Green Goes with Anything

Decreasing Environmental Impact of Digital Libraries at Virginia Tech

Alex Kinnaman
Virginia Tech
United States
alexk93@vt.edu
0000-0001-8943-8946

Alan Munshower
Virginia Tech
United States
alanmun@vt.edu
0000-0002-2878-5896

Abstract – This paper examines existing digital library practices at Virginia Tech University Libraries, and explores changes in documentation and practice that will foster a more environmentally sustainable collections platform.

Keywords – sustainability, digital libraries, digital preservation, archives, appraisal
Conference Topics – Environment

I. Introduction

As digital library practitioners, we are investigating ways to guide digital curation practices more broadly across Virginia Tech University Libraries (VTUL), while prioritizing considerations for environmental sustainability. In doing so, we explore university and professional standards and ethics, using the 2019 article "Towards Environmentally Sustainable Digital Preservation" [1] as a guide to focus on immediate areas that we can address in our digital library workflows. We investigate our workflows for appraisal, digitization, fixity checking, and storage choices to identify areas of improvement that find balance between best practices and environmental sustainability. This topic aligns with the conference theme Environment, and seeks to understand the environmental impact of VTUL's digital preservation choices on the community in which we live.

II. Literature Review

While federal action in the United States specifically addressing climate change has only emerged since the early 1990's, libraries and archives have been attentive to growing concern decades prior [2]. C. Durham writes, "all cultural institutions are vulnerable to other aspects of the Climate Emergency…[and] need to prepare and adapt for the world humanity has created for itself, and they need to prepare quickly" [3]. There is an evident impact of digital preservation activities on the environment.

Beginning in the 1980's, innovative concepts such as natural air-conditioning of paper materials underlied the environmentally friendly mission-specific work of lending and efficient management of physical materials [4]. Similar practices spread internationally to address conservation by using structural, rather than artificial means, to control the environment [5]. The digital age, and the accelerating proliferation of technology, has removed digital content managers from a similar physical awareness of their environmental impact in day-to-day work. Large datasets and complex digital objects are primary responsibilities of cultural heritage institutions, often with many parties involved in the accessioning, processing, and management. However, the effects of not triaging these processes through audit or inventory can be compounding. These necessary actions may be in conflict with an environmentally-sustainable approach to collection management.

Missions and Statements of Shared Value from professional organizations are valuable resources, as we look to others for guidance on a charge

iPres 2022: The 18th International Conference on Digital Preservation, Glasgow, Scotland.

DOI: 10.1145/nnnnnnn.nnnnnnn

towards more environmentally sustainable digital curation practices at VTUL. The Society of American Archivists (SAA) makes a clear case for green-focused practice, charging members to "Devise environmentally sustainable techniques for preserving collections and serving communities" [6]. There is an understanding of the balance of the ever-present dialogue with environmental considerations: "[D]eveloping acquisition, processing, storage, and service models—must necessarily involve an ongoing awareness of the impact of archival work on the environment" [6].

While not addressing Environmental Stewardship directly in its 2018 Declaration of Shared Values, the Digital Preservation Services Collaborative (DPSC) has listed sustainability as a core value. Partnering sustainability and affordability in the list core values, DPSC is presenting sustainability as a general duty in providing services, though the key value of accountability may also serve to guide decision making on climate policies and renewable energy options [7].

The National Archives and Records Administration of the United States (NARA), has created a climate action plan, specifically aimed at addressing "one of the most significant issues impacting…long term continuity" [8]. Among the plan's five action items is the "strengthen[ing] of NARA's climate resilience by leveraging cloud-based solutions." Benefits outlined include the safeguarding against weather events, a more secure data supply chain, and notably that a move to cloud systems "may ultimately reduce GHG [greenhouse gas] emissions due to consolidated cooling and controlling of the data centers" [8]. This is presented mainly as a hypothetical in the plan, not offering evidence for greenhouse gas reductions, other than demonstrating that a shift from in-person to virtual reading room practices will generally contribute to less emissions.

There is a growing corpus of scholars interested in further exploring the challenges of environmental stewardship and digital preservation. Most fundamental for the purposes of this article, is the work by K. Pendergras et al. "Toward Environmentally Sustainable Digital Preservation [1]." Critically, the authors parse out different types of sustainability efforts in the field, focusing their scholarship on environmental sustainability and digital preservation practices.

This comprehensive look at current practices provides a framework for organizations to shift towards environmentally sustainable goals. "[Cultural Heritage Organizations] need to reduce the amount of digital content that they preserve while reducing the resource-intensity of its storage and delivery. To do so, cultural heritage professionals must reevaluate their basic assumptions of appraisal, permanence, and availability of digital content" [1].

Recommended approaches for this paradigm shift include addressing appraisal, permanence, determination of acceptable loss, fixity check methods and frequency, choice of storage technologies, file format migration policies, and the number of redundant copies.

While K. Pendergrass et al. [1] offer a number of avenues to explore in their paradigm shift, much of the existing additional literature has an emphasis on storage and the raw energy consumption of large data sets. This concern frames the immediacy of the need to create sustainable practices.

> "Every decision to acquire, preserve, or replicate a byte of data is, essentially, a commitment to put some amount more carbon into the earth's atmosphere. This reality should prompt a meaningful though difficult conversation about whether the survival of knowledge into the distant future will be primarily dependent on deliberately preserving less of it at lower quality" [9].

Virginia Tech itself is located in Montgomery County, Virginia. Virginia has a long history of coal mining as a major economic backbone. Beginning in the late-18th century coal has been mined in portions of Montgomery and Pulaski counties [10] after which production of coal ebbed and flowed until it climaxed in 1943-44 and continued well into the 1960's. The worst but not only disaster on record occurred in April 1946 when a mine in McCoy, Virginia exploded from a methane leak and killed 12 miners, orphaning 51 children [11]. The relationship between the economy and industrial energy extraction in Virginia and in Montgomery County has lasted 250 years, and continues to be a primary source of income for the state[1] and a major cultural hub of the community.

[1] Virginia Coal: httpsrgy.vt.edu/coal.html

As a cultural institution in the middle of the primary location for coal mining in southwest Virginia, Virginia Tech plays a role in tracking energy consumption for the University. The Virginia Tech mission statement is as follows: "Inspired by our land-grant identity and guided by our motto, Ut Prosim (That I May Serve), Virginia Tech is an inclusive community of knowledge, discovery, and creativity dedicated to improving the quality of life and the human condition within the Commonwealth of Virginia and throughout the world." Improving the quality of life and the human condition applies to many facets of the University, including environmental sustainability. The Energy Patterns and Trends Electronic Database provides an authoritative resource on Virginia energy consumption. This supports the Virginia Department of Mines, Minerals and Energy and the Virginia Center for Coal and Energy Research in "responding to information requests from the general public and legislative bodies."[2]

Virginia Tech's Division of Campus Planning, Infrastructure, and Facilities' Office of Energy Management has established energy efficiency design guidelines to reduce electric and water usage during facility construction on campus.[3] They have also developed a 5 Year Energy Action Plan that ended in 2020 and supported the current iteration of the Virginia Tech Climate Action Commitment, which aims to set the university on a path to carbon neutrality by 2030. Virginia Tech releases Sustainability Annual Reports[4] to track progress on various sustainability projects. Progress is measured using the The Sustainability Tracking, Assessment & Rating System[5] from the Association for the Advancement of Sustainability in Higher Education.

Virginia Tech has a responsibility to engage with our history of industrial energy extraction and build better sustainability strategies into each aspect of our university. While also being a campus building consuming similar energy to other facilities on campus, VTUL is unique in its management of multiple stores of data in our institutional repositories, digital libraries, Special Collections and University Archives, and data repository. It is with this history and context in mind that we explore the current environmental impact of our digital library choices and recommendations for decreasing this impact through changes in our workflows.

III. Methodology

A. Appraisal and Digitization

Newly created digital collections at VTUL are mediated by a team of stakeholders from across library departments who review project proposals. Once approved, projects are managed by a dedicated Digital Imaging Coordinator. The core goal of content creation in the Digital Imaging Lab is to create Preservation Digital Objects (a TIFF) to serve as a surrogate to the original object. These goals are informed by the Federal Agencies Digital Guidelines Initiative (FADGI), Metamorfoze and ISO imaging guidelines. This includes not only resolution (PPI) and sharpness (sampling efficiency) requirements of the above standards but also the color accuracy and tonal accuracy requirements. By following FADGI guidelines, the Digital Imaging Lab strives to achieve consistent, repeatable, measurable digital files in an efficient and scalable manner. The FADGI Standard contains specific technical guidelines for a variety of formats. Below are the general guidelines for the TIF files captured in the Digital Imaging Lab which represent the majority of output as stored data.

Preservation File TIFF
File Type: Uncompressed TIFF
Color Depth: 24 bit Color RGB
File Compression: None
Bit Depth: 16 bit
PPI: 400
Color Profile: AdobeRGB (1998)

The latest approved revision of the FADGI guideline does not explicitly address environmental sustainability in the creation of preservation standards. On the limitations of its guidelines, the initiative defers that its quality standards are "…appropriate for most cultural heritage imaging projects, and takes into consideration the competing requirements of quality, speed of production, and cost [12]."

[2] Virginia Energy: https://vept.energy.vt.edu/index.html

[3] VT Energy Efficiency Design Guide: https://www.facilities.vt.edu/energy-utilities/energy-reduction-efforts/energy-efficiency-design-guidelines.html

[4] VT Sustainability Annual Reports: https://www.facilities.vt.edu/sustainability/sustainability-reports/virginia-tech-sustainability-annual-reports.html

[5] STARS: https://stars.aashe.org/

The Digital Imaging Lab has a production server that is backed up nightly, and upon completion, transfer the working file to the appropriate department for either metadata cleanup or deposit into the Digital Library Platform. With the variety of projects, some which may be hosted and managed by VTUL, and some that may not, there is a likelihood of redundancy in the transfer of ownership. More copies in more places is a tenant of digital preservation, but where do diminishing returns in the realms of security and preservation cross into harmful environmental practices?

B. Fixity

In addition to evaluating archival practices, we evaluated our digital preservation choices regarding fixity, including frequency and algorithm, and storage, including number of copies and general redundancy, and their relationship to one another. Both fixity checking and mid to long-term storage are ongoing services that result in continued energy consumption. Everything in a digital preservation and access system is by name, digital, and therefore requires some form of power. Ingest, fixity, restoration, migration, distributed storage, virus checking, file format verification, access, are all functions we include in our preservation system. When we evaluate the balance between what is important to us in our preservation system, we find that fixity and distributed storage are both necessary functionalities that may also allow for flexibility that could help decrease our carbon footprint. We choose these factors because we may not be able to control factors like necessity to migrate and number or frequency of access, but we can choose fixity frequency, appraisal of content, and the number of copies we choose to maintain.

According to the 2017 NDSA Fixity Survey, 84.1% of respondents indicated that they did utilize fixity information at some point in their workflows, though the methods, schedules, and reasons are widely varied [13]. Many digital preservationists have agreed that checksum computations are an intensive energy activity [1], and may not need to be performed as frequently as the field has been practicing [14]. This is because fixity checks need to open and read the entire file to produce an accurate checksum. While there is consensus that fixity should be performed regularly, neither the NDSA Levels of Preservation [15] nor the DPC's Digital Preservation Handbook [16] provide a best practice on the optimal frequency for scheduled fixity checks, but agree that any situation where a file is moved from one location to another should always have a fixity check. More frequent fixity checking leads to faster repair, but is energy intensive and can be cost-prohibitive especially in the cloud environment [17]. Comparatively, LOCKSS runs continuous fixity checks and uses a non-canonical fixity store [18], which requires less bandwidth as it relies on the multiple copies to self-heal rather than retrieving the entire document for a fixity check [14] to notify a manager of an error.

The Virginia Tech Digital Library Platform generates fixity at multiple points in the data lifecycle; pre-ingest, on ingest, and on a regular schedule.[6] We have two local servers, one of which is synced to Amazon Web Services (AWS) nightly, and one as-needed. We use the MD5 hash[7] because this is what AWS requires. Currently our AWS instance is set to run fixity on ingest and every 90 days. Our preservation storage services are the Academic Preservation Trust (APTrust)[8] and the MetaArchive Cooperative,[9] both with their own independent fixity policies. MetaArchive is built on LOCKSS, which runs fixity as needed in a non-canonical, self-healing fixity store [17]. We also use Figshare[10] to store our data repository. Figshare contracts with Chronopolis for preservation, and we ingest our datasets into APTrust.

The following section will refer to several energy units including millijoule (mj), watt-second (W*s), watts-hour (kWh), and megatonne (MT). It will also refer to carbon dioxide equivalent as CO2e. With an understanding of our fixity triggers and frequency, we investigated the estimated energy consumed from generating an MD5 hash. In a study examining energy measurements of standard security functions, [19] found that of a series of hash algorithms they explored, MD4 and MD5 were the least energy-consuming hash algorithms. This study examined the type of hash and the size of file, noting that "consumption increases with the size of the files." They found that a hash for a 10kb file consumed approximately 5mj and grew to

[6] Fixity Policy: https://apps.es.vt.edu/confluence/display/LIBDPLD/Fixity+Policy

[7] MD5 Message Digest algorithm: https://datatracker.ietf.org/doc/html/rfc1321

[8] APTrust: https://aptrust.org/

[9] MetaArchive: https://metaarchive.org/

[10] Figshare: https://figshare.com/

approximately 40mj for a 1mb file. Energy consumption is also dependent on the energy source, meaning coal, natural gas, petroleum, or other, with coal having the highest impact at 54% of energy in the United States in 2020 [20]. The schedule of fixity checking also affects energy consumption, as running ongoing tasks during peak hours will consume more energy than running them during off-hours, such as in the middle of the night.

In an analysis of quality and energy efficiency in hashing algorithms of mobile devices, [21] highlighted the importance of low-energy hash functions' effect on battery life and found a 29% difference in battery life between choosing the highest and least energy-consuming hash. They concluded that changing the algorithm to reduce energy consumption without losing security functionality is possible. Reference [22] noted that Reference [21] did not focus on the energy consumption of hashing "from an algorithmic perspective" [22] but also concluded that MD5 is the least energy consuming algorithm. We applied this research to our own fixity practices.

C. *Storage*

The energy consumption of fixity checking is intertwined with digital storage choices and the number of copies. Storage is a necessary but energy-exhaustive component of preservation systems. Robust digital preservation means distributed digital preservation storage, preferably with administrative diversity, and multiple copies. The NDSA Levels of Preservation V2 recommends a minimum of 3 copies [15] and LOCKSS maintains 5-7 copies.[11]

AWS is one of VTUL's primary storage locations. AWS claims to have a 72% reduction of carbon emissions from their data centers when compared to other enterprise data centers [23]. They have instituted multiple initiatives for renewable energy, water stewardship, supporting other organizations to increase their own sustainable initiatives.[12] Reference [24] and a team of researchers have attempted to test these claims by building a dataset of CO2e emissions from AWS's EC2 hardware to attempt to estimate the impact of EC2 hardware on carbon emissions. They found that it was difficult to measure the distribution of emissions over time due to the limited lifespan of a server, but ultimately produced a dataset available for revalidation and manipulation. Others have claimed that cloud computing and storage is significantly more energy-consuming than saving to a disk [25], but that any security-driven disk server will consume more energy than an energy-saving disk server [26].

Other similar work in determining the electricity usage of a storage system is at the University of Houston Libraries where Bethany Scott inventoried all of the hardware components of their access and preservation infrastructure [27]. She concluded that focusing on ZFS fixity checking and decreasing file format resolutions would be the best way to optimize their local hardware to decrease environmental impact.

D. *Limitations*

This paper is scoped to archival appraisal, digitization workflows, fixity frequency, and storage options. We are not exploring the energy consumption of migrations, data transfers, VTUL hardware energy consumption; we are also not examining other cloud computing actions that occur, although there is significant interest in green computing.

We are using approximate numbers to determine a broad sense of approximate impact that is not based on hard numbers and relies on others' research. Our paper is highly qualitative and meant to provide direction for exploring changes in our digital library practices. Isolating our research to the defined scope may alter the ultimate environmental impact of our practices, but still provides insight on what we may be able to modify in the short term.

IV. Results

A. *Appraisal and Digitization Methods*

Using the plainly-stated charge of Pendergrass et al. to reduce digital content overall, appraisal and digitization practices are areas which should be scrutinized. A well defined collecting policy will help control the scope and prioritize the collecting efforts. The Special Collections and University Archives at VTUL has a mission to provide access to materials in their original form, and to offer

[11] LOCKSS FAQ: https://www.lockss.org/about/frequently-asked-questions

[12] AWS Sustainability: https://sustainability.aboutamazon.com/environment/the-cloud?energyType=true

materials in digital format "when possible".[13] This language has an allowance for familiar constraints to cultural heritage institutions such as time and funding. It may be beneficial to directly name environmental considerations in a future revision. The proliferation of born-digital collections presents an amplified challenge, and may require a modified collecting policy to address sustainability.

Clarity in how collections are prioritized internally for digitization is another area to address. VTUL has an Advisory Council for Digital Collections which prioritizes library and community projects for the Digital Imaging Lab. The committee's rubric for selection focuses on mission-specific projects and works through requisite technical details. This committee could be a logical check on unsustainable digital projects in the pipeline. This scrutiny should also exist within submitting library departments prior to review by the Advisory Council. Departments should clarify how collections are prioritized internally. Selection decisions may be made around privacy, access restrictions, copyright, uniqueness, as well as time and effort required. The impact of a project on sustainability goals should be given ample consideration in this list. It may also be beneficial to create a list of collections that specifically will not be digitized.

Among digitization practices, organizations should identify areas where changes can be made. In some cases this will mean going against industry standards of resolution or bit depth. While the biggest results will come from a reevaluation of standards contributing to file size, simple cleanup to digitized material can play a role in sustainability goals. For visual materials, this could mean addressing duplicate or blank pages. For audio-visual materials, editing dead air and trimming commercial/non-relevant content from digitized sources prior to repository ingest is valuable work.

B. Fixity Estimations

Given our context in the libraries, we assume most of our files will be on the larger end of the range tested by Fournier et al. [18]. If we operate under the assumption that the average MD5 hash consumes at least 40mj per hash, or 0.05 W*s, this equates to 1.111111111E-8 kW*h. This is too small to translate to CO2e emissions, but if we calculate 1 terabyte (TB) of content, we get the following approximate results in Table 1.

TABLE I
Estimated energy consumption and carbon emission of hashing 1TB of data

Size	Millijoule	Watt-second	Watt-hour	Kilogram (kg)
1 MB	40mj	0.05 W*s	1.111111111E-8 kW*h	--
1 TB / 1,000,000 MB	40,000,000mj	40,000 W*s	11.111111 kW*h	2.59 kg CO_2e

The number of storage locations, varying workflows, varying fixity frequencies, and general flow of storage to make exact calculations difficult. If we simplify it to whole numbers as our total TB and take into account the following table of each Virginia Tech storage location and the number of approximate TBs in each location, we find the following approximation for running fixity one time on each storage space in Table 2.

TABLE II
Estimated carbon emissions of VTUL storage spaces based on size

Storage Location	Fixity Freq	Size in TB	Kilogram CO_2e
Local high speed server	Nightly	11	28.59 CO_2e
Local NAS server	Nightly	10	25.9 kg CO_2e
AWS East Region	Every 90 days / ~4 times a year	1	2.59 kg CO_2e
AWS West Region	Every 90 days / ~4 times a year	.5	1.3 kg CO_2e
APTrust	Every 90 days / ~4 times a year	6	15.54 kg CO_2e
MetaArchive (LOCKSS)	Every 90 days / ~4 times a year	5	12.95 kg CO_2e
			86.87 kg CO_2e

The final result is simply, the environmental impact of running fixity is very complicated to define. Our results are extremely broad and validating these results would involve a time-intensive research study in collaboration with our IT division, vendors, and preservation vendors, which is a goal that we currently do not have the support or bandwidth to perform. Despite this, if our final result of 86.87 kg CO2e for running a single fixity check on all of our

[13] VT Special Collection and University Archives: https://spec.lib.vt.edu/about/index.html

approximate data is even close to accurate, this is cause for concern and an impetus to refine our workflows.

C. Storage Considerations

VTUL has designated 4 levels of preservation.[14] Not all content will be maintained at all levels. In terms of the number of storage locations, the levels are as follows: Level 0 is no preservation action taken; Level 1 basic preservation is 1-2 local copies, 1 cloud copy; Level 2 extended preservation is 2 local and 2 cloud copies; and Level 3 Advanced preservation is 2 local copies, 2 cloud copies, and ingest into one of our two distributed storage locations, APTrust or MetaArchive. Most of our content is designated at a Level 2.

To review, VTUL uses a combined storage system of the following contracted storage vendors and the approximate number of copies as shown in Table 3.

TABLE III
Overview of VTUL storage locations

Storage Location	Medium	Geo-graphical Location	Purpose	Copies
Local NAS	disc	Virginia	Working/ staging server	1
AWS East Region	Cloud	Virginia	Primary cloud storage	1
AWS West Region	Cloud	Oregon	Secondary cloud storage	1
APTrust	Cloud	Virgina, Oregon	Preservation storage, admin diversity	3
MetaArchive (LOCKSS)	various disc	varies	distributed preservation	5
Chronopolis (Figshare)	Cloud + disc	Virginia, Oregon, California	Figshare preservation	1

This list is not scoped to include additional data points, such as our institutional repository which is run on a local DSpace instance, our learning object repository in Omeka, any other Omeka instances VTUL hosts, our Confluence spaces, GitHub instance, or our Google Drive storage. We clearly rely heavily on AWS for our own primary and secondary storage, through APTrust, and indirectly through Chronopolis. The question we asked ourselves was whether administrative diversity between multiple services benefited us enough on a security level to justify this reliance and the number of copies we maintained. Lots of copies and lots of checksums do keep stuff safe, but can we articulate the value of these choices and still account for the carbon footprint?

Our answer is yes, but with modifications. Our local servers are for creating and staging content for ingest and our AWS serves as both a backup for our disc servers as well as access and preservation for our digital library. All of our preservation options are for geographically distributed preservation storage, administrative diversity, and technology diversity, all of which are considered good practice in the digital preservation community. Actually defining the environmental impact of all of our storage locations is complicated due to the various workflows, number of copies, distribution of copies, and independent needs of the collection. We commit to the number of copies we maintain and the storage we have chosen, but the amount of space and energy we consume is dependent on our appraisal system, both pre-digitization and for preservation. We also found that we have not determined what Reference [1] describes as acceptable loss - the “level of acceptable loss in collection under [our] care” [1] to make better use of what resources we do have.

AWS recently released a new feature called the Customer Carbon Footprint Tool, available to all customers. This tool allows users to track their carbon emissions over time and over geographic location, specially measuring Scope 1 and Scope 2[15], or direct emissions and indirect emissions, of content in AWS [28]. With the aid of our digital library’s Software Engineer, we obtained results from our development server in AWS from January 2020 through November 2021. The results as seen in Figures 1 and 2, indicates that we emitted 0.6 MTCO2e and claims that we have saved 0.4 MT CO2e as compared to “on-premises computing equivalents.” S3 is the feature generating the most carbon emissions. As it is a new feature we are still learning how to read the information and understand the true impacts of the numbers, and

[14] VTDLP Preservation Policy: https://apps.es.vt.edu/confluence/display/LIBDPLD/VTDLP+Preservation+Policy

[15] EPA: https://www.epa.gov/climateleadership/scope-1-and-scope-2-inventory-guidance

we will continue to monitor it as we increase activity in AWS.

Figure 1 Virginia Tech's AWS carbon emissions summary

Service	Carbon emissions	%
EC2	0.1 MTCO2e	16.67%
S3	0.4 MTCO2e	66.67%
Other	0.1 MTCO2e	16.67%
Total	0.6 MTCO2e	100%

Figure 2 Virginia Tech's AWS carbon emissions by service

One general concept that we encountered in our work is that energy efficiency is predominantly measured by financial cost rather than environmental cost. The issue with increasing environmentally friendly systems is that it can have little to no impact on cost [29], which is a primary concern in most organizations and is often the focus of energy sustainability benefits over the environmental impact itself. The Customer Carbon Footprint Tool, for example, is found through the AWS Billing Console under Cost & Usage Reports, emphasizing financial cost. Cost is a major factor in all digital curation systems and cannot be overlooked, but it seems to be a mistake for us to only rely on financial cost as a measure of our energy sustainability.

V. Recommendations

Based on our preliminary results, we recommend the following actions to help reduce the carbon footprint of the VTUL digital library.

- **Include climate considerations in appraisal of digital collection projects:** The long term environmental impact of digitizing a collection should be considered alongside other factors in collection selection.
- **Revisit collection policies and institutional mission**: We recommend adjusting an existing collection policy or mission to reflect a commitment to sustainably manage digital resources and guide future decision making.
- **Decrease redundancy of working files**: We recommend streamlining the transfer process to minimize the multi-department storage redundancy of working files. Understanding that redundancy is a necessity, determine which stages of the collection management process should be the most secure. Schedule a process for deletion after migration and quality assurance.
- **Reduce ongoing fixity checks**: We recommend reducing scheduled fixity checks of all AWS objects from every 90 days to every 120 days or possibly more, increase spot-checking fixity from a randomly selected subset of files in each digital collection, and to increase test restorations to account for the decreased fixity checking.
- **Determine acceptable loss**: Reducing security will reduce energy consumption. We need to determine acceptable loss for each of the storage vendors we contract with and alter our workflows with mechanisms for faster healing to compensate for any loss.
- **Preservation appraisal**: We have defined our own levels of preservation, but we recommend modifying them to include more direct appraisal strategies and a determination of acceptable loss for each level.
- **Investigate smaller object sizes:** The size of a digital object directly impacts the energy consumed in running fixity, transferring between storage locations, and ongoing storage maintenance. We recommend exploring collections or data

types where there are options for creating lower resolution or otherwise smaller objects.

- **Sustainability commitment:** As an organization, we recommend that VTUL develop a Sustainability Statement for the Digital Libraries at VTUL to scope our work and to emphasize not only the importance of but the immediate need for greener digital library curation strategies.
- **Community training:** The Libraries are responsible for keeping up with digital trends and practices and educating the University and larger community. We recommend regular Professional Development Network training sessions on ensuring good practices in personal and professional archiving that also emphasize environmental sustainability.

VI. Next Steps

There are several next steps we want to pursue after this preliminary research. Exploring time and money spent on all of these steps to reinforce the areas where we need improvement on multiple levels. We also hope to explore other preservation activities including migration, restoration, transfer and syncing, file format verification, alternate storage opportunities, and appraisal.

1. REFERENCES

[1] K. Pendergrass, W. Sampson, T. Walsh, and L. Alagna, "Toward environmentally sustainable digital preservation," The American Archivist, vol. 82 no 1, pp. 165-206, 2019. https://doi.org/10.17723/0360-9081-82.1.165

[2] United Nations, "United Nations Framework Convention on Climate Change," United Nations Treaty Collection, 1992. https://treaties.un.org/pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XXVII-7&chapter=27&Temp=mtdsg3&clang=_en

[3] C. Durham, "The necessity of environmentally sustainable digital preservation and its effects on preservation workflow," Johns Hopkins University Sheridan Libraries, 2019. https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/62122/Durham%2c%20Curtis%20J.pdf?sequence=1&isAllowed=y

[4] H. Stehkämper, "'Natural' air conditioning of stacks," Restaurator 9, no. 4, pp. 162-177, 1988. doi.org/10.1515/rest.1988.9.4.163.

[5] Rowoldt, Sandra. "Going archivally green: Implications of doing It naturally in Southern Africa archives and libraries," South African Journal of Libraries and Information Science, vol. 66, no. 4, 2014. doi:10.7553/66-4-1425.

[6] Society of American Archivists, "SAA Core Values Statement and Code of Ethics," SAA, 2020. https://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics

[7] Digital Preservation Services Collaborative, "Digital Preservation Declaration of Shared Values." https://dpscollaborative.org/shared-values_en

[8] National Archives and Records Administration, "Climate action plan," National Archives, 2021. https://www.archives.gov/files/about/plans-reports/sustainability/climate-action-plan.pdf

[9] B. Goldman, "It's not easy being Green(e): Digital preservation in the age of climate change," in Archival Values: Essays in Honor of Mark Greene, Society of American Archivists, 2018. https://scholarsphere.psu.edu/resources/381e68bf-c199-4786-ae61-671aede4e041

[10] W. R. Hibbard, Virginia Coal: An Abridged History, Virginia Polytechnic Institute and State University, Ed. T. J. Clutter, 1990, pp. 20-24. http://hdl.handle.net/10919/90196

[11] R. Freis, "The day the earth shook 50 years later, Montgomery's biggest tragedy resonates still," in The Roanoke Times, 18 April 1996, pp. 1. https://scholar.lib.vt.edu/VA-news/ROA-Times/issues/1996/rt9604/960418/04180017.htm

[12] Federal Agencies Digital Guidelines Initiative, "Technical Guidelines for Digitizing Cultural Heritage Materials," September 2016. http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

[13] S. Barsness, A. Collie, M. Gallinger, C. Kussman, S. Schaefer, and G. Truman, "2017 fixity survey report," National Digital Stewardship Alliance, 2017. https://ndsa.org/documents/Report_2017NDSAFixitySurvery.pdf

[14] D. Rosenthal, "How few copies?," DSHR's Blog, April 2016. https://blog.dshr.org/2016/04/how-few-copies.html

[15] Digital Preservation Coalition, "Digital preservation handbook," 2nd Ed., 2015. https://www.dpconline.org/handbook

[16] J. Bailey, "File fixity and digital preservation storage: More results from the NDSA storage survey," The Signal, 2012. https://blogs.loc.gov/thesignal/2012/03/file-fixity-and-digital-preservation-storage-more-results-from-the-ndsa-storage-survey/

[17] N. Taylor, "Lots of checksums keep stuff safer," CNI Spring Membership Meeting Proceeding, 2019. https://www.cni.org/wp-content/uploads/2019/04/CNI_Flexibility_Taylor.pdf

[18] B. Fournier, V. Tong, and G. Guette, "Accurate measurement of the energy consumption of security functions," Scitepress, 2021. https://www.scitepress.org/Papers/2021/105446/105446.pdf

[19] U.S. Energy Information Administration, "How much of U.S. carbon dioxide emissions are associated with electricity generation?," Frequently Asked Questions, 2020. https://www.eia.gov/tools/faqs/faq.php?id=77&t=3

[20] R. Damasevicius , G. Ziberkas, V. Stuikys, and J. Toldinas, "Energy consumption of hash functions," Elektronika ir Electrotechnika, vol. 18, no. 12, pp. 81-84, April 2012. https://pdfs.semanticscholar.org/7f43/7af19922bda1fe845b663290440e359d21ff.pdf

[21] P. Harish, “Towards designing energy-efficient secure hashes,” University of North Florida Digital Commons Graduate Theses and Dissertations, 2015. https://digitalcommons.unf.edu/etd/598/

[22] National Digital Stewardship Alliance, “Levels of digital preservation,” version 2.0, 2019. https://ndsa.org/publications/levels-of-digital-preservation/

[23] D. Bizo, “The carbon reduction opportunity of moving to Amazon Web Services,” AWS, October 2019. https://sustainability.aboutamazon.com/carbon_reduction_aws.pdf

[24] B. Davy, “Building an AWS EC2 carbon emissions dataset,” Medium, September 2021. https://medium.com/teads-engineering/building-an-aws-ec2-carbon-emissions-dataset-3f0fd76c98ac

[25] J. Adamson, “Carbon and the cloud,” in Stanford Magazine, Ed. N. Mmonatau, May 2017. https://stanfordmag.org/contents/carbon-and-the-cloud

[26] S. Yin, M. I. Alghamdi, X. Ruan, M. Nijim, A. Tamilarasan, Z. Zong, X. Qin, and Y. Yang, “Improving energy efficiency and security for disk systems,” in 2010 12th IEEE International Conference on High Performance Computing and Communications, 2010. DOI 10.1109/HPCC.2010.26

[27] B. Scott, “Estimating energy use for digital preservation, part 1,” Bloggers!, October 2020. https://saaers.wordpress.com/2020/10/06/estimating-energy-use-for-digital-preservation-part-i/

[28] J. Barr, “New – customer carbon footprint tool,” AWS News Blog, 2022. https://aws.amazon.com/blogs/aws/new-customer-carbon-footprint-tool/

[29] D. Rosenthal, “Why is green preservation hard?” in Green Bytes: Sustainable Approaches to Digital Stewardship at the National Digital Information Infrastructure and Preservation Program partners meeting, July 2013. https://digitalpreservation.gov/meetings/documents/ndiipp13/Rosenthal.pdf