

Don't wait until it's too late: Can preservation influence the design of complex digital publications?

Catherine Nicole Coleman

Stanford Libraries, Stanford University, Stanford, California, USA.

E-mail address: cnc@stanford.edu

Jasmine Mulliken

Stanford University Press, Stanford University, Stanford, California, USA.

E-mail address:



Copyright © 2020 by Catherine Nicole Coleman and Jasmine Mulliken. This work is made available under the terms of the Creative Commons Attribution 4.0

International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Digital scholarship has found a new expression: book-length, peer-reviewed, interactive, multimodal, web-based publications with no print corollary. For the past five years, Stanford University Press and Stanford Libraries have been engaged in a grant funded partnership to pave the way for others, too, to publish this new form of scholarship. These new digital projects push the bounds of the growing interest of other academic presses in the “enhanced ebook.” Stanford University Press has already published six such projects with more on the way. These works take full advantage of the rich multi-media potential of digital delivery to communicate scholarly arguments in both narrative and form. And they present a new set of challenges for access and preservation.

Though we call them ISW's, an acronym for interactive scholarly works, our LOCKSS team has a different interpretation of the acronym: “Impossible to store and work with.” What makes interactive multi-modal projects particularly challenging for preservation is that the scholarship is embedded in the digital form. In other words, the medium in which the work is delivered to the reader, its unique format, is an essential part of the argument. To take the work out of its environment — to rebind it, as it were — would be to remove half of what the author created. The author is writing more than the argument and the text, she is writing the functionality of the object itself.

University presses and libraries have well-established protocols and processes for print publication, many of which are rooted in our assumptions about the durability and longevity of the printed word. ISW's also need to find a place within the academy, one that is at least persistent enough to be considered for a scholar's hiring, promotion, or tenure evaluation. The Stanford University Press has worked closely not only with Stanford Libraries but also with projects developing the technologies surrounding digital authorship and online publication to establish guidelines and standards to help mitigate the fragility inherent in software dependencies. The tools used in authorship can complicate already complex content permissions, and even upend our understanding of what constitutes a published work.

A published work that requires maintenance to ensure access begs the question: Who will be responsible for maintenance, and for how long?

What we have learned so far is that this is a critical moment for scholars, publishers, libraries, and vendors to share ideas and expertise about how to address sustainability of the knowledge created in this new form. This is a case where the expertise of library professionals can influence the design of tools for authorship that will in turn help us shape a sustainable authoring and publication process. To facilitate that cross-functional conversation, we have created a diagram comparing the life of a print publication to that of an electronic book to that of an ISW, from the author's inspired proposal through to preservation.

Keywords: Digital Humanities, scholarly communications, preservation, multimodal, archiving

DON'T WAIT UNTIL IT'S TOO LATE: CAN PRESERVATION INFLUENCE THE DESIGN OF COMPLEX DIGITAL PUBLICATIONS?

INTRODUCTION

In 2014, Stanford University Libraries and Stanford University Press launched a digital publishing program for post-monograph scholarly arguments and products that was intended to make book-length, peer-reviewed digital projects into first-class academic publications. In addition to establishing a portfolio of published works, the goal of the collaboration has been to build the program into a sustainable model for other academic publishers to consider. The partnership between library and publisher leverages the complementary strengths, resources, and experience that each brings to the collaboration: Stanford University Press's imprint and long experience in academic publishing of peer-reviewed scholarly work; and the archiving, digital delivery, and scholarly functions development programs of the Stanford University Libraries.

The move from print to digital publications is part of a larger cultural shift that still carries notions that digital is faster, cheaper, and easier. In fact, digital scholarship has distinct advantages over print: larger audiences and a quicker path to publication which results in more citation opportunities for scholars. (Borgman, 2010) But authors, in large part, have not let go of the expectations that scholarly publishing promises both validation and *longevity*. Longevity and the digital are at odds with each other. Abbey Rumsey addresses the problem of digital longevity in her book *When We Are No More* (2016). We, as a society, writes Rumsey, need to catch up with our own ambitions. If new digital forms of publication are here to stay,—and we believe they are—then we need to figure out what kind of sustainability is appropriate and what is possible.

Six projects have already been published with the press and yet preserving these unique works remains a mostly unsolved problem. Print converted to digital suffers from the fundamental problem of bit rot. Electronic publications add to bit rot the problem of link rot,

as digital publication begets online citation. Interactive Scholarly Works, a term which in this paper we use interchangeably with digital projects, are subject to bit rot, link rot as well as code churn, deprecated format standards, disappearing software platforms, and dissolution of the companies who create them. Digital projects are not sets of ascii files. They are digital organisms with many interdependent parts.

a) Why are digital projects important to 21st century scholarship?

Even for those who are unfamiliar with all of the details of publishing a print book, the materiality of the accomplishment reaches the senses: When the cover catches your eye, when you pick it up in your hand and feel its weight, when you crack it open and smell the ink and paper, when you leaf through the pages, everything from the running head to the page number is elegantly in proportion, where you expect it to be. All of the conventions that have evolved over time with the printed book to make it accessible — the table of contents, the index, the running heads, even the colophon, acknowledgements, and preface — are about framing or positioning the content, preparing the reader for the encounter. Digital projects are different.

The experience of navigating through Anna Tsing's *Feral Atlas* is fluid, unpredictable, and even dangerous (2020). Returning to the opening page warns: "Revert at your own risk." Alisa Lebow structures *Filming Revolution* as a network, drawing the connections within a diverse and complex mix of voices and actions (2018). Samuel Liebhaber's project about the oral tradition of the Mahra is presented as a dynamic graphic guide to exploring their poetic expression (2018). *Chinese Deathscape* uncovers layers of history through digital maps (2019). *Constructing the Sacred* invites the reader into the three-dimensional space of Saqqara, Egypt necropolis (2020). And in *Black Quotidian*, Matthew Delmont takes the reader through the archives of Black newspapers to provide a deep and rich history of Black America that undergirds the stories of singular heroic figures (2019). *Enchanting the Desert*, like the lantern slides that inspired it, presents a faceted view of the human geography of the Grand Canyon (2016).

When project teams convened in 2017 for an All Projects meeting at the Mellon Foundation, it was an opportunity for the Stanford team to add the needs of authors to the considerations of those who were working on authoring, publishing, and reading platforms. With the other projects, discussion of preservation and archivability began and ended with either ePub, HTML, or, in the case of Scalar, RDF wrapped in XML. Preservation for the digital projects we are publishing is a commitment to the unique form of scholarly communication that the project presents to readers, not just the technology format or platform. The experience and intellectual import of the work may be lost if we are only preserving the discrete content elements and code. The whole is greater than the sum of the parts of an interactive scholarly work. And, as Doug Armato, Director of the University of Minnesota Press, pointed out at the meeting, the ultimate purpose of the presses is to serve the authors. It is easy to get caught up in the details of a specific technology challenge and lose sight of the bigger picture.

Scholars are finding that multimodal digital formats afford important new opportunities for presenting their interpretation and making arguments. Projects conceived as digital rather than conceived as print in a digital format explore a rhizomatic structure. The new structure re-defines reading as well as writing—or making. The works tend to be non- or multi-linear,

encourage interaction, and integrate sources (data) and argument. These qualities encourage an experiential learning and knowing. In a digital project, for example, a map is not merely an included figure. The author is also sharing all of the data with the reader, as well as a platform for exploring relationships in the data: showing the reader a richer view of what the author sees. This rich context for the argument can be captivating, but it also provides the reader with more opportunities for critical engagement and reinterpretation.

When the Andrew W. Mellon Foundation funded the Stanford University digital projects program, it was clearly understood to be unlike any of the other efforts within the funding group. In fact the Mellon's Scholarly Communications director described Stanford's ISW project more than once as an outlier. What makes the program so different is that we are taking interactive digital projects built on *any* platform with the goal of publishing peer-reviewed, book-length works with clearly demonstrated intellectual merit. There are no restrictions placed on the authoring platform used because the emphasis is on the communication of ideas that cannot be expressed in print.

The digital publishing program honors the axiom from Johanna Drucker: "*The design of digital tools for scholarship is an intellectual responsibility, not a technical task* (emphasis in the original)" (2009). Though Drucker's exhortation was delivered in the context of tools and platforms for research, it is equally applicable to this new form of scholarly communication. And yet, there are technical constraints for those of us concerned with the longevity of these projects. That includes the authors, publishers, librarians and archivists. Despite the frictionless feel of digital projects, the computing environment in which they are created, served, stored, and preserved is very physical. Compute takes time, space, and energy. The fluidity of the online experience belies the materiality of its production, maintenance, and preservation.

b) The production process

The production, or making, of a digital publication is very different from that of a traditional monograph. In publishing terms, production is a process conducted and coordinated by a press in which an author's completed text is copyedited, indexed, laid out for print, registered and cataloged for distribution, printed and bound. It's a process that can only be carried out once the manuscript is finished and the argument finalized. It's essentially preparation of a book for material assembly, and that assembly has been informed by centuries of practice and a set of conventions now expected by readers. But also inherent in the design of a book is the limitation of its lifespan. A book must be constructed of lasting material if it is expected to persist on a shelf or an archive.

Likewise, a publication meant to persist in digital form must be built with longevity as a key pillar of its form. But whereas a publisher has nearly complete control of how a book is laid out, coded, and prepared for physical construction--and can do all of this after the content has been almost fully written, reviewed, and edited--when a digital project is designed by an author in concert with its content, a publisher enters the production process much later in the game. An author has often spent years considering, developing, and implementing the interactive delivery of the argument because that form is itself part of the argument. In some cases this development coincides with the press's consultation, but oftentimes it's been under way, with the aid of funding that has already been poured into it, long before the press

can inform its development. In this sense, by the time a project is acquired, reviewed, and transmitted from acquisition to production in a publishing workflow, it is often too late to manage the traditional aspects of production in the typical connotation of the term. Instead, production takes on a new meaning, and it's one that's ineluctably tied to what it means to preserve both the form and content of this new form of scholarship.

A press's production process for digital publications is therefore not so much a design of the final format as it is an *intervention* on what forms have already been developed by the author so that they can be supported with minimal updating by the press's hosting infrastructure and then rolled over to archiving agents, like libraries and archives, with little to no remediation for use by future researchers. Ideally, these two destinations would be able to support the same architecture, but in reality the potentials and limitations of each are very different, making it difficult to produce a publication that is both highly interactive and also amenable to current digital archiving systems. The requirements of preservability often put tension on authors' desire to exploit the potentiality of web applications and readers' expectations for immersive reading in digital modes. Therefore the publisher must balance a razor's edge between enabling innovative, provocative, challenging scholarly argument and ensuring a durable, persistent, sustainable scholarly record.

A key example of this tension can be found in the difference between optimizing a publication for the live web versus optimizing it for web archiving. A publication that is developed with the aim of quick execution in a browser making calls to a server will use minified scripts that may themselves be generated dynamically as a user interacts with the content, ensuring a machine can quickly present the desired content. The effect of this for web archiving, though, can be debilitating: a script that is generated dynamically on a server may need to be rewritten for a web archive player that might not have all the components of a url necessary to find and display the embedded assets. Most javascript is itself a challenge for web archiving frameworks even though it's at the heart of many interactive web applications. Likewise a database framework may offer convenience for authoring and editing but may bog down a server that, when the content is finalized and no longer needs to be edited, still needs to call up and reconstruct various nodes of information that could be served more quickly from a single static data file. But when it comes to human readability both formats leave something to be desired for a researcher digging through a digital repository collection a decade or century down the road when modes of reading have evolved far beyond today's web frameworks altogether.

So the question becomes: if the production process is meant to be the process by which a publication's final output is materialized, which output should be guiding that process? The immediate publication which will be hosted on the publisher's server, a platform which itself is ever evolving but could theoretically be emulated or virtualized to provide future readers with the full experience of the publication (Rosenthal, 2015)? The web archive which seems to be the most logical solution to persisting the high-fidelity experience of web-based content, but whose form itself imposes limitations on the use of highly interactive web functionality? The digital repository collection made up of the thousands of media and code files and configurations which must be reassembled by a researcher with the help of exhaustive documentation that must be produced alongside the publication and anticipate the questions and needs of future readers?

With limited time and resources available in a press publishing workflow, we have so far needed to prioritize the immediate hosted version in the production process but have developed guidelines and assessments that anticipate preservation formats.

The first stage of what is considered the production process for digital publications is an assessment of the hosting needs and archivability of a project that has been accepted for publication. The assessment takes the form of a questionnaire developed in consultation with both Stanford Libraries and Reclaim, SUP's contracted hosting provider. It solicits information on a range of technical specifications. We ask for samples of code or access to any existing repos, a listing of platform or framework dependencies, external data or media resources embedded in the project, and configurations for the development server. Further questions survey the status of any existing backups, use of unique state urls, adherence to accessibility standards, availability of site maps, and level of documentation all in an effort to anticipate any additional work the press will need to do during production to ensure archiving efforts will be well under way by the time of publication. The answers to all these questions are often unknown to the authors themselves, and completion of the form usually necessitates conversation with the author's development team or service providers, foregrounding the collaborative nature of digital projects. So aside from providing technical specs, the questionnaire serves as an entry point to a network of contributors who will need to be consulted as production and archiving work continues over the following months.

The questionnaire is only one part of a guidelines package that we provide to authors when their project is nearing readiness for production work. These guidelines serve less as a demand for certain requirements to be met (again, much of the *making* has already been done) and more as a statement on best practices and caveat for the longevity an author might expect for their project in relation to the choices made during development. The package—which contains individual guides for color, file names and formats, fonts, images, links, labor and data citation, and browser and display interoperability—is prefaced by an overview on archivability, which begins by appealing to the author to consider their own role in the expected longevity of their project:

Stanford University Press shares with you the belief that digital academic arguments deserve treatment in the scholarly record equal to that of traditional print media, and we have a serious commitment to ensuring the longevity of your work through responsible approaches to sustainability and archiving. Your conscientious authoring and design choices are the first step toward sustaining the long-term fidelity of your work. But because we acknowledge that even the most sustainably designed web interfaces are susceptible to normal decay, we must anticipate the necessity of preserving your project's content so that readers and researchers can access it even after changing web environments begin to disrupt the project's dynamic and interactive features. To that end, we intend to take multiple approaches to archiving your content and preserving documentation of its intended presentation.

Unlike a typical monograph publication, for which the publisher implements the material architecture that contains the content, you, the author, hold the creative responsibility for your project's format, design, and execution. Nevertheless, as your publisher, we share the responsibility of the sustainability of your work and therefore recommend, encourage, and require certain technology and design standards. We

have compiled these archivability guidelines, along with the enclosed style sheets to help you make informed decisions on the types of technological considerations that will impact your project's longevity, and its amenability to current web archiving standards and practices.

In addition to your use of these recommendations, it is important you complete the technical questionnaire so we may begin developing the best publishing plan for your project, including its most suitable archiving environment. (Stanford University Press, 2017)

With the technical specifications cataloged and the recommendations shared, the production process then proceeds with migration of the project to the press's publication server where further work, including copyediting, technical cleanup, and assignment and implementation of registration identifiers will take place. The closest corollary to migration in the traditional publishing workflow is transmittal, the process by which a manuscript is transferred to the production editor who will then code the text for use in the press systems for copyediting, design, composition, registration, and finally distribution.

The migration process necessarily looks different for every project. *Chinese Deathscape*, which uses a Ruby on Rails framework, for example, took a collaborative effort between production, the author's development team, and even the founder of the press's contracted hosting service. It involved the developers writing a deployment script that could be executed by the press repeatedly as project edits continued. It meant also moving external media assets to the local project directories and adding branding and other design tweaks to the code base.

In the case of *Constructing the Sacred*, migration involved installing and configuring the Scalar content management system and loading the exported RDF-XML file from the development version of the project. This project also required ingestion of hundreds of media and data files into the Stanford Digital Repository, an approach that meant these assets were being called into the live publication from an external source but that those assets were also already pre-loaded into what would eventually serve as part of the the publication's static archive collection at the library. Still another project, *Feral Atlas*, requires its database backend be pared down to a static JSON data file that can be read by a revised frontend that was originally coded to manipulate data hosted and organized by a proprietary content management company. An added benefit of the static output is archive harvesting becomes somewhat easier as content is not being dynamically populated from multiple sources at once.

In all of these examples, what is exposed is that most authors and developers of digital content use, for better or worse, the systems available to them at their institutions and/or those that fall within their technical comfort zones. Many are not considering the dependencies inherent in those systems or the limitations they impose on archiving processes like crawling or content transfer necessary for web archiving or accessioning into long term storage for access by researchers. Though ideally migration presents an opportunity to make a limited number of structural revisions, it is still difficult to impose too many new demands on a project at such a late stage in the publishing workflow. But what edits *are* manageable are cataloged for technical cleanup, a process that happens later in

production once even more necessary changes are exposed through the processes of copyediting and proofing.

Some of the other production processes retain similarities to the traditional book production process: text is copyedited (though it's done in a platform, not usually in Word, requiring copyeditors with a level of digital literacy beyond typical editing tools), media captions edited to meet permissions and use requirements, cover and design implemented (though by freelance web designers, not in-house cover artists and compositors), and registration acquired (though not from Library of Congress but instead DOI providers and OCLC catalogers). Metatags for web indexing are integrated into HTML heads and terms of use pages loaded, processes somewhat analogous to preparing CIP and copyright statements in books. And at the heart of most of these endeavors is the goal of fixity. We are ultimately working to finalize something that resists finality. Whereas the digital publication presents itself as a kind of performance in the theater of contemporary reading technologies, our goal is to capture its presentation in a way that allows repeatable and sustained use.

c) Getting out in front of a new form

In about 500 BCE in Kermanshah province in modern day Iran an inscription was carved high up on a rock face, large enough to be visible from the road — an ancient highly travelled road connecting the capitals of Babylonia and Media. The text is presented in three languages: Old Persian, Elamite and Babylonian. It includes illustrations. We might say that the Bisotun inscription has all of the characteristics that define what authors want from print publications: prominence, available in translation, durable, with illustrations. The digital projects that authors now want to produce are far more likely to meet the fate of the pioneer Vectors Journal projects:

“Due to technological obsolescence and other factors, a number of *Vectors Journal* projects have become unavailable since their original publication.”

That notice on the website includes a link URL for <http://vectorsjournal.org/> that is no longer owned by the Vectors Journal but points to something suspiciously illegitimate but also called Vectors Journal.

There has been significant continuity over the many changes in print since the inscription at Bisotun. Born-digital documents have quietly disrupted the process from production through publication and preservation. The notion of ‘final copy’ for example, is changing. Tracking digital files remains a challenge. And, in some instances, Amazon.com is relied upon as the archive. (Can we find a print copy and scan it?) The move to interactive digital scholarly communication that we have described above may carve out an entirely new vein of academic publication that is not about longevity and durability, but that embraces the generative, renewable, evolving and uncertain qualities that this new format affords.

Authors see publication as legitimizing and, somehow, keeping their work alive. But for how long? This is a new question for authors. How can libraries and archives anticipate and influence the preservability? We have demonstrated how these new works redefine the notion of publication. Digital projects also confuse our assumptions about what it means to

archive versus maintain a project. And the projects introduce new conditions for preservation: How do we mitigate inevitable security concerns as software evolves? How can we manage the cost of keeping it alive (care and feeding)? Who is going to pay for it?

References

Anderson, S., & McPherson, T. (2011). Engaging digital scholarship: Thoughts on evaluating multimedia scholarship. *Profession*, 136-151. <https://doi.org/10.1632/prof.2011.2011.1.136>

Bauch, N. (2017). Guest editor's note: Process in digital geohumanities. *International Journal of Humanities and Arts Computing*, 11(1), 16-19.

Bauch, N. (2016). *Enchanting the desert*. Stanford University Press.
<http://doi.org/10.21627/2016etd>

Borgman, C. L. (2010). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4).

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital Humanities*. MIT Press.

Burdick, A., & Willis, H. (2011). Digital learning, digital scholarship and design thinking. *Design Studies*, 32(6), 546-556. <https://doi.org/10.1016/j.destud.2011.07.005>

Delmont, M. (2019). *Black quotidian: Everyday history in African-American newspapers*. Stanford University Press. <http://doi.org/10.21627/2019bq>

Drucker, J. (2009). Blind spots. *Chronicle of Higher Education*, 55(30).

Lebow, A. (2018). *Filming revolution*. Stanford University Press.
<http://doi.org/10.21627/2018fr>

Lebow, A. (2016). Seeing revolution non-linearly: www.filmrevolution.org. *Visual Anthropology*, 29(3), 278-295.

Liebhaber, S. (2018). *When melodies gather: Oral art of the Mahra*. Stanford University Press. <http://doi.org/10.21627/2018wmg>

Liebhaber, S. (2010, January). Written Mahri, Mahri fuṣ ḥ ā and their implications for early historical Arabic. In *Proceedings of the Seminar for Arabian Studies* (pp. 227-232). Archaeopress.

McPherson, T. (2017). From Vectors to Scalar: A brief primer for applied media studies. In *Applied Media Studies* (pp. 48-59). Routledge.

Mullaney, T. S., Henriot, C., Snyder-Reinke, J., McClure, D. W., & Worthey, G. (2019). *The Chinese deathscape: Grave reform in modern China*. Stanford University Press.
<http://doi.org/10.21627/2019cd>

Rosenthal, D. S. (2015). Emulation & virtualization as preservation strategies. *Report commissioned by The Andrew W. Mellon Foundation.*

Rosner, D. K. (2018). *Critical fabulations: Reworking the methods and margins of design.* MIT Press.

Rumsey, A. S. (2016). *When we are no more: How digital memory is shaping our future.* Bloomsbury Publishing USA.

Stanford University Press. (2017). Project preparation: Archivability.
<https://www.sup.org/digital/authors/current/docs/Archivability.pdf>.

Sullivan, E. A. (2020). *Constructing the sacred: Visibility and ritual landscape at the Egyptian necropolis at Saqqara.* Stanford University Press.
<http://doi.org/10.21627/2020cts>

Sullivan, E. A. (2017). Seeking a better view: Using 3D to investigate visibility in historic landscapes. *Journal of Archaeological Method and Theory*, 24(4), 1227-1255.

Tsing, A. (2020). *Feral atlas: The more-than-human anthropocene.* Stanford University Press. <http://doi.org/10.21627/2020fa>.

Tsing, A., Deger, J., Keleman, A., & Zhou, F. (2019). Feral Atlas and the more-than-human anthropocene. In *conference Art and the Anthropocene, June* (pp. 7-9).