ENSURING ENDURING DIGITAL ACCESS

PRINCIPLES AND APPROACHES TO DIGITAL PRESERVATION

Trevor Owens, 2019 Federal Depository Library Conference,
Arlington VA, October 21


United States
Atlantis
2
SIDE 1
3

LEAD

What We Post Online Is Forever, and We Need a Reminder.

What you post is tied to you forever.

By Meredith Fineman *Founder, FinePoint* **@MeredithFineman**

Science, Tech & Environment

Scientists warn we may be creating a 'digital dark age'

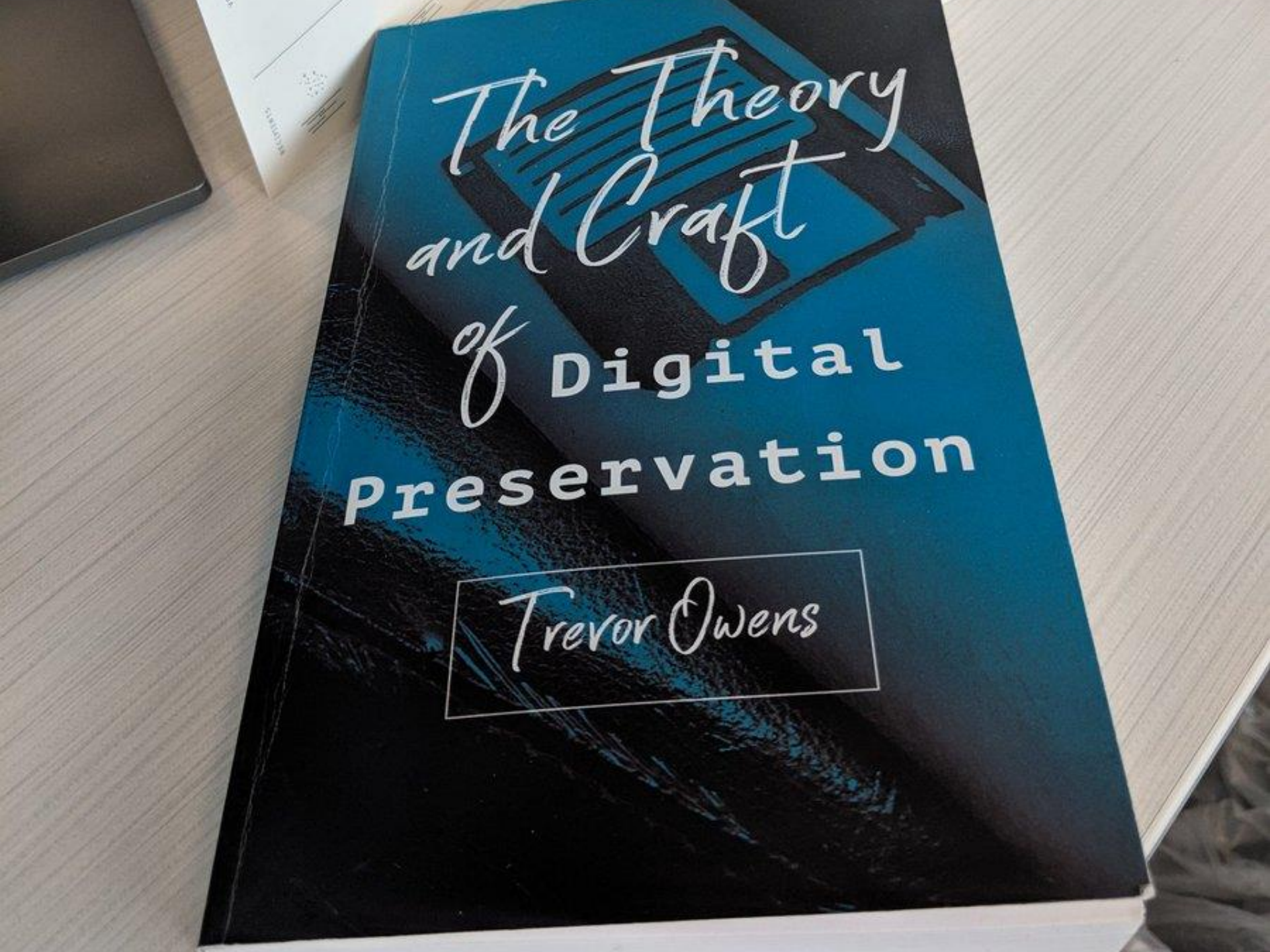
Science Friday

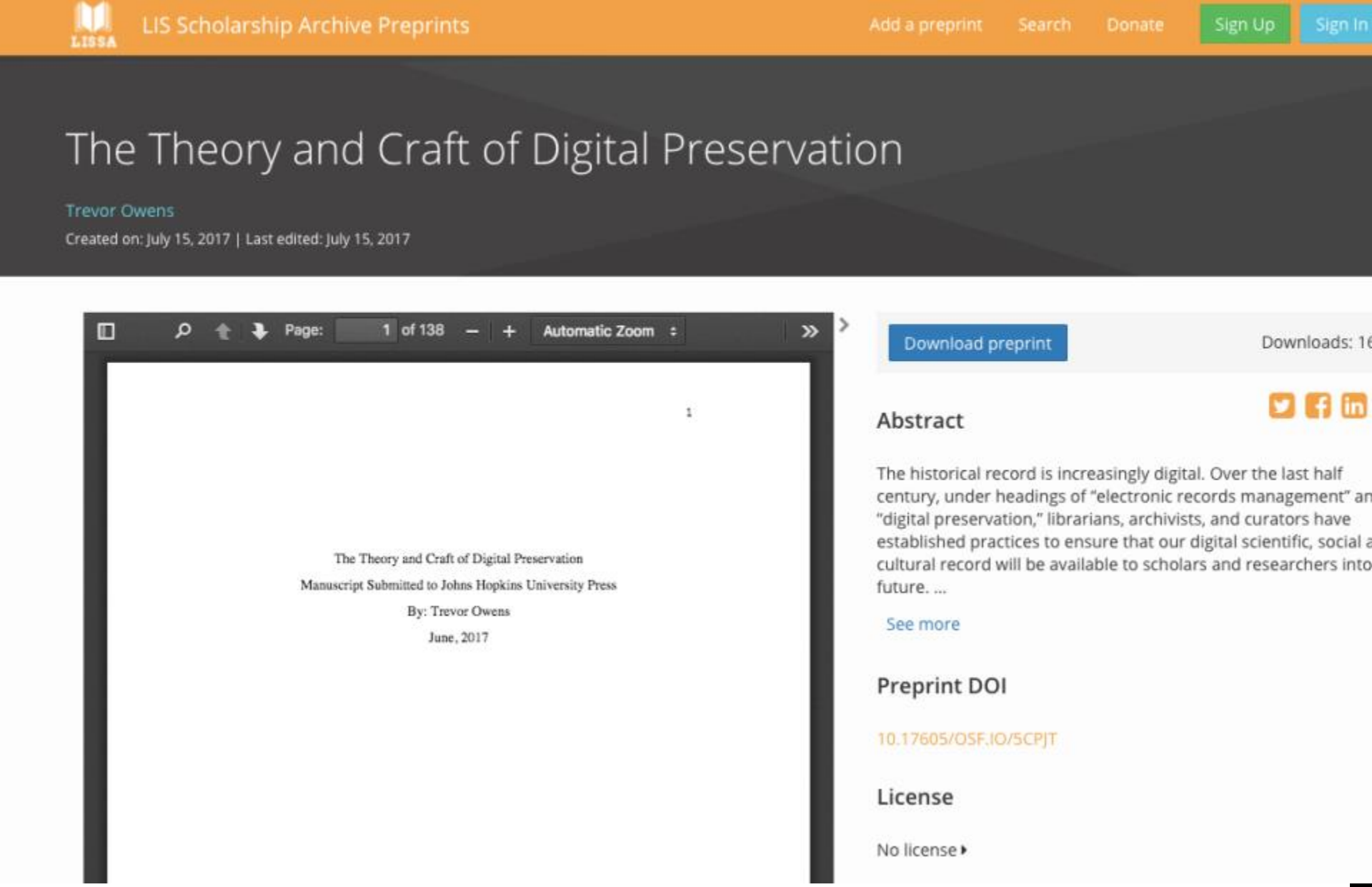
January 01, 2018 · 11:30 PM EST

Writer Adam Wernick

Producer Christopher Intagliata

Listen to the full interview.


The Theory and Craft of Digital Preservation
Trevor Owens



The Theory and Craft of Digital Preservation
https://osf.io/preprints/lissa/5cpjt/

4363

ress Cataloging-in-Publication Data

evor, author.
d craft of digital preservation / Trevor Owens.
e : Johns Hopkins University Press, 2018. | Includes
ences and index.
010715 | ISBN 9781421426976 (paperback : alk. paper) |
perback : alk. paper) | ISBN 9781421426983 (electronic)
tronic)
servation. | Digital libraries—Management.
C65 O94 2018 | DDC 025.8/4—dc23
lccn.loc.gov/2018010715

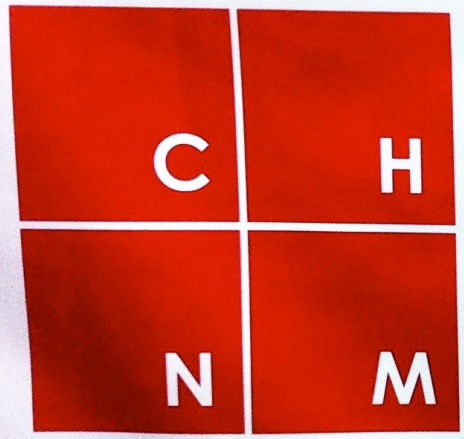
available from the British Library.

bulk purchases of this book. For more information,
16-6936 or specialsales@press.jhu.edu.

vironmentally friendly book materials, including
at least 30 percent post-consumer waste,

TALK ROADMAP

- My background and orientation to digital preservation
- A short definition of digital preservation
- 16 axioms for digital preservation
- A short list of steps and resources people around the world can use to improve digital preservation.



Center for History and New Media

Building a Better Yesterday, Bit by Bit

chnm.gmu.edu

NATIONAL DIGITAL STEWARDSHIP ALLIANCE

working together to preserve
the nation's digital information

digitalpreservation.gov/ndsa

TREVOR OWENS SELECTED AS IMLS SENIOR PROGRAM OFFICER, NATIONAL DIGITAL PLATFORM

Monday, January 5, 2015
FOR IMMEDIATE RELEASE

IMLS Press Contact
202-653-4799
Giuliana Bullard gbullard@imls.gov

Trevor Owens Selected as IMLS Senior Program Officer, National Digital Platform

Owens to head National Digital Platform responsibilities across programs at IMLS

Washington, DC—Institute of Museum and Library Services (IMLS) Director Susan H. Hildreth announced today that Trevor Owens has been selected to be the Senior Program Officer with responsibility for the national digital platform in the Office of Library Services.

IMLS Deputy Director for Library Services, Maura Marx, said, "We are delighted that Trevor has accepted this position. His hands-on experience working with online platforms and delivery systems, his background in community building and engagement, and his strategic vision will be great assets to our Discretionary and


It is possible for every libra
the country to leverage and
benefit from the work of oth
libraries in shared digital
services, systems and
infrastructure.
IMLSfocus

Q&A with Trevor Owens, LC Head of Digital Content Management

By Lisa Peet on December 21, 2017 Leave a Comment

In August, Trevor Owens became the first head of digital content management in library services at the Library of Congress (LC). The role at LC represents something of a full circle for Owens; before serving in various roles at the Institute of Museum and Library Services (IMLS) from 2015–17, most recently as acting associate deputy director for libraries, Owens was a digital archivist at LC's Office of Strategic Initiatives since 2010. The position, on the other hand, is brand new—both to Owens and to LC.

Trevor Owens
Photo by Shawn Miller

Library Journal caught up with Owens to find out more about how his career path brought him back to LC and his plans for the library's new division.

***LJ*: What does the head of digital content management do?**

RECENT & POPULAR

Recent Posts | Most Commented | Most Viewed

A Baby's Bones

The Week Ends with Book Awards and Title Announcements | Book Pulse

The Heart of Transformation:

https://lj.libraryjournal.com/2017/12/people/qa-with-trevor-owens-lc-head-of-digital-content-management/#_

DIGITAL HISTORY METHODS

HIST 477/677 American University

MARCH 31, 2019 BY SJONES

Curation and Organization of Digital Archives

"Big Data, Little Narration" is a transcript on digital archives and presentation by Dragan Espenschied, who defines himself as an "electronic musician and internet artist."

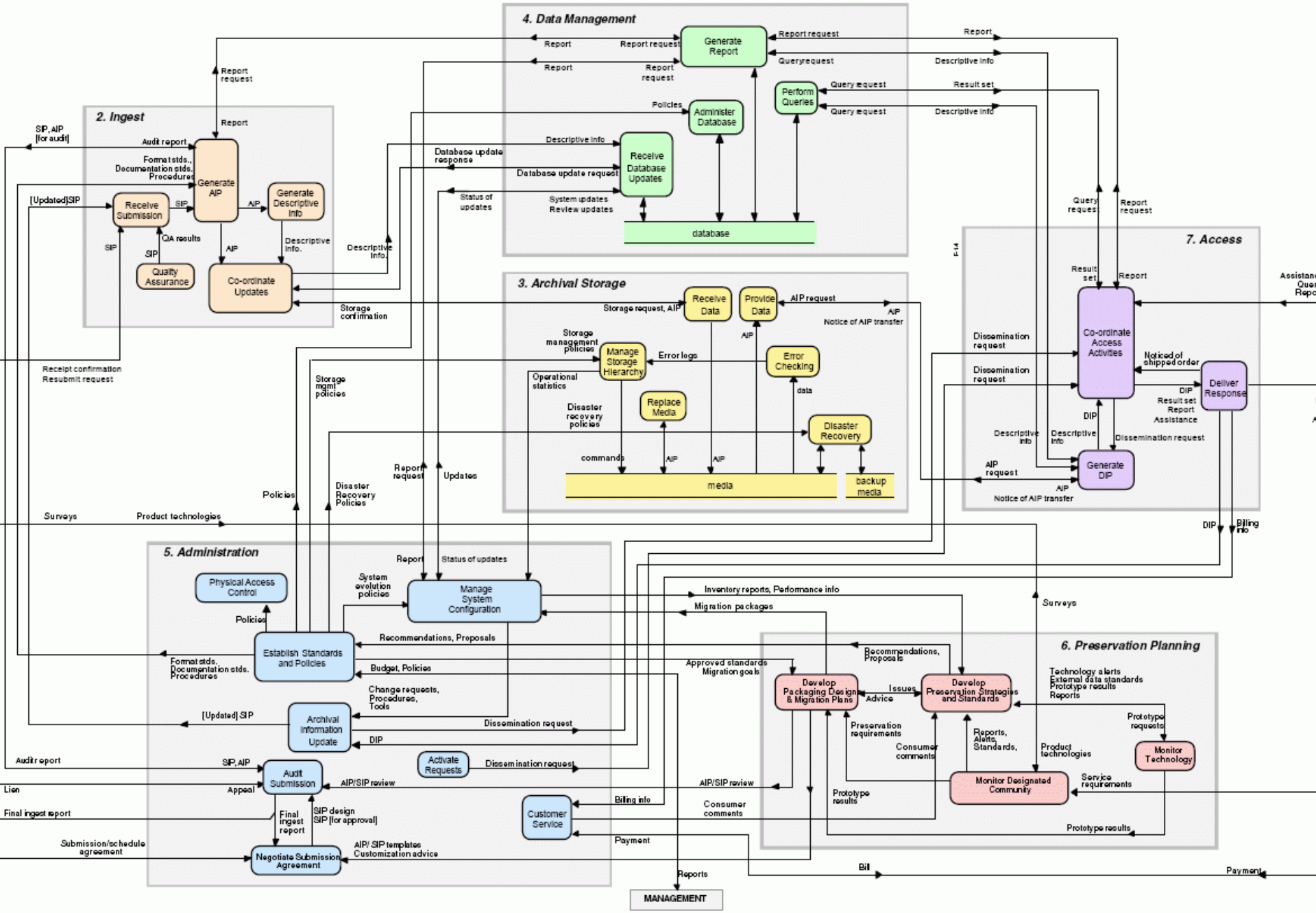
In the transcript, Espenschied discusses the differences between performance and activity, with performance being the thing that the computer does, and the activity being what the user does. However, somewhere between the two, Espenschied highlights that there can often be a disconnect. In demonstrating this, the author uses a

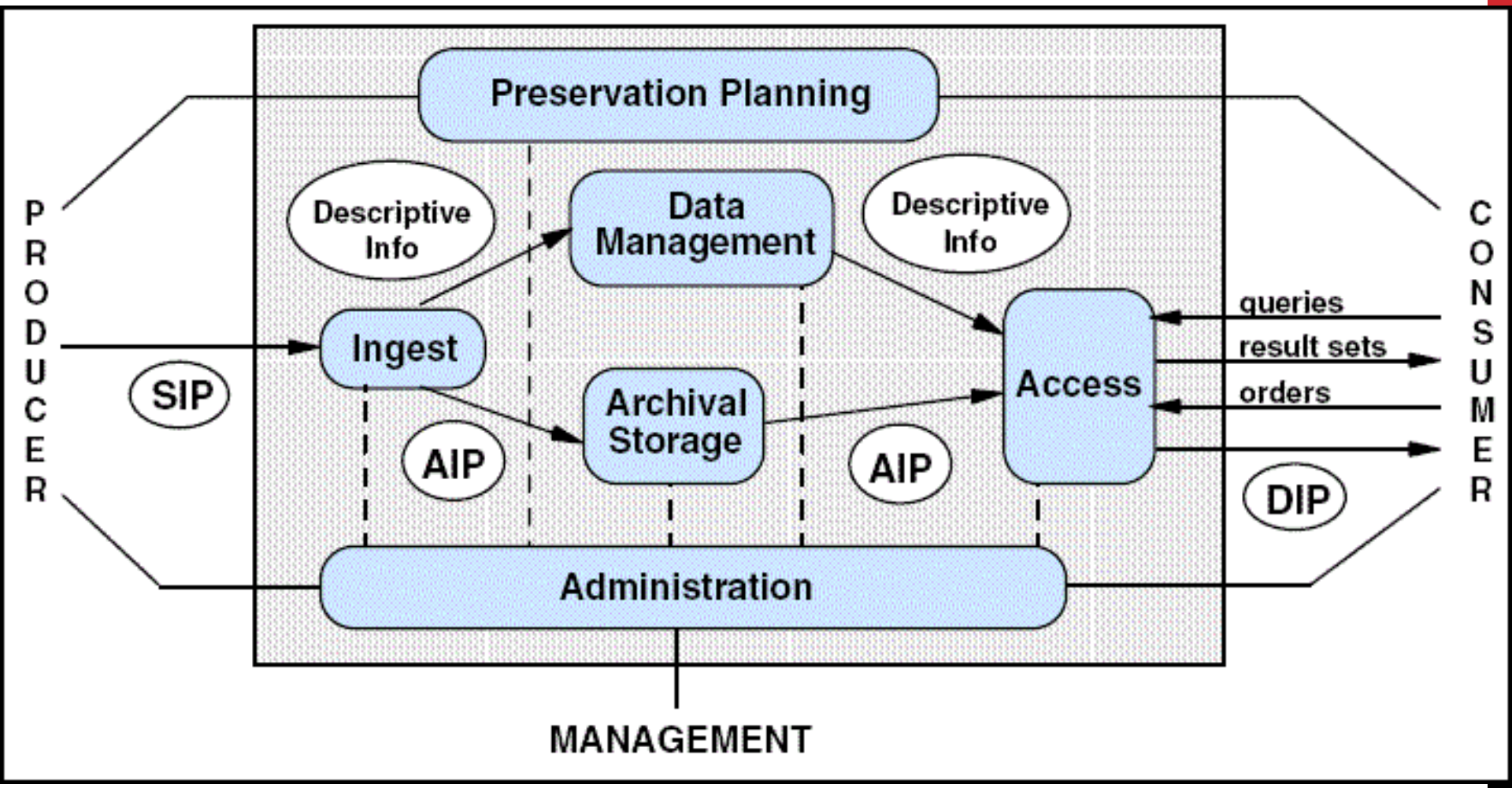
DIGHIST IS...

The course blog for *Digital History Metho* graduate seminar at American University the explicit goals of this course is for us t develop as communicators on the public please do join our conversation, but pleas respectfully. We are all learning how to d together.

Header image Highsmith, Carol M, *Play s at a children's computer center in Rockvi Maryland*.

WHAT IS AND ISN'T DIGITAL PRESERVATION


4. Data Management
Generate Report
Report request
Report
Query request
Descriptive info
Perform Queries
Result set
Administer Database
Policies
Receive Database Updates
Descriptive info
Database update response
Database update request
Status of updates
System updates
Review updates
database
2. Ingest
Report request
SIP, AIP [for audit]
Audit report
Format stds., Documentation stds., Procedures
Generate AIP
Generate Descriptive Info
[Updated]SIP
Receive Submission
SIP
AIP
QA results
Quality Assurance
Descriptive Info.
Co-ordinate Updates
Storage confirmation
Receipt confirmation
Resubmit request
3. Archival Storage
Receive Data
Provide Data
AIP request
Storage request, AIP
Notice of AIP transfer
Storage management policies
Manage Storage Hierarchy
Error logs
Error Checking
Operational statistics
Replace Media
Disaster recovery policies
Disaster Recovery
commands
media
backup media
Storage mgmt policies
Report request
Updates
Disaster Recovery Policies
7. Access
Result set
Report
Co-ordinate Access Activities
Dissemination request
Noticed of shipped order
Deliver Response
DIP
Result set
Report
Assistance
Descriptive info
Generate DIP
AIP request
Notice of AIP transfer
Surveys
Product technologies
Billing info
5. Administration
Physical Access Control
System evolution policies
Manage System Configuration
Inventory reports, Performance info
Migration packages
Policies
Establish Standards and Policies
Recommendations, Proposals
Format stds. Documentation stds. Procedures
Budget, Policies
Change requests, Procedures, Tools
Archival Information Update
[Updated] SIP
Dissemination request
DIP
Activate Requests
Audit report
SIP, AIP
Audit Submission
AIP/SIP review
Appeal
Lien
Final ingest report
Final ingest report
SIP design
SIP [for approval]
Submission/schedule agreement
Negotiate Submission Agreement
AIP/ SIP templates
Customization advice
Customer Service
Billing info
Payment
Consumer comments
6. Preservation Planning
Develop Packaging Design & Migration Plans
Develop Preservation Strategies and Standards
Issues
Advice
Approved standards
Migration goals
Recommendations, Proposals
Technology alerts
External data standards
Prototype results
Reports
Preservation requirements
Consumer comments
Reports, Alerts, Standards,
Prototype requests
Monitor Technology
Product technologies
Monitor Designated Community
Service requirements
Prototype results
Prototype results
Reports
MANAGEMENT
Bill
Payment


Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
AIP
Archival Storage
AIP
Access
queries
result sets
orders
DIP
CONSUMER
Administration
MANAGEMENT

DIGITAL PRESERVATION IS WORKING TO ENSURE ENDURING ACCESS TO DIGITAL CONTENT

16 AXIOMS OF DIGITAL PRESERVATION

1. A REPOSITORY IS NOT A PIECE OF SOFTWARE

Software cannot preserve anything. Software cannot be a repository in itself. A repository is the sum of financial resources, hardware, staff time, and ongoing implementation of policies and planning to ensure long-term access to content. Any software system you use to enable you preserving and providing access to digital content is by necessity temporary. You need to be able to get your stuff out of it because it likely will not last forever. Similarly, there is no software that “does” digital preservation.

2. INSTITUTIONS MAKE PRESERVATION POSSIBLE

Each of us will die. Without care and management, the things that mattered to us will persist for some period of time related to the durability of their mediums. With that noted, the primary enablers of preservation for the long term are our institutions (libraries, archives, museums, families, religious organizations, governments, etc.) As such, the possibility of preservation is enabled through the design and function of those institutions. Their org charts, hiring practices, funding, credibility, etc. are all key parts of the cultural machinery that makes preservation possible.

3. TOOLS CAN GET IN THE WAY JUST AS MUCH AS THEY CAN HELP

Specialized digital preservation tools and software are just as likely to get in the way of solving your digital preservation problems as they are to help. In many cases, it's much more straightforward to start small and implement simple and discrete tools and practices to keep track of your digital information using nothing more than the file system you happen to be working in. It's better to start simple and then introduce tools that help you improve your process then to simply buy into some complex system without having gotten your house in order first.

4. NOTHING *HAS BEEN* PRESERVED, THERE ARE ONLY THINGS *BEING* PRESERVED.

Preservation is the result of ongoing work of people and commitments of resources. The work is never finished. This has significant ramifications for how we think about staffing and resourcing preservation work. If you want to evaluate how serious an organization is about digital preservation don't start by looking at their code, their storage architecture, or talking to their developers. Start by talking to their finance people.

5. HOARDING IS NOT PRESERVATION.

It is very easy to start grabbing lots of digital objects and making copies of them. This is not preservation. To really be preserving something you need to be able to make it discoverable and accessible and that is going to require that you have a clear and coherent approach to collection development, arrangement, description and methods and approaches to provide access.

6. BACKING UP DATA IS NOT DIGITAL PRESERVATION.

If you start talking about digital preservation and someone tells you “oh, don’t worry about it, we back everything up nightly” you need to be prepared to explain how and why that is not digital preservation. It’s important to develop your explanation of what the differences are. Many of the aspects that go into backing up data for current use are similar to aspects of digital preservation work but the near term concerns of being able to restore data are significantly different from the long term issues related to ensuring access to content in the future.

7. THE BOUNDARIES OF DIGITAL OBJECTS ARE FUZZY.

Individual objects reference, incorporate and use aspects of other objects as part of their everyday function. You might think you have a copy of a piece of software by keeping a copy of its installer, but that installer might call a web service to start downloading files in which case you can’t install and run that software unless you have the files it depends on. You may need a set of fonts, or a particular video codec, or any number of other things to be able to use something in the future and it is challenging to articulate what is actually inside your object and what is external to it.

8. ONE PERSON'S DIGITAL COLLECTION IS ANOTHER'S DIGITAL OBJECT IS ANOTHER'S DATASET.

In some cases the contents of a hard drive can be managed as a single item, in others they are a collection of items. In the analog world, the boundaries of objects were a little bit more straightforward or at least taken for granted. The fuzziness of boundaries of digital objects means that the concept of "item" and "collection" is less clear than with analog items. For example, a website might be an item in a web archive, but it is also functionally a serial publication which changes over time, it is also a collection of files.

9. DIGITAL PRESERVATION IS MAKING THE BEST USE OF RESOURCES TO MITIGATE THREATS AND RISKS.

You are never done with digital preservation. It is not something that can be accomplished or finished. Digital preservation is a continual process of understanding the risks you face for losing content or losing the ability to render and interact with it and making use of whatever resources you have to mitigate those risks.

10. THE ANSWER TO NEARLY ALL-DIGITAL PRESERVATION QUESTION IS "IT DEPENDS."

In almost every case, the details matter. Deciding what matters about an object or a set of objects is largely contingent on what their future use might be. Similarly, developing a preservation approach to a massive and rapidly growing collection of high-resolution video will end up being fundamentally different to the approach an organization would take to ensuring long-term access to a collection of digitized texts.

11. IT'S LONG PAST TIME TO START TAKING ACTIONS.

You can read and ponder complicated data models, schemas for tracking and logging preservation actions, and a range of other complex and interesting topics for years but it's not going to help "get the boxes off the floor." There are practical and pragmatic things everyone can and should do now to mitigate many of the most pressing risks of loss. So be sure to prioritize doing those things first before delving into many of the more open-ended areas of digital preservation work and research.

12. HIGHLY TECHNICAL DEFINITIONS OF DIGITAL PRESERVATION ARE COMPLICIT IN SILENCING THE PAST

Much of the language and specifications of digital preservation have developed into complex sets of requirements that obfuscate many of the practical things anyone and any organization can do to increase the likelihood of access to content in the future. As such, a highly technical framing of digital preservation has resulted in many smaller and less resource rich institutions feeling like they just can't do digital preservation, or that they need to hire consultants to tell them about complex preservation metadata standards when what they need to do first is make a copy of their files.

13. THE AFFORDANCES OF DIGITAL MEDIA PROMPT A NEED FOR DIGITAL PRESERVATION TO BE ENTANGLED IN DIGITAL COLLECTION DEVELOPMENT.

Digital media affords significant new opportunities for engaging communities with the development of digital collections. When digital preservationists take for granted that their job is to preserve what they are given, they fail to help an organization rethink what it is possible to collect. Digital preservation policy should increasingly be directly connected to and involved in collection development policy. That is, the affordances of what can be easily preserved should inform decisions about what an organization wants to go out and collect and preserve.

14. ACCEPT AND EMBRACE THE ARCHIVAL SLIVER.

We've never saved everything. We've never saved most things. When we start from the understanding that most things are temporary and likely to be lost to history, we can shift to focus our energy on making sure we line up the resources necessary to protect the things that matter the most. The ideology of "the digital" makes it seem like we could or should attempt to save everything. However, this comes from the mistaken thinking that digital preservation is primarily a technical challenge instead of a social and ethical one.

15. THE SCALE AND INHERENT STRUCTURES OF DIGITAL INFORMATION SUGGEST WORKING MORE WITH A SHOVEL THAN WITH A TWEEZERS.

While we need to embrace the fact that we can't collect and preserve everything, we also need to realize that in many cases the time and resources it takes to make decisions about individual things could be better used elsewhere. It's often best to focus digital preservation decision making at scale. This is particularly true in cases where you are dealing with content that isn't particularly large. Similarly, in many cases it makes sense to normalize content or to process any number of kinds of derivative files from it and keep the originals.

16. DOING DIGITAL PRESERVATION REQUIRES THINKING LIKE A FUTURIST

We don’t know the tools and systems that people will have and use in the future to access digital content. So if we want to ensure long term access to digital information we need to, at least on some level, be thinking about and aware of trends in the development of digital technologies. This is a key consideration for risk mitigation. Our preservation risks and threats are based on the technology stack we currently have and the stack we will have in the future so we need to look to the future in a way that we didn’t need to with previous media and formats.

WHAT TO START TODAY?

SIX THINGS TO DO STARTING TODAY

1. **Identify what digital Stuff you've got that you need to keep**
2. **Get the digital boxes off the floor**
3. **Schedule out a plan for improving things and checking in**
4. **Read the NDSA levels of Digital Preservation Paper**
5. **Join the communities of practice**
6. **Read the Theory and Craft of Digital Preservation**

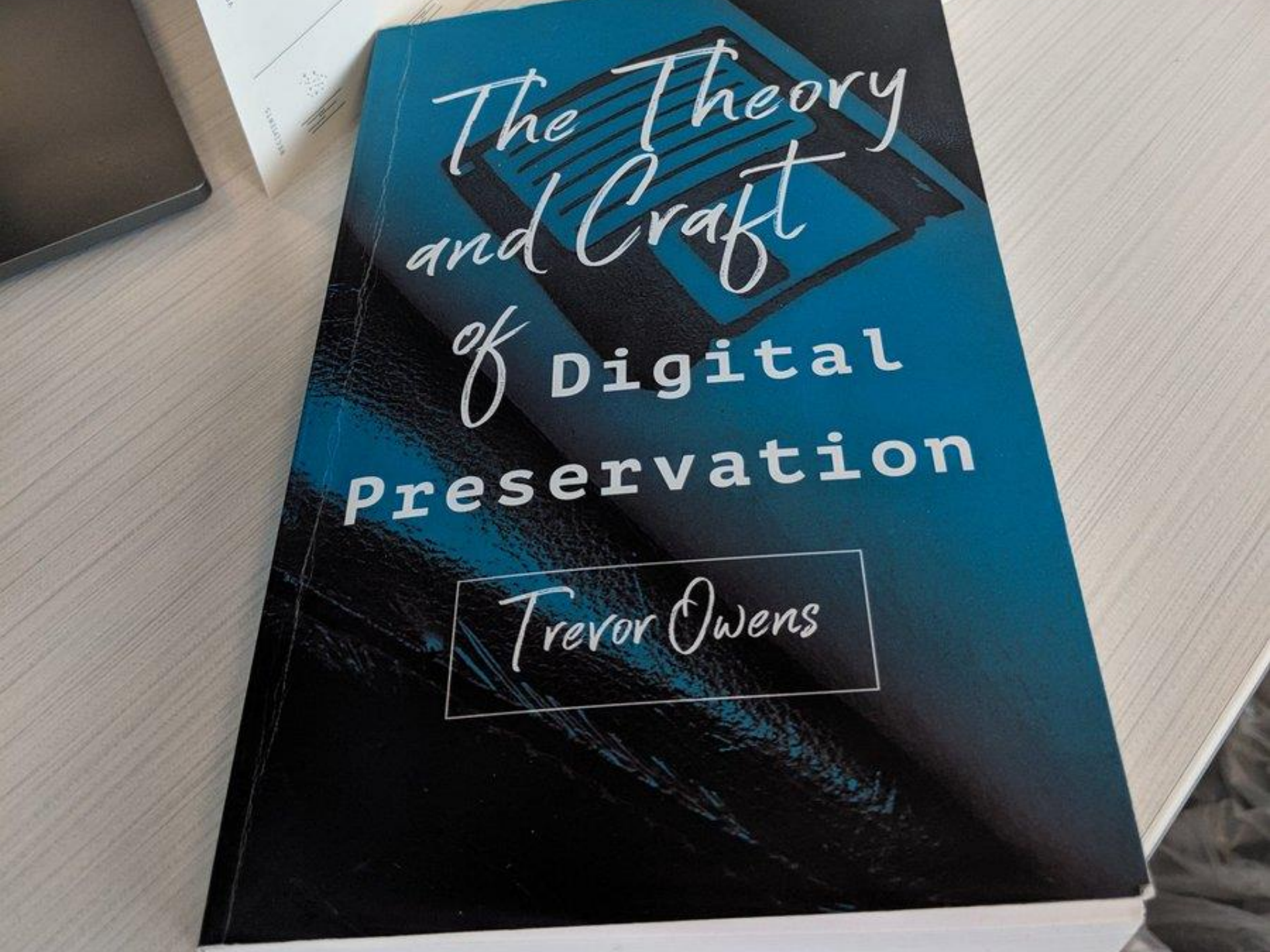
Levels of Digital Preservation

The "Levels of Digital Preservation" (PDF; *Proceedings of the Archiving (IS&T) Conference,* April 2013, Washington, DC) are a tiered set of recommendations for how organizations should begin to build or enhance their digital preservation activities. A work in progress by the NDSA, it is intended to be a relatively easy-to-use set of guidelines useful not only for those just beginning to think about preserving their digital assets, but also for institutions planning the next steps in enhancing their existing digital preservation systems and workflows. It allows institutions to assess the level of preservation achieved for specific materials in their custody, or their entire preservation infrastructure. It is not designed to assess the robustness of digital preservation programs as a whole since it does not cover such things as policies, staffing, or organizational support. The guidelines are organized into five functional areas that are at the heart of digital preservation systems: storage and geographic location, file fixity and data integrity, information security, metadata, and file formats.

It is expected that the Levels of Digital Preservation will be updated over time as additional feedback is received, experience is gained implementing its recommendations and as empirical research provides detailed information about data loss. For this reason, each iteration of the Levels will be versioned. Version 1 (PDF) is shown below.

Table 1: Version 1 of the Levels of Digital Preservation

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if	- Check fixity on all ingests - Use write-blockers when working with original media	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on	- Check fixity of all content in response to specific events or activities - Ability to


The Theory and Craft of Digital Preservation
Trevor Owens