

# Installing BitCurator: Getting Started with the Virtual Machine

Discussion questions to pair with the screencast

<a href="#">Authors</a>	<a href="#">1</a>
<a href="#">Description</a>	<a href="#">1</a>
<a href="#">Learning object type</a>	<a href="#">1</a>
<a href="#">Learning objectives</a>	<a href="#">1</a>
<a href="#">Screencast</a>	<a href="#">2</a>
<a href="#">Discussion Questions</a>	<a href="#">2</a>
<a href="#">Answer Key</a>	<a href="#">4</a>
<a href="#">Tools and Resources Mentioned in This Document</a>	<a href="#">7</a>

## Authors

Cal Lee, Hannah Wang

## Description

These discussion questions can be used to encourage student engagement with the BitCurator screencast, [Getting Started with the Virtual Machine](#). The questions can also be used for discussion accompanying a live demonstration, a guided hands-on exercise, or independent exploration of the BitCurator Environment.

## Learning object type

Lesson plan/materials

## Learning objectives

## BitCuratorEdu Learning Object

This learning object might be used in a lesson to satisfy the following learning objectives:

- Determine the technical requirements and dependencies for installing open-source tools to process born-digital materials.
- Practice using tools in the BitCurator Environment.
- Test and evaluate tools for use in born-digital archiving workflows.
- 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.

## Screencast

<https://youtu.be/R2CldxWIXA>

## Discussion Questions

*These discussion questions can be used to encourage student engagement with the BitCurator screencast linked above. The questions can also be used for discussion accompanying a live demonstration, a guided hands-on exercise, or independent exploration of the BitCurator Environment. Video timestamps are included in parentheses, where applicable.*

1. The video explains that the BitCurator environment is distributed as both (1) a virtual machine (VM) and (2) a live ISO to be run on a dedicated machine. What are the differences between these two options?
2. What is VirtualBox, and what role is it playing in the setup and use of the BitCurator environment?
3. What are the advantages of running the BitCurator environment as a VM rather than a dedicated operating system (OS)?
4. What are the advantages of running the BitCurator environment as a dedicated OS rather than a VM?
5. The BitCurator VM is available as a .tar.gz file (3:48). What is a .tar.gz file?

## BitCuratorEdu Learning Object

6. Why do you think the environment is distributed as a .tar.gz file?
7. What is the software doing to extract the files (3:58)?
8. What are the .vdi and .vbox files (4:27)?
9. The settings menus in VirtualBox allow you to change technical characteristics of the VM (7:04). Why would you want to change these settings?
10. One of the settings you can change is how much memory the guest (VM) operating system will use and how much memory it will leave for the host operating system. Why isn't this just set at an optimal level when the VM is distributed?
11. Why are many of the settings greyed out when the VM is running, and how should you address this issue?

### Answer Key

*These discussion questions can be used to encourage student engagement with the BitCurator screencast linked above. The questions can also be used for discussion accompanying a live demonstration, a guided hands-on exercise, or independent exploration of the BitCurator Environment. The questions **(in bold text)** ask students to analyze the social and technological context of the BitCurator Environment and the tools packaged in the distribution. Example answers are given (in regular text), though some questions are subjective and answers may vary, depending on the knowledge of the student and the scope of the class. Video timestamps are included in parentheses, where applicable.*

**The video explains that the BitCurator environment is distributed as both (1) a virtual machine (VM) and (2) a live ISO to be run on a dedicated machine.**

**What are the differences between these two options?**

As a virtual machine, the BitCurator environment runs as a “guest” on top of your computer’s normal operating system (the “host”). The VirtualBox software provides a hypervisor that translates Linux system calls into system calls within the host operating system (OS). When you run programs within the VM, they operate just as they would if they were running on top of Linux installed directly onto the computer.

By contrast, the ISO file allows you to install the BitCurator environment as the OS directly on the machine itself. This could be the only OS on the computer, or it could be running on one of several partitions, each of which could run its own OS.

**What is VirtualBox, and what role is it playing in the setup and use of the BitCurator environment?**

Oracle VM VirtualBox is an open-source hypervisor that allows you to run one OS on top of another OS. Rather than being a specific application, the BitCurator environment is an entire specialized distribution (based on Ubuntu) of the Linux operating system. VirtualBox allows you to use all of the functionality of the Linux OS without having to install Linux directly onto your computer.

### **What are the advantages of running the BitCurator environment as a VM rather than a dedicated operating system (OS)?**

- It can often be easier to get administrative approval and support to install VirtualBox and run a VM within it. Once you've demonstrated how and why the BitCurator environment is vital to your work, you can potentially make the case for setting it up as the native OS (either as the sole OS on the machine or on a dedicated partition).
- A VM can be a sandbox that you just delete and replace if you run into stability or security issues (e.g. viruses).
- The VM is easier to set up for testing, education, and demonstration. For example, students in a class or workshop can run the VM on whatever computer they would normally be using, provided that it meets the system requirements (see the QuickStart Guide at <https://github.com/BitCurator/bitcurator-distro/wiki/Releases#quickstart-guide>).

### **What are the advantages of running the BitCurator environment as a dedicated OS rather than a VM?**

Running the BitCurator environment as a VM introduces an extra layer in between the hardware and the OS. Exchange of data and instructions between the guest and host usually works quite smoothly, but it can sometimes introduce problems with reading peripheral devices that are attached to the host. The input/output (I/O) speed between the guest and the hardware can also be significantly slower, so you would not want to image a large disk (e.g. 2 TB) from within the VM. Instead, you can create the disk image within the host OS and then move the disk image into the VM (see Setting up Shared Folders, [https://youtu.be/35Pa\\_PTfMn8](https://youtu.be/35Pa_PTfMn8)).

### **The BitCurator VM is available as a .tar.gz file (3:48). What is a .tar.gz file?**

Tar allows you to package a set of files and folders into a single file (often called a tarball). In this way, it's quite similar to ZIP. Computer professionals often refer to both tar and ZIP files as "archives," meaning that they store data that has been pulled out of a live system and set aside together for storage or transfer. The name tar is based on "tape archive"; the format was originally developed for writing data to devices (including tape) that did not have their own filesystem. In contrast to

## BitCuratorEdu Learning Object

directly copying a directory of files, writing them into a tar file retains a variety of filesystem attributes including names, directory structures and timestamps. If you've ever zipped up a set of files, you may have noticed that the software usually compresses the data, in addition to packaging it together. Tar doesn't compress data, so there's often an additional step of applying GNU zip (gzip) compression. Extracting the files from the .tar.gz package then involves two steps: uncompressing and then unpacking.

### **Why do you think the environment is distributed as a .tar.gz file?**

The tar package ensures that multiple files in the same folder will remain associated with each other (in this case the VBOX and VDI files). GNU Zip compresses the data, so the download is smaller.

### **What is the software doing to extract the files (3:58)?**

It's first uncompressing the data (compressing using GNU Zip), and then it's pulling the data out of the tar package and writing it to the local filesystem as individual files.

### **What are the .vdi and .vbox files (4:27)?**

The Virtual Disk Image (VDI) file is the VM itself. Note that it's much bigger than the .vbox file. The .vbox file contains settings for the VM (e.g. how much memory from the host computer should be devoted to the guest VM). The two files should have the same prefix (name before the extension) and sit in the same directory/folder.

### **The settings menus in VirtualBox allow you to change technical characteristics of the VM (7:04). Why would you want to change these settings?**

Students could provide many different answers. For example, one might want to adjust the CPU or memory resources devoted to the VM; how it handles ports and peripheral devices; what IP addresses the VM uses for internet activities; how much storage is allocated to the VM's virtual drive; and set up shared folders.

### **One of the settings you can change is how much memory the guest (VM) operating system will use and how much memory it will leave for the host**

### **operating system. Why isn't this just set at an optimal level when the VM is distributed?**

The distributors of the VM don't know how much memory your computer has, so the VM has a default setting based on a minimum value. If your computer has more than that minimum, you can allocate more of it for the VM.

### **Why are many of the settings greyed out when the VM is running, and how should you address this issue?**

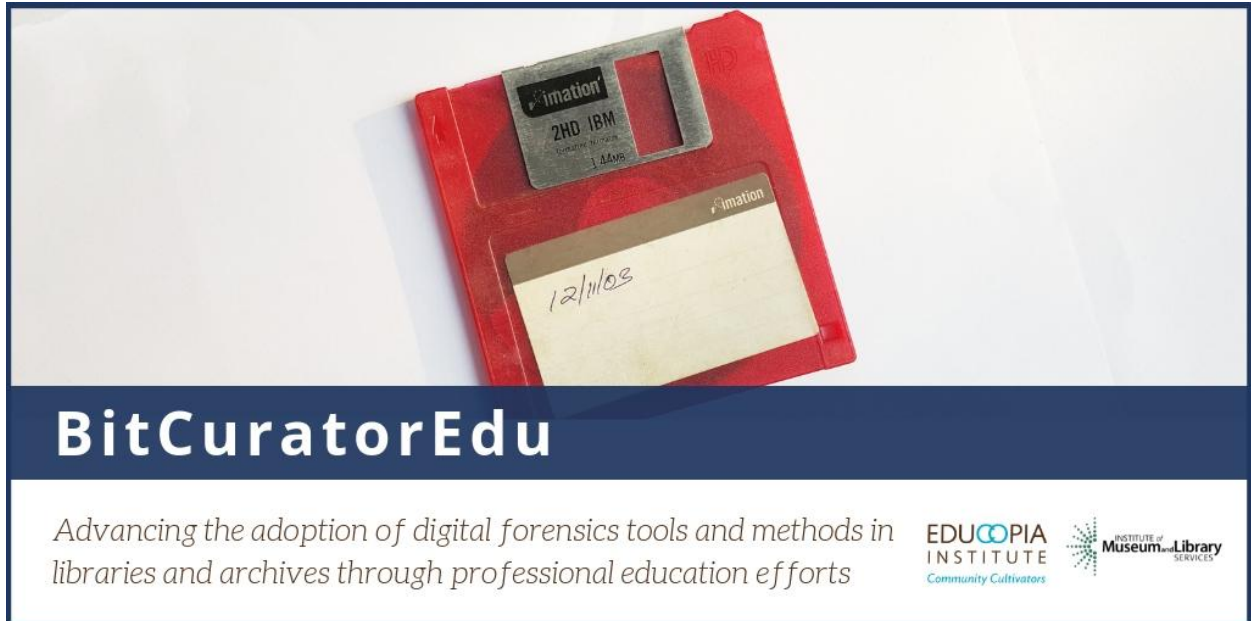
VirtualBox consults the settings in the VBOX file before it starts up the VM, so it knows which resources to devote to the VM. It's essentially creating an imaginary computer that the VM runs on. Just as, for example, you can't change the amount of RAM on your computer while it's still running, you can't do this with the VM. You need to shut down the VM before you can make many of the setting changes.

## Tools and Resources Mentioned in This Document

BitCurator Consortium. (2019, June 4). BitCurator Software: Setting up Shared Folders. [https://www.youtube.com/watch?v=35Pa\\_PtFmN8](https://www.youtube.com/watch?v=35Pa_PtFmN8)

Tar. (2021). In *Wikipedia*. [https://en.wikipedia.org/wiki/Tar\\_\(computing\)](https://en.wikipedia.org/wiki/Tar_(computing))

VirtualBox: <https://www.virtualbox.org/>



This resource was released by the BitCuratorEdu project and is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

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.

*The [BitCuratorEdu project](#) is a three-year effort (2018-2021) funded by the [Institute of Museum and Library Services \(IMLS\)](#) to study and advance the adoption of digital forensics tools and methods in libraries and archives through professional education efforts. This project is a partnership between [Educo피아 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.*