Protecting the Pathology Commons in the Digital Era

Benjamin L. Mazer, MD, MBA; Nathan Paulson, MD; John H. Sinard, MD, PhD

Whole-slide imaging (WSI) technology and machine learning (ML)–aided diagnostics are primed to bring us new ways of seeing, learning, and understanding human pathology.[1] Physicians, investors, and the public are being promised that these technologies will provide "faster, more accurate diagnosis of disease"[2] and allow pathologists to "make better decisions for their patients."[3] We worry, however, that these technological advances also come with practical and ethical risks. One particular concern is that WSI and ML are being accompanied by a monopolization of slides, technology, and intellectual property. Some digital monopolies in this arena are de facto whereas others are legal, but both may lead to a loss of shared access to current and future knowledge. We see 3 long-term risks to the monopolizing forces developing in anatomic pathology: the privatization of seeing, the privatization of experience, and the privatization of understanding. To counter these trends, we recommend that our profession create a digital pathology commons.

PRIVATIZED SEEING

The glass slide has many limitations because of its physicality: compared with digital information, it is difficult to store, transport, and copy. But its physicality is also a benefit. If I have a slide, I know I can view it. Many kinds of microscopes are commercially available, and one could even build one's own. It is worth contrasting this openness with the complex legal and technological environment we must navigate with WSI. Take the first Food and Drug Administration–approved WSI device for primary diagnosis: the IntelliSite Pathology Solution (Philips, Amsterdam, Netherlands).[4] This device produces images in the proprietary iSyntax image format.[5] Compared with the interoperability of the humble glass slide, this image format, and the corresponding libraries and viewer necessary to access it, is far more restrictive. Most users of the technology will purchase all of the hardware and software directly from Philips. The company does provide a software package to allow third-party developers to access this image format; however, any third-party users must agree to an end-user license agreement that stipulates the software package is only for academic research or for commercial use limited to in-house laboratory-developed tests (LDTs).[6,7] Developers cannot use the package to create publicly available software or hardware, and thus cannot compete directly with Philips. By forbidding derivative commercial applications, these licensing terms encourage vendor lock-in, especially if laboratories develop LDTs that rely on the platform. Accordingly, even open-source projects like OpenSlide (Mahadev Satyanarayanan, Carnegie Mellon University) and Bio-Formats (Open Microscopy Environment, University of Dundee), whose aim is to allow universal access to WSI and image analysis, do not support iSyntax.[8,9] Future access to the iSyntax format depends upon the company continuing to maintain this software and these licensing terms.

We do not mean to deride a specific company for restrictive terms. We would expect any commercial vendor to pursue some form of lock-in—it's just good business. But as pathologists, it is our job to look beyond a single vendor or single point in time to ensure continued access to the images we depend on for our clinical and scientific work. We should not discount the value of an open, interoperable viewing system that connects pathologists around the world and across decades.

PRIVATIZED EXPERIENCE

Pathology education is a volume business. We don't learn a diagnosis based on a single image in a textbook. In residency and beyond, we gain skill by viewing diverse examples of diagnostic entities. Similarly, access to an adequate case volume is how we ensure pathology's observational research methods are powered to produce accurate results.[10] Patients, mostly unknowingly, have enabled the education and scholarly work of generations of pathologists by allowing their tissue to remain within institutional archives.

As the digital era dawned, it appeared that pathologists would more easily be able to access pathology archives. Eminent teaching sets like those of Juan Rosai[11] and Johns Hopkins (Baltimore, Maryland)[12] were made public online. Websites like WebPathology[13] and PathologyOutlines[14] offered freely available pathology encyclopedias. Pathologists took to social media to share educational images.[15]

We may now be seeing a reversal of this trend. Memorial Sloan Kettering Cancer Center (MSKCC; New York, New York), for example, is home to a rich archive of more than 25

million slides.[3] This venerable trove has taught generations of trainees and produced innumerable scientific studies. Yet MSKCC has recently courted controversy by granting an exclusive license for use of its slides in computational pathology to a single start-up company. The monetization of these slide sets has garnered strong objection from even some of the hospital's own pathologists, according to public reporting.[16]

We do not intend to criticize any specific organization. Memorial Sloan Kettering Cancer Center is not unique in recognizing the value of patient tissue samples. Although its digital slides may still be available for use outside of computational pathology, other organizations could license slides under more restrictive terms. Machine learning methods rely on processing large data sets, so the appeal of securing access to institutional archives is obvious. Indeed, through the power of a data set as large as MSKCC's, Paige.AI (New York, New York) has already produced impressive sensitivity in detection of some types of cancer.[17] Digitizing libraries of slides is also costly for hospitals, which understandably want to recoup these expenses. The opportunity costs, however, must also be considered. Diverse clinical, scientific, and technological advances can all be derived from a single pathology archive. Although it is easy to see the profits available from these archives, it is difficult to quantify the downstream effects of limiting access. We feel these effects will be significant to patients and practitioners. As adoption of digital pathology increases among academic institutions, the field faces an important turning point: access to data sets for learning and scientific discovery could increase substantially or could decrease in pursuit of monopoly profits.

PRIVATIZED UNDERSTANDING

Although a histologic image doesn't change, the way we understand it does. When a pathologist views a slide, he or she relies on a variety of conceptual frameworks for interpretation. These concepts evolve over time through experience, research, and innovation. Pathologists have been viewing histologic sections of gastric ulcers for centuries, for example, but only since the work of J. Robin Warren, MBBS, and Barry Marshall, MBBS, do we understand that the helical bacteria often present in these ulcers represent their cause.[18] The development of ML technologies promises to accelerate and multiply the conceptual frameworks used for histopathologic interpretation. Each new ML model developed represents a unique way of understanding a histologic image. ML techniques, however, often produce algorithms that remain a "black box" to the human user.[19] Unlike simpler technologies that replicate a human practice (eg, counting mitoses or scoring human epidermal growth factor receptor 2 stains), opaque ML algorithms can lead to a natural monopoly. Growing private investment in commercial ML enterprises also suggests that many of these systems will be proprietary.

We recognize the potential for automated ML diagnostics to improve pathology's accuracy and reproducibility. Ironically, as these digital systems improve, the possibility of vendor lock-in increases. If a proprietary ML system offered superior prognostic or predictive interpretation, physicians would feel ethically obligated to offer it to patients. One real-life example of this ethical dilemma is the growing use of the OncotypeDX test (Genomic Health, Redwood City, California) for guiding treatment of early breast cancer.[20] Although this test has improved care, it has also centralized tumor prognostication within a single for-profit company. Monopolization of pathologic understanding threatens equity and access for underserved patient populations, even if it improves outcomes for some.

To understand the implications of these trends, we can look to an analog example: the commercialization of clinical questionnaires. Recently, some entrepreneurial clinical assessment creators have retained the intellectual property to their survey instruments in order to charge licensing fees—sometimes upwards of thousands of dollars—for their use. These practices have caused "licensing headache[s]" for the scientific community, ultimately leading to study retractions and legal conflicts because fees weren't paid.[21] If these behaviors become widespread, they could have a chilling effect on practice and research. Pathologists would balk at paying a licensing fee every time they provided a Gleason score to a patient, for example, yet this becomes a real possibility as the field moves to ML-assisted diagnostics, whose use can more easily be controlled by creators.

CREATING A DIGITAL PATHOLOGY COMMONS

In the rush to innovate and profit, we risk losing the pathology commons. The "commons" is a term used in legal, ethical, and historical contexts to refer to "shared resources in which each stakeholder has an equal interest."[22] We are specifically defining the pathology commons as the set of openly shareable and modifiable knowledge the profession relies on for practice and scholarship. This can encompass histologic images, diagnostic frameworks, scientific information or techniques, and other resources. Law professor Lawrence Lessig, JD, has pointed out that the "key freedom" of knowledge not restricted by intellectual property law is not that we may access it free of cost, but that knowledge becomes "permission free."[23] Scientific progress depends on the ability of scholars and practitioners to discuss, modify, and innovate without friction. Today, pathologists may take this for granted. If we read a new study suggesting a set of histologic criteria to diagnose a tumor, we can immediately try these criteria out in our own practice, perform a follow-up validation study, or discuss the criteria on any medium we choose. Appropriate attribution is required, of course, but we have no need to call up the authors to beg for permission to practice, research, or debate this new piece of knowledge. It's hard to imagine the same freedom with proprietary slide images and machine-learning algorithms.

How do we create a digital commons? Here are the steps we feel are most important to take. First, the profession must ask that any technologies use an open storage format (common seeing). Whole-slide images should be stored in a nonproprietary, royalty-free format that can be independently reimplemented. Second, the profession must encourage the hospitals and physicians who care for pathology slides not to sequester them through exclusive contracts (common experience). All scholars and commercial entities that wish to analyze large pathology data sets in order to produce new knowledge should be granted access under reasonable conditions and costs. We don't mean to imply free access is required, but prices should not exclude the majority of scholars or companies and could be scaled to the user's resources. Third, any new algorithms or concepts derived from these data sets should in turn be accessible to other scientists and businesses (common understanding).

The pathology commons is fundamentally decentralized, but deliberate tools and practices can enhance it. Some academics may be familiar with Creative Commons (CC) licenses, which were developed to facilitate sharing and are now the primary means of licensing open-access scholarly publications.[24] Creative Commons licenses allow intellectual property holders to easily and clearly delineate how the public may use their work. This clarity and consistency encourage knowledge dissemination and even iteration. Creative Commons licenses also allow an optional "share alike" clause to ensure derivative works remain open. Stakeholders can therefore enhance the pathology commons by promoting use of CC licenses (or equivalents) for digital resources such as images and diagnostic algorithms.

We believe this philosophy is supported by statements from stakeholders like the Digital Pathology Association (DPA). The organization has written that "leverag[ing] patient data in order to commercialize artificial intelligence…raises ethical and legal concerns regarding data ownership and intellectual property."[25] The DPA goes on to encourage regulators and vendors to "work together to set standards and increase interoperability" within computational pathology, and that these standards "should be formalized via regulatory guidance." A group of stakeholders has created a pathology-specific version of an open imaging standard called Digital Imaging and Communications in Medicine (DICOM); however, the DPA warns us that "in the current digital pathology landscape, whole slide imaging systems store image data in proprietary file formats…the proprietary nature of data formats and interfaces create vendor lock-in and impede data access."[26] The Philips system claims that it "will allow migration of the data in the platform to or from alternative platforms using standards such as DICOM,"[27] which is an important step toward interoperability. As the DPA suggests, however, "there is a practical difference…between exporting single slides as DICOM, as opposed to routinely transmitting every slide as DICOM."[28]

BARRIERS TO A DIGITAL COMMONS

One counterargument to a permissive intellectual property scheme is that it will lack the profit incentives needed to generate innovation. We don't dispute that financial incentives and industry-academic partnerships can foster innovation. On the other hand, we believe that many of anatomic pathology's foundational diagnostic advances have been the product of an open era. As pathologists decide whether or how to license slide archives and intellectual property, the profession will inevitably face a difficult balancing act between funding innovation and maintaining access.

We recognize that the concerns we raise could feel theoretical at the moment, that we may seem opposed to technological progress, or that we are simply too idealistic. We suspect most pathologists see the advantage of adopting a universal WSI format, for example, yet even this has not been achieved. There are multiple potential mechanisms available to protect the pathology commons, from best-practice guidelines to open-source consortia to regulatory levers. Our goal in this editorial is not to prescribe definitive solutions but to broaden the dialogue from digital pathology specialists to include all practicing pathologists. Effective practice and policy changes will require buy-in from commercial, government, academic, and community stakeholders. This is no easy feat.

In the beginning, WSI and ML may merely supplement analog technologies, but over time they could entirely replace them. We personally believe ethical and regulatory protections should be in place before that occurs. If pathologists come to rely on closed systems for practice and scholarship, reversing these changes will be difficult. Digital pathology is promising convenience, innovation, and profit, but pathologists must simultaneously protect our tradition of open learning and ensure patient data is handled safely and equitably.

References

1. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253–e261. doi:10.1016/S1470-2045(19)30154-8
2. PathAI Web site. https://www.pathai.com. Accessed January 15, 2020.
3. Lunden I. Paige.AI nabs $25M, inks IP deal with Sloan Kettering to bring machine learning to cancer pathology. TechCrunch Web site. http://social.techcrunch.com/2018/02/05/paige-ai-nabs-25m-inks-ip-deal-with-sloan-kettering-to-bring-machine-learning-to-cancer-pathology/. Published February 5, 2018. Accessed January 15, 2020.
4. Philips receives FDA clearance to market Philips IntelliSite Pathology Solution for primary diagnostic use in the US. https://www.usa.philips.com/a-w/about/news/archive/standard/news/press/2017/20170413-philips-fda-clearance-intellisite-pathology-solution-primary-diagnostic.html. Published April 13, 2017. Accessed January 15, 2020.
5. Philips' iSyntax for digital pathology. *Pathologist*. November 2016. https://thepathologist.com/fileadmin/issues/App_Notes/0016-024-app-note-Philips__iSyntax_for_Digital_Pathology.pdf. Accessed January 15, 2020.
6. End user license agreement for research use. Philips Open Pathology. https://www.openpathology.philips.com/eula-research/. Accessed January 15, 2020.
7. End user license agreement for limited commercial use. Philips Open Pathology. https://www.openpathology.philips.com/eula-limited-commercial-use/. Accessed January 15, 2020.
8. Philips format. OpenSlide. https://openslide.org/formats/philips/. Accessed January 15, 2020.
9. Supported formats. Bio-Formats 6.3.1 documentation. https://docs.openmicroscopy.org/bio-formats/6.3.1/supported-formats.html. Accessed January 15, 2020.
10. Mazer BL, Homer RJ, Rimm DL. False-positive pathology: improving reproducibility with the next generation of pathologists. *Lab Invest*. 2019;99(9):1260–1265. doi:10.1038/s41374-019-0257-2
11. Juan Rosai's collection of surgical pathology seminars (1945–present). https://www.rosaicollection.org/. Accessed January 15, 2020.
12. Johns Hopkins surgical pathology unknown conference. https://apps.pathology.jhu.edu/sp/. Accessed January 15, 2020.
13. Ramnani D. WebPathology. https://www.webpathology.com/. Accessed January 15, 2020.
14. PathologyOutlines.com. https://www.pathologyoutlines.com. Accessed January 15, 2020.
15. Oltulu P, Mannan AASR, Gardner JM. Effective use of Twitter and Facebook in pathology practice. *Hum Pathol*. 2018;73:128–143. doi:10.1016/j.humpath.2017.12.017
16. Ornstein C, Thomas K. Sloan Kettering's cozy deal with start-up ignites a new uproar. *New York Times*. September 20, 2018. https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html. Accessed January 15, 2020.
17. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301–1309. doi:10.1038/s41591-019-0508-1
18. Warren JR, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet*. 1983;321(8336):1273–1275. doi:10.1016/S0140-6736(83)92719-8
19. Wang F, Kaushal R, Khullar D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Ann Intern Med*. 2020;172(1):59. doi:10.7326/M19-2548
20. Andre F, Ismaila N, Henry NL, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: ASCO clinical practice guideline update—integration of results from TAILORx. *J Clin Oncol*. 2019;37(22):1956–1964. doi:10.1200/JCO.19.00945
21. Marcus A. Pay up or retract?: drug survey spurs conflict. *Science*. 2017;357(6356):1085–1086. doi:10.1126/science.357.6356.1085
22. Research on the commons, common-pool resources, and common property. Digital Library of the Commons. http://dlc.dlib.indiana.edu/dlc/contentguidelines. Accessed January 15, 2020.
23. Lessig L. Recrafting a public domain. *Yale J Law Humanit*. 2013;18(3):56–83.

24. Open access. Creative Commons. https://creativecommons.org/about/program-areas/open-access/. Accessed January 15, 2020.

25. Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol*. 2019;249(3):286–294. doi:10.1002/path.5331

26. Herrmann M, Clunie D, Fedorov A, et al. Implementing the DICOM standard for digital pathology. *J Pathol Inform*. 2018;9(1):37. doi:10.4103/jpi.jpi_42_18

27. Hulsken B. Philips IntelliSite open pathology platform. *Pathologist*. November 2016. https://thepathologist.com/fileadmin/issues/App_Notes/0016-025-app-note-Philips_IntelliSite_Open_Pathology_Platform.pdf. Accessed January 15, 2020.

28. Clunie D, Hosseinzadeh D, Wintell M, et al. Digital imaging and communications in medicine whole slide imaging Connectathon at Digital Pathology Association Pathology Visions 2017. *J Pathol Inform*. 2018;9(1):6. doi:10.4103/jpi.jpi_1_18