BITCURATOR GUIDE
FOR EDUCATORS

BitCuratorEdu Project

AUTHORS
Hannah Wang
Jessica Farrell
Christopher A. Lee
Katherine Skinner

EDITORS
Rhiannon Bettivia
Paul Conway
Caitlin Perry
Alyce Scott
Kim Schroeder

CONTRIBUTORS
Devan Donaldson
Karen Gracy
Grant Hurley
Kam Woods
Jane Zhang

EDUCOPIA INSTITUTE
Community Cultivators

UNC
SCHOOL OF
INFORMATION AND
LIBRARY SCIENCE

INSTITUTE of
Museum and Library
SERVICES

Table of Contents

Introduction

In recent years, archivists, librarians, and other information professionals have borrowed and adapted concepts from digital forensics to provide authentic, reliable, and meaningful versions of digital content. The open-source BitCurator environment and its thriving user community reflect and support significant evolution in digital curation workflows to better meet the following objectives:

- capturing data in ways that do not inadvertently change it (e.g., creating disk images, capturing timestamps of files);
- reflecting provenance and original order of materials;
- documenting chain of custody (recording actions of people and software); and
- locating, reviewing, and potentially acting on (e.g., redacting, flagging, filtering) sensitive information

Professionals charged with the curation of born-digital materials must make numerous decisions about which tools and methods to adopt in particular contexts. Professional education should provide supportive learning environments for students to directly apply tools and methods to digital materials; problem solve and clearly express technical challenges; and consider which professional practices would be most appropriate in different social, organizational, and technical settings.

This *BitCurator Guide for Educators* is designed to provide instructors with advice, ideas, and materials to support their inclusion of digital forensics concepts, tools, and methods into courses related to the curation of born-digital materials. This *Guide* is intentionally geared to a broad spectrum of uses by readers with varying levels of digital forensics knowledge and technical acumen. It includes core context, as well as designs and lesson plan examples and assignments that we hope will appeal to and provide scaffolding for different types of teaching and learning environments.

In this *Guide*, you will find a range of materials that support "low-tech" conceptual/theoretical framing, including how to talk about the ethical questions and dilemmas raised by digital forensics. You will also find a section containing and explaining how to provide "higher-tech" hands-on exercise examples, even if you are nervous about your own technical skill level (e.g., using the BitCurator software to identify and locate personally identifiable information, or PII). We have also provided information chronicling how to structure and support student self-guided explorations.

We know that many faculty and instructors are intimidated by the challenge of coaching students through technical problems across many different devices and platforms. We address this through a section devoted to troubleshooting that chronicles the most common errors or

failures and advises how to gauge your time and facilities to ensure success. We also know that concerns about accessibility hinder some from adopting BitCurator and other tools for class-based assignments or lessons, and we provide guidance and examples of ways to address these concerns and provide appropriately equitable learning environments for students.

The *Guide* begins by describing the BitCurator environment, which has been adopted by archives, libraries, and museums across the globe. We work to root this *Guide* in this concrete, existing set of open-source tools, while simultaneously laying a strong theoretical foundation to help the lessons extend to future tools and environments.

What is Digital Forensics?

Digital forensics is the process of recovering, validating, interpreting, and reporting on digital content that is found on a broad range of containers, including floppy disks and zip drives, as well as hard drives and servers. The digital forensics industry emerged during the 1990s in response to computer crime. Today, digital forensics is primarily used in criminal investigation, corporate security, and civil law arenas. This work has been dramatized in books, television shows, podcasts, and movies, bringing it into the public imagination as an area of criminal and private investigative science.

Digital forensics tools and methods have also become an important aspect of curating born-digital collections. They provide ways for libraries, archives, museums, and other collecting institutions to acquire, move, manage, and recover digital objects safely and with strong validation markers that help to verify the authenticity of those objects and the chains of custody that have managed them over time.

Intention marks a key difference between digital forensics in industry and criminal investigative realms and the application of digital forensics tools and methods in collecting institutions. Security and criminal investigations most often gather and use information in very specific and limited ways; the digital content is valued primarily as evidence that can demonstrate and document actions (often criminal). By contrast, collecting institutions use these tools to acquire and prepare born-digital content for long-term preservation and to enable future public access.

What is the BitCurator Environment?

The open-source BitCurator environment is used by collecting institutions, including libraries, archives, and museums, to move born-digital materials that are stored on removable media (e.g. floppy disks, flash drives, CD-ROMs, hard drives) into more sustainable preservation environments.

The BitCurator project series (2011-2018), led by the UNC School of Information and Library Science (UNC-SILS) and funded by the Mellon Foundation, has developed, documented, and

disseminated the open-source BitCurator environment to help curators ensure the completeness, authenticity, and availability of born-digital content while also simplifying and enhancing access.

Building on open-source digital forensics tools and methods deployed in the field of criminal investigation, a team from UNC-SILS and the Maryland Institute for Technology in the Humanities (MITH) designed the BitCurator environment to address two unique, foundational needs for collecting institutions: 1) incorporating these tools and methods into archival and library collection management workflows and 2) using them to provision public access to the data.

The BitCurator environment is built on a stack of free and open-source tools and associated software components, modified and packaged for increased accessibility and functionality for collecting institutions. The BitCurator software is freely distributed under an open source license. It can be installed as a Linux environment; run as a virtual machine on top of most contemporary operating systems; or run as individual software tools, packages, support scripts, and documentation. Tools in the BitCurator environment – both those produced by the project team and those from third-party developers – help advance core digital curation activities, including (but not limited to):

- Reducing the risk of inadvertent changes to content through software-based write blocking
- Creating authentic copies of content through disk imaging and cryptographic hashing
- Mounting forensically packaged disk images to view & export contents
- Reflecting original order of materials through capture of file system metadata
- Establishing trustworthy chains of custody through documentation of curatorial actions (log files, PREMIS records)
- Generating reports that characterize the contents of disks and directories
- Identifying and documenting duplicate files
- Discovering and exposing associated contextual information
- Finding and identifying sensitive information for further review, redaction, or removal, including PII
- Exporting contents of disks and directories for inclusion in Archival Information Packages and Dissemination Information Packages
- Building access environments that allow users to access data from disks using a web browser or emulation platform

Because institutions and individuals are not required to register to download and use the software, we can only estimate the size of the current BitCurator user community. In 2022 (at the time of writing), there are 47 organizations, composed of 367 practitioners, that are members of the BitCurator Consortium (BCC). The BCC has built a community of organizations that support practitioners responsible for the curation of born-digital materials. We know this is

only a small percentage of the institutions using the software. There are more than 400 individuals in the BitCurator Users group; the BitCurator Twitter account has almost 1,000 followers; and the annual BitCurator Users Forum drew more than 300 registrants in 2021.

Why Teach with the BitCurator Environment?

Born-digital materials have become a major focus and responsibility of collecting institutions of all kinds. Libraries, archives, museums, and similar organizations now regularly engage with locally created born-digital content (including their own policies and records) and with increasing numbers of externally created born-digital acquisitions (e.g., an author's digital manuscripts and email, or a composer's digital draft of a score) that are transferred to the collecting institution through a variety of methods, including removable media. Concerted effort and specialized techniques are needed to assure data integrity, preserve technical metadata like timestamps, and identify personally identifying or sensitive information. If these measures are not taken, institutions risk harming the materials, losing context, and exposing the private information of individuals reflected in their collections.

As more collecting institutions come into contact with born-digital materials, more practitioners need the tools and knowledge to approach the inherently fragile process of born-digital acquisition, even if born-digital archiving is not their primary role. Students in library, archives, and museum studies in particular need to understand born-digital materials both on a theoretical and practical level. This knowledge will not only be useful for practitioners who are regularly acquiring content from removable media, but also for future managers and administrators who must understand the level of resourcing and support necessary to do this work.

Education helps to advance new professional practices like the incorporation of digital forensics tools and methods into digital curation workflows; it also helps to prepare students to meet emerging needs that curating institutions are likely to fill with new positions. However, training opportunities in library and information science (LIS), archives, and museum studies curricula have lagged significantly behind the widespread adoption of these digital forensics tools and methods in curating institutions over the last decade.

Seeking to address this gap, the BitCuratorEdu project (2018-2022), has worked to identify and address factors that have stymied the inclusion of digital forensics concepts, tools, and methods in LIS graduate school and continuing professional development environments. Led by UNC-SILS, the Educopia Institute, the BitCurator Consortium, and the Council of State Archivists (COSA) and funded by the Institute of Museum and Library Services (IMLS), this project has engaged a broad array of academic institutions, professional associations, faculty, and students to understand what factors influence adoption of digital forensics concepts, tools, and methods in learning environments, and to build a resource and support infrastructure to encourage broader implementation, including a series of modular learning objects.

BitCuratorEdu Curriculum Philosophy Statement

This statement was created by the BitCuratorEdu project and partners to direct our research and development of learning objects:

> *Digital curation is a broad, interdisciplinary field of scholarship and practice, which includes, among other areas, the application of digital forensics tools and methodologies when acquiring and processing born-digital materials. Teaching with the BitCurator Environment, a suite of open-source digital forensics and data analysis tools, not only builds capacity within the cultural heritage sector for working with digital materials, but is also an effective way to introduce students to informational structures and levels of representation of digital information, whether or not they intend to pursue a digital-focused career. The BitCuratorEdu curriculum aims to provide access to hands-on learning experiences and to offer an approach to these experiences that is grounded in digital curation scholarship. The goal of the curriculum is to produce competencies in when, why, and how to use digital forensics tools in the processing of born-digital materials to assure continued access and preservation.*

BitCuratorEdu Learning Objectives

BitCuratorEdu learning objects provide support for instructors with the following learning objectives for their students:

Objective #1

Understand how to apply digital forensics tools and methods to born-digital materials.

Learning outcomes:
- Determine the technical requirements and dependencies for installing open-source tools to process born-digital materials
- Practice using tools in the BitCurator Environment
- Design a born-digital archiving workflow
- Diagnose and troubleshoot issues that arise in born-digital archiving workflows

Objective #2

Determine when to use digital forensics tools and methods within born-digital processing workflows.

Learning outcomes:
- Identify the appropriate tools to: safely acquire born-digital materials from storage media and other modes of transfer; assist in the appraisal of born-digital materials; scan for

sensitive information in born-digital materials; and package born-digital materials for preservation and access
- Test and evaluate tools for use in born-digital archiving workflows

Objective #3

Comprehend why digital forensics tools and methods are used and the larger context around their implementation.

Learning outcomes:
- Articulate the relationship between actions taken during acquisition and processing of born-digital materials and their long-term preservation and access
- Translate the major archival functions in traditional processing to their equivalent tools and methods in born-digital processing
- Differentiate between born-digital curation and digitization/digital projects
- Analyze the ethical considerations for tools and methods used in born-digital curation
- Advocate for the use of community-built and -supported open-source tools in digital curation

How to Use This Guide

In this *Guide*, we outline three broad modes of teaching with the BitCurator environment, which progress from less to more "hands-on":

1. Conceptual Framing Lessons and Activities
2. Self-Directed Student Exploration with the BitCurator Environment
3. Hands-on Exercises and Assignments

Conceptual Framing Lessons and Activities provide conceptual and theoretical background of digital forensics, BitCurator, and associated digital curation workflows, without necessarily requiring hands-on interaction with the software itself. Self-Directed Student Exploration with BitCurator covers lessons and activities that require students to explore the BitCurator environment through documentation, videos, or hands-on engagement, often accompanied by a chance for critical reflection. Hands-on Exercises and Assignments allow students to directly engage in digital curation tasks in an environment that is more forgiving than working with real primary sources in a "production" environment.

By "exercise," we mean an activity initiated and completed within a single learning unit. For a synchronous (in-person or remote) learning experience, this often means 1-2 class sessions. By "assignment," we mean an activity that spans multiple learning units, e.g., a semester-long group project. Both exercises and assignments can require students to submit deliverables to the instructor, but we will presume that only exercises can be completed within the context of a specific learning unit.

BitCuratorEdu

For each of these three pedagogical approaches, we offer a brief overview of why and how the approach might be used; outline prerequisites for the instructor and students and technical requirements; and list practices and examples created by the BitCuratorEdu project and our partners to achieve certain learning objectives and outcomes.

Learning objects created by the BitCuratorEdu project are intentionally modular, reflecting a stated desire from many instructors to have access to "raw materials" for teaching, rather than extensive and fully packaged learning modules. Most resources from the BitCuratorEdu project are intentionally left with basic formatting and without project branding. We encourage educators, practitioners, and students to adapt these materials as much as needed and share them widely. As the technology changes and the field adapts to teaching new tools and concepts, we hope that these learning objects may have a longer life by not being tied to a particular moment in time or version of the software.

The last three sections of this *Guide* are for general reference in teaching, and they cover Troubleshooting, Using the BitCurator Environment with Assistive Technology, and Resources for educators and students created by the project and the BitCurator Consortium.

About the BitCuratorEdu Project

The BitCuratorEdu project was an effort (2018-2022) funded by the Institute of Museum and Library Services (IMLS) to study and advance the adoption of digital forensics concepts, tools and methods in libraries and archives through professional education. This project was a partnership between Educopia Institute and the School of Information and Library Science at the University of North Carolina at Chapel Hill, along with the Council of State Archivists (CoSA) and several Masters-level programs in library and information science, including:

- The Catholic University of America
- Indiana University
- New York University
- San José State University
- Simmons University
- University of Maryland
- University of Michigan
- University of Texas at Austin
- Wayne State University

Conceptual Framing Lessons and Activities

Overview

Conceptual framing lessons provide conceptual and theoretical background for digital forensics, BitCurator, and associated digital curation workflows. These lessons do not necessarily require downloading, installing, or using the BitCurator environment. Examples of these types of lessons and concepts include:

- An introduction to the history of digital forensics in law enforcement and its use in the cultural heritage sector
- A discussion and mapping of archival principles of provenance, original order, and chain of custody to digital forensics methods
- A discussion of how the use of digital forensics tools and methods assist in the end goal of preserving and providing access to digital collections
- An overview of the challenges posed by hardware, file systems, operating systems
- A survey of digital curation workflows and how digital forensics tools and methods fit into those workflows
- An introduction to ethical issues associated with born-digital collections such as privacy, security, and donor relations

Prerequisites

This type of lesson and class activity can act as a helpful primer for students who might have a background in the theory and concepts of LIS, but who may not have the technical expertise and experience (and access to equipment) to jump directly into hands-on exercises.

Instructor and Student Knowledge and Preparation

Students should have a solid grounding through coursework or practical experience in archival principles and concepts such as:

- Appraisal, to understand the process of evaluating the significance of digital records and selecting for long-term retention and preservation
- Accessioning, to understand the process by which digital materials become part of an institution's collections
- Chain of custody, to understand the importance of documentation in demonstrating the authenticity and trustworthiness of digital records
- Authenticity, to understand how trustworthiness is established and demonstrated in digital records creation, transfer, preservation, and access

- Original order, to understand the importance of maintaining or reflecting order for comprehending digital records
- Provenance, to understand the importance of retaining information about the context of creation for digital records

It may also be helpful for students to have additional coursework or the equivalent experience that covers information organization, information systems, and electronic records management.

Instructors should also have solid theoretical and practical grounding in archival principles, information organization, information systems, and electronic records management. Even though instructors may not necessarily demonstrate these tools or create hands-on activities, they should have familiarity and experience with one or more digital curation tools to ground the theory in a practical context of the skills and labor required to do digital curation work. The key to successful teaching and learning about these tools is a willingness to experiment, problem solve, and work with user communities. Consult resources from the BitCurator Consortium for workflow examples, slide decks, and screencasts (see [Resources](#)).

Equipment and Technology

There are no specific equipment or technology requirements for these types of lessons for either instructors or students. These lessons are well-suited to programs and learning environments of any technical level, including where there are little to no provisions for student technology. These types of lessons also lend themselves most easily to adoption in either asynchronous or synchronous online teaching. For workflow modeling activities, it is useful to have sticky notes, index cards, or an interactive digital equivalent.

Practices and Examples

Articulating the relationship between archival principles and digital forensics tools and methods

One of the most compelling cases for using digital forensics tools and methods for cultural heritage materials is the relationship between these methods and the archival principles of provenance, original order, chain of custody, and privacy. Discussing these relationships can be a good place to start when teaching digital curation to archival practitioners or students who have taken archival coursework.

Lesson: [Introduction to BitCurator](#) (Cal Lee)
This slide deck includes a section mapping major archival principles to digital forensics methods. In introducing this concept, you might choose to talk about each of these principles and ask the questions:

- What is the importance of this principle in archival practice?
- How is it enacted for paper-based records?
- How might it be enacted for born-digital records?
- Are there any barriers to enacting this principle with born-digital records?

Lessons and Exercises: [BitCurator Screencasts and Discussion Questions](#) (Cal Lee, Hannah Wang)

These discussion questions can be used to encourage student engagement with the BitCurator software. While the BitCurator technology will evolve and the screencasts themselves must be updated to reflect that evolution, the discussion questions mostly pertain to the topic of the video more generally and can be adapted as needed.

Topics and tools covered include:

- [Introduction to the BitCurator Software Environment](#)
- [Installing BitCurator: Getting Started with the Virtual Machine](#)
- [Setting Up Shared Folders](#)
- [Creating a Disk Image Using Guymager](#)
- [Safely Mounting Devices](#)
- [Using fiwalk to Generate Filesystem Metadata](#)
- [Using the BitCurator Reporting Tool](#)

These learning objects present opportunities to explore the application of a digital forensics tool for specific purposes, without necessarily downloading or using the tool itself. Using the Guymager screencast, for example, you could make this connection explicit by focusing on the discussion questions:

- At what point(s) in a digital curation workflow might you expect to use Guymager?
- Why might you want to create a disk image rather than only copying the files?

These questions can prompt students to discuss accessioning best practices, chain of custody, provenance, and original order.

Understanding and creating digital curation workflows

We have assembled a number of conceptual framing activities for teaching students how to compare, analyze, and draft digital curation workflows.

Lesson and exercise: [Workflow Modeling](#) (Cal Lee)

This example lesson and exercise begins by introducing workflow models and process modeling, discussing common tasks that are often included in digital curation workflows, then presenting a scenario for students to create a hypothetical workflow model using sticky notes or

a digital equivalent. This is followed by a post-mortem discussion reflecting on the process of creating and modeling workflows.

Exercise: Cross-Institutional Workflow Analysis (Rhiannon Bettivia, Hannah Wang)
Students are asked to look at two examples of real-world digital curation workflow diagrams, answer a series of questions comparing and contrasting the workflows, and discuss their comparative efficacy as artifacts and models.

Exercise: OSSArcFlow Guide to OAIS Mapping (Hannah Wang)
This exercise asks students to walk through the steps outlined in the OSSArcFlow Guide to Documenting Born-Digital Archival Workflows. The guide assists collecting institutions to begin documenting their born-digital workflows and map them to functional entities in the Reference Model for an Open Archival Information System (OAIS). The discussion questions encourage students to think about the differences between an OAIS model and an OSSArcFlow model in terms of their functions and audiences.

Self-Directed Student Exploration with the BitCurator Environment

Overview

This *Guide* defines self-directed student explorations as lessons and activities that require students to explore the BitCurator environment through documentation, videos, or hands-on engagement, often accompanied by a chance for critical reflection. These types of lessons offer students the opportunity to learn about the technology and reflect on why various digital forensics interventions might be appropriate. While this *Guide* focuses on lessons that make use of tools in the BitCurator environment, many of them could be applied as frameworks for building out lessons around other digital curation tools.

Given the more "hands-off" approach of these types of lessons and activities, self-directed student explorations are accessible to a wide group of instructors, students, and curricula. But the wide range of options for activities may require the instructor to first assess students' access to technology and capacity in order to effectively select and scope lessons (see [Digital Curation Pre-Class Survey](#) for an example assessment tool).

Self-directed student explorations can take many shapes, from hands-on testing of the BitCurator software to reviewing and reflecting on BitCurator technical documentation. They can be designed for students of varying skills and access to technology, including a stable internet connection. With thoughtful planning, instructors have the opportunity to provide different options for engagement based on individual students' access to or readiness for installing and using the BitCurator software. These lessons also may be expanded or contracted to accommodate different amounts of instruction time, from a single unit or part of a unit in the semester to incorporation into a longer series of activities.

For courses that incorporate the BitCurator environment into only one unit, we recommend lessons that do not require students to install the BitCurator environment on their own computers. Some instructors offer the opportunity to download and experience the BitCurator environment as an optional activity for the particularly curious and technically capable. For the sake of parity, we recommend that instructors refrain from offering extra credit for activities like this unless they are sure that all students have access to the technology required to succeed.

The distinguishing feature of this type of lesson is that it prioritizes experiential learning and reflection. The primary learning outcome is to gain experience understanding the practical uses of the tool. Practitioners develop problem-solving and reflective skills through these types of activities. It is important to give students the opportunity to explore and to safely fail or succeed

in educational settings, in order to build the confidence to carry out similar tasks in a "real-world" environment.

Prerequisites

Instructor and Student Knowledge and Preparation

Students need to have some conceptual understanding of digital curation before diving into these types of activities to understand the purpose of their exploration and to examine it critically. Additionally, a basic understanding of the command line is useful for students and instructors alike.

Instructors need to foster critical engagement with the activity or tool. While teaching with a static resource like a pre-recorded video or technical manual may be a less complicated method of delivering digital forensics education than hands-on activities, the technical understanding that underpins digital curation tasks is necessary to support effective student self-guided exploration. The term "self-guided" is meant to signal that in teaching this type of lesson, troubleshooting technology required to engage in the lesson is mostly out of scope. But it does not mean that instructors will be completely hands-off in the learning process. These activities do require some guidance from the instructor, and critical engagement with the activity is a key to ensuring successful learning outcomes for students.

Our educational partners on this project generally agreed that a conceptual foundation is necessary to teach these types of lessons, because they are the lowest-barrier building blocks to a bridge between theory and practice. A conceptual foundation is important for engaging critically with students and their reflections on how and why the technology works the way it does.

Equipment and Technology

Equipment and technology needs will vary widely depending on the lesson. These activities can be conducted in a lab or students may work independently on their own devices. For self-guided hands-on explorations, instructors may want to communicate to students that troubleshooting is part of the learning experience, and troubleshooting support will be provided by helping the students find the right documentation to support their exploration. Instructors can also encourage peer technical support as part of class participation. See [Troubleshooting](#) and [Resources](#) for more information on troubleshooting resources available for instructors and students alike.

Because these lessons can be more flexibly interpreted by the student, they are particularly useful in remote instruction where the instructor has little control over students' computing environments. To encourage a level playing field, instructors should survey their students at the

beginning of the course to determine each student's comfort with and access to the associated technologies, including questions about the machine they will be using for the course (see Digital Curation Pre-Class Survey for an example assessment tool). Some programs have baseline requirements for student machines, and it is helpful for instructors to be aware of how those specifications align with the system requirements for software used in class. Some programs have opportunities to borrow "loaner" laptops for students who need them, though not all of these laptops will meet the hardware requirements for installing the BitCurator environment as a virtual machine or allow students the requisite permissions to install software. It is a good idea to check in first with your program about provisioning equipment as well as your students. Doing this preparation will reduce troubleshooting time with students later, alert relevant IT support staff of the activities, and help instructors choose appropriate activities from the outset, and avoid the need to develop new options during the course. .

Practices and Examples

Evaluating tools for use in born-digital archiving workflows

Assignment: Digital Curation Technologies (Devan Donaldson)
This assignment engages students in digital curation technology through several self-guided explorations culminating in a presentation. Students have the option of choosing the BitCurator environment as a research subject. Students answer a series of questions that mirror questions a practitioner responsible for making a recommendation on adopting a new technology, repository, or service must ask themselves.

Articulating the relationship between archival principles and digital forensics tools and methods

Lessons and Exercises: BitCurator Screencasts and Discussion Questions (Cal Lee, Hannah Wang)
These discussion questions encourage student engagement with the BitCurator software. Although they were designed to be used in concert with the BitCurator screencasts, they would also pair well with a self-guided or instructor-guided exploration of the BitCurator environment.

While the BitCurator technology will evolve and the screencasts themselves must be updated to reflect that evolution, the discussion questions mostly pertain to the topic of the video more generally and can be adapted as needed.

Topics and tools covered include:

- Introduction to the BitCurator Software Environment
- Installing BitCurator: Getting Started with the Virtual Machine
- Setting Up Shared Folders

- [Creating a Disk Image Using Guymager](#)
- [Safely Mounting Devices](#)
- [Using fiwalk to Generate Filesystem Metadata](#)
- [Using the BitCurator Reporting Tool](#)

Assignment: [BitCurator and the DCC Lifecycle Model](#) (Devan Donaldson)
This assignment uses the Digital Curation Centre's lifecycle model as a conceptual foundation and asks students to identify where they see opportunities to take the Sequential Actions outlined in the model using the BitCurator environment.

Articulate the relationship between actions taken during acquisition and processing of born-digital materials and their long-term preservation and access

Assignment: [Digital Forensics Paper and Project](#) (Jane Zhang)
This is a paper and project assignment that asks students to review literature on digital forensics' applications in digital archives, and gain hands-on experience using digital forensics tools to acquire and manage born-digital archival materials.

Hands-on Exercises and Assignments

Overview

Hands-on exercises and hands-on assignments allow students to engage directly in digital curation tasks in an environment that is more forgiving than working with real primary sources in a "production" environment. These activities can be in-person or online; they can be implemented synchronously, asynchronously, or both. In contrast to self-directed student exploration, exercises and assignments involve a higher degree of instructor guidance and expectation of outcomes.

Students always need to use some specific software to perform actions in a digital environment, but digital curation educational objectives are rarely tool-specific. Hands-on exercises and hands-on assignments can advance at least three different categories of digital curation learning objectives:

1. understanding digital curation functions and potential role of supporting technologies
2. understanding underlying technology, and
3. technical proficiency

Understanding Digital Curation Functions and the Potential Roles of Supporting Technologies

Instructors often incorporate the BitCurator environment into university courses that are framed around broader themes such as archives and records management (ARM) or digital curation. In other cases, instructors incorporate the BitCurator environment into classes about specific archival functions (e.g., records management, description, preservation, appraisal). In both categories of courses, a central priority can be helping students to make connections between any hands-on tasks and the digital curation functions supported by those tasks.

For example, a class module could highlight the role, value and limitations of incorporating cryptographic hashes (checksums) into digital curation workflows. An associated hands-on exercise might ask students to create md5 hashes of a set of files, make changes to the files, and then attempt to verify the hashes (confirming that a bitstream change results in a different hash). In this exercise, students have an opportunity to directly experience important relationships between bitstreams and hashes, and hopefully reflect on what these relationships mean for digital curation. The learning objectives can be the same, regardless of whether the students perform the tasks with the md5deep command at the Linux command line, use the graphic user interface (GUI) of GtkHash in Linux, run QuickHash on a Windows machine, or run the files through an online hashing service.

Understanding Underlying Technology

A great deal of digital curation work happens at the boundaries of systems and language communities. This requires considerable translation work between those systems and communities. Between systems, this requires a fundamental understanding of how technological platforms operate, their associated dependencies, and strategies for promoting interoperability across space, time, and context.

Successful digital curation requires commitments and coordinated activities of many different stakeholders. Digital curation professionals can benefit significantly from learning the terminology of those stakeholders. One set of important stakeholders are associated with the technical implementation of digital curation decisions. When communicating with an IT manager about how to configure and support a new digital repository platform, for example, one who does not understand any of the language associated with IT administration can be at a serious disadvantage.

Hands-on exercises and projects can highlight, demonstrate, and reinforce concepts and terminology related to underlying technologies. For example, performing tasks in the BitCurator environment allows students to directly experience various characteristics of operating systems, file systems, and virtual machines. Instructors can ask students to express in their own words what is happening technically in the environment, as a way to demonstrate and refine their learning.

Technical Proficiency

Much as digital curation conversations build upon specific terms and concepts, digital curation tasks also build upon specific skills and capabilities. Hands-on exercises and assignments can help students to develop, hone, and reinforce those skills and capabilities. Within professional education, it is desirable to provide lessons that students can generalize to a variety of settings. An instructor of a unit on extracting metadata from disk images, for example, may ask students to run [Brunnhilde](#) at the Linux command line within the BitCurator environment. While the tasks involve typing specific commands, the exercise can provide an opportunity to demonstrate more general command line concepts, including arguments, pipes, switches, wildcards, and directory paths.

Prerequisites

Instructor and Student Knowledge and Preparation

The prerequisite knowledge required of students and instructors can vary dramatically based on the nature of the exercises/assignments and learning objectives.

Students learning how tools in the BitCurator environment can potentially advance professional values, principles, and objectives must first understand those values, principles, and objectives. Several instructors who have introduced the BitCurator environment into their courses, for example, have done so within the context of an ARM curriculum that requires students to take an "Introduction to ARM" class early in their experience. If the instructor of a course knows the students have already taken such a foundational class, the exercises and assignments can focus on specific professional, ethical, and technical implications rather than needing to explain the underlying values, principles, and objectives.

Both **students and instructors** may need to already be proficient at the command line when an advanced course is offered with in-depth exercises that use a set of command-line tools. However, it is important to distinguish between knowledge students must have before engaging in an exercise/assignment and knowledge they must have by the time they complete the exercise/assignment. For example, if one of the primary objectives of an exercise is to teach students how to use the command line, the instructor will need to have command-line proficiency, but the students will not. In this case, students may learn command line skills through the in-depth forensics exercises so long as an instructor can point students to basic primers on the Linux or Windows command line and bring their own knowledge to bear, so students can adapt their skills along the way.

Equipment and Technology

The BitCurator environment is an entire operating system (OS) based on Ubuntu Linux. One can set up a designated computer that runs the BitCurator environment as its native operating system — either as its only OS or by booting into a dedicated BitCurator partition. Alternatively, one can run the BitCurator environment as a virtual machine (VM) on top of another OS (usually Windows or MacOS). The VM is referred to as the "guest" OS and the user's native operating system is referred to as the "host." For instructions and system requirements, see the [QuickStart Guide](#).

The VM option is the most widely adopted by instructors and students, because it requires minimal changes to the user's host OS. Installation requires downloading VirtualBox and extensions; running the VirtualBox installation wizard; downloading the BitCurator VM package (.tar.gz file); extracting the contents of the VM package to a local directory; and adding the VM to VirtualBox. See [Troubleshooting](#) for information on troubleshooting issues that may arise while installing the VM.

Practices and Examples

Practice using tools in the BitCurator Environment

Assignment: [Installation of BitCurator VM and Introduction to Disk Imaging](#) (Karen Gracy)

BitCuratorEdu

This assignment introduces students to the BitCurator environment by asking them to download and install the virtual machine, create a disk image using Guymager, and analyze the contents of the disk image .INFO file. It is a hands-on activity that familiarizes students with resources like the BitCurator Quick Start Guide.

Lesson and Activity: [Curating Potentially Sensitive Information in Digital Collections](#) (Cal Lee, Kam Woods, Simson Garfinkel)
In this lesson, students learn why and how to identify personally identifiable information (PII) and other sensitive information within digital collections, as well as methods for redacting the information. The unit includes two hands-on exercises: 1) identifying potentially sensitive information within a disk image and 2) redacting information that meets specified criteria within a PDF file**.**

Activities: [BitCurator Hands-On Exercises and Sample Data](#) (Cal Lee, Kam Woods)
This series of hands-on exercises, adapted from the SAA "Advanced Digital Forensics" class, can be used by instructors to introduce a number of tools and concepts in the BitCurator environment, including:

- [Basic Linux Commands](#)
- [Disk Images](#)
- [Exiftool](#)
- [Extracting Data from Office Documents](#)
- [File System Attributes](#)
- [National Software Reference Library (NSRL)](#)
- [PRONOM, Siegfried, and Brunnhilde](#)
- [Windows Artifacts](#)

The [sample data](#) needed to complete these exercises is available through GitHub, and linked from each exercise. Instructors can either distribute this sample data via flash drive or ask students to download the data directly from GitHub.

Troubleshooting

Troubleshooting can be essential to hands-on digital curation lessons or demonstrations. Instructors and students have faced heightened troubleshooting responsibilities since the onset of the COVID-19 pandemic in early 2020, as many classes transitioned from in-person to online environments. This has required development of new course materials, as well as attempts to adapt existing materials to the constraints of specific learning management systems (e.g., Canvas, Sakai, Moodle).

In addition to enabling students to engage in other educational tasks, troubleshooting can itself be a valuable learning experience for instructors and students. Troubleshooting can yield many lessons about problem-solving, as well as building confidence to take on other novel challenges. Like other professional activities, one can get better at troubleshooting through practice.

Prepare to fail

Digital curation practices, technologies and types of digital materials all evolve over time. Digital curation education should prepare students to participate in that evolution. This includes providing them opportunities to experience failure and practice problem-solving to address technical challenges.

Consider sharing useful documentation in a lesson early within a course or a one-pager of troubleshooting tips, such as:

- Common low-effort troubleshooting methods like typing questions into a search engine
- How to read, interpret, and share error messages
- Where to find the BitCurator FAQ and other BitCurator documentation
- Where to find Linux and Ubuntu documentation

Make sure that students understand that no single person knows the answers to all the technological problems they will encounter. The best way they can prepare themselves to efficiently troubleshoot software throughout their careers is to build knowledge upstream and to cultivate a strong understanding of where resources, including community voices, reside and how they can help.

Build knowledge upstream

By familiarizing yourself with both the OS and (if applicable) virtualization software, you will be in a better position to assess the situation when there is an error. In the case of the BitCurator environment, you may want to offer resources to students about how Linux works, especially in comparison to MacOS and Windows. You will also want to reassure them that the best way to

become familiar with an OS is to use it. The problem solving involved in downloading and setting up the BitCurator environment is representative of many similar tasks digital curation professionals encounter regularly in their work.

Cultivate a community

You feel like you know nothing about the OS and the "power users" all seem to have spent many years internalizing a vast array of experiences and lessons that are completely foreign to you. This is a great time to ask for help.

Cultivating a community of support is vital in an ever-changing profession. Students should be ready to build a community of support for themselves as they enter the profession. Students and working professionals can build their own personal networks of individuals they can turn to for a fresh eye on their technical questions, or join communities like the BitCurator Consortium where they can spread this labor across many supporters. Digital curation education should prime this process by prompting students to consider where they will go for help when they run out of ideas for next steps in troubleshooting.

Consider sharing communities to turn to when you assign lessons requiring the use of BitCurator, such as:

- Classroom online discussion forums, if applicable
- Asking colleagues through social media
- Reaching out through professional discussion lists (e.g., BitCurator Users Google Group for BitCurator questions, or Digital Curation Google Group for other archival software troubleshooting questions)
- StackExchange

Reaching out for technical assistance not only allows students and professionals to build valuable social connections, but it also gives them practice in asking constructive technical questions. Whether asking a work colleague, fellow student or instructor for help, a statement such as "It's not working for me," is not particularly helpful. One must know (1) basic characteristics of the environment (e.g., RAM, OS, CPU), (2) what specifically someone was doing when encountering the issue (commands, sources of input), (3) what the specific result was (e.g., unexpected output, error message), and ideally, (4) what specific steps they have tried to address the issue. While the final item is not strictly necessary for troubleshooting in general, it's vital for constructive learning. Rather than simply asking for answers, students should learn to convey what answers they have come to themselves, even if they feel wrong or incomplete.

Prevent burnout: Reward troubleshooting

Note that troubleshooting does take significant time. Students should receive recognition for troubleshooting time and not just the successful "completion" of an assignment. Ensure that students are rewarded for time spent troubleshooting and tinkering. You may offer credit for students replying to each others' questions on a message board, or the opportunity to share all of the things they tried instead of the successful outcome desired.

In general, success should not be measured solely by the outcome but also by the learning objectives addressed through the process. Once students enter the field in a position using these skills, they may have a variety of tools to choose from or be restricted from using open-source software altogether. In addition to learning about the functionality of the software, students can also learn strategies for successfully setting up and navigating software in an unfamiliar environment.

Troubleshooting VM Installation Issues

The following table summarizes the most common issues instructors and students encounter when installing the BitCurator VM, along with responses to those issues. See further discussion of these issues in the QuickStart Guide and the BCC's Troubleshooting and FAQ page. Some instructors have repurposed elements of this guidance to create documents specific for their classes or programs. For example, Justine Cucchi and Katie Tapia-Lynch at Wayne State University developed a succinct "Installing and Running BitCurator" document for use in their School of Information Sciences course.

Issue	Response
Insufficient random-access memory (RAM) on host machine	Use a different host machine
Insufficient storage available on host machine	Free up space on the host drive (if practical) or use a different host machine
Central processing unit (CPU) is not powerful enough to run the software	Use a different host machine
Host machine has enough RAM, storage, and/or CPU resources, but VirtualBox does not know this	Within VirtualBox (ensuring first the VM is shut down), change the VM settings to provide sufficient RAM, storage and CPU from the host to the guest
User doesn't have the right permissions to install or execute VirtualBox on the host	In some cases, it is possible to temporarily escalate one's privileges (e.g., "Run as Administrator" in Windows), but this often is

machine (most common with work or school machines that have been locked down for security/stability)	not possible. When aware of this issue in advance of a class activity, one can ask an administrator for the credentials to make the changes. Otherwise, you need to switch to a different host machine.
VM package file is corrupted (did not fully download)	Rerun the download. To verify that it was successful, generate a checksum of the downloaded file and verify that it matches the checksum from the GitHub download page for the VM.
The VM is missing functionality including support for USB 3 devices	Be sure that you installed not just VirtualBox but also the extension pack
After downloading the VM, the user is not able to see or add the VM within VirtualBox because the contents of the .tar.gz file are not extracted	Use a tool built into the host OS or a third-party tool (e.g., 7-Zip) to uncompress and then untar the contents of the file into a directory of files that VirtualBox can read
Virtualization between the guest and host isn't working properly (particularly on laptop computers)	Many laptops come with required virtualization features turned off by default. You will need to enter the BIOS and enable this feature (called V-tx on Intel machines).

Troubleshooting Tasks within the BitCurator Environment

Once students have downloaded, installed and successfully run the BitCurator environment, they can encounter issues within the Linux environment. The following table summarizes several common issues along with responses.

Issue	Response
Student gets confused by whether tasks are being performed in the host or guest OS (e.g., trying to perform a task within the VM on a set of files that are stored only on the host machine)	Remind students to consider whether they are performing a task in the right place and point out features of the VirtualBox interface that allow you to distinguish between the host and guest. Students can also benefit from setting up a shared folder between the host and guest.
User is prompted for a password to complete a task	Some tasks require the user to switch to a higher administrative status, so Linux prompts the user for a password. The default password and login are both bcadmin.

Students unsure of the meaning of terms used within the environment	The proper response depends on which tools are involved. For Linux interface labels, one can consult the Ubuntu documentation. For the core tools developed for the BitCurator environment (e.g., BitCurator Reporting Tool, BitCurator disk image access Tool), see the QuickStart Guide and BCC documentation. For any other tools running in the environment, see the documentation for those specific tools (e.g., Bulk Extractor user manual to identify specific scanners, in addition to information on the BCC wiki — https://confluence.educopia.org/display/BC/Understanding+Bulk+Extractor+Scanners)
Students unable to perform command-line tasks	Explain the general structure of command-line operations and how to fix errors.

The final item in the table warrants further discussion. Some students hit an impasse at the command line, because they do not feel comfortable even getting started. In these cases, the best remedies are quick tutorials (in class or provided as external resources) about how command-line operations work and what challenges one is likely to encounter.

Command-line tasks generally involve the name of an operation, any switches applied to the operation (configuration options), any special arguments (data values), sources of input and any output paths (e.g., to the screen, to a file, or hand-off to another operation). Once one learns the basic structure, one can begin to interpret existing commands and construct new ones. Command-line programs usually provide little or no feedback to let the user know when a task completed successfully, so students need to know that generally "no news is good news," i.e., when there is no error message, this usually means the task was successful. Because human typing is error-prone, it is also helpful for students to learn about common syntax errors, as well as tricks that reduce the amount of typing they need to do. These include use of the tab key to autocomplete words in a command and, especially, using the up and down keys to move through previously typed commands. It can be much easier to change just one part of a command and run it again than to completely reconstruct the command again by hand.

More resources for troubleshooting and tool-specific guides can be found in Resources.

Using the BitCurator Environment with Assistive Technology

Introduction

The BitCurator environment is an Ubuntu-derived Linux distribution. BitCurator can be installed and run as a virtual machine (VM) on an existing computer using Oracle VM VirtualBox. Since the VM is more likely to be used in an instruction environment (rather than a dedicated install), this section documents accessibility considerations and assistive technologies available when installing and running BitCurator as a VM.

VirtualBox Accessibility

The BitCurator environment is distributed as a VirtualBox VM. Oracle VM VirtualBox allows the user to run multiple operating systems on your host operating system (the OS of the physical computer where you have installed VirtualBox). The VirtualBox Manager is a graphical user interface that is described in the BitCurator Installation documentation on the BitCurator Consortium Wiki. Since it is an application run on your own desktop, the accessibility of the Oracle VM VirtualBox will depend to some extent on your host operating system and the assistive technologies that you are using. The VirtualBox User Manual provides somewhat limited documentation on accessibility features available in the GUI. An ongoing list of support tickets and feature requests related to accessibility is available on the VirtualBox website.

Oracle VM VirtualBox User Manual (stable)
VirtualBox support tickets and feature requests with the keyword "accessibility"

Ubuntu Accessibility

When run as a VM, the guest operating system (the OS that runs inside the VM) is Ubuntu. The Ubuntu desktop (GNOME) includes assistive technologies to support users with visual impairments and other accessibility needs. At the time of writing, the BitCurator environment runs on Ubuntu 18.04LTS — accessibility documentation for this version of Ubuntu is available on the Ubuntu documentation site.

Ubuntu 18.04 Universal Access Documentation
Ubuntu Accessibility Documentation (stable)

BitCuratorEdu

Resources

There is a wealth of resources available for educators and students from the BitCuratorEdu project and the BitCurator Consortium. Bear in mind that links may change over time after the publication of this document.

BitCuratorEdu Lesson Plans and Materials

BitCuratorEdu Project Page (BitCurator Consortium Website)
https://bitcuratorconsortium.org/bitcuratoredu/
All resources and presentations produced by the BitCuratorEdu project (IMLS, 2018-2022) can be found on this page. Although this *Guide* describes a number of learning objects and teaching materials gathered and created as a part of the project, many more are available to explore on this page.

BitCuratorEdu Project Page (Educopia Website)
https://educopia.org/bitcurator-edu/
Historical information about the BitCuratorEdu project outputs and personnel can be found on this page.

BitCuratorEdu Bibliography
https://www.zotero.org/groups/2454255/bitcuratoredu/library
The BitCuratorEdu project created this bibliography of readings in digital curation and digital forensics which might be used in courses on these subjects.

BitCurator Documentation

BitCurator Wiki
https://confluence.educopia.org/display/BC/BitCurator+Environment
The BitCurator Wiki, maintained by the BitCurator Consortium Documentation & Training Committee, is the technical documentation hub for the BitCurator environment. Here you will find detailed step-by-step guides for installation and using tools in the environment.

BitCurator on GitHub
https://github.com/BitCurator/bitcurator-distro
The latest releases of the BitCurator environment are available for download on GitHub. Be mindful of the technical requirements and type of distribution – in most cases, educators and students will want to download the VM release.

Quick Start Guide
https://github.com/BitCurator/bitcurator-distro/wiki/Releases#quickstart-guide

BitCuratorEdu

The Quick Start Guide provides a step-by-step walkthrough of the installation process with screenshots.

BitCurator Consortium YouTube
https://www.youtube.com/channel/UCtehN0wVGPbylXgZ0Do1n-g/featured
The BitCurator Consortium maintains a series of YouTube screencast tutorials demonstrating different tools and concepts in the BitCurator environment.

BitCurator Consortium Resources
https://bitcuratorconsortium.org/resources/
The BitCurator Consortium maintains a list of resources generated by and for the BitCurator user community. These resources include Teaching Materials, Workflows, Documentation, Case Studies, Instructional Videos, Presentations, and Event Recordings.

BitCurator Users Google Group
https://groups.google.com/g/bitcurator-users
The BitCurator Users Google Group is an open mailing list to post questions, discuss issues, and connect with other members of the BitCurator user community. It can be an especially useful resource for troubleshooting.

OSSArcFlow Workflows
https://educopia.org/ossarcflow-as-is-all-workflows/
The OSSArcFlow Project (IMLS, 2017-2020) documented the born-digital archiving workflows of 12 organizations, many of which used the BitCurator environment in their workflows. These offer practical real-world examples of how tools in the BitCurator environment are used at collecting institutions, and where pain points appear in connecting different parts of their workflows.

BitCurator Consortium Presentations
https://bitcuratorconsortium.org/resources/#presentations
This set of recordings and slide decks from presentations at the BitCurator Users Forum and other events are an invaluable resource for educators and students who are interested in learning about recent research and innovations in digital curation.

Sample Data

BitCurator Consortium Datasets Library
https://confluence.educopia.org/display/BC/Resources#Resources-datasets-libraryDatasetsLibrary
This library, which was started by the BitCuratorEdu project, collects a number of publicly available datasets that can be used for educational, testing, and research purposes.

BitCuratorEdu

BCC Digital Forensics Sample Data

https://github.com/BitCurator/bcc-dfa-sample-data/

This GitHub repository contains sample data originally collected or created for Cal Lee and Kam Woods's SAA "Advanced Digital Forensics" class, which can be used on its own or in conjunction with the BitCurator Hands-On Exercises compiled as part of the BitCuratorEdu project.