



Implementing BitCurator at Rice University: *Baby Steps to Digital Preservation Glory*

Rebecca Russell and Amanda Focke
Archivists. Detectives.





Implementing BitCurator at Rice University: *Baby Steps to Digital Preservation Glory*

Rebecca Russell and Amanda Focke
Archivists. Detectives.



About us

Rebecca Russell
Archivist/Special
Collections Librarian



Amanda Focke
Asst. Head of Special
Collections

Fellow digital detectives / archivists at Rice: **Dara Flinn & Norie Guthrie**
"It takes a village!"



The Year of Digital Preservation

- Departmental goals for 2014
 - put in place the needed
 - institutional support
 - policies
 - workflows
 - software, hardware, and
 - staff skills
- *to have a true working digital preservation program.*
- BitCurator is an important tool in this process



The year of Digital Preservation

Step 1: Developing a digital preservation plan

Reading / Training including

- SAA's Arrangement and Description of Electronic Records, and SAA's Digital Forensics.

Starting with high level policy, **in baby steps**

- added digital preservation to our mission statement
- published a high-level DP policy using NEDCC template
- this shows institutional commitment
- Details of our efforts on a publicly accessible wiki (<http://bit.ly/1pjrcZW>)



The year of Digital Preservation

Step 2: Getting real with our electronic holdings

- significant legacy holdings in electronic formats on a variety of unbacked-up original storage media scattered in the collections
- constant flow of new electronic accessions.

Baby steps:

- conducted an inventory of e-holdings in our collections
- made guidelines for new donors of e-formats



The year of Digital Preservation

Step 3: Adapting our traditional processes to include the digital

Baby step: edited our existing processing manual

Tracking & Storing the Accession (reword into OAIS terms of SIP and AIP)

Assign each "item" a **unique identifier #** (possibly a **new filenames convention for disk images**, such as MS356di_001 for a disk image from MS 356) in Archivists Toolkit as a digital object)

As with physical archives, it is not always practical to describe materials at the item level – it will not be possible to describe each digital file. Archivists must decide on materials at what level it makes sense to assign a number that identifier numbers can have suffixes to

messy &
incomplete –
but that's ok!

... disk image # on Fondren Library's project server
(fonli) ... can ... SIPs (Z:\WRC\1_SIPs) which are redundantly
backed up by Fondren with incremental nightly back-ups and monthly full back-ups.
Watch out for the overall WRC folder size limit of 300GB.

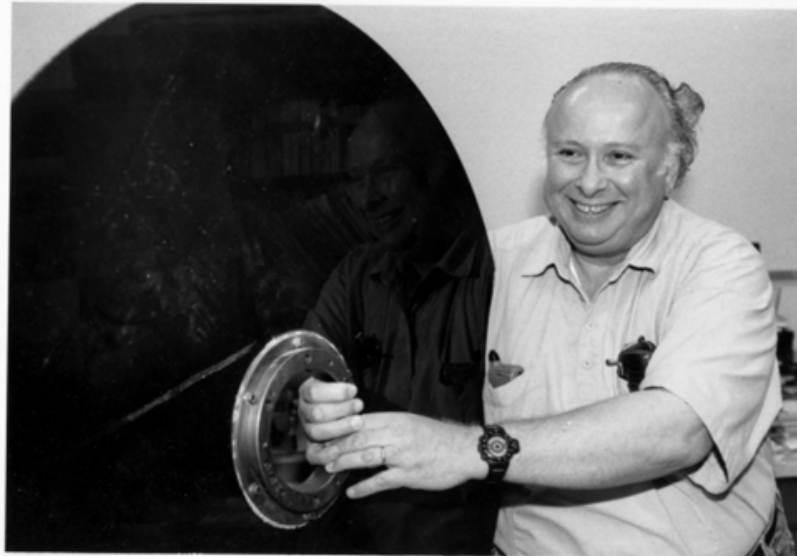
Main additions:

- creating a **disk image**
- extracting info for an **OAIS compliant "package"**



Step 3: Adapting our traditional processes to include the digital, cont...

What's a disk image?



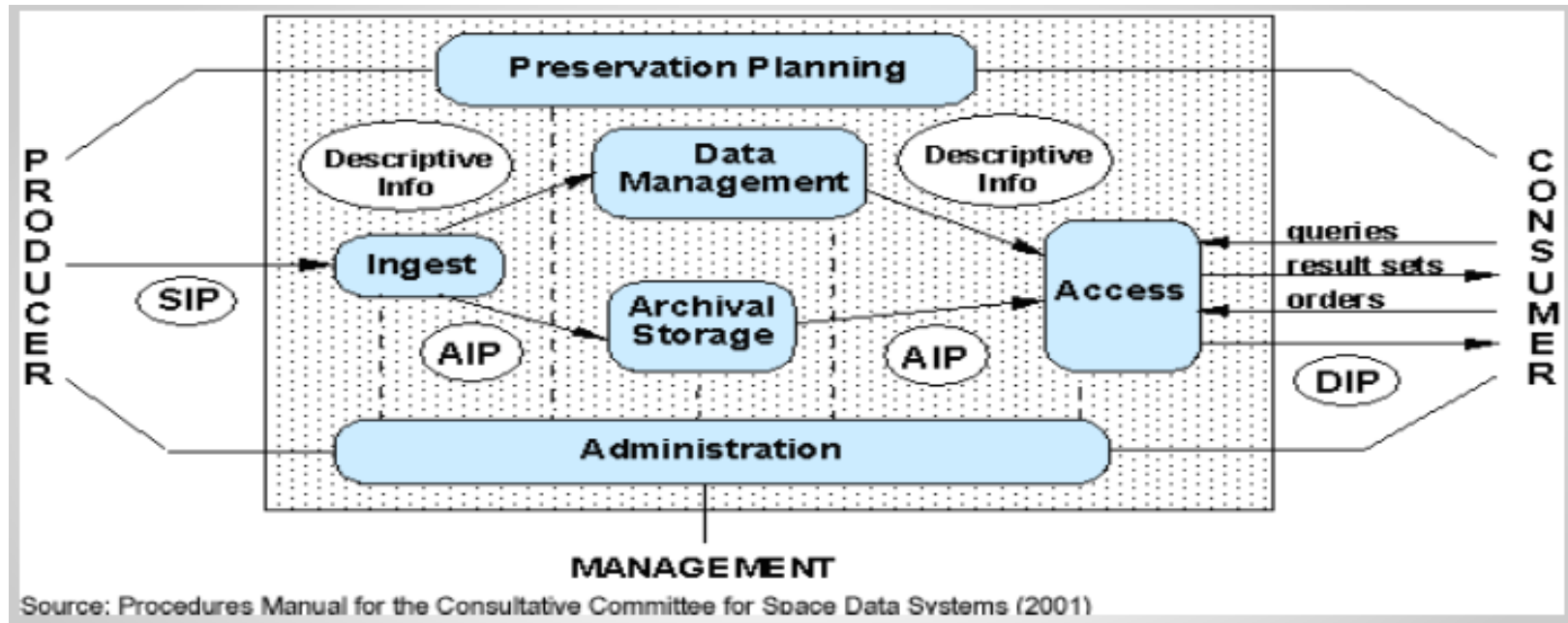
Joel Cyprus with disc platter from "Bryant bulk memory installed" in enclosed module, Rice University Computer Project, <http://scholarship.rice.edu/handle/1911/71814>

Disk imaging is the process of extracting unaltered bitstreams from digital media; it creates a perfect capture of a device's file structure and all contents (including hidden files and fragments) into one file, the "disk image"



Step 3: Adapting our traditional processes to include the digital, cont...

What is OAIS?



- Open Archival Information System – a functional model
- Natural /easy in the analog world
- Requires some deep thought in the digital world (in baby steps!)



The year of Digital Preservation

Step 4: Identify tools to help us implement a robust OAIS compliant preservation & access system

Baby step:

Realizing BitCurator could specifically help us in:

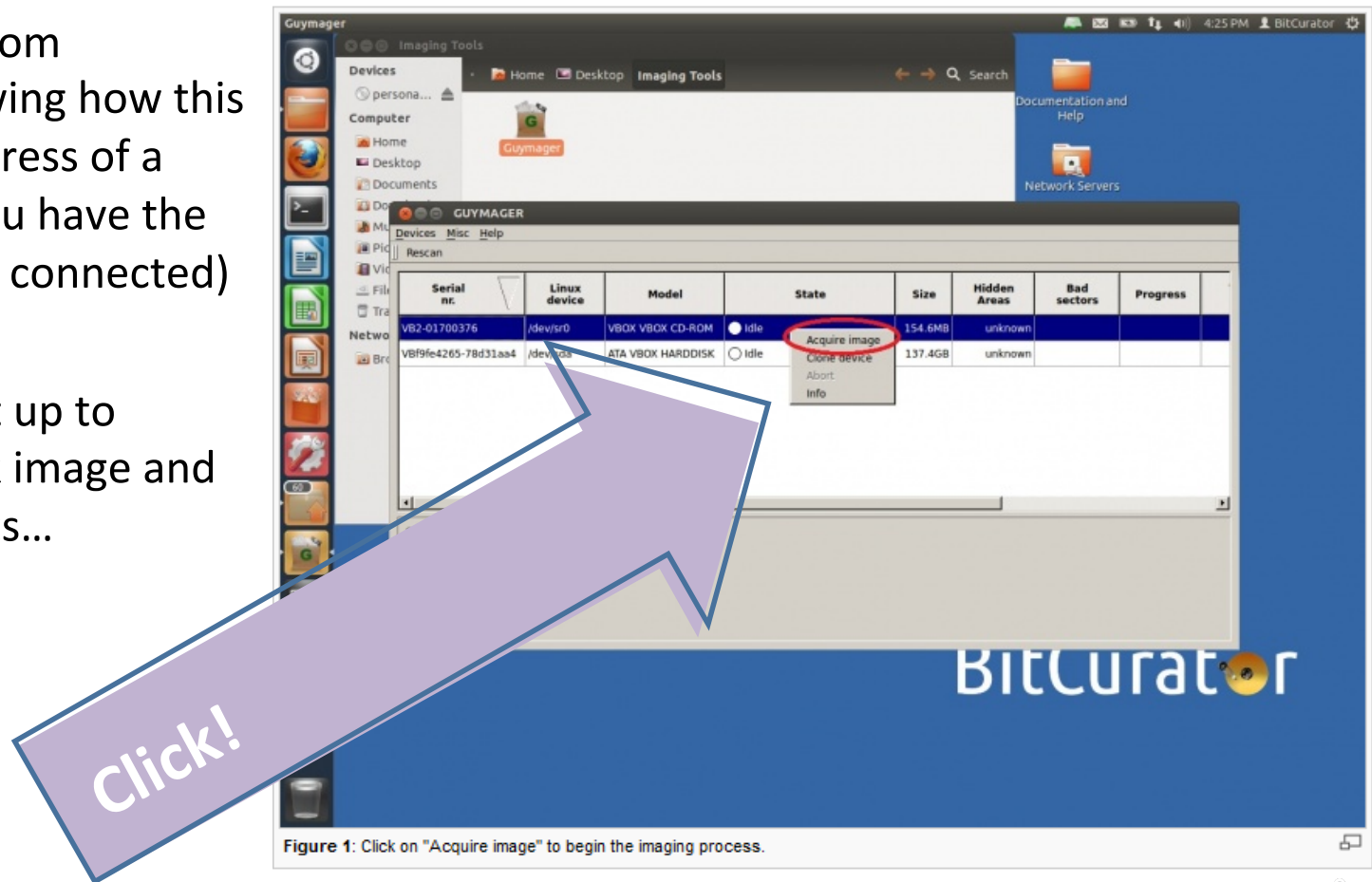
- 1. creating disk images** for more reliable preservation of bitstreams
- 2. extracting the needed information to create OAIS-compliant information packages.**



Step 4: Identify tools to help us implement a robust OAI compliant system, cont...

Making a disk image

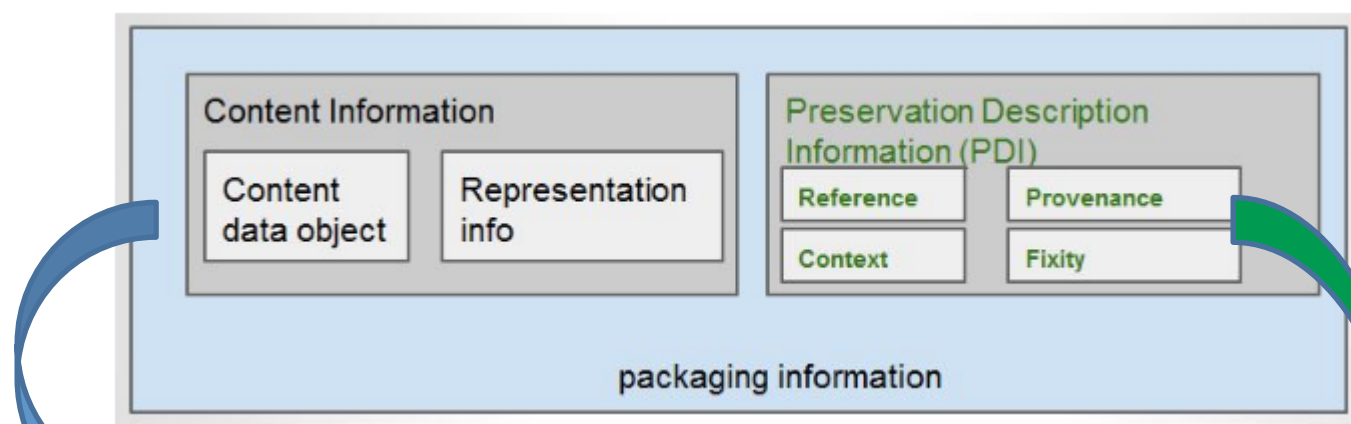
- A screen shot from BitCurator showing how this done with the press of a button (once you have the drive mounted/ connected)
- Then we are set up to analyze the disk image and generate reports...



Step 4: Identify tools to help us implement a robust OAIS compliant system, cont...

BitCurator reports

- Relate directly to “Information Package” parts



Content data object = Disk Image
Representation Info = summary of file formats etc

PDI (Preservation Description Information)

Reference = our unique identifier for the package, which is also the root of the disk image file name

Context =

fiwalk.xls contains the MAC times, file characterization

fiwalk.xml contains the file system type

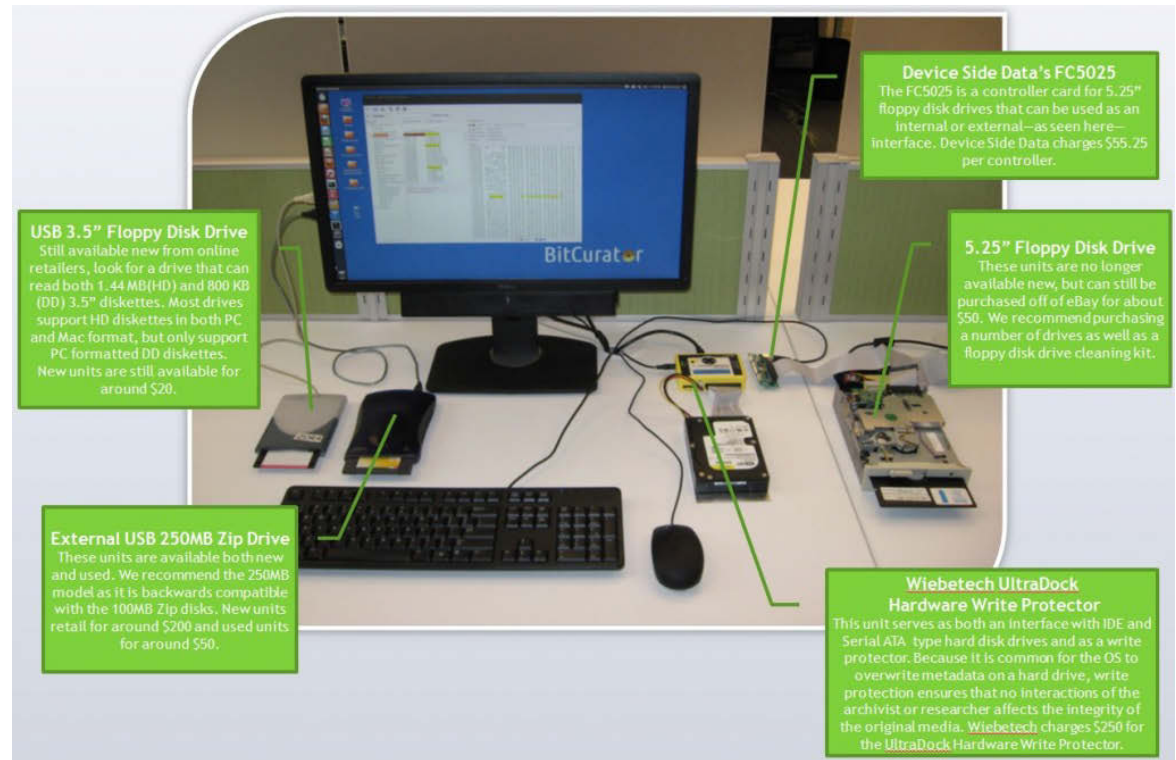
Provenance = written by WRC staff

Fixity = hash values / checksums in fiwalk.xls report



How did we do it?

- **Training:** Digital Forensics workshop and a site visit helped a lot, but the BitCurator wiki is also great
- **Community:** in touch with fellow BitCurator explorers in our state and in the Google group
- **MAINLY:**
Our simple **BABY STEP** set-up: **Working from “live disc environment” as the simplest way to begin**



Porter Olsen's blog post on a forensics set-up includes equipment & prices (<http://www.bitcurator.net/author/porter/page/5/>)



Bumps along the way

- **Archivists' concerns**
 - **Overwhelming nature of so much change** in our procedures at once – baby steps are necessary
 - **What if our simple set up is too limited?** We would like a virtual machine and/ or a dedicated machine...
 - **Are we really going to have the time** to go back and apply these procedures to legacy media?



Bumps along the way

- IT support /concerns
 - “**Why** do you need to do forensics all of the sudden?”
 - “**We aren’t comfortable** with you operating in a Virtual Machine”
 - “You need 1 year minimum of training to do any forensics work and **you would ask us for significant support...**”
 - “How **are we going to store and back up all those disc images** you think you need?”



Jimmy Fallon as Nick Burns, “Your Company’s Computer Guy,” in a recurring “Saturday Night Live” sketch.

NBC image. URL: <http://blogs.wsj.com/digits/2009/07/31/are-you-celebrating-sysadmin-day/>, accessed June 23, 2014

Note: Our IT team is great, but this alarmed them.



Wrap-up / Future?

- We have moved from mission to policy to workflows and tools (and glory)
- BitCurator is a key tool in our program
- We have implemented it in baby steps – starting with the live disc environment, getting comfortable there, refining our workflow, and making the next steps once we more clearly see what they are.



Resources

- Olsen, P. (2012). “Digital Curation Workstation.”
<http://mith.umd.edu/digital-curation-workstation/>
- BitCurator Wiki. (2014).
http://wiki.bitcurator.net/index.php?title=Main_Page
- Digital Preservation Support at Fondren Library. (2014).
<https://digitalriceprojects.pbworks.com/w/page/44763477/Digital%20Preservation%20Support%20at%20Fondren%20Library>



Thanks!

Contact Us!

Rebecca Russell
Amanda Focke
Dara Flinn
Norie Guthrie

All reachable via woodson@rice.edu

Woodson Research Center



About us

Rebecca Russell
Archivist/Special
Collections Librarian



Amanda Focke
Asst. Head of Special
Collections

Fellow digital detectives / archivists at Rice: **Dara Flinn & Norie Guthrie**
"It takes a village!"



The Year of Digital Preservation

- Departmental goals for 2014
 - put in place the needed
 - institutional support
 - policies
 - workflows
 - software, hardware, and
 - staff skills
- *to have a true working digital preservation program.*
- BitCurator is an important tool in this process



The year of Digital Preservation

Step 1: Developing a digital preservation plan

Reading / Training including

- SAA's Arrangement and Description of Electronic Records, and SAA's Digital Forensics.

Starting with high level policy, **in baby steps**

- added digital preservation to our mission statement
- published a high-level DP policy using NEDCC template
- this shows institutional commitment
- Details of our efforts on a publicly accessible wiki (<http://bit.ly/1pjrcZW>)



The year of Digital Preservation

Step 2: Getting real with our electronic holdings

- significant legacy holdings in electronic formats on a variety of unbacked-up original storage media scattered in the collections
- constant flow of new electronic accessions.

Baby steps:

- conducted an inventory of e-holdings in our collections
- made guidelines for new donors of e-formats



The year of Digital Preservation

Step 3: Adapting our traditional processes to include the digital

Baby step: edited our existing processing manual

Tracking & Storing the Accession (reword into OAIS terms of SIP and AIP)

Assign each "item" a **unique identifier #** (possibly a **new filenames convention for disk images**, such as MS356di_001 for a disk image from MS 356) in Archivists Toolkit as a digital object)

As with physical archives, it is not always practical to describe materials at the item level – it will not be possible to describe each digital file. Archivists must decide on materials at what level it makes sense to assign a number that identifier numbers can have suffixes to

messy &
incomplete –
but that's ok!

disk image # on Fondren Library's project server
(fondren) can be found in the **SIPs** (Z:\WRC\1_SIPs) which are redundantly backed up by Fondren with incremental nightly back-ups and monthly full back-ups. Watch out for the overall WRC folder size limit of 300GB.

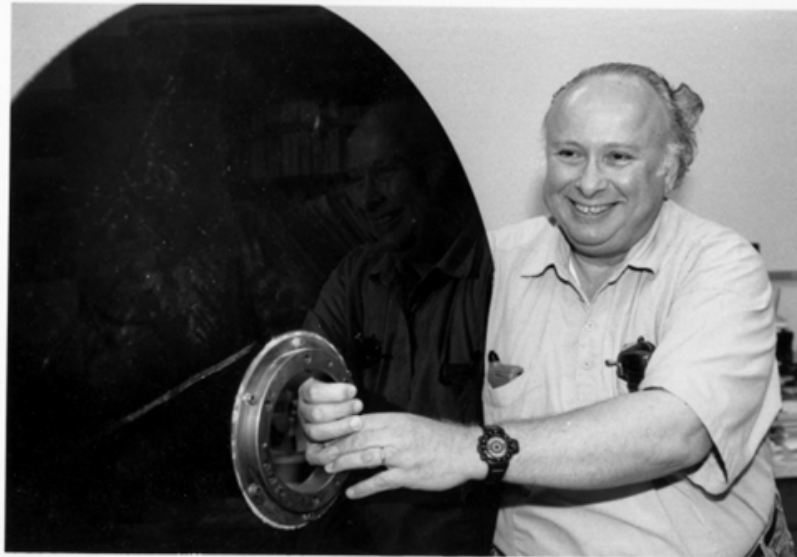
Main additions:

- creating a **disk image**
- extracting info for an **OAIS compliant "package"**



Step 3: Adapting our traditional processes to include the digital, cont...

What's a disk image?



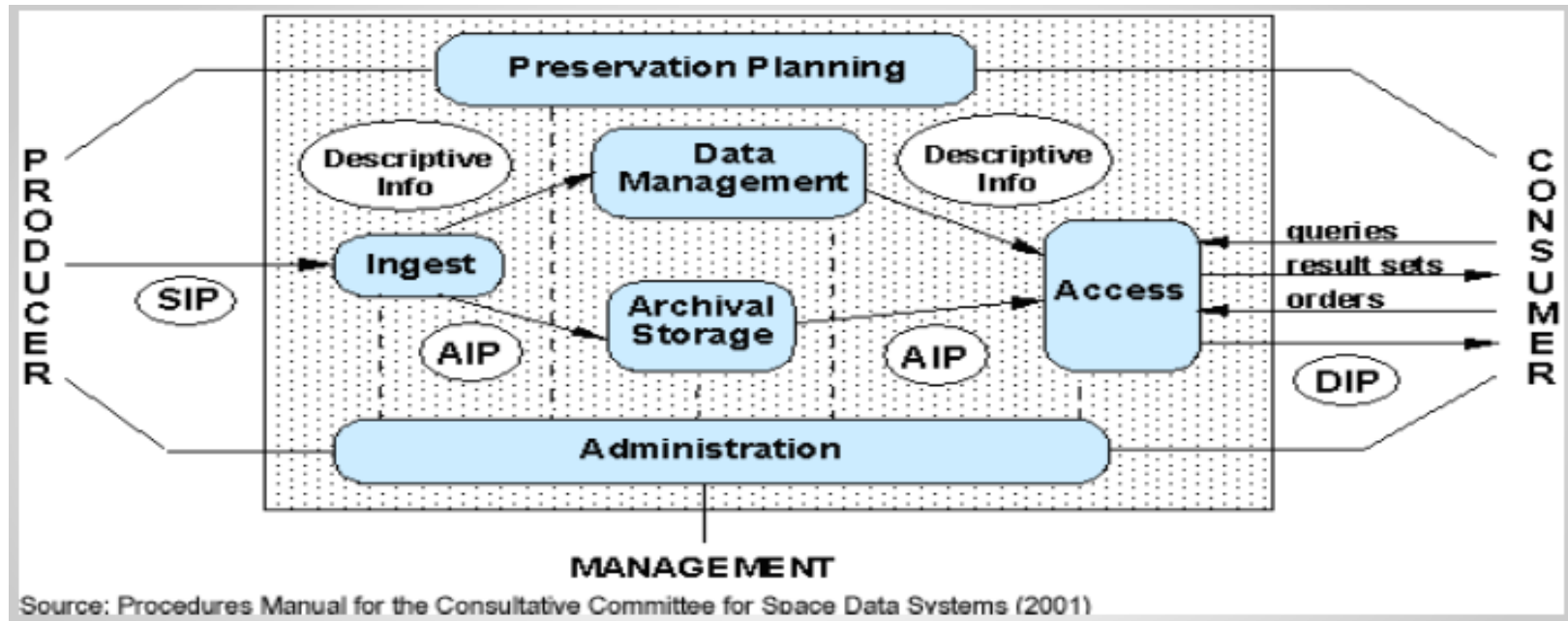
Joel Cyprus with disc platter from "Bryant bulk memory installed" in enclosed module, Rice University Computer Project, <http://scholarship.rice.edu/handle/1911/71814>

Disk imaging is the process of extracting unaltered bitstreams from digital media; it creates a perfect capture of a device's file structure and all contents (including hidden files and fragments) into one file, the "disk image"



Step 3: Adapting our traditional processes to include the digital, cont...

What is OAIS?



- Open Archival Information System – a functional model
- Natural /easy in the analog world
- Requires some deep thought in the digital world (in baby steps!)



The year of Digital Preservation

Step 4: Identify tools to help us implement a robust OAIS compliant preservation & access system

Baby step:

Realizing BitCurator could specifically help us in:

- 1. creating disk images** for more reliable preservation of bitstreams
- 2. extracting the needed information to create OAIS-compliant information packages.**



Step 4: Identify tools to help us implement a robust OAIS compliant system, cont...

Making a disk image

- A screen shot from BitCurator showing how this done with the press of a button (once you have the drive mounted/ connected)
- Then we are set up to analyze the disk image and generate reports...

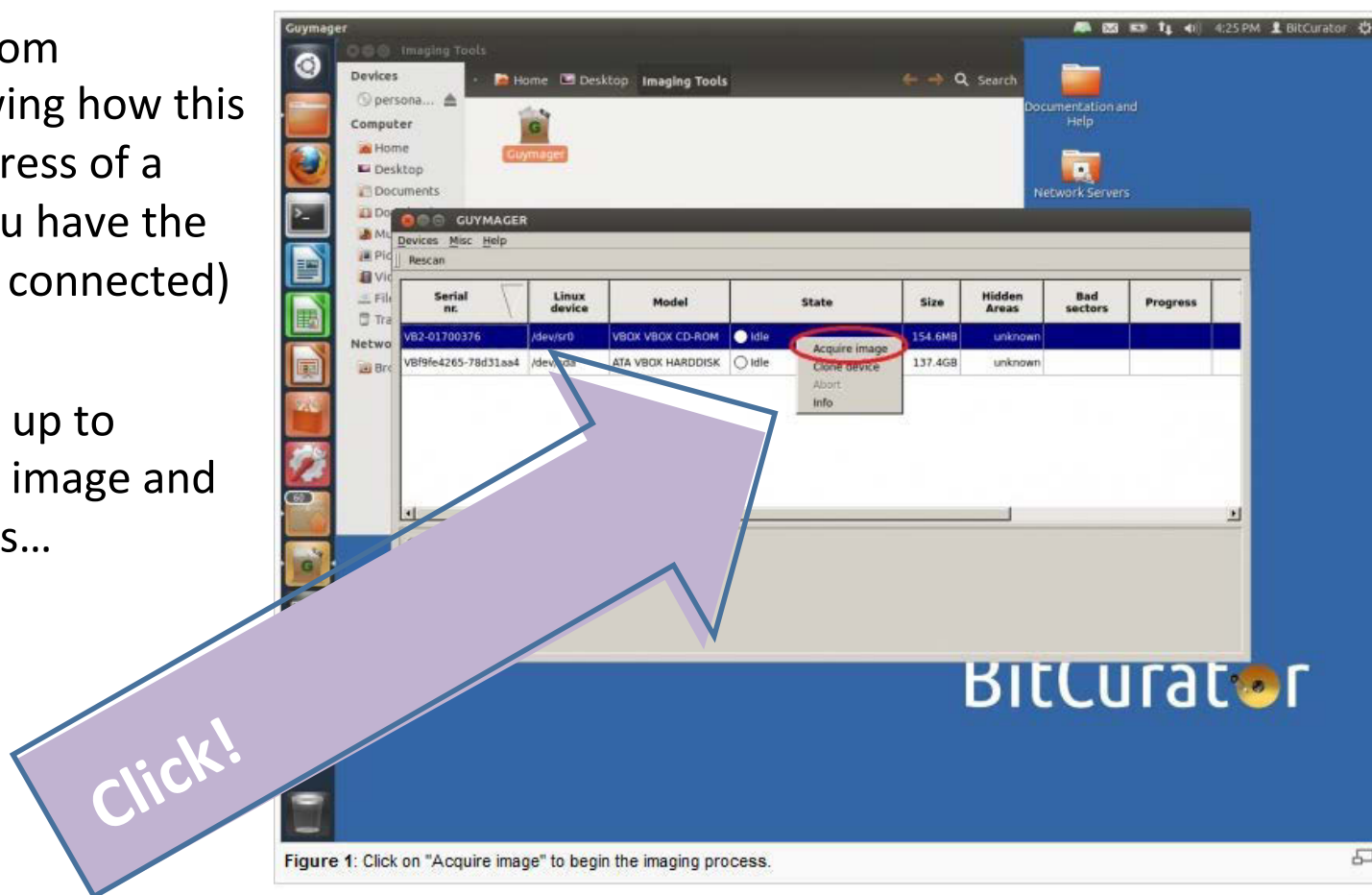


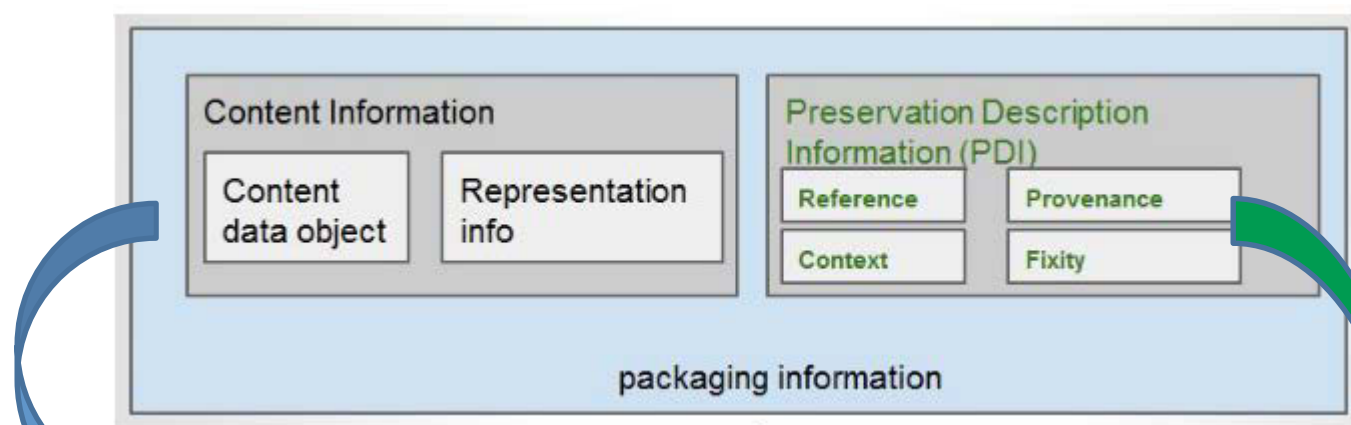
Figure 1: Click on "Acquire image" to begin the imaging process.



Step 4: Identify tools to help us implement a robust OAIS compliant system, cont...

BitCurator reports

- Relate directly to “Information Package” parts



Content data object = Disk Image
Representation Info = summary of file formats etc

PDI (Preservation Description Information)

Reference = our unique identifier for the package, which is also the root of the disk image file name

Context =

fiwalk.xls contains the MAC times, file characterization

fiwalk.xml contains the file system type

Provenance = written by WRC staff

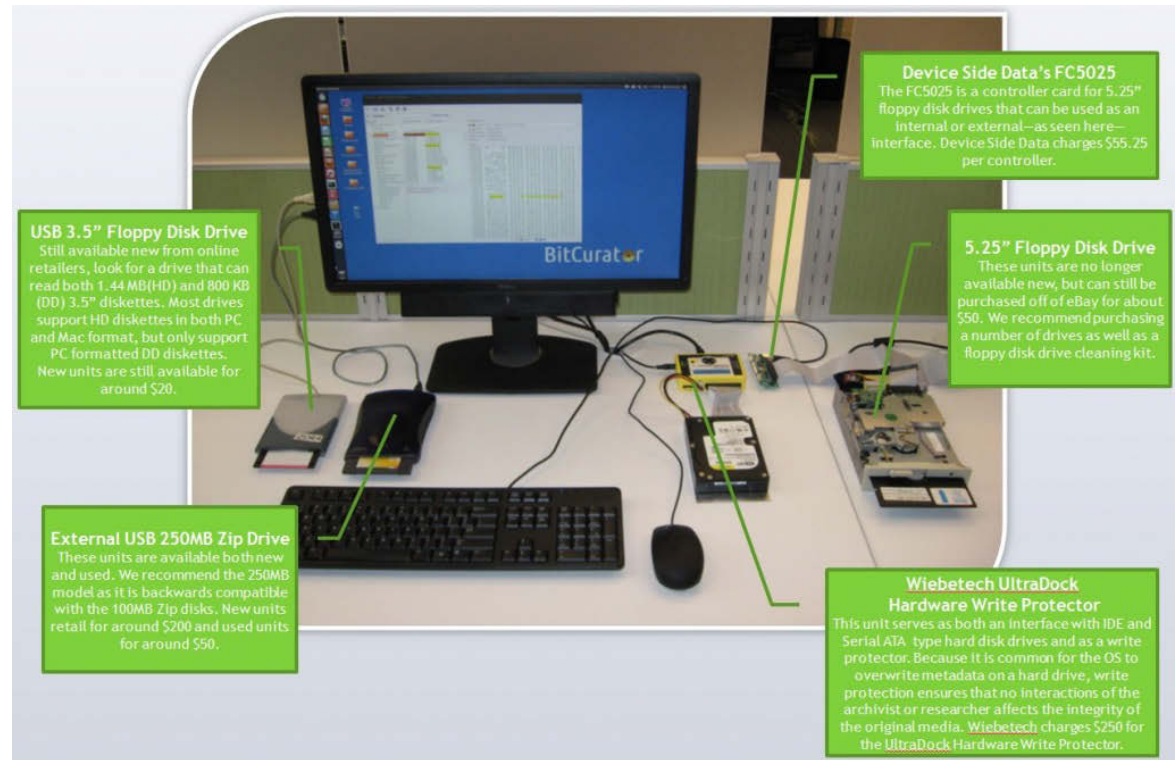
Fixity = hash values / checksums in fiwalk.xls report



How did we do it?

- **Training:** Digital Forensics workshop and a site visit helped a lot, but the BitCurator wiki is also great
- **Community:** in touch with fellow BitCurator explorers in our state and in the Google group

• **MAINLY:**
Our simple **BABY STEP** set-up: **Working from “live disc environment” as the simplest way to begin**



Porter Olsen's blog post on a forensics set-up includes equipment & prices (<http://www.bitcurator.net/author/porter/page/5/>)



Bumps along the way

- **Archivists' concerns**
 - **Overwhelming nature of so much change** in our procedures at once – baby steps are necessary
 - **What if our simple set up is too limited?** We would like a virtual machine and/ or a dedicated machine...
 - **Are we really going to have the time** to go back and apply these procedures to legacy media?



Bumps along the way

- IT support /concerns
 - “**Why** do you need to do forensics all of the sudden?”
 - “**We aren’t comfortable** with you operating in a Virtual Machine”
 - “You need 1 year minimum of training to do any forensics work and **you would ask us for significant support...**”
 - “How **are we going to store and back up all those disc images** you think you need?”



Jimmy Fallon as Nick Burns, “Your Company’s Computer Guy,” in a recurring “Saturday Night Live” sketch.

NBC image. URL: <http://blogs.wsj.com/digits/2009/07/31/are-you-celebrating-sysadmin-day/>, accessed June 23, 2014

Note: Our IT team is great, but this alarmed them.



Wrap-up / Future?

- We have moved from mission to policy to workflows and tools (and glory)
- BitCurator is a key tool in our program
- We have implemented it in baby steps – starting with the live disc environment, getting comfortable there, refining our workflow, and making the next steps once we more clearly see what they are.



Resources

- Olsen, P. (2012). “Digital Curation Workstation.”
<http://mith.umd.edu/digital-curation-workstation/>
- BitCurator Wiki. (2014).
http://wiki.bitcurator.net/index.php?title=Main_Page
- Digital Preservation Support at Fondren Library. (2014).
<https://digitalriceprojects.pbworks.com/w/page/44763477/Digital%20Preservation%20Support%20at%20Fondren%20Library>



Thanks!

Contact Us!

Rebecca Russell
Amanda Focke
Dara Flinn
Norie Guthrie

All reachable via woodson@rice.edu

Woodson Research Center

