

Scaling Born-Digital Appraisal

Adriane Hanson and Emmeline Kaser, University of Georgia (USA)

DPC Reappraising Appraisal Event

September 4, 2024

Introductions



University Libraries
UNIVERSITY OF GEORGIA

- Adriane Hanson, Head of Digital Stewardship
- Emmeline Kaser, Digital Archivist, Digital Stewardship Unit

Presentation Outline

Context

Born-digital appraisal in practice

Scripting to scale

Looking ahead

Context

Institutional context

- Unified born-digital workflow for two special collections libraries
 - Hargrett Rare Book and Manuscript Library
 - Richard B. Russell Jr. Library for Political Research and Studies
- Archivist(s) for each library do collecting and donor relations
- Digital Stewardship has technical expertise and does the collections work
- Started in 2012, overhauled in 2021 with more emphasis on appraisal
- 5 TB Born-digital, non-AV materials: small but complex

Working definitions

Accessioning: getting new collections safely stored and minimally documented so they can be managed while in the backlog

Appraisal: deciding what to keep and not keep

Reappraisal: revisiting a previous decision to keep something based on new information, such as research use or labor needed to keep it usable. Appraisal is an iterative process.

Born-digital: created digitally and donated to us, as opposed to digitized from analog. Born-digital AV is handled by another unit.

Why appraisal is important



Saves storage space



Saves time monitoring and addressing risks



Improves researcher experience

Why scaling is important

(and scripting is integral)



Small size may contain large number of files



Annual accession amounts increasing



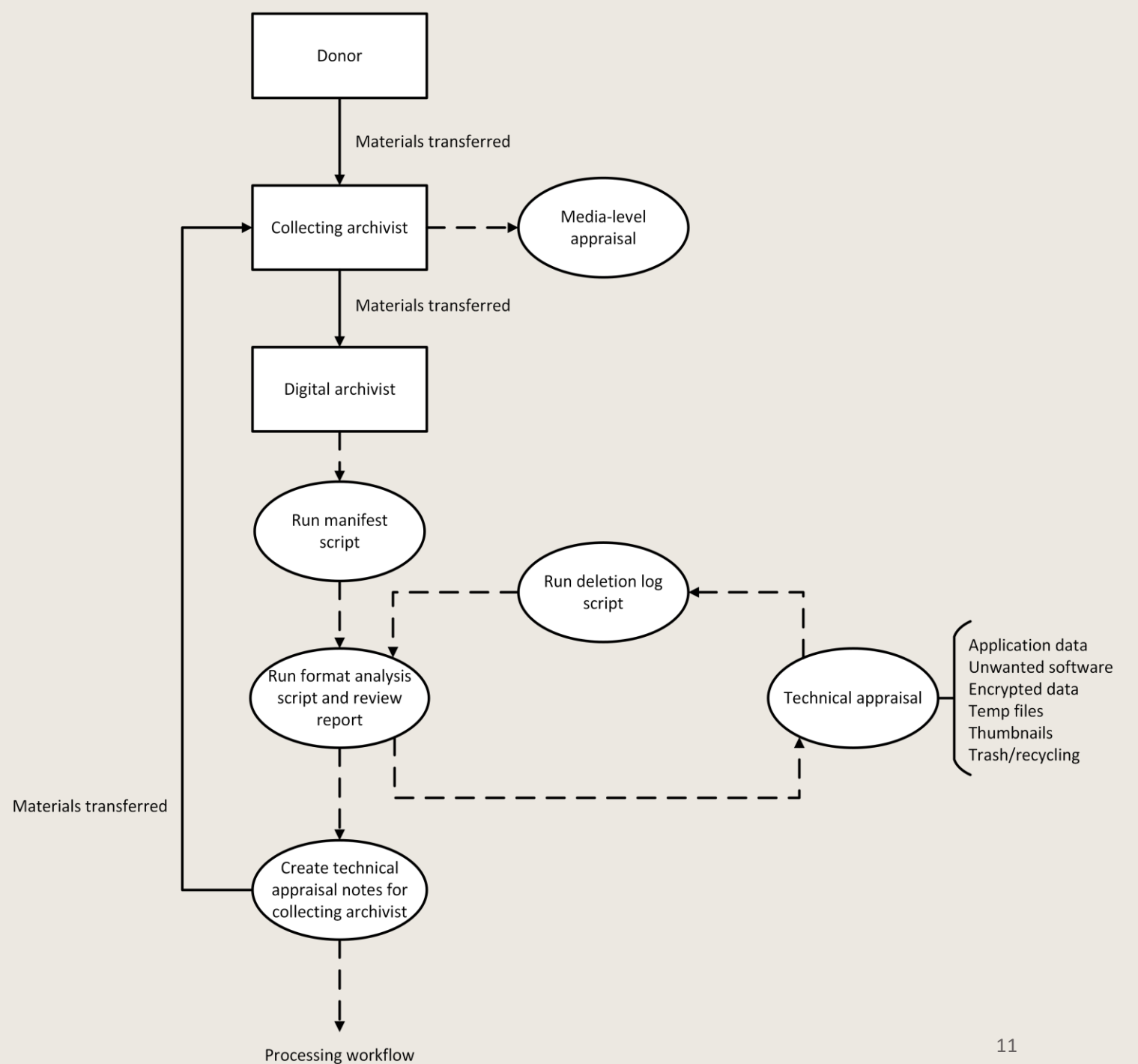
Don't burn out our staff

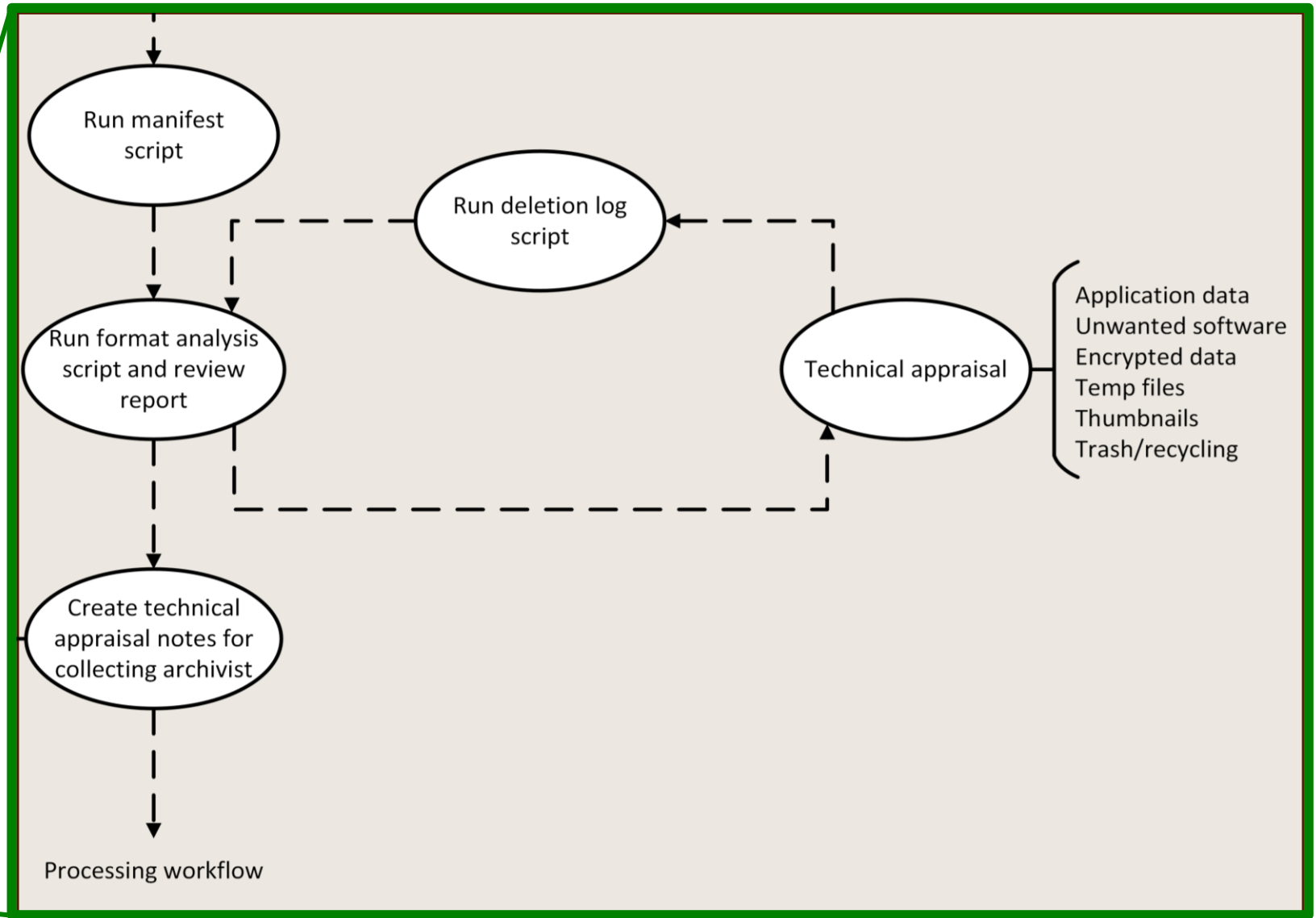
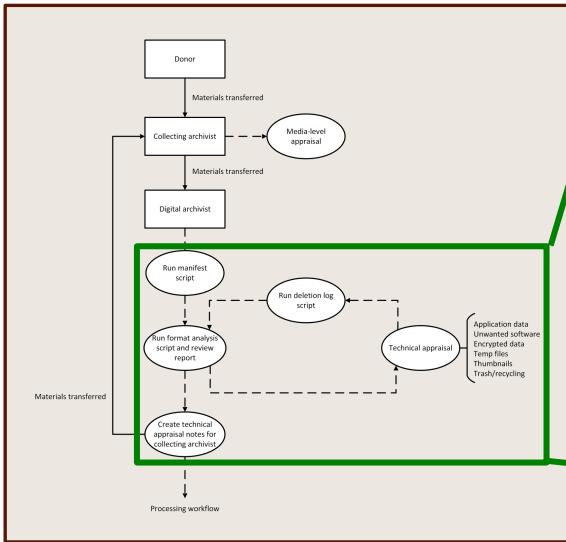
Born-digital appraisal in practice

A two-pronged approach to iterative appraisal

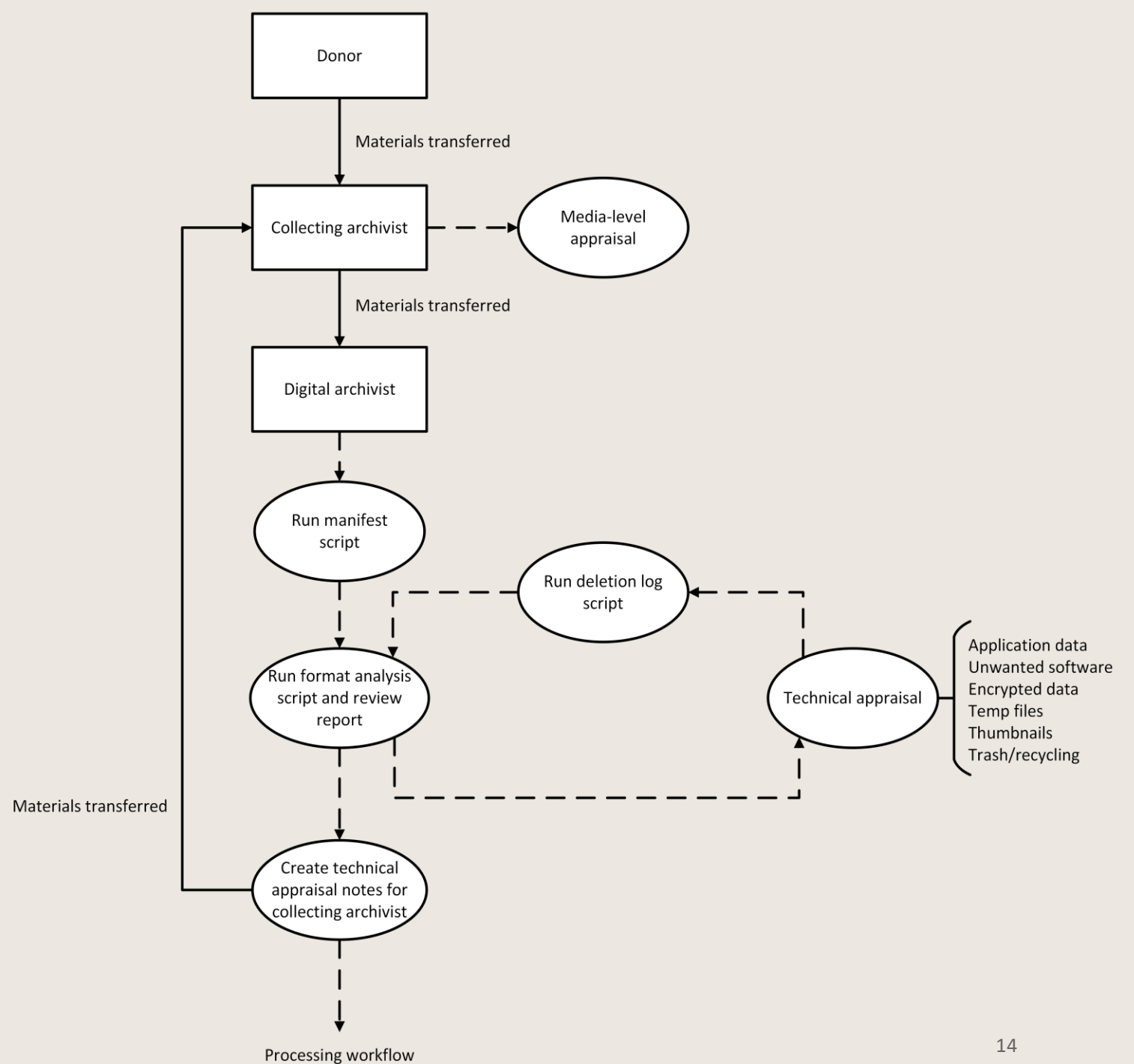
	Technical Appraisal	Content Appraisal
Assesses:	<ul style="list-style-type: none">• the materials' usability and preservation needs	<ul style="list-style-type: none">• the materials' research value and uniqueness
Informed by:	<ul style="list-style-type: none">• knowledge of digital preservation best practice and current capabilities• digital preservation policies	<ul style="list-style-type: none">• collection expertise and knowledge of researcher interest• collecting policies
Results in:	<ul style="list-style-type: none">• removal of incidental and unwanted data• a pared-down set of files to share with collecting archivist for content appraisal	<ul style="list-style-type: none">• removal of out-of-scope materials• a finalized set of files that is ready to be processed and ingested

Iterative Workflow





Iterative Workflow



Reflecting on our process: successes

- Saves storage space
- Reduces complexity going into processing
- Gives digital archivist a better understanding of the collection
- Format-based complexity now factors into content appraisal
- Removes some of our most common high-risk formats (executables)
- Automated deletion logging improves documentation, reduces human error

Reflecting on our process: challenges

- Asynchronous collaborative workflow has created bottlenecks in processing queue
- Limits to technical appraisal's scope
- Mindset shift of appraising born-digital archives when folks are used to appraising paper

Scripting to scale

<https://github.com/uga-libraries/accessioning-scripts>

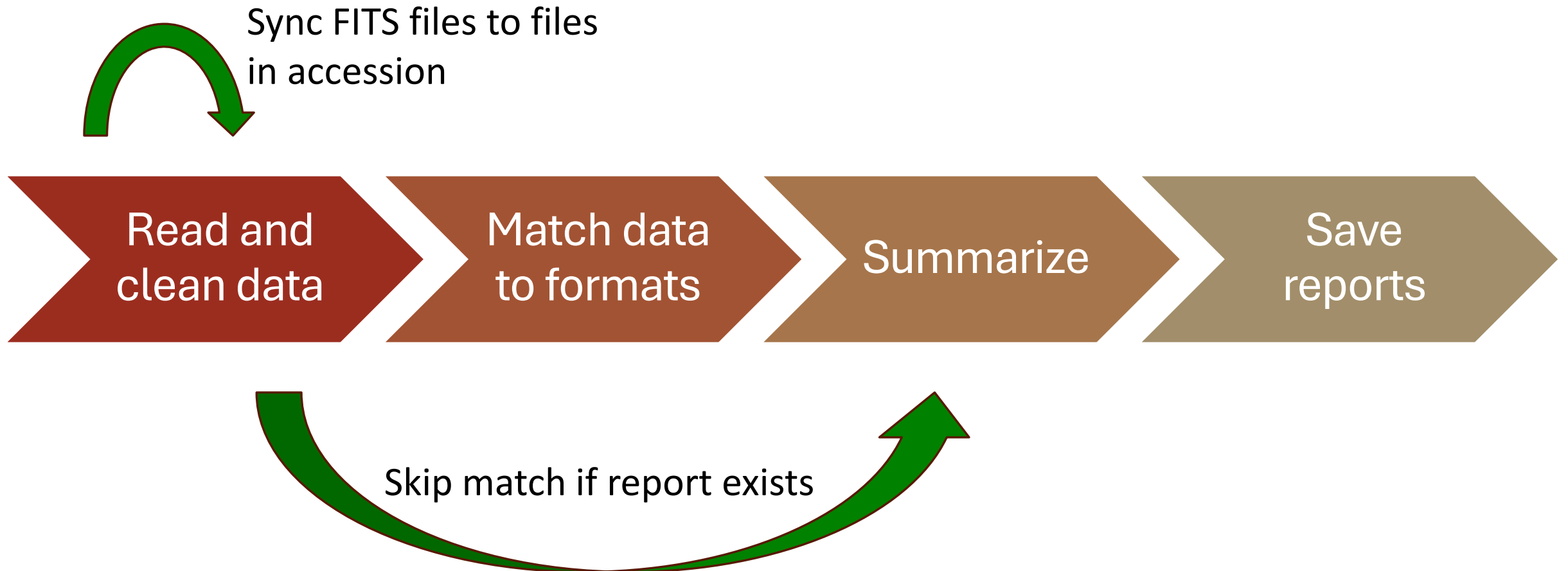
Input Data



Script workflow



Supporting iteration



Looking ahead

What's next for automation?



Refine script for better matching and more useful summaries



Finishing script that includes collection summaries



Dabbling with text analysis



Scripting other aspects of the workflow to free time for appraisal

What's next for our workflows?



Addressing workflow bottlenecks with new communication methods



Encouraging more donor communication and up-front info gathering



Creating a list of media that is no longer accepted



Creating reappraisal opportunities through regular audits of the born-digital backlog

Thank you!

- Related resources
 - [Accessioning scripts repo \(GitHub\)](#)
 - [Hub monitoring repo \(GitHub\)](#)
 - [Appraisal workflow \(GitHub\)](#)
 - [List of format IDs for technical appraisal \(GitHub\)](#)
 - BitCurator Forum 2023 Presentation: [Collaborative Appraisal of Born-Digital Archives](#)
- Contact us!
 - ahanson@uga.edu
 - ekaser@uga.edu