic Analysis, Inc.  
WESLEY L. HARRIS, NAE, Massachusetts Institute of Technology  
W. CARL LINEBERGER, NAS, University of Colorado, Boulder  
AJAY P. MALSHE, NAE, Purdue University  
C. KUMAR N. PATEL, NAS/NAE, Pranalytica, Inc.  
ELSA REICHMANIS, NAE, Lehigh University  
FRANCISCO F. ROBERTO, NAE, Newmont Corporation

### *Staff*

AZEB GETACHEW, Senior Program Assistant  
EVA LABRE, Administrative Coordinator  
JAMES MCGEE, Director





## Acknowledgment of Reviewers

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report:

Sanjeev Arora, NAS,<sup>1</sup> Princeton University,  
Luis A. Escobar, Louisiana State University,  
William D. Gropp, NAE,<sup>2</sup> the University of Illinois, Urbana-Champaign,  
Kevin B. Johnson, NAM,<sup>3</sup> Vanderbilt University School of Medicine,  
Peter K. Lee, NAM, Microsoft Corporation,  
Kyran D. Mish, Sandia National Laboratories,  
Padma Raghavan, Vanderbilt University, and  
Bin Yu, NAS, University of California, Berkeley.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report nor did they see the final draft before its release. The review of this report was overseen by Robert F. Sproull, NAE, University of Massachusetts, Amherst. He was responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

---

<sup>1</sup> Member, National Academy of Sciences.

<sup>2</sup> Member, National Academy of Engineering.

<sup>3</sup> Member, National Academy of Medicine.



## Contents

SUMMARY	1
1 THE CHARGE TO THE PANEL AND THE ASSESSMENT PROCESS	4
2 INFORMATION ACCESS DIVISION	6
3 SOFTWARE AND SYSTEMS DIVISION	14
4 STATISTICAL ENGINEERING DIVISION	24
5 CROSSCUTTING CONCLUSIONS AND RECOMMENDATIONS	31
APPENDIXES	
A Acronyms	35
B Biographical Sketches of Panel Members	37



## Summary

At the request of the Director of the National Institute of Standards and Technology (NIST), in 2020 the National Academies of Sciences, Engineering, and Medicine formed the Panel on Assessment of Selected Divisions of the National Institute of Standards and Technology (NIST) Information Technology Laboratory and established the following statement of task for the panel:

The National Academies shall appoint three panels to assess independently the scientific and technical work performed by the National Institute of Standards and Technology (NIST) Information Technology Laboratory, Physical Measurement Laboratory, and Center for Neutron Research. Each panel will review technical reports and technical program descriptions prepared by NIST staff and will visit the facilities of their respective NIST laboratory. Visits will include technical presentations by NIST staff, demonstrations of NIST projects, tours of NIST facilities, and discussions with NIST staff. Each panel will deliberate findings in closed session panel meetings and will prepare a separate report summarizing its assessment findings. The Panel on Assessment of Selected Divisions of the Information Technology Laboratory at the National Institute of Standards and Technology will review the following divisions of the NIST Information Technology Laboratory: Information Access, Software and Systems, and Statistical Engineering. This panel will not access restricted information; the report summarizing its assessment will contain only public release information.

The NIST Director requested that the panel focus its assessment on the following factors:

1. The organization's technical programs;
2. The portfolio of scientific expertise within the organization;
3. The adequacy of the organization's facilities, equipment, and human resources; and
4. The effectiveness with which the organization disseminates its program outputs.

To accomplish the assessment, the National Academies assembled a panel of 24 volunteers whose expertise matched that of the work performed by the Information Technology Laboratory (ITL) staff.<sup>1</sup> On June 21-24, 2021, the panel conducted a virtual review (via Internet media) of the Information Access Division (IAD), the Software and Systems Division (SSD), and the Statistical Engineering Division (SED). During a plenary session, the panel received overview presentations by the acting NIST Director and the Director of the ITL. Subsequently, the panel spent approximately 1.5 days receiving presentations from and engaging in discussions with the staff at the three divisions reviewed. On the third day, the panel met in a closed session to deliberate on its findings and to define the contents of this assessment report. The panel met with NIST management on the fourth day to clarify open questions. The choice of projects to be reviewed was made by the ITL. The panel applied a largely qualitative approach to the assessment. Given the non-exhaustive nature of the review, the omission in this report of any particular ITL project should not be interpreted as a negative reflection on the omitted project. Crosscutting conclusions and recommendations are presented in this Summary. Additional conclusions and recommendations specific to individual ITL divisions are also presented in the body of this report.

---

<sup>1</sup> See the NIST Information Technology Laboratory at <https://www.nist.gov/itl>, accessed May 17, 2021.

## TECHNICAL QUALITY OF THE WORK

The technical quality of the work is generally excellent, addressing the assigned mission areas. IAD, SSD, and SED continue to make significant contributions to address the ITL and NIST missions, national needs, and the needs of government, industry, and academic stakeholders. Amidst continuing technological and societal changes, there are opportunities for increased collaborations across projects to address common challenges in such areas as artificial intelligence, machine learning, health information technology, data science, and statistical methodology.

Trustworthy AI is an open problem. IAD has done a commendable job in laying out cross-cutting, high-level elements of trustworthiness, and there are additional cross-cutting contributions that it can make in terms of guidelines. However, standards will be highly dependent on the application area, and getting community trust depends on incorporating communities and their input at the design phase.

## TECHNICAL EXPERTISE OF THE STAFF AND ADEQUACY OF STAFFING

Technical staff generally possess expertise adequate to perform their task. Some are among the best in the world in their areas of research. However, evolving needs in some areas (e.g., AI, ML, and data science) will produce expertise gaps that will have to be filled.

In terms of human resources, recruitment of highly qualified full-time staff has been identified as an overall challenge at ITL due to competition with industry and academia, which offer salaries that exceed the limits available to the ITL. It may be difficult for small teams to be effective due to lack of critical mass in some areas. The relatively large proportion of the permanent workforce at retirement age in some areas raises the potential need for succession planning to ensure maintenance of competence in core areas while meeting the demand for new areas of competency. Current and future ITL needs cannot be met without the addition of new competencies to complement existing strengths. New hires, to a large extent, will define the future and new activities of the SED.

The NIST Domestic Guest Researcher Program provides access for technically qualified U.S. citizens to NIST facilities and equipment while working with NIST staff on projects of mutual interest. Exchange programs would provide additional opportunities for collaboration and could serve as additional means of enticing recruits to work at NIST.

**RECOMMENDATION: ITL should apply an aggressive, imaginative focus on hiring to replace retiring staff, to address important growth areas such as artificial intelligence, machine learning, and data science, and to fill specific gaps in the divisions. This effort should aspire to diversity targets.**

**RECOMMENDATION: ITL should plan and implement effective ways to recruit and retain a diverse workforce to ensure the appropriate staffing in areas of significant interest to national welfare and security, and to address severe competition from industry in areas such as artificial intelligence, cybersecurity, and the Internet of Things.**

**RECOMMENDATION: ITL should establish exchange programs with relevant government laboratories, academic institutions, and industry consortia to stimulate new ideas and problem areas, enhance competencies, and facilitate collaboration.**

Properly managed, ITL's diversity, equity, and inclusion strategy can add a helpful element to the recruiting process. The staff has performed commendably during the COVID-19 pandemic, collaboratively addressing critically important responses to the pandemic.

## ADEQUACY OF FACILITIES AND EQUIPMENT

Post-pandemic planning gives NIST opportunities to continue some remote work, making NIST a more attractive environment for staff and giving NIST a needed edge in recruiting. Also, creative thinking about the new work environment can lead to more productive use of facilities, which could be considered during NIST's ongoing laboratory and office renovations.

The ITL laboratory facilities are generally adequate and support well the activities of the divisions. There is a critical need, however, for improved computing capabilities that are needed to support complex computation and analysis. This is particularly the case for the growing number of projects that involve state-of-the-art machine learning models, which call for ever larger training sets and computational resources.

**RECOMMENDATION: ITL should take steps to insure adequate resources, especially computing to support AI/ML and data science at sufficient scale.**

**RECOMMENDATION: To get access to the most modern resources, ITL should seek collaborations with other organizations in the public and private sectors, including other Government agencies. To achieve collaborative access, the ITL should examine its potential contributions to partnerships.**

## EFFECTIVENESS OF THE DISSEMINATION OF OUTPUTS

Each division disseminates its outputs widely, but the divisions vary in the relative emphasis placed on dissemination vehicles (e.g., publications, workshops, data repositories, standard reference data, and educational programs). Careful, systematic, and continuing analysis of the needs of specific stakeholder communities would improve the effectiveness of the dissemination of ITL's outputs.

**RECOMMENDATION: ITL should broaden its impact to non-technical stakeholders, policy makers, and the public.**

**RECOMMENDATION: ITL should improve messaging aimed at non-technical audiences such as policy makers, media, and the general public for the outputs of the Information Access, Software and Systems, and Statistical Engineering Divisions.**

This would benefit NIST by providing greater acceptance of and support for NIST efforts.

## The Charge to the Panel and the Assessment Process

At the request of the National Institute of Standards and Technology (NIST), the National Academies of Sciences, Engineering, and Medicine has, since 1959, annually assembled panels of experts from academia, industry, medicine, and other scientific and engineering environments to assess the quality and effectiveness of the NIST measurements and standards laboratories, of which there are now six,<sup>1</sup> as well as the adequacy of the laboratories' resources.

At the request of the Director of NIST, in 2020 the National Academies formed the Panel on Assessment of Selected Divisions of the National Institute of Standards and Technology (NIST) Information Technology Laboratory and established the following statement of task for the panel:

The National Academies shall appoint three panels to assess independently the scientific and technical work performed by the National Institute of Standards and Technology (NIST) Physical Measurement Laboratory, Information Technology Laboratory, and Center for Neutron Research. Each panel will review technical reports and technical program descriptions prepared by NIST staff and will visit the facilities of their respective NIST laboratory. Visits will include technical presentations by NIST staff, demonstrations of NIST projects, tours of NIST facilities, and discussions with NIST staff. Each panel will deliberate findings in closed session panel meetings and will prepare a separate report summarizing its assessment findings. The Panel on Assessment of Selected Divisions of the Information Technology Laboratory at the National Institute of Standards and Technology will review the following divisions of the NIST Information Technology Laboratory: Information Access, Software and Systems, and Statistical Engineering. This panel will not access restricted information; the report summarizing its assessment will contain only public release information.

The NIST Director requested that the panel focus its assessment on the following factors:

1. The organization's technical programs;
2. The portfolio of scientific expertise within the organization;
3. The adequacy of the organization's facilities, equipment, and human resources; and
4. The effectiveness with which the organization disseminates its program outputs.

The context of this technical assessment is the mission of NIST, which is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve the quality of life. NIST laboratories conduct research to anticipate future metrology and standards needs, enable new scientific and technological advances, and improve and refine existing measurement methods and services.

NIST specified that three of the six divisions of the Information Technology Laboratory (ITL) would be reviewed: the Information Access Division (IAD), the Software and Systems Division (SSD),

---

<sup>1</sup> The six NIST laboratories are the Engineering Laboratory, the Physical Measurement Laboratory, the Information Technology Laboratory, the Material Measurement Laboratory, the Communication Technology Laboratory, and the NIST Center for Neutron Research.



and the Statistical Engineering Division (SED). The following ITL divisions were not reviewed because they had previously been reviewed (NIST may request their review in the future): the Applied and Computational Mathematics Division, the Computer Security Division, the Applied Cybersecurity Division, and the Advanced Network Technologies Division. In order to accomplish the assessment, the National Academies assembled a panel of 24 volunteers, whose expertise matches that of the work performed by ITL staff.<sup>2</sup>

On June 21-24, 2021, the panel conducted a virtual review (via Internet media). During a plenary session, the panel received overview presentations by the acting NIST Director and the Director of the ITL. Subsequently, the panel spent approximately 1.5 days receiving presentations from and engaging in discussions with the staff at the three divisions reviewed. These presentations and discussions were attended by the following three separate teams of panel members, with the chair, Philip Neches, attending portions of each session:

- For the IAD: Ruzena K. Bajcsy, Duane Blackburn, Stephanie A. Schuckers, Peter Tu, Mari Ostendorf, Christopher Manning, Mary Ellen Zurko, and Juan E. Gilbert;
- For the SSD: Mark Edward Dean, Suzanne Bakken, Don Eugene Detmer, J. Marc Overhage, Peter M. Kogge, Dorin Comaniciu, Phillip Colella, Ming C. Lin, Gerard J. Holzmann, and James Darnell;
- For the SED: Stephen B. Vardeman, Alyson G. Wilson, Janes O. Berger, Wing Hung Wong, and Peter A. Beling.

On the third day, the full panel met in a closed session to deliberate on its findings and to define the contents of this assessment report. The full panel met with NIST management on the fourth day to clarify open questions.

The panel's approach to the assessment relied on the experience, technical knowledge, and expertise of its members. The panel reviewed selected examples of the technical research performed at ITL; because of time constraints, it was not possible to review ITL programs and projects exhaustively. The examples reviewed by the panel were selected by ITL. The panel's goal was to identify and report salient examples of accomplishments, challenges, and opportunities for improvement with respect to the factors suggested above by the Director of NIST. These examples are intended collectively to portray an overall impression of the laboratory, while preserving useful suggestions specific to the projects and programs that the panel examined. The panel applied a largely qualitative, rather than quantitative, approach to the assessment.

Given the necessarily broad but not exhaustive nature of the review, omission in this report of any particular ITL program or project should not be interpreted as implying any negative reflection on the omitted program or project.

---

<sup>2</sup> See the NIST Information Technology Laboratory homepage for information on organization and programs at <https://www.nist.gov/itl/about-itl/>, accessed May 17, 2021.

## **Information Access Division**

The mission of the Information Access Division (IAD) is to foster trust in technologies that make sense of complex information relating to human action, human behavior, and human characteristics. The IAD includes the following four groups: the Image Group, Multimodal Information Group, Retrieval Group, and Visualization and Usability Group. IAD is an effective organization with prodigious output in terms of reports, publications, standards development, and other outputs.

### **TECHNICAL QUALITY OF THE WORK**

IAD personnel provide leadership in various information technology (IT) fields through a collaborative process that helps define and quantify problems and then build communities of interest that can address them. Some of the efforts lead to new evaluation metrics or data sets and proposed standards. IAD has led in expanding NIST's approaches beyond traditional metrology to understanding human factors in IT success and failures. Overall, IAD research is of very high quality.

IAD benefits from broad community interactions and pushes into emerging areas involving support of other agencies. Through collaborative processes, IAD has visibly driven the creation or advancement of important technical fields. IAD has facilitated benchmarking, using best known methods for measurement to accurately assess the state of the art in several fields.

### **Assessment of Individual Projects**

NIST Information Technology Laboratory (ITL) voting work, centered on the development of the Voluntary Voting Systems Guidelines (VVSG), including the much-desired 2021 release of VVSG 2.0 for human factors, accessibility, and usability, is an example of a NIST project requiring assembling expertise from multiple computer science divisions and working with multiple outside constituencies, including community outreach via public working groups. The division contributed high-quality human factors technical expertise in improving the accessibility and usability of voting systems with critical work on improving system understandability by poll workers and work on providing equal access to all voters, including voters with disabilities, through the use of universal design methods. While voting system security receives more public focus, these human-computer interaction aspects are equally vital to trouble-free and fair elections.

The Retrieval Group and its longstanding flagship Text Retrieval Conference (TREC) are the best-in-the-world group for evaluating information access from unstructured data. An IAD TREC organizer recently received external recognition as a fellow of the Association for Computing Machinery (ACM) and of the Washington Academy of Sciences, and by election to ACM SIGIR's (Special Interest Group on Information Retrieval's) Academy as well as internal recognition by appointment as a NIST fellow in 2021. In the past 2 years, the group has continued to innovate with well-chosen new evaluation foci, such as Health Misinformation and Fair Ranking, and there are plans for a new Trustworthy AI

Conference (TRUC), suggested at a workshop at the National Academies of Sciences, Engineering, and Medicine.

Sister benchmark evaluations to TREC take place under the Text Analysis Conference (TAC) and the video retrieval evaluation tracks of TREC (TRECvid). TAC defines and runs evaluations for advanced natural language processing technologies such as question answering, knowledge base population, and summarization. This activity has played a beneficial role in codifying new NLP tasks and supporting evaluations for the Defense Advanced Research Projects Agency (DARPA) and the National Institutes of Health, although perhaps not achieving the standout recognition of TREC. TRECvid has jumpstarted a large amount of academic research in content-based video retrieval, for which the researchers received an Institute of Electrical and Electronics Engineers (IEEE) PAMI prize in 2018. In 2021, TRECvid supported the following six major tracks: ad hoc video search, activities in extended video, instance search, video to text, video summarization, and disaster scene description and indexing. Disaster scene description is an example of IAD focusing its resources in a way that exploits government data sources, focuses on an area with lagging performance not of commercial interest, and builds important national capabilities.

The Multimodal Information Group has a long history and deep expertise in evaluating spoken language and multimedia technologies, machine translation, and knowledge representation. It has an excellent reputation for statistical rigor in the design and implementation of evaluations, which has led to a large demand from outside agencies. It has a decades-long history of important contributions to evaluation of speech technology, which has helped drive the tremendous advances made in speech and speaker recognition. The work has evolved to include language processing, with considerable impact on machine translation, leading to important efforts on cross-lingual language processing tasks. While many of these efforts are initiated by outside agencies, work continues with open evaluations in several areas.

With the goal of developing a secured reference architecture for big data applications, the Big Data Public Working Group has focused on big data definitions and taxonomies, use cases and requirements, security and privacy, specification of a reference architecture, and a standards roadmap. A comprehensive list of 437 requirements were extracted from 51 use cases, and 35 aggregated general requirements have been divided into 7 categories. Various NIST Big Data Interoperability Framework (NBDIF) documents describing the proposed reference architecture, its interfaces, and the issues associated with adoption and modernization have been published. Five out of seven of the NBDIF documents have become ISO/IEC standards.

The group working on explainable artificial intelligence (AI) has released a white paper and has conducted a NIST workshop with the goal of endowing AI algorithms with the capacity to provide explanations regarding their various inferences and outputs. Principles of explainable AI have been proposed and are in the process of garnering community consensus. Connections between users, policy makers, industry, and academia have been nurtured. This is a prime example of NIST's capacity to provide thought leadership.

The Image Group continues to provide best-in-the-world technology evaluations and to support the development of a variety of biometric technologies. The group has long hosted multiple highly impactful evaluations in face and iris recognition, with worldwide participation and impact. The face recognition studies benchmark an order of magnitude decrease in error rates over the past 3 years. The group's prominent study on differential performance of face recognition across age, ethnicity, and gender has been presented approximately 160 times in multiple venues (industry and academic forums, congressional testimony, press coverage), including being featured on the CBS show *60 Minutes*. For fingerprints, technologies associated with credentialing, law enforcements, and forensics have been supported via standardization of template representations as well as routines associated with fingerprint image segmentation. The use of standardized minutia records has resulted in a significant reduction in storage and truly effortless swapping between formats. The group also supports technologies associated with iris recognition, contactless fingerprint capture, and tattoo recognition. All of these efforts are supported by a high-performance computing infrastructure at the Biometrics Research Laboratory (BRL).

The group working on biometric forensic science continues to advance this frontier of forensic science capabilities. With the goal of anchoring the latent fingerprint community, the team has established a new latent fingerprint dataset with 200 subjects and are working on minutiae correspondences. The team has performed experiments with human examiners focused on facial imagery captured in uncontrolled circumstances. A critical finding is again evidence of cross-race differential bias. The team has also conducted experiments to support the efficacy of forensic iris recognition without the need for specialized image capture devices. The societal and legal implications of this work are significant.

A collection of public safety communication research projects has focused on the technology needs of the first responder community. A survey with 7,182 respondents clearly identifies the communication needs of this community. Automated stream analysis methods have been vetted, and mechanisms that allow end users to evaluate such technology using their data in a secure manner have been developed. This illustrates IADs drive to enable the development of technology and support its adoption by society.

IAD's usable security research is prominent within the security research community, and it has the potential for broader impacts beyond experts. For example, the researchers have established the concept of security fatigue. Their Smart Home work led to a recent publication in a top-tier cybersecurity venue. Fifteen government and business organizations expressed interest in evaluating or implementing Phish Scale, a method for rating human phishing detection difficulty; the method has also appeared in academic journals and the technology press. NIST Special Publication 800-63,<sup>1</sup> which includes a data-driven usability chapter covering this work, has been downloaded nearly 420,000 times since its update in June 2017. This publication is well known in both usable security research and password compliance communities.

### Assessment Across the Division

It does not appear that IAD has specific, defined objectives to measure success against. Rather, IAD has a purpose statement and a collection of individual projects. IAD is under-focusing on strategic and stakeholder communications aspects of technical program management.

**RECOMMENDATION: IAD needs to define answers to the following questions: Who are our stakeholders and what do they need? How do we organize projects (individually and collectively) to meet those needs? How do we assess how well we're meeting those needs? How are we communicating our findings to those entities in understandable and usable ways?**

Bias in AI is becoming increasingly recognized as a key issue in trustworthiness and is an opportunity for IAD. Bias is only one aspect of trustworthy AI—the entire set of considerations of trustworthy AI needs to be embraced by NIST, including a quality-controlled process of making standards using data and in context of real-world deployments, with stability and robustness to the human judgment calls as a central concern.

IAD has made commendable recent investments in social science expertise, leading to the very recent publication of a NIST Special Publication on identifying and managing bias in AI,<sup>2</sup> and NIST has done definitive work in characterizing demographic differentials in face recognition. However, other examples of differential performance that have been highlighted for voice and language have not yet been addressed; IAD needs to take a more proactive role in providing data that facilitates more such analyses.

---

<sup>1</sup> National Institute of Standards and Technology (NIST), 2017, *Digital Identity Guidelines*, SP 800-63-3, June, <https://pages.nist.gov/800-63-3>.

<sup>2</sup> NIST, 2021, *A Proposal for Identifying and Managing Bias in Artificial Intelligence*, Draft NIST Special Publication 1270, June, <https://doi.org/10.6028/NIST.SP.1270-draft>.

Moreover, at this stage, IAD seems insufficiently engaged with bias and fairness stakeholders, such as groups who participate in the ACM FaCCT (Fairness, Accountability, and Transparency) community. There is still a need to more extensively complement IAD's excellent technical metrology work with social science expertise focusing on the societal dangers of potentially biased AI tools. While IAD has turned the corner in valuing qualitative research, there is still more to do here.

Trustworthy AI is an open problem. IAD has done a commendable job in laying out cross-cutting, high-level elements of trustworthiness, and there are additional cross-cutting contributions that it can make in terms of guidelines. However, standards will be highly dependent on the application area, and getting community trust depends on incorporating communities and their input at the design phase. Even if the division grows, the number of people in IAD working on this will be limited, and it will be important to focus efforts, taking into account both stakeholder needs and areas of IAD expertise.

IAD is well known for its work on the design and implementation of open evaluations for technologies grounded in human characteristics and behavior. As individual research groups, companies, and collectives of researchers are increasingly posting open challenges and shared tasks to evaluate related technologies, IAD will need to continue to evolve to maintain its leadership. The Image Group has organized online and ongoing evaluations whereby participants can submit at any time and a leaderboard tracks performance. The Retrieval and Multimodal Information Groups also need to move in this direction, where appropriate, in areas such as old data sets or already well-defined tasks or opportunities making for data sets openly downloadable. The IAD could do more to extend the reach of its metrology expertise to the broader community—by consulting, hosting tutorials, or playing an important role in establishing best practices for setting up informative evaluations.

Human-agent interaction and human-in-the-loop systems represent a growing area of AI, particularly for speech, language, and multimodal technologies. Evaluation in this space is an open problem, and there is an opportunity for IAD to have an impact here by drawing on the expertise in multiple groups. It would be a significant undertaking if staff are already spread thin, so where to focus needs to be carefully considered.

## TECHNICAL EXPERTISE OF THE STAFF

IAD staff has done an excellent job in building the division research projects. They have published that research broadly, including at top-tier venues. The face and fingerprint recognition technology evaluation within the biometrics group is one of the best in the world. Research teams bring together social scientists, engineers, and computer scientists, producing stronger research through this combination of disciplines. In the past, the social scientists found it a challenge to explain the importance of qualitative research to their engineer and scientist colleagues. However, now they are receiving requests for that expertise from several initiatives.

Many of the challenges faced by ITL's mission would benefit from broader application of the social science expertise of IAD. While that expertise is being leveraged in the trustworthy AI and public safety projects, it needs to be recognized and made use of more broadly across other projects. As other projects recognize their need for social science expertise, it seems likely that more demands will be put on those resources.

Given the importance of the research, IAD needs the ability to compete for top talent, who are offered generous salaries for their expertise in industry.

**RECOMMENDATION: IAD should explore compensating benefits such as flexibility and work/life balance.**

**RECOMMENDATION: IAD should consider more external exposure of its ongoing diversity, equity, and inclusion initiatives to benefit hiring.**

## ADEQUACY OF RESOURCES

IAD has unique facilities that support the breadth of its activities, including the BRL, the Usability Testing Laboratory, the Assessor Laboratory, and the Sequestered Data Testing Laboratory. In particular, the BRL is an excellent resource for the work involving controlled unclassified information. The Usability Testing Laboratory was probably underutilized during the pandemic, but it will be an important resource moving forward, particularly given that usability is important for trust in AI.

IAD researchers provide expertise in multiple disciplines, including mathematics, engineering, computer science, social science, and IT. This is critical for supporting the different work that those researchers do. In areas where they need additional expertise, IAD staff collaborate with staff in other divisions in ITL and across NIST. ITL is actively supporting continuous learning for staff members, including online learning and graduate work.

IAD significantly contributed to research in reports aimed at understanding shortcomings in the areas of diversity, equity, and inclusion (DEI) at NIST. IAD is responding by instituting new procedures for performance assessments, hiring, and leadership training. Since 2016, the percentage of women has increased to roughly a third of their staff, including women in leadership roles. They also have a collaboration with Morgan State University for recruiting Professional Research Experience Program students.

Work on visualization was missing from the presentations and division documentation provided to the panel, which raises the question of a gap in expertise of the Visualization and Usability Group. IAD leadership explained that this was a historical artifact associated with challenges in changing the name. It does not focus on visualization work, but instead relies on the expertise of colleagues in other divisions. Visualization is an effective means to communicate information in an accessible manner to most people. This gap needs to be filled at NIST.

Like many organizations, IAD recognizes that there will be challenges and opportunities in moving back to in-person work after an extended period of remote work during the pandemic. Allowing continued teleworking for some staff members may be attractive to some staff and helpful for recruiting and retention, and it offers the potential for expanding efforts within existing space. However, it will be important to consider ways to include remote workers in opportunities for socialization that are important for team building and generating new ideas.

There are growing needs in the areas of privacy, usable security, bias, and AI, and this work will be especially important in the next few years.

Given its expertise, IAD is already getting more requests than it can take on, and such requests are likely to increase. The groups are spread thin with the many projects they currently have.

**RECOMMENDATION: To maintain their high standards, IAD should prioritize projects with all of its stakeholders in mind.**

Expanding the projects that it takes on will require expanding the facilities and increased human resources.

There is increasing recognition that it is important to assess demographic differential performance. While differences in face recognition performance have attracted a lot of media attention, similar differences have been observed for speech recognition. Assessing such performance differences can involve tracking sensitive information. Further, as face and speaker recognition technology has improved, much audio and video data has become personally identifiable information and thus potentially sensitive.

**RECOMMENDATION: In order to continue its important work on audio and video data, IAD should expand the facilities for handling controlled unclassified information and should consider streamlining the privacy process for obtaining and providing access to sensitive data.**

In the vision and language processing communities, there is a trend of having leaderboards for assessing performance, since technology is evolving so quickly. IAD could beneficially consider whether it wants to provide more support for ongoing assessment, and if so, it is important to be forward thinking about computing resources.

All of ITL will be facing challenges in hiring, given the high demand for IT expertise in industry now and the fact that government salaries are not competitive. This will be a particular challenge for IAD if it is to expand its staff to meet the growing demands for its expertise in privacy, security, and trustworthy AI. It will be important to identify strategic advantages that IAD might have. Offering flexibility of telecommuting is important, but many companies are recognizing this. The support for continuing education, the potential for societal impact, and the development of an inclusive community are possible selling points. Building relationships with people by supporting interns, postdoctoral researchers, and visiting scholars can also support the recruiting process.

IAD has acknowledged that there are DEI issues to address. It has taken the first steps in doing so, but it would be helpful to develop an explicit strategy for recruiting and retention with measurable objectives.

The large percentage of other agency funding is an opportunity and a challenge, since it is risky to commit to permanent employees.

The BRL may need to be expanded to handle growing privacy concerns related to speech and video.

Providing reference data is an important aspect of standards for performance measurement. IAD plays a vital role in providing biometric data sets and in reporting demographic differences in system performance. Efforts in speech and multimedia standards and reference material could be expanded.

## EFFECTIVENESS OF DISSEMINATION OF OUTPUTS

IAD's technical products include the depth and specificity required to drive capability advancement, and there are numerous examples that demonstrate how its work positively impacts the nation.

IAD has identified "researchers who are developing technology" as its primary stakeholder, and their research and dissemination of findings are clearly driven by and support this group. Their work appears to be in the appropriate intersection of investigations that this community requires, and which is appropriate for NIST to perform. The range of outputs provided, such as published papers, data sets, technical briefings, tools, and guidelines, vary properly to meet the needs of individual technology areas so that IAD is positively impacting their advancements to support government and non-government needs. No substantive changes are necessary in these commendable efforts.

It is apparent that IAD staff inherently monitor how researchers use their outputs (as evidenced by the selective use of a range of outputs and how these evolve over time), but it does not appear that IAD has taken on such assessments formally. Doing so would provide insights to further improve the scoping of their activities and how results are presented, as well as provide management benchmarks on which to measure progress and IAD's influence on this stakeholder's priorities.

IAD identified Congress and policy makers as its secondary stakeholder and recipients, users, and consumers of technology as its third. For the most part, IAD's outputs are not driven by, nor are they effective for, these stakeholders. Core visualization and usability work is driven by stakeholders and users.

Indeed, these entities are usually required to analyze IAD's technical outputs themselves and determine their relevance and meaning, often resulting in misinterpretations, because these very technical outputs are not understandable by a variety of non-expert audiences.

**RECOMMENDATION: IAD should take action to better support nonexpert stakeholders, including policy makers and the general public.**

The gap in priority between IAD's primary and other stakeholders is not as wide as its current focus suggests, and it will further narrow in the future. Actions taken to address this important gap ought not diminish the technical depth and specificity of IAD's reports, which need to continue for the division to support capability advancement. Appendices within technical reports, or adjunct publications, devoted to meet the specific needs of each stakeholder group (policy makers in Congress or the White House and those that attempt to influence them, the public and press, technology-specific communities) need to be considered instead.

In all cases, measuring effectiveness and impact needs to be modernized and assessed against each stakeholder group. The metrics currently used are artifacts (numbers of papers published, awards received, or briefings provided) or are indicators of a community's advancement (performance improvements over time, more participants in technology evaluations, adoption of the technology) rather than IAD's specific influence. Developing and using metrics that assess IAD's impacts for each of its stakeholder constituencies will drive needed cultural changes, help overcome the division's strategic planning gaps, and provide data-driven evidence for use in promoting IAD to its stakeholders.

## GENERAL CONCLUSIONS

The work and staff across IAD can generally be characterized as excellent. IAD researchers address ITL objectives. The subject matter that IAD is charged with is of great importance to the nation, and IAD provides high-quality research and service to the nation. This effort will be enhanced if IAD can integrate the individual projects with a common thread and recruit the proper staff who can address IAD's evolving objectives. IAD has a purpose statement and a collection of outstanding individual projects. However, IAD is not adequately focusing on strategic and stakeholder communications and needs to identify stakeholder needs, organize projects to address those needs, communicate findings to stakeholders in understandable and usable ways, and assess how well it meets stakeholder needs.

The fact that there is no integrated objective does not give the whole IAD the credit that it otherwise deserves. For example, bias in AI is becoming increasingly recognized as a key issue in trustworthiness and is an opportunity for IAD. IAD has made commendable recent investments in social science expertise, leading to the publication of a NIST Special Publication on identifying and managing bias in AI, and IAD has done definitive work in characterizing demographic differentials in face recognition. However, other examples of differential performance that have been highlighted for voice, language have not yet been addressed, and IAD needs to take a more proactive role in providing data that facilitates more such analyses. IAD seems insufficiently engaged with bias and fairness stakeholders, such as groups who participate in the ACM FaCCT community. There is still a need to more extensively complement IAD's excellent technical metrology work with social science expertise focusing on the societal dangers of potentially biased AI tools. While IAD has turned the corner in valuing qualitative research, there is still more to do.

Trustworthy AI is an open problem. IAD has done a commendable job in laying out cross-cutting, high-level elements of trustworthiness, and there are additional cross-cutting contributions that IAD can make in terms of guidelines. However, standards will be highly dependent on the application area, and getting community trust crucially depends on incorporating communities and their input at the design phase. Even if the division grows, the number of people in IAD working on this will be limited, and it will be important to focus efforts, taking into account both stakeholder needs and areas of IAD expertise.

IAD is well known for its work on the design and implementation of open evaluations for technologies grounded in human characteristics and behavior. As individual research groups, companies, and collectives of researchers are increasingly posting open challenges and shared tasks to evaluate related technologies, IAD will need to continue to evolve to maintain its leadership. The Image Group has organized online and ongoing evaluations, whereby participants can submit at any time and a leaderboard tracks performance. The Retrieval Group and Multimodal Information Group could also move in this



direction, where this is appropriate. There may be more opportunities for data sets to be openly downloadable, such as old data sets or already well-defined tasks. IAD groups could do more to extend the reach of their metrology expertise to the broader community, such as by consulting, hosting tutorials, or playing an important role in establishing best practices for setting up informative evaluations. Human-agent interaction and human-in-the-loop systems represent a growing area of AI, particularly for speech, language, and multimodal technologies. Evaluation in this space is an open problem, and there is an opportunity for IAD to have an impact here, drawing on the expertise in multiple groups. It would be a significant undertaking when staff are already spread thin, so where to focus needs to be carefully considered.

## Software and Systems Division

The main purpose of the Software and Systems Division (SSD) is to inspire and cultivate trust and confidence in software, systems, and their measurements. The division comprises the following four groups: Software Quality Group, Information Systems Group, Systems Interoperability Group, and Cyberinfrastructure Group. Application domains within the division include digital forensics, health care, imaging, biosciences, voting, smart grid, Internet of Things (IoT), cloud computing, materials genome initiative, and scalable computing.

### TECHNICAL QUALITY OF THE WORK

The work of SSD demonstrates multiple areas of unique resources and competencies. These include digital forensics, voting, and characterization of imaging processes covering multiple modalities. In addition, there is promising, newer work in several additional areas, including category theory—foundations of systems semantics, SSD Bugs Framework; software metrology; and artificial intelligence (AI) in imaging and natural language processing across application domains that include health, science, and engineering.

### Assessment of Individual Projects

Digital forensics comprises retrieving, storing, and analyzing electronic data from computers, hard drives, mobile phones, and other storage devices that can be useful in criminal investigations. This is a unique niche with multiple notable accomplishments. The National Software Reference Library won the SSD Judson French Award for work combating child sexual exploitation. The library enables digital forensics examiners to eliminate files of non-interest during analysis, enabling them to focus on user artifacts; it is much easier to find a needle in a haystack if that haystack can typically be reduced by 90 percent. The library serves as a unique source for the digital evidence community.

The Federated Testing Project aids laboratories with a method to test tools (e.g., forensic string search, mobile forensic data extraction) and facilitates the sharing of tool test reports that follow the SSD test protocol with the forensics community. This is a highly effective resource for increasing the quality assurance in digital laboratories that seek to produce quality results but lack the resources to support a dedicated, internal testing team.

Computer Forensic Reference Data Sets (CFReDS) is a highly effective tool for examiners and laboratories. CFReDS posts extractions from computers and related devices that can be used in a myriad of ways, including competency testing, proficiency testing, and training.

There is not a minimum performance threshold for tools tested by the Computer Forensics Tool Testing Program. This presents an opportunity for the digital forensics program to collaborate with digital forensics community organizations to set a threshold for performance rather than simply enumerating issues. Potential collaborators include the Scientific Working Group on Digital Evidence and the Digital Evidence Subcommittee and the Organization of Scientific Area Committees for Forensic Science.

The work of the Software Quality Group spans a broad range of well-chosen topics, and evinces impressive results. Especially noteworthy is the work on preparing a response to Executive Order on Improving the Nation’s Cybersecurity,<sup>1</sup> which was issued May 12, 2021. The draft response can have a significant impact on software development practice. SSD itself is mentioned no less than 21 times in the executive order, so it clearly has a major role to play.

Progress of the Software Quality Group in some areas seems to have stalled. An example of this is the work on the static analysis tool exposition (SATE) initiative, which is a non-competitive study of static analysis tool effectiveness aimed at improving tools and increasing public awareness and adoption. The initiative provides sample test suites to vendors of static code analysis tools, who then apply their tools to the test suites and report results. The results are then discussed at workshops. Between 2008 and 2014, five such workshops were held, but since then just one more event was held in 2019. With the growing importance of static code analysis as part of a routine software development practice, resuming the pace of regular SATE tool exposition events guided by SSD can be beneficial. The tool assessment could also benefit from a more quantitative approach, using measurable benchmarks for static analysis tools.

Similarly, the last update to the Software Assurance Reference Dataset (SARD) seems to have occurred in 2017.<sup>2</sup> Several cases were added in 2020. SARD provides users, researchers, and software security assurance tool developers with a set of known security flaws. Comparable data sets maintained—for example, for CVE (enumerating known cybersecurity vulnerabilities)<sup>3</sup>—have seen significant growth in the past few years, so it is surprising that the SARD data set has not experienced similar growth. The creation of this data set is itself an important initiative that deserves praise. The SARD data set itself, if kept up to date, would also be an important source of input for the tool evaluation workshops (SATE). The data set could help in the creation of a reference set of code that can be used as a target for an objective, quantitative assessment of the performance of software analysis tools, challenging tool vendors to gradually improve over time.

Work on the development of a framework for classifying software bugs (Bugs Framework) that started in 2016 can similarly have significant impact. It could provide new insights if applied to large reference sets, such as CVE or SARD, to quantify the frequency of different categories of bugs. This would be a Herculean task without substantial community support, given the large size of the data sets and the relatively small size of the internal SSD team. This could be a valuable opportunity for public outreach and stronger external collaborations.

### **Assessment Across the Division**

Given that inspiring trust in software tools and measurements is among its key objectives, SSD could take a leading role in the development of a neutral platform for tool evaluation (e.g., SATE) and in the articulation of a vendor-neutral standard output format for static analysis tools in general. An emerging standard that is currently being developed, called SARIF (static analysis results interchange format),<sup>4</sup> is supported by a subset of the commercial tool vendors but does not seem so far to have benefited from the leadership or participation of NIST. There is a clear role for NIST to play here.

Described in 1985 as the integration of people, processes, and technology with connectable devices and sensors to enable remote monitoring, status, manipulation, and evaluation of trends of such

---

<sup>1</sup> Executive Office of the President, 2021, “Improving the Nation’s Cybersecurity,” E.O. 14028, May 12.

<sup>2</sup> NIST Software Assurance Reference Dataset Project website, <https://samate.nist.gov/SARD/index.php>, accessed July 12, 2021.

<sup>3</sup> CVE, “CVE List Home,” <https://cve.mitre.org/cve/>, accessed July 12, 2021.

<sup>4</sup> Static Analysis Results Interchange Format (SARIF) website, <https://sarifweb.azurewebsites.net/>, accessed July 12, 2021.

devices,<sup>5</sup> the IoT continues to evolve through the integration of advanced control systems, real-time analytics, machine learning, commodity sensors, high-speed wireless communications, and embedded systems with high computational capabilities. Real-world applications of IoT include smart homes, wearables, smart/connected cars, industrial controls/automation, smart cities, smart retail, health care, energy management, and agriculture—each of which require some unique approaches to technology integration to support effective and efficient operations.<sup>6</sup>

Many of the tasks performed by an IoT network require devices and systems to have a synchronized method of time for their correct operation. A main reason for a highly accurate synchronized method of time is to provide highly accurate data collection that can be correlated from different sensors and actuators. This highly accurate time synchronization will support reliable and accurate analysis of events, accurate real-time actuation of controls across highly distributed embedded systems, and efficient data communications across a complex networked system. Given the diversity of IoT applications, there are many trade-offs to consider when developing a time-synchronization solution. There is no single time synchronization methodology that satisfies all IoT application requirements, making the development of technologies, standards, and testing methodologies challenging.

The scientific, engineering, and commercial development communities agree with SSD in concluding that integrating state-of-the-art time-transfer methods into modern cyber-physical infrastructure is needed for data and system synchronization. SSD is correct in concluding that timing infrastructures need to continuously evolve to transfer a common timescale across all nodes at ever increasing accuracy as demanded by many types of complex systems, including power systems, robotics, advanced manufacturing, and quantum networks. Even though timing technology is evolving quickly, there is still a lag in the research, development, integration, and standardization of new technologies to support synchronized time references for advanced IoT deployments. Some emerging IoT environments required picosecond timing stability for accurate operation. Manufacturing, energy, transportation, health care, and advanced research environments are pushing the limits of existing time synchronization methods.

SSD is working with several organizations (e.g., the Institute of Electrical and Electronics Engineers [IEEE], Electric Power Research Institute, North American SynchroPhasor Initiative, American Society for Testing Materials, and Department of Homeland Security) on time synchronization methods, standards, and related testing in support of high-resolution calibration techniques, data collection, data analysis, and device control in real-time environments. The key for these advanced real-time environments is the need for highly accurate correlation of events. SSD is contributing knowledge to many standards development activities, including the IEEE 1588 Precision Timing Protocol. SSD is also engaging with stakeholders to design and establish accurate infrastructure to provide measurement time references capable of under 200ps timing stability.

SSD is continuing to make significant contributions to IoT technologies and standards. However, demand for more advanced time synchronization methods is outstripping supply. The need for sub-picosecond time synchronization is nearing, and effective methodologies must be developed. SSD has the opportunity to provide the necessary leadership in the development of these methodologies and their resulting standards. The main challenges are allocation of the resources, identifying partnerships, and establishing the focus needed to address the needs of the diverse IoT applications.

SSD laboratories have multiple separate IoT facilities. To be able to handle the expected growth in IoT systems requires an increase in the available SSD IoT infrastructure, including a common IoT research testbed with advanced capabilities that includes a core setup plus contributing IoT systems from

---

<sup>5</sup> “History of IoT,” Applications of Internet of Things (IoT) in the Market, <https://sites.google.com/site/whatisiot2017v3/history>, accessed July 12, 2021.

<sup>6</sup> Analytics Vidhya, 2016, “10 Real World Internet of Things (IoT) Explained in Videos,” August 26, <https://www.analyticsvidhya.com/blog/2016/08/10-youtube-videos-explaining-the-real-world-applications-of-internet-of-things-iot>.

each laboratory. This would allow SSD groups to leverage the sum total of the resources available to tackle bigger projects.

**RECOMMENDATION: SSD should establish a common Internet of Things research testbed with advanced capabilities (e.g., commercial-off-the-shelf devices, high-speed network, and state-of-the-art test/measurement equipment).**

Without standardization, performance and integration of IoT systems will be limited in performance, have suboptimal functionality, and be filled with vulnerabilities.

**RECOMMENDATION: SSD should drive more standardization across the emerging Internet of Things applications.**

**RECOMMENDATION: SSD should collaborate with other laboratories and research teams to develop time synchronization methods to support near sub-picosecond performance.**

The Systems Interoperability Group develops advanced testing infrastructures and contributes to standards development for ensuring the robustness and interoperability of health information technology (IT) systems, thus removing technical obstacles to implementation and interoperability and accelerating the adoption of cost-effective health IT. The quality of its portfolio is variable. Some projects are very innovative, but a few are unlikely to bring the return on investment consistent with the Information Access Division's (IAD's) expectations for increasing value and making a significant impact on the nation. Some projects focus on theory relating to important dimensions of managing the computing and communications environment. Another portion of the portfolio is more practically focused. One part of the interoperability program, which helped the nation deal with COVID-19, is commendable, influential, and meaningful despite having limited longer-range value once the nation has established the technical infrastructure to monitor and manage public health emergencies, including pandemics.

NIST provides the necessary conformance tests, test tools and techniques to advance health care IT standards that are complete and testable. The focus on the automatic generation of artifacts for conformance testing is appropriate and valuable. Establishing and validating approaches that reduce the requirement for human intervention is critical for adopting standards at scale.

However, that portion of the portfolio focused on health, and health care does not seem to have evolved since the National Academies of Sciences, Engineering, and Medicine review of 2015<sup>7</sup> despite adding a nurse informatician to the workforce and supporting a physician through the University of Maryland Professional Research Experience Program (PREP) program. While adding this expertise to the team was a good step, the division still has not undertaken a comprehensive review of all health systems-related work to take advantage of the new perspectives, knowledge, and skills they brought.

Supplying a substantial voluntary workforce for Health Level (HL) 7 activities cuts into the time and talent available for work offering a more significant impact. HL 7 undertakes many initiatives, and some are likely to have a transformative effect on the health care system, while others are less likely to be influential. The current health care activities seem unlikely to contribute to the strength and visibility of SSD.

Substantial opportunities exist for SSD to impact health care. Three examples illustrate how SSD could refocus, extend, and integrate its health system-oriented priorities and activities to more relevant emerging issues. Given the visibility and importance of SSD, its programmatic work could make a profound and national impact.

---

<sup>7</sup> National Academies of Sciences, Engineering, and Medicine, 2015, *Review of Three Divisions of the Information Technology Laboratory at the National Institute of Standards and Technology: Fiscal Year 2015*, Washington, DC: The National Academies Press.

The first example relates to the health impact of global warming. There are sure to be significant challenges relating to needed measures and standards as this existential issue is addressed. Recently, the combined National Academies have given climate change a very high priority, and the likelihood of a significant federal investment in infrastructure could supply substantial multi-year funding.

A second example relates to the impact that the Internet has had on the availability of scientific and non-scientific information becoming a ubiquitous resource for citizens as well as health care professionals. Citizen science is motivating society to focus on social determinants (i.e., upstream factors) of health rather than simply addressing medical care issues. For example, SSD could creatively support metrics and software to allow citizens to partner with universities and scientists to apply valid and timely scientific information to create innovative technology-based solutions for addressing social determinants of health. This approach would also fit with the IoT work in SSD.

A third example comes directly from health care delivery. What was already a crisis was made more acute due to the demands and pressures presented by the COVID-19 pandemic. Over the past few years, the National Academy of Medicine, Center for Medicare and Medicaid Services (CMS), Office of the National Coordinator for Health Information Technology, the American Medical Association, the American Nurses Association, and the American Medical Informatics Association have each focused on factors that influence clinician burnout, including clinical documentation, which is three times more demanding of time in the United States than in other developed health care systems. Part of this relates to demands for administrative chores rather than care-related work. Similarly, an effort to deal with care authorization, billing, and payment data could be easier to automate effectively and efficiently. Key enabling domains—such as privacy preserving activity recognition and natural language understanding, clinical activity categorization, and approaches to the management of unstructured clinical data—could be advanced by SSD. This example also serves to illustrate how SSD might use a central theme on which to center its systems interoperability efforts to magnify the effects of work being undertaken and create synergies between otherwise disparate efforts. Consistent with SSD’s identification of support for smart health care as a target for expansion of SSD competence and scope among issues within the National Security Commission on AI report,<sup>8</sup> such an effort could tie together work in privacy, IoT, AI methods, and more with systems interoperability.

During the past decade, CMS has supported innovation by establishing the CMS Innovation Center, which oversees a portfolio testing various payment and service delivery models designed to achieve better care for patients, smarter spending, and healthier communities.<sup>9</sup> Consideration might be given to the value of having such an explicitly designed and operated unit within SSD to achieve the goals of influencing health care system quality and cost in a way that is not occurring or likely to occur given the current activities. Partnerships between SSD and CMS might approach issues like those highlighted above.

Our society today is undergoing a massive technological shift toward automation at scale. AI and machine learning (ML) have been widely viewed as among the most transformative technologies that will revolutionize the way we live. Given the potential impact on the economy and national security, these technologies must be developed and used in a trustworthy and responsible way. Characteristics to support trustworthiness include accuracy, explainability and interpretability, reliability, privacy, robustness, safety, security (resilience), and mitigation of harmful bias. Principles such as transparency, fairness, and accountability need to be considered during deployment and use.

SSD has had a relatively light investment in AI historically, but SSD leadership has identified AI to be an important area to focus on for both opportunities and challenges it presents. In a relatively short period of time, SSD has made some important accomplishments in AI, at least from the perspectives of leading initiatives and setting standards and requirements for AI systems. SSD plays an important role in

---

<sup>8</sup> National Security Commission on Artificial Intelligence, 2021, *Final Report*, <https://www.nsc.ai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.

<sup>9</sup> Centers for Medicare and Medicaid Services, “Where Innovation Is Happening,” <https://innovation.cms.gov/>, accessed July 12, 2021.

co-chairing the National Science and Technology Council Machine Learning and AI Subcommittee and participating at policy discussions with its senior technical member on detail to the National Security Commission on AI. These are important services to the United States and can help in creating requirements that are flexible enough to be meaningful for applications, while sufficiently concrete to have meaningful impacts. SSD has also participated in congressional testimonies and hearings.

In collaboration with other federal agencies and academic and industry communities, SSD has led the way to develop guidance for assurance, governance, and practice improvements, as well as techniques for enhancing communication among different stakeholder groups on bias in AI. Bias is not unique to AI. The goal is to identify, understand, measure, manage, and reduce bias in AI systems. Standards and guides are needed for terminology, measurement, and evaluation of bias. SSD can play an important role in this aspect and has already made initial and outstanding efforts to develop trustworthy AI. Trustworthiness and interpretability of machine learning and artificial intelligence is also a topic worthy of attention.

For AI applications, the choices of natural language processing (NLP), especially text processing and imaging, are timely, and the quality of work is excellent with immediate and broad social implications.

Given that AI and ML have been widely regarded as some of the most key technologies to invest in across different fields, one of the biggest challenges faced by SSD will be in recruiting a more diverse workforce that is well trained across multiple disciplines for foundational research and use-inspired development. Although the NIST PREP with Morgan State University, designated a historically Black college and university, is a good idea, it may not be adequate to address this issue. Partnering with other nearby academic institutions, along with scholarships or fellowships for the underrepresented groups that are coupled with internship opportunities at SSD, could be considered to broaden the pool of potential candidates. SSD and IAD could collaborate in this area to further expand the potential impact they can make together on foundational research in AI and ML, in terms of setting the guidelines and standards for data sets, reliability measurements, AI trustworthiness, system requirements, and mass communication to the broader public.

**RECOMMENDATION: SSD should partner with academic institutions and collaborate with other communities on recruiting future workforce, expanding impact, and setting guidelines on artificial intelligence/machine learning system requirements and data sets.**

The Imaging Group has a long tradition of leading the field in the space of novel instruments and calibration protocols covering optical, electron, neutron, and magnetic resonance modalities. Although small, the group is very visible and succeeds in producing highly relevant work by focusing on its core expertise.

In image analytics, the focus of the group is on trusted and reproducible measurements over terabyte-sized images. Image sources can cover a wide range of physical scales—nano to centimeter—corresponding to different underlying physical or biological processes. As a result, it is often difficult to establish measurement accuracy, uncertainty, reproducibility, and interoperability, making the work of the SSD imaging group critical to many in industry and academia. With the emergence of AI-based measurement models, the complexity is increased, and the validation of models becomes a critical component of any imaging system.

SSD's work has inspired, for example, the QUAREP-LiMi initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy.<sup>10</sup> A very relevant work quantifies the variability in microscopy image analysis for COVID-19 drug discovery.<sup>11</sup>

The TrojAI program is an example of the high-quality-imaging AI work pursued in the division. It is focused on methods of detecting hidden behavior in AI models, prior to widespread deployment. The topic is very relevant for the current status of AI, and the TrojAI team shows a mastery of the work, including partnerships, impact, and dissemination of results.<sup>12</sup>

The larger the image data, the more important is the role of the underlying computing infrastructure. The AI-based processing, training, and validation of measurement methods on large image collections require specialized computing infrastructure covering both the parallel computational power (graphics processing units [GPUs]) and the ability to store and transfer data with a very high bandwidth (flash storage). It seems that the scientific productivity of the group would be greatly supported by an advanced computing infrastructure, which is ideally specialized for imaging and related data.

The group has the opportunity to broaden and grow the testing and validation of emerging AI-based multimodal image measurements, an activity that is very much needed by industry. Related to this, the group has the opportunity to drive and stimulate the formation of emerging standards in AI for imaging, a topic that is well aligned with SSD's mission.

The high-performance computing (HPC) team has a track record of delivering high-impact capabilities in image processing using HPC and GPU-based systems, reducing the time to process images by up to 4 orders of magnitude. This has a substantial impact on the SSD science mission through the team's close collaboration with internal SSD stakeholders in materials science and biology that use its software. The team's current work on the development of an abstraction layer aimed at improving the productivity of HPC software developers for this domain is well-thought out, with the data flow design informed by members' experience in image processing. The team's approach is correctly scoped to its resources, since it is leveraging HPC numerical kernels from the vendors and focusing its development effort on the data pipeline, which has been a key component of past success. It is also collaborating with the very high-quality and well-established University of Utah HPC research group to extend the abstraction layer, using the latter's Uintah framework to provide support for distributed computing based on the message passing interface standard.

One facilities issue for the HPC team. Exacerbated by the current global supply shortage of GPU hardware, is the difficulty in obtaining early access to the latest GPU hardware. This lack of access is a significant handicap, causing delays in deploying the software to users when new systems become more widely available.

**RECOMMENDATION: SSD management should consider obtaining early access to systems through the large supercomputer centers housed at the Department of Energy, the Department of Defense, and the National Science Foundation, which generally have better access to such systems.**

---

<sup>10</sup> U. Boehm, G. Nelson, C.M. Brown, S. Bagley, P. Bajcsy, J. Bischof, A. Dauphin, et al., 2021, QUAREP-LiMi: A community endeavor to advance quality assessment and reproducibility in light microscopy, *Nature Methods*, <https://doi.org/10.1038/s41592-021-01162-y>.

<sup>11</sup> M. Simon, S. Yu, J. Nagarajan, P. Bajcsy, N.J. Schaub, M. Ouladi, S. Prativadi, N. Hotaling, 2021, "Quantifying Variability in Microscopy Image Analyses for COVID-19 Drug Discovery," in pp. 3801-3809 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops 2021*, <https://openaccess.thecvf.com>.

<sup>12</sup> P. Bajcsy, N.J. Schaub, and M. Majurski, 2021, Designing Trojan detectors in neural networks using interactive simulations, *Applied Sciences* 11(4): 1865.



## TECHNICAL EXPERTISE OF THE STAFF

Two of SSD's strategic goals are related to technical expertise of the staff—(1) expanding competence and scope to new domains consistent with SSD's and the Information Technology Laboratory's strategic investments and (2) leveraging shared competencies with other SSD units and external organizations.

SSD highlighted the potential to become a leader in addressing the research and development issues identified in the 2021 National Security Commission on AI report.<sup>13</sup> There is a high level of competency across programs. Nimbleness/agility for new approaches was identified by SSD as a core competency for staff. SSD targeted six issues within the National Security Commission on AI report as plausible areas for expansion of SSD competence and scope—big data, massive information, and large knowledge bases; innovative approaches to software quality; support for smart health care; engineering biology; mobile computing; and cyber-physical social systems.

SSD is a valued venue for guest researchers, Ph.D. students, and postdoctoral trainees. Recruitment of such personnel is facilitated by ongoing relationships with multiple universities in the United States and internationally. Successful recruitment of several recent postdoctoral fellows into permanent positions suggests that the mentoring efforts of the permanent staff benefit the broader community and also facilitate a pipeline of talented researchers for hire at SSD.

Competition with industry for talent is a challenge for maintaining and expanding the technical expertise of the staff. There is a need for general data science competencies across all areas. The program developed by the National Library of Medicine to expand the data science competencies of its staff may provide a model and resources for SSD staff wishing to expand their data science competencies.

## ADEQUACY OF RESOURCES

A third SSD Strategic Goal is to ensure that the division has adequate human and computing resources. SSD has 45 full-time staff, including two National Research Council postdoctoral fellows, and 57 associates; the latter includes students, guest researchers, contractors, and others. In response to a perceived insufficiency of clinical expertise in the Systems Interoperability Group, SSD has hired a clinical informatician with a nursing background and a physician. SSD identified the need for full-time staff and funding in the areas of advanced information modeling, including semantics; software testing, system verification, and formal methods; AI; computational science; and biomedical informatics.

In recognition of the growing demands in the division, SSD has contributed to a proposal that delineates the short- and long-term plans to increase computational resources and infrastructure across SSD.

The diversity of projects in the SSD portfolio strains existing human and computational resources. In terms of human resources, recruitment of highly qualified full-time staff has been identified as an overall challenge at SSD due to competition with industry. Within SSD, it may be difficult for small teams to be effective due to lack of critical mass in some areas. The relatively large proportion (about one third) of the permanent workforce at retirement age raises the potential need for succession planning to ensure maintenance of competence in core areas while meeting the demand for new areas of competency.

Current and future SSD needs cannot be met without the addition of new competencies to complement existing strengths.

**RECOMMENDATION: SSD should establish exchange programs with relevant government laboratories (e.g., Army Research Laboratory), academic institutions, and**

---

<sup>13</sup> National Security Commission on Artificial Intelligence, 2021, *Final Report*, <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.

**industry consortia to stimulate new ideas and problem areas, enhance competencies, and facilitate collaboration.**

**RECOMMENDATION: SSD should plan and implement effective ways to recruit and retain a diverse workforce to ensure the appropriate staffing in areas of significant interest to national welfare and security, and to address severe competition from industry in areas such as artificial intelligence, cybersecurity, and the Internet of Things.**

Existing projects in several areas are outgrowing current computational resources. For example, for IoT, expressed needs include a common IoT research testbed, a collection of commercial off-the-shelf devices that can be shared and reused across the laboratories, a common set of hardware and software applications that can be used by interested parties at SSD, a high-speed network, and a pool of available general computing resources.

There are opportunities for innovative approaches, such as public-private partnerships and sharing computational resources across national laboratories or other agencies with supercomputers, to complement the purchase of relevant high-performance resources.

**RECOMMENDATION: SSD should establish a multi-faceted, multi-phased approach to enhance computational resources and infrastructure (e.g., programmable logic for optimization and acceleration) through public-private partnerships, sharing with other national laboratories, and direct purchase.**

Partnering with government agencies such as the Department of Energy (DOE), Department of Defense (DoD), and the National Science Foundation would ensure early access for developers to every high-performance computing platform to be ahead of the industry curve. It would open the door to a larger opportunity for both traditional HPC platforms, but also for collaboration on emerging technologies in computing that are showing promise in HPC settings, and also in other computing areas such as AI, ML, and related data science technologies such as encryption and large-scale storage and retrieval. DoD and DOE already collaborate successfully on both the high-precision supercomputing and the lower-precision ML/AI computational technologies.

**RECOMMENDATION: SSD should partner with government agencies such as the Department of Energy, the Department of Defense, and the National Science Foundation to ensure early access for developers to every high-performance computing platform to be ahead of the industry curve.**

**SSD should also consider partnering with extra-U.S. agencies where allowable and appropriate.**

## **EFFECTIVENESS OF DISSEMINATION OF OUTPUTS**

The volume of outputs is impressive across the groups and projects and includes scholarly products, code, software, and participation or leadership in standards development organizations. Dissemination of outputs includes press releases, technical reports, books, journal publications, conference presentations, events such as Connectathons, and repositories such as GitHub for code and software.

For scholarly products, code, and software, the effectiveness of dissemination is evaluated through measures such as frequency of downloads. Effectiveness of leadership and participation in standard development organizations such as HL 7, IEEE, and International Standards Organization is evaluated by volume of leadership, participation, and standards to which SSD has contributed. High

potential impact of Executive Order on Improving Nation’s Cybersecurity<sup>14</sup> gave SSD several tasks, including developing minimum standards for software testing to support federal procurements.

SSD outputs are clearly delineated, but it is more difficult to ascertain effectiveness and impact for multiple reasons, including missing metrics or lack of clarity in metrics. In the area of leadership and participation in standards development organizations, it is difficult to ascertain the cost-versus-benefit ratio of such work related to other job responsibilities, although SSD participation is clearly valued due to its perceived neutral stance.

Matching the needs of the market and users to product development is an essential foundation for effectiveness and impact. A tool or other product may be considered high quality from a technical perspective but not meet specific market or user needs. For example, in terms of system interoperability in health care, FHIR (Fast Healthcare Interoperability Resources), a draft standard at HL 7, is driven by policy and market factors that promote its use, while some SSD tools remain focused on the technical aspects of conformance testing in the HL 7 version 2 context.

The strategies for dissemination of SSD outputs appear to be limited to primarily scientific and technical audiences. To increase the effectiveness and impact of SSD efforts, a broader variety of tailored communication strategies is needed to reach other target audiences such as policy makers.

The effectiveness and impact of SSD’s substantial outputs is suboptimal.

**RECOMMENDATION: Building on its current strengths in engagement with target communities, SSD should establish strategies to enhance its understanding of market and user needs to ensure that products are not only of high technical quality, but also useful to relevant stakeholders.**

**RECOMMENDATION: SSD should consider using the National Security Commission on AI target issue of support for smart health care as an organizing framework to increase the impact of their health-related work. This could be facilitated by establishing a center of innovation.**

**RECOMMENDATION: SSD should implement a broader variety of tailored communication strategies for dissemination of outputs to reach important target audiences (e.g., policy makers) beyond scientific and technical audiences.**

## GENERAL CONCLUSIONS

Consistent with its purpose of inspiring and cultivating trust and confidence in software, systems, and their measurements, the overall technical quality of SSD’s unique and promising new programs is excellent, given human and computational resources. There is a high-level of technical competency among staff across SSD programs. There are four areas requiring additional focus. First, SSD’s capacity to handle the expected growth in IoT systems is insufficient. Second, SSD performance is increasingly limited in several areas by lack of adequate computational resources and infrastructure. Third, current and future SSD needs cannot be met without the addition of new competencies to complement existing strengths. Fourth, the effectiveness and impact of SSD’s substantial outputs is suboptimal.

---

<sup>14</sup> Executive Office of the President, 2021, “Improving the Nation’s Cybersecurity,” E.O. 14028, May 12.

## Statistical Engineering Division

The mission of the Statistical Engineering Division (SED) is to develop and apply statistical and probabilistic methods supporting research in measurement science and technology; implement methods for experimental design, data analysis, statistical modeling, and probabilistic inference; study and apply best practices for the characterization of measurement uncertainty; and disseminate sound statistical methods to U.S. industry and the scientific community. SED conducts statistical research, consulting, and collaboration in metrology, standard development, forensics, and other areas fundamental to NIST's mission. While maintaining core competence in experimental design, statistical modeling, and data analysis, the division is expanding into the areas of machine learning (ML), artificial intelligence (AI), and data science. The activities of SED staff not only lead to advances in statistical methodology; they also bring significant benefits to the scientific programs of their many collaborators in government and industry. The division's Boulder Statistics Group focuses on collaborations with NIST scientists in the Boulder, Colorado, campus. The Gaithersburg Group focuses on collaborations with NIST scientists in the Gaithersburg, Maryland, campus.

### TECHNICAL QUALITY OF THE WORK

SED has considerable scientific expertise in the three distinct areas needed to accomplish its mission: metrology and inter-laboratory studies, collaborative research and education in support of other NIST programs, and standing as a neutral and trusted arbiter in areas of statistical controversy. However, advanced statistical computing techniques (beyond some software development) seems to be for the most part absent from the activities, priorities, and interests of the SED. Some software developments were reported, but the role and importance of computing algorithms go beyond this. The strong interlink between statistics methods and research to computer science methods and research are an integral part of modern statistics.

The area of metrology and inter-laboratory studies has been a core technical component of SED since its founding in 1947. NIST appears to be the best in the world among all the organizations engaging in these activities. The activities are of two types. The first is publishing papers in applied science and metrology outlets to try to improve statistical practice. For instance, SED staff indicated that standard metrology assumes independence of data (a generally wrong assumption for metrology data). A SED staff member led the development of an international standard (ISO/CD 24185) that shows how to deal with dependence properly. This was not new statistics, but it shows NIST doing its job to continually improve metrology. Now statistical methodology development is expanding beyond the modeling stage to include the whole data science life cycle including data formulation and data cleaning, with consistent documentation and code repository including the data cleaning process.

Obtaining inter-laboratory agreements (necessary for national and international laboratories to reach consensus on metrology issues) is an interesting statistical and political problem. The political aspect is that even laboratories that are hopelessly wrong ought not to simply be excluded from the analysis, because rejection of their results would be politically problematical. The SED chief scientist explained that disagreement could be addressed through Bayesian hierarchical modeling, whereby the wrong results would have little effect on the final answer, but the confidence intervals for the incorrect

laboratories would still include the correct result, which would not then be rejected. This is a creative use of modern Bayesian analysis.

In the area of collaborative research and education in support of other NIST programs, SED strongly supports other NIST programs that are in need of statistical support. SED seems to be among the best government and industry collaborative statistics groups. Sometimes this support is routine, as in the frequent need of other NIST scientists for assistance in standard experimental design (a SED strength). Even here, NIST is forward looking with developments on the leading edge of design. For instance, the ABACUS chemical analysis package was developed to provide all chemical analysis groups at NIST a way to generate optimal designs for their experiments. The package is a leading-edge hierarchical Bayes design package that could beneficially be made widely available to the chemical industry. It would be useful to compare Bayesian model conclusions with machine learning conclusions unless the Bayesian modeling approach is scientifically vetted and well documented. The qualitative scientific conclusions should not depend on whether a NIST statistician is Bayesian or not.

Often the collaborations involve development of novel statistical methodology. One such example was the LANTERN methodology, which focused on distilling large-scale genotype-phenotype measurements into an explainable low-dimensional representation, while using a modern Gaussian process implementation to deal with the highly variable response surface arising from the measurements. Across several large-scale benchmarks (including ML and AI), LANTERN's predictive performance was outstanding, providing interpretable scientific insights concerning the way that the genotypes and phenotypes affect the analysis.

SED also provides numerous short courses and internal individual training opportunities in statistics for NIST staff. Such technology transfer is an important part of enhancing the overall scientific expertise at NIST. Often, this technology transfer occurs on an individual level. For example, a SED staff member developed a very complex statistical design/analysis implementation for problems a particular NIST researcher faced; by the end of the project, the scientist had internalized this very complex analysis and subsequently did not need statisticians to help address such problems.

In the area of standing as a neutral and trusted arbiter in areas of controversy, following its long-standing role as the setter of standards in metrology, SED is also seeking to establish standards in other areas of statistical controversy. An illustration of this was the work on the reporting of forensic evidence in court.<sup>1</sup> A typical current practice for DNA evidence is to report likelihood ratios to represent evidential weight, but this has become a default practice rather than a carefully reasoned methodology, and as such is quite controversial. The SED effort contribution was to go back to the roots of the problem and try to identify methods of presenting forensic evidence that all could agree with.

A 2015 National Academies of Sciences, Engineering, and Medicine report<sup>2</sup> encouraged SED to broaden its engagement with the academic community, both to augment the expertise at NIST but also to further raise the profile of SED. A number of steps were taken to do this and need to be pursued. However, a basic problem is that very few academic statisticians are involved with SED. Academics can spread the word that NIST is a great place to work and with which to collaborate. Academic involvement can also help address the deficiency that SED staff have very few publications in statistics journals (only 17 since 2015)—a limiting factor in terms of visibility to the external community. This is to be expected for metrology and inter-laboratory agreements, as the audiences for these works are primarily non-statistical.

Most of the collaborative research being done by SED is of enough novelty that it could be published in mainstream statistical journals. The challenge to doing so is time; the primary audience for the work is typically the discipline in which the supported project originated, and publishing in that

---

<sup>1</sup> National Institute of Standards and Technology, Information Technology Laboratory, "Likelihood Ratios and Evidence Communication," presentation to the panel on June 22, 2021.

<sup>2</sup> National Academies of Sciences, Engineering, and Medicine, 2015, *Review of Three Divisions of the Information Technology Laboratory at the National Institute of Standards and Technology*, Washington, DC: The National Academies Press.

discipline has priority at NIST. Abstracting from the work of a supported project the novel methodology for a statistical journal can be time consuming, and SED staff often reasonably choose to spend their time elsewhere (e.g., on a new collaborative project). Academic engagement will help to develop these methodologies more fully. Additionally, this could be helpful in influencing the academic community to take on research topics and challenges of long-term interest to NIST, such as the development of methods at the intersection of statistics and machine learning.

**RECOMMENDATION: SED should strengthen its engagement with the academic community and should consider assigning an academic statistician on a project to take the lead in the production of statistics journal articles.**

There seems to be excellent synergies occurring between traditional statistics and data science, as reflected in the LANTERN project, which involved a very productive synergy between the two. The incorporation of data science within the existing activities of SED staff could become a challenge if data science were to replace the core statistical elements of SED. On the other hand, this may not become a problem if data science staff were added to SED as part of the overall Information Technology Laboratory (ITL) data science initiative.

**RECOMMENDATION: SED should expand engagement with the data science, machine language, and artificial intelligence researchers at NIST.**

There are opportunities for NIST to leverage its reputation as a neutral arbiter in directions other than metrology, inter-laboratory agreements and forensics. For instance, SED is ideally positioned to propose a common language for ML, AI, data science and statistics, although it may well be that those ships have sailed.

**RECOMMENDATION: SED should maintain and enhance its role as an impartial and trusted arbiter of statistical issues.**

## **TECHNICAL EXPERTISE OF THE STAFF**

SED scientists are highly capable at accomplishing this mission and supporting NIST's diverse programs and objectives, both inside and outside of ITL. Division members are well trained statistical generalists, each capable of supporting data collection and analysis efforts across a spectrum of technology applications. In some cases, these applications involve non-trivial existing methods that may be unknown to non-statisticians. In others (e.g., basic metrology), SED staff members improve well established, but not the best, statistical practice methods and advocate the use of appropriate methods. New science in NIST programs often produces the need for new theoretically sound and practically effective statistical methodology. SED effectively provides excellent support in all of these kinds of activities.

SED's long-standing core responsibility to support measurement science has grown substantially since the division's 2015 National Academies review. For example, its role in the development and documentation for every Standard Reference Material available from NIST is now quite substantial. It is making important national and international contributions to measurement and standards organizations.

In their responsibility to support new science and technology, SED staff members work as close collaborators with scientists and engineers, educating their partners and developing deep personal domain interest. They provide breadth of statistical expertise and domain engagement that impact outcomes across NIST. This includes both relatively small/focused and more major/broad initiatives like forensics and evaluation of the current state of diversity, equity, and inclusion.

AI and ML are areas of substantially increased interest at NIST, including ITL. Probabilistic modeling and advanced statistical methodology provide perspectives and tools with potential to greatly advance these fields. AL and ML experts typically have limited background in statistics, and not all statisticians have experience in these areas. Although some progress has already been made, there is need for additional SED staff that can very soon interact substantially with computer scientists and engineers in AI/ML. Part of the challenge here is staff size and the number and size of the efforts that SED is already supporting. Part of the challenge is to find statistical generalists who have existing experience in these areas.

**RECOMMENDATION. In hiring of technical staff, SED should continue its practice of hiring statistical generalists (trained individuals with good understanding of the theory and the applications of current and proven statistical methodologies) and pay first attention to maintaining expertise to support its existing missions and broad statistical expertise. SED should also seek to recruit excellent statisticians with prior experience in artificial intelligence (AI)/machine learning (ML), targeting some recruiting efforts at good statistics and computer science Ph.D. programs with strong AI/ML components.**

**RECOMMENDATION: As key elements in its hiring plans, SED should seek individuals with expertise in statistical computing, artificial intelligence, machine learning, and research publication.**

## ADEQUACY OF RESOURCES

SED staff reported that the division has sufficient hardware, software, and information technology support for their mission. It has implemented the recommendation from the 2015 National Academies report to develop stronger ties with the statistical research community for its staff. Staff have been engaged with the American Society for Quality, the American Statistical Association, and the International Statistical Engineering Association; NIST hosted the 2019 Fall Technical Conference and participated regularly in the Defense and Aerospace Test and Analysis Workshop (DATAWorks). It sponsored a Virginia Tech Computational and Data Analytics Capstone Project course. SED has also participated in NIST's efforts to expand diversity, equity, and inclusion and has participated in both the Inclusivity Network Analysis as a First Step to Harness Human and Social Capital for Innovation at NIST and Assessing Inclusivity of Women at NIST projects.

One of the key challenges for SED is how to recruit and retain the next generation of talent. Over 60 percent of SED staff are eligible for retirement. SED is one of the very best groups in the world in metrology and inter-laboratory experimentation and among the best government or industry applied collaborative statistics groups. As SED develops its strategy for growth, it will need to ensure that it maintains excellence in these areas.

**RECOMMENDATION. SED's staffing strategy should continue to support its excellence in metrology, reference material development and calibration, documentary standards development, and inter-laboratory comparisons. It should also include statisticians who can support broad scientific collaboration at NIST.**

SED has a growth opportunity. AI and ML are a strategic technical focus for NIST. Statisticians are key players in AI/ML, and data science more broadly, and can make substantial contributions to collaborative projects. As an example, the LANTERN project focuses on developing an interpretable hierarchical Bayesian model to distill large-scale, genotype-phenotype landscape measurements into an explainable low-dimensional representation. Across several large-scale benchmarks, LANTERN's predictive performance equals or outperforms all alternative approaches while providing interpretable

scientific insights. Examples such as the LANTERN project illustrate the value that SED can bring to NIST in AI/ML and suggest that SED should leverage this opportunity to increase its staffing in these important areas.

**RECOMMENDATION: SED should leverage the growth opportunity in artificial intelligence and machine learning to increase its staffing in these areas and in data science broadly.**

SED is not well known among students and postdoctoral researchers in the statistics community. In addition, only 16 percent of SED's technical staff are women. These factors suggest that developing new recruiting strategies needs to be a high priority for SED.

SED might consider adapting a model like that used in the Statistical Sciences Group at the Los Alamos National Laboratory (LANL).<sup>3</sup> Starting over 20 years ago, that Statistical Sciences Group identified faculty from across the United States with expertise in areas of strength or strategic growth. These faculty were invited to spend portions of their summers or sabbaticals at LANL with their travel supported, and their students were encouraged to apply for internships. Faculty and students worked on projects, with particular emphasis on helping to prepare results for publication. Faculty were able to provide peer-review for internal technical documents, and students often chose to extend the initial project work into part of a master's thesis or Ph.D. dissertation, which continued their collaborations into the academic year. Many students, after exposure to the deep scientific collaboration provided by collaborative projects, chose careers at the national laboratories. Intentionally focused engagement with the academic community may enable both recruitment and retention by allowing SED staff to continue their professional development.

This form of recruitment pipeline is widely used in both Department of Energy (DOE) and Department of Defense (DoD) agency settings. The DOE laboratories have especially well-developed internship programs attached to a variety of U.S. universities, including LANL's and the Lawrence Livermore National Laboratory's well-established historical ties to the University of California system's various campuses. These internship and related faculty outreach activities provide not only an existing successful exemplar for science, technology, engineering, and mathematics (STEM) student recruitment, but also a risk-mitigation strategy for the problem of effective recruitment efforts being compromised by ineffective retention methods. Hiring student interns from U.S. universities provides a cost-effective means to evaluate the fit of potential candidates while helping students and faculty learn about the unique culture of these federal institutions.

Faculty and student outreach efforts also provide a means to enrich the diversity of the NIST workforce while remedying the problem of aging out of the NIST staff ranks. A good way to develop young STEM talent for a federal agency is through university outreach, and especially when the outreach collaborative efforts are focused on universities that recruit and retain high-quality students with diverse backgrounds. The collaboration with Morgan State University mentioned in the Information Access Division and SSD chapters above is a good start on solving the problem of the recruitment of under-represented minorities, and many federal agencies have formed similar collaborations with universities that have high proportions of U.S. citizens among their student communities, as well as strong STEM programs that attract high-quality, technically oriented students. These universities are often found in large states with diverse populations. For example, Texas has several universities with diverse student bodies. The development of intern programs between NIST and these universities would provide a time-tested strategy for recruitment and retention of high quality diverse staff and management talent. If funding can be developed to support the year-round academic studies of these hiring candidates, the strategy only improves its utility by creating strong intellectual and personal bonds between NIST and key university faculty who recruit domestic students.

---

<sup>3</sup> Los Alamos National Laboratory, "Statistical Sciences," <https://www.lanl.gov/org/ddste/aldsc/computer-computational-statistical-sciences/statistical-sciences/index.php>, accessed July 12, 2021.



**RECOMMENDATION: SED should develop a strategy to broaden its workforce pipeline.**

In addition to recruiting and retaining staff, SED has an opportunity to expand its impact with more effective development and deployment of software products, particularly SED-developed web-based tools. An example of this kind of tools is ABACUS, a web-based tool that automates the design and analysis of data from quantitative chemical measurement procedures. In general, SED staff do not have the development experience to make robust, deployable tools, and they would benefit from collaboration with other ITL staff. In addition, SED staff reported that NIST policies and procedures currently make it difficult to deploy an outward facing software product. SED could consider software options that include open source software, third-party repositories (whether github or an academic/commercial cloud system), applying resources to some level of user support, and investing in developing and presenting tutorials, documentation, and training.

**RECOMMENDATION: ITL should identify resources and create processes that make software and tools developed by SED easily available both within and outside NIST.**

**EFFECTIVENESS OF DISSEMINATION OF OUTPUTS**

The majority of SED's research activities are applied in character and are characterized by close support of disciplinary research groups in NIST and engagement with standards bodies in the proper use of statistics. As noted earlier, the division also engages in methodological research. The research portfolio is driven by the needs of stakeholders across NIST and the standards communities. Research outcomes tend to be either contributions to disciplinary research products or methodological in nature. As such, appropriate levels of technology transfer and the dissemination of research results can be accomplished through publication of peer-reviewed journal and conference papers, publication of standard reference materials and documentary standards, presentation at scientific conferences and workshops, and dissemination of computational tools. SED has demonstrated continuing high-levels of dissemination activity that over the past 5 years includes publication of 287 academic papers, reports, and book chapters, as well as characterization of numerous reference materials. SED researchers have presented at numerous scientific conferences and have hosted dozens of short courses and tutorials for the benefit of NIST colleagues, other government agencies, and the academic community. Additionally, division researchers have a number of active collaborations with members of the academic community that include cooperative research agreements and hosting visiting faculty, postdoctoral researchers, and students.

While SED is highly active in publishing in subject-matter journals and conferences, its dissemination of research results, as well as its standing and familiarity within the academic community, could be improved through increasing publication in top-shelf statistical journals. The scope of most top journals tends to be on foundational theory and novel methodology, whereas most SED research activity is applied and aims to support the practice of statistical analysis. A move to increase publication in statistics journals might be achieved by collaborations with academic statisticians to further develop and publish on the novel statistical issues arising from these applications. Additionally, it would be beneficial if SED staff expanded their publication and research presence beyond statistics communities to machine learning journals such as the *Journal of Machine Learning Research* and conferences such as the Annual Conference on Neural Information Processing Systems (NeurIPS) and the International Conference on Machine Learning.

**RECOMMENDATION: SED should continue its highly active dissemination in subject-matter venues while increasing publication in highly ranked statistics journals.**

The dissemination of software products by the division appears to be centered on open-source publication by the research groups themselves. Dissemination and technology transfer could be improved if these efforts were supported by either a dedicated software group that could manage and implement a pipeline for the translation of research tools into products with robustness and usability attributes appropriate for broader distribution and use, or by partnering with other units within NIST to establish such a pipeline for disseminating software.

**RECOMMENDATION: SED should commit to providing translational software support for research products with the goal of disseminating software with usability at near-commercial levels.**

## GENERAL CONCLUSIONS

SED has maintained outstanding expertise to support core missions of NIST and has provided close and productive collaborations with other NIST divisions, other government agencies, and industry. It serves as a neutral and trusted arbiter on the interpretation of statistical evidence and the validity of statistical methods. With a relatively small staff size, SED faces a challenge of maintaining core competence in statistical design, modeling, data cleaning, and uncertainty measurement, while at the same time growing new competence to support NIST initiatives in areas such as AI. There is need and opportunity for improving the technical quality of programs, scientific expertise, resource development, and dissemination efforts. A common theme underlying many of these areas is the need to strengthen ties with the statistical community to improve the recruitment, retention, and professional development of staff, and to enhance visibility through collaborative publications with academic statisticians.

## Crosscutting Conclusions and Recommendations

### TECHNICAL QUALITY OF THE WORK

The technical quality of the work is generally excellent. Information Access Division, Software and Systems Division, and Statistical Engineering Division continue to make significant contributions to address the Information Technology Laboratory (ITL) and NIST missions, national needs, and the needs of government, industry, and academic stakeholders. Amidst continuing technological and societal changes, there are opportunities for increased collaborations across projects to address common challenges in such areas as artificial intelligence, machine learning, health information technology, data science, and statistical methodology.

### TECHNICAL EXPERTISE OF THE STAFF AND ADEQUACY OF STAFFING

Technical staff generally possess expertise adequate to perform their task. Some are among the best in the world in their areas of research. However, evolving needs in some areas (e.g., artificial intelligence, machine learning, and data science) will produce expertise gaps that will have to be filled.

In terms of human resources, recruitment of highly qualified full-time staff has been identified as an overall challenge at ITL due to competition with industry and academia, which offer salaries that exceed the limits available to the ITL. It may be difficult for small teams to be effective due to lack of critical mass in some areas. The relatively large proportion of the permanent workforce at retirement age in some areas raises the potential need for succession planning to ensure maintenance of competence in core areas while meeting the demand for new areas of competency. Current and future ITL needs cannot be met without the addition of new competencies to complement existing strengths.

**RECOMMENDATION: ITL should apply an aggressive, imaginative focus on hiring to replace retiring staff, to address important growth areas such as artificial intelligence, machine learning, and data science, and to fill specific gaps in the divisions. This effort should aspire to diversity targets.**

**RECOMMENDATION: ITL should plan and implement effective ways to recruit and retain a diverse workforce to ensure the appropriate staffing in areas of significant interest to national welfare and security, and to address severe competition from industry in areas such as artificial intelligence, cybersecurity, and the Internet of Things.**

**RECOMMENDATION: ITL should establish exchange programs with relevant government laboratories, academic institutions, and industry consortia to stimulate new ideas and problem areas, enhance competencies, and facilitate collaboration.**

Properly managed, ITL's diversity, equity, and inclusion strategy can add a helpful element to the recruiting process. The staff has performed commendably during the COVID-19 pandemic, collaboratively addressing critically important responses to the pandemic.

## ADEQUACY OF FACILITIES AND EQUIPMENT

Post-pandemic planning gives NIST opportunities to continue some remote work, making NIST a more attractive environment for staff and giving NIST a needed edge in recruiting. Also, creative thinking about the new work environment can lead to more productive use of facilities, which could be considered during NIST's ongoing laboratory and office renovations.

ITL facilities are generally adequate and support well the activities of the divisions. There is a critical need, however, for improved computing capabilities that are needed to support complex computation and analysis. This is particularly the case for the growing number of projects that involve state-of-the-art machine learning models, which call for ever larger training sets and computational resources.

**RECOMMENDATION: ITL should take steps to insure adequate resources, especially computing to support artificial intelligence, machine learning, and data science at sufficient scale.**

**RECOMMENDATION: To get access to the most modern resources, ITL should seek collaborations with other organizations in the public and private sectors, including other government agencies**

## EFFECTIVENESS OF THE DISSEMINATION OF OUTPUTS

Each division disseminates its outputs widely, but the divisions vary in the relative emphasis placed on dissemination vehicles (e.g., publications, workshops, data repositories, standard reference data, and educational programs). Careful, systematic, and continuing analysis of the needs of specific stakeholder communities would improve the effectiveness of the dissemination of ITL's outputs.

**RECOMMENDATION: ITL should broaden its impact to non-technical stakeholders, policy makers, and the public.**

**RECOMMENDATION: ITL should improve messaging aimed at non-technical audiences such as policy makers, media, and the general public for the outputs of the Information Access, Software and Systems, and Statistical Engineering Divisions.**

This would benefit NIST by providing greater acceptance of and support for NIST efforts.

## **Appendixes**



## A

### Acronyms

AI	artificial intelligence
BRL	Biometrics Research Laboratory
CFReDS	Computer Forensic Reference Data Sets
CMS	Center for Medicare and Medicaid Services
CVE	enumerating known cybersecurity vulnerabilities
DARPA	Defense Advanced Research Projects Agency
DATAWorks	Defense and Aerospace Test and Analysis Workshop
DEI	diversity, equity, and inclusion
DoD	Department of Defense
DOE	Department of Energy
FaCCT	fairness, accountability, and transparency (Association of Computing Machinery)
FHIR	fast healthcare interoperability resources
GPU	graphics processing unit
HL	Health Level
HPC	high-performance computing
IAD	Information Access Division
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
ISO	International Standards Organization
IT	information technology
ITL	Information Technology Laboratory
LANL	Los Alamos National Laboratory
ML	machine learning
NBDIF	NIST Big Data Interoperability Framework
NIST	National Institute of Standards and Technology
NLP	natural language processing
PREP	Professional Research Experience Program
SARD	Software Assurance Reference Dataset
SARIF	static analysis results interchange format
SATE	static analysis tool exposition

SED	Statistical Engineering Division
SSD	Software and Systems Division
STEM	Science, Technology, Engineering, and Mathematics
TAC	Text Analysis Conference
TREC	Text Retrieval Conference
TRECvid	video retrieval evaluation tracks of TREC
TRUC	Trustworthy AI Conference
VVSG	Voluntary Voting Systems Guidelines



## B

### Biographical Sketches of Panel Members

PHILIP M. NECHES, *Chair*, is a venture partner at Entrepreneurs Roundtable Accelerator in New York City. He is also an independent consultant working with early-stage companies in the information technology and communications service industries on their technical, market, and business strategies as an advisor, board member, and investor. A world-renowned authority on databases, he was the founder and chief scientist of Teradata Corporation, he pioneered the application of parallel processing to commercial applications. Dr. Neches was elected to the National Academy of Engineering (NAE) in recognition of this work. He began his career at Transaction Technology Inc., a subsidiary of Citigroup. He led analysis of consumer banking networks, including the first large-scale deployment of automated teller machines (ATMs) in the United States. After Teradata, as senior vice president and chief scientist for NCR Corporation, he led both the repositioning of NCR's computer product family and the product plan for a merger with AT&T. Dr. Neches served as vice president and chief technical officer of Multimedia Products and Services Group at AT&T Corporation. He has served on the board of directors of ExpandBeyond, Inc. and International Rectifier, and on the advisory boards of EarthLink, Tacoda Systems, Luxtera, and the Technology Group of Merrill Lynch. Other prior directorships include MCC, Semitech, Dayton Public Radio, DemoGraFx, MediaMap, PeopleLink, and VendQuest. He serves on the California Institute of Technology (Caltech) board of trustees and sits on its executive committee. Dr. Neches holds B.S., M.S., and Ph.D. degrees in computer science from Caltech.

RUZENA K. BAJCSY received her master's and Ph.D. degrees in electrical engineering from Slovak Technical University, Bratislava, Slovak Republic, in 1957 and 1967, respectively, and a Ph.D. in computer science from Stanford University in 1972. She is a professor of electrical engineering and computer sciences and the NEC chair holder at the University of California, Berkeley (UC Berkeley), and director emeritus of the Center for Information Technology Research in the Interest of Science (CITRIS). Prior to joining UC Berkeley, she was a professor of the Computer Science and Information Department at the University of Pennsylvania. There she founded the General Robotics and Active Perception laboratory in 1979, which is flourishing now. In 1999, she was appointed to head the Computer and Information Science and Engineering Directorate at the National Science Foundation (NSF). In 2001, after she finished her stay at NSF, she retired from University of Pennsylvania and joined the faculty at UC Berkeley. Dr. Bajcsy is a member of the NAE and the National Academy of Medicine (NAM) as well as a fellow of the Association for Computing Machinery (ACM), the Institute for Electrical and Electronics Engineers (IEEE), and the American Association for Artificial Intelligence. Her election to NAE/NAM was predicated on the invention of computer registration of an anatomy atlas to CAT images of the human brain (1978-1979) using theory of elasticity to account for local deformations. This principle led to many applications and improvements in brain analysis, pursued by her student Jim Gee since the 1990s. At UC Berkeley, in collaboration with University of California, San Francisco (UC San Francisco) she pursued assessment of mechanical properties of musculoskeletal system of human bodies. This work led to implementation of various prosthetic wearable devices. In 2001, she received the ACM/Association for the Advancement of Artificial Intelligence Allen Newell Award. Since 2008, she has been a member of the American Academy of Arts and Sciences. She is the recipient of the Benjamin Franklin Medal for Computer and Cognitive Sciences (2009) and the IEEE Robotics and Automation Award (2013) for her contributions in the field of robotics and automation. She received the 2016 NAE Simon Ramo Founders Award for her life achievements. Dr. Bajcsy's son, Peter Bajcsy, is employed by

NIST as a researcher in the Software and Systems Division's Information Systems Group within the Information Technology Laboratory.

SUZANNE BAKKEN, Ph.D., RN, FAAN, FACMI, FIAHSI, is the Alumni Professor of Nursing and a professor of biomedical informatics at Columbia University. Her program of research has focused on the intersection of informatics and health equity for more than 30 years and has been funded by the Agency for Healthcare Research and Quality, National Cancer Institute, National Institute of Mental Health, National Institute of Nursing Research (NINR), and the National Library of Medicine (NLM). Dr. Bakken's program of research has resulted in more than 300 peer-reviewed papers. At Columbia Nursing, she leads the NINR-funded Precision in Symptom Self-Management (PriSSM) Center and Reducing Health Disparities Through Informatics (RHeaDI) Pre- and Post-doctoral Training Program. She is a fellow of the American Academy of Nursing, American College of Medical Informatics, International Academy of Health Sciences Informatics, and a member of the NAM. Dr. Bakken has received multiple awards for her research, including the Pathfinder Award from the Friends of the National Institute of Nursing Research, the Nursing Informatics Award from the Friends of the NLM, the Sigma Theta Tau International Nurse Researchers Hall of Fame, and the Virginia K. Saba Award from the American Medical Informatics Association. Most recently, she was the first nurse recipient of the Francois Gremy Award from the International Medical Informatics Association. Dr. Bakken currently serves as editor-in-chief of the *Journal of the American Medical Informatics Association* and as a member of the board of regents of the NLM. Dr. Bakken has served on multiple consensus committees, most recently a committee focused on returning of biomarker results to research participants. As the NAM Nurse Scholar-in-Residence, she helped staff "Communities in Action: Pathway to Health Equity." She is also a member of the Health Literacy Roundtable.

PETER A. BELING is a professor in the Grado Department of Industrial and Systems Engineering and associate director of the Intelligent Systems Laboratory in the Hume Center for National Security and Technology at Virginia Polytechnic Institute and State University (Virginia Tech). Dr. Beling's research interests lie at the intersections of systems engineering and artificial intelligence (AI) and include AI adoption, reinforcement learning, transfer learning, and digital engineering. His research has found application in a variety of domains, including mission engineering, cyber resilience of cyber-physical systems, prognostics and health management, and smart manufacturing. Prior to joining Virginia Tech in 2021, he was a professor of systems engineering at the University of Virginia (UVA) and directed the UVA site of the Center for Visual and Decision Informatics, an NSF Industry/University Cooperative Research Center, and the Adaptive Decision Systems Laboratory. Additionally, he serves on the Research Council of the Systems Engineering Research Center, a University Affiliated Research Center for the Department of Defense (DoD). Dr. Beling has served as editor and reviewer for many academic journals and regularly serves as a panel member for the National Academies of Sciences, Engineering, and Medicine and NSF. Dr. Beling received his Ph.D. in operations research from UC Berkeley.

JAMES O. BERGER is the Arts and Sciences Distinguished Professor Emeritus of Statistics at Duke University. His current research interests include Bayesian model uncertainty and uncertainty quantification for complex computer models. Dr. Berger was president of the Institute of Mathematical Statistics from 1995-1996 and of the International Society for Bayesian Analysis during 2004. He was the founding director of the Statistical and Applied Mathematical Sciences Institute, serving from 2002-2010. Dr. Berger was elected as a foreign member of the Spanish Real Academia de Ciencias in 2002, elected to the U.S. National Academy of Sciences (NAS) in 2003, was awarded an honorary doctor of science degree from Purdue University in 2004, and became an honorary professor at East China Normal University in 2011.

DUANE BLACKBURN serves as deputy director of MITRE's Center for Data-Driven Policy, which brings objective, evidence-based, nonpartisan insights to government policymaking. Before joining

MITRE, he served as an assistant director of the White House Office of Science and Technology Policy and coordinated science and technology (S&T) policy for homeland security, law enforcement, and identity matters. He led the development and implementation of government-wide S&T strategies on a variety of subjects and influenced the conceptualization and oversight of national strategies, policies, regulatory oversight, and federal systems throughout the formative stages of the nation's homeland security enterprise. He has also served as a research and development (R&D) program manager at the Federal Bureau of Investigation, National Institute of Justice, and DoD Counterdrug Technology Development. Mr. Blackburn holds bachelor's and master's degrees in electrical engineering from Virginia Tech. He serves on the advisory boards for the Bradley Department of Electrical and Computer Engineering at Virginia Tech and the U.S. Government Accountability Office's Science, Technology Assessment, and Analytics team.

PHILLIP COLELLA is senior mathematician and group leader for the Applied Numerical Algorithms Group in the Computing Sciences Directorate at the E. O. Lawrence Berkeley National Laboratory. He is a leader in the development of mathematical methods and computer science tools for science and engineering. His research has been in the area of high-resolution and adaptive methods for partial differential equations. He has also applied numerical methods in a variety of scientific and engineering fields, including shock dynamics, low-mach number and incompressible flows, combustion, porous media flows, and astrophysical flows. Dr. Colella is a member of the NAS and received A.B., M.A., and Ph.D. degrees from UC Berkeley, all in applied mathematics.

DORIN COMANICIU serves as senior vice president for artificial intelligence and digital innovation at Siemens Healthineers. His scientific contributions to computational imaging and machine intelligence have translated to multiple clinical products focused on improving the quality of care, specifically in the fields of diagnostic imaging, image-guided therapy, and precision medicine. Dr. Comaniciu is a member of the NAM and a Top Innovator of Siemens. He is a fellow of the IEEE, ACM, MICCAI Society, and American Institute for Medical and Biological Engineering. He is recipient of multiple honors, including an honorary doctorate and the IEEE Longuet-Higgins Prize for fundamental contributions to computer vision. Dr. Comaniciu is listed on Wikipedia's list of prolific inventors with 300 granted U.S. patents on health-care technology. He has co-authored 350 peer-reviewed publications, which have received 51,000 citations, with an h-index of 80. He is an advocate for technological innovation that saves and enhances lives, addressing critical issues in global health.

JAMES DARNELL is a 21-year veteran of the U.S. Secret Service, having served in protective and investigative assignments in Las Vegas, Nevada, Washington, DC, and Tulsa, Oklahoma. Mr. Darnell holds the position of assistant to the special agent in charge and is currently assigned to the National Computer Forensics Institute where he administers the Service's digital evidence research activities and runs a digital evidence laboratory. Mr. Darnell began his career with the Secret Service in 1999 and served as a special agent in the Las Vegas Field Office where he conducted criminal investigations, worked protective assignments, and received extensive training in the field of computer forensics. In 2005, he transferred to the Criminal Investigative Division in Washington, DC, where he served as the Service's program manager for computer forensics. Mr. Darnell is the current vice chair of the Scientific Working Group on Digital Evidence (SWGDE) and former chair of both SWGDE and the NIST Organization of Scientific Area Committees for the Forensic Science Digital Evidence Subcommittee. He provides presentations at the annual American Academy conference, periodically instructs at the National Computer Forensics Institute and for the agencies of the Treasury Computer Forensics Training Program, is a member of the ASTM E30 Committee, and is an adjunct professor at Oklahoma State University. He received his bachelor of science degree from the University of Nevada, Las Vegas.

MARK EDWARD DEAN has 34 years of industry experience as a design engineer, researcher, manager, and executive in computer technology R&D at IBM, including being an IBM fellow, vice president, and

chief technology officer. He also has 6 years of academic experience as a professor, researcher, and interim dean of the Tickle College of Engineering at the University of Tennessee, and 2 years as a consultant in advanced technology development and commercialization. Dr. Dean presently has 44 patents and is a member of the Inventors Hall of Fame. He is a member of the NAE and has numerous awards recognizing his career contributions to the computer technology industry, including Black Engineer of the Year, NAE member, IEEE fellow, and member of the National Academy of Inventors.

DON DETMER is a University Professor of Health Policy (Emeritus) and professor of medical education in the Department of Public Health Sciences at the University of Virginia School of Medicine, senior advisor to the American Medical Informatics Association, and visiting professor at the Center for Health Informatics and Multiprofessional Education (CHIME) at the University College of London. Dr. Detmer's research interests include national and international health information and communications policy, quality improvement, administrative medicine, vascular surgery, education of clinician-executives, and leadership of academic health sciences centers. He is the founder and co-chair of the Blue Ridge Academic Health Group, chair of the board of MedBiquitous, associate editor of AMIA's *Standards*, and a director of the Corporation for National Research Initiatives. He is a member of the NAM, a lifetime associate of the National Academies, a recipient of the NAM's Walsh McDermott medal for lifetime contributions, and a fellow of American Association for the Advancement of Science (AAAS), American College of Medical Informatics, American College of Surgeons, and American College of Sports Medicine (emeritus). Dr. Detmer is the immediate past president and CEO of the AMIA and is a past chairman of the NAM Board on Health Care Services, NLM board of regents, and the National Committee on Vital and Health Statistics. He sat on the Strategic Plan Work Group of the Policy Advisory Committee to the Office of the National Coordinator for Health Information Technology. He chairs the steering committee of the AMIA Global Partnership Program. Faculty appointments have been held at University of Wisconsin, Madison, University of Utah, University of Virginia, and Cambridge University. He served as vice president for health sciences at Utah and Virginia. His education and training include work at the University of Kansas, Johns Hopkins University, the National Institutes of Health, Duke University, the NAM, and Harvard Business School. He earned an M.A. from Cambridge University and an M.D. from the University of Kansas.

JUAN E. GILBERT received his M.S. and Ph.D. degrees in computer science from the University of Cincinnati in 1995 and 2000, respectively. He also received his B.S. in systems analysis from Miami University in Ohio in 1991. Dr. Gilbert is currently the Andrew Banks Family Preeminence Endowed Professor and Chair of the Computer and Information Science and Engineering Department at the University of Florida where he leads the Human Experience Research Lab. He has research projects in election security/usability/accessibility, advanced learning technologies, usability and accessibility, human-centered AI/machine learning and ethnocomputing (culturally relevant computing). He is an ACM fellow, a fellow of the AAAS and a fellow of the National Academy of Inventors. In 2012, Dr. Gilbert received the Presidential Award for Excellence in Science, Mathematics, and Engineering Mentoring from President Barack Obama. He also received the AAAS 2014 Mentor Award, the 2021 ACM SIGCHI Social Impact Award, and the 2018 Computer Research Association's A. Nico Habermann Award. Dr. Gilbert has served on three National Academies committees on projects such as "The Role of Authentic STEM Learning Experiences in Developing Interest and Competencies for Technology and Computing," "The Science of Effective Mentoring in Science, Technology, Engineering, Medicine, and Mathematics (STEMM)," and "The Future of Voting: Accessible, Reliable, Verifiable Technology."

GERARD J. HOLZMANN has expertise in software reliability and safety, code analysis, formal verification, software tools, and logic model checking. He is currently founder and consultant at Nimble Research. He was formerly fellow at NASA Jet Propulsion Laboratory (2003-2017), and before that, director of Computing Principles Research, Bell Laboratories (1983-2003). He is a member of the NAE (elected 2005). He has received the ACM Software Systems Award (2001), the NASA Software of the

Year Award (2013), and is an ACM fellow (2012). He was software lead on the National Highway Traffic Safety Administration/NASA investigation of unintended acceleration in Toyota vehicles (2010-2011). He holds eight U.S. patents and has written four technical monographs and approximately 80 technical papers. He has given approximately 120 invited talks at conferences, universities, and companies and has served on approximately 130 program committees. He was associate editor for *IEEE Software*, served on the review board of HRL Laboratories (2018) and the review panel of the NIST/ITL (2015), and was an invited expert at the U.S. Nuclear Regulatory Committee hearing in Washington, DC in 2011. He holds a Doctor h.c. from Twente University, The Netherlands (2006).

PETER M. KOGGE is the Ted H. McCourtney Professor of Computer Science and Engineering at the University of Notre Dame, a retired IBM fellow, and a founder of Emu Solutions, Inc. He is a fellow of both the IEEE and AAAS. His research interests are in massively parallel computing paradigms, processing in memory, and the relationship between massive non-numeric applications, emerging technology, and computer architectures. He holds over 40 patents and is author of two books, including the first text on pipelining. His Ph.D. thesis led to the Kogge-Stone adder used in many microprocessors. Other projects included EXECUBE—the world’s first multi-core processor and first processor on a DRAM chip, the IBM 3838 Array processor which was for a time the fastest floating point machine marketed by IBM, and the IOP—the world’s second multi-threaded parallel processor, which flew on every space shuttle. In 2008, he led the Defense Advanced Research Project Agency’s (DARPA’s) Exascale technology study group, which resulted in a widely referenced report on technologies and architectures for exascale computing. His startup, Emu Solutions, has demonstrated the first scalable system that utilizes mobile threads to attack large-scale big data and big graph problems. Dr. Kogge has received the Daniel Slotnick best paper award (1994), the IEEE Seymour Cray award for high-performance computer engineering (2012), the IEEE Charles Babbage award for contributions to the evolution of massively parallel processing architectures (2014), the IEEE Computer Pioneer award (2015) (highest award from IEEE Computer Society), and the Gauss best paper award for high-performance computers (International Supercomputing Conference 2015). Dr. Kogge holds a Ph.D. in electrical engineering from Stanford University (1973), an M.S in systems and engineering sciences from Syracuse University (1970), and a B.S. in electrical engineering from the University of Notre Dame (1968).

MING C. LIN is a Distinguished University Professor of Computer Science, a Dr. Barry Mersky and Capital One Endowed Professor, and a former Elizabeth Iribe Chair of Computer Science in the Departments of Computer Science (CS) and Electrical and Computer Engineering (ECE), and the University of Maryland Institute for Advanced Computer Studies (UMIACS) of the University of Maryland (UMD) at College Park. Before joining UMD, Dr. Lin was the John R. and Louise S. Parker Distinguished Professor at University of North Carolina, Chapel Hill (UNC), where she was a faculty for 20 years. Dr. Lin’s research interests include AI, robotics, computer graphics, virtual reality (VR), visualization, human-computer interaction, data analytics, scientific computing, modeling and simulation, with applications to health care, traffic analytics, design and prototyping. She has authored or co-authored more than 300 refereed scientific publications and five books in these areas. She has received several honors and awards, including the NSF Young Faculty Career Award, UNC Hettleman Award for Scholarly Achievements, Beverly W. Long Distinguished Term Professor, IEEE VGTC VR Technical Achievement Award, Washington Academy of Sciences Distinguished Career Award, and several best paper awards. She is a fellow of ACM, IEEE, and Eurographics, and ACM SIGGRAPH Academy. She is a member of CRA-WP board of directors, former chair of IEEE Computer Society (CS) Computer Pioneer Awards Committee and of IEEE CS fellows committee, and founding chair of ACM SIGGRAPH Outstanding Doctoral Dissertation Award Committee. She is an editor-in-chief emerita of *IEEE Transactions on Visualization and Computer Graphics*, a former IEEE CS board of governors member and a former chair of the IEEE CS Transactions Operations Committee. She has served on the steering/executive committees of many international conferences, NSF board of visitors, several other technical advisory boards for the government agencies, universities, industry, and the international

scientific research community. She has consulted for Amazon, Intel, and Microsoft; she was also the co-founder of Impulsonic, Inc., which was acquired by Val and its audio technology, Phonon, has been incorporated into SteamAudio for use by VR and other real-time applications.

CHRISTOPHER MANNING is the Inaugural Thomas M. Siebel Professor in Machine Learning in the Departments of Linguistics and Computer Science at Stanford University, director of the Stanford Artificial Intelligence Laboratory (SAIL), and an associate director of the Stanford Human-Centered Artificial Intelligence Institute (HAI). His research goal is computers that can intelligently process, understand, and generate human language material. Dr. Manning is a leader in applying deep learning to natural language processing, with well-known research on the GloVe model of word vectors, question answering, tree-recursive neural networks, machine reasoning, neural network dependency parsing, neural machine translation, sentiment analysis, and deep language understanding. He also focuses on computational linguistic approaches to parsing, natural language inference and multilingual language processing, including being a principal developer of Stanford dependencies and universal dependencies. Dr. Manning has co-authored leading textbooks on statistical approaches to natural language processing (Manning and Schütze 1999) and information retrieval (Manning et al., 2008), as well as linguistic monographs on ergativity and complex predicates. He is fellow of the ACM, the Association for the Advancement of Artificial Intelligence, and the Association for Computational Linguistics (ACL) and a past president of the ACL (2015). His research has won ACL, Coling, EMNLP, and CHI Best Paper Awards. He has a B.A. (Hons) from the Australian National University and a Ph.D. from Stanford University in 1994, and he held faculty positions at Carnegie Mellon University and the University of Sydney before returning to Stanford. He is the founder of the Stanford NLP group (@stanfordnlp) and manages development of the Stanford CoreNLP software.

MARI OSTENDORF, NAE, is a professor of electrical and computer engineering and an Endowed Professor of System Design Methodologies at the University of Washington, also serving as associate vice provost for research. Dr. Ostendorf's research focus is on machine learning for speech and language processing, particularly in conversational contexts. She received B.S., M.S. and Ph.D. degrees in electrical engineering (1980, 1981, 1985, respectively) from Stanford University and is a member of the NAE. Dr. Ostendorf's research area is known for having technology evaluation competitions, and many of the competitions associated with government-funded research programs were run by the National Institute of Standards and Technology (NIST). As such, she is familiar with some of the work of the NIST Information Technology Laboratory, particularly the Information Access Division. She also has extensive experience in serving on institute and center review panels in the United States and internationally.

J. MARC OVERHAGE, NAM, is the chief health information officer for Anthem, Inc., and previously served as the chief medical informatics officer for Cerner and Siemens Health Services. He is an internationally recognized expert in health information modeling, standards, and interoperability as well as clinical decision support, health services research, and implementation science. He is a member and fellow of the NAM, a master of the American College of Physicians, a member of the American College of Medical Informatics, and a fellow of the International Academy of Health Science Informatics. Dr. Overhage was awarded the Donald A.B. Lindbergh Award for Innovation from the American Medical Informatics Association in 2018. He served on a previous National Academies review panel for NIST as well as participating in a variety of National Academies panels, reports, and workshops.

STEPHANIE A. SCHUCKERS is the Paynter-Krigman Endowed Professor in Engineering Science in the Department of Electrical and Computer Engineering at Clarkson University and serves as director of the Center of Identification Technology Research (CITeR), an NSF Industry/University Cooperative Research Center. She received her doctoral degree in electrical engineering from University of Michigan. Dr. Schuckers' research focuses on processing and interpreting signals that arise from the human body.

Signals include the electrocardiogram and biometric signals like fingerprints and iris. Methods involve classic signal processing, statistical techniques, pattern recognition, algorithm development and evaluation, data mining, and image processing. Analysis is focused on data collected from cadaver, human, and animal studies. Her work is funded from various sources, including NSF, the Department of Homeland Security, and private industry, among others. She has started her own business, testified for Congress, and has over 40 journal publications as well as over 60 other academic publications.

PETER TU is GE Research's chief scientist for artificial intelligence. Dr Tu's primary interest is the field of computer vision. More recently, he has investigated a number of third-wave AI topics, including the grounding problem and emergent cooperation. In 1995, Dr. Tu received his doctorate in computer vision from Oxford University. He has over 75 peer reviewed publications and 50 patents. In 2015, Dr. Tu participated in a National Academies panel review of NIST research.

STEPHEN B. VARDEMAN is currently a university professor (through May 15, 2021, and emeritus professor thereafter) of statistics and industrial engineering at Iowa State University, where he has served as major advisor or co-major advisor for 32 Ph.D. graduates and 53 M.S. graduates. His professional interests and expertise include statistical machine learning, engineering and natural science applications of statistics, statistics and metrology, directional data analysis, process monitoring and control, reliability, industrial applications of statistics, statistical education, and the development of new statistical theory and methods. He holds degrees in statistics (Ph.D. 1975 from Michigan State University) and mathematics (M.S. and B.S., 1973 and 1971, from Iowa State University). He is a fellow of the American Statistical Association (ASA), an elected member of the International Statistical Institute, served on the ASA board of directors 2001-2003, was ISU LAS Kingland data analytics faculty fellow 2017 through 2019, and was editor of *Technometrics* from 1993 through 1995. He has authored or co-authored five textbooks, including the winner of the 1994 ASEE Meriam-Wiley Distinguished Author Award for an outstanding new engineering textbook, and co-edited *Statistical Methods for Physical Science, Volume 28*, in the Academic Press series *Methods of Experimental Physics*. He has published nearly 80 authored or co-authored refereed papers. He served on the National Academies Panel on Information Technology in 2009, 2011, and 2015 and on the ASA Task Force on Statistical Significance and Reproducibility 2020-2021.

ALYSON G. WILSON is the associate vice chancellor for national security and special research initiatives at North Carolina State University (NC State). She is also a professor in the Department of Statistics and principal investigator for the Laboratory for Analytic Sciences. Her research interests include statistical reliability, Bayesian methods, and the application of statistics to problems in defense and national security. Prior to joining NC State, Dr. Wilson was a research staff member at the IDA Science and Technology Policy Institute (2011-2013); an associate professor in the Department of Statistics at Iowa State University (2008-2011); a technical staff member in the Statistical Sciences Group at Los Alamos National Laboratory (1999-2008); and a senior statistician and operations research analyst with Cowboy Programming Resources (1995-1999). Dr. Wilson has served on several committees for the National Academies, including most recently the Committee on Improving Defense Acquisition Workforce Capability in Data Use, the Committee on Risk Analysis for Nuclear War and Nuclear Terror, and the Committee on Applied and Theoretical Statistics. Dr. Wilson is a fellow of the ASA and the AAAS. She received her Ph.D. in statistics from Duke University, her M.S. in statistics from Carnegie-Mellon University, and her B.A. in mathematical sciences from Rice University.

WING HUNG WONG obtained his Ph.D. in statistics from the University of Wisconsin, Madison, in 1980. He has held teaching positions at the University of Chicago, the Chinese University of Hong Kong, University of California, Los Angeles, and Harvard University. In 2004, he joined the faculty of Stanford University, where he is currently a professor of statistics, professor of biomedical data science, and holder of the Stephen R. Pierce Family Goldman Sachs Professorship in Science and Human Health. His

research interests include (1) mathematical statistics, where he studied the large sample properties of sieve maximum likelihood estimates in general spaces; (2) Bayesian statistics, where he developed sampling-based algorithms for Bayesian computational inference; and (3) computational biology, where he developed tools for the analysis of microarrays and sequencing data, and applied them to study gene regulatory systems. Dr. Wong was the winner of the COPSS Presidents' Award in 1993. He was elected to the NAS in 2009. He served on the editorial board of *Proceedings of the National Academy of Sciences U.S.A.* from 2009-2015. He had participated in the writing of the National Research Council report *Mathematics and 21st Century Biology* (2005).

MARY ELLEN ZURKO is a technical staff member at Massachusetts Institute of Technology (MIT) Lincoln Laboratory. Ms. Zurko has experience in cybersecurity architectures, cybersecurity analysis, product development and prototyping, and applied research. She holds over 20 patents. She defined the field of user-centered security in 1996. She was security architect doing DevSecOps of one of IBM's earliest clouds. Her previous research includes authorization policies, high assurance virtual machine monitors, the web, and PKI. She was a founding member of the National Academies' Forum on Cyber Resilience and serves as a Distinguished Expert for NSA's Best Scientific Cybersecurity Research Paper competition. Ms. Zurko received S.B and S.M. degrees in computer science from MIT.