https://doi.org/10.2352/issn.2168-3204.2020.1.0.12

Digitization for Everybody (Dig4E)

Paul Conway; University of Michigan; Ann Arbor, USA

Abstract

Digital imaging, as an archival practice, is not a "solved problem' for the cultural heritage community. As Google, publishers, and other content providers digitize and deliver resources at scale, there is an increasingly pressing demand from users to digitize the rich resources in library special collections, archival institutions, and the vast array of invaluable content in private collections. This paper introduces a research and learning initiative (Dig4E-Digitization for Everybody) designed to bridge the knowledge gap that presently exists between well-established or emergent international standards derived from imaging science, on the one hand, and local practices for digital reformatting of archival resources. The paper describes the rationale for the education and training initiative and summarizes the intellectual structure and the technical platform of an innovative sequence of self-paced online resources that can be adapted for a variety of audiences.

Introduction

Digital imaging, as a matter of archival theory and practice, is not a "solved problem' for the cultural heritage community. Librarians, archivists, and museum professionals face three major barriers to providing high quality digital surrogates of their static (largely paper-based) collections: time, money, and knowledge. Time is the enemy of fragile media. There will never be enough money to rescue, extend the lifespan, and/or digitally reformat everything of long-term value [1]. The knowledge gap, however, is pressingly real and possibly solvable. For imaging scientists and archivists who are committed to fostering archival quality digitization in the cultural heritage sector, the primary challenge is to reach and convince a wide-ranging audience to cross a learning bridge between the technical complexities of international digital conversion standards and local adaptive practices that require higher levels of knowledge and skill than are present in most cultural heritage organizations, even after two decades of efforts to develop best-practice guidelines [2][3].

As Google, publishers, and other content providers digitize and deliver resources at scale, there is an increasing demand from users to digitize the rich resources in library special collections, archival institutions, and the vast array of invaluable content in private collections. In the area of still image digitization, we know that user services in cultural heritage organizations turn on the existence and availability of digital image surrogates of cultural heritage resources [4]. Research by Paul Conway and Ricardo Punzalan found that scholars, professional researchers, and expert avocational researchers have overwhelmingly strong preferences for the affordances of archival quality digital images and digital data [5]. We are less certain about the implications of wholesale digitization of cultural heritage resources as both a technology and a social force [6]. A deeper understanding of the ways that international imaging standards are and are not being applied in practice may well inform the future development of imaging standards and extent to which digitization is a "knowledge organization practice" in libraries and archives [7].

Digitization for Everybody (Dig4E) addresses a basic need of nearly the entire cultural heritage community for knowledge on how to master, adapt, and apply existing and emergent technical standards. Dig4E is in part a form of advocacy for embracing the technical qualities embedded in relevant preservation-oriented standards. The project also aspires to make technical standards intellectually accessible and understandable to professionals who are not themselves technical experts. In this way Dig4E responds to Don Williams and Peter Burns' decade-old call for "image literacy," requiring of students and professionals a form of conceptual and technical knowledge that enables them to interpret image science standards and assess their application in practice [8].

Framing Research and Imaging Standards

Steadily over the past twenty-five years, libraries, archives, and museums have made the transition from imaging practices based on locally derived experimentation to processes influenced by the science of imaging [9] and the power of statistical process control [10]. Although it is outside the scope of this paper to recount the history of digitization, it is important acknowledge the pioneering work of image scientists dedicated to promulgating archival quality digitization in the cultural heritage sector. Technical standards and guidelines informed by imaging science research now define the governing principles of good practice.

The fundamental assumption of Dig4E is that the underlying international standards for still image digital transformations (including metadata) are largely in place and quite stable. In 2017, the International Standards Organization (ISO) issued the technical specification for an image science driven method of assessing the image quality of reflective original sources typically held by cultural heritage organizations [11][12]. An accompanying technical report approved the same year outlines in fairly rigorous and technical language recommended best practices for digitizing cultural heritage materials [13]. Both ISO documents utilize a standardized vocabulary of terms [14]. Together, these three documents developed through the efforts of Working Group 26 of ISO Technical Committee 42 synthesize a host of prior standards and, most important, largely reconcile competing guidelines for archival quality imaging developed through more than two decades of imaging practice in Europe and the United States [15][16].

Technical standards are generally difficult to understand, use, and interpret in practical local contexts and in educational settings. Current guidelines developed by the Federal Agencies Digital Guidelines Initiative (FADGI) [17] and the National Library of the Netherlands (Metamorfoze) [18] are dense and challenging to adopt, in part because they represent an admirable effort to embed knowledge of image science metrics in workflows previously dependent on subjective visual inspection [19]. Applying these guidelines in practice requires making the connections between the characteristics of original source materials, the technical performance of the entire digitization workflow, and the purposes to be made of the created in the digitization process [20]. Equally problematical for students and professionals who wish to embrace the underlying technical standards for archival quality imaging is the

lack of open access to standards documents maintained and published by ISO and other national and international standards bodies. With few exceptions, standards publications are held behind a steep paywall.

In contrast to the emergence of mature imaging science as the central feature of digitization best practice, a deeper theorization of digitization's praxis and impact is only emerging. Mats Dahlström has highlighted the importance of human agency in establishing the technical parameters of the digital products created through digital conversion [21]. Zinaida Manžuch emphasizes that the agency of decision-making introduces ethical issues that "influence decisions and the organization of processes at different stages of digitization, from selection and organization of information to developing information systems and online communication of digitized content" [22]. Andrea Sartori argues that ultimately without a full understanding of how different user groups perceive and interpret digitized collections, "our understanding of the digitization process is going to remain largely speculative" [23].

Some research seeks to generalize digitization practice. For example, Zack Lischer-Katz's exploratory study of the information practices of digitizers inside the lab theorizes digitization as an information practice as opposed to merely a set of technical specifications. Lischer-Katz coins the term "improvised self-documentation" to describe the range of materials produced locally in the digitization process, such as application profiles, user guides, and logs of digitized materials. Lischer-Katz finds that digitization standards tend to be background noise – a sort of contextual frame that governs the details of practice [24].

Efforts to translate this research into teaching resources has exposed a major gap between the technical complexity, sophistication, and maturity of the international standards for archival quality digitization and the implicit knowledge required to accomplish digitization in the cultural heritage sector. One consequence, we find, is that existing and emerging standards are not applied as well or as consistently as they could or should be. Technical documentation is rich, while supporting explanatory material is weak or non-existent. The absence of ubiquitous adoption of existing standards and guidelines for archival quality digitization is a major conundrum, because a primary justification for the long-term preservation of digital assets turns in part on the value of the content to be preserved. Standards-based digitization is one of the principal mechanisms for creating content worth preserving [20]. The most important goal of Dig4E is to highlight and advocate for digital imaging standards while explaining how to read, understand, and interpret the technical specifications.

Approach

Digitization for Everybody (Dig4E) is informed by nascent digitization theories but organized to draw upon the increasingly well-defined science of imaging. In developing and publishing online learning resources for understanding digitization standards and workflows, the Dig4E project embraces the notion that distinctions are collapsing between flexible and structured online course support content, on the one hand, and the relatively fixed nature of ePub and print-on-demand formats. Both are digital, modular; both now support multimedia content, and both facilitate self-paced, reinforceable, and assessable learning. Dig4E adopts the model of learning support from Massive Open Online Courses (MOOC) [25], while focusing the content on the application of digitization standards for local practice.

The two-year (2019-2020) self-contained project (Dig4E – Digitization for Everybody) creates, vets with a group of highly experienced experts and consultants, and then publishes modular learning resources that support standards-based digitization of cultural heritage resources. The project website is a gateway to the dynamically produced learning resources [26].

Audiences

The audiences and primary beneficiaries for Dig4E are teachers and learners, broadly defined, who wish to understand and then apply digital imaging standards and the associated core technical metadata elements. Dig4E operationalizes the notion of "Everybody" by creating a flexible, scaffolded learning process with a distinctive approach to defining technical jargon so as to demystify imaging science concepts.

- *Teachers*: Information Schools teach imaging technology and practice sparingly at the graduate level, in a few stand-alone courses, or more typically as week-long modules in courses on collection development, digital curation, and library/archives technology. Instructors who may not themselves be technically expert need teaching resources explicitly developed for the classroom.
- *Workshop Instructors*: Workshops by professional associations largely convey best practices at a generalizable level, rather in a way that encourages and supports the adaption of guidelines in local settings. Dig4E is developing and testing educational materials for learners in workshops or the workplace, along with self-assessment components that learners can use to judge their progress through the materials.
- *Managers and Trainers*: The training of temporary workers in in-house digitization labs is a major and costly challenge for many libraries, archives, and museums. Dig4E is generating educational materials on underlying digital imaging standards and guidelines that can be adapted for training students and temporary workers in local contexts. For digitization projects that involve outsourcing of work to third-party providers of conversion services, the project grounds the learner in digitization standards that can be worked into requests for proposals (RFP). Dig4E also has a special focus on estimating digitization costs, which typically is one of the most challenging elements of effective project grant-writing.
- *Independent Learners*: Community archives and archivists are embracing digitization and audio-visual transformations, largely to support sharing and wider online access o often privately-held archival resources. Many community-based programs focus their energies on the content and value of their materials, often at the expense of technical matters that support exchange, interoperability, and long-term preservation. A wide understanding of digitization and metadata standards provides knowledge and incentives that communities need to adopt quality standards and best practices. Some of the case studies built into the individual learning units help independent learners understand how and when principled compromise on archival standards may be appropriate to reduce costs or manage project with minimal support.

Learning Unit Structure and Design

The Imaging module, one of three modules in the Dig4E project, is structured as fourteen "Learning Units" that cover the territory of best practices. The Imaging module is self-paced; individual learning units are repeatable and independent, supported by open, auto-graded quizzes designed to telegraph learning objectives and provide feedback on the mastery of key ideas. Each "Learning Unit" features one or more case studies of good and not-

so-good outcomes of digitization process processes, with particular reference to points in a workflow where problems and errors arise due to a combination of material defects, technology system failures, or post digitization processing. Each "Learning Unit" is independent, self-paced, and repeatable.

Table 1: Dig4E-Imaging Learning Unit Outline

1. Vision and Digital Imaging	We learn about visual perception and the digital encoding of reflected light.
2. Imaging Standards Overview	We learn about the state of international imaging standards and guidelines.
3. Physical Care & Handling	We learn about handling materials safely and discuss some myths about light exposure.
4. Imaging Technology Options	We learn about the fundamentals of scanners, viewing conditions, and lighting.
5. Imaging Workflow	We consider imaging procedures and the ethics of producing archival quality images.
6. System Performance	We explore assessing imaging system capabilities using standardized test charts.
7. Resolution	We learn about specifying and measuring Spatial Frequency Response.
8. Tone Response	We learn about assessing image tone/exposure and white balance/neutrality.
9. Color Accuracy	We introduce the fundamentals of visible color spaces and color encoding accuracy.
10. Image File Formats	We learn about the TIFF and JPEG2000 file formats.
11. Image Processing	We explore ethics and techniques for normalizing master images and producing derivatives.
12. Image Metadata Standards	We learn document the technical features of digital images.
13. Outsourcing Digital Imaging	We learn how to estimate imaging costs and work with third-party vendors.
14. Future of Digital Imaging	We consider the technical trends that are changing digital imaging.

Table 1 is a list of learning units for the Imaging module as presently configured. In numbering the units, the module suggests a path forward, and yet each learning unit is discrete. Several learning units serve to frame archival quality digital imaging in the broader context of image science. The majority of the units is mapped to international standards or components of the FADGI guidelines.

The online components of each learning unit include a variable combination of textual explanations of the relevant standards, illustrated as appropriate with graphic material either created explicitly for the project provided by organizations as public domain content or through open Creative Commons licensing. The online textual and graphic materials (also downloadable in PDF format) are supplemented by short audio-only or audiovisual presentations that demonstrate technical processes, the physical condition of source materials, digitization equipment, or working through a step-by-step workflow document. Dig4E Imaging allows for the provision of supplemental materials in audiovisual formats, the automatic scoring of quizzes that accompany each Learning Unit, and hyperlinks to glossary terms. Dig4E also provides the capacity to integrate pre-existing content provided and licensed by project collaborators.

Delivery Platform

The operative metaphor for the modular components of Dig4E and the associated code required to assemble the components dynamically is an online, multimedia, interactive book. To enable the broadest adoptability of open and free books, the technical platform for the project provides structured access to ancillary materials, including presentation slides, videos, or assignments, and provides interactive software elements to support self-assessment of learning. Additionally, the open standards learning environment such as we are building for Dig4E requires a software framework and hosting platform (GitHub + open standards web servers) that makes the learning materials and software available using open content exchange standards such as IMS Learning Tools Interoperability. In the Dig4E project we are building reusable libraries that make it relatively easy for the project's open content and applications to support interoperability.

This work is based on the Tsugi project (Japanese for "next") [27]. An example of combined book + materials + assessments is Python for Everybody (www.py4e.com). This dynamically-produced web site functions as a learning object repository, application store, standalone Massive Open Online Course (MOOC). Web content and Tsugi tools are organized into "reusable learning objects" within a learning module using the Koseu "Lessons" tool. The resources become items in a Learning Object Repository (LOR), typically housed in GitHub repositories that support the Tsugi learning objects framework. The Koseu (Korean for "course") platform provides a number of analytics options that allow us to track overall usage and be informed when there are problems with the content as experienced by the learners.

Learning Assessment

Assessment of learning is fundamental to the Dig4E project. Typically in large-scale online learning environments, evaluation of learning outcomes takes the form of summative assessment, including establishing pre-requisite knowledge, a demonstration of obtained mastery, or the gateway through which one progresses up a "ladder" of sequential courses toward a degree or certification of competency. In the context of Dig4E, assessment is iterative and formative rather than summative in nature.

Each "Learning Unit" includes an auto-graded quiz with multiple- choice questions that serves multiple purposes simultaneously. A quiz helps the learner determine whether the learning unit might be useful to pursue. A quiz serves as an explicit statement of learning outcomes for the unit. Quizzes allow learners to assess their own learning and reinforce the important points to remember. The self-assessment components are designed for immediate scored feedback in an online environment. The primary pedagogical goal of the multiple choice quizzes is to give learners a sense of the material covered within each unit and cumulatively in

the Imaging module. Learners are encouraged and allowed to take the quizzes repeatedly to help them master the material.

A future benefit of the quiz structure is the potential to cluster learners into a group and moderate or guide the learning with the assistance of an instructor. This particular feature, already built into the coding of the delivery platform, will require the implementation of a login function that is not enabled in the present open content iteration.

Reuse and Sustainability

All Dig4E content created for the learning initiative is freely available and distributed through the project website under a Creative Commons BY-NC-SA license, which lets others use, revise, and build upon the work non-commercially, as long as they attribute their work properly and release any revisions under an identical CC license. This type of license maximizes reuse capabilities while acknowledging the creators of the original content. In terms of permissions, all audio and video components created by the project team are open and freely available. All illustrated materials will be created by the project team or drawn from online resources (such as Flickr or Wikimedia Commons) where an open Creative Commons license has already been assigned. Any preexisting user licenses will be carried forward into the Imaging module.

The University of Michigan will preserve all educational content and any associated computer code in its open digital repository, Deep Blue Data, which is the university's preservation service for digital research data built upon the Samvera/Fedora platform.

Conclusion

The Dig4E project is ambitious in terms of the quantity, variety, and diversity of content that is being created and published to support learning about digitization standards. We aspire to offer some really well developed and well-vetted learning materials to help practitioners across a range of audiences embrace the value and present maturity of digitization standards and apply them to create well-formed archival-quality digital surrogates of some of the most common cultural heritage resources.

References

[1] P. Conway, "Preservation in the age of Google: Digitization, digital preservation, and dilemmas," Lib. Q., vol. 80, no.1, pp. 61-79, 2010.

[2] M. Alghnimi, "Digitization guidelines for static & non-static (Audiovisual) media: Compliance & challenges in academic libraries," Ph.D. dissertation, Univ. Wisc., Milwaukee, 2018. [Online]. Available: https://dc.uwm.edu/etd/1969

[3] P. Conway, "Best practices for digitizing photographs: A network analysis of influences," in Proc. IS&T Archiving 2008, (Berne, Switzerland), (June 24-27), 2008, pp. 94-102.

[4] P. Conway, "Digital transformations and the archival nature of surrogates," Arch. Sci., vol. 15, no. 1, pp. 51-69, 2015. doi: 10.1007/s10502-014-9219-z

[5] P. Conway and R. Punzalan, "Fields of vision: Toward a new theory of visual literacy for digitized archival photographs," Archivaria, vol. 71, pp. 63-97, 2011.

[6] P. Gooding, M. Terras, and C. Warwick, "The myth of the new: Mass digitization, distant reading, and the future of the book," Lit. and Ling. Comp., vol. 28, no. 4, p. 632, 2013. doi: 10.1093/llc/fqt051

[7] M. Dahlström, J. Hansson, and U. Kjellman, "'As We May Digitize': Institutions and documents reconfigured," LIBER Q., vol. 21, nos. 3-4, pp. 455–474, 2012. doi: 10.18352/lq.8036

[8] D. Williams and P. D. Burns, "Preparing for the image literate decade," Proc. IS&T Archiving 2009, (Arlington, VA), (May 4-7), pp. 124-127, 2009.

[9] F. Frey & S. Süsstrunk, "Image quality issues for the digitization of photographic collections," Proc. IS&T 49th Annual Conference, (Minneapolis, Minn.), (19-24 May), pp. 349-353, 1996.

[10] M. Stelmach and D. Williams, "When good scanning goes bad: A case for enabling statistical process control in image digitizing workflows," Proc. IS&T Archiving 2006, (Ottawa, Canada), (23-26 May), pp. 237-243, 2006.

[11] Photography — Archiving systems — Image quality analysis — Part 1: Reflective originals, ISO/TS 19264-1:2017, International Standards Organization, Geneva, Switzerland, 2017.

[12] D. Wueller and U.B. Kejser, "Standardization of image quality analysis – ISO 19264," Proc. IS&T Archiving 2016, (Washington, DC), (19-22 Apr.), pp. 111-116, 2016.

[13] Photography — Archiving systems — Part 1: Best practices for digital image capture of cultural heritage material, ISO/TR 19263-1:2017, International Standards Organization, Geneva, Switzerland, 2017.

[14] Photography – Archiving Systems – Vocabulary, ISO 19262:2015, International Standards Organization, Geneva, Switzerland. [Online}. Available: https://www.iso.org/obp/ui/#iso:std:iso:19262:ed-1:v1:en

[15] Image Access GmbH, Quality Controlled Scanning FAQ 1.0, 2017. Accessed: 15 April 2020. [Online]. Available: https://support.imageaccess.de/downloads/product_manuals/FAQ/FAQ-Quality-Controlled-Scanning.pdf

[16] Federal Agencies Digital Guidelines Initiative, Still Image Group, "Guidelines: Digital imaging framework," April 2, 2009. Accessed: 15 April 2020. [Online]. Available: http://www.digitizationguidelines.gov/guidelines/digitize-framework.html

[17] Federal Agencies Digital Guidelines Initiative, Still Image Group, "Technical guidelines for digitizing cultural heritage materials, September 2016. Accessed: 15 April 2020. [Online]. Available: http://www.digitizationguidelines.gov/guidelines/digitize-technical.html

[18] H. van Dormolen, "Metamorfoze preservation imaging guidelines, version 2.0," Proc. IS&T Archiving 2019, (Lisbon, Portugal), (14-7 May), 2019, pp. 9-11. doi: 10.2352/issn.2168-3204.2019.1.0.3

[19] M. Hoffmann, "Quality assurance: Visual inspection of digitized images," Proc. IS&T Archiving 2018, (Washington, DC), (17-20 April), pp. 110-114, 2018. doi: 10.2352/issn.2168-3204.2018.1.0.24

[20] P. Conway, "6.4. The relevance of preservation in a digital world," NEDCC Preservation Leaflets, Andover, MA: Northeast Document Conservation Center, 1999. [Online]. Available: https://www.nedcc.org/free-resources/preservation-leaflets/6.-reformatting/6.4-the-relevance-of-preservation-in-a-digital-world

[21] M. Dahlström, "Critical editing and critical digitization," in Text Comparison and Digital Creativity, W. Th. van Peursen, E. Thoutenhoofd, and A. van der Weel, eds., Brill Online, 2010.

[22] Z. Manžuch, "Ethical issues in digitization of cultural heritage," J. of Cont. Arch. Stud., vol. 4, no. 4, 2017. [Online]. Available: https://elischolar.library.yale.edu/jcas/vol4/iss2/4

[23] A. Sartori, “Towards an intellectual history of digitization: Myths, dystopias, and discursive shifts in museum computing,” Dig. Schol. in the Hum., vol. 31, no. 2, p. 437, 2016.
[24] Z. Lischer-Katz, “Digitization as information practice,” Proc. iConference 2014, (Berlin, Germany), (4-7 March), p. 1104, 2014. [Online]. Available: http://hdl.handle.net/2142/47417
[25] C. Severance, "MOOCs: An insider's view," Computer, vol. 46, no. 10, pp. 93-96, October 2013.
[26] “Digitization for Everybody (Dig4E),” University of Michigan School of Information. http://www.Dig4E.org (accessed 15 Apr. 2020).
[27] N. Galanis, M. Alier, M.J. Casany, E. Mayol, and C. Severance, “TSUGI: A framework for building PHP-based learning tools,” Proc. TEEM ’14, (Salamanca, Spain), (Oct 1-3), pp. 409-413, 2014. doi: 10.1145/2669711.2669932

Acknowledgements

The Andrew W. Mellon Foundation provided generous funding for the development and the development and deployment of Dig4E project learning resources. Don Williams and Tom Rieger provided invaluable technical advice. Jordan Gorzalski designed the project website. Charles Severance is contributing his programming expertise to constructing the delivery system infrastructure.

Author Biography

Since 2006 Paul Conway has been associate professor at the University of Michigan School of Information. His professional career encompassed archivist at the National Archives and Records Administration, Preservation Program Officer for the Society of American Archivists, and senior administrator for the libraries at Yale University and Duke University.