https://www.ingentaconnect.com/contentone/ist/ac/2011/00002011/00000001/art00022

Accurate Color? A Preliminary Investigation into the Color Gamut of Selected Special Collection Library Objects

F. Barry Wheeler; Library of Congress; Washington, DC/USA
Michael J. Bennett; University of Connecticut Libraries; Storrs, CT/USA

Abstract

As cultural institutions continue to digitize their collections' objects, millions of images now exist in TIFF, JPEG, and JPEG 2000 color still image formats. Commonly, however, the colors of the original objects are not accurately reproduced when such digital image files are rendered on a computer monitor or on a print [1]. Such renderings are acceptable to a degree based upon user intent, but a direct comparison of the original and the image rendering will show easily visible differences upon closer inspection. The chain from the object itself, to the digital imaging device through color encoding, storage, color decoding, and finally to presentation is good but clearly needs improvement if archival storage of color critical materials is desired. This paper presents an examination of the first step in this chain and provides a more precise indication of what colors may be present in documents held by cultural institutions.

A number of initiatives are currently underway to improved color accuracy of scanned images. Best practices workflows now are guided by physical color targets so that captured image files can be evaluated against known color standards. In addition, multispectral techniques are being studied. A CIE/ISO Standards Archival Color Committee is currently investigating ways of encoding and storing color data in image archives [2]. However, to date there have been few investigations that provide measurements of colors taken directly from original objects in cultural heritage collections.

This study provides such measurements of sample colors taken directly from the surface of a wide selection of objects from the Library of Congress' collections – maps, prints, photographs, books, rare books, sheet music, and manuscripts. The Library's collections are so vast and complex that no attempt was made to select a statistically valid representative document set - but the selection was sufficiently large and varied to give an indication of the range of colors. Following a procedure used by Williams and Burns [3], direct spectrodensitometer measurements were made of a set of colors from each selected object. Per object these sets of colors were imported into CHROMiX ColorThink Pro software through which specific colors can be displayed in graphical form. This allowed for the comparative overlay of the color gamut for a variety of device specific color spaces – such as the sRGB gamut commonly available on computer monitors – to provide information about the ambiguities involved in capturing, storing, and rendering the color values accurately.

Colorimetric measurements are the focus – a given object's colors were characterized using the device independent CIELAB color space. Analysis was then done to characterize the colors currently being stored in common RGB TIFF files and to provide guidelines for selecting appropriate color spaces when digitizing different kinds of objects. Comparisons of the document colors and the colors commonly used to build ICC color profiles are shown. These comparisons indicate that using a broader set of profiling colors – particularly colors similar to those in the original documents – may provide more accurate color in digital images.

Background

The Library of Congress and its contractors create hundreds of thousands of digital images each year. The Library's collections are incredibly varied –books, maps, manuscripts, musical scores, prints, photographs, and visual materials are comprised of a huge variety of paper, film, and colorants. A wide range of equipment, workflows, and processing techniques are used to accomplish such reformatting, and maintaining consistently high quality is difficult.

One way of improving and maintaining the quality of digital images is to plan the digital capture processes based on the actual colors present in the original documents. Significant research on the spectral response of many high-value documents has been done, but such technical data has not yet been incorporated into most routine operational imaging workflows. The Library calibrates its scanners using a chart of sample colors – usually using one of the Macbeth color charts. But currently there is no broad database of the actual colors present for a wide range of cultural heritage materials that can guide the Library's operations. In turn, there is currently no way to know how well such sample colors relate to the actual colors in the Library's documents. Thus the purpose of this paper is to undertake a survey of the colors present in a diverse set of documents held in the Library's collections and to analyze how closely the document colors relate to the sets of sample target colors that are used to create the Library's device profiles.

Color Issues

Color Space Choice

Digitization requirements may specify a color space – or may not. If a space is required, usually one of three spaces are specified – sRGB has traditionally been used as the common denominator for many CRT and LCD display devices and is generally the default space for most scanners and cameras. AdobeRGB(1998) has been specified as a broader space tailored toward printed output and may be specified to insure that colors acquired from the imaging device are not lost or unnecessarily adjusted. Occasionally, ProPhotoRGB may be specified as a broad gamut that essentially contains all the colors in the visible spectrum. In many instances AdobeRGB(1998) and ProPhotoRGB are selected under the assumption that the digitized master images may safely be converted to a narrower gamut for any derivative images produced with minimal color fidelity loss.

When a color space is not specified, sRGB is generally assumed. Baseline TIFF v6.0 images do not have a color space tag – many scanners only use the "PhotometricInterpretation" tag (tag 264) to indicate RGB. Cameras and some scanners may formally specify the sRGB color space through the 40961 tag in the Exif block. Any of the three specified color spaces are best defined by the inclusion of a device specific ICC profile in the 34675 tag.

A difficulty of using one of the larger color spaces is the need for increased bit depth in order to avoid posterization or banding. Large color space images are best specified in 16-bit per channel to provide sufficient color values so that similar colors are not lumped together under one value. This requirement inflates filesize significantly. The larger color space and bit depth also may make file processing more difficult. Current monitors receive color data in 8-bit per channel format and only the most expensive monitors can display colors outside sRGB but that are within AdobeRGB(1998) or ProPhotoRGB.

Such larger color spaces may also be unnecessary. If the actual colors of the original documents are all contained within the sRGB color space, a wider color space employed during conversion is not helpful, and any additional bit depth may or may not show finer color gradations.

Reference Colors Used for Calibration and Profiling

A related issue is the choice of reference colors used for calibration and profiling. An x-rite ColorChecker is one color set traditionally used for device calibration and profiling. But the 24 patches with 18 distinct colors and 6 shades of gray were chosen by target manufacturers to improve the pleasing output for consumer photography. While other, larger, color sets may be used, the authors know of no established relationship between the colors used in device profiling and the colors contained in cultural heritage documents.

Recent research has suggested that document colors mathematically close to the reference colors in the profiling set should be more accurately displayed while distant colors may show higher variability in color accuracy [3]. In turn, the choice of a wide gamut color space may exaggerate this effect.

Objectives

The objective of this study is to begin a database of the actual documents' colors read as CIELAB values by a spectrodensitometer directly from documents' surfaces. This database will provide guidance in selecting the appropriate color space for imaging document sets, and for selecting the color set – or additional individual colors to augment a standard color set – used to create device-specific color profiles.

Prior Studies

Williams and Burns have most recently noted that current digitizing guidelines for testing color image fidelity tend to 'teach to the test' by using the same imaging target for both calibration and color fidelity evaluation. This approach can yield unrealistic results, and may not measure important color regions. While they made detailed measurements on only one document, they note the small gamut and significant distances between their document's measurements and the color checker reference patches. They suggest adopting collection-specific color test targets for digital collections where detection of small differences in material properties is important [4].

It is worthwhile noting that there is an increasing awareness of color-gamut moderation in digitizing collections for cultural heritage applications. While adopting a wide color gamut appears *de rigueur*, this may actually serve to diminish color fidelity by increasing the encoded quantization interval between neighboring colors. Larger quantization intervals can reduce color and tone discrimination, which makes it difficult to render subtle near neutral tones in slowly varying image areas [5]. Berns and Frey also note in their 2005 benchmarking report that for digital collections with limited color-gamut, or where detection of small differences in material properties is important, adopting collection-specific color test targets is advisable. This can be used to reduce metamerism, and to sample the device signal space in the most important regions [6]. Trumpy [7] corroborates this point, "…the color of the object can be colorimetrically distant from the patches of the target and therefore the interpolation process required results in a low level of precision."

Methodology

Team

For document selection in each of the Library's divisions a division curator (or an interested division content specialist) and a digital conversion specialist knowledgeable of the division's collections were enlisted. A Conservation specialist and a Library Digital Scan Center engineer were also asked to provide input as appropriate. This team worked together to select the document set and indicate the specific spots for measurement.

Materials selection

Each division's collection is so large and so diverse that it was impossible to select a statistically representative document set within the scope of this project. In turn, up to ten items of curatorial interest were selected based on the documents':

- Significance / importance
- Usage
- Color
- Graphics
- Format

Following this initial survey, additional document sets may be selected for analysis if any of the color data indicates a more focused and detailed dataset would be useful.

Preparation

All documents selected were reviewed for handling restrictions, and a conservation appraisal was conducted if appropriate. Each document was identified and described using all available existing Library catalog data. Page measurements were used to prepare a Mylar cover sheet for each individual page. This cover sheet was cut to size, labeled, and placed on top of the document. Following the curator's guidance, up to 25 points of interest were marked. Holes were punched and numbered at each point. All resulting Mylar sheets were stored so that time-based data can be collected at later dates.

Data Collection Procedures

Each document was laid on (or backed by) a neutral white support and covered by the Mylar sheet. An x-rite 530 Spectrodensitometer connected to a computer was used to take measurements at each point of interest, and resulting data was placed in an Excel spreadsheet. A first pass was made to collect colorimetric data in the *CIELAB* format– this data was studied to determine the gamut and range of colors within the gamut.

Data Analysis

The *CHROMiX ColorThink Pro* software application was used to plot all data showing the relationship of each color to the sRGB and AdobeRGB(1998) color gamuts. The colors were also compared to the colors of the Macbeth color charts used to calibrate the Library's imaging devices.

Project review

After the collected data was reviewed and analyzed, the authors discussed the results with the division curator, the digital specialist, and interested division staff. This provided the authors additional information about the document colors and allowed for plans for further studies as appropriate.

Results

Initially ten (10) CIELAB datasets were measured from the Library's general collection documents. Each data point represented an image color as La*b* values. Then ten (10) datasets were collected from the Prints and Photographs collections. As of this writing, 4 datasets have been collected from maps in the Geography and Maps collection. Additional datasets are being collected from six (6) maps, and work is underway for obtaining datasets from the manuscript, music, and rare book collections.

The current datasets [8] are from documents ranging from a 1492 atlas to a 1966 identification guide to gems and minerals. An MS Excel spreadsheet was created for each collection. Within each spreadsheet, Sheet 1 is an index with sufficient bibliographic information to identify each item within the Library's collections and to give the reviewer an indication of the type of material and its content. Sheet 1 also provides references to each document's data sheet, to its exported *ColorThink* raw data text file, and all related JPEG images used in the data analysis.

Each succeeding Excel sheet contains the La*b* values for each color point selected on that document. Then, because *ColorThink* cannot read the Excel format, the values are exported as a text file. *ColorThink* computations are additionally presented as 2-dimension graphs which have been converted to JPEG images. Library staff and digitization contractors can readily work with the data when it is prepared in this illustrative manner.

Figure 1 shows the dataset from a full page color illustration in an 1814 travel book. Figure 2 shows the dataset from a 1948 illustrated guide to exotic birds, and Figure 3 shows the dataset from the 1966 photographs tipped into the gems and minerals guide. The shield outline on each figure shows the sRGB gamut. The outer shield around Figure 3 shows the AdobeRGB(98) limit.

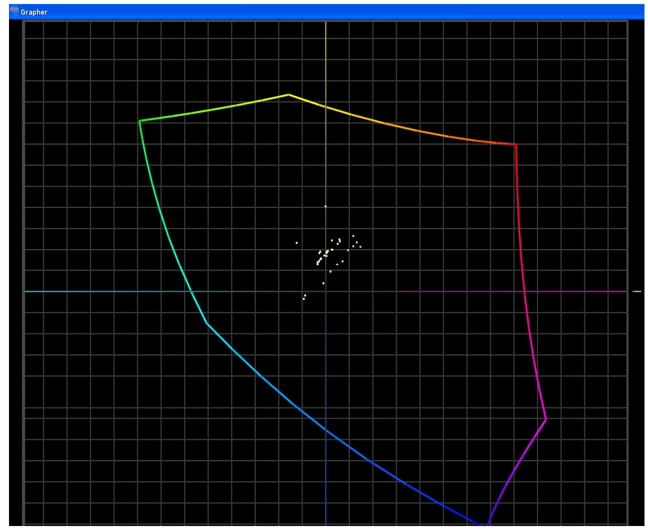

Figure 1. *The CIELAB values of an 1814 travel book color illustration fit within the sRGB color gamut.*

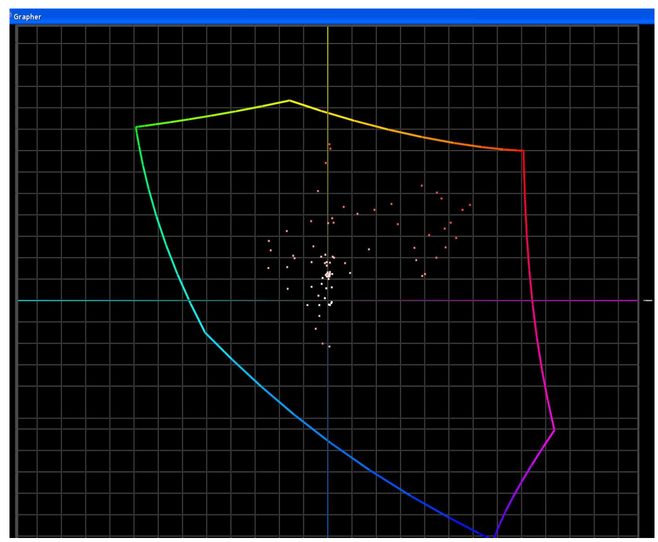

Figure 2. *The CIELAB values of the vivid colors used in a 1948 exotic birds guide illustration fit within the sRGB gamut.*

Currently, a total of 622 spot readings of La*b* values have been taken from all documents. Of the 622 reading, only the two from the gems and minerals photographs shown in Figure 3 have been outside the sRGB gamut.

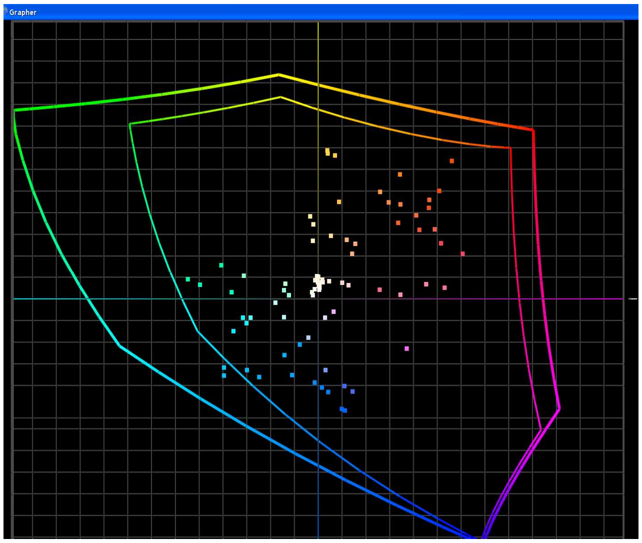

Figure 3. *The CIELAB values for photographs from a 1966 gems and minerals guide do not all fit within the sRGB gamut but are contained within the AdobeRGB(98) gamut.*

Figure 4 shows the sRGB gamut limits, the color points for an 1883 Geological Survey map (square color points), and the color points of a 24 patch *ColorChecker* target used to create a camera ICC profile (circles.)

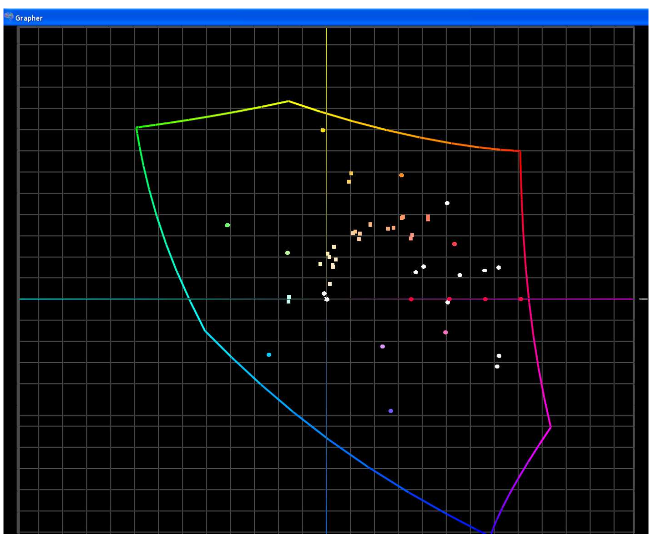

Figure. 4. *The CIELAB values of a 24 patch ColorChecker target (circles) are generally distant from the CIELAB values of an 1883 Geological Survey map's colors (squares).*

Analysis

All documents except one could be imaged using the sRGB color space without color clipping or adjustment. The one exception could be imaged within the AdobeRGB(98) color space. The authors expect that additional documents will be found to require the use of the AdobeRGB(98) color space, and that the Library will be able to then better develop guidelines that help prepare imaging requirements that include a more appropriate color space.

The *ColorThink* diagrams show that the 24 patch *ColorChecker* does not provide reference colors for ICC profile calibration that are close to the colors contained in the study's source documents. To reiterate, prior research suggests that more distant profile reference points result in less accurate color [9].

The Library's data suggests – but does not establish – that these distant color references do indeed cause real issues in color display accuracy. A set of copyright catalog cards, for example, was selected for analysis because the converted image of one card appears much yellower than the original. Though the spectrodensitometer showed that the card in question did produce slightly more yellow La*b* values than the others it appears that the scanner, a high speed, two sided, sheet fed office unit, exaggerated the yellow as if it was calibrated using a single yellow value similar to the yellow of the *ColorChecker*. Figure 5 shows the catalog card data as two squares near the central white point (the illustration has been edited to show the squares more clearly in the space of this paper). Almost all of the catalog cards are represented by the values similar to the lower patch. The card with the very yellow image contains many values shown in the square slightly above. The 24 patch *ColorChecker* values are shown as circles. Note the single highly saturated yellow value directly above, but very distant from the card patches. It appears reasonable to hypothesize that the generic device color profile is distorting the yellows in the captured image because of the lack of less saturated yellow reference values. However, such phenomenon deserves further study.

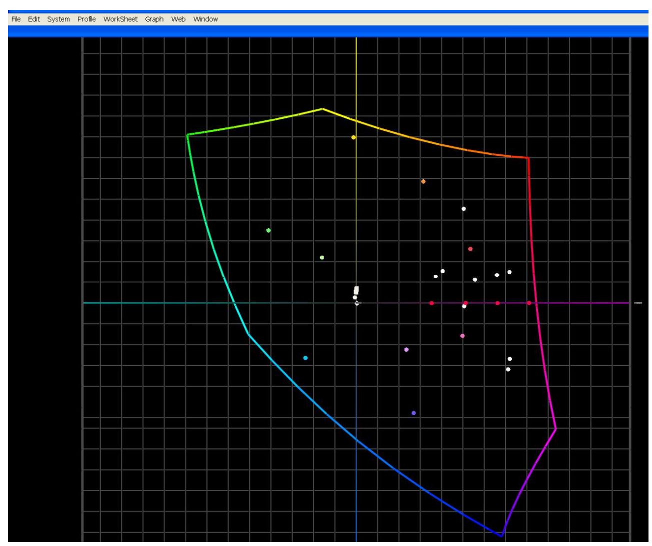

Figure 5. *The yellower copyright catalog card readings are shown as squares slightly above the xy axis intersection. The 24 patch ColorChecker target values are the circles surrounding the card values.*

Figure 6 shows the 1948 exotic bird guide illustration data overlaid with the 140 patch *ColorChecker SG* profiling target for digital cameras. Again, the data suggests but does not establish that using this chart to produce device-specific ICC color profiles might visibly improve the color accuracy of the document images. But even the *ColorChecker SG* patches leave important color regions without close patches.

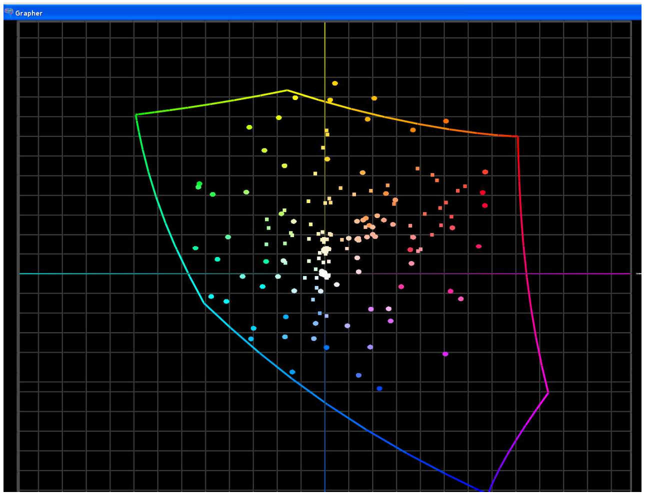

Figure 6. *The CIELAB values of the 140 patch ColorChecker SG target (circles) are generally much closer the CIELAB values of the 1948 exotic birds guide illustration colors (squares)*

Conclusions

The results of this study can be used to help determine the appropriate color space that should be used when imaging specific document collections. This is not a trivial task. Knowledgeable digital specialists have frequently concluded that many color problems are due to the intensity of document colors which fall out-of-gamut. At this point, the authors' data suggest additional factors such as the distances among object colors to profile reference colors may also play a key role in such inaccuracies.

The data can also be utilized to indicate if using a target with significantly more colors, such as the *ColorChecker SG,* to prepare an ICC device profile might help improve color accuracy. In cases where this also may be insufficient, color patches similar to specific document colors might be printed and used to provide closer reference colors.

Finally, an additional benefit of the data gathering process emerged while working with curators and digitization specialists. Each collection presented unique issues and color requirements. Print curators were concerned when the blue roof in an original print appeared green in the image – they wanted improved color accuracy. Map specialist found it difficult to locate on the digital map a specific shade of brown indicated in the original's map legend – they wanted consistent color but were less concerned with the color accuracy. Manuscript specialists wanted to be able to discern specific changes in the base paper's "white" to determine and track the document's condition – other colors were of less concern. In turn, using direct document color measurements can provide for the ability to build custom color sets that tailor the imaging workflow to meet the specific needs of each collection.

References

[1] Roy Berns and Franziska Frey, "Direct Digital Capture of Cultural Heritage-Benchmarking American Museum Practices and Defining Future Needs-Final Report 2005," 2005, (retrieved 3.19.11), http://msc.mellon.org/researchreports/Direct%20Digital%20Capture%20of%20Cultural%20Heritage.pdf/.

[2] Robert Buckley, "CIE – International Commission On Illumination - Technical Committees - Division 8-09 - Archival Colour Imaging", n.d. (retrieved 3.28.11) http://www.cie.co.at/index.php/Technical+Committees.

[3] Giorgio Trumpy, "Digital Reproduction of Small Gamut Objects: A Profiling Procedure based on Custom Color Targets," in *CGIV 2010, Final Program and Proceedings*, vol. 5 (presented at the IS&T CGIV 2010, Joensuu, Finland: Society for Imaging Science and Technology, n.d.), pg. 143

[4] Don Williams and Peter D. Burns, "Capturing the Color of Black and White," *Society for Imaging Science and Technology, Archiving 2010, Final Proceedings*, 2010. pg. 96.

[5] Ibid., 97.

[6] Roy Berns and Franziska Frey, "Direct Digital Capture of Cultural Heritage-Benchmarking American Museum Practices and Defining Future Needs-Final Report 2005," 2005, (retrieved 3.19.11), http://msc.mellon.org/researchreports/Direct%20Digital%20Capture%20of%20Cultural%20Heritage.pdf/. pg. 99-100.

[7] Giorgio Trumpy, "Digital Reproduction of Small Gamut Objects: A Profiling Procedure based on Custom Color Targets," in *CGIV 2010, Final Program and Proceedings*, vol. 5 (presented at the IS&T CGIV 2010, Joensuu, Finland: Society for Imaging Science and Technology, n.d.), pg. 143.

[8] F. Barry Wheeler and Michael J. Bennett, "Dataset for: Accurate Color? A Preliminary Investigation into the Color Gamut of Selected Special Collection Library Objects," (2011). (retrieved 4.7.11), http://digitalcommons.uconn.edu/libr_pubs/35

[9] Giorgio Trumpy, "Digital Reproduction of Small Gamut Objects: A Profiling Procedure based on Custom Color Targets," in *CGIV 2010, Final Program and Proceedings*, vol. 5 (presented at the IS&T CGIV 2010, Joensuu, Finland: Society for Imaging Science and Technology, n.d.), pg. 143.

Author Biography

F. Barry Wheeler is Digital Projects Coordinator in the Office of Strategic Initiatives at the Library of Congress. Previously he has worked as a web services specialist at the FDA Medical Library, and as a Professor of Library Science at The Catholic University of America. He holds a Masters in Library Science from Catholic University and a Doctorate in Instructional Computing from The University of Tennessee.

Michael J. Bennett is Digital Projects Librarian & Institutional Repository Coordinator at the University of Connecticut. There he manages digital reformatting operations while overseeing the University's institutional repository. Previously he has served as project manager of Digital Treasures, a digital repository of the cultural history of Central and Western Massachusetts and as executive committee member for Massachusetts' Digital Commonwealth portal. He holds a BA from Connecticut College and an MLIS from the University of Rhode Island.