

"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"



"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"
"STATUS REPORT"



from 15 may to 15 august 2025, the University of Manitoba Libraries' Research Strategy and Data Services (RSDS) unit partnered with JSTOR's new Digital Stewardship Services (hence DSS) in order to address the digital asset backlog in the UM Archives & Special Collections' (hence "the Archives") care. this was done through the use of Seeklight AI, a generative artificial intelligence tool for generating descriptive metadata packaged with JSTOR DSS.

the digital asset backlog refers to the large volume of born-digital assets acquired by the Archives which have not been preserved or described at an item level. without addressing issues of format and



access, assets in the backlog are at immediate risk of dissociation and data degradation.

the main step in the workflow I worked on was uploading files to Janus, a um-secure server for storing assets prior to further preservation. the maximum amount of space on Janus is 28 terabytes (TB). on 2 july (when i began compiling files for this piece), 27 assets equal to 16.5 TB (653 324 files in 153 822 folders) were stored on Janus. this number rose to 56 assets equal to 16.6 TB (984 138 files in 154 172 folders).

physical storage devices for digital assets are located throughout the the annex,

annex basement, archives vault, archives workroom, and rare book room. according to the digital asset register, an index for known digital assets, there are 180 digital assets corresponding to 258 individual accessions. this zine compiles examples from the initial review of assets in the backlog as a case study of digital preservation workflows. many of these assets have not been examined since the time of their accrual. this "STATUS REPORT" highlights particular concerns related to the accumulation of digital material. these include the risk of dissociation, data degradation, and data bloat.

to the final point of data bloat, the 16.6 TB of accessible files on Janus fails to include



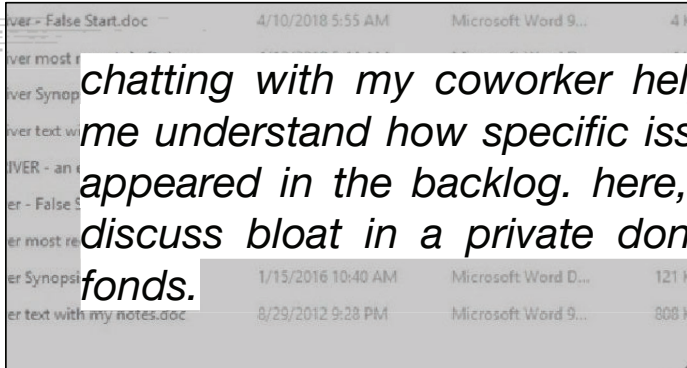


~1.662 TB of data when compared to the server's actual disk size of 18 264 567 992 320 bytes. while some of this figure includes attached documentation, this is the bloat—likely system files or mis-uploaded items that, nonetheless, take up space.

if the data bloat were stored entirely in double-sided 3½-inch floppy disks (the most popular of which could hold up to 1.44 MB each), it would require 1 154 167 individual disks to contain it all.

Windows Photo Viewer can't open this picture because the file appears to be damaged, corrupted, or is too large.

Are they similar to the 4 KB files at the top of this directory? They look like they're related files that were still on the system in some way when we imaged the content.



It's not technically a deaccession because we'd keep parts of the collection, but maybe we could clarify with the head archivist if we could create a list of categories for files they'd be ok with us deleting as we come across them, where we could just delete the files when we come across it and include things like files the donor deleted, but didn't fully wipe from their system on that list.



im reading some floppy disks to update the DAR

are floppy disks always this loud

ive never used one before



... yes. Consider it digital preservation asmr





MSS 326 a2010-112 EL
25, Box 2, William O.
Pruitt Jr. fonds, University
of Manitoba Archives and
Special Collections,
Winnipeg, Manitoba.

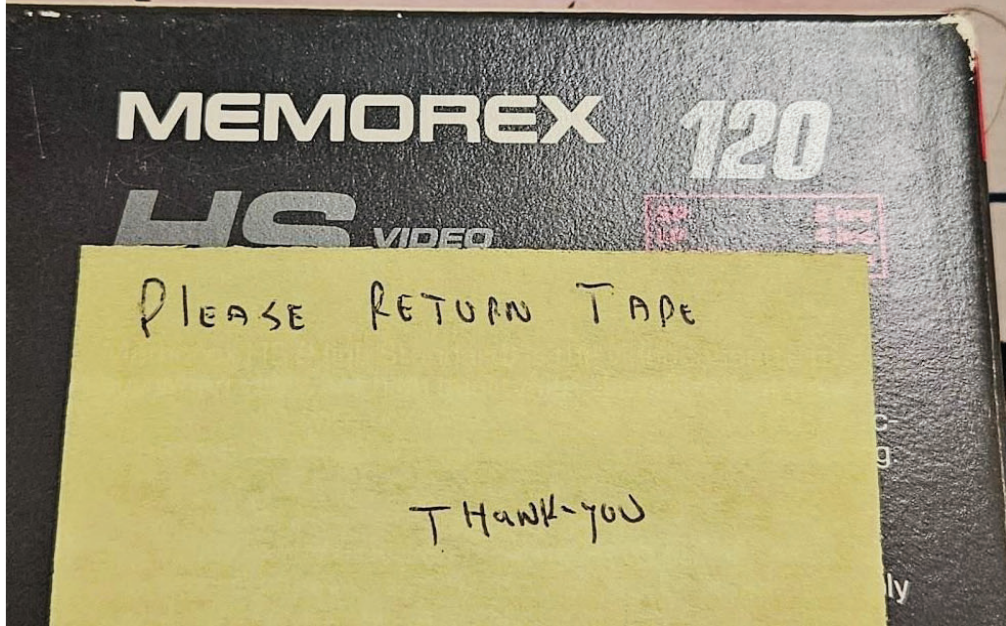
i learnt that gold CDs used to be archival standard and that there was a big rush to digitize stuff when that technology first became available. my conception of digital preservation sought forever, not extension, like my forebears. forever's impossible. i wonder how many people worked under the pretense it was achievable in a digital world. furthermore, i'm operating on a windows OS, but there's no guarantee that donors were using the same system software. furthermore, while the oldest assets in the register were donated in 1991, some individual items predate this year (for instance, the William O. Pruitt Jr. fonds contained disks from 1988. this might change as more digital assets are found and/or re-identified.*

my parents were born in 1980. if digital was supposed to be for forever, does that mean i was born afterwards? are we both living past it, right now? i never saw the towers fall.

another issue posed by the digital asset backlog is that, sometimes, donors attach requests for items to be returned after being processed. because the backlog is left unprocessed, these requests may not reach the archives.

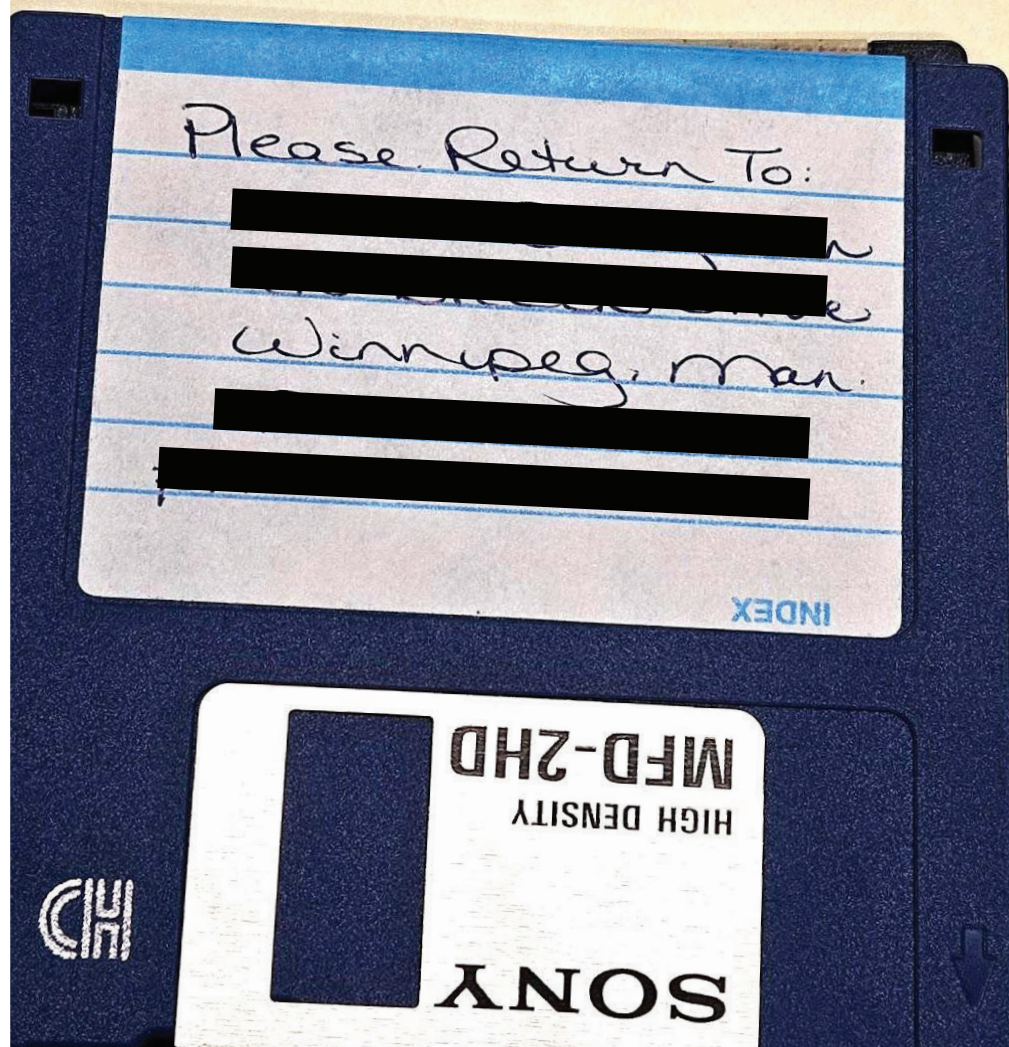
A.99-66 corresponds to the 1999 accrual of mostly paper administrative documents from the Manitoba Cattle Producers Association (A1999-066). photographed is a Memorex 120-minute HS video tape with a note requesting it to be returned, found twenty-six years later.

A.99-66



another example. EL 17, two 3.5" floppy disks in Betty Nordrum's fonds (MSS 309 a1999-034), contains files related to the Government of Manitoba's Women's Advisory Committee and were donated in the same year as A.99-66.

Pictured below, both have deteriorated to the point of being unopenable on a USB floppy disk drive.



a significant portion of work done throughout the internship involved locating the physical storage locations of digital assets in order to edit the digital asset register to more accurately reflect their contents.

in some cases, these digital assets are missing. for instance,

PC 260	MSS 205	Frederick G. Stambrook
PC 259	A98-012	School of Nursing

accurately locating digital assets is of the utmost importance for manual review and the digital preservation workflow writ large. file dissociation poses an immediate issue, but data decay/degradation is a concern of greater priority for digital assets originating from the 2000s and earlier.

take a closer look at the pattern on this page used throughout this zine.



the pattern is a modified image taken of a wordperfect document stored on a floppy disk since the mid-90s, unmodified here. its contents have since completely degraded.

COMMONS.BK!, 1996, MSS 326 a2010-112 EL 25, Box 2, Item 91, William O. Pruitt Jr. fonds, University of Manitoba Archives and Special Collections, Winnipeg, Manitoba.

* i am going to catalogue the webcam images from UM Info at the container level because i dont want to describe individual photos taken at 10 minute intervals over the course of 6 months

* i think this is fair



* Ya that's totally fine. Another thing to bring to the archivist when she's back - "do you want a photo per minute..."

* I hope it was only while they were actively doing work and not hours of photos from the idle of the night where nothing changed at least..



* i have what can only be called bad news.



in other cases, digital assets may contain near-duplicates or individual file components of a larger, amalgamated project. the following pages represent Information Services and Technology's unique solution of rigging a webcam to take images every ten minutes in order to record the construction of the Faculty of Engineering building through the end of 2002 and beginning of 2003.

webcam_image_021101_08000, 1
November 2002, a2006-091, Box 1,
Item 1, Information Services and
Technology fonds, University of
Manitoba Archives and Special
Collections, Winnipeg, Manitoba



webcam_image_021101_082700, 1
November 2002



webcam_image_021107_145400, 7
November 2002



webcam_image_021101_111100.jpg, 1
November 2002



webcam_image_021107_150400, 7
November 2002

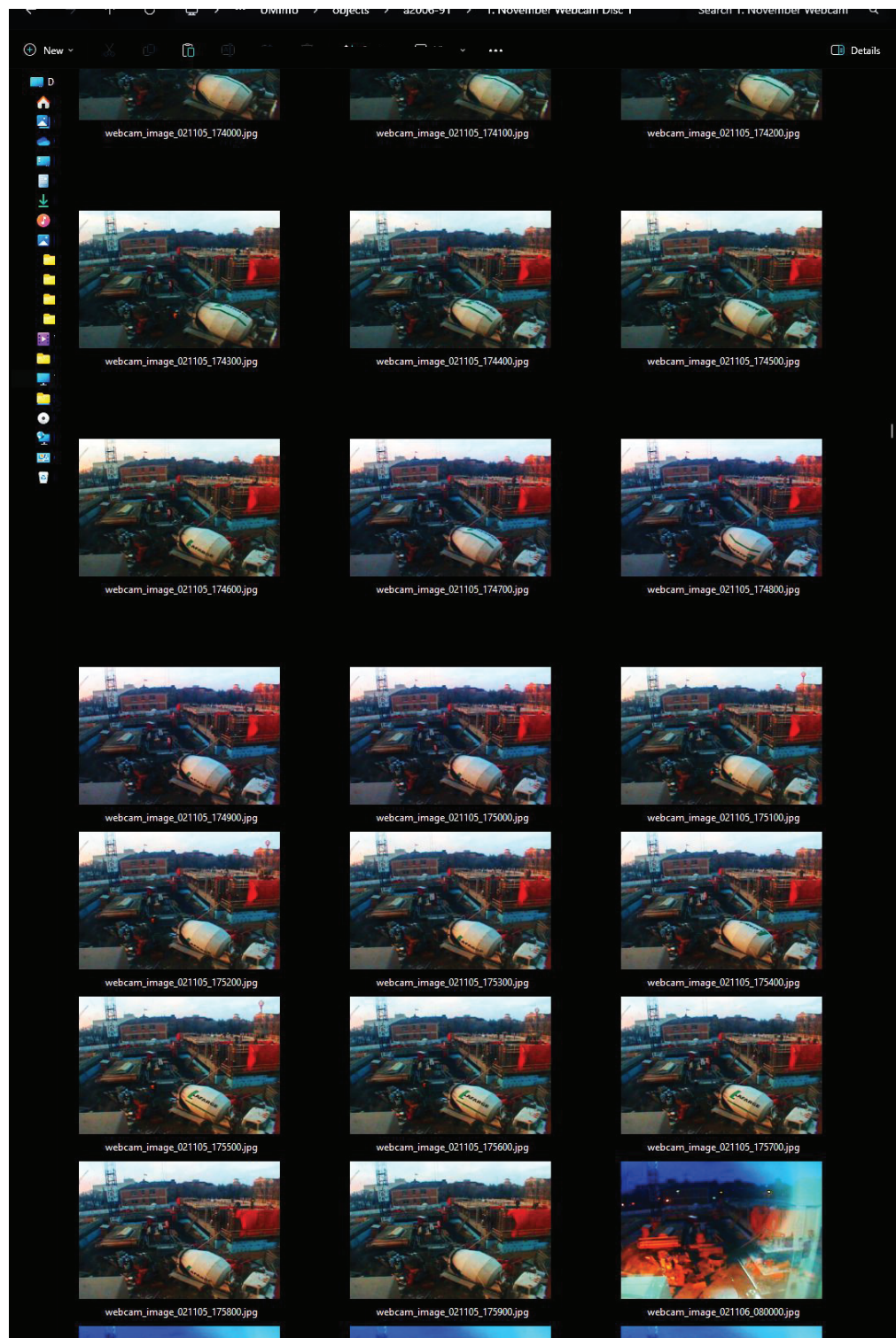


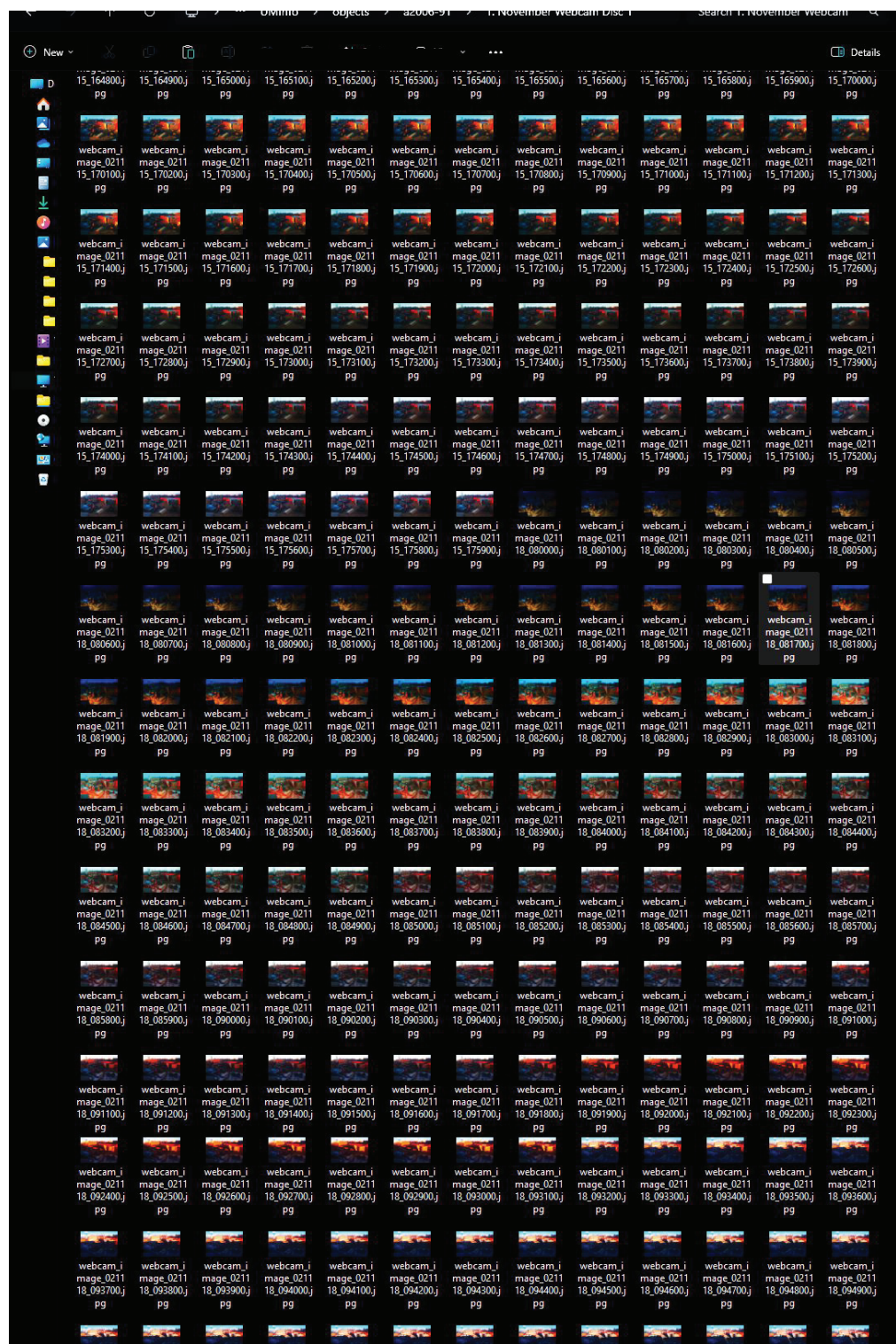


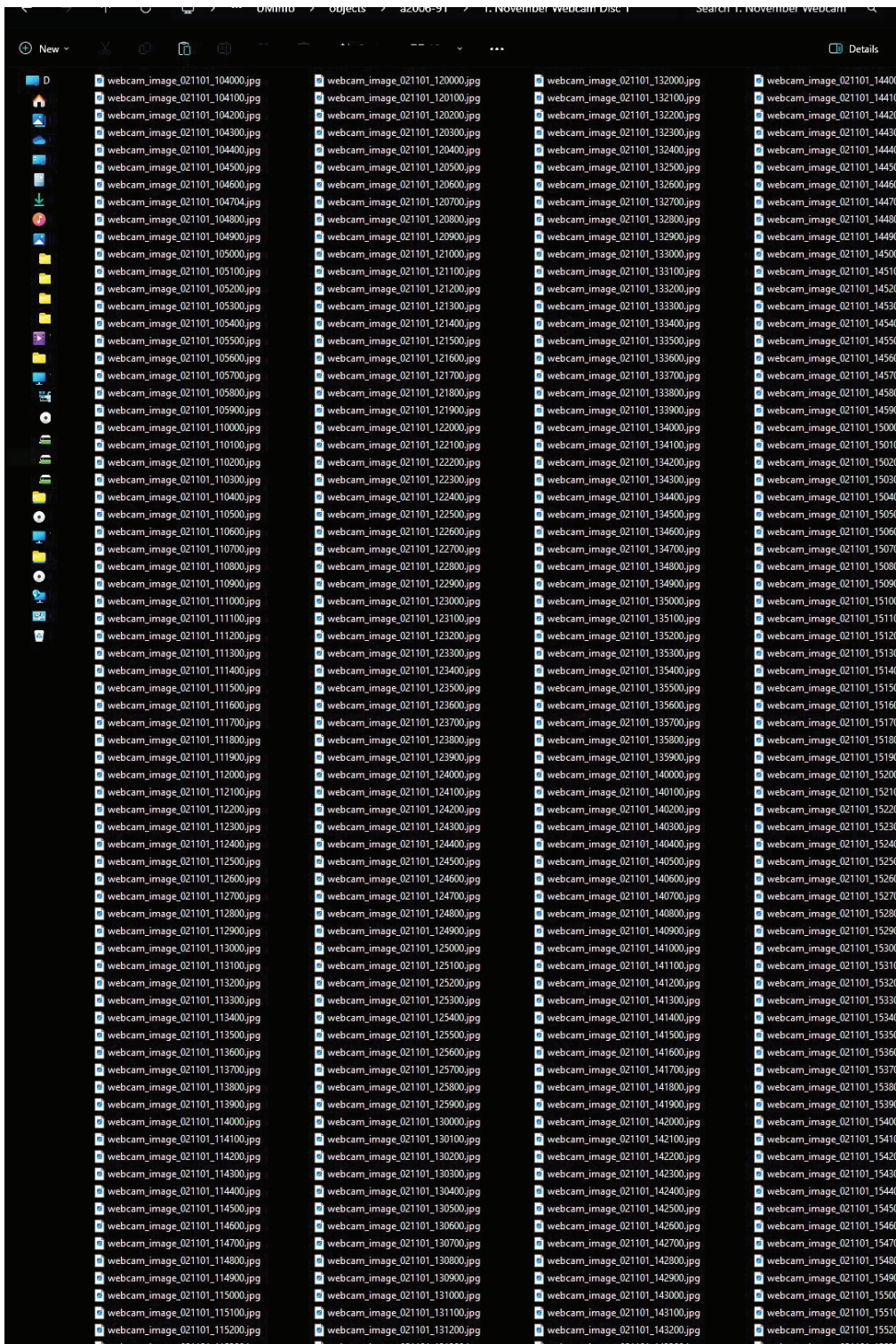
webcam_image_021108_110300, 8
November 2002



webcam_image_021108_123600, 8
November 2002









what do i do? i thought, with this hollinger full of CDs and cart full of hollingers, halfway through my internship with nothing to show for it. the position with RSDS was a crash course in archival work—i was an anthropology undergrad starting the archival studies program in fall. while i had done a project in the College of Medicine Archives last year, i felt woefully out of my depth. this is something i usually feel, though.

it is very easy to work in this kind of environment. labour can be a coping mechanism for imposter syndrome far more effective than mindfulness or stress regulation or therapy. you would think that my growing awareness of this fact would

would be enough to drive positive change (and, for the most part, it does). so why is it still so hard to stop working mindlessly? or worse, heartlessly?

the notion that the work one does is important to their life is also important to me. it's probably why i have such a terrible work-life balance and it's a hill i die on often. how is it possible to separate one's life into distinct parts without them all bleeding into each other? to keep the parts of it organized in a way that another person can understand by a precursory overview or a passing glance? not only for the content but for the methodological basis of the construction of a narrative

my previous therapist asked me what it was like, once, to have an imaginary graveyard in my head filled with real people.

what do you do when it is painful and

arduous to care when you can't stop caring? isn't it too much weight to carry something like that? isn't that strange? everyone has one of those. why ask me? besides, that's not the thesis here. this is barely an academic text. a curated exhibit, if anything. i wanted to have something to show for you.

i also realize that my knowledge presents only a partial understanding of the breadth of the issues and concerns posed by the digital asset backlog. the full scale of the bloat, decay, and dissociation (discussed herein) are barely described. how ironic. cataloguing all of our problems is part of the catalogue's problem—we don't know what preservation a digital asset requires until it is opened, so accessing them is necessary. Seeklight is supposed to take on that task and generate metadata accordingly.



it works fine. it knows archival terminology, but generates over-generalized metadata. an AI might be able to identify spider's silk in a photograph, but not why such an image would be a taken. thus, a human archivist is still necessary when using AI tools. what should be processed in this way? what shouldn't?

i found my use of Seeklight drastically shift over the course of my brief internship from actively involving it in preservation workflows to selecting particular fonds that might work well with it. typically, fonds containing high volumes of text or image files or with some form of pre-organization can provide the most specific data for. even then, files with AI-generated metadata required extensive revision in order to follow the format of finding aids (where available) and item order. Seeklight is only able to examine JPG and PDF files,



0001 Spiders Web, 13 March 2005,
a2010-098, Box 1, Item 1, Catherine
Thexton fonds, University of
Manitoba Archives and Special
Collections, Winnipeg, Manitoba

