IT Professional’s 20th Anniversary
IEEE COMPUTER SOCIETY: Be at the Center of It All

IEEE Computer Society membership puts you at the heart of the technology profession—and helps you grow with it.

Here are 10 reasons why you need to belong.

1. Training that sharpens your edge in Cisco, IT security, MS Enterprise, Oracle and more.

2. Certifications and exam preparation that set you apart.

3. Industry intelligence, including Computer, myCS, Computing Now, and myComputer.

4. Over 200 annual conferences and technical events held worldwide.

5. 300-plus chapters and more than 30 technical committees keep you connected.

6. Access to hundreds of books, 13 technical magazines and 20 research journals.

7. Deep discounts on magazines, journals, conferences, symposia and workshops.

8. Opportunities to get involved through speaking, publishing and volunteering opportunities.

9. Scholarships awarded to computer science and engineering student members each year.

10. A robust jobs board, plus videos, articles and presentations to help you land that next opportunity.

IEEE Computer Society—keeping you ahead of the game. Get involved today.

www.computer.org/membership
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Retrospectives and Reflections: 20 Years of IT Professional</td>
<td>Sorel Reisman, Wushow &quot;Bill&quot; Chou, Frank E. Ferrante, Simon Y. Liu, and San Murugesan</td>
</tr>
<tr>
<td>27</td>
<td>Effectiveness of Test-Driven Development and Continuous Integration: A Case Study</td>
<td>Chintan Amrit and Yoni Meijberg</td>
</tr>
<tr>
<td>36</td>
<td>Automatic Annotation of Text with Pictures</td>
<td>J. Kent Poots and Ebrahim Bagheri</td>
</tr>
<tr>
<td>45</td>
<td>The Design of a Software Engineering Lifecycle Process for Big Data Projects</td>
<td>Yen-Tai Lin and Sun-Jen Huang</td>
</tr>
<tr>
<td>53</td>
<td>Improving Transparency and Efficiency in IT Security Management Resourcing</td>
<td>Knut Haufe, Srdan Dzombeta, Knud Brandis, Vladimir Stantchev, and Ricardo Colomo-Palacios</td>
</tr>
<tr>
<td>5</td>
<td>FROM THE EDITORS</td>
<td>IT Professional Enters its 20th Year with Confidence</td>
</tr>
<tr>
<td>9</td>
<td>IT TRENDS</td>
<td>Computer Science Education in 2018</td>
</tr>
<tr>
<td>15</td>
<td>INTERNET OF THINGS</td>
<td>Introducing the Internet of Things Department</td>
</tr>
</tbody>
</table>
64  CYBERSECURITY
Bad Security Metrics Part 1: Problems
David Flater

69  DATA ANALYTICS
The Role of a Customer Data Platform
Seth Earley

77  LIFE IN THE C-SUITE
Life in the C-Suite 2.0
Stephen J. Andriole

80  STUDENT FORUM
Introducing the Student Forum Column
Gustavo Rossi

83  IT ECONOMICS
Introducing the IT Economics Department
Nir Kshetri

Also in this Issue
3  Masthead
88  CS Information

For more information on computing topics, visit the Computer Society Digital Library at www.computer.org/cSDL.
Looking for the BEST Tech Job for You?

Come to the Computer Society Jobs Board to meet the best employers in the industry—Apple, Google, Intel, NSA, Cisco, US Army Research, Oracle, Juniper...

Take advantage of the special resources for job seekers—job alerts, career advice, webinars, templates, and resumes viewed by top employers.

www.computer.org/jobs
IT Professional Enters its 20th Year with Confidence

Welcome to the 20th year of IT Professional! The magazine is entering 2018 as an established, highly reputable publication for developers and managers of enterprise information systems. Thanks to our highly selective team of editors and the exceptional support of the IEEE Computer Society (CS) staff, IT Pro is growing and thriving. As the incoming editor in chief (EIC), I would like to present an overview of IT Pro’s current state and where we are headed.

EDITORIAL BOARD

IT Pro has a diverse, highly reputable, and very active editorial board (EB). All EB members support the magazine by managing the peer review process of articles assigned to them and lending their expertise to all areas of the magazine. Some EB members have traditionally served as associate editors in chief (AEICs), column/department editors, special issue guest editors, and conference liaisons. This year, we are adding new roles, including issue management editor, outreach editor, web editor, history and analytics editor, secretary, and awards chair.

For 2018, the CS Publications Board (PB) approved the following AEICs:

- Reza Djavanshir, John Hopkins University (AEIC Regular Papers);
- Charalampos Z. Patrikakis, Piraeus University of Applied Sciences (AEIC Special Issues); and

In addition, the IT Pro EB welcomes four new members, also approved by the PB.

Maria José Escalona, co-editor of the new Student Forum department along with Gustavo Rossi, is a full professor in the Department of Computer Languages and Systems and director of the Web Engineering and Early Testing research group at the University of Seville. Her research interests include requirements engineering, web system development, model-driven engineering, early testing, and quality assurance. Escalona received a PhD in computer science from the University of Seville. She is a member of INPRO, the public company of the Seville government for IT. Contact her at mjescalona@us.es.

Nir Kshetri, editor of the new IT Economics department, is a professor at the University of North Carolina–Greensboro and a research fellow at Kobe University. He is also a consultant for the Asian Development Bank and has participated as a lead discussant at the peer review meeting of the UN’s Information Economy Report. Kshetri received a PhD in business administration from the University of Rhode Island. He is the author of seven books and more than 110 articles in various publications. His book Global Entrepreneurship: Environment and Strategy
(Routledge, 2014) was selected as an Outstanding Academic Title by Choice magazine. Kshetri is the winner of 2016 Bryan School Senior Research Excellence Award, and a two-time winner of the Pacific Telecommunication Council’s Meheroo Jussawalla Research Paper Prize. Contact him at nbkshetr@uncg.edu.

Kincho H. Law is a professor of civil and environmental engineering at Stanford University. His research interests include the innovative use of computational and information science in engineering. Law received a PhD in civil engineering from Carnegie Mellon University. He has authored and co-authored more than 400 articles in journals and conference proceedings and has received best research paper awards from the American Society of Civil Engineers (ASCE), the American Society of Mechanical Engineering (ASME), IEEE, and the Digital Government Society. Law received the ASCE Computing in Civil Engineering Award in 2011, and was elected as a Distinguished Member of the ASCE and as a Fellow of the ASME in 2017. Contact him at law@stanford.edu.

Gianfranco Politano, web editor along with EB member Arpan Pal, is an assistant professor in the Department of Control and Computer Engineering at Politecnico di Torino. His research interests include the analysis of gene regulatory networks (GRNs) with particular emphasis on networks topological properties analysis and GRN simulation. Politano’s efforts are mostly focused on pattern analysis for recurring post-transcriptional regulatory motif searching in GRNs, and topological analysis of GRNs for highlighting potential molecular targets for new drugs (pharmacogenomics). Politano received a PhD in computer engineering and information technology from Politecnico di Torino. He has been part of the organizing committee of several IEEE conferences. Contact him at gianfranco.politano@polito.it.

SUBMISSIONS

IT Pro strives to publish only exceptional, innovative articles that highlight advances in IT, discuss their effect on our work and lives, and address related challenges and risks. Over the past three years, IT Pro has had 354 article submissions and accepted 69 papers, with an approximate acceptance rate of 19 percent.

In addition to regular papers, IT Pro publishes special issues (SIs), each of which is curated by a team of subject-matter experts as guest editors (GEs).

In 2017, IT Pro published two issues with regular papers and the following four SIs:

- Mobile Data Analytics (May/June 2017),
- Cognitive Computing (July/August 2017),
- Trusting the Internet of Things (September/October 2017), and
- Graph Databases (November/December 2017).

All published issues can be accessed from the History tab at www.computer.org/it-professional or directly from IEEE Xplore or the CS Digital Library (CSDL).

There are three SIs planned for 2018:

- Financial Technologies and Applications (March/April 2018),
- Cyberthreats and Security (May/June 2018), and
- Connected and Autonomous Driving (November/December 2018).

Submissions for the first two are in review, and the call for papers is open for the third until April 2018 (see publications.computer.org/it-professional/2017/10/24/connected-autonomous-driving).

COLUMNS AND DEPARTMENTS

IT Pro EB members write or invite authors to write opinion or research-based articles in the following columns and departments.
FROM THE EDITORS

• Columns: From the Editors, IT and Future Employment, IT Trends, Mastermind, Life in the C-Suite, Student Forum (new)
• Departments: Cybersecurity (new name for Securing IT), Data Analytics, Extreme Automation (new), Internet of Things (new), IT Economics (new), Spotlight

DELIVERY METHODS

IT Pro is delivered through myCS (epubs, PDF, and HTML), the myComputer app, the CSDL (PDF and HTML), and IEEE Xplore (PDF and HTML). The magazine’s best articles are also featured in Computing Now and ComputingEdge, as well as on IT Pro’s website (www.computer.org/it-professional).

COLLABORATION WITH CONFERENCES

IT Pro has partnered for three years with COMPSAC (www.compsac.org) to cosponsor and organize the IT in Practice (ITiP) symposium. Submissions and participation have been solicited for the fourth ITiP, which will be held in July 2018 in Tokyo, Japan. IT Pro also plans to partner with the Software Technology Conference (STC) and to implement the J1C2 (journal first, conference second) and C1J2 (conference first, journal second) publication models.

IN THIS ISSUE

To celebrate the magazine’s 20 successful years, this issue includes retrospectives and reflections from the Advisory Board Chair Sorel Reisman, the inaugural EIC Wushow “Bill” Chou, and former EICs Frank E. Ferrante, Simon Y. Liu, and San Murugesan. The issue also introduces the new IT Economics and Internet of Things departments, the new Student Forum column, and the new editor of Life in the C-Suite. In addition, this issue presents four feature articles.

In “Effectiveness of Test-Driven Development and Continuous Integration: A Case Study,” Chintan Amrit and Yoni Meijberg present a case study where a Dutch small-to-medium enterprise (SME) implemented test-driven development and continuous integration. They discovered a higher number of defects compared to a baseline case study, and an increase in the focus on quality and test applications, while considering customer acceptance. The company now has infrastructure in place to further evaluate other software process improvement initiatives.

In “Automatic Annotation of Text with Pictures,” J. Kent Poots and Ebrahim Bagheri address the challenges related to interpreting large volumes of data, especially for users who need cognitive assistance. The authors describe the techniques that help users understand text through automated text picturing and explain the constituent steps of knowledge extraction, mapping, and scene rendering. Finally, they present the application areas and compare eight “real-world” picturing systems with defined use cases for each.

In “The Design of a Software Engineering Lifecycle Process for Big Data Projects,” Yen-Tai Lin and Sun-Jen Huang present a software engineering lifecycle process for big data projects that is based primarily on the international standard ISO/IEC 15288:2008–Systems and Software Engineering–System Lifecycle Processes. The designed process has been evaluated through big data projects from banking and retail industries in Taiwan.

Finally, in “Improving Transparency and Efficiency in IT Security Management Resourcing,” Knut Haufe, Srdan Dzombeta, Knud Brandis, Vladimir Stantchev, and Ricardo Colomo-Palacios present a resource management process for information security management systems that is based on the international standards of the ISO/IEC 27000 family. They state that the process can be used to more transparently plan and assign costs of controls and can be implemented by any organization regardless of type, size, or nature. The authors also present results from expert evaluation of the process, summarize the main findings, and discuss future work.
IT Pro is highly appreciative to all authors, reviewers, and readers for their contributions. The magazine would not exist without the strong support from its EB and the CS staff. We welcome your comments and suggestions for sustaining IT Pro as the premier forum for IT professionals. I look forward to serving as the EIC of IT Pro.

ABOUT THE AUTHOR

Irena Bojanova is a computer scientist at NIST. Previously, she was a program chair at University of Maryland University College (UMUC), an academic director at Johns Hopkins University’s Center for Talented Youth (JHU-CTY), and a co-founder of OBS Ltd. (now CSC Bulgaria). She earned a PhD in mathematics/computer science from the Bulgarian Academy of Sciences. Bojanova is a Senior Member of IEEE and serves as editor in chief (EIC) of IT Professional, co-chair of the IEEE Reliability Society's Internet of Things Technical Committee, and a founding member of the IEEE Special Technical Committee (STC) on Big Data. Previously, she was the founding chair of the IEEE CS Cloud Computing STC, interim EIC of IEEE Transactions on Cloud Computing, and Member at Large and Integrity Chair of the IEEE CS Publications Board. Contact her at irena.bojanova@computer.org.
Computer Science Education in 2018

For the January 2018 issue of Computer, six senior computer science educators participated in a Virtual Roundtable where they were asked about how universities are preparing students to deal with contemporary IT challenges, including social networking, false information, and other subjects we see in the news today. For this installment of IT Trends, we asked the same six educators—Michael Lewis (College of William and Mary), Keith Miller (University of Missouri–St. Louis), Shiuhpyng Shieh (National Chiao Tung University), Phillip A. Laplante (Pennsylvania State University), Jon George Rokne (University of Calgary), and Jeff Offutt (George Mason University)—three questions about the current state of computer science education, software engineering, and licensing software engineers using a “Body of Knowledge” approach.

What are today’s core classes in computer science (CS) education? Are they generally uniform in most universities and colleges? How do they compare with those in the early days of CS education (1970s and 1980s)?

LEWIS: Looking at various programs, I was a bit surprised to see the extent to which vestiges of the old core classes are still around: discrete math, data structures and algorithms, programming languages, computer organization or computer architecture, and some sort of software development course. There is also a fairly ubiquitous math requirement of two to three semesters of calculus plus linear algebra.

It’s surprising how static the core of the curriculum seems to be, given how much CS has changed over the years. Of course, the content of the core CS classes has changed in varying degrees, and the electives are far more varied and numerous than what we had in the past. Indeed, many of the topics of electives today did not exist in the 1970s and 1980s.

ROKNE: There has not been a general agreement on what exactly is the core of CS, neither today nor in the past. However, one can probably say that today’s core classes in CS are generally focused on procedural programming, data structures, and algorithm analysis supported by courses in mathematics and statistics, and that in the early days of CS there was a greater emphasis on lower-level programming and courses dealing with computer hardware, and fewer courses on data structures and software engineering.

SHIEH: Today’s core classes include programming (algorithms, programming languages, and data structure), mathematics (linear algebra and discrete mathematics), and system design (computer architectures and operating systems). Although the objectives of the core classes remain the same, the content varies significantly in comparison with that of the 1970s and 1980s. It in-
cludes many new techniques we take for granted today, such as multitasking, just-in-time compilation, networking, and artificial intelligence. Moreover, the core classes need to cover new requirements.

MILLER: One way to answer this question is to reference the ACM/IEEE Computer Society Computer Science Curricula 2013 (CS2013; www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf). It’s not the only approach to the computing curriculum available on the world stage (for example, the European Union is working on the Bologna Process, which includes computing); however, CS2013 approaches curriculum not on the basis of “classes” but on the basis of “hours.” An hour is meant to be the amount of material covered in an hour of “lecture,” although the CS2013 document takes pains to not endorse lecture as the preferred method of pedagogy.

See Table 1, which summarizes CS2013 and two previous curriculum guidelines from the same group.

<table>
<thead>
<tr>
<th>Knowledge area</th>
<th>CS2013 Tier 1</th>
<th>CS2013 Tier 2</th>
<th>CS2008 Core</th>
<th>CS2001 Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL: Algorithms and complexity</td>
<td>19</td>
<td>9</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>AR: Architecture and organization</td>
<td>0</td>
<td>16</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>CN: Computational science</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DS: Discrete structures</td>
<td>37</td>
<td>4</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>GV: Graphics and visualization</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>HCI: Human–computer interaction</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>IAS: Information assurance and security</td>
<td>3</td>
<td>6</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IM: Information management</td>
<td>1</td>
<td>9</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>IS: Intelligent systems</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>NC: Networking and communications</td>
<td>3</td>
<td>7</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>OS: Operating systems</td>
<td>4</td>
<td>11</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>PBD: Platform-based development</td>
<td>0</td>
<td>0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>PD: Parallel and distributed computing</td>
<td>5</td>
<td>10</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>PL: Programming languages</td>
<td>8</td>
<td>20</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
Although they aren’t defined as such, it seems clear that several classes familiar to most CS graduates can be identified by locating large numbers of hours in Table 1: algorithms (AL), computer architecture (AR), discrete structures (DS), programming languages (PL), software development fundamentals (SDF), and software engineering (SE) stand out to me. Personally, I agree that these six areas, covered by some collection of courses, would be fundamental for any undergraduate degree in CS. I wouldn’t think these six would be sufficient, but they certainly would be a base upon which to build.

We can go back to 1968 to compare early CS curricula to the 2013 recommendations. An ACM task force in 1968\(^2\) recommended eight courses in the core curriculum.

1. Introduction to Computing
2. Computers and Programming
3. Introduction to Discrete Structures
4. Numerical Calculus
5. Data Structures
6. Programming Languages
7. Computer Organization
8. Systems Programming

We see strong similarities with the 2013 document, but several noteworthy differences. Calculus, though discussed in the 2013 document, is not explicitly mentioned in the 2013 core, though discrete structures are. Software engineering appears in 2013, but is not explicit in 1968. Perhaps most striking is how many hours in 2013 are given over to topics not explicitly covered in the 1968 core, including social issues and practice, parallel and distributed computing, and intelligent systems. In the decades between 1968 and 2013, it isn’t surprising that new topics gained importance; perhaps it’s more surprising how similar many of the emphases are.

**LAPLANTE:** I’ve been teaching in computer science and software engineering programs since the mid-1980s. Over the years, I have seen curricular changes to make room for breadth courses unrelated to CS, and to make the programs more accessible to those who are not strong in mathematics. I believe that these changes come at the expense of a deeper understanding of computation.

ABET (the Accreditation Board for Engineering and Technology) accredits CS programs through guidelines provided by the CSAB (Computer Science Accrediting Board)—a joint effort of the IEEE Computer Society and ACM. The current guidelines say that a CS curriculum must have:

| SDF: Software development fundamentals | 43 | 0 | 47 | 38 |
| SE: Software engineering | 6 | 22 | 31 | 31 |
| SF: Systems fundamentals | 18 | 9 | -- | -- |
| SP: Social issues and professional practice | 11 | 5 | 16 | 16 |
| **Total core hours** | **165** | **143** | **290** | **280** |
| **All Tier 1 + All Tier 2** | | | | **308** |
| **All Tier 1 + 90% of Tier 2** | | | | **293.7** |
| **All Tier 1 + 80% of Tier 2** | | | | **279.4** |
• coverage of the fundamentals of algorithms, data structures, software design, and concepts of programming languages and computer organization and architecture;
• an exposure to a variety of programming languages and systems;
• proficiency in at least one higher-level language; and
• one year of science and mathematics.

Most CS programs comply with these guidelines with little variation.

Notice that the ABET recommendations omit operating systems, compiler theory, automata theory, database theory, and more—courses that were traditionally taught in most CS programs through at least the late 1990s and that remove the “magic” from how computers and software applications really work. The ACM/IEEE Computer Society 2016 curricular recommendations (www.acm.org/education/curricula-recommendations) provide more depth beyond the ABET criteria, but I’m not sure how widely adopted these are.

OFFUTT: ABET accredits most CS programs in the US, and as a result, the requirements, at least in North America, tend to be fairly uniform. Students from the 1980s (like me) still recognize many of the year one through three courses. Courses like introductory CS, introductory programming, data structures, computer organization and assembly, operating systems, computability, and algorithms are still standard. At the higher level, we see newer topics like security, web and mobile app development, game development, and big data analysis alongside standbys like database, graphics, networks, and AI. I just looked at the requirements at my university (George Mason) and the coursework we had in 1980 would come very close to satisfying current requirements. I find that surprising.

“The brightest computer science graduates are often heavily self-taught due to their passion for this area.” Is this statement true? How often do you experience cases where the students know more than their professors?

MILLER: In my experience, that statement is true, although I don’t think it’s always the brightest students who are heavily self-taught. I think it’s often the most passionate students who are heavily self-taught. And this self-teaching is often in specific, mostly concrete areas. Students might know quite a bit about how to program particular machines or systems, even though their understanding of algorithms in general (for example) might not be particularly sophisticated. I don’t think I can quantify how often this happens to me, but surely in a class of 25 undergraduates it happens several times in a semester where at least one student knows some detail about a particular system or programming language that I don’t know (or don’t remember).

LEWIS: I agree that the brightest CS students generally learn a lot on their own or by working with equally bright students. These students often participate in extra-curricular activities such as personal programming projects and contests, hackathons, and various jobs and internships. This plays a critical role in the development of CS students.

In my experience it’s pretty common for students to know more about particular software tools or frameworks than their professors do. Students frequently return from internships with all manner of knowledge such as web programming frameworks. Also, in their personal software projects they end up doing things that faculty might not typically do (such as hacking Bluetooth drivers on cellphones).

On the other hand, CS faculty tend to know more about the science part of computer science than students, so we’re probably worth keeping around.

LAPLANTE: My experience is that the best young computer scientists have a solid undergraduate education that’s supplemented with lots of hands-on experience obtained through some combination of internships, part-time jobs, and self-study. There are so many open source projects to work on, free tools, and low-cost small platforms—such as Arduino and Raspberry Pi—to play with that there’s no excuse for young computer scientists to not have lots of hands-on experience by the time they graduate. The best students take advantage of these opportunities.

As for students knowing more than their professors—in one area or another, my students know more than me all the time. I can’t be an expert in every programming language, development environment, application domain, or piece of hardware. I constantly learn from my students.
SHIEH: In computer security discipline, both attackers and ethical hackers are often self-taught. More and more self-learning resources are now available on the Internet. The knowledge and implementation skills in CS can be easily digitalized and distributed online, and therefore passion can motivate students to polish their skills through self-learning. As CS domain knowledge expands and grows rapidly, it might be a challenge for professors to keep up and in particular to provide cross-disciplinary hands-on experience. For example, Mark Zuckerberg of Facebook took a year to build a smart home that can follow speech commands, control switches, and even tell a joke. However, in my personal opinion, the academic way of thinking and problem solving can still be the key to the success of new technologies.

ROKNE: It’s certainly true that some of brightest computer science students are heavily self-taught. However, their knowledge base tends to have gaps that need to be filled in through more formal education processes. This was exemplified by one of our very successful undergraduate students who was challenged in his graduate work elsewhere by the advanced theoretical CS courses expected of a graduate student. This student had avoided most theoretical courses offered in his undergraduate studies. It’s not difficult for a bright computer science student to know more than a professor if knowledge is measured by detailed knowledge of specific software or hardware products. For a deep understanding of CS, though, professors are seldom challenged by students.

OFFUTT: I’ve only seen two or three students in my 30-year career who are primarily self-taught to be software engineers. Many are self-taught to be programmers, but most were bad programmers. Certainly not engineers. And many students have very high self-efficacy, believing that they already know everything because they’ve written a few Android apps. I see a few seniors surpass their good teachers, and always find that exhilarating. Unfortunately, I see more students surpass teachers because their teachers do not know much.

The IEEE Computer Society has developed a Software Engineering Body of Knowledge (SWEBOK) and a related examination to become a licensed software engineer. Do you see any impact from this now or in the future?

OFFUTT: Not personally. The last time I looked at the SWEBOK it struck me as out of touch with modern software engineering. It looked like something that might have been appropriate in 1995, not 2017. That’s just my opinion, of course.

ROKNE: It’s difficult to assess the impact of the SWEBOK in a CS department because it’s specifically aimed at software engineering (incidentally, a term coined by F.L. Bauer). CS as we think of it encompasses software engineering but includes other disciplines too. Splitting off new departments of software engineering from CS would mean less interaction and cross-disciplinary research in the overall CS area.

LEWIS: I have not seen any impact of this, even though my state (Virginia) was one of those that first asked for a licensure system. It’s interesting how many of the core topics in the SWEBOK are those core CS courses discussed in the first question (discrete math, data structures and algorithms, software development, and computer architecture). This supports the view that the CS curriculum really has an identifiable core.

SHIEH: The SWEBOK contains necessary knowledge in a summarized form. It will be very helpful to new graduates. In this case, the licensing test might be a good way of ensuring the fundamental knowledge of developers.

LAPLANTE: I led the effort to create the professional engineer (PE) licensing exam for software engineers in the US, so I can answer this question from a unique perspective. For several reasons a different body of knowledge, similar to SWEBOK, had to be created for the licensing exam. We put a tremendous amount of effort into surveying hundreds of professionals, creating the BOK, and writing (and maintaining) the exam. The reasoning and process behind the exam are described in “A Principles and Practices Exam Specification to Support Software Engineering Licensure in the United States of America.”

Unfortunately, since its introduction in 2013 there have been few exam takers (less than 100). There are several reasons for this. First, (with some exceptions) deans, department chairs, and
faculty seem generally uninterested in promoting the exam. Secondly, state departments of engineering have not been uniformly requiring software PEs for public works that contain software. Without this requirement, there is little demand for licensed software engineers. Finally, the path to licensure is difficult because candidates must pass the Fundamentals of Engineering exam. This examination is very broad and expects the candidate to be knowledgeable in areas that most engineers would study, but most software engineers would not.

**MILLER:** The issue of licensing software engineers has a long and contentious history. The issue of whether software engineers should, or will, be licensed has certainly not been settled, despite decades of hard work by several organizations. For a contrary view in the popular press, see “Programmers: Stop Calling Yourselves Engineers.”

Because of the lack of consensus among professional organizations, I don’t think that licensing will be important in the next few years. It might gain momentum if there are many spectacular disasters. Meanwhile, the SWEBOK might improve our understanding of the principles important to our profession, and might have a more immediate and direct effect on curricula and practice.

**CONCLUSION**

Along with the Virtual Roundtable in the January 2018 issue of *Computer*, we hope this column helps readers get a better understanding as to where computer science education is today. We were pleased to see differing viewpoints to what we believe were challenging questions. We leave it to the readers to make judgments based on their own experiences.

**REFERENCES**


**ABOUT THE AUTHORS**

Jeffrey Voas is a computer scientist at NIST and an IEEE Fellow. Contact him at j.voas@ieee.org.

Rick Kuhn is a computer scientist at NIST and an IEEE Fellow. Contact him at kuhn@nist.gov.

Celia Paulsen is an IT specialist at NIST. Contact her at celia.paulsen@nist.gov.

Kim Schaffer is an IT specialist at NIST. Contact him at kim.schaffer@nist.gov.
Introducing the Internet of Things Department

Welcome to the Internet of Things, a new department with the mission of presenting fresh ideas and applications from a practitioner point of view.

We’re interested in showcasing articles about real, implemented Internet of Things (IoT) systems—not theoretical treatments or laboratory-based proofs of concept. There are all too many of the latter and not enough of the former.

This is especially true with academic papers about potential applications that seem very distant from the realities of implementation.

The articles we’re interested in featuring can take many different forms, but here are some examples:

- Descriptions of deployed IoT systems, particularly in “surprising” application domains.
- Reports from workshops, panel discussions, and practice-focused roundtables of professionals and researchers.
- Surveys and reviews of tools for building IoT systems, including examples of deployed systems built using the tools.
- Thoughtful discussions of societal, legal, and ethical issues surrounding the deployment of IoT applications.

When building real systems, failure is inevitable. But we often learn more from these failures than from our successes. Therefore, we invite articles that share unsuccessful experiences or lessons learned from failed (or moderately successful) IoT system deployment. We realize that it might be difficult to expose failures to the public domain, but there are positive ways to do so.

IOT FOR HEALTHCARE PANEL DISCUSSION

The following is an example of one type of article we’d like to showcase. It’s a report based on a two-hour panel discussion on IoT in healthcare, held at the NIST offices in Gaithersburg, Maryland on 30 August 2017. The panel was part of a one-day workshop on IoT sensors hosted by NIST and the IEEE Sensors Council. Column editor Phil Laplante was the moderator, and participants included column editor Ben Amaba and other experts with experience building and/or sponsoring deployed IoT healthcare applications. The other panelists were:

- Seth Carmody, cybersecurity program manager for the Center for Devices and Radiological Health (CDRH), serving as co-chair of CDRH’s Cybersecurity Working Group;
- Venky Karuppnan, CEO of Teezle, a leading IoT platform company;
- Mansur Hasib, cybersecurity leader, keynote speaker, author, and media commentator;
Marc Wine, subject matter expert in many areas including federal health policy and technology innovation, health IT and informatics, mobile health applications, and the Nationwide Health Information Network; and

Ken Blount, infrastructure project lead at Program Executive Office Healthcare Management Systems, Department of Defense.

The intention of this panel discussion was to harvest mindshare from these practitioners to provide guidance for those building IoT healthcare applications. The discussion consisted of opening statements, a set of prepared questions, and closing statements. For brevity, here we provide a few of the questions and highlights from the answers.

**Where Is IoT for Healthcare on the Gartner Hype Cycle?**

Panelists observed that the stages of Gartner’s Hype Cycle for Emerging Technologies (see Figure 1) represent the changing status of a new technology from its introduction (technology trigger) through various aspects of overstated expectations and disillusionment with the slow pace of realizing the expectations, to the realization of the technology’s true capabilities and realities (enlightenment), to productive creation and deployment of systems using the technology.

![Figure 1. Simplified Gartner Hype Cycle.](image)

Panelists suggested that while IoT for healthcare is still somewhere between disillusionment and enlightenment, we need common standards, security (especially regarding human–machine interaction), interoperability, silo breakdowns, and a clearer definition of horizontal and vertical application layers to reach the plateau of productivity. Furthermore, we need a good data governance program, better requirements and architectures, and an improved understanding of human behavior.

IoT applications generate a lot of data, and the healthcare domain has a history of very advanced data collection, as opposed to other application domains that are beginning to enjoy the benefits of IoT. Panelists discussed the best uses of new kinds of data being collected by IoT applications. For example, there is clearly value in predictive, proactive analytics, so many companies will exploit this reality. But is there a business case for real-time reactive analytics that companies will pursue?
What Successes Have Been Achieved and What Lessons Have Been Learned?

Panelists noted that there have been many successes for IoT in the healthcare space. For example, patients can have certain organs such as the heart connected to the Internet through various types of monitors. The US National Kidney Registry uses IoT-enabled transport containers to track donor organ movement from harvest to implantation. Home telemedicine for veterans is being widely used by the US Department of Veterans Affairs and has potential for significant growth.

But several panelists wondered if we’ve done enough to create “trustworthy” healthcare systems. For example, it was recently shown that IoT-enabled devices that contain accelerometers for motion and location sensing can be disrupted through acoustic attacks.

What Challenges are Ahead?

Panelists agreed that there are many challenges, but that there’s also a great deal to be learned from these challenges. For example, all panelists agreed that security, particularly for personal identifying information, is a big problem. One panelist noted that if an application employs “user id” and “password” on IoT healthcare applications, the consequences of loss or theft of this information could be deadly. To prevent unintended consequences, when an IoT medical device moves out of its intended context, alarms should be set off. A lesson learned from these observations is that we need continuous diagnostics in mitigation of cybersecurity.

Blockchain is widely thought to be a potential technology to provide security, privacy, and reliability in the IoT space. But panelists wondered if blockchain is ready for prime time.

Others noted that a huge challenge for IoT-enabled healthcare applications is opposition from organizations reluctant to expose their products and devices to the Internet because of security and intellectual property concerns. None of the medical IoT devices “speak the same language,” which also arises partly from defending proprietary boundaries.

Medical devices need to be tamper-proof out of the box, but medical device experts shouldn’t have to become cybersecurity experts. Similarly, every medical device and IoT healthcare solutions provider shouldn’t need to be full-blown software companies. All of this means that we need off-the-shelf software components and solutions. One of the panelists suggested that licensing of software engineers working on critical infrastructure systems, as is done in many US states, would be essential for IoT-connected medical devices to ensure patient safety and privacy.

What Is Needed from Government/Industry/Academia to Move the Ball Forward?

Panelists pointed out that, in general, what is needed from all players comprises three categories: operational efficiency, better services, and applications that are closer to the customer. To achieve these goals, better information and experience sharing is essential. All 50 US states are supposed to share such information, but this isn’t happening. Moreover, we need to share models (such as semantics models for data collection) and not just anecdotes, and we need to use this information in a meaningful way.

Platform standardization is another important area where progress needs to be made. It’s unclear which platform vendor will emerge as the leader, but that leader will need to help companies build applications and monetize them, thus enhancing the virility of the platform.

Finally, we need to eliminate fragmentation among solutions and providers. Achieving this goal requires open data and open architecture standards. An iterative, actively engaged solutions development community is also needed.
CONCLUSION

Many of the panel’s recommendations—the need for information sharing, open components and solutions, standardization, and learning from failures—reinforce the mission of our new department. We’d love to consider your contributions along these lines, but please query us before sending an article. Articles should be around 2,500 words (including figures and tables, which are considered 250 words each). Articles are reviewed by us and we might ask other experts to review it as well—submission is not a guarantee of publication. Because we have very limited editorial capability, submitted columns must be ready to go; that is, they must be well written and grammatically correct, and they must conform to the Computer Society’s style guide (including all references in the correct format); see www.computer.org/cms/Computer.org/Publications/docs/2016CSStyleGuide.pdf. We look forward to reading your submissions!

REFERENCE


ABOUT THE AUTHORS

Phillip A. Laplante is a professor of computer science at Penn State. Contact him at plaplante@psu.edu.

Ben Amaba is a Global Executive for the IBM Hybrid Cloud Division. Contact him at baamaba@us.ibm.com.
Retrospectives and Reflections

20 Years of *IT Professional*

**Sorel Reisman**  
California State University

**Wushow “Bill” Chou**  
North Carolina State University

**Frank E. Ferrante**  
College of William & Mary

**Simon Y. Liu**  
Agricultural Research Service

**San Murugesan**  
BRITE Professional Services

Commemorating the 20th anniversary of *IT Professional*, the advisory board chair and four past editors in chief reflect on the first two decades of the magazine.

To celebrate 20 successful years of *IT Professional*, we asked the current advisory board chair, Sorel Reisman, and four former editors in chief (EICs)—Wushow “Bill” Chou, Frank E. Ferrante, Simon Y. Liu, and San Murugesan—to reflect on the history of the magazine so far and where it’s headed.

**SOREL REISMAN, ADVISORY BOARD CHAIR**

Twenty or so years ago it seemed like the Computer Society’s (CS’s) primary focus was on topics from the discipline of computer science—the theoretical underpinnings of what computing industry practitioners worked on in their professional lives. Folks who were industry practitioners, and even some like me who were faculty in IS or IT academic departments (which I define as the practitioner end of the computing spectrum) felt marginalized by this focus. But rather than abandon the CS and IEEE, since most of us were or had been electrical
engineers, we plotted to broaden the CS’s theoretical focus so that more practitioners like us could find value in CS membership.

If we look back at what was going on in the industry 20 years ago, you’d think our ambitions would have been easily realizable. From a practitioner and even a consumer standpoint, it was an exciting time. Some 1997 industry highlights: IEEE introduced 802.11 (the wireless local area network standard), Steve Jobs came back to Apple, Windows 98 was released, Yahoo Mail was introduced, world chess champion Garry Kasparov lost to IBM’s supercomputer Deep Blue, Microsoft acquired Hotmail, DVDs went on sale for the first time, lithium-ion batteries were first sold commercially, and both Facebook and Netflix came online. What a year! From our perspective, the most significant event of that year was the inception of *IT Professional*, with delivery scheduled for January 1999. This was a very short gestation period, atypical of IEEE publications.

Thinking back on the planning meetings that eventually led to IEEE’s approval of the magazine, the selection of the first editor in chief (Bill Chou), and the formation of the first editorial board, I remember our many animated yet friendly debates about the structure and direction of the magazine. Despite the often heated discussions, we all remained friends—even today. In fact, after Bill, with whom I had many of these debates, subsequent EICs appropriately came from the ranks of our first editorial boards.

Over the ensuing years, under the leadership of Arnold (Jay) Bragg and then Simon Liu, *IT Professional* itself really became professional, formalizing and improving its review process so that the magazine is now an indexed and referenceable publication competing qualitatively with other IEEE publications. More recently, under the outgoing EIC San Murugesan, the magazine is an annual sponsor of the IT in Practice (ITiP) track of the CS’s signature COMPSAC conference.

Our computer technology–driven society has changed a lot since 1997, and *IT Pro*, which was established to explain the implications of emerging computing technologies to professional practitioners and managers, has evolved too—both in terms of format and content. For example, in our first year of publication, we published seven articles about Y2K. Remember that? Our most recent issues are more diverse, still continuing to address contemporary IT issues (for example, mobile computing, security, cloud computing, and deep learning) of concern to today’s IT practitioners and managers. Two decades later, we are still here to help IT professionals learn and understand real, practical, and social issues related to emerging IT products and services. Stay with us. We have many more anniversaries to celebrate!

*Sorel Reisman* is a professor of information systems at California State University, Fullerton, and managing director of the MERLOT system at the California State University Office of the Chancellor. He is a past president of the CS (2011); a member of the IEEE Publications, Products, and Services Board; chair of the *IT Professional* advisory board; COMPSAC standing committee chair; and a Fulbright Specialist. Contact him at sreisman@computer.org.

**WUSHOW “BILL” CHOU, EIC 1999–2001**

*IT Pro* was initially intended to attract professionals who manage enterprise computer and communications systems, and to entice them to join the CS as members. (In spite of working closely with computers, few of these practitioners were IEEE or CS members.)

However, what was required to make *IT Pro* successful for these intended objectives did not fit well with the CS’s culture, philosophy, environment, and operational mode. There were many challenges and difficulties.
Identifying Topics and Content

The first challenge was to identify key and timely topics, to provide a suitable amount of technical information that was tutorial in nature and could enhance readers’ skills in managing and operating enterprise computer and communications systems, and to convince qualified authors to write such articles pro bono. For every article in the inaugural issue, a volunteer editor was responsible for choosing the title, writing a summary, and negotiating with a qualified individual to write the article without remuneration. This process was very time consuming for unpaid volunteers.

Finding Authors

The next challenge was identifying authors with intimate knowledge of the technology, management, or operation of enterprise computer or communications systems who were capable of writing for the intended audience, and were willing to write without compensation. Based on personal contacts, some editors were able to recruit a few authors. The difficulty was that such efforts were not sustainable in the longer term.

Finding Editors

Another challenge was finding individuals who had established professional status and were willing to serve as an editorial board member or advisory board member. I was fortunate enough to be able to appoint many highly qualified individuals as editors and advisory board members from my professional contacts. Among them were a member of the US National Academy of Engineering, a Knight of the Legion of Honor (France), a recipient of the Queen Elizabeth Prize of Engineering, two inductees of the Internet Hall of Fame, Taiwan’s minister of technology, a university president and a university dean, CEOs of IT corporations, CIOs of large enterprises, developers of IT systems, past presidents of the CS, a former EIC of an IT magazine, and experienced IT consultants. The difficulty was that most of these individuals were only willing to serve for a limited term, and people with such backgrounds were in short supply.

Operating within the CS’s Framework

Metrics that are typically used to judge the quality of an IEEE or CS magazine are the quantity of paper submissions, the percentage of submissions accepted for publication, and the number of citations of published papers. The challenge was to convince the CS Publications Board that IT Pro was a different kind of magazine than typical CS publications, and this board can be difficult to convince.

Attracting Intended Readers

We also needed to provide articles that would appeal to intended readers who were used to getting such articles for free or at a lower cost. As mentioned earlier, under the CS’s operating environment, it was unrealistic to provide such articles on a regular basis. It was even more difficult—actually impossible—not to charge a fee higher than what these potential readers were willing to pay.

Prognosis

Before the end of my term as EIC, it became clear that it was unrealistic to count on unpaid authors and editors to provide enough articles that appealed to the intended readers, to expect that the intended readers would be willing to pay a nontrivial amount for a subscription fee on
top of the membership fee, and to expect that the Pubs Board would allow the magazine to operate without a low acceptance rate and a high citation rate. It became clear that the focus of readership needed to be shifted toward our traditional membership (those who were already members), that the editors needed to be CS staff members, that the sources of the articles needed to depend on submissions, and that the prestige of the magazine needed to place emphasis on a low acceptance rate and a high citation rate. I am glad that Frank Ferrante, Jay Bragg, Simon Liu, and San Murugesan steered IT Pro into this new direction and evolved it into what it is today.

I feel very privileged to have worked with so many talented editors, advisors, and CS editorial staff. In particular, I was extremely impressed with Angela Burgess and Janet Wilson.

**Wushow “Bill” Chou** is a professor emeritus at North Carolina State University. He was the inaugural vice president of telecommunications at Network Analysis Corp., the founding director of Computer Studies and a professor at North Carolina State University, the inaugural chief information officer/deputy assistant secretary at the US Department of the Treasury, founding EIC of the *Journal of Telecommunications Networks*, and founding EIC of *IT Professional*. He was an IT consultant to more than 30 organizations, is an IEEE Fellow, and is a recipient of the US Treasury Meritorious Service Award. He has written more than 100 publications, given 100 public speeches, and visited 90 countries/territories. Chou received a PhD in electrical and computer engineering from the University of California, Berkeley. Contact him at chou@ncsu.edu.

**FRANK E. FERRANTE, EIC 2002–2005**

During my time as the second editor in chief of *IT Professional*, I knew the magazine was passing through a tough time. However, I was optimistic and felt *IT Pro* could maintain its readership and subscriptions if we could meet the needs of our readers.

**Learn from Our Examples**

My concerns at that time were the magazine’s themes and clear topics. As Bill Chou pointed out in the last month of his three-year term as our founding EIC, we needed to focus on “…delivering technical articles, rather than news and product information,” to give our readers an “…unbiased view of emerging technologies,” and to present easy-to-understand papers that could not be found in the CS’s detailed transactions. I knew emerging technology researchers out in the field had detailed material that we could present to our readers, but we needed to knock it down a level so that all readers could learn how the material applied to their needs.

**Themes and Topics**

Throughout my term as EIC, I promoted offering both themes and topics. With the help of our editorial staff and board members, we developed a clear theme for each issue but also looked for floating articles or topics that either extended information from past themes or introduced readers to new innovations (what we called “hot” topics of interest). This allowed us to offer new subjects so readers could see how these topics could be used in their own projects or to influence a future theme.
This approach might not have resulted in a massive increase in subscribers, but I did recall that it helped sustain our readership, leading to the continuation of this terrific magazine for our society.

**Top Articles**

Looking back on the articles published during my EIC term, I felt that the leading papers were ones that addressed current events such as 9/11, Hurricane Katrina, and jobs moving offshore. Wireless LAN standards, RFID techniques, grid computing systems, and population privacy were other hot topics. After my term ended, EIC Simon Liu requested that the board members vote on the top 10 papers from the first 10 years of the magazine, and he then identified the top 10 papers from the board’s votes.

He also drew data from the CSDL (the CS’s digital library) to get a list of the most downloaded articles.

Between the two lists, I found these articles from my EIC term to stand out (in no particular order):

- R. Weinstein, “RFID: A Technical Overview and its Application to the Enterprise,” vol. 7, no. 3, 2005, pp. 27–33 (on both lists);
- “Grid Computing 101: What’s All the Fuss About?,” vol. 6, no. 2, 2004, pp. 25–33;

After seeing all of these terrific papers and more, I knew that *IT Professional* would grow and be maintained for many more decades, especially with all of the outstanding editorial board members and CS staff. *IT Pro* will always be a success for the CS.

**Frank E. Ferrante** is an executive partner and adjunct professor in the Mason School of Business at the College of William & Mary, and served as an associate faculty member at Johns Hopkins University’s Carey Business School. He has extensive experience in telecommunications system design, implementation, and management. Ferrante was technical director of R&D for the Northrop-Page Communications Engineers and senior manager and electro-magnetic compatibility analyst for the Atlantic Research Corporation. He has served on the CS Board of Governors as treasurer and a member of the Publications Board, as well as EIC of *IT Professional*. Ferrante received an MSEE from Syracuse University and an MS in engineering and public policy from Carnegie Mellon University. He’s a Life Senior Member of IEEE. Contact him at frankfwmbg@gmail.com.
During my time as EIC, I gained a much deeper level of appreciation for the exceptional support of CS staff and the outstanding volunteer advisory members, editors, and reviewers. They inspired me to believe that IT Pro represents the IT community at its best, striving to significantly impact the world of IT professionals.

In 2011, IT Pro was enlisted in the Institute for Scientific Information (ISI) index—the number of submitted manuscripts doubled, the number of article downloads doubled, and IT Pro became more geographically and technically diverse than ever, all while maintaining high-quality content and decreasing our submission-to-publication time. As I think about this period, four accomplishments stand out:

1. Diversifying advisory and editorial boards. We rejuvenated the advisory and editorial boards with new members, and identified and recruited new board members to reflect balance and diversity in geographic areas, gender, discipline, and background (academic and industry). We realigned editorial subject areas to better reflect the needs of IT Pro readers and to recruit outstanding subject area editors from around the world. By 2013, we had editors from nine countries, six female editors, and an equal number of editors from academia and industry.

2. Establishing a strong leadership team. We recruited three new associate EICs as part of succession planning; delegated authority and responsibility to associate EICs to handle various activities; and conducted regular meetings or calls with the EIC and associate EICs to develop and plan activities, review progress, identify improvement opportunities, and refine goals and visions if needed. Two associate EICs, San Murugesan and Irena Bojanova, later became EICs of IT Pro.

3. Enlisting in the ISI index. In 2009, IT Pro was not listed by the ISI index—this limited the magazine’s attraction and lowered incentives for contributions from the academic community. To address this, we established an ISI membership task force led by editorial board member Morris Chang. The task force analyzed ISI index requirements and inclusion processes and provided the necessary documentation to ISI. In 2011, IT Pro was successfully included in the ISI index.

4. Increasing agility to respond to reader needs. A review of the 20 most downloaded papers in 2009 revealed that most of the articles were on emerging technologies and practical tutorials. To respond to reader needs, a few columns and departments were added to increase submissions, augment hot topic coverage, elevate visibility, and improve readership and circulation. For instance, a column for emerging technologies was added to provide articles on the newest topics, even if the articles were not extremely deep technically. Another example was establishing a learning department for tutorial/survey articles that led readers to interesting things going on the field. By 2013, the number of article downloads had doubled.

I sincerely believe that the IT professional community will benefit from IT Pro’s timely response to emerging topics and excellent scholarship for many years to come. I am grateful to my predecessors Bill Chou, Frank Ferrante, and Jay Bragg for establishing a solid and strong foundation as well as my successors San Murugesan and Irena Bojanova for continuously exploring new frontiers and charting new directions.

Simon Y. Liu is the associate administrator of the Agricultural Research Service, a research agency with more than 2,000 research scientists. He previously served as director of the National Agricultural Library, the world’s largest and most accessible research library specializing in agriculture. Before that, Liu served as associate director and chief information officer of the National Library of Medicine. He also served as the director of information management and security at the US Department of Justice and the chief IT architect of the US Treasury Department. Liu attended 10 universities in three countries and holds two PhDs and three master’s degrees in computer science, education, government management, business administration, and mathematics. Contact him at simonyliu@yahoo.com.
IT PROFESSIONAL’S 20TH ANNIVERSARY

SAN MURUGESAN, EIC 2014–2017

2018 marks IT Professional’s 20th anniversary, and this issue is its 121st (you can see past issues, articles, and article summaries conveniently aggregated at www.obren.nl/ieeeit). This is a significant milestone in the magazine’s journey and evolution, and it is a good time for recollection, reflection, and celebration.

IT Pro Is Unique

Since the first issue, much has changed in the technology and applications landscape, and also in our life, work, and profession as a result of advances in and adoption of IT. IT Pro has traced these changes well and remains a source of trusted information for IT professionals to update their skills and knowledge. The magazine continues to be informative, timely, stimulating, and helpful to a broad spectrum of readers interested in IT and its applications.

IT Pro is a unique, scholarly IT magazine specifically targeted to the IT community at large—including young and inexperienced students, academics, IT researchers, applications developers, and IT managers and executives. It’s unique primarily because it covers a wide spectrum of topics of interest and relevance—not only key emerging technologies but also their novel applications, key development challenges and issues, socioeconomic impacts, and more. The magazine also examines technology trends and highlights contributions of pioneers who formed the foundation for additional development and offered inspiration to advance our field.

Thanks to all those who conceived this unique magazine, shaped it, supported and promoted it, and steered it to its current stage.

Looking Back

I consider myself fortunate and privileged to be closely associated with this great magazine that IT professionals can rely on. My four-year tenure as EIC, which recently ended, has been a rewarding experience for me both professionally and personally. It gave me an opportunity to serve a community of IT and computer professionals, academics, researchers, and students (the future professionals) by offering them information of interest and relevance.

Over the past four years, IT Pro has published several special issues on emerging topics, including mobile commerce, wearable computing, data analytics, digital innovation and transformation, cybersecurity and privacy, the Internet of Things, cognitive computing, and graph databases. We also launched new departments and columns, such as Life in the C-Suite, Mastermind, Data Analytics, and IT and Future Employment.

IT Pro has performed consistently, making significant progress in terms of coverage, readership, citations, article downloads, and impact factor. The magazine is now indexed by major indexing services including ISI. Even 10 years after publication, many visionary and tutorial articles are still being widely read and cited. Readers, stakeholders, and industry professionals continue to appreciate and look forward to every new issue.

In 2014, Sorel Reisman, Carl Chang, and I realized that many practitioners considered COMPSAC and other CS conferences to be academic in nature and hence had shown little interest in participating. With the objective of bringing together researchers and practitioners to interact, share their knowledge, and discuss areas that need further research and development, IT Pro (in collaboration with COMPSAC) launched COMPSAC’s IT in Practice (ITiP) Symposium in 2015. The symposium is now an annual event. This year, COMPSAC and the ITiP Symposium will be held in Tokyo from 23–27 July (for details, visit www.compsac.org).
We’ve also affiliated *IT Pro* with other international and regional conferences and symposia relevant to IT practitioners, professionals, and managers, including the Software Technology Conference (STC). Furthermore, to focus attention on the importance of information systems governance, *IT Pro* organized an Information Systems Governance conference at NIST in 2015.

Despite a few challenges, difficulties, and barriers that we addressed satisfactorily, my journey as EIC has been relatively smooth and enjoyable. This is primarily because I had the support of several great people—editorial and advisory board members, editorial staff, guest editors, reviewers, authors, subscribers, readers, well-wishers, and the CS Board of Governors and executives who helped raise the magazine to greater heights. I acknowledge their contributions and encouragement.

**Looking Ahead**

Significant shifts are ahead, driving us to a new digital landscape and a new information age. We can’t yet fully visualize this future or anticipate its potential impact, so IT professionals must protect themselves from “technological threats”—that is, threats and issues posed to them by continuing rapid and widespread advances in IT. These include technological obsolescence, technological irrelevance, and a lack of skills and knowledge that are relevant now and will continue to be in the future.

To thrive in this dynamic environment, as I highlighted in my final EIC editorial (“Stay Professionally Fit, Always,” vol. 19, no. 6, 2017, pp. 4–7; bit.ly/2ArV6DG), IT professionals, academics, and researchers must continue to be professionally fit. They should be T-shaped professionals (or funnel-shaped professionals)—that is, professionals who have broad knowledge in several related areas, and in-depth knowledge and skills in one or more specific areas of interest. Then they will be able to really make a difference, helping change the world for the better, embracing advances in technology, and addressing challenges and issues that will arise in the new, interconnected, smarter information age. *IT Pro* will continue to assist in this endeavor by providing ongoing updates on IT research and practice and general insights of interest and relevance to IT professionals and executives, researchers, and students. The magazine, however, will need your support and contributions to further enrich its value to readers in the exciting and challenging times ahead—the new information age. Continue to enjoy reading it and benefiting from it.

The magazine has a long, challenging road ahead. To reach further heights and to remain a preeminent magazine for IT professionals for years to come, it has to remain informative, timely, relevant, and valuable to its readers. It has to effectively engage with its readers, whose demographics and preferences will continue to change, and has to continue to be innovative in meeting readers’ varied and changing needs and expectations. My best wishes to *IT Pro* and its editorial board as they further the magazine’s journey.

**San Murugesan** is the director of BRITE Professional Services and an adjunct professor at the University of Western Sydney, Australia. His research interests include cloud computing, green IT, the Internet of Things, and IT applications. Murugesan served in several senior positions in academia and industry in Australia and India. He was EIC of *IT Professional* from 2014–2017 and is co-editor of *Encyclopedia of Cloud Computing* (Wiley, 2014). Murugesan is a Fellow of the Australian Computer Society and the Institution of Electronics and Telecommunication Engineers (IETE). Contact him at san@computer.org.
Effectiveness of Test-Driven Development and Continuous Integration

A Case Study

In a case study where a Dutch small-to-medium enterprise (SME) implemented test-driven development and continuous integration, researchers observed that the SME discovered a higher number of defects compared to a baseline case study, and that there was an increase in the focus on quality and test applications.

Although many older development methodologies include a separate lengthy testing phase at the end of development, test-driven development (TDD) comes with a new paradigm: always test before coding. Continuous integration (CI), on the other hand, involves running integration builds and automated tests on the code committed by the developers. CI combines development and testing by enabling unit and functional tests while profiling the application code. CI was designed to improve the number of testing cycles and the resulting application quality while decreasing the amount of time it takes to find problems and reducing the cost to fix them.

Both TDD and CI, however, cost the development company in terms of time and money. Hence, well-grounded empirical evidence is required to determine the applicability of these practices in actual industrial settings. In response to this, some empirical studies were conducted on the effectiveness of TDD implementations in academic as well as industrial settings (see the sidebar for a literature review). The outcomes of these empirical studies seem to indicate an increase in code quality along with an inconclusive effect on developer productivity. Yahya Rafique and Vojislav B. Misic mention in their extensive literature review that some experiments with a more detailed design obtained better results on TDD implementation. This was verified in a recent project where TDD was implemented successfully. However, this claim has not been adequately or quantitatively tested.
in an industrial setting, nor has it been discussed widely. Furthermore, most papers do not provide a detailed description of the TDD implementation.

In this article, we describe a case study of both TDD and CI in a Dutch small-to-medium enterprise (SME). The contributions of this article are multi-fold. While previous research dealt with few metrics for accessing the cost and quality, we use multiple metrics. Along with an evaluation of the technical quality (which was also done in previous studies), we propose a quantitative evaluation of the impact of TDD and CI implementation and provide a detailed account of the case setting. To aid future research, we list a set of adherence metrics to measure the extent to which TDD/CI principles have been applied. Our article contributes to the growing body of literature on TDD and CI implementation, as well as its evaluation in an industrial setting. Similar to other studies, we cannot isolate the effect of the sole application of TDD from the effect of CI.

THE CASE CONTEXT

The data for this research originates from a two-case comparison of software implementation at a Dutch SME. The two cases were consecutive software development projects at the software company. The first case followed a development process without any methodology alterations nor any involvement by the researchers, while the second project followed a process that was refined by TDD and CI principles. The second author of this article was present as both an information analyst and chief methodology implementer of the second project. We use the first project as a reference case to compare the potential differences in the outcome of the two projects.

Both software development projects dealt with healthcare claim handling. Healthcare claims are mostly digital transactions (involving the transfer of messages, digital invoices, or money) between healthcare providers and insurers. As soon as an individual receives some form of healthcare from a provider (for instance, surgery), the provider usually makes several transactions that need to be appraised by the insurer. On average, each individual in the Netherlands is involved in about 10 transactions per year. This implies that managing these digital transactions is a high-volume industry. As these transactions are highly standardized by the Dutch government, they are well suited for automation. The Dutch SME in this case study makes software applications that provide such automation. The software project described in this article is based on one such software application. For the sake of clarity, we denote the software development project in which TDD and CI testing was implemented as case study 2 (CS2), and the previous version where it was not implemented as case study 1 (CS1).

CS1 served mainly as an entry portal for incoming healthcare transactions from healthcare providers at an insurance intermediary. Its main goal was to structure, check, and help in the appraisal of incoming transactions. After this process, the transaction was entered into the financial database.

CS2’s software application was installed at an invoicing intermediary, which takes care of everything around healthcare transactions for a set of healthcare providers. This is done so that the healthcare providers can concentrate on providing proper care instead of getting bogged down with the many financial transactions that are required. However, in terms of functionality, CS1 and CS2 were very similar.

The applications of the two cases were installed at different points of the healthcare chain, and thus satisfied different stakeholders. They were similar in terms of functionality, as they interacted (indirectly) within the same industry while applying common transaction standards. The CS2 implementation was, in particular, based on the CS1 architecture and most of the CS2 modules were refactored CS1 modules. As 92.5 percent of the modules overlapped between the two architectures, we can safely assume that the number and complexity of the features were similar. Note that this overlap is more than the 75 percent overlap reported by Roberto Latorre.

Table 1 lists the similar project characteristics found in the two cases in terms of context factors and software product measure factors. We notice that both kinds of factors, especially the contextual factors, are comparable and rather similar between the two case studies.
## Table 1. Description of the context and product measures of case studies 1 and 2.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Case study 1 (CS1)</th>
<th>Case study 2 (CS2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>Transactional system for technical verification and appraisal, as well as administrative handling and payment of healthcare declarations</td>
<td>Transactional system for technical verification, routing, and appraisal of healthcare declarations</td>
</tr>
<tr>
<td>Customer</td>
<td>Authorized insurance intermediary</td>
<td>Invoicing intermediary</td>
</tr>
<tr>
<td>Duration (time)</td>
<td>8 months</td>
<td>10 months</td>
</tr>
<tr>
<td>Average monthly effort (man days)</td>
<td>78</td>
<td>73</td>
</tr>
<tr>
<td>Total effort (man days)</td>
<td>732</td>
<td>731</td>
</tr>
<tr>
<td>Team size</td>
<td>&lt;10</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Team people overlap</td>
<td>N/A</td>
<td>&gt;4</td>
</tr>
<tr>
<td>Experience level (&lt;5 years, 6-10 years, &gt;10 years)</td>
<td>Most members &lt; 5 years’ experience</td>
<td>Most members &lt; 5 years’ experience</td>
</tr>
<tr>
<td>Project manager’s expertise</td>
<td>&gt; 5 years’ experience</td>
<td>&gt; 5 years’ experience</td>
</tr>
<tr>
<td>Applied technology</td>
<td>C# ASP.NET MSSQL NHibernate</td>
<td>C# ASP.NET MSSQL NHibernate</td>
</tr>
<tr>
<td><strong>Product measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source LoC</td>
<td>28,049</td>
<td>21,340</td>
</tr>
<tr>
<td>Maintainability index average</td>
<td>79.20</td>
<td>85.32</td>
</tr>
<tr>
<td>Cyclomatic complexity average</td>
<td>280.53</td>
<td>297.71</td>
</tr>
<tr>
<td>Depth of inheritance average</td>
<td>02.62</td>
<td>04.70</td>
</tr>
<tr>
<td>Class coupling average</td>
<td>57.59</td>
<td>77.65</td>
</tr>
</tbody>
</table>

### MEASUREMENT METRICS

To measure the effectiveness of TDD and CI, we considered three perspectives: defect reduction, defect lead and throughput, and development productivity.

Defect reduction is whether TDD and CI helped in reducing the number of defects; for example, if CS2 has fewer defects compared to CS1. This includes the pre-release defect level (the number of defects detected in the pre-release software per KLOC) and post-release defect level (the number of defects occurring in the post-release software per KLOC).9,10
Defect lead and throughput is whether TDD and CI helped in reducing the time to find and fix the defects. This includes defect resolution duration—the duration between the discovery and closure of defects, given by

\[
\frac{\sum d_{\text{closing}} - d_{\text{reporting}}}{|D|}
\]

where |D| is the number of defects. This also includes defect pre/post-solvability—the proportion of defects uncovered prior to and after the release given by \[\frac{|D^{\text{pre}}|}{|D|}\] per pre-solvability release and \[\frac{|D^{\text{post}}|}{|D|}\] per post-solvability release.

Development productivity is the development productivity and rework rate in development per release, given by \[\frac{\text{KLOC}}{\text{Effort}}\], where the KLOC is the amount of KLOC added per release.

RESULTS

We first analyzed our data using descriptive statistics and then considered inferential statistics. Thirty-nine versions of the CS1 software and 24 versions of the CS2 software were released in the period from the start of the two projects up to when the data for this study was collected. Both projects used the Mantis bug tracker, and developers and customers submitted bug reports. We analyzed the CS1 and CS2 Mantis bug trackers and removed duplicates and issues that were not really bugs. CS1 had 414 viable bug reports for analysis, while CS2 had 474 viable bug reports. The bugs were prioritized in both CS1 and CS2 in terms of defect severity. But this was not done explicitly in both projects. The developers and project leaders knew which bugs to resolve first, based on which part of the code they occurred in, and this knowledge was tacit (not made explicit in the bug tracker). In the case of CS1, the prioritization and resolution were done manually. In both projects, the bugs that caused build failures were fixed first and the rest were fixed in the order of prioritization. In the case of CS2, CI helped in this regard, as the CI-related best practice of “executing all tests with every build and making a single failed test fail the build” was followed.\(^1\)

When we asked the project managers of both projects to compare the list of bugs (in terms of the number of high-priority bugs and the severity of the defect), they agreed that the bugs from both the projects were comparable.

Descriptive Statistics

The box plot of the distribution of pre- and post-defect resolution data is shown in Figures 1a and 1b; the numbering of the figures that follow is based on this categorization. In Figure 1a we notice that the defect level of CS2 is consistently higher (before the release) than the defect level of CS1. This could be the result of implementing the test-first approach in CS2, and not necessarily because the CS1 software had fewer bugs.

This seems to be verified by Figure 1b, where we notice that CS1 has more bugs than CS2 after the release. Figure 1b also demonstrates that the medians of the post-release number of bugs per KLOC are nearly the same for CS1 and CS2. The smaller inter-quartile range of CS2 shows that CS2 had a more consistent number of bugs per release, and fewer releases with a large number of bugs (as was the case with CS1).

In Figures 1c and 1d we can see the results of the defect lead and throughput category metrics. Figure 1c indicates that the defect resolution duration (the time in days between discovery and resolution) is mostly higher for CS1 than CS2 (although the medians are nearly the same). Furthermore, CS1 has a much longer tail, indicating that a few bugs required more fixing time compared to CS2.
In Figure 1d we see the defect pre-/post-solvability (the proportion of defects uncovered before release). CS2 again outperforms CS1 because a consistently higher percentage of bugs was found before a major release. The developers in CS2 also accurately flagged a bug as solved as soon as they had fixed a defect. This was done to prevent colleagues from fixing the same defect again, and to show the progress to the customer. Hence, this data can be considered reliable.

In Figure 1e we see the comparison of development productivity between the two projects. The inter-quartile range of CS1 is almost double that of CS2, while the medians are nearly the same. This implies that while the development productivity was nearly the same, there were instances when the number of added KLOCs was very large for the same number of man days. When we approached the team members for an explanation, they thought it could be due to the introduction of large pre-coded components (containing several hundred code lines apiece).

Figure 1. The descriptive statistics for the metrics described in this article. (a) Pre-release defect level. (b) Post-release defect level. (c) Defect resolution duration in days. (d) Defect pre-/post-solvability. (e) Development productivity.
Inferential Statistics

We first tried to ascertain the difference in means between CS1 and CS2, with respect to the metrics listed in this article. The common statistical method used to test the difference in the distributions of two samples is the Student’s t-test. However, the prerequisites for using the t-test are that the two distributions must have the same variance and both populations should follow the normal distribution. We could not use the Student’s t-test as Levene’s test showed that some of the metrics from CS1 and CS2 did not have equal variances, while the Shapiro–Wilk test for normality demonstrated that normality was violated by almost all the metrics. The reason behind this was the variation in the denominator in all the metrics, namely KLOC. When comparing the KLOC between CS1 and CS2, Levene’s test showed a significance of 0.38 (indicating that the population variances are equal) and Shapiro–Wilk showed a significance of 0.00 for both projects (indicating that they are not normally distributed). As the distribution was not normal for both samples, we considered the Mann–Whitney–Wilcoxon (MWW) test, which is non-parametric.

Table 2. Mann–Whitney–Wilcoxon test results for the metrics in this article.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mann–Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
<th>Asymptotic significance (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-release defect level</td>
<td>258</td>
<td>1038</td>
<td>-0.62</td>
<td>0.009</td>
</tr>
<tr>
<td>Post-release defect level</td>
<td>388</td>
<td>619</td>
<td>-0.33</td>
<td>0.739</td>
</tr>
<tr>
<td>Defect resolution duration</td>
<td>55555.5</td>
<td>154790.5</td>
<td>-2.57</td>
<td>0.010</td>
</tr>
<tr>
<td>Defect pre-/post-solvability</td>
<td>312</td>
<td>942</td>
<td>-1.84</td>
<td>0.067</td>
</tr>
<tr>
<td>Development productivity</td>
<td>413</td>
<td>713</td>
<td>-0.78</td>
<td>0.436</td>
</tr>
</tbody>
</table>

Table 2 groups the results based on the relevant metric category. The p-value of the asymptotic significance of the pre-release defect level and defect resolution duration is less than 0.05. Hence, these two metrics demonstrate that CS2 outperforms CS1. Regarding the defect pre-/post-solvability metric, the p-value is just over 0.05, whereas for the post-release defect level and development productivity metrics, we find no significant difference between CS1 and CS2. Note that the MWW test can only determine whether the two distributions are indeed significantly different, and not which distribution has a higher median (or other measures of dispersion). Therefore, we combined the data from Table 2 with Figures 1a and 1c. We see that the number of pre-release defects found in CS2 is significantly greater than in CS1 (Figure 1a). We also see that the defect resolution duration of CS1 is significantly larger than that of CS2. Though the U value of the defect resolution duration appears to be rather large, it is more or less what we can expect given the large number of samples: 414 for CS1 and 474 for CS2 (an expected estimator of U is multiplying half of the sample sizes of the distributions). In the case of defect pre-/post-solvability, we see that the MWW test p-value just exceeds 0.05, though Figure 1d shows that the number of defects uncovered before release is much larger in CS2 than CS1. Hence, CS2 outperforms CS1 in this metric. With respect to the post-release defect level, the MWW test shows that the distribution of the number of bugs per KLOC of the two projects is not significantly different. We see from Figure 1b that the medians of the two distributions are indeed quite close, though the CS1 upper quartile is much larger than that of CS2, so there were
more defects occurring in CS1 post-release. This is also the case with the development productivity metric; in Figure 1e, we see that the medians are similar but the lower and upper quartiles are largely different, implying that the development effort put into a few releases in CS1 was much larger than in CS2.

CONCLUSION

The team members at the Dutch SME did perceive an increase in the focus on quality and in the application of tests, while considering customer acceptance. The company now has an infrastructure in place to further evaluate other software process improvement (SPI) initiatives.

Though the inferential statistics are not conclusively in favor of CS2, the descriptive statistics point to an overall improvement in not only finding more defects (defect reduction), but also in shortening the time required to fix the defects (defect lead and throughput). One of the limitations of this research could be the use of KLOC in most of the metrics (in the denominator), as well as the use of KLOC for measuring development productivity (see Figure 1e). Emad Shihab and his colleagues mention that metrics like cyclomatic complexity—and a combination of software metrics—outperform LOC in estimating effort, whereas using solely LOC underestimates effort. However, if the same metric is used to compare different projects, the underestimation of effort could balance out—as it is underestimated for both cases, and we are only interested in the relative comparison of development effort and not the exact development effort of each project. Yet, applying KLOC limited our use of different statistical techniques, as the variance it introduced made the distributions non-normal.

The contributions of this research are multiple. We have added to the small but hopefully growing body of empirical literature on agile implementation in an industrial setting. We have endeavored to provide a rich description of the project setting (contextual factors and product measures in Table 1) that we think is useful to recognize potential validation errors. Finally, we have added to the existing set of metrics, and we think our new metrics give a richer and more detailed description of the effects of agile implementations in general.

SIDEBAR: LITERATURE OVERVIEW

We performed a meta-analysis of the literature reviews on TDD published in the past five years. In the table below, internal quality relates to the quality of the software design (measured by OO metrics, code density, cyclomatic complexity, etc.); external quality refers to the number of pre/post-release defects per given code size; and productivity refers to developer productivity (measured using development time, total LOC divided by total effort, hours per feature/development effort per LOC, etc.). In Table 3, N/A refers to the fact that the particular construct was not analyzed.

<table>
<thead>
<tr>
<th></th>
<th>C. Desai et al.</th>
<th>B. Turhan et al.</th>
<th>A. Causevic et al.</th>
<th>S. Kollanus</th>
<th>Y. Rafique and V.B. Misic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal quality</td>
<td>Little evidence</td>
<td>Little evidence</td>
<td>Moderate evidence</td>
<td>Inconclusive</td>
<td>N/A</td>
</tr>
</tbody>
</table>

This research adds to the existing set of metrics, and we think our new metrics give a richer and more detailed description of the effects of agile implementations in general.
Table 3 broadly suggests there is little to moderate evidence that implementation of TDD in an academic or industrial setting is accompanied by an increase in code quality. However, the evidence for an improvement in productivity is largely inconclusive and the literature indicates, to some extent, that TDD implementation is accompanied by decreased productivity.

REFERENCES


ABOUT THE AUTHORS

Chintan Amrit is an assistant professor in the Department of Industrial Engineering and Business Information Systems (IEBIS) at the University of Twente. His research interests include software development, business intelligence using machine learning, data analysis methods, supply-chain logistics, and mining software repositories. Amrit received a PhD in information systems from the University of Twente. He is an associate editor of IEEE Access, a coordinating editor of Information Systems Frontiers, and a regular track chair at the European Conference on Information Systems (ECIS). Contact him at c.amrit@utwente.nl.

Yoni Meijberg is a managerial change agent at Topicus. His research interests include software project management and business research methods. Meijberg received a master’s degree in organizational change from RSM Erasmus University and another in industrial engineering from the University of Twente. Contact him at yoni.meijberg@topicus.nl.
Automatic Annotation of Text with Pictures

J. Kent Poots  
York University

Ebrahim Bagheri  
Ryerson University

The vast array of information available on the Internet makes it challenging to quickly determine the importance and relevance of content. Text picturing is a cognitive aid that can help with text understanding, as it helps users decide if the text deserves a closer look by showing relevant pictures along with the text.

Readers of business reports, newspapers, and social media face the challenge of interpreting large volumes of text in a short amount of time. Further, information loss can be experienced by people who have trouble with text-only documents, those using hand-held devices, and those who might not be able take in written material at a standard pace.

The trend for information sharing on the Internet is to offer details at just the right length. The emergence of the phrase “too long; didn’t read” (TL;DR) shows the tendency of readers to look for short yet informative text where they can quickly determine its relevance to their interests.

To tackle the challenge of interpreting text and to help breach the “information-assimilation wall” of cognitively deficient readers, a novel application has emerged that automatically adds pictures to text. This application is called automatic text annotation with pictures, concept picturing, or text picturing. The objective is to reduce a reader’s cognitive information processing load by showing relevant pictures along with input text. A reader can quickly and easily decide if the text information deserves a closer look.

TEXT PICTURING

For writers, bloggers, and editors, linking text with media content is crucial. Visual content helps readers follow the crux of a discussion or identify the core theme of an article. In efforts to organize and display helpful images, content managers, news writers, and user-interface designers perform page layout, assigning images to concepts. To improve image-retrieval accuracy, a number of issues must be addressed. For instance, journalists trying to illustrate a point are required to match a story’s natural language to the available images. Accordingly, matching a news story to images typically relies on matching text to a tagged image library.

Despite online photo sites that imply they can easily (and automatically) label images, one of the important text-picturing research issues is the accurate and automated tagging of images. This
The tagging process is modeled as multiclass image classification. Research has produced commercial applications, including MetaMind and Clarifai, which continue to improve their results. However, automated image tagging still faces constraints, including the limited number of identifiable objects, sometimes poor image quality, limited image library breadth, and the need for reliable training sets.

Before reviewing the technical underpinnings of text picturing, we first consider some important application areas.

APPLICATION AREAS

Researchers have long placed considerable value on including pictures with text. In 1911, Evening Journal editor Arthur Brisbane crafted a now-popular maxim when he told members of the Syracuse Advertising Men’s Club to “use a picture. It’s worth a thousand words.” In fact, educators argue that pictures help us build mental models that aid in text comprehension and enhance our understanding.3

Pictures as a Cognitive Aid

Cognitive processes are related to acquiring knowledge, paying attention, remembering, judging, and reasoning.4 In cognitive science, “cognition” usually refers to an information-processing view of an individual’s psychological functions.4 A cognitive aid improves information flow and retention by ensuring compatibility with users’ competencies. A ready example is the pictures used in classrooms to help students grasp new concepts.2

Cognitive science suggests that the mind processes each portion of language as a mix of memory chunks and rule-governed assemblies.4 Cognitive science advocates a similar mechanism for the mind’s analysis of images. To automatically annotate text with pictures, we need to mimic, at least at a high level, our human ability to extract and match concepts from multiple types of media.

Annotating text with pictures can be described as multimedia adoption or generation for the purpose of improving information transfer through better alignment with cognitive processes.5

Assistive Communication Devices

Pictures can be incorporated into assistive communication devices.6 Considering substituting pictures for text as a form of translation, a real-world application can be found in augmentative and alternative communication (AAC) tools. AAC tools are used by the cognitively impaired to decipher written expressions.6 These communication devices compensate for gaps in communication skills. Picture-based communication tools can provide a dynamic representation of concepts to reinforce a word-picture link.

AAC devices using text to pictograph (specialized symbol) translation can assist individuals with complex communication challenges. Current technologies that focus on pictographic translation systems for assistive communication include Bliss symbols, Minspeak, and Sclera.6 To decrease complexity, most assistive communication pictographic languages are cleansed of natural language features such as grammar quirks and nesting of descriptive phrases. Meaning extraction is usually limited to a shallow analysis requiring only dictionary lookup, and then direct mapping to pictographs.6

AAC-related tool surveys show that some concept symbols might be overly simplified for general purposes, but the assistive symbol sets can still be useful in conveying basic messages for
many applications. Interestingly, educators affirm that visuals are essential for effective learning in general.

**Adding Pictures to Text Advertising**

Advertising is more effective when ads contain a combination of words and pictures. We understand visual information more readily than text—sometimes at least twice as fast. A text picturing system could be used to deliver near-real-time content to a handheld device; for example, in response to a sentiment expressed in a social media post. That sentiment could be brand-specific where the pictures could show an alternate brand. Images could, in turn, be linked to additional webpages promoting that alternate brand, with the hope of increased sales or customer retention.

**SYSTEM FEATURES**

Text picturing can be considered an analog of a human information retrieval task, where a machine reads text input and transcribes that input to a set of visual outputs using information extraction and matching techniques.

The components of a generic text picturing system can typically include the following.

1. **Language input.** The input text can be obtained from the output of a speech-to-text device or directly taken from a webpage. Language input is provided as one or more groups of linguistic units (one or more words or one or more sentences).
2. **Knowledge resources.** These are used to add related context to the input. The inclusion of additional context to the input text (called *semantic expansion*) often proves to be very effective in retrieving higher-relevance images.
3. **Knowledge extraction.** To effectively retrieve the most relevant image for text, an essential step is to identify the main themes or terms in the text.
4. **Picture matching.** Once the main themes are identified, a similarity metric matches the input text to the descriptors (tags) of candidate images.
5. **Output rendering.** Selected images are rendered in one or more visual scenes.

Systems that perform text picturing use some or all of these five components. Recent text picturing systems have been built and can be categorized according to this component list.

**General Approach and Primary Technical Challenges**

Text picturing systems transform input text into pictures, usually implementing a processing pipeline where text is first processed into basic linguistic units (such as sentences), knowledge is extracted, and pictures are found.

Knowledge extraction presents the primary set of technical challenges. For this purpose, knowledge can be extracted using distributional techniques (such as named entity extraction), using linguistic techniques (like ontology-assisted knowledge extraction) or a hybrid approach. Knowledge extraction using distributional techniques requires sizable collections of labeled examples and statistical analysis tools, along with hours of hand-labeling effort.

With the evolution of computing capability, the required statistical analysis tools are now within reach of researchers. This was not the case even 10 years ago. Linguistic tools have also evolved to the point where off-the-shelf components (such as those for parsing) can be used in the text analysis pipeline.

**Language Input**

Many details of the text picturing process are determined by the intended knowledge resolution. Knowledge resolution is a measure of the amount of knowledge transcribed from the linguistic unit of input text to the output scene. Three levels of knowledge resolution are often considered: paragraph, sentence, and word. In paragraph-level resolution, one or more output images are
generated per paragraph of input text. Sentence-level resolution generates one or more output images per sentence of input text, and word-level resolution functions similarly per word of the input text.

Figure 1 shows word-level knowledge resolution\textsuperscript{11} (with words as a basic linguistic unit) and sentence-level knowledge resolution,\textsuperscript{12} which focuses on sentences as a basic linguistic unit.

![Figure 1. Comparing word-level picturing for the noun “Victoria” (a) versus sentence-level picturing for “People from Canada celebrate the birthday of Queen Victoria” (b).](image)

For word-level resolution, we see multiple images for a single word. For sentence-level resolution, we see multiple images for one sentence. Word-level picturing could help a teacher when presenting a word with multiple meanings (see Figure 1a). Sentence-level resolution gives context for a collection of words, allowing identification of relationships. Figure 1b shows multiple concepts related to Queen Victoria’s birthday.

**Knowledge Resources**

Knowledge resources are external sources that help clarify meaning. Visual resources (collections) provide output images. Examples of text resources include Wikipedia and WordNet. Examples of visual resources include Flickr, Google Image, and Wikimedia images.

**Text Resources**

Wikipedia, which provides considerable detail about specific topics, can be used as a reference for basic knowledge about input text.\textsuperscript{11} Links between articles can provide a hierarchy (taxonomy) of topics that can be helpful when searching for images in domains lacking detail.

WordNet is a specialized database where words are grouped according to their semantic relationships.\textsuperscript{13} Influenced by theories of human lexicon memory, WordNet offers alternatives by assembling nouns, verbs, adverbs, and adjectives into sets of synonyms as part of a semantic lexical network. WordNet entries can be used to expand knowledge of related and important terms and concepts, providing added context for identifying images.

In the literature, Wikipedia and WordNet have been the predominant textual knowledge resources that help identify the main themes and terms within text.

**Image Resources**

Examples of visual resources include Wikimedia images, Flickr images, images from search engines (Google Images), and domain-specific collections. Wikimedia images are stored in topic-
specific “infoboxes,” enabling automated retrieval by software applications. Flickr has been another widely used source, where photos are manually labeled (“tagged”) for later retrieval. A number of applications use Flickr tags for the purpose of matching images to concepts. The web offers collections of domain-specific images that can be used for text annotation. For instance, medicine is represented by databases containing medical images, such as X-rays (see the Centers for Disease Control Image Library; phil.cdc.gov/phil/home.asp), while business sector topics are represented by more generic commercial databases such as Shutterstock. A useful review of image sources is provided in “A Survey of Current Datasets for Vision and Language Research,” which describes captioned image resources such as the Pascal Dataset, Microsoft Coco, and Flickr datasets.

Knowledge Extraction

Knowledge extraction refers to the steps in extracting structured information from unstructured text. There are several techniques for this purpose. An intuitive approach is grammatical analysis. For example, in the sentence “John ate the red apple,” grammar rules say that the subject is John, the action (or predicate) is ate, and the object (or target of the action) is the red apple. The knowledge we can extract is who (John) did what (ate) to whom or what (the red apple). Then we can use this to find appropriate images.

In addition to grammatical analysis, more exotic techniques can be employed—for example, graphical analysis, statistical techniques, and keyphrase (or keyword) extraction. Graphical representations are often used in text analysis. Keyphrase extraction techniques are typically used in online content indexing and searches.

Finally, several existing tools are intended specifically to extract textual knowledge. These tools are targeted mostly at researchers, have a learning curve, can require considerable machine resources, and can be slow to extract knowledge.

Picture Matching to Input Text

Picture matching requires the semantic match of input text concepts and output image concepts. The knowledge resolution parameter determines the number of pictures for each basic linguistic unit (word/sentence/paragraph). It is assumed that pictures are also represented by text (for example, a caption or a short paragraph) so that the matching exercise means matching input text to image description text. Image descriptions can come from four sources: URLs, surrounding text, user tags, and analysis of visual content. An example of matching for “John ate a red apple” would be a picture of a man (picture tags “man” and “human”), a picture of someone eating food (picture tags “person eating,” “person,” “hungry”), and a picture of an apple.

Matching can be automated by searching for images that have the largest number of terms in common with the input text. To improve results, external knowledge sources can provide related terms (semantic expansion). These additional terms usually improve the chance of a good match. There are additional matching measures based on more formal natural language processing techniques. For instance, some approaches match terms by comparing the distribution of similar terms, and others use matching rules.

Output Rendering

The final category used for comparing text picturing systems relates to output scene composition. One or more output scenes are created for each unit of linguistic input where the number of output scenes is determined by the number of input text concepts. For each scene, pictures need
to be laid out on a canvas to represent the input text; this is typically a left-to-right ordering, corresponding to the same ordering as the input text.

Rather than assemble a scene from several images, it is possible to find images that illustrate multiple concepts, perhaps “man eating (an) apple.” These complex images could be used as the entire output scene.

PICTURING TECHNOLOGY: FEATURE COMPARISON

Here, we provide a detailed comparison of eight well-known picturing systems.

1. WordsEye, a text-to-3D-scene rendering system, parses text into a semantic representation and creates a 3D rendering (a scene).
2. Story Picturing Engine extracts keywords and matches words to a labeled image database as part of the process of story illustration.
3. ROC-MMS for Complex Sentences produces a multimodal (text and pictures) summary of sentences.
4. Simple Sentence Picturing is an effort to investigate the idea that pictures can convey simple sentences across language barriers.
5. Enriching Textbooks matches sections of a textbook to Internet images.
7. Text to Pictographs for AAC is related to so-called translation efforts, where pictures are displayed to assist those with cognitive impairment.
8. Picturing Microblogs is an effort to summarize microblog (such as Twitter) content using representative pictures.

Table 1 summarizes the features of each system. The rows represent various picturing techniques and the columns denote the dimension of our framework.

**Table 1. Comparison of text picturing technology.**

<table>
<thead>
<tr>
<th>Picturing tool</th>
<th>Language input</th>
<th>Knowledge resources</th>
<th>Knowledge extraction</th>
<th>Picture matching</th>
<th>Output rendering</th>
<th>Notable features</th>
<th>Use cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordsEye Scene</td>
<td>Scene description in a few sentences</td>
<td>WordNet, 3D models</td>
<td>Extracts subject, verb, object (SVO) using dependency parse</td>
<td>Rules match terms to models</td>
<td>3D scene per description; uses 3D models</td>
<td>Represents some abstraction</td>
<td>Scene creation without photorealism</td>
</tr>
<tr>
<td>Story Picturing Engine</td>
<td>A paragraph of text</td>
<td>WordNet, image libraries</td>
<td>Extracts nouns and ranks images</td>
<td>Matches nouns to pictures</td>
<td>One image per sentence</td>
<td>Innovative image-ranking</td>
<td>Summarizing pages, paragraphs, documents</td>
</tr>
<tr>
<td>ROC-MMS for Complex Sentences</td>
<td>One sentence</td>
<td>Rules, Wikipedia</td>
<td>Extracts SVO using dependencies</td>
<td>Matches nouns to pictures</td>
<td>&gt;1 image per sentence</td>
<td>Simplifies text</td>
<td>Summarizing multi-phrase sentences</td>
</tr>
<tr>
<td>Simple Sentence Picturing</td>
<td>One- or two-phrase sentences</td>
<td>WordNet, PicNet images</td>
<td>Extracts SV using semantic labels</td>
<td>&gt;1 image per sentence</td>
<td>Creates PicNet image dictionary</td>
<td>Summarizing simple text without photo-realism</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Enriching Textbooks</td>
<td>One paragraph of text</td>
<td>Internet search</td>
<td>Extracts keyphrases, (nouns, adjectives)</td>
<td>One to four images per paragraph</td>
<td>Uses text mining</td>
<td>Visual summary of textbook sections</td>
<td></td>
</tr>
<tr>
<td>Word2Image</td>
<td>Word being pictured</td>
<td>Flickr images</td>
<td>Extracts keywords, ranks pictures</td>
<td>Matches keywords to pictures</td>
<td>Finds several images per word</td>
<td>Uses text mining, term clustering</td>
<td>“Picturing” multiple meanings of a word</td>
</tr>
<tr>
<td>Text to Pictographs (for augmentative and alternative communication; AAC)</td>
<td>One to two sentences</td>
<td>Local icons</td>
<td>Extracts terms, simplifies</td>
<td>Matches terms to pictures</td>
<td>&gt;1 image per sentence</td>
<td>Rendering uses frequent image cache</td>
<td>Picturing for one to two simple sentences</td>
</tr>
<tr>
<td>Picturing Microblogs</td>
<td>Microblog entries with pictures</td>
<td>Crowdsourced user tags, embedded images</td>
<td>Extracts best images from series of posts</td>
<td>Graph-based ranking</td>
<td>Several images per concept</td>
<td>For Twitter, Tumblr, Facebook, etc.</td>
<td>Summarizing social media streams</td>
</tr>
</tbody>
</table>

**Unique Features**

The notable features column in Table 1 highlights the unique features of each picturing system.

- **WordsEye** includes a method of representing abstraction (for example, representing an idea as a light bulb), can connect near and far linguistic references, and can represent figures of speech.
- The **Story Picturing Engine** uses an innovative image-ranking algorithm to generate a semantic representation. This scheme could see reuse in similar applications.
- **ROC-MMS** simplifies sentences before detailed analysis. Simplification is used for longer sentences that include several modifying (propositional) phrases.
- **Simple Sentence Picturing** constructs a picture dictionary called PicNet to prepare for picturing. PicNet contains images for an indexed subset of terms from WordNet, including nouns and verbs.
- **Text to Pictographs** (for AAC) uses natural language processing in the form of part-of-speech tagging, lemmatization, and rules of grammar. A cache of frequently used word/image mappings is also used. Output scene rendering uses a combination of drawings from an AAC symbol collection.
- The **Picturing Microblogs** system visualizes content from a microblogging platform like Twitter. This implementation affirms that picturing can be applied to social media content with unconventional characteristics.
CONCLUSION

IT professionals are familiar with visual interfaces and how they can help user productivity. Text picturing applications similarly help users’ understanding of textual content and can help productivity. Several applications can benefit from text picturing, including AAC tools, classroom material, visual summary of reports, advertising, news, and social media.

Text picturing remains an active topic of research. There are many opportunities to realize its potential in real-world applications to help with text understanding.

Finally, we would like to point out that a related dual problem to the one introduced in this article is the task of automatic image caption generation. While deferring detailed discussion of this to others, we note that there is an important role for solutions to the dual problem; that is, the neural-network image-decoding approaches currently found in image-interpreting solutions. These solutions to the dual problem might find their way into solutions for text picturing.

ACKNOWLEDGMENTS

We thank the late professor Nick Cercone for the opportunity to get started in this field.

REFERENCES


**ABOUT THE AUTHORS**

**J. Kent Poots** is a PhD student at York University. His research interests include computational intelligence and biomedical engineering. Poots is a student member of IEEE and a professional engineer in the Province of Ontario. Contact him at kpoots@cse.yorku.ca.

**Ebrahim Bagheri** is an assistant professor at Ryerson University, where he leads the Laboratory for Systems, Software and Semantics (LS3). His research interests include large-scale software reuse and semantic web technologies. Bagheri is a Senior Member of IEEE and a professional engineer in the Province of Ontario. Contact him at bagheri@ryerson.ca.
The Design of a Software Engineering Lifecycle Process for Big Data Projects

There is currently no development process standard for big data projects. With the increasing number of such projects, the authors designed a new software engineering lifecycle process for big data projects, primarily based on ISO/IEC standard 15288:2008.

There is great divergence between traditional data analysis and big data analysis. The project goal, scope, and functional requirements in traditional data analysis or software projects are considered to be relatively more explicit than those of big data projects. There are four major characteristics used to determine whether a project involves big data: volume, velocity, variety, and variability (4V). In other words, a software project with the 4V characteristics is regarded as a big data project.

Currently, big data projects fail to reach a high completion rate: the incompletion rate is about 55 percent, whereas the incompletion rate for general software projects is about 38 percent. The difference can be attributed to inaccurate scope and value of the results. Among the 4V characteristics, variety is found to be the most essential and interesting as it brings about higher uncertainty and unexpected elements in data relation. In particular, when external data are added to big data, innovative data relation is created, which further improves the value of the results—for example, the application of big data in real-time traffic operations, decision making, and crime hot spots.

However, variety creates risks and difficulties in the management of big data projects as well. Problems occur in how the clients and the development team regard the results, which is also considered the striking difference between big data projects and general data analysis projects. When we studied big data projects, we found that they all contained standard processes found in software engineering, project management, journals, books, and public documents of global IT vendors. Considering the effect that variety has on big data, we designed a process suitable for
big data projects using inductive analysis and comparison. In this article, we explain the characteristics and benefits of the new process. We expect that it can be effectively applied in big data projects, helping to promote the value of information, establish a consensus between the development team and clients, and increase the completion rate of projects.

THE LACK OF A TYPICAL BIG DATA PROJECT LIFECYCLE PROCESS

So far, IEEE has not provided a development process standard for big data projects; only for software projects. Figure 1 is drawn from ISO/IEC 15288:2008—Systems and Software Engineering—System Life Cycle Processes, an international standard that clearly defines the process for general software projects.11 Data science and big data analysis are growing rapidly. A good process—more specifically, one that is suitable for big data projects—is needed.

Figure 1. ISO/IEC standard 15288:2008—Systems and Software Engineering—System Life Cycle Processes.11
SOFTWARE ENGINEERING LIFECYCLE FOR BIG DATA

VARIETY

The concept of variety in terms of big data was first proposed by Doug Laney in 2001 along with the following terms: inconsistency resolution, XML-based “universal” translation, application-aware enterprise application integration (EAI) adapters, data access middleware and ETLM (extract, transform, load, and management), distributed query management, and metadata management. Today, “variety” is interpreted in a somewhat different way by NIST, which describes the organization of the data and whether the data is structured, semi-structured, or unstructured.

To cope with variety in big data projects, it is necessary to search for data relevance in structured and unstructured data pools. Well-known business needs or uses are marked in the structured data and regarded as relatively easy to find. Thus, it is common to start from the known themes and structured data, followed by extending the data through integration with the unstructured data. However, discovering unstructured data directly or plugging in external data step by step proves to be the best way to acquire the most valuable results—in other words, building data hierarchies and statistics to discover the information relation and data trends. This is the most valuable part of a big data project, yet it is also regarded as one of the riskiest and most uncertain aspects of project management.

A process suitable for a big data project—especially when faced with variety—will therefore be required. Furthermore, the process should serve to change the ways clients and project teams define the scope and value of a project. It is expected that in the future, big data projects will operate even more smoothly with project risks reduced and completion rates increased.

DATA INNOVATION

For general data analysis projects or software projects, definite project goals or functions serve as the requirements for specification, followed by the work plan and implementation. However, variety in big data projects makes it impossible to completely verify the results of information application—the goal, in accordance with variety, should involve data-innovative processing and corresponding approaches. Since data innovation can lead to special value in results and uncertain factors, the data-innovation rate of a project should not account for 100 percent, but instead should be limited to a rate that can be controlled, such as 20 percent. As for the rest of the results, 80 percent should be limited, defined, and controllable.

When working with data innovation, it is advised to not be overly limited by certain factors such as goal orientation, data readability, data integrity, and information quality. The implementation of data innovation should instead seek any possible data trends and relations through different views, ranges, properties, and dimensions, or by other scientific techniques such as statistics and multivariable methods.

Figure 2. Goals and objectives of big data projects should be 80 percent data analysis and 20 percent data innovation.
Data innovation has three essential aspects. First, the development team and clients reach a consensus, getting clients to realize the importance of data innovation and then redefine the innovation rate of the project scope. Second, the data-innovation process is added to data processes when necessary. And finally, cycle steps for data innovation are established.

DESIGNING A LIFECYCLE PROCESS FOR BIG DATA PROJECTS

There are four major elements involved in designing a process suitable for big data projects: one characteristic, one concept, and two processes. The characteristic refers to data variety, the concept is data innovation, and the processes are software engineering and data analysis.

![Figure 3. Major elements of the big data project lifecycle process.](image)

To deal with variety in big data projects, it is advised to establish the processes described below.

- **Data value, result, and innovation process (agreement process).** It is regarded as a risky undertaking for big data project contracts to only define project goals without including the data scope. Project risks can be prevented by first defining and controlling the data scope.
- **Domain specialist resource management process (organizational project-enabling process).** Due to the variety in big data, it is likely to grow more complicated in the management of interdisciplinary personnel. There should be an independent process set aside to be checked by a specialist, with the resources coming from clients or external experts.
- **Data inventory process (data process).** After the data is collected, the data inventory is performed for management. Data information is supposed to contain data format, type, source, amount, timestamp, states, renew period, owner, and so on.
- **Data requirement analysis process (data process).** This is undertaken for realizing the expected results and value.
- **Data cleaning process (data process).** To prevent data variety from being deleted, it is advised to clean the data after the data-innovation process is done.

To cope with data innovation, it is advised to establish the following processes.
• Data value, result, and innovation process (agreement process). Agreement processes are supposed to be used to make changes. It is required that 80 percent of the deliverables be confirmed for target results, and 20 percent should be reserved for data innovation. This 20 percent will likely offer clients the maximum value.

• Data innovation process (data process). This is considered a task force where data analysts seek any possible data trends and relations, and are never limited by scope or goals. It is necessary to resume the data integration process, data inventory process, or data collecting process as soon as new data innovation takes place. See Figure 4 for the steps and flow cycle of the data innovation process.

Figure 4. Data innovation process and cycle.

It is advised that data processes serve as an independent process from the project processes and technical processes. Also, data processes should contain the following processes: data collecting, data inventory, data requirement analysis, data integration, data verification, data analysis, data modeling, data simulation, data prediction, data innovation, data validation, data cleaning, and data maintenance.

To deal with data processes, it is advised to establish the following technical processes.

• Data automation and monitoring process (technical process). These processes are mainly concerned with the establishment of a mechanism, by means of technical approaches, to collect and monitor data automatically and continually. The mechanism is supposed to be able to prevent the data source from being anomalous so that it gets only the right results.

• Data visualization process (technical process). Data visualization deserves great emphasis, as it is viewed as a very important part of a big data project. It is also important to make sure that results can be integrated with a visual tool or platform.

• Data decision support process (technical process). Most data projects are applied in the decision support for businesses or government. This process is mainly concerned with analysis and application of the results.

We used the above-mentioned processes, along with ISO/IEC 15288:2008, to design a lifecycle processes for big data projects, shown in Table 1.
Table 1. Big data project lifecycle processes.

<table>
<thead>
<tr>
<th>Agreement processes</th>
<th>Project processes</th>
<th>Data processes</th>
<th>Technical processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data value, result, and innovation process</td>
<td>Project planning process</td>
<td>Data collecting process</td>
<td>Stakeholder requirement definition process</td>
</tr>
<tr>
<td>Acquisition process</td>
<td>Project assessment and control process</td>
<td>Data inventory process</td>
<td>Requirement analysis process</td>
</tr>
<tr>
<td>Supply process</td>
<td>Decision management process</td>
<td>Data requirement analysis process</td>
<td>Architectural design process</td>
</tr>
<tr>
<td>Risk management process</td>
<td>Data integration process</td>
<td></td>
<td>Data automation and monitoring process</td>
</tr>
<tr>
<td>Configuration management process</td>
<td>Data verification process</td>
<td></td>
<td>Data visualization process</td>
</tr>
<tr>
<td>Information management process</td>
<td>Data analysis process</td>
<td></td>
<td>Data decision support process</td>
</tr>
<tr>
<td>Measurement process</td>
<td>Data modeling process</td>
<td></td>
<td>Implementation process</td>
</tr>
<tr>
<td>Organizational project-enabling process</td>
<td>Data simulation process</td>
<td></td>
<td>Integration process</td>
</tr>
<tr>
<td>Lifecycle mode management process</td>
<td>Data prediction process</td>
<td></td>
<td>Verification process</td>
</tr>
<tr>
<td>Infrastructure management process</td>
<td>Data innovation process</td>
<td></td>
<td>Transition process</td>
</tr>
<tr>
<td>Project portfolio management process</td>
<td>Data validation process</td>
<td></td>
<td>Validation process</td>
</tr>
<tr>
<td>Domain specialist resource management process</td>
<td>Data cleaning process</td>
<td></td>
<td>Operation process</td>
</tr>
<tr>
<td>Human resource management process</td>
<td>Data maintenance process</td>
<td></td>
<td>Maintenance process</td>
</tr>
<tr>
<td>Quality management process</td>
<td></td>
<td></td>
<td>Disposal process</td>
</tr>
</tbody>
</table>
CONCLUSION

To evaluate its feasibility and performance, the process described in this article has been applied to some big data projects in the banking and retail industries in Taiwan. The organizations have promised to collect process-related experiences, information, and artifacts derived from executing the processes. These include work products, measurement results, lessons learned, and process improvement suggestions. The process-related experiences and assets will be collected and stored in the organizations’ measurement repository and process asset library so that we can further evaluate the performance of the processes and improve them regularly. We will also make these experiences and assets available to those who are planning to use these processes in the near future.

REFERENCES


ABOUT THE AUTHORS

Yen-Tai Lin is a researcher in the Department of Information Management at the National Taiwan University of Science and Technology. He has more than 15 years of experience in system integration and information management. Lin’s research interests include manage-
ment information systems, cloud computing, the Internet of Things, and big data applications. He received a master’s degree in interactive multimedia systems from Liverpool John Moores University. Contact him at D10309203@mail.ntust.edu.tw.

Sun-Jen Huang is a professor in the Department of Information Management at the National Taiwan University of Science and Technology. His research interests include software engineering and project management, software process improvement, software measurement and analysis, and software quality management. Huang received a PhD in software engineering from La Trobe University. Contact him at huangsj@mail.ntust.edu.tw.
Improving Transparency and Efficiency in IT Security Management Resourcing

The authors propose a resource management process for information security management systems to more transparently plan and assign costs of controls. The process relies on and is compliant with international standards of the ISO/IEC 27000 family and can be implemented by all organizations regardless of type, size, or nature.

Nearly all organizations today depend on appropriate secure and compliant information processing. This has been stated practically in relevant standards and frameworks as well as in the literature.1,2 Furthermore, there is a growing consensus that such standards need to be properly operationalized with unequivocal policies, processes, and metrics.3 Standards for the management of information security have been developed and established, with those devoted to the development and operation of an information security management system (ISMS) published in the ISO/IEC 27000 series.4,5

IT governance, although recognized as an important topic in academic and professional environments, often is not a top priority of executive management.6,7 In this scenario, commoditization and novel technology delivery models such as cloud computing8 often lead to an increase in the number of IT stakeholders, making IT governance even more complex. This environment requires novel governance models such as participatory and shared governance.9 These developments make the planning, implementation, and maintaining of ISMS processes and controls even more costly, keeping in mind that information security and data protection are traditionally considered to be cost drivers. Security projects have often been justified on fear, uncertainty, and doubt.10 Over the past few years, cost-benefit discussions have influenced information security
The value of information must justify protection and personnel costs and—particularly in the context of shared, participatory governance—this justification must be even clearer. Adjustment and cost-effectiveness are key elements of cost governance and prerequisites for a successful ISMS, while adequate funding of ISMS processes and controls is important and necessary to reach ISMS objectives.

ISMS budgets are often stressed, and additional funding is often requested to implement necessary controls. However, in many cases controls are not realized due to budget problems. Often, it is assumed that controls are paid for from the ISMS budget. This results in a nontransparent and irrational allocation of costs. Instead, costs should be divided among the individual cost centers according to the information security requirements. As a prerequisite for this, an adequate differentiation between costs for specific controls and costs for running and maintaining the ISMS processes is necessary. Such a differentiation would allow the ISMS to assign costs for controls to the owners whose assets determine information security or data protection requirements.

This problem is still relevant and unsolved because in the current standards, there is no clear differentiation between ISMS processes and controls. For instance, ISO/IEC 27002 contains several controls regarding information security incident management and information security reviews, which are ISMS processes and not controls. Available methods (such as ROI and ROSI) to decide how much to invest in security and privacy are not generally accepted because of their complexity. To address this open research question, we propose a resource management process for information security management systems to more transparently plan and assign costs of controls. The proposed process relies on international standards of the ISO/IEC 27000 family and is compliant with them. The proposed process can be implemented by all organizations regardless of type, size, or nature.

STATE OF THE ART IN INTERNATIONAL STANDARDS

ISO and IEC formed a joint technical committee (ISO/IEC JTC 1) for experts to come together to develop worldwide information and communication technology (ICT) standards for business and consumer applications. The subcommittee SC 27 has a working group (WG 1), which develops and facilitates international standards for ISMSs. ISO/IEC 27001, as the international standard from ISO/IEC JTC 1 SC 27 WG 1 for ISMSs, is the security standard in enterprises.

ISO/IEC 27001 contains the requirements for planning, implementing, operating, and improving an ISMS. Requirements are formulated in a general manner to fit all organizations independent of their size, objectives, business model, location, and so on. Absolutely no requirements are formulated for any specific technology in ISO/IEC 27001, but the standard contains requirements for ISMS core processes.

The ISO/IEC 27000 series does not consist only of 27001. The second common standard for information security in the 27000 series is 27002, which contains the controls that should be implemented with an ISMS. 27002 is linked with 27001 with an annex of 27001 listing the controls of 27002. Further standards of the 27000 series are as follows.

- 27000—Information Security Management Systems
- 27003—Information Security Management Systems—Guidance
- 27005—Information Security Risk Management
- 27006—Requirements for Bodies Providing Audit and Certification of Information Security Management Systems
- 27007—Guidelines for Information Security Management Systems Auditing
- 27008—Guidance for Auditors on Information Security Controls
- 27010 and following—Information Security Management for Intersector and Interorganizational Communications
- 27030 and following—Standards for Technical Controls and Guidelines for Controls of ISO/IEC 27002
According to ISO/IEC 27000/27001, an ISMS process (which needs to be designed) is the resource management process, which ensures that necessary resources are determined and provided. Nevertheless, no further information on how to define or establish such a process is provided in the standards.

The information security risk treatment process is a process to select and implement controls to mitigate risk. Controls are determined during the process of risk treatment, rather than being selected from Annex A of ISO/IEC 27001 (see www.isaca.org/COBIT/Pages/COBIT-5-Enabling-Processes-product-page.aspx).

**THE PROCESS**

Governance deals with value delivery to the stakeholders, risk optimization, and resource optimization (see www.isaca.org/Knowledge-Center/cobit/Documents/COBIT4.pdf). Governance is also the responsibility of executives and boards of directors, and consists of the leadership, organizational structures, and processes that ensure that the organization’s strategies and objectives are reached. As leadership, risk assessment, risk treatment, and resource management are elements of the ISO/IEC 27001, the concept of cost governance in IT security management is integrated in ISO/IEC 27001. Here, we focus on the proposed resource management process as part of cost governance in IT security management.

The resource management process is the process to identify, allocate, and monitor required resources to run the ISMS core processes, as well as to implement and run the selected controls. So, a resource management process is key to efficiently use limited resources as part of the governance of IT security costs.

A resource management process is also part of the ISMS planning process. However, we focus on the management process for the resources necessary to operationally run the ISMS or security controls.

The resource management process needs to be carried out on a regular basis because it is integrated in the ISMS and continuously supports the ISMS processes as well as the controls by identifying, allocating, and monitoring the required resources. So, this is not a one-time task.

Table 1 and Figure 1 contain a description of the proposed process.

<table>
<thead>
<tr>
<th>Process name</th>
<th>Resource management process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process categorization</td>
<td>Information security management system (ISMS) core process</td>
</tr>
<tr>
<td>Brief description</td>
<td>The resource management process identifies, allocates, and monitors required resources to run the ISMS core processes and to implement and run the selected controls. It enables the governance of costs for IT security management.</td>
</tr>
<tr>
<td>Objectives/purposes</td>
<td>Ensure that the resources for the ISMS and the controls are available</td>
</tr>
<tr>
<td></td>
<td>Appropriate management of ISMS resources</td>
</tr>
<tr>
<td></td>
<td>Ensure efficiency of resource usage</td>
</tr>
<tr>
<td>Input</td>
<td>From risk treatment process: drafts and final list with selected controls</td>
</tr>
<tr>
<td></td>
<td>From ISMS planning project: management approval for initiating the ISMS and documented business case for the ISMS/project proposal</td>
</tr>
<tr>
<td>Output</td>
<td>From purchasing department: list of suppliers, framework contracts, and terms and conditions of purchasing</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
| Activities/functions | For information security risk treatment process: estimation of necessary resources to implement controls  
For communication process: estimation of necessary resources to operate the ISMS core processes and reports regarding resource usage of ISMS core processes  
For information security customer relationship management process: reports on resource usage |
| Activities/functions | Initially plan necessary resources to implement and run the controls  
Categorize controls: a differentiation is made between controls funded by the ISMS budget and controls funded by other departments  
Communicate necessary resources to the information security risk treatment process to implement and run the controls; if necessary, repeat this step and the planning of necessary resources  
Communicate necessary resources to the communication process regarding the ISMS controls (necessary to operate the ISMS core processes)  
Allocate necessary resources for approved controls funded by the ISMS  
Permanently monitor ISMS resource usage, and, if necessary, update resource allocation  
Develop and communicate reports regarding resource usage of ISMS core processes to the information security officer (ISO) |
| Sample metrics | Key goal indicators:  
Count of activities/functions of the ISMS that were not performed due to a lack of resources  
Key performance indicators:  
Resources to implement the selected controls  
Planned but not used resources and resource trends  
Count of controls with budget overrun  
Ratio between ISMS-funded controls and otherwise-funded controls |
| Owner | ISO |
| Manager | ISO or resource manager |
| Actors | ISO or resource manager  
Purchase and human resources department  
Consultants and specialists |
| Interfaces | This process is interlinked with the information security risk treatment process and all other ISMS core processes as well as processes of the purchasing and human resources departments. |
Figure 1. Resource management process flow chart.
Key to the proposed resource management process is the categorization of controls that must be funded by the ISMS and controls that should be funded by other departments.

The ITIL framework\textsuperscript{21} integrates the idea to charge costs for IT services to the functional units that use IT services. Considering this idea, it sounds obvious to charge costs for information security measures to the functional units demanding those measures. To realize this, a clear differentiation should be made between ISMS processes financed by the chief information security officer (CISO) and other security measures financed by other cost centers. As a result, information security costs are made transparent and can be better managed.

We propose using this categorization as the main criterion to differentiate controls funded by the ISMS and controls funded by other departments. Only ISMS core processes should be financed by the ISMS. All other controls should be charged to functional units, departments, or cost centers demanding those controls.

Core processes deliver apparent and direct customer value and are derived from the core competencies of an organization. From the perspective of the ISMS, a core process is the core competency of the ISMS and delivers apparent and direct value to the stakeholder (generally the organization implementing the ISMS and the management of the organization).

Thinking about processes in general as well as ISMS processes means asking what needs to be done on a regular basis.\textsuperscript{22} Therefore, ISMS core processes are operational processes (mainly, but not limited to, part of the “Do” and “Check” steps of the PDCA cycle) that need to be repeated regularly. Furthermore, a criterion for the categorization as an ISMS core process is that the process is owned by the information security officer (core competency of the ISMS). This means that the information security officer is accountable for the whole process and defines objectives and goals for the process derived from the objectives of the ISMS, which will be defined by senior management.

In many practical cases, especially in small organizations, the information security officer will also be the process manager responsible for the process operation. This could also be used as a criterion for ISMS core processes.

We propose the following basic criteria to identify ISMS core processes.

1. Regularity—interrelated and interacting tasks are repeated on a regular basis.
2. Transformation—inputs are transformed into outputs.
3. Operation—process is carried out while operating the ISMS.
4. Accountability/responsibility—the information security officer is the process owner or process manager and the process is a core competency of the ISMS.
5. Value generating—delivers apparent and direct value to the stakeholder.

All criteria must be fulfilled to qualify a process as an ISMS core process. Criteria 1 and 2 are basic criteria to identify processes, but they are necessary to exclude one-time tasks like obtaining management’s commitment to establish an ISMS or tasks within a process. Criterion 3 is mainly a basic process criteria, but is necessary to exclude processes that are not part of the operation of the ISMS. An example is the initial ISMS planning process, which is carried out during the ISMS planning phase. Criterion 4 is the main criteria to decide whether the process is ISMS-owned and requires ISMS core competencies. Criterion 5 is necessary to exclude supportive processes like the documentation and records control process from the ISMS core processes.

These criteria should not be used in a binary way (fulfilled or not fulfilled). Instead, the level of fulfilment of a criterion (for example, how much value is generated by the process) should be used to discuss and decide whether the process can be understood as an ISMS core process.
EVALUATION RESULTS

The proposed resource management process was implemented as a pilot project in a software development company specializing in healthcare. The company is located in Germany (one main office and an additional branch office), has a staff of 50, and uses an outsourced datacenter.

Altering the resource management process was part of a project to implement ISO/IEC 27001 compliance. A preliminary project to implement an ISMS had failed the year prior to the new project (budget for external consultants: 60 person days). From our viewpoint, the main reason for this was that the ISMS project tried to compensate for unavailable information within the project budget. For example, in the failed project it was identified that the IT documentation was out of date and incomplete, so about 20 person days from the project budget were spent to gather information about the IT. This not only unnecessarily reduced the limited project budget, but it was also unsustainable because the reason for the inappropriateness of the documentation was not discovered.

IT documentation and the change-management process were not part of the new ISMS project. We also recognized a strong barrier from the IT staff when they were only told that the IT documentation was not appropriate, what the requirements were, and that they needed to solve the problem within the IT budget. We needed to communicate additional requirements and to convince management and every administrator why it is necessary and beneficial (for every individual) to maintain appropriate IT documentation.

The existing risk-management processes were altered to integrate interfaces for the new process, which was not in place at all. The company had no idea what it cost to operate an ISMS because costs for security controls were not differentiated. Another main challenge of the new process was that not all risk-treatment options and security controls were approved by the information security officer. In the new process, the risk owner is generally accountable for deciding how to mitigate the risk. This is not new, but the clear differentiation between general security risks (mitigated within ISMS core processes) and specific security risks outside the ISMS (such as inappropriate IT documentation) helped focus the ISMS budget on the establishment and operation of the ISMS.

The main challenge regarding resource-usage monitoring within the ISMS was to identify significant metrics like time spent for risk analysis and to continually gather the necessary data for those metrics. The company is still improving this ISMS resource-usage monitoring. It was also difficult to include processes with a low maturity level in the resource-usage monitoring. The company decided to implement an information-security-awareness process at a repeatable maturity level because they felt that all employees were highly skilled. So, they focused on informal and ad-hoc awareness measures instead of planned, documented trainings. Gathering resource usage data for processes like this is still a challenge for the company.

The new resource management process is now in place and has been operating for six months at the time of this writing. During this time, feedback from senior management and the information security officer was gathered continually during formal meetings and informal discussions.

From the viewpoint of senior management, two main advantages of the new process were recognized compared to the former processes:

1. More transparency of information security costs. Due to the separation of costs for the ISMS and costs for individual controls, senior management feels better informed about the use of resources for information security. This also positively influences discussions about security controls among management, departments, and the information security officer.
2. Improved decision making. Senior managers feel more confident in the decision-making process regarding information security controls and have a better understanding of where costs are coming from.

From the viewpoint of the information security officer, two main additional advantages of the new resource management process were recognized compared to the former processes:
1. Improved image. Because of the transparency of cost allocation, information security is no longer recognized as a global cost driver. This objectifies budget discussions.

2. Improved efficiency. Time spent for budget discussions was dramatically reduced. Because of the clear differentiation of ISMS costs and costs for controls, the information security officer no longer had to fight for budget for the controls aside from the ISMS core processes. Instead, this task was assigned to the business units. Also, nearly 90 percent of the security requirements gathered in the failed project were reduced due to the risk owner’s new understanding that they needed to fund the security controls.

A direct measurement and comparison of the costs before and after implementing the new process is difficult because no process was in place when the project began and the ISMS itself was not complete.

CONCLUSIONS AND FUTURE WORK

The pilot implementation of the resource management process proved that a differentiation between costs for the ISMS and costs for individual controls can shift the image of information security from a cost driver to a transparent and manageable success factor. Furthermore, it provides a precise answer to the question, “Who owns IT?” in the specific area of ISMSs. IT governance can be improved by the process as it provides cost transparency even in complex shared or participatory governance settings.

Future work is envisioned in three main steps:

- Further evaluation of the proposed process. Results of the evaluation of the resource management process will be analyzed empirically. Changes and consequences of implementing the process must be specifically measured to allow a more reliable evaluation of the process. Consequences for the long-term budgets for ISMS and information security controls are of special interest.

- Development of an ISMS core process framework. To the best of our knowledge, there is not a specific process framework for security management that clearly differentiates between ISMS processes and the security measures controlled by ISMS processes. Furthermore, a detailed description of ISMS processes and their interaction as well as the interaction with other management processes does not exist. This ISMS process framework should be developed and data across industries should be gathered to demonstrate the viability and effectiveness of the proposed framework.

- Development of a method to adjust and make costs transparent for operating the ISMS core processes. Transparency of information security costs could be further improved by tailoring the maturity level of ISMS processes to the requirements of the organization. Assuming that not every ISMS process needs the same level of maturity, an approach should be developed to identify the appropriate level of maturity using a proper maturity model. This model, together with the specific assessment tools to determine the appropriate maturity model for an organization, will optimize information governance costs. Furthermore, costs can be attributed properly to specific services with respect to their configuration and thus provide more transparent cost information in the marketplace for cloud services.

REFERENCES


ABOUT THE AUTHORS

Knut Haufe is managing director and lead expert for information security management systems at PwC Cybersecurity Services. His research interests include process reference models, integrated management systems, and maturity levels. Haufe received a PhD in information technology from the Universidad Carlos III de Madrid. He is a member of the Standards Committee on Information Technology and Applications (NIA) 043-01-27-01 of the DIN (German Institute for Standardization). Contact him at khaufe@pwc-cybersecurity.com.

Srdan Dzombeta is a partner at PwC Cybersecurity Services. His research interests include internal control systems and cloud systems. Dzombeta received a PhD in information technology from the Universidad Carlos III de Madrid. Contact him at sdzombeta@pwc-cybersecurity.com.

Knud Brandis is a partner at PwC Cybersecurity Services. His research interests include information security and risk management. Brandis received an MBA in financial management from the University of Wales. Contact him at kbrandis@pwc-cybersecurity.com.

Vladimir Stantchev is the executive director of the Institute of Information Systems at SRH University Berlin, where he is also a research professor. He is also a professor at the University of Granada and an affiliated senior researcher with the Networking Group at the International Computer Science Institute (ICSI). Stantchev’s research interests include IT governance, cloud computing architectures, IT strategy, and methods for service and software engineering. He received a PhD in system architectures from the Berlin Institute of Technology. Stantchev is a Senior Member of the IEEE Computer Society and ACM. Contact him at stantchev@computer.org.

Ricardo Colomo-Palacios is a full professor in the Computer Science Department at Østfold University College. His research interests include applied research in applied information systems including IT governance and service management. Colomo-Palacios received a PhD in computer science from the Universidad Politécnica of Madrid. Contact him at ricardo.colomo-palacios@hiof.no.
Take the CS Library wherever you go!

IEEE Computer Society magazines and Transactions are available to subscribers in the portable ePub format.

Just download the articles from the IEEE Computer Society Digital Library, and you can read them on any device that supports ePub, including:

- Adobe Digital Editions (PC, MAC)
- iBooks (iPad, iPhone, iPod touch)
- Nook (Nook, PC, MAC, Android, iPad, iPhone, iPod, other devices)
- EPUBReader (Firefox Add-on)
- Stanza (iPad, iPhone, iPod touch)
- ibis Reader (Online)
- Sony Reader Library (Sony Reader devices, PC, Mac)
- Aldiko (Android)
- Bluefire Reader (iPad, iPhone, iPod touch)
- Calibre (PC, MAC, Linux)
  (Can convert EPUB to MOBI format for Kindle)

www.computer.org/epub
Bad Security Metrics

Part 1: Problems

Security metrics are numerous and in high demand. Unfortunately, measuring security accurately is difficult and many security metrics are problematic. The problems with security metrics can be complicated and subtle. However, using measurement theory, it’s possible to determine quickly that many metrics are unfit for the purposes for which they are used without venturing into subtle or subjective analysis.

This two-part series doesn’t call out questionable metrics that are in use. Instead, it focuses on defining the problem conceptually and revealing a path forward for improving both security metrics and how people use them.

SCALE THEORY

Scale theory is a small part of the broad discipline of measurement theory. It was popularized in 1946 by Stanley S. Stevens. Scale theory isn’t without problems, but is so well known that it’s often the most practical way to frame a discussion. The scales needed for this discussion are introduced briefly in Table 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio (and absolute)</td>
<td>The scale of most physical measurements of length, mass, time, and so on, where units can be converted through simple multiplication. A ratio scale with a supposedly unique, “natural” unit is sometimes called the absolute scale.</td>
</tr>
<tr>
<td>Interval</td>
<td>The scale of Celsius and Fahrenheit temperatures, which have meaningful units but arbitrary zero points.</td>
</tr>
<tr>
<td>Ordinal</td>
<td>The scale of relative ranks and grades, and generally of ordered values with no meaningful units, where the subtraction of one value from another does not yield a meaningful quantity.</td>
</tr>
<tr>
<td>Nominal (and dichotomic)</td>
<td>The scale of naming or classifying things. Nominal values have neither magnitude nor order. The two-valued nominal scale of a generic two-</td>
</tr>
</tbody>
</table>
way partitioning (usually yes/no or true/false) has been called the dichotomic or dichotomous scale.

The following sections identify six characteristics of bad security metrics and explain how they are problematic.

MAKE-BELIEVE MEASURAND

The *International Vocabulary of Metrology*[^3] defines “measurand” as the “quantity intended to be measured,” and it defines “quantity” as a “property of a phenomenon, body, or substance where the property has a magnitude that can be expressed as a number and a reference.” For example, “5 m” means 5 times the standard unit of measure, which is the meter.

The first problem we encounter with security metrics is simply that they are often put forth as measurements of security as if security were a quantity to be measured. A statement such as “X is 5 times as secure as Y” is meaningless on its face unless the general term “security” is given a specific interpretation. For example, if X is claimed to be 5 times as secure as Y because an attack against it is expected to take 5 times as long, then security has been interpreted to mean a specific quantity of time for which a standard unit of measurement already exists.[^4] Without such an interpretation, there can be no unit of security, and security can’t be a quantity in the metrological sense.

Furthermore, security isn’t a property of an IT artifact in anything remotely like the way that mass is a property of a physical artifact. It isn’t inherent or intrinsic to the artifact. As described in *A Rational Foundation for Software Metrology*:

> To be meaningful, security for any system, software or physical, must be defined according to some specification of the protection that it will provide in a particular threat environment. Assumptions must be made regarding the capabilities and ingenuity of the adversary and the pace of future technological progress, all of which are unknown.[^5]

Unlike a normal physical measurement that captures information about the present state of a system, an assessment of security is primarily a forecast about what can or will happen in the future.[^6] It can be an observation of present facts only in the case that security is absent.

MISUSE OF ORDINAL SCALE

The non-comparability of differences on an ordinal scale makes them unfit for the purpose of making security-relevant decisions. Consider the example of choosing an email filter to keep out phishing, and consider the two different metrics shown in Table 2 that could be used to evaluate competing products. Assume we have narrowed the choices to filter A and filter B, with the attributes shown in Table 3.

If we only had metric 2, we would have an indication that filter A performed better than filter B, but we would have no way to evaluate the magnitude of the performance advantage and no rational way to decide whether it was worth the money. Only a wealthy fool would buy the top-performing tool without looking at the price–performance tradeoff.

This example was simplified by the assumption that a better metric existed, but the argument holds generally: ordinal metrics give no indication of the magnitudes of differences, and therefore can’t be used to evaluate the tradeoffs of security against cost, usability, and so on that are critically important to real-world security planning.
Table 2. Example metrics for email filters.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Proportion of phishing attempts blocked, for a representative sample</td>
<td>Ratio/absolute</td>
</tr>
<tr>
<td>2</td>
<td>Relative ranking of the products’ performance (1 means best, 2 means second-best, and so on)</td>
<td>Ordinal</td>
</tr>
</tbody>
</table>

Table 3. Product choice scenario.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Metric 1</th>
<th>Metric 2</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>83%</td>
<td>1</td>
<td>$10,000</td>
</tr>
<tr>
<td>B</td>
<td>82%</td>
<td>2</td>
<td>Free</td>
</tr>
</tbody>
</table>

FALSE PRECISION

In Table 3, results were stated without any indication of uncertainty. One might assume that this means they are sufficiently accurate for any reasonable use to which one would put them. This often isn’t the case.

Without an indication of the uncertainty of results, we have no reason to believe that the difference in performance was actual. Generally accepted measurement practice calls us to test the products not with one sample but with many different ones, and then report the mean performance with a confidence interval. The results would then look something like Table 4.

This rendering is consistent with the first, but the confidence intervals reveal that the difference between products isn’t statistically significant. The results without uncertainty, therefore, could be worse than useless, potentially misleading someone into paying $10,000 extra for performance that’s actually worse.

Table 4. Product choice scenario with confidence intervals.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Metric 1</th>
<th>Metric 2</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(83 ± 3)%</td>
<td>1 to 2</td>
<td>$10,000</td>
</tr>
<tr>
<td>B</td>
<td>(82 ± 5)%</td>
<td>1 to 2</td>
<td>Free</td>
</tr>
</tbody>
</table>

MISLEADING SCALE

A misleading scale is a way of expressing a result that leads the reader to make inferences that are unsound or untrue. For example, the numerical range of an ordinal measure is completely arbitrary and meaningless, but often they’re scaled to produce a maximum value of 10 or 100. This leads the audience to confuse them with probabilities and proportions and then use them as if they were on a ratio scale.

Another common misleading practice is using a count of rule violations or nonconformities as a “badness” metric. The problem here is the induced assumption that having \( n \) rule violations is \( n \) times as bad as having a single rule violation, or alternately, that violating \( n \) different rules is \( n \) times as bad as violating a single rule. The count is legitimately on a ratio or absolute scale, but only while it’s interpreted literally as a count of rule violations and nothing more. As soon as it’s interpreted as a metric of badness, it becomes misleading. For example, if two different rules address two different attack vectors, either of which would allow for complete system compromise with high likelihood and at negligible cost to the attacker, it’s hard to argue that
violating both rules simultaneously is even approximately twice as bad as violating only one. It’s somewhat analogous to multiplying an infinite value (infinite badness) by a constant.

COMBINING DISPARATE MEASURES

Attempts to synthesize a metric for nontrivial qualities often resort to weighted sums or more complex functions of disparate measures that in principle aren’t even comparable with one another as quantities. The combination functions have no theoretical basis or measurement principle to justify their derivation. The mathematical notation might lead a reader to assume that the functions are derived using sound principles, but it might instead be a veneer that hides flaws. The functions might merely have been designed to produce the desired results for the cases tested.

Combining disparate measures can make it difficult or impossible to produce a valid metric. Properties that might be lost include:

- Correlation: in repeated use on different objects, the measured quantity value should exhibit a consistent correlational relationship to the measurand.
- Tracking: a change in the measurand should always result in the measured quantity value moving in the same direction.
- Consistency: measured quantity values should always preserve a relative ordering by the measurand.
- Predictability: for metrics that are used in forecasting, prediction error should always be within the specified range or tolerance limit.
- Discriminative: high and low values of the true quantity should be distinguishable from the measured quantity values.


NAIVE USE OF HUMAN INPUT

When humans are the measuring instruments, subjectivity and human factors introduce distortion. For example, a respondent who is asked to evaluate something on a five-point scale might perceive the five levels as covering different-sized intervals or might “grade on a curve” based on a belief about what the distribution of results is supposed to be. To mitigate such distortion, social scientists have been migrating to more robust models, such as the graded response model within item response theory, whose validity depends on less onerous assumptions. Unfortunately, those advances have not transferred to analogous activities in IT risk assessment, such as estimation of likelihoods and impacts. Unless human factors are taken into consideration, the uncertainty associated with various human-produced scores will be large.

CONCLUSION

This article, the first of a two-part series, provided background on scale theory and explored different types of problems that afflict security metrics. Part two will continue with an explanation of how to avoid these issues, and finish with conclusions.

ACKNOWLEDGMENTS

Thanks to Paul Black and Barbara Guttman of NIST for their helpful review comments and suggestions.
REFERENCES

ABOUT THE AUTHOR
David Flater is a computer scientist in the Software and Systems Division of NIST’s Information Technology Laboratory. Contact him at david.flater@nist.gov.

Disclaimer: Products may be identified in this document, but such identification does not imply recommendation by the US National Institute of Standards and Technology or the US Government, nor that the products identified are necessarily the best available for the purpose.
The Role of a Customer Data Platform

At the end of the day, everything boils down to providing value to our customers. The more we know about our customers, the better we can provide that value and stand out from our competitors. Although this objective sounds simple, many challenges arise along the way. The entire customer experience relies on data, so creating a positive and compelling experience depends on the quality and types of available data to understand the needs of the customer. The challenges to getting the right data fall into four broad categories related to systems that produce and consume customer data: the ways that data is modeled, the data itself, gaining insights from the data, and acting on the data.

Customer data platforms (CDPs) aggregate data from many different sources to provide a 360-degree view of the customer. These platforms are designed to be managed and used directly by marketers, and they eliminate the need to access multiple systems to create customer profiles, develop marketing campaigns, test the effectiveness of marketing strategies, and predict customer behavior.

WHAT IS A CDP?

According to the Customer Data Platform Institute (CDPI), a CDP is “a marketer-managed system that builds a unified, persistent customer database that is accessible to other systems.” The CDP acts as a centralized clearinghouse and repository for all sorts of data from various internal and external systems. Consider any place where a customer interaction is recorded, tracked, or managed. Past purchases constitute a big category of customer behavior, of course. But so do social media interactions and website visits, even when nothing is actually purchased. Collectively, this data produces signals that can be thought of as “electronic body language.”

Some data is reasonably straightforward (such as name, address, and demographic details). However, some information requires processing and interpretation. Clickstream data, for example, tracks part of the customer journey and can be very informative, but understanding what it means requires effort and human intervention. Data about website behavior can be stored in a CDP, but the dataset is large and has numerous components that are time- and context-dependent. Before the data can be meaningfully acted upon, it must be analyzed and interpreted.

Ability to Summarize Data or Surface Trends

A CDP can be used to summarize certain types of data to present a trend or characteristic that can then be exploited. Rather than require a marketer to sift through hundreds of log lines from a
website visit, certain events or behaviors can be segmented and acted upon. These could include offers, promotions, content that similar users find valuable, or other relevant signals derived from large datasets.

**Integration of Varying Data Formats and Structures**

Another valuable function of CDPs is their ability to accommodate different data types and formats that might have varying structures and naming conventions. Data might come into the CDP through a live feed via an API or web service layer, or might be input on a batch basis through a file transfer. Formats can be structured by transaction such as accounting and purchase data or unstructured content such as chat logs or Facebook conversations, tweets, and even images from Instagram. This varied data can then be put into a consistent format (through rules retained in the CDP) that can be more easily interpreted by a marketer or acted upon by other systems without repeated manual running of extraction, translation, and load scripts.

**Ability to Cleanse and Process Data**

Data might have redundant records and content; for example, it might be missing details (fields or values) or contain incorrect data that has to be reconciled with another system. When new data is available, only certain values can be updated. Other data sources can be used to enrich the data or append missing information. Rules for cleansing, enriching, appending, and correcting data can reduce the cost and complexity of data hygiene by automating remediation.

**Exposing Data for Use by Other Systems**

A major function of the CDP is to act as a centralized location so other systems can access and act upon customer data. The CDP becomes a broker or orchestration layer that can take the outputs from one customer-facing application, process it, convert the format, and export it or make it available via an API for a personalization engine. The personalization engine might then send data back to the CDP, which can use the results to inform or drive another process.

**Doesn’t My Marketing Automation System Do This?**

A CDP can provide some of the functionality of other marketing systems and customer engagement platforms, but it is fundamentally different in design and function. Marketing automation systems can integrate with other tools but usually in a limited fashion to suit a narrow set of use cases. CDP tools are designed from the ground up to talk to other systems. They also retain details from other systems that the engagement or automation tool does not. This is valuable for trend analysis, predictive analytics, and recommendations that can leverage large amounts of historical data.

The various functions of a CDP, from data ingestion and normalization to identification of relevant signals and output to downstream systems, are illustrated in Figure 1.
CUSTOMER DATA MODEL CHALLENGES

A customer data model identifies the factors that a marketer believes to be important in understanding and predicting a customer’s behavior. Without a model, there is no way to systematically segment customers into groups and test the effectiveness of different marketing strategies. A data model captures a variety of data, from unambiguous, clearly defined attributes like name, address, and demographic details to data that can be derived and inferred through interactions and by processing data produced by other systems.

The data model contains attributes that might be created when an event occurs—for example, when a customer’s purchases exceed a certain threshold. Or an attribute can be based on who the customer is and where they live. Segmentation models can be based on a combination of explicitly defined data points, such as purchase history, user-declared values (for example, an expressed preference), externally referenced information (subscription or membership information), and attributes and values that are inferred by comparison to large numbers of customers with similar characteristics. Some techniques find hidden or latent attributes, or create relationships based on numerous subtle data signals.

Data models can be mathematical, rules-based, visual, or based on a list of the relevant factors that are believed to produce certain behavior; for example, all people under the age of 40 will prefer slim-cut jeans and those over the age of 40 will prefer a looser fit. (This is not a valid general statement, but it could form the basis of a hypothesis that could then be tested.)
A data model represents the customer and the collective insights and understanding of their real-world needs.

Creating a data model is a valuable exercise for non-technical specialists because it allows marketers to use language to describe what they know, believe, or can infer about their customer. These characteristics are converted to a structure that the system leverages (either capturing the details or defining rules and algorithms for inferring them). Data models can also inform marketers about the types of metrics they should track, and how well their strategy and specific campaigns are performing relative to a particular customer segment or characteristic.

Customer data is usually collected from a large variety of systems that come from different vendors or, if homegrown, are created by different groups. Therefore, they will have varying formats, architectures, and naming conventions. As a result, customer data models could be inconsistent, which makes it challenging to create a unified model that incorporates meaningful data in actionable formats. For example, one system might define one customer at the individual level and another at the household level. If one system totals all of the purchases for multiple members and another tracks individual purchases, the analysis of sales per customer will produce different results.

Despite variations in the data, the model must contain enough detail and the correct attributes to support advanced functionality such as effectively predicting purchase patterns or recommending an appropriate product that meets the precise needs of the customer. A customer data model is analogous to a content data model (typically called a content model). Customer profiles contain attributes that are used by other systems to improve their outputs. For example, for personalization to work correctly, the model needs to provide signals to customer engagement systems that tell those systems how to differentiate the customer’s experience—what content to present, what products or solutions to offer, and the overall experience that will move them forward in their journey. What is it about the customer that can be captured as metadata (or attributes) in the customer data model and represented in the details of their profile that will drive a unique interaction? It might be the customer’s age, or whether they were active on social media, or whether they had children. The CDP stores data about the customer that can be leveraged by various downstream systems to predict and influence the customer’s behavior.

These signals can come from many sources. Some are based on explicit attributes such as demographics, content preferences, and account information (see Figure 2). Others come from subjective or behavioral attributes (see Figure 3). These might be interest profiles, past purchase behaviors, social media patterns, loyalty scores, and real-time website behaviors.

![Figure 2. Customers can be described with explicit metadata from a variety of source systems.](image)
DATA ANALYTICS

Figure 3. Implicit metadata about a customer is based on judgment and/or derived from other data sources.

ATTRIBUTES, METADATA, AND SIGNALS

Attributes are descriptors that define all the things we know about the customer and describe their characteristics, needs, and behaviors. They determine how systems and tools provide an experience that helps customers meet their objectives and solve their problems. For systems and tools to meet user needs, these attributes have to be represented as metadata in the customer experience applications and in back-end systems that contain transaction and interaction histories.

Metadata supports the machinery of back-end operational and front-end customer-facing technology. Metadata describes the “is-ness” and “about-ness” of a customer. Are they consumers or are they businesses? The metadata field of “customer type” contains the attribute of consumer.

A signal is the data produced by an action of the customer when interacting with an application. The applications can fall into multiple categories of functionality including those that generate explicit, objective, or applied metadata (see Figure 2). This metadata and associated data is typically sourced from customer relationship management systems, enterprise resource planning, e-commerce, sales automation, order management, and external data sources like credit bureaus, data aggregators, and credit card processors (some of which can also fall into the derived metadata category). Examples of this type of metadata include:

- customer type (consumer, business, or nonprofit),
- demographic (age, gender, language, location, income level),
- master customer data (account, name, address, contact phone, email),
- account details (products, service plans, billing, rate plans, credit information),
- relationship details (trouble tickets, call history, account access details),
- content preferences (product updates, technical, communities, topics, offerings), and
- equipment (devices and configurations).

Applications that generate implicit, subjective, or derived metadata (see Figure 3) are sourced by analyzing outputs from diverse customer experiences and operational and social media applications, and by applying conventional analytics and machine-learning algorithms to create
new ways of understanding and describing the customer. Examples of this type of metadata include:

- social graph (LinkedIn, Facebook, Twitter, Instagram),
- marketing applications (website score, email opens, campaign history),
- strategic segmentation (high value, high growth),
- social media (forum discussions, participatory marketing, social conversations),
- loyalty attributes (predicted lifetime value, likelihood to recommend value, length of relationship), and
- behavioral segmentation (purchase and motivational behavior).

The overall information architecture (for customer data models and content models) needs to be aligned so that specific pieces of information can be surfaced to the user depending on the real-time signals from customers’ digital body language. This step requires human understanding—which pieces of information most contribute to a solid model for behavior. While some of these features can be defined in advance, many are based on upstream and downstream system architecture and algorithms.

MINING THE DATASTREAM

What does customer data reveal about what the customer needs? Many different data sources are available. For instance, social media data might contain information about a preference. How is that captured? Clickstream tells us something about how customers are consuming content and traversing the website: whether they click through an offer, whether they respond to a promotion, or whether they are able to complete their purchase. The data tells a story—the question is how to understand that story.

Every tool and technology in the digital engagement ecosystem produces datastreams that need to be interpreted to be acted upon. The challenge lies in identifying what data is important, understanding what it is saying, and determining what to do with it. These questions must be asked over and over again to keep the focus on the purpose of collecting and analyzing data, which might change over time.

Metrics fall into several classes, including search, behavior, utilization, content, and response. Each category of metric can have dozens of details and reports. The goal of understanding these metrics is to drive an action to optimize or improve an outcome. When users browse to a certain point and then leave the site, they were unable to complete their task. What can be changed to impact this behavior? Experiments need to be designed to find the best combination of user outcomes. These are the insights that lead an organization to change its strategies for campaigns and offers. The CDP not only allows ready access to all the metrics associated with customer behavior, but also provides the ability to execute an appropriate response based on the data, sometimes by accessing and triggering actions in external systems.

DERIVING CUSTOMER INSIGHTS

A good sales associate knows both their customer and their products and solutions, and makes recommendations that meet the customer’s needs. In the digital world, organizations need a way to capture and act on the insights that come from a digital representation of this knowledge. This means first interpreting the signals, which arrive in a different form from those a sales associate receives. Marketers and merchandisers have been dealing with the issue of signals for years, but in a different context—the signals were in the form of customer feedback, market intelligence, and sales trends captured over a period of weeks or months and across broad segments of customers rather than at the individual level.

The new pieces of the puzzle are the scale of the challenge, the velocity of commerce, and the number of variables that need to be interpreted as well as the granularity of responsiveness to customer needs. It is now possible to act on detailed understanding of customer needs based on signals from data; however, it is not possible to manually design and create the combinations of
products, services, and solutions that meet each customer’s unique needs. In the physical world, this is what a great salesperson does—they know the customer and offer solutions based on that knowledge. Digital technology is the stand-in for the best salesperson in an organization.

The process begins with a hypothesis about what offers the customer will respond to and what components of that offer can be recombined. It could be as simple as shopping basket analysis, where purchase history combinations are mined and presented to customers exhibiting similar signal patterns. Other sources of variables can be mined from human experts, product engineering documentation, maintenance manuals, and support call chat logs.

Personalizing search results can be based on customer data models that are related to product data and content models mapped through an ontology. The ontology tells the search engine how to interpret a search query based on interests, industry, or other derived profile data. For example, “mold stripping” means one thing to a manufacturing engineer and another thing to a construction manager. In the first case, the task is related to an injection mold maintenance procedure. In the second, it is related to an approach for renovating a flooded building. The ontology can disambiguate a search query by associating the context of the term with the industry or task of the user. This requires a customer data model that captures information about the user’s world at a more granular level.

Marketing organizations need to understand users and their tasks and journeys at a level of specificity that allows them to anticipate needs and determine what offer, product combination, solution, advice, or content will move the customer closer to their objective. Marketers have to ask, “What does it mean to personalize my customer’s experience? How can I differentiate between one type of customer and another, and what does that mean for how they interact with us? What content and information can be served that will be different? Why is it important?”

An analysis might reveal that for a specific segment of customers, a particular combination of products leads to increased conversions under certain circumstances. The hypothesis is that extending the product breadth with new audiences will also lead to increased revenue. To test this, the CDP needs to integrate with downstream systems that orchestrate the user experience. These systems include content management systems, e-commerce applications, product information systems, and the CDP.

Once the insights are gained, the challenge becomes one of converting knowledge into action.

**ACTING ON INSIGHTS FROM CUSTOMER DATA**

The goal of understanding the customer is to take action based on that understanding. How does the organization interpret the interactions, preferences, experiences, and all the signals that stream from every customer relationship? What do you do with the data? This is the “what next” question that has to be continually asked and answered. What do we do when we know our customers’ needs? The obvious answer is that we try to meet those needs as cost-effectively as we can. The question is always, “How do we act on the information?”

Human intervention is critical at this stage. Acting on insights requires a human to interpret the data and recognize a pattern, the ability to test hypotheses about actions that will create the desired behavior, and the machinery to operationalize the confirmed (or fine-tuned) hypothesis. The end result is dynamic functionality derived from a combination of human judgment, expertise, and creativity. However, this outcome cannot be achieved unless the foundational information, data, and content architecture is in place for the digital machinery.

Because data comes from different processes and applications, there are likely varying constraints on the data—from usage and permissions to downstream enrichment. In many cases, owners of data in one application might not be aware of the downstream impact of their decisions about customer data. They also might not be impacted by quality issues that would be
critical to another use case at the downstream orchestration layer. In that case, there would be no incentive to fix the data and no business case to invest in remediation. These complexities make CDPs an enterprise priority that requires executive support, sponsorship, and funding that considers the entire customer data ecosystem.

CONCLUSION

CDPs are increasingly essential to integrated digital marketing programs. Deploying these technologies reveals gaps and challenges throughout the entire enterprise and the digital supply chain serving the customer. If the customer cannot be fully understood from every point of view of the enterprise, it is not possible to serve them optimally. These gaps and challenges can only be remediated with board- and C-level resources and attention. If you are not serving your customers optimally, they will go to your competitor.

Successful use of customer data requires the development of a robust model, judicious selection of data, careful interpretation of analytics, and the ability to act on the results. Each of these steps poses its own challenges. By providing access to data from numerous systems in one database and supporting the systems that can produce an appropriate customer experience, the CDP overcomes the limitations imposed by fragmented point solutions and presents a holistic approach to customer interactions.

ABOUT THE AUTHOR

Seth Earley is CEO of Earley Information Science (www.earley.com). Contact him at seth@earley.com.
Life as a C-level executive is complicated. It’s part political and part financial, part operational and part strategic. It’s quiet and chaotic, fun and tragic. It’s not for the faint of heart: introverts need not apply. C-suiters also have stakeholders. Some are internal, like boards of directors, and some are external, like supply chain partners, vendors, and customers. C-suite life is also tenuous: some executives keep their posts for a short while, and others for decades. Skills, competencies, relationships, financial realities, and experience can all determine how long the lease on the C-suite runs. Many of these are well beyond the control of the suite’s residents.

Digital technology is still too often a black box to members of the C-suite, and digital ignorance at any level will terminate a C-suite lease. Many executives know a little about technology, but unless they work for a technology company, they seldom know a lot. Many need help with the basics such as the Internet and mobile computing, with assessing emerging technologies such as blockchain, and with managing the vast technology infrastructures that power their global operations. Executives need to understand emerging and disruptive technology trends—the new industrial revolution called “digital transformation.”

In light of this digital revolution, this column is being reborn. We want to help C-level executives understand the vast digital world in which they live, and how they should leverage digital technology into their business processes and business models. But we’re not going to lecture our readers: we’re going to communicate the basics and the complexities of digital technology. We’ll forecast the future, interview C-suite residents, identify the technologies most likely to transform the C-suite, and determine the steps necessary to elevate the understanding, communication, and application of emerging and disruptive digital technologies throughout the digital enterprise—which is now every enterprise on the planet. In addition, we’ll identify C-suite best practices for the assessment and deployment of emerging and potentially disruptive digital technologies.

This column will explore five main topics in 2018, each addressed through the lens of C-suite issues and challenges:

- technology trends,
- digital transformation and competitiveness,
- governance and organization of technology,
- data breaches and cybersecurity, and
- digital education and training.

Business is becoming increasingly digital, and C-level executives must therefore become increasingly knowledgeable about technology trends, opportunities, and threats. At the highest level, this column can be summarized as a continuous technology SWOT (strengths, weaknesses, opportunities, and threats) analysis.
In this column, we’ll identify and assess the technologies that will impact business processes and models the most. These include artificial intelligence (AI) and machine learning, blockchain, location-based platforms, the Internet of Things (IoT), virtual/augmented/mixed reality, and augmented analytics. Some of these technologies will be discussed more comprehensively than others, but they will all be assessed according to their potential impact.

AI/machine learning is already impacting every vertical industry. C-level executives should understand how AI and machine learning can transform functional activities like planning, diagnosis, and all flavors of deductive problem solving. They should also understand how they should track and even invest in AI/machine-learning startups: should they generously fund their corporate venture capital teams?

Executives should understand the broad transactional power of blockchain and the growing use of cryptocurrency that blockchain enables.

Location-based services will fuel the complete customer journey—the Holy Grail of retail. They will also enable all forms of mobile computing. C-level executives need to appreciate how pervasive location-based services will become.

The IoT will transform industries and cities. Connectivity will change the way products and services are defined and delivered. Are the connections secure and capable? What happens when 5G is widely deployed? Are C-level executives ready for always-on, fully connected mobile devices?

The experiential impact of alternative realities will change marketing, training, education, travel and gaming, and other areas. How well do C-level executives understand virtual/augmented/mixed reality?

Augmented analytics is where big data analytics meets AI/machine learning. All the location and IoT data will be integrated and converted into predictions and prescriptions that will redefine personal and professional experiences and transactions. Executives must understand the reach of augmented (intelligent) analytics and adjust their business processes and models accordingly.

Life in the C-Suite will identify, describe, and assess a variety of technologies for C-level executives who might not be as versed in the technologies as they perhaps should be.

It would be difficult, if not impossible, to find a Board of Directors that hasn’t asked its executive team about digital transformation. At the same time, not every executive understands digital transformation the same way. Digital transformation seeks to improve or disrupt existing business processes and models with emerging technology. C-level executives understand that digital transformation can impact their competitiveness in their existing markets against their traditional competitors. But it can also threaten whole business models from industry outsiders, like how Airbnb and Uber disrupted the hotel and transportation industries.

Digital transformation can be bottom-up or top-down. It can be enterprise-wide or confined to a business unit or two. It can be funded by the C-suite or by business units or even departments. It can be declared an emergency or just the next competitive step. But there are landmines in the digital transformation process that C-level executives must appreciate before they launch expensive projects. Remember when ERP was all the rage? Executives need to assess the probability of digital transformation project failure. Is it as high as the probability of ERP project failure (which remains frighteningly high)? Most importantly, digital transformation requires sustained, unambiguous leadership from the C-suite. But leadership must be informed to be effective. This column will help.

How should digital technology be organized and governed? C-level executives must confront profound changes in the control of technology, which often begins as a consumer product before entering the enterprise. Technology now enters companies through many different doors. How
can all the doors be monitored? What new technology adoption rules should C-level executives approve? Is technology standardization still a relevant goal? Is centralized control at all realistic in the age of cloud delivery? How many C-level executives (besides the CIO, CISO, and CTO) have even thought about all this? There are questions about the location of emerging and disruptive technology. Funding is also important. Who pays the bills for technology pilots? How shared should governance be in the 21st century? Should the new governance just let the technology flowers bloom? Where is the cost-effective line? Should there even be a line?

Data breaches threaten every C-suiter in the world. The largest data breaches have compromised business models and executive careers. What happens next? There’s no question that the number and severity of data breaches will increase dramatically over time. Most professionals in the field believe there will never be perfectly secure computing environments. If this is true (and all evidence suggests it is), then the best C-level executives can do is brace for the worst and try to control and minimize the damage. Although investments in disaster recovery and business resumption planning are necessary, they’re not sufficient to avoid data breaches. C-level executives must also worry about liability as breaches become more expensive. Is cybersecurity the answer? Executives might believe that a well-funded cybersecurity program and a new CISO will keep them safe. They won’t. What kind of damage control plans and procedures are needed? Will corporate warfare become digital? How can C-level executives win digital wars with local, regional, national, and global competitors? How should companies immediately handle breaches?

Life in the C-Suite will explore these sobering issues where there are no easy approaches or answers. In fact, digital security is a ticking time bomb sitting right below every C-suite in the world. This column will help executives navigate continuously troubled waters and learn what to do when the bombs explode.

Finally, we’ll look at how C-level executives can stay current. We’ll discuss the internal and external education and training programs necessary to keep C-level executives aware of technology trajectories, important use cases, and the pace of strategic investments in emerging technologies. Education and training program content will be suggested.

This column will be about the exploding role technology plays in life and business. It SWOTs five areas on behalf of C-level executives who need as much insight as possible. So, stay tuned for some lively discussions about how digital technology is rearranging the furniture—and the careers—in the C-suite. We’ll begin answering what C-level executives need to know about for each of the five topics in the next issue. We’ll also look at some of the most obvious “headline” challenges (such as net neutrality, “fake news,” and technology mega-mergers) impacting the technology industry and how those challenges will affect specific industry sectors. We’re hoping that this column becomes a must-read for senior executives and all those who create, sell, and apply technology. We’re also hoping it becomes a forum for the discussion of the positive and negative roles digital technology play in our personal and professional worlds.

ABOUT THE AUTHOR

Stephen J. Andriole is the Thomas G. Labrecque Professor of Business Technology in the Villanova School of Business at Villanova University, where he teaches courses in strategic technology and innovation and entrepreneurialism. He conducts applied research on technology management best practices, social media, big data analytics, cloud computing, and emerging (“ready”) technology adoption. Andriole is the recipient of the US Department of Defense Meritorious Civilian Service Award for his work at DARPA, and is a Charter Member of the US Senior Executive Service (SES). Contact him at steve@andriole.com.
Introducing the Student Forum Column

Welcome to the Student Forum, a new column that intends to connect IT students with industry, government, and academia. In this introductory article, we explain the rationale of this new column, discuss its aims, and provide readers with different ways to contribute.

The role of students in IT has become increasingly important over the past 20 years. Today, students are not only our future colleagues in industry and academia, but they also (and perhaps more importantly) constitute the workforce of a very relevant part of the software industry.

STUDENTS IN INDUSTRY AND GOVERNMENT

Traditionally, graduate students devoted their time to completing their PhD thesis. Today, they participate intensively in joint projects with industry and government, who often fund these projects. Previously, graduate students often ended up working in academia after earning their PhD. Today, due to the tremendous growth of the software industry (and major budget cuts in academia), an increasing number of PhD graduates work in industry. They are more aware of technology trends, sometimes because their PhD projects solve concrete industry needs. But they are also active in exploring job opportunities, and many times they work in industry while finishing their thesis.

The most shocking change is in the number of undergraduate students involved with industry. The increasing need for human resources has completely changed how undergraduate students interact with industry, and therefore how they interact with IT professionals. Most large companies (like Google and Microsoft) have initiated internship programs where undergraduate students can experience the corporate world. How much this experience has shaped software companies is being studied, but it’s a fact that such companies try to establish links with students as early as possible.

In many emerging countries, the explosion of the software industry has led to a shortage of professionals, so students are constantly being hired. Undergraduate students often begin working as developers before they’ve finished their computing education. The impact of this situation has not been completely addressed yet. However, some observations are that students take longer to finish their education and that fewer students go into research as their career paths are being defined earlier. These consequences are not necessarily negative, but they need to be assessed.
Technology changes rapidly—there are new kinds of jobs being created in both industry and government. For example, it was reported that 4.4 million IT jobs would be created globally to support big data by 2015.1 Because programming languages evolve rapidly, there’s a lively discussion about new types of careers, such as coding boot camps. This new scenario might affect students positively but it’s still in its infancy.2

Considering the circumstances of the “birth” of some of the key software players (like Google and Facebook), one might argue that the explosion of the software industry is related to students’ involvement in the software industry.

STUDENTS IN RESEARCH AND ACADEMIA

Graduate students have been one of the key drivers in academia. They perform research while pursuing their PhDs, participate in research projects, publish in journals and magazines, and attend and speak at events like doctoral consortiums and student volunteer programs. The impact of students’ participation in scientific conferences is similar to the impact of their participation in industry activities.

Students also help with other typical academic tasks: they teach and mentor younger students, review papers, help prepare research proposals, and get involved in seeking funding.

In a different but related context, years ago the European Commission launched the Erasmus+ Program (ec.europa.eu/programmes/erasmus-plus/node_en), which provides opportunities for more than 4 million Europeans to study, train, gain experience, and volunteer abroad. Though not restricted to IT, most computer and engineering-related organizations have extensively used this program.

STUDENTS AND IEEE

IEEE was early to recognize the importance of students, as seen with IEEE Student Branches and the Computer Society’s Student Chapters (www.computer.org/student-chapters); there are branches and chapters on all continents. These serve not only as the liaison between IEEE and IT students, but they also organize activities ranging from courses and seminars to conferences. They produce educational materials, help students meet with industry leaders, and help them develop relationships for internships and job opportunities.

HOW TO CONTRIBUTE

This column will be a vehicle for IT students to build stronger relationships within industry, government, and academia. IT Pro will aim to serve as the connecting hub for these relationships by publishing articles about subjects in which students are involved. These areas of interest include (but are not restricted to):

- experiences and opinions/details on internship programs (in general or in particular);
- IT university curricula and its impact on how students relate to industry;
- new types of IT careers outside academia;
- the role of IT graduate and undergraduate students in industry and government;
- IT students in academic and scientific programs and activities;
- IEEE Student Chapter activities, experiences, and outcomes;
- specific students’ experiences in their work;
- government and industrial programs related to IT students; and
- IT learning and the impact of distance learning on IT students.

Submit your ideas for articles to gustavo@lifia.info.unlp.edu.ar or mjescalona@us.es. Columns are relatively short—no more than 2,500 words (including figures and tables, which are considered 250 words each). For style guidelines see www.computer.org/cms/Computer.org/Publications/docs/2016CSStyleGuide.pdf.
REFERENCES


ABOUT THE AUTHOR

Gustavo Rossi is a full professor at Universidad Nacional de La Plata and a principal researcher at Conicet. Contact him at gustavo@lifia.info.unlp.edu.ar.
Introducing the IT Economics Department

Welcome to the new IT Economics department, which seeks to advance the understanding of various microeconomic and macroeconomic issues that IT managers need to examine in their decisions to adopt and implement information and communications technology (ICT)-related systems, services, processes, and practices.

Microeconomic forces operate at the level of an adopting unit or organization. These forces influence how likely or unlikely organizations are to adopt a particular system, service, process, or practice. Put differently, they affect how organizations evaluate the costs and benefits of ICT-related systems, services, processes, and practices. For instance, when making a decision to implement an ICT system, IT managers need to take into account costs and benefits associated with ownership of the system, making effective use of the system, learning and switching to the system, and the system’s compatibility with existing systems. Of course, the costs and benefits of implementing an ICT solution will vary across organizations of different sizes and in different industries.

Macro-level forces also influence the adoption of ICT-related systems, services, processes, and practices. These are external forces such as the regulations, rules, and policies enacted by national, regional, and local governments. Thus, the political economy of development and use of ICT-related systems, services, processes, and practices will be one of the key focus areas of this department. Other major macro-level forces include the availability of ICT infrastructures and skills, consumer preferences, and external threats (for example, from malware and cybercriminals). This department will also consider intermediate-level forces such as those related to actions of competitors and pressure from trade association and industry groups as well as from value-delivery networks (for example, supply-chain partners).

A related focus of this department concerns increasing returns and externalities. W.B. Arthur noted that “increasing returns are … mechanisms of positive feedback that operate—within markets, businesses, and industries—to reinforce that which gains success or aggravate that which suffers loss.” This approach would help explain how firms, innovations, industries, and the environment influence one another. The law of increasing returns argues that economies of scale, decreasing costs, and feedback mechanisms lead to further success for already successful entities. Articles published in this department, for instance, can explore the effects of increasing returns in an ICT industry and analyze whether self-reinforcing feedback provided by institutions, industry, and the market might allow an IT system to gain an edge over competing systems.
Many small and medium-sized enterprises (SMEs) don’t allocate sufficient time, resources, and efforts to secure their IT systems. They tend to think that cybersecurity investments involve high costs and low benefits. Surveys have reported that SMEs often believe they have no data of interest to cyberattackers. Unfortunately, this view is often misguided and associated with a lack of true understanding of the evolving nature of cyberthreats.

According to a 2017 study by the insurance company Zurich, 875,000 SMEs in the UK faced at least one cyberattack over the past 12 months, 10 percent of which reported losses over $70,000. Likewise, according to the National Cyber Security Alliance, one in three small businesses in the US become cybercrime victims every year and 60 percent of them will close within six months of experiencing a cyberattack.

Poor Cybersecurity Orientation among SMEs

Despite these problems, increasing the level of preparedness to defend themselves against cyberattacks has not been a major priority for most SMEs. Experian’s annual data breach preparedness study found that 45 percent of SMEs across the UK have no contingency plan in place to deal with a potential data breach. A survey of more than 1,000 SMEs in the UK indicated that about half planned to spend about $1,000 on cybersecurity annually. According to the National Cyber Security Alliance, 87 percent of small businesses in the US had no formal cybersecurity plans.

Many SMEs tend to believe that investments in employee cybersecurity training and awareness activities will have a low return on investment. Surveys and anecdotal evidence indicate that SMEs lack initiatives to provide adequate training and support to enhance their employees’ cybersecurity competence. According to a 2011 survey conducted by Zogby International, 77 percent of US small businesses lacked a formal written cybersecurity policy for employees and 49 percent lacked even an informal policy. The survey also found that 45 percent of small businesses provided no cybersecurity training to employees. Moreover, 56 percent lacked Internet usage policies to clarify websites and web services that employees can use.

Similar findings have emerged from studies of SMEs in other industrialized countries. According to a survey conducted among SMEs in the UK manufacturing sector, 46 percent of manufacturers failed to increase their investment in cybersecurity over the past two years, and one-fifth failed to make their employees aware of cyber risks. Only 56 percent of the respondents said that cybersecurity was given serious attention by their board, and 36 percent had an incident response plan in place.

These findings are of concern and have important implications for SMEs’ cyber-defense capabilities, especially because recent studies have suggested that 40 percent of SMEs experienced a security breach resulting from employee visits to malware-hosted websites.

Poor cybersecurity orientation among SMEs can be mainly attributed to the perceived high costs of cyber-defense measures and the relative newness of cyberthreats. According to the UK government’s Cyber Streetwise campaign (www.cyberstreetwise.com), which aims to change how cybersecurity is viewed, a quarter of UK SMEs reported that cybersecurity is too expensive to implement. The survey also found that a fifth of the respondents did not know “where to start.”
IT ECONOMICS

Changing Cost-Benefit Structure for SMEs

SMEs are facing greater regulatory pressures from governments to strengthen cybersecurity, which would affect the cost-benefit calculus of cybersecurity investments. A notable example is the Department of Financial Services (DFS) in New York, which regulates banks and insurance companies in the state. New guidance issued by the DFS in December 2014 specified stricter rules in corporate governance, login security, management of third-party vendors, cybersecurity insurance, and others. The DFS asked financial sector firms to explain the processes and mechanisms used to track potential vulnerabilities at their third-party vendors and suggested that they develop more cybersecurity expertise on their boards.15 The head of DFS also urged financial companies to invest in cyber insurance.

Regulators are using not only sticks but carrots as well. For example, in 2015, the UK government announced a cybersecurity innovation voucher scheme that offers micro, small, and medium-sized businesses money for specialist advice to strengthen their cybersecurity.16 Some governments are taking measures to strengthen SMEs’ cybersecurity. For instance, in the UK, Cyber Essentials (www.cyberstreetwise.com/cyberessentials) requires an organization to complete a self-assessment questionnaire, and the responses are reviewed independently by an external certifying body. This program was developed as part of the UK’s National Cyber Security Program in close consultation with industry. It’s been reported that most viruses, spyware, or malware found in commonly detected cyberattacks can be prevented if SMEs are Cyber Essentials certified.17 Since October 2014, any UK government tenders are required to hold Cyber Essentials accreditation.18

There’s an increasing tendency among organizations to evaluate the cybersecurity practices of value-delivery networks such as distribution channels and supply-chain partners. The goal is to make sure that supply-chain partners have at least the same cybersecurity standard that companies set for themselves with compliance mandated in contracts.19 In a survey conducted by KPMG among the UK’s procurement managers at large organizations across several sectors, 94 percent of respondents said that suppliers’ cybersecurity standards were important when awarding contracts to SMEs. Seventy percent of the respondents were of the view that SMEs could do more to protect valuable data, and 86 percent noted that SME suppliers that suffer a data breach would be removed. Two-thirds of the responding organizations asked their suppliers to demonstrate accreditations such as Cyber Essentials or the Payment Card Industry Data Security Standard (PCI DSS).

In addition, SMEs are facing more demanding customers that place high importance on business partner companies’ cybersecurity measures. It was reported that about a quarter of medium-sized businesses in the UK had been asked by a current or prospective customer about their cybersecurity measures.8 Likewise, in a 2014 survey conducted among US adults by Princeton Survey Research Associates International, 45 percent of respondents with credit or debit cards indicated that they would “definitely or probably avoid” retailers that experienced a data breach.20

Figure 1. Development of technological capabilities to successfully attack and defend.
Finally, an additional mechanism that might force SMEs to strengthen cybersecurity measures is the rank effect. The idea here is that the deployment of cyber-defense mechanisms tends to diffuse from large to small organizations. As large companies develop stronger defense mechanisms against cyberattacks, SMEs are more likely to become cyberattack targets. This is illustrated in Figure 1, where Firm A is a large firm and Firm B is a small firm.

The asymmetric nature of cyberattacks means that entities with limited financial and technical resources can compromise high-value targets. This means that the attacker is likely to have a dramatically higher success probability compared to the two firms’ probabilities of defending themselves.

An implication of the rank effect is that the level of threats faced by a firm is also a function of cybersecurity measures taken by other firms. In Figure 1, both Firm A and Firm B increase cybersecurity investments over time, but at any point Firm A invests more resources than Firm B. Between time P and time Q, while Firm A’s cybersecurity investment might be sufficient to defend itself from the attacker, Firm B might not be able to do so.

CONCLUSION

It’s critical for SMEs to consider the rapidly changing nature of the cost-benefit structure associated with strengthening cybersecurity. SMEs must proactively track new cyberthreats and develop formal and informal policies to deal with them. They must provide information, guidance, education, and training to employees about cybersecurity. Overall, SMEs need to do more on the cybersecurity front than just complying with the demands of external stakeholders such as regulators, certification agencies, supply-chain partners, and customers.

I would love to consider your contribution in this department along the above-mentioned lines. The requirements for IT Pro columns in terms of length, format, writing style, and other criteria have been discussed in this issue (see “Introducing the Internet of Things Department”). Please feel free to contact me with your ideas, thoughts, and questions.

REFERENCES


ABOUT THE AUTHOR

Nir Kshetri is a professor of management at the Bryan School of Business and Economics at the University of North Carolina at Greensboro and editor of IT Pro’s IT Economics department. Contact him at nbkshetr@uncg.edu.
PURPOSE: The IEEE Computer Society is the world's largest association of computing professionals and is the leading provider of technical information in the field.

MEMBERSHIP: Members receive the monthly magazine Computer, discounts, and opportunities to serve (all activities are led by volunteer members). Membership is open to all IEEE members, affiliate society members, and others interested in the computer field.

COMPUTER SOCIETY WEBSITE: www.computer.org

OMBUDSMAN: Direct unresolved complaints to ombudsman@computer.org.

CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

AVAILABLE INFORMATION: To check membership status, report an address change, or obtain more information on any of the following, email Customer Service at help@computer.org or call +1 714 821 8380 (international) or our toll-free number, +1 800 272 6657 (US):

- Membership applications
- Publications catalog
- Draft standards and order forms
- Technical committee list
- Technical committee application
- Chapter start-up procedures
- Student scholarship information
- Volunteer leaders/staff directory
- IEEE senior member grade application (requires 10 years practice and significant performance in five of those 10)

PUBLICATIONS AND ACTIVITIES

Computer: The flagship publication of the IEEE Computer Society, Computer, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

Periodicals: The society publishes 13 magazines, 19 transactions, and one letters. Refer to membership application or request information as noted above.

Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.

Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.

Technical Committees: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

Conferences/Education: The society holds about 200 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The society offers two software developer credentials. For more information, visit www.computer.org/certification.

NEXT BOARD MEETING

1-2 February 2018, Anaheim, CA, USA

EXECUTIVE COMMITTEE

President: Hiroonori Kasahara
President-Elect: Cecilia Metra; Past President: Jean-Luc Gaudiot; First VP.
Publication: Gregory T. Byrd; Second VP, Secretary: Dennis J. Frailey, VP;
Member & Geographic Activities: Forrest Shull; VP, Professional &
Educational Activities: Andy Chen; VP, Standards Activities: Jon Rosdahl;
VP, Technical & Conference Activities: Hausi Muller; 2018-2019 IEEE
Division V Director: John Walz; 2017-2018 IEEE Division VIII Director:
Dejan Milojicic; 2018 IEEE Division VIII Director-Elect: Elizabeth L. Burd

BOARD OF GOVERNORS

Term Expiring 2018: Ann DeMarle, Sven Dietrich, Fred Dougis, Vladimir Getov, Bruce M. McMillin, Kunio Uchiyama, Stefano Zanero
Term Expiring 2019: Saurabh Bagchi, Leila DeFloriani, David S. Ebert, Jill I. Gostin, William Gropp, Sumi Helal, Avi Mendelson
Term Expiring 2020: Andy Chen, John D. Johnson, Sy-Yen Kuo, David Lomet, Dimitrios Serpanos, Forrest Shull, Hayato Yamana

EXECUTIVE STAFF

Executive Director: Angela R. Burgess
Director, Governance & Associate Executive Director: Anne Marie Kelly
Director, Finance & Accounting: Sunny Hwang
Director, Information Technology & Services: Sumit Kacker
Director, Membership Development: Eric Berkowitz
Director, Products & Services: Evan M. Butterfield

COMPUTER SOCIETY OFFICES

Phone: +1 202 371 0101 • Fax: +1 202 728 9614
Email: bq.ofc@computer.org
Los Alamitos: 10662 Los Vaqueros Circle, Los Alamitos, CA 90720 Phone: +1 714 821 8380
Email: help@computer.org

MEMBERSHIP & PUBLICATION ORDERS

Phone: +1 800 272 6657 • Fax: +1 714 821 4641 • Email: help@computer.org
Asia/Pacific: Watanabe Building, 1-4-2 Minami-Aoyama, Minato-ku, Tokyo 107-0062, Japan
Phone: +81 3 3408 3118 • Fax: +81 3 3408 3553
Email: tokyo.ofc@computer.org

IEEE BOARD OF DIRECTORS

President & CEO: James Jefferies
President-Elect: Jose M.F. Moura
Past President: Karen Bartleson
Secretary: William P. Walsh
Treasurer: Joseph V. Lilke
Director & President, IEEE-USA: Sandra “Candy” Robinson
Director & President, Standards Association: Forrest D. Wright
Director & VP, Educational Activities: Witold M. Kinser
Director & VP, Membership and Geographic Activities: Martin Bastiaans
Director & VP, Publication Services and Products: Samir El-Ghazaly
Director & VP, Technical Activities: Susan “Kathy” Land
Director & Delegate Division V: John W. Walz
Director & Delegate Division VIII: Dejan Milojicic
COMPSAC is the IEEE Computer Society Signature Conference on Computers, Software and Applications. It is a major international forum for academia, industry, and government to discuss research results and advancements, emerging challenges, and future trends in computer and software technologies and applications. The theme of COMPSAC 2018 is Staying Smarter in a Smartening World.

Computer technologies are producing profound changes in society. Emerging developments in areas such as Deep Learning, supported by increasingly powerful and increasingly miniaturized hardware, are beginning to be deployed in architectures, systems, and applications that are redefining the relationships between humans and technology. As this happens, humans are relinquishing their roles as masters of technology to partnerships wherein autonomous, computer-driven devices become our assistants. What are the technologies enabling these changes? How far can these partnerships go? What will be our future as we deploy more and more “things” on the Internet of Things - to create smart cities, smart vehicles, smart hospitals, smart homes, smart clothes, etc.? Will humans simply become IoT devices in these scenarios and if so, what will be the social, cultural, and economic challenges arising from these developments? What are the technical challenges to making this all happen - for example, in terms of technologies such as Big Data, Cloud, Fog, Edge Computing, mobile computing, and pervasive computing in general? What will be the role of the ‘user’ as the 21st Century moves along?

COMPSAC 2018 is organized as a tightly integrated union of symposia, each of which will focus on technical aspects related to the “smart” theme of the conference. The technical program will include keynote addresses, research papers, industrial case studies, fast abstracts, a doctoral symposium, poster sessions, and workshops and tutorials on emerging and important topics related to the conference theme. A highlight of the conference will be plenary and specialized panels that will address the technical challenges facing technologists who are developing and deploying these smart systems and applications. Panels will also address cultural and societal challenges for a society whose members must continue to learn to live, work, and play in the environments the technologies produce. Authors are invited to submit original, unpublished research work, as well as industrial practice reports. Simultaneous submission to other publication venues is not permitted. All submissions must adhere to IEEE Publishing Policies, and all will be vetted through the IEEE CrossCheck Portal.

Standing Committee Chair: Sorel Reisman, California State University, USA
Steering Committee Chair: Sheikh Iqbal Ahamed, Marquette University, USA
General Chairs: Shinichi Honiden (NII, Japan)
Roger U. Fuji, Fuji Systems, 2016 IEEE Computer Society President
Program Chairs in Chief:
Jiannong Cao (Hong Kong Polytechnic University, Hong Kong)
Stelvio Cimato (University of Milan, Italy)
Yasuq Okabe (Kyoto University, Japan)
Sahra Sedighsarvestani (Missouri University of Science & Technology, USA)
Workshop Chairs: Kenichi Yoshida (University of Tsukuba, Japan)
Ji-Jiang Yang (Tsinghua University, China)
Hong Yu Leong (Hong Kong Polytechnic University, Hong Kong)
Chung Hong Lung (Carleton University, Canada)
Local Chair: Hironori Washizaki (Waseda University, Japan)

Important Dates
Workshop proposals
Due date: 15 October 2017
Notification: 15 November 2017
Main Conference papers
Due date: 15 January 2018
Notification: 31 March 2018
Workshop papers
Due date: 10 April 2018
Notification: 1 May 2018
Camera Ready and Registration
Due date: May 15, 2018
Now there’s even more to love about your membership...

Read all your IEEE Computer Society magazines and journals your way on myCS, the digital magazine portal from IEEE Computer Society.

Finally…go beyond static, hard-to-read PDFs. Our go-to portal makes it easy to access and customize your favorite technical publications like Computer, IEEE Software, IEEE Security & Privacy, and more. Get started today for state-of-the-art industry news and a fully adaptive experience.

LEARN MORE AT: mycs.computer.org