ANYTHING IS BETTER THAN NOTHING

Society of California Archivists / 2019 Annual General Meeting / Long Beach, CA

MINIMUM VIABLE ACTIONS FOR BORN DIGITAL ACCESSIONING

1. PROCESSING WORKSTATION

Emily Vigor

2. APPRAISAL

Kate Dundon

3. CAPTURE & PRESERVE

Steve Duckworth

4. STABILIZE

Steve Kutay

5. ACTIVITY

6. Q&A

1

PROCESSING WORKSTATION

What do I need?

Emily Vigor
Environmental Design Archives

MINIMUM REQUIREMENTS

Computer

Storage

Open source software

CURRENT WORKSTATION

Grant funded

Supplies under $2000

IT Support :)

SUPPLIES & MATERIALS

COMPUTER [$1000]

Requires an internal optical media drive to read CDs and DVDs, as well as several USB ports to connect disc drives to

EXTERNAL HARD DRIVE(S) [$200]

1TB (minimum) hard drives to be formatted for Mac and PC

OPEN SOURCE SOFTWARE [$0]

Make sure you have the IT support or training to help guide installation

FLOPPY DISK CONTROLLER CARD & DISK DRIVE [$200]

Helps newer machines read media connected via obsolete disk drives

WRITE-BLOCKER [$300]

Protects legacy media from being overwritten once connected to a computer for imaging

MEMORY CARD READER [$25]

Supports reading flash drive formats such as SD, CompactFlash and MicroSD

Ben Goldman post:
https://practicaltechnologyforarchives.org/issue2_goldman/

OTHER THINGS TO CONSIDER

Access to IT

Storage

Priorities

NO MONEY? NO PROBLEM!

Open Source Tools

ePADD

ArchiveIt

California Revealed

Research and Training

Literature review

Webinars

Build internal documents

Raise $$$

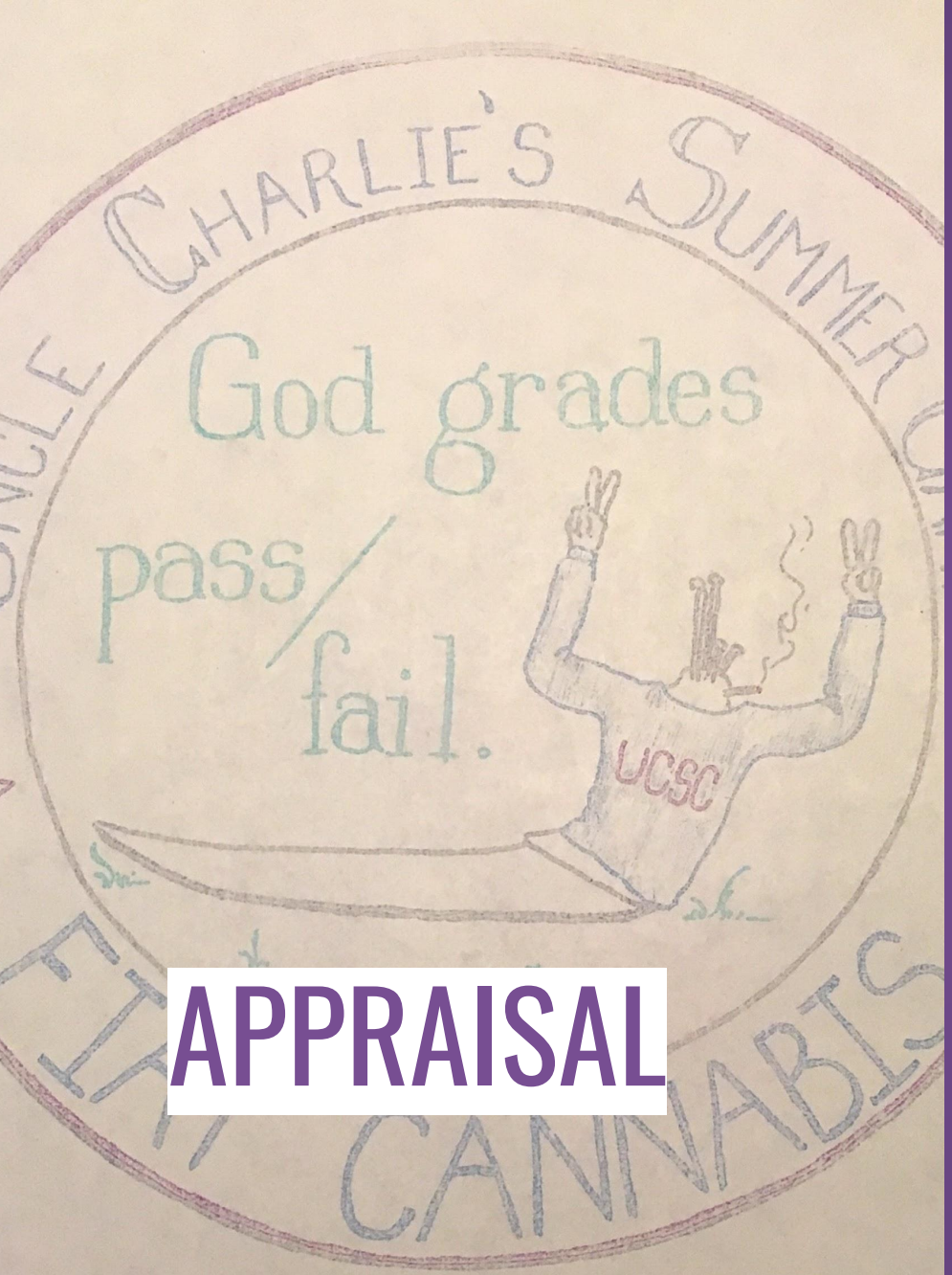
Grants

Crowdfunding campaigns

Collaborate

Put a call out on Listservs

Check your own institution



APPRAISAL

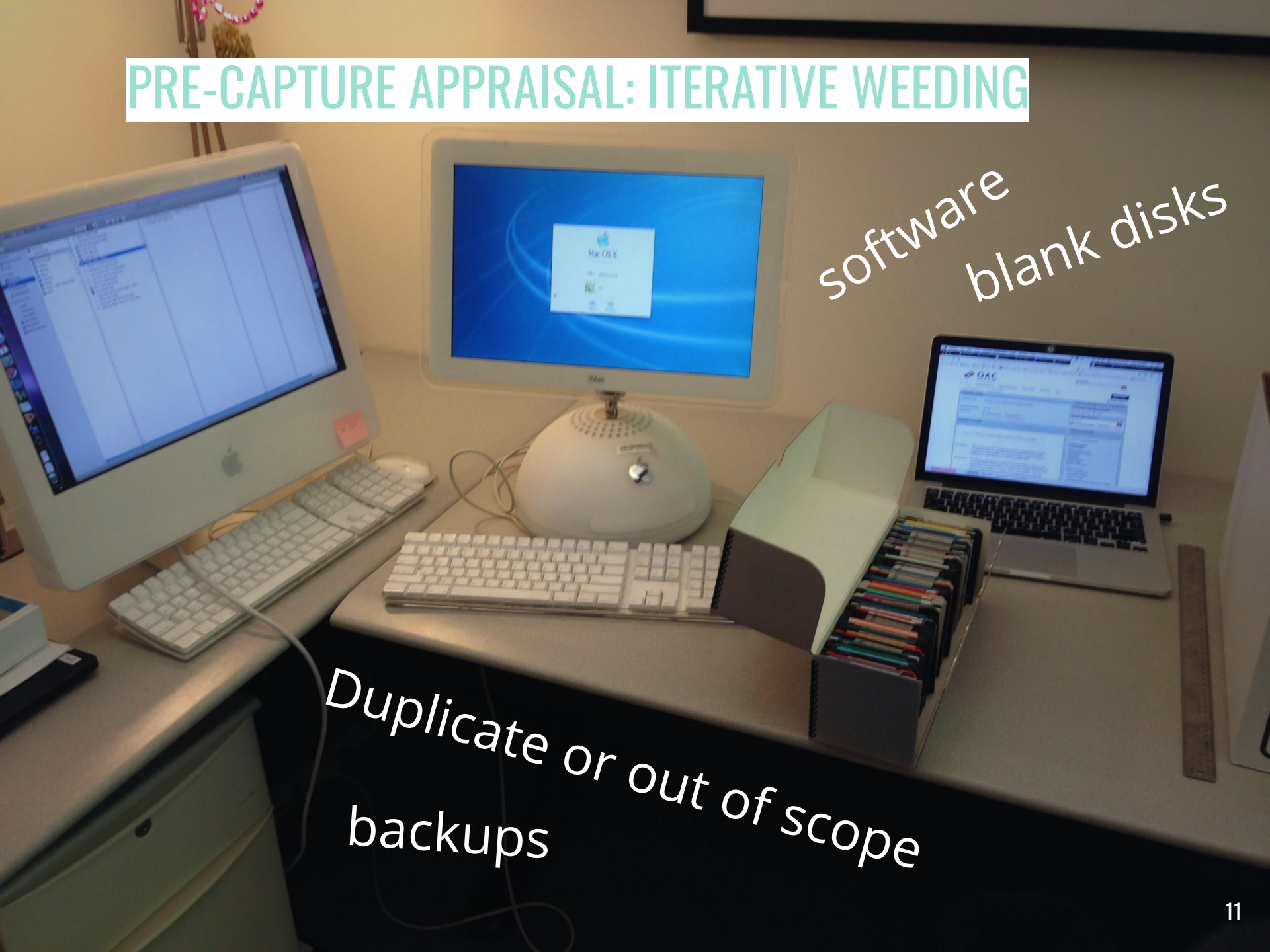
2

Kate Dundon
UC Santa Cruz Special Collections & Archives

PRE-ACQUISITION APPRAISAL

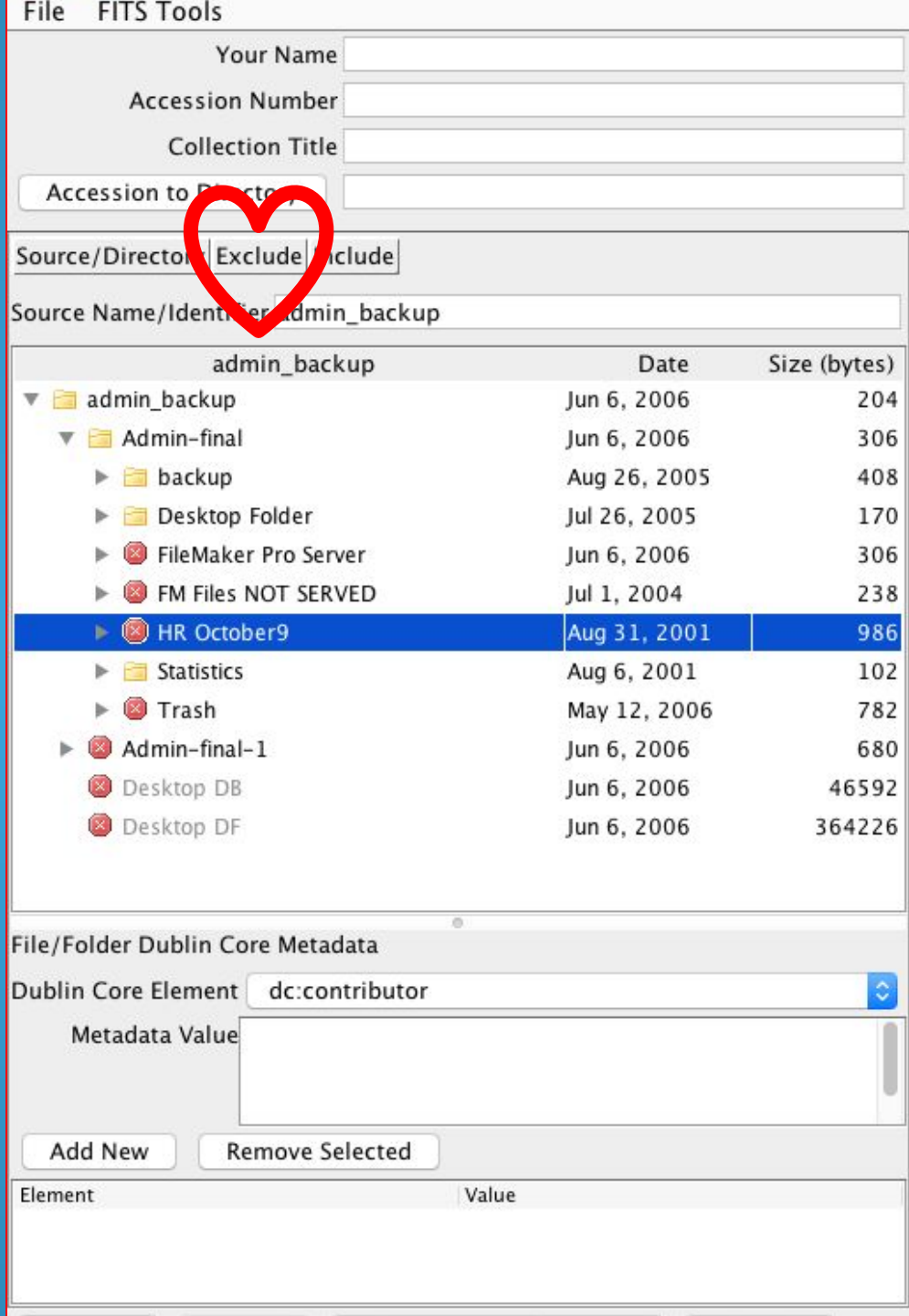
- ✓ Collect born digital intentionally
- ✓ Consider your technical capacity
- ✓ Exclude out of scope content

PRE-CAPTURE APPRAISAL: ITERATIVE WEEDING



INTRA-CAPTURE APPRAISAL

Use Data Accessioner to exclude unwanted content from your transfer



POST-TRANSFER APPRAISAL

Manually identify and delete PII and clutches of non-collection content

Use DROID csv export and Excel to identify duplicate files

NAME	SIZE	TYPE	EXT	LAST_MODIF	MD5_HASH
TranssaccadicInteractionMotion		Folder		2015-08-15T	09:33:15
BiberIlgRemapMAE.pdf	492181	File	pdf	2011-01-27T	f05d48a2bf3
ChaChong2014.pdf	530082	File	pdf	2015-02-27T	9adbbfb078a
DemeyerTranssaccadic	687382	File	pdf	2011-01-02T	c583aee7bd5
WatsonKekelberg2009.	116337	File	pdf	2015-07-09T	7397976092
TranssaccadicMasking		Folder		2015-08-15T	09:33:21
DePisapiaetal2010.pdf	514264	File	pdf	2014-04-18T	78fc3103e97
TasMooreHollingworth	439686	File	pdf	2014-02-18T	113730436f4
White1976.pdf	787319	File	pdf	2015-08-14T	69162b958b
Uli-VSS2016-poster-fina	2E+06	File	pdf	2016-05-24T	76dc3386eb
VSS 2016 Lateral Gaze I	605871	File	pdf	2016-05-21T	1526e6ddd7
VSS 2016poster copy.p	513786	File	pptx	2016-05-07T	5fcc789d91f
VSS 2016poster.pptx	491923	File	pptx	2016-05-07T	7fcab061d71
binocular		Folder		2014-06-21T	10:41:44
3D glasses.doc	26624	File	doc	2013-03-15T	b48155192e
BBC stereo vision.weba	2E+06	File	webarch	2013-01-29T	a1aca6e6c3b
BBC_Hugo2.doc	47616	File	doc	2012-05-30T	da5f2d66816
Binocular Bruce.doc	34304	File	doc	2012-03-15T	c1fc3bdf843
Binocular Bruce.docx	68317	File	docx	2014-02-03T	bf9e53986fd
BinocularReviews.docx	147099	File	docx	2014-01-31T	9731fe45a12
GuardianStereo.webarc	3E+06	File	webarch	2013-06-18T	66fa3a5047
OVS13567-2.pdf	1E+06	File	pdf	2014-02-04T	011f60107e8
OVS13567R1 - please re	4938	File	rtf	2014-02-19T	21d04839c6
Stereo biography		Folder		2014-06-07T	10:44:25
AcqiringStereoAbstract	30720	File	doc	2013-11-19T	fd88651788
Acquiring Adult Stereop	28672	File	doc	2013-11-19T	da45bd54f13
Acquiring Stereo Vision	63488	File	doc	2014-02-11T	4b7779b474
AcquiringStereoTable.d	23552	File	doc	2013-11-19T	d4a08381e1
BinocCleanR2.docx	129148	File	docx	2014-02-20T	4ee0b841a6
BinocReviewReply.docx	160576	File	docx	2014-02-11T	7b14de7790
BinocularAbstract.docx	59849	File	docx	2014-02-20T	188e5280e1
BinocularBruceBio.doc	51712	File	doc	2013-02-20T	e13d85b6d9
BinocularBruceBioClear	135514	File	docx	2014-05-26T	0e88d0f976
BinocularBruceBiograp	72704	File	doc	2014-02-04T	80e48f3d1b
BinocularTitle.docx	46679	File	docx	2014-02-20T	a60fa62bdf6

APPRAISAL: TAKEAWAYS

APPRAISAL IS ITERATIVE

Take many passes

DOCUMENTATION IS ESSENTIAL

Record pre and post-processing extents

BUILD A KNOWLEDGE BASE

Maintain continuity by processing both paper and digital

RESPECT YOUR RESOURCES

It is resource-intensive to steward digital content

3

CAPTURE & PRESERVE

Accessible control and preservation

Steve Duckworth, @archivesteve

Oregon Health & Science University

ACCESSIONING: real talk

Moving/copying digital records can be scary

What if I mess up the metadata and ruin everything forever?

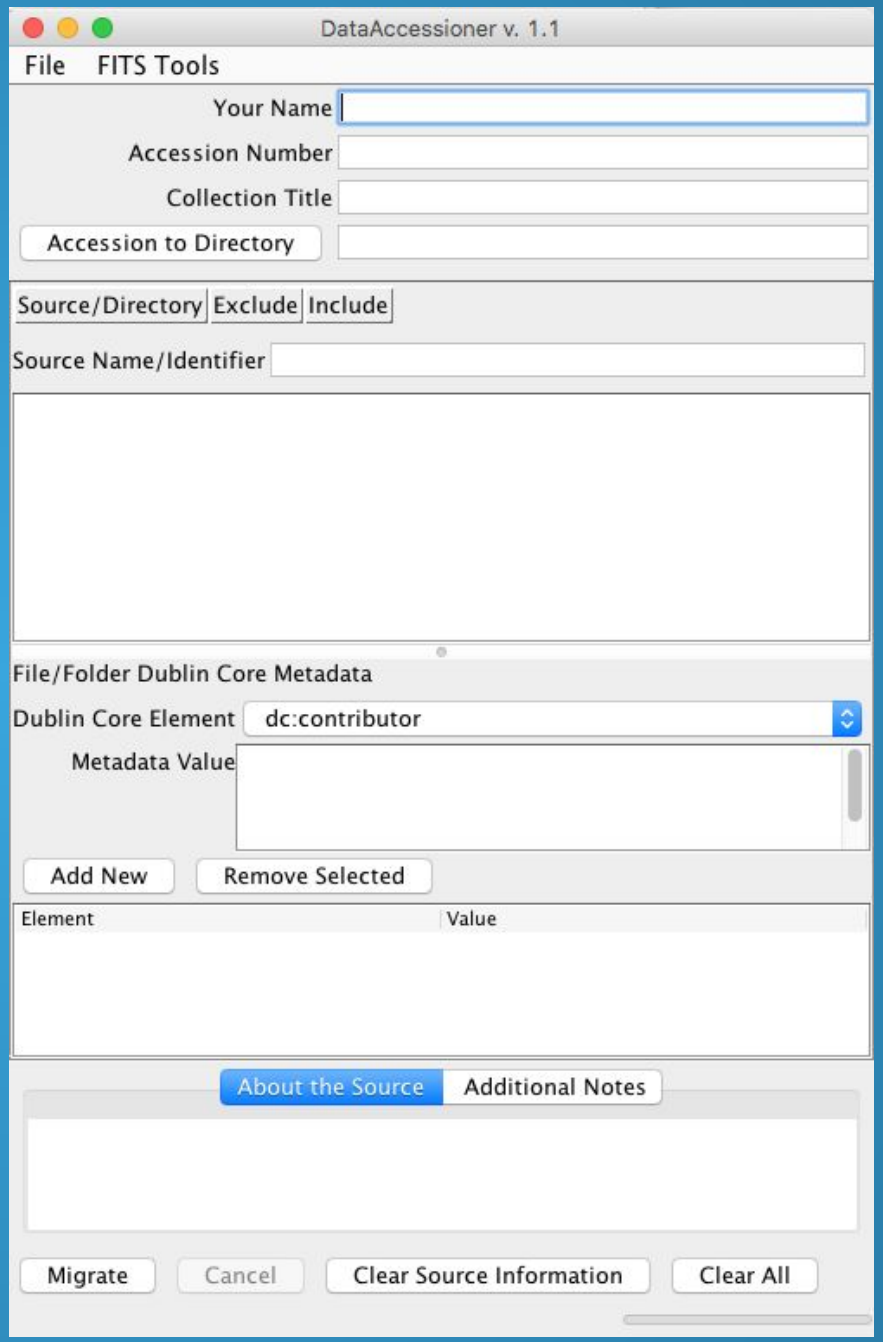
Leaving everything on deteriorating media is scarier

Move it or lose it.

INGESTING RECORDS with DATA ACCESSIONER

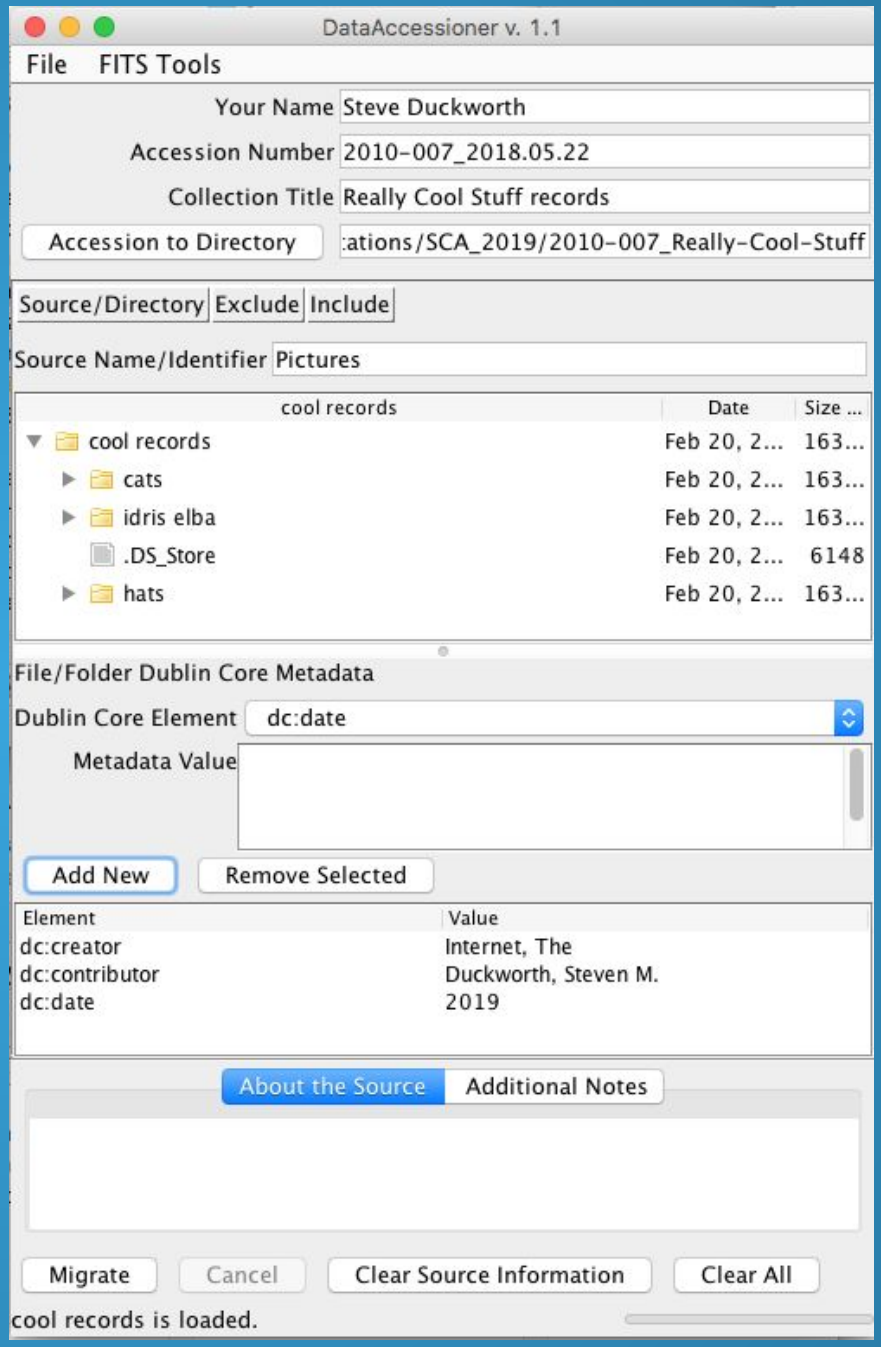
- migrate data
- capture metadata
- create preservation data (checksums)

- assists with description



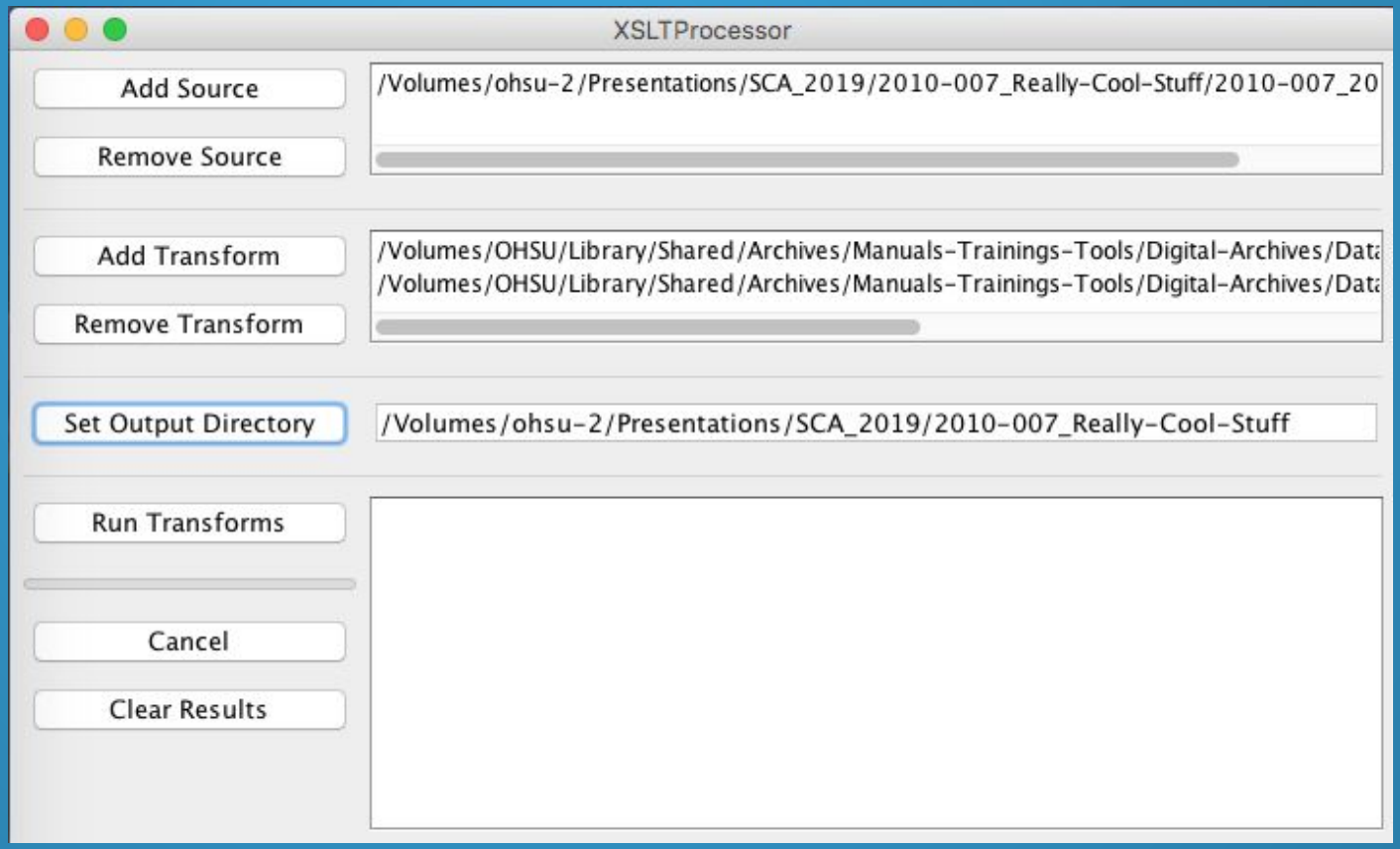
INGESTING RECORDS with DATA ACCESSIONER

- fill in your metadata
- “Migrate” does all the work for you



BONUS TALK:
DATA ACCESSIONER METADATA TRANSFORMER

- use the XML file from the migration to create more useful documents



BONUS TALK:
DATA ACCESSIONER METADATA TRANSFORMER

- CSV file can help with quick descriptions

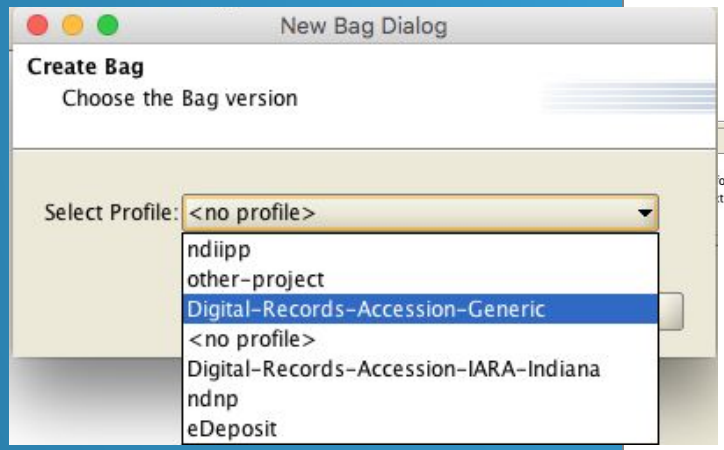
directory path	file name	file extension	last modified	size (bytes)	md5	file format	format version	format registry key/PUID	Identified By	MIME Type	ID Warning
Pictures/cats/	cool-cat-with-sunglasses-768x410.jpg	jpg	2019-02-20T	37118	b71ce62288b4ba00b99cfb08616e831c	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/cats/	moustache-cat(1).jpeg	jpeg	2019-02-20T	67519	8837e8508cc0a3a352042a171ec17c7f	JPEG EXIF	2.2.1	fmt/645	Droid v6.1.5	image/jpeg	DROID Identification Warning:
Pictures/cats/	f5902de28b467619409f16e75449aa4a--shark-hat-the-shark.jpg	jpg	2019-02-20T	8285	a6904085d180a007366029ed8446511a	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/cats/	5bad360e260000360080d8a2.jpeg	jpeg	2019-02-20T	56176	b808cb7451d86b4f5cf7f0fb159c99c2	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/idris elba/	Idris_Elba-4580_(cropped).jpg	jpg	2019-02-20T	5220431	fddd9ab0840532b0e5c73e79c24d591e	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/idris elba/	Unknown.jpeg	jpeg	2019-02-20T	4601	e489e266623c54033d6cea8935549b66	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/idris elba/	Unknown-1.jpeg	jpeg	2019-02-20T	6239	aa79b8adbd379d7a03ba73bfa5953a0b	JPEG File Interchange Format	1.01	fmt/43	Droid v6.1.5	image/jpeg	
Pictures/	.DS_Store	DS_Store	2019-02-20T	6148	c136b888c5835a190671bef774e0986b	DS_store file (MAC)		fmt/394	Droid v6.1.5	application/(	DROID Identification Warning:
Pictures/hats/	fox.jpg	jpg	2019-02-20T	129022	3fd51c054376fc293257df43e8e43720	jpeg			ffident v0.2	image/jpeg	
Pictures/hats/	article-2336777-1A2D308B000005DC-196_634x808.jpg	jpg	2019-02-20T	96785	a94df9db61a3fab1dd9e0dd8286be1e8	JPEG File Interchange Format	1.01		Exiftool v10.:	image/jpeg	
Pictures/hats/	brains.jpg	jpg	2019-02-20T	137047	74a2596e520f2b013b57153190228337	jpeg			ffident v0.2	image/jpeg	

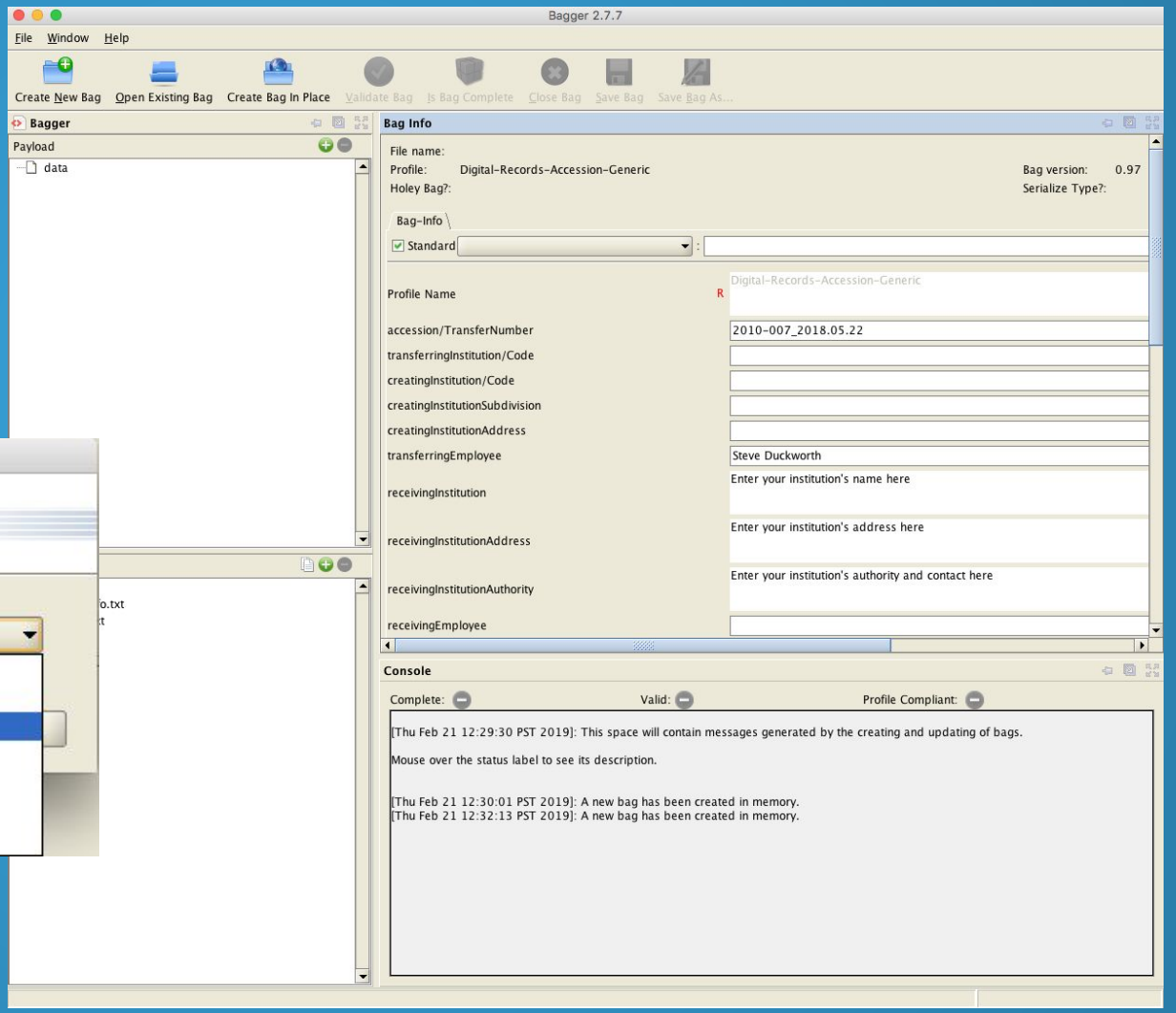
PRESERVATION with BAGGER

- group together your data (records) and metadata (description)
- like a digital archival box
- not compressed

PRESERVATION with BAGGER

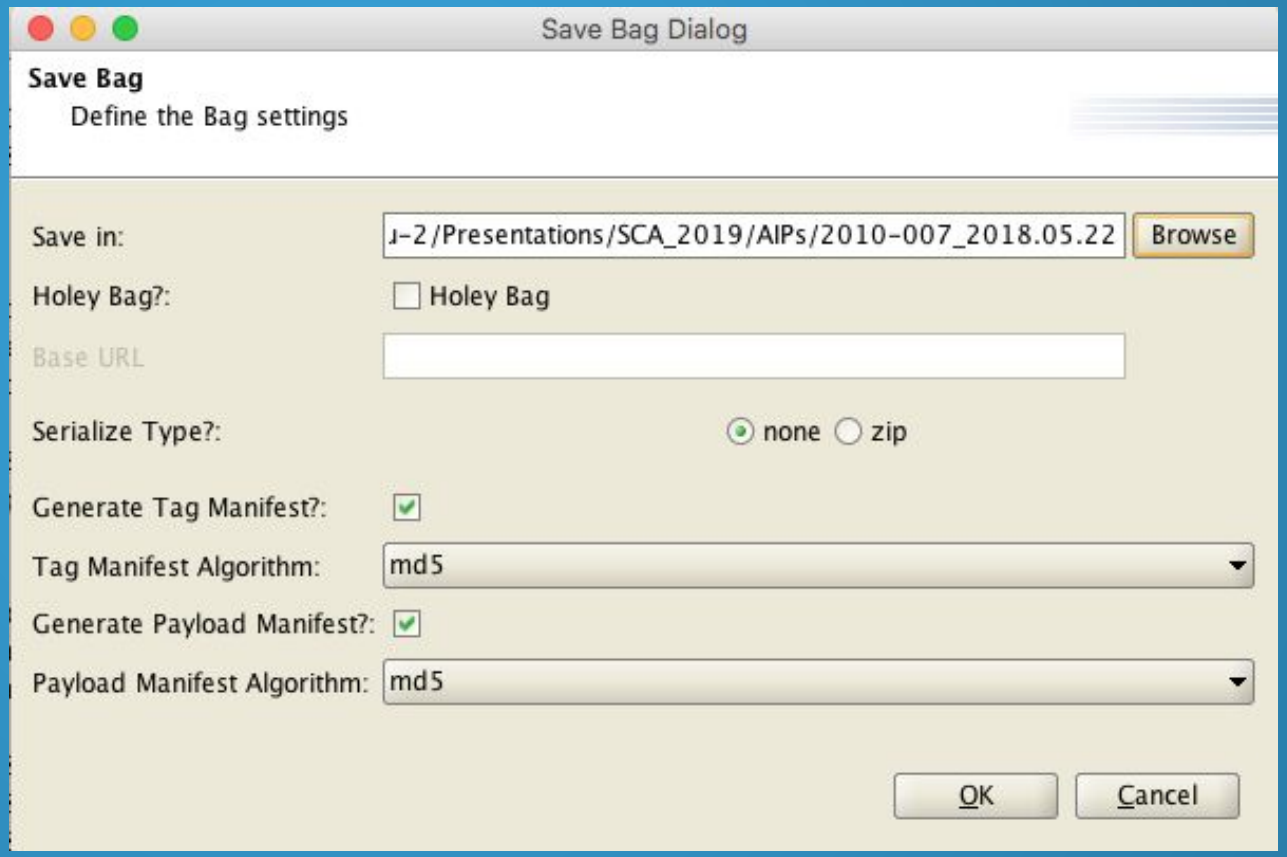
- more metadata





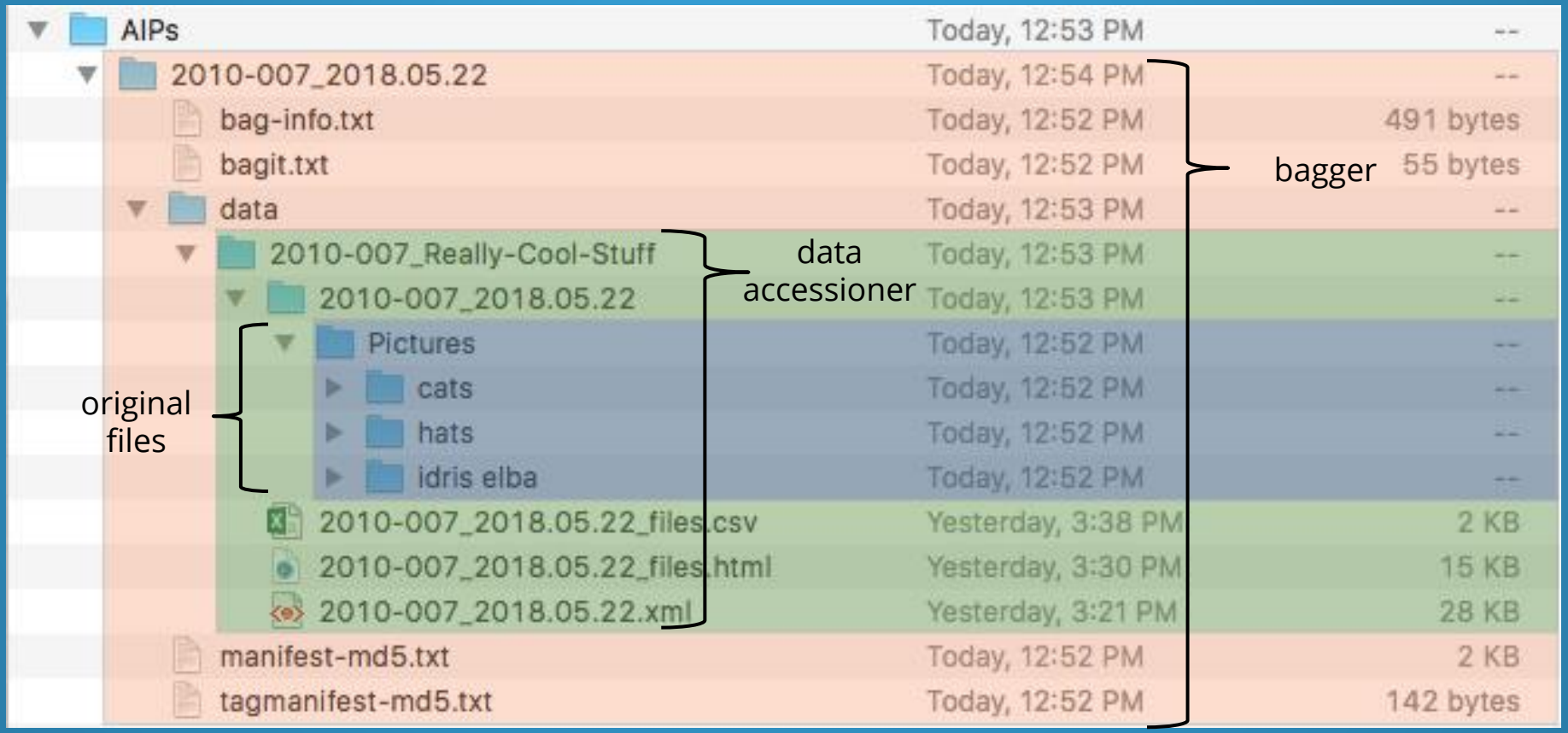
PRESERVATION with BAGGER

- save that bag!



OVERVIEW

- structure of the process



ACCESSIONING: KEY TAKEAWAYS

DATA ACCESSIONER

transfer records without messing them up

capture metadata

DATA ACCESSIONER METADATA TRANSFORMER

reuse already captured metadata

less work for you

BAGGER

group files and metadata together for preservation and portability

4

STABILIZE

Migrating from storage media

Steve Kutay, California State University, Northridge

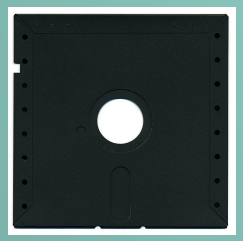
BARRIERS TO ACCESS
A REVIEW

LESS RISK

RELIABLE

OBSOLETE &
HARDWARE
DEPENDENT

ACTIVE &
HARDWARE
INDEPENDENT



FAILURE PRONE

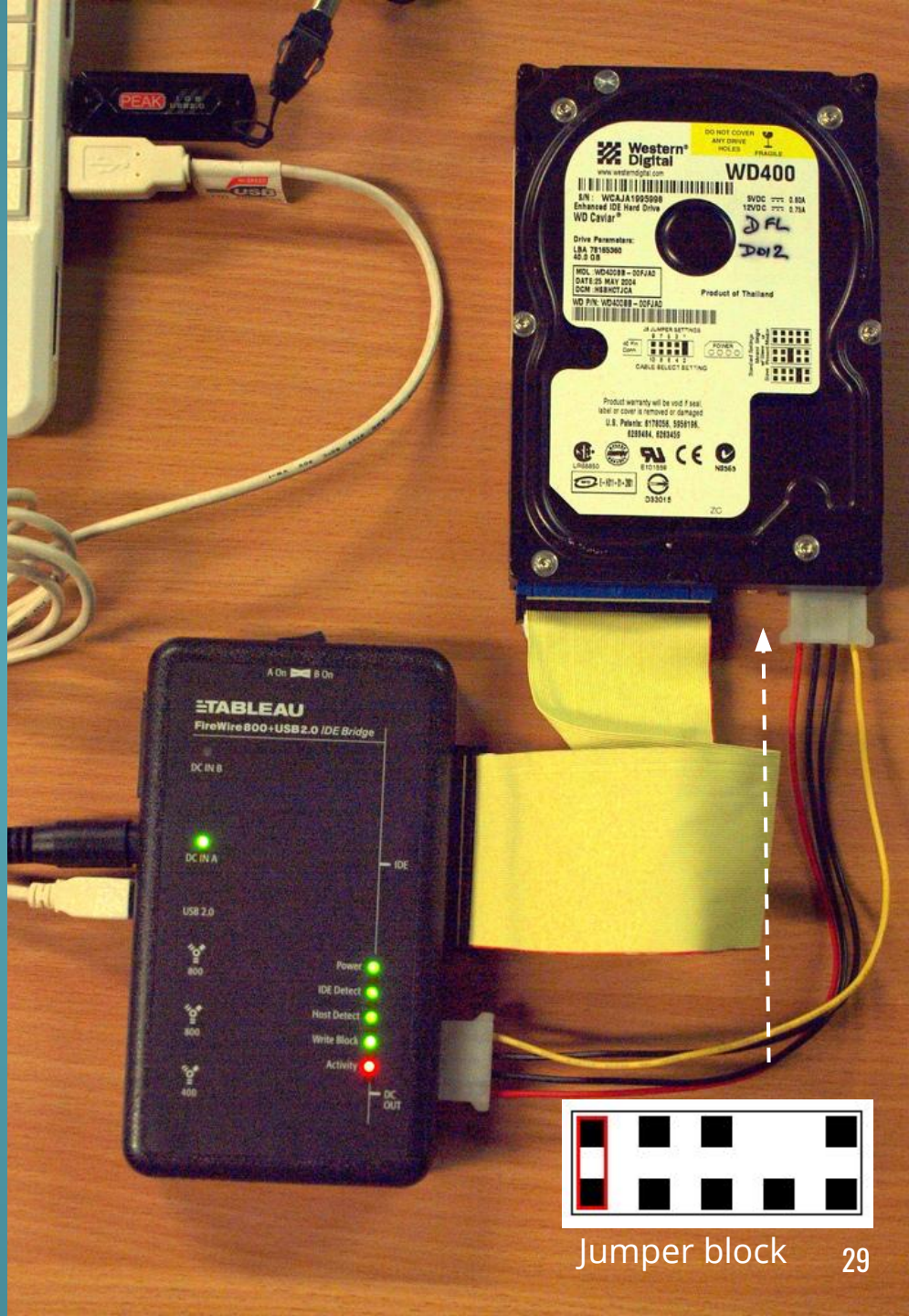
MORE RISK

MIGRATION WORKFLOW

Creating a preservation package from source to vault

PART 1: PREPARE MEDIA

1. Assess drive and jumper block (if necessary)
2. Take photos
3. Document
4. Connect
5. Create destination directory



Jumper block

Jumper_01.png by Zulucho [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]

PART 2: DESCRIBE and ANALYZE

USING SOFTWARE:

1. Add contextual and administrative metadata regarding files
2. Extract technical metadata
3. Generate fixity and other analytic reports
4. OPTION: Consider a *forensic disk image* for further analysis and storage

FORENSIC ANALYSIS:

- Virus scans
- Checksums
- Format validations
- Hashes/de-duplication
- File inventories
- Directories/structure
- Private information
- Hidden/deleted files
- Web history
- And more

* THE FORENSIC DISK IMAGE

(optional copy method)

WHAT IS IT?

A bit-for-bit copy (bitstream) of all files (visible and hidden), folders, and free space as a single file. A stable and mountable preservation format.

WHEN DO I USE IT?

Acquire images when access to hidden information in the storage medium is desired. Most useful for media containing system files and logs. *Consider donor privacy.*

HOW DO I CREATE IT?

Open source and proprietary software applications. Some are format-specific.

PART 3: EXPORT

EXPORT ALL AS A PACKAGE

Data packaging software (e.g. BagIt, Bagger) bundles disk image or source files with metadata, reports, and other sidecar files for transmission as part of an archival information package (AIP) of the collection.

CONFIRM/MAINTAIN FILES

- Check destination folders for accuracy
- Duplicate preservation file package for use as *access files*, but remove redactions*
- Write-protect your access drive, or purge and replace access package after each use*

*access policies will vary by institution

PART 4: PRESERVE

SECONDARY MIGRATION

Migrate AIP (disk image, files, metadata, reports) to long-term storage destinations.

REDUNDANCY

Protect against human/mechanical errors and catastrophic damage by storing 2-3 copies in 2 or more locations.

PHYSICAL MEDIA PLAN

Create policy a policy regarding the retention or disposition of the physical storage media once redundant preservation is achieved.

STILL WANT MORE?

Tools to consider to develop your capacity...

BitCurator

Grant-funded, open source virtual machine bundles apps for disk imaging, integrity checks, analysis, redaction, directory reports, deduplication, and more.

FORENSIC WORKSTATIONS

Assembled for forensic computing. Contains numerous write-protected ports and storage bays for acquiring, analyzing, and storing disk images.
$5000 - $8000

ACTIVITY

Small group discussion

5

THANKS

Emily evigor@berkeley.edu

Kate dundon@ucsc.edu

Steve D duckwors@ohsu.edu

Steve K stephen.kutay@csun.edu

Anything is better than nothing: Minimum viable actions for accessioning born-digital

Resource list

Getting started

- *You've Got to Walk Before You Can Run: First Steps for Managing Born-digital Content Received on Physical Media* https://www.oclc.org/research/publications/library/2012/oclcresearch-managing-born-digital-physical-media.html
- *Walk This Way: Detailed Steps for Transferring Born-digital Content from Media You Can Read In-house* https://www.oclc.org/research/publications/library/2013/oclcresearch-transferring-born-digital.html
- *The No-Nonsense Guide to Born-Digital Content* https://www.alastore.ala.org/content/no-nonsense-guide-born-digital-content
- Digital POWRR Project (Preserving digital Objects With Restricted Resources) https://digitalpowrr.niu.edu
- Outfitting a Born Digital Archives Program https://practicaltechnologyforarchives.org/issue2_goldman/
- Museum of Obsolescent Media https://obsoletemedia.org/
- *Archival application of digital forensics methods for authenticity, description and access provision.* Christopher A. Lee (Author). https://ils.unc.edu/callee/p133-lee.pdf

Processing and description

- *Digital Processing Framework* https://ecommons.cornell.edu/handle/1813/57659
- *UC Guidelines for Born-Digital Archival Description* https://github.com/uc-borndigital-ckg/uc-guidelines

Workflow planning

- AIMS White Paper https://dcs.library.virginia.edu/aims/white-paper/
- OSSArcFlow “as-is” workflows https://educopia.org/ossarcflow/

Education

- *Moving Beyond the Lone Digital Archivist Model Through Collaboration and Living Documentation* https://blogs.princeton.edu/techsvs/2017/03/28/moving-beyond-the-lone-digital-archivist-model-through-collaboration-and-living-documentation/
- DigiPET: A Community Built Guide for Digital Preservation Education + Training https://docs.google.com/document/d/16KGPT-IsQ39PSFN1p3eu1jWHQccAydB6uyWJmWfyA7E/edit
- DLF Born Born Digital Access Bootcamp https://www.diglib.org/groups/born-digital-access-group/
- BitCurator Users Forum 2019 https://bitcuratorconsortium.org/bitcurator-users-forum-2019

Tools referenced in presentation

- Bagger https://github.com/LibraryOfCongress/bagger
- Bitcurator https://bitcurator.net/
- Duke Data Accessioner http://dataaccessioner.org
- Data Accessioner Metadata Transformer http://dataaccessioner.org/da-mt.htm
- DROID http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

Activity questions

1. Describe your current capacity for accessioning born digital content, focusing on the processes of appraising, capturing, and ingesting files into your repository.

2. Describe one thing you need/want to be able to do in accessioning born digital. Pick something specific, big or small.

3. What do you need to accomplish this at your institution?

4. Describe one thing that you learned today that you can realistically apply to help you achieve this goal.

Discuss your answers to these questions with your group. Consider: what is standing in your way of achieving your goal in question 2?