Journal of Digital Imaging

Toward Clinically Relevant Standardization of Image Quality

Ehsan Samei, Ph.D.,[1] Alan Rowberg, M.D.,[2] Ellie Avraham,[3] and Craig Cornelius,[3]

In recent years, notable progress has been made on standardization of medical image presentations in the definition and implementation of the Digital Imaging and Communications in Medicine (DICOM) Grayscale Standard Display Function (GSDF). In parallel, the American Association of Physicists in Medicine (AAPM) Task Group 18 has provided much needed guidelines and tools for visual and quantitative assessment of medical display quality. In spite of these advances, however, there are still notable gaps in the effectiveness of DICOM GSDF to assure consistent and high-quality display of medical images. In additions the degree of correlation between display technical data and diagnostic usability and performance of displays remains unclear. This article proposes three specific steps that DICOM, AAPM, and ACR may collectively take to bridge the gap between technical performance and clinical use: (1) DICOM does not provide means and acceptance criteria to evaluate the conformance of a display device to GSDF or to address other image quality characteristics. DICOM can expand beyond luminance response, extending the measurable, quantifiable elements of TG18 such as reflection and resolution. (2) In a large picture archiving and communication system (PACS) installation, it is critical to continually track the appropriate use and performance of multiple display devices. DICOM may help with this task by adding a Device Service Class to the standard to provide for communication and control of image quality parameters between applications and devices, (3) The question of clinical significance of image quality metrics has rarely been addressed by prior efforts. In cooperation with AAPM, the American College of Radiology (ACR), and the Society for Computer Applications in Radiology (SCAR), DICOM may help to initiate research that will determine the clinical consequence of variations in image quality metrics (eg, GSDF conformance) and to define what constitutes image quality from a diagnostic perspective. Implementation of these three initiatives may further the reach and impact of DICOM toward quality medicine.

KEY WORDS: Display quality, display performance, display calibration, DICOM, AAPM, luminance response, image quality

In recent years, significant progress has been made on standardization of medical image presentation in two areas. First, the Digital Imaging and Communications in Medicine (DICOM) Grayscale Standard Display Function (GSDF) (see Fig 1) has proven to be an effective means of achieving visual consistency of medical images on a variety of hardcopy and softcopy display systems.[1] In the last four years, a cross-vendor testing of GSDF implementation at the Radiological Society of North America–Healthcare Information and Management Systems Society (RSNA–HIMSS) "Integrating the Healthcare Enterprise" (IHE) initiative has shown that consistent image quality performance is possible through the DICOM GSDF.[2] However, the DICOM grayscale consistency solution resolves viewing consistency only from a luminance and grayscale perspective.

In parallel with the DICOM effort, the American Association of Physicists in Medicine (AAPM) Task Group 18 has provided much needed detailed guidelines and tools for easy and

[1]*From the Department of Radiology, Physics, and Biomedical Engineering, Duke University, Durham, North Carolina, 3302, USA.*

[2]*From the Department of Radiology, University of Washington, Seattle, Washington, USA.*

[3]*From the Eastman Kodak Company, Rochester, New York, USA.*

Originally printed at the DICOM 20th Anniversary Conference, Baltimore, MD, Sept. 2003.

Correspondence to: Ehsan Samei, Ph.D., tel: 919-684-7852; fax: 919-684-7122; e-mail: samei@duke.edu

Online publication 25 November 2004
doi: 10.1007/s10278-004-1031-5

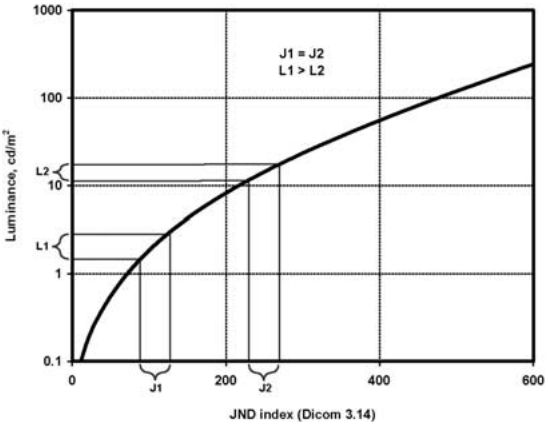

Fig 1. The Barten curve of DICOM Part 14 defines the Grayscale Standard Display Function (GSDF). Equal perceptual steps translate to different luminance ranges as the absolute light level varies.

reproducible measurement of display hardware characteristics in both qualitative and quantitative terms.[3] The AAPM initiative includes all key aspects of display performance, including resolution, noise, glare, reflection, geometrical distortion, and chromaticity, as well as luminance. In addition to testing methods, the AAPM recommends specific minimum acceptance criteria for display performance. However, because the AAPM is not a standardization body, the offered guidelines are only professional recommendations and are not binding as "standards."

In spite of current DICOM and AAPM efforts, much potential for improved image quality has not been realized in clinical settings. The DICOM standard has been limited to grayscale consistency; other display characteristics have not been standardized or implemented clinically, and it is still unclear how display technical data relate directly to clinical and diagnostic usability and user performance. This article proposes three specific ways to enrich the methods and solutions for improving the display image quality and consistency beyond the luminance and grayscale consistency currently afforded by DICOM GSDF, taking into consideration other key elements of display quality, their clinical significance, and their clinical implementation.

PROPOSALS

In this work, we propose three specific collaborative steps that DICOM and the American Association of Physicists in Medicine (AAPM) can take in cooperation with the American College of Radiology (ACR) and the Society for Computer Application in Radiology (SCAR) to bridge the gap between technical performance and clinical use issues. All three proposals relate to image quality. It is recognized that both acquisition and output devices contribute to the ultimate perceived quality of an image. Even though most of our proposals can be implemented taking into account the image quality attributes of the acquisition devices and other possible elements in the imaging chain, this article primarily focuses on softcopy display and hardcopy presentation of medical images.

DICOM Beyond Luminance Response

Display quality is one of the main factors that influence the quality of softcopy viewing of medical images. Display quality is comprised of multiple factors, including luminance response, reflection—specular and diffuse—spatial resolution, noise, geometrical distortions, display chromaticity, veiling glare, and temporal response (eg., temporal noise-and flicker.) The DICOM standard currently covers only the luminance response aspect of display quality through the DICOM GSDF (Barten model, defined in DICOM PS 3.14). This specifies a standard function for mapping the display device dynamic range in a "perceptually linear" manner to Human Visual Contrast Sensitivity such that images look similar on different display devices (Fig 1). While GSDF provides a reasonable approximation for luminance consistency across multiple display devices, it does not provide acceptance criteria and procedures to evaluate the conformance of a display device to the standard, nor does it address other important display quality characteristics, the evaluation of which has recently been defined by the AAPM.

We propose the addition of a Display Image Quality (DIQ) initiative to DICOM, extending beyond luminance consistency afforded by GSDF, adding measurable and quantifiable elements of the AAPM display performance procedures. The new DIQ initiative will include testing methodologies as well as defined limits

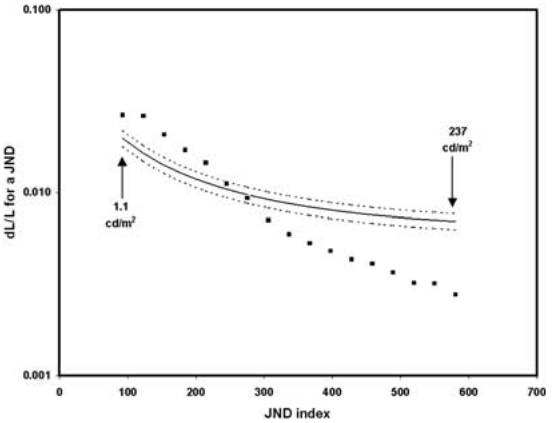

Fig 2. The contrast response according to the Barten curve of DICOM Part 14 (solid curve) compared to that of the contrast response of a display device not calibrated to the DICOM standard. Dashed lines specify the ±10% deviation from the Barten curve.

for acceptable clinical and diagnostic performance (as recommended by the AAPM guidelines). The extension may also include visual performance quantification using simple test images, specific observer protocols, and relative acceptance indicators. The DIQ will form the basis of quantitative and visual-based display quality control protocols.

Softcopy Display Factors

The DIQ will include a number of key display factors listed above. The details associated with a few of these factors are outlined below. In the paragraphs that follows.

The luminance response of a display is well described in the DICOM PS 3.14. DIQ may provide specific guidelines for quantifying the extent of the compliance based on the percent deviation from the contrast dictated by the GSDF curve, as defined in the AAPM TG18 document (Fig 2). Quantitative minimum criteria in terms of minimum and maximum luminance may also be provided. In addition, the TG18-CT and the TG18-MP test patterns (Fig 3) may be considered for the visual evaluation of the luminance response.

Reflection of ambient room light from the display device surface directly impacts display performance. Therefore it is important to evaluate the reflection characteristics of the device

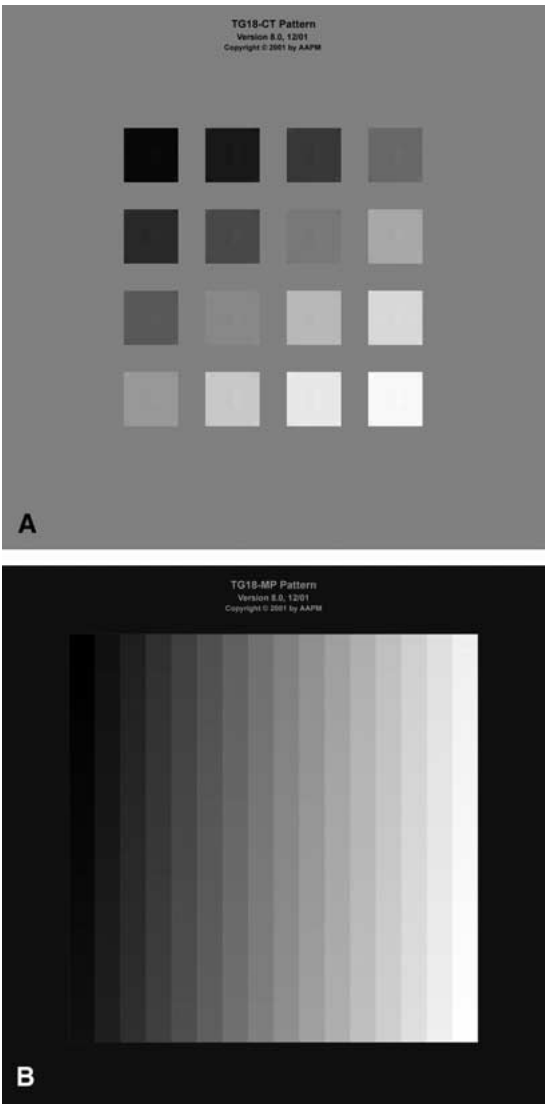

Fig 3. The test patterns TG18-CT (a) for the visual evaluation of contrast/luminance response and TG18-MP (b) for the evaluation of bit-depth/continuous grayscale performance of display devices.

and the maximum ambient illumination that can be used in the reading area without significantly affecting the display presentation. The reflection exists in two types: specular (which creates a virtual mirror image) and diffuse (which creates a haze or matte image view). The display reflection can also be visually and quantitatively evaluated for both forms, providing the maximum allowable room lighting for a display device at a specific minimum

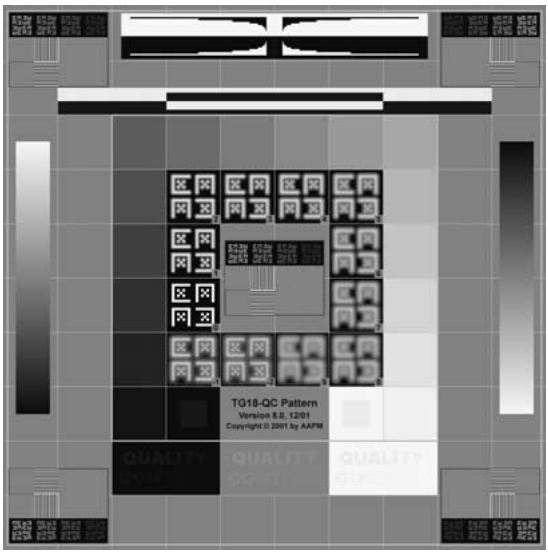

Fig 4. The comprehensive TG18-QC test pattern for the evaluation of key display characteristics including resolution, luminance, and geometrical distortions.

luminance level based on specular and diffuse reflections.

The spatial resolution is the quantitative measure of the ability of the display device to produce separable images of different points, best quantified in terms of the Modulation Transfer Function (MTF). Although specialized equipment is required for calculation of the MTF, display resolution can be visually evaluated by assessing the "CX" patterns in the TG18-QC (or TG18-CX) test pattern (Fig 4). More quantitative measurements are possible using a digital camera equipped with a macro lens in conjunction with single line test patterns.

The geometric distortion of a display system can be visually ascertained or quantitatively measured using the TG18-QC test pattern (Fig 4). Note that the TG18-QC test pattern incorporates multiple test elements enabling the evaluation of many display factors. The TG18-QC test pattern is expected to replace an earlier comprehensive test pattern developed by The Society of Motion Picture and Television Engineers.[4]

Hardcopy Display Factors

Several factors affect the hardcopy image quality when printed on transmissive media for viewing on light boxes or reflective media for direct viewing. These include the intrinsic quality of the media (eg, freedom from visible coating variations), quality of the printer itself (eg, freedom from visible variations in exposure or processing), fidelity of the entire hardware/software chain in interpreting DICOM parameters to yield the desired grayscale densities, printed density versus pixel value conformance with the GSDF, the theoretically available total number of just noticeable differences (JNDs) for a given ambient light level considering the printable density range, the number of true monotonically increasing gray levels, the printable matrix size, absence of artifact, fidelity in terms of spatial frequency response, and quality of viewing conditions. These image quality factors can be methodically quantified using appropriate test patterns and measurable tools, such as a film digitizer, a densitometer, and image processing software.

Color Extensions

Although the DICOM GSDF explicitly addresses only grayscale images, it is important to note that GSDF is not limited to monochrome display devices. Indeed, standardizing the tonescale for presentation of grayscale images on color displays is not only feasible, but it is also highly desirable, especially for review and capture console displays. This standardization can and should be extended to the luminance dimension of nominally color images.

Color consistency is a second area for standardization because current DICOM does not define any way to characterize and standardize the calibration of color display and print devices, nor is there a DICOM standard for consistent presentation of colors. It is noted that DICOM Working Group 11 currently is exploring how color can be standardized. It is expected that this is most important for consistent presentation of DICOM's Visible Light Image Information Objects, including endoscopic, microscopic, and photographic images.

Image Quality Communication Protocol

With the increasing regulation of all of health care, imaging is under pressure to do even more

quality management, and to document the results of quality programs. Image quality and its optimization is a combination of many elements, making troubleshooting image quality problems a complex task. It is necessary to continually track the calibration and other image quality performance factors of output devices, including modality consoles as well as PACS softcopy and hardcopy output. While display and print devices are increasingly calibrated automatically, quality control documentation and record keeping is still largely a manual process. At this time, however, DICOM offers little help with these technical and management tasks.

A New DICOM Service Class

We propose a new DICOM Display Performance Service Class (DPSC) standard for communication and control of image quality information factors between applications and display devices. Services would include query of display capabilities, hardware attributes, and calibration values, as well as setting configurable display attributes. Such information may be used for device management and quality control (QC), to determine fitness for use, or to compute what, if any, image processing is required to achieve standard results. Through this initiative, the current DICOM Printer Configuration Retrieval Service Object Pair (SOP) Class may also be extended to include additional image quality information factors such as spatial metrics. Coupling display systems more closely with other operational components of PACS via DPSC will improve both the efficiency and the effectiveness of image quality management.

Through DPSC, a Service Class User (DPSCU) would query the display/print capabilities of a Display Performance Service Class Provider (DPSCP). Some configurable parameters of displays could be set by a DPSCU, such as the luminance range, the display function to be used, and other parameters (Fig 5). Queryable information would include the number of displays, the matrix size and physical size of each display, the display function installed, the number of grayscale and color levels available, minimum and maximum luminance values, time

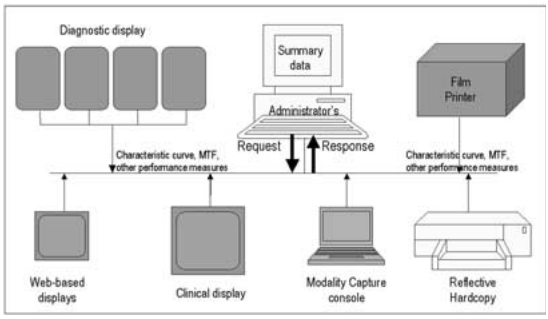

Fig 5. Administrative application requesting and receiving image quality performance data in the role of Display Performance Service Class User (DPSCU).

of last calibration, ambient light conditions, and, if available, the display system's characteristic luminance curve. Additional information should include spatial characteristics such as MTF, and luminance uniformity, including results automatically measured and others evaluated by a human observer, for example the score of the TG18-CX pattern or artifact evaluation. The work proposed for the DICOM standard would include defining appropriate data elements to be retrieved and set by applications, setting up new query rules, and identifying the behavior of appropriate SCUs and SCPs for this new service.

Use Cases for Display Performance Class

1. **Standardized output.** An application acting as a DPSCU requests the luminance characteristic curve from its workstation's display system. The DPSCU determines if the display is standardized, and, if needed, computes an internal image grayscale correction, effectively giving GSDF standardized output (Fig 6). Similarly, a printing application would retrieve the matrix size and MTF of a film printer to determine what type of magnification, if any, should be applied to the image to give the least amount of artifact.
2. **Image Quality Management for Display and Print Devices.** A centralized DPSCU application queries all display devices daily for their last calibration date, luminance characteristics, ambient light settings, and other image

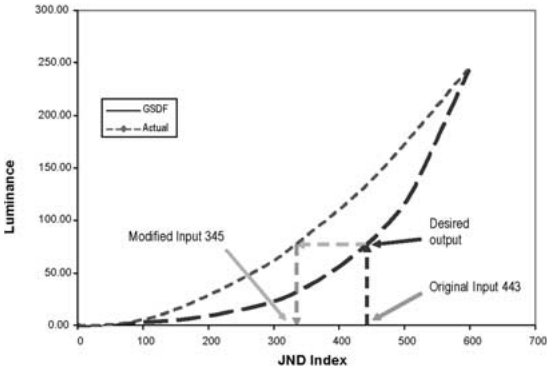

Fig 6. Correcting an actual luminance curve to simulate a GSDF standardized output. The units of the x-axis are Just Noticeable Difference (JND) indices.

quality performance data (Fig 5). The application records the results in a central database, automatically creating lists of displays and printers that require special attention or routine maintenance. Other applications use the database to periodically produce regulatory and management reports. The PACS administrator regularly reviews stability and lifetime statistics of the displays to plan schedules and budgets for replacing and upgrading displays.

3. **Consultation.** The DPSC will help physicians improve their consultations when images are involved. For example, a referring physician calls the radiologist because he cannot identify the lesion on the review workstation. The radiologist displays the same image on the diagnostic workstation, where the lesion is visible. Then the radiologist retrieves the device profile of the review workstation, notes its contrast characteristics, and temporarily transfers this profile to the diagnostic workstation, observing that the lesion is no longer visible. The radiologist then adjusts the window width and window center to make the lesion more visible, and advises the referring physician to use this new setting, and to turn out the room lights.
4. **Standardization and Quality Control of Capture Consoles.** The technologist operating a digital X-ray acquisition device observes the image just acquired to ensure its diagnostic content. Because the acquisition's console has been standardized, as verified by automatically gathered image quality performance data, the technologist releases the image to the hospital's PACS with confidence that the console accurately reflects how it will be later seen by both radiologists and attending physicians.

Clinical Significance of Image Quality

Both clinical and regulatory arenas are placing increasing demands on image quality, and there is a growing need to measure and document image quality at all steps from acquisition through display, printing, and viewing. This process must include not only physical measurements but also performance measures that include psycho-visual, environmental, and system considerations. In addition to their use in a quality management program, such image quality measures must be available to the users of images throughout the image interpretation process so the professionals who use such images can understand the factors that influence and limit their ultimate presentation.

The question of clinical significance of image quality metrics has rarely been addressed. The focus of most prior efforts dealing with display quality has been the physical performance aspects of display devices, assuming that the physical measures would "somehow" translate to diagnostic quality, a foundational assumption that has not been rigorously tested by clinical trails. The relationship between display physical metrics and diagnostic performance should be researched to determine the clinical consequences of variations in image quality metrics (eg, GSDF conformance) on specific diagnostic tasks (Fig 7). In cooperation with the AAPM, ACR, and SCAR, DICOM may be able to gather the necessary information on what constitutes image quality from a diagnostic perspective. The limits identified might form the basis for a joint standard based on standardized test patterns, procedures, and clinical use cases.

The proposed activity may be directed by an initial inter-society committee formed to design a specific research project with representatives from all involved professional organizations.

The tentative elements of such a project might be as follows: In cooperation with ACR and

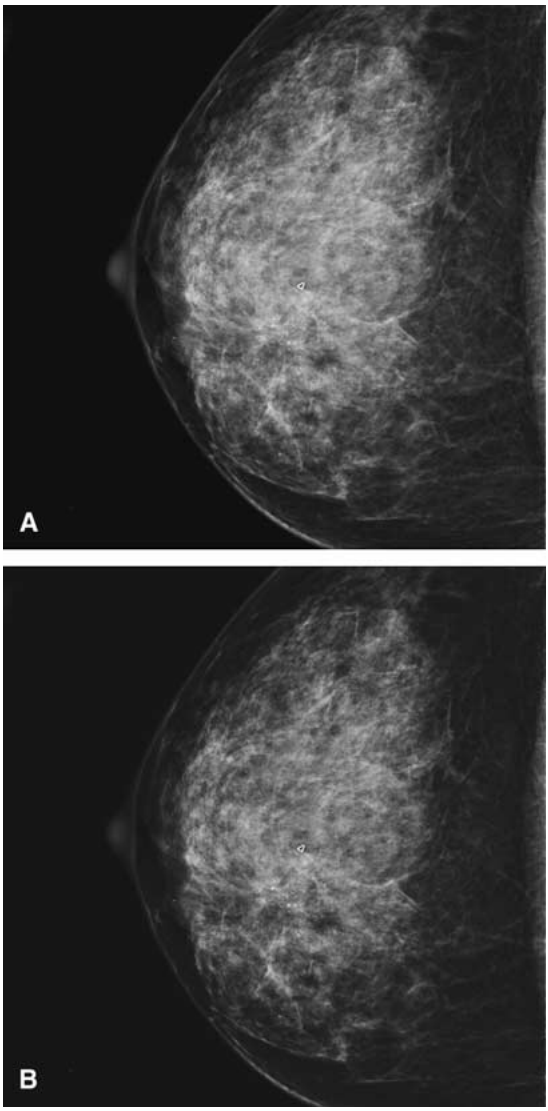

Fig 7. The impact of variations in the luminance response of display devices on the appearance of a mammogram (TG18-MM1 test pattern). Figure (a) is rendered with DICOM GSDF, while figure (b) is not (image' quality may be suboptimal due to print rendering). The clinical impact of these types of image quality variations on diagnostic accuracy for specific clinical tasks has not been established. The proposal outlined provides a strategy toward that goal.

SCAR, specific de-identified digital images from three radiographic modalities (eg, chest, musculoskeletal, and mammography) will be gathered from multiple medical centers. The images will be expected to be within an acceptable range of quality in terms of noise, resolution, and contrast presentation. Both normal and abnormal cases will be collected. The abnormal chest images will contain subtle nodules, the musculoskeletal images subtle fractures, and mammograms subtle masses and microcalcifications. All images will be stored at a central database.

Multiple versions of each image will be presented, each with altered grayscale, resolution, noise, and/or contrast reduction characteristics (reminiscent of display glare and reflection) of the image to simulate deviations from a standardized presentation. Viewing and scoring the images will be done via a graphical user interface. At upcoming RSNA and SCAR meetings, radiologists may be invited to take part in an observer performance experiment in viewing areas set up specifically for this purpose. The images will be viewed and scored by the observers on CRT and LCD display devices that closely conform to the DICOM GSDF and the AAPM guidelines. The results will be analyzed by ROC methodologies. The findings will be used to establish the clinical impact of specific image quality variations in terms of common diagnostic tasks in these three demanding radiographic modalities. The results will be reported in refereed publications, and may subsequently be reflected in upcoming DICOM standards and ACR practice guidelines.

CONCLUSION

In spite of unprecedented success, there are still notable gaps in the effectiveness of DICOM GSDF to assure consistent and high-quality display of medical images. The implementation of the steps suggested may further the reach and impact of DICOM toward quality medicine. We expect that the proposed additional DICOM elements and suggested activities to DICOM proposed have will unleash the creativity of vendors, application developers, and users, providing substantial benefits for PACS users, administrators, and, ultimately, the patients.

REFERENCES

1. NEMA-PS3.14: Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function. Rosslyn, VA: National Electrical Manufacturers Association (NEMA), 2000

2. Consistent Presentation of Images (CPI)–IHE Radiology Technical Framework, Vol I and Vol II, Revision 5.5, HIMSS/RSNA, 2003 (http://www.rsna.org/IHE/tf/ihe_tf_index.shtml)

3. Samei E, Badano A, Chakraborty DP, et al. Assessment of Display Performance for Medical Imaging Systems. American Association of Physicists in Medicine (AAPM) Task Group 18, 2004 (http://deckard.duhs.duke.edu/tg18)

4. SMPTE-RP133: Specifications for Medical Diagnostic Imaging Test Pattern for Television Monitors and Hardcopy Recording Cameras. White Plains, NY: Society of Motion Picture & Television Engineers (SMPTE), 1991