Command-line Tools for Digital Preservation

Mark Phillips
TSLAC Austin, TX
June 13, 2011
Command-line Tools for Digital Preservation Curation

Mark Phillips
TSLAC Austin, TX
June 13, 2011
Motivation
Most tools for digital preservation involve the command-line
Some have a graphical user interface (GUI)
But most are designed for use on the command-line
This is a good thing.
JHOVE2
file
md5sum
Imagemagick (mogrify, convert, identify)
Sometimes there are GUI interfaces available
Other times there just aren't any
Today I want to gently introduce you to the command-line and its use in digital curation
Demonstrate basic curation concepts
Start to build a foundation for future work
Allow you to see some of the “magic” that Unix and Linux users have had access to for years.
(yes there is a command prompt in Windows, but this is different, we will talk about why)
At the end of today, you won't...
Be an expert
Learn “Linux”
Be ready to be a system administrator
Those are different things and outside of the scope of this workshop.
We will work from a very beginner level
With some examples and tasks that might not seem relevant.
But trust me they are.
We are going to do some of the following.
Create a list of filenames
Sort lists
Count occurrences in a list of terms
Work with a variety of basic *nix tools
Tools like:
pwd, ls, cd, sort, uniq, find, file, grep, sed, tr, wc, less, head, tail
To name a few
Then we will work these into the digital curation context
We will create a digital preservation “package”
We will discuss and demonstrate file characterization with fits, identify and pdfinfo
We will create and verify an md5 hash for a file
We will do basic metadata quality control activities
All on the command-line
All of which can be used for working with digital content of all types.
Agenda

9:00-9:30      Introductions
9:30-10:00     Getting Started
10:00-10:45    Basic Concepts and Tools
10:45-11:00    Break
11:00-12:00    Building Blocks of Digital Preservation
12:00-1:00     Lunch
1:00-2:30      Examples and Exploring
2:45-3:45      Next Steps show and tell
3:45-4:00      Wrap up
Quick Introductions:

Name
Institution
Experience Level
What you want to get from the workshop
First example
Metadata Quality Control and Analysis
UNT Scholarly Works
Total number of authors
5 most frequent authors
how?
Questions?