Extracting Data from Office Documents Exercise

BitCuratorEdu

Last Updated: January 18, 2022

About This Exercise

Author
Cal Lee

Description
This hands-on exercise is meant to introduce students to methods for extracting hidden data from Microsoft Office files. These slides are excerpted from Cal Lee's SAA "Advanced Digital Forensics" class. The sample data referenced in these slides is available here: https://github.com/BitCurator/bcc-dfa-sample-data/

Learning object type
Lesson plan/materials

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

- Practice using tools in the BitCurator Environment.

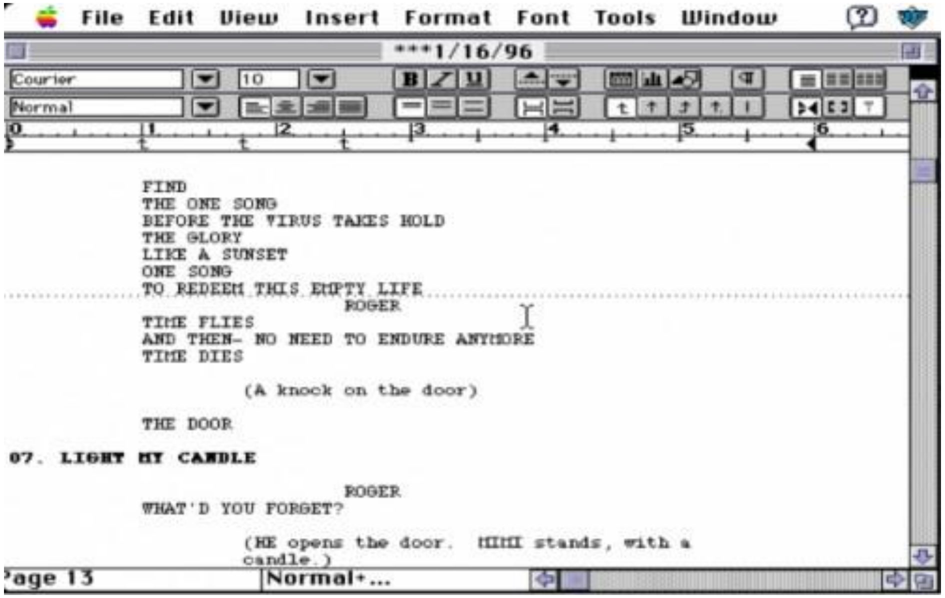
Office Documents

- Are the “new” office formats (ODF and OOXML) better or worse for forensics?
- What kinds of information can you get out of them?
- What sorts of approaches might you take to view and/or extract the information?

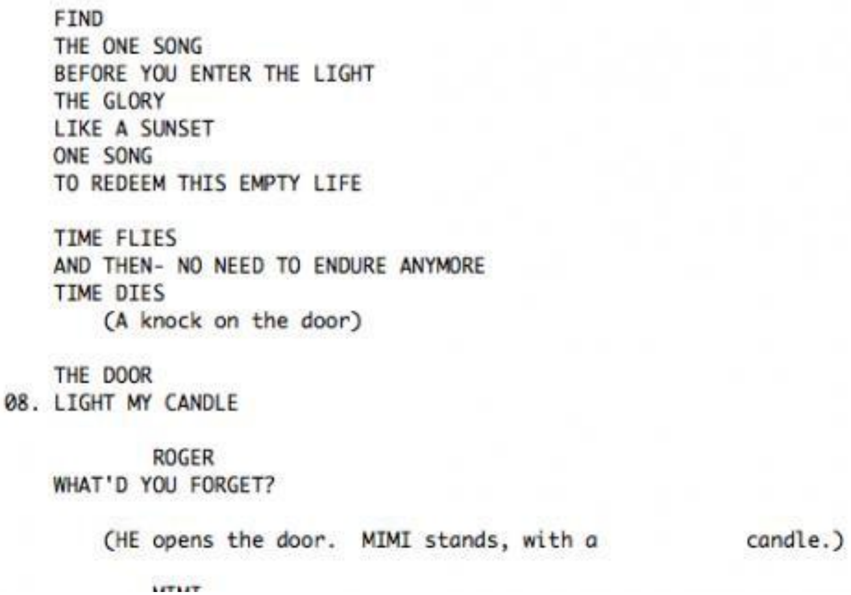
Office Documents - PPTX File Example

- Download (or copy for your flash drive) the following to your desktop: https://github.com/BitCurator/bcc-dfa-sample-data/blob/main/The%20NDSA%20Levels%20of%20Digital%20Preservation_3.pptx
- Change the file extension to .zip
- Open it with 7-Zip or WinZip
- Extract all the files
- Examine the contents of the resulting directory
  - □ Can you find a thumbnail of the first slide?
  - □ Where are the slides stored?
  - □ Where are embedded images stored?
  - □ Can you determine who created the file?

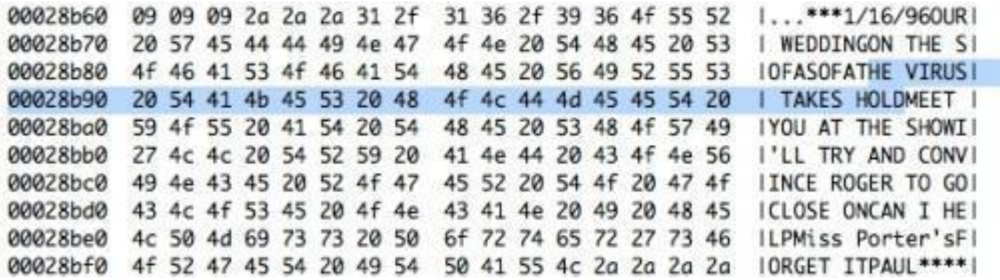
Jonathan Larson Fast Save Example



```
          FIND
          THE ONE SONG
          BEFORE THE VIRUS TAKES HOLD
          THE GLORY
          LIKE A SUNSET
          ONE SONG
          TO REDEEM THIS EMPTY LIFE
                              ROGER
          TIME FLIES
          AND THEN- NO NEED TO ENDURE ANYMORE
          TIME DIES

                  (A knock on the door)

          THE DOOR

07. LIGHT MY CANDLE

                              ROGER
          WHAT'D YOU FORGET?

                  (HE opens the door.  MIMI stands, with a
                  candle.)
```



```
    FIND
    THE ONE SONG
    BEFORE YOU ENTER THE LIGHT
    THE GLORY
    LIKE A SUNSET
    ONE SONG
    TO REDEEM THIS EMPTY LIFE

    TIME FLIES
    AND THEN- NO NEED TO ENDURE ANYMORE
    TIME DIES
        (A knock on the door)

    THE DOOR
08. LIGHT MY CANDLE

            ROGER
    WHAT'D YOU FORGET?

        (HE opens the door.  MIMI stands, with a           candle.)
```



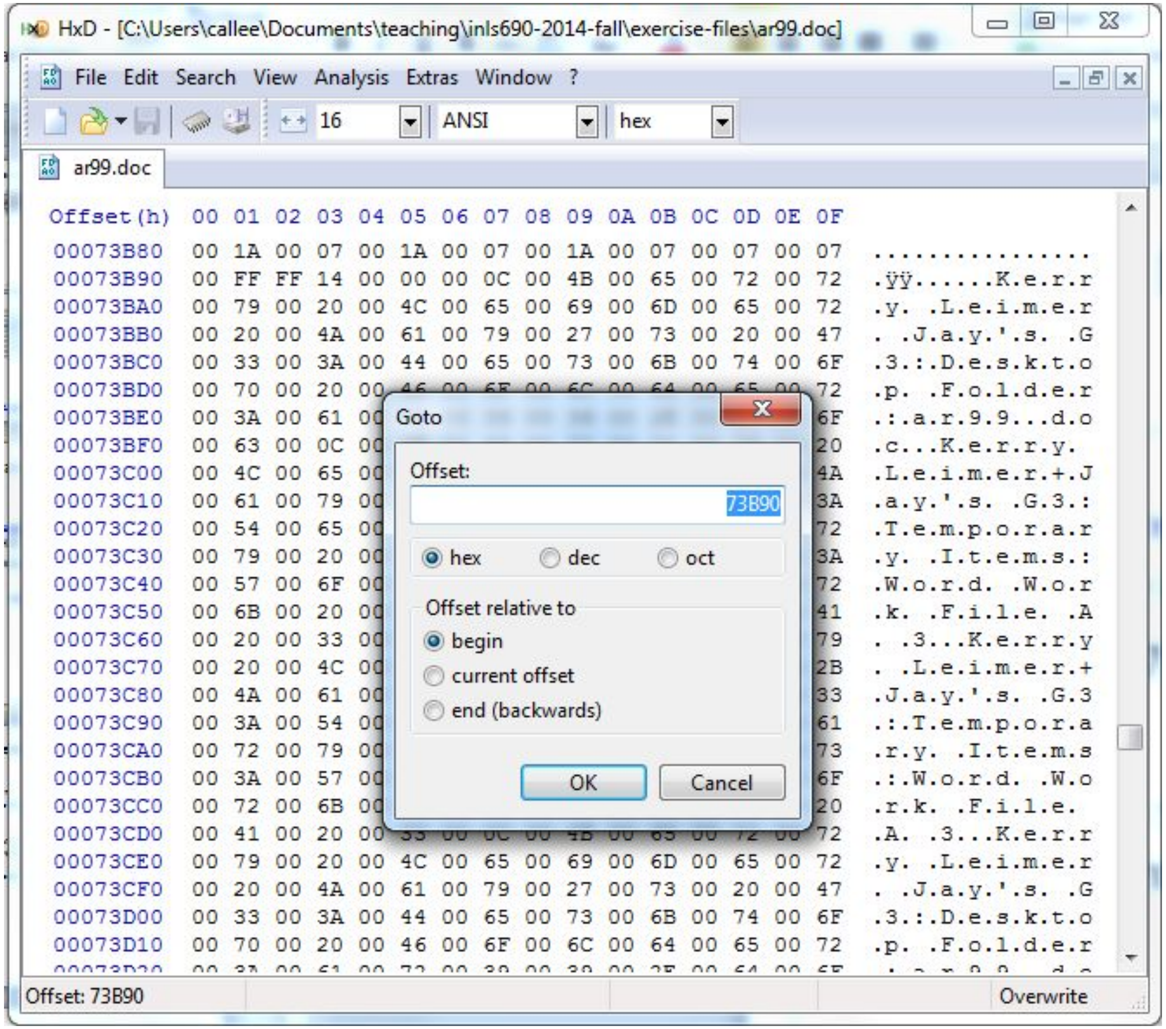
```
00028b60  09 09 09 2a 2a 2a 31 2f  31 36 2f 39 36 4f 55 52  |...***1/16/96OUR|
00028b70  20 57 45 44 44 49 4e 47  4f 4e 20 54 48 45 20 53  | WEDDINGON THE S|
00028b80  4f 46 41 53 4f 46 41 54  48 45 20 56 49 52 55 53  |OFASOFATHE VIRUS|
00028b90  20 54 41 4b 45 53 20 48  4f 4c 44 4d 45 45 54 20  | TAKES HOLDMEET |
00028ba0  59 4f 55 20 41 54 20 54  48 45 20 53 48 4f 57 49  |YOU AT THE SHOWI|
00028bb0  27 4c 4c 20 54 52 59 20  41 4e 44 20 43 4f 4e 56  |'LL TRY AND CONV|
00028bc0  49 4e 43 45 20 52 4f 47  45 52 20 54 4f 20 47 4f  |INCE ROGER TO GO|
00028bd0  43 4c 4f 53 45 20 4f 4e  43 41 4e 20 49 20 48 45  |CLOSE ONCAN I HE|
00028be0  4c 50 4d 69 73 73 20 50  6f 72 74 65 72 27 73 46  |LPMiss Porter'sF|
00028bf0  4f 52 47 45 54 20 49 54  50 41 55 4c 2a 2a 2a 2a  |ORGET ITPAUL****|
```

http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files

Hidden Data Exercise – Using a Hex Editor

Download (or copy from your flash drive): https://github.com/BitCurator/bcc-dfa-sample-data/blob/main/ar99.doc

- Open the file in HxD
- Go to offset 73B90 (use Search > Goto or just Control+G)
- What do you see there?What does it tell you about the document?



HxD - [C:\Users\callee\Documents\teaching\inls690-2014-fall\exercise-files\ar99.doc]

```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00073B80  00 1A 00 07 00 1A 00 07 00 1A 00 07 00 07 00 07  ................
00073B90  00 FF FF 14 00 00 00 0C 00 4B 00 65 00 72 00 72  .ÿÿ......K.e.r.r
00073BA0  00 79 00 20 00 4C 00 65 00 69 00 6D 00 65 00 72  .y. .L.e.i.m.e.r
00073BB0  00 20 00 4A 00 61 00 79 00 27 00 73 00 20 00 47  . .J.a.y.'.s. .G
00073BC0  00 33 00 3A 00 44 00 65 00 73 00 6B 00 74 00 6F  .3.:.D.e.s.k.t.o
00073BD0  00 70 00 20 00 46 00 6F 00 6C 00 64 00 65 00 72  .p. .F.o.l.d.e.r
00073BE0  00 3A 00 61 00                               6F  .:.a.r.9.9...d.o
00073BF0  00 63 00 0C 00                               20  .c...K.e.r.r.y.
00073C00  00 4C 00 65 00                               4A  .L.e.i.m.e.r.+.J
00073C10  00 61 00 79 00                               3A  .a.y.'.s. .G.3.:
00073C20  00 54 00 65 00                               72  .T.e.m.p.o.r.a.r
00073C30  00 79 00 20 00                               3A  .y. .I.t.e.m.s.:
00073C40  00 57 00 6F 00                               72  .W.o.r.d. .W.o.r
00073C50  00 6B 00 20 00                               41  .k. .F.i.l.e. .A
00073C60  00 20 00 33 00                               79  . .3...K.e.r.r.y
00073C70  00 20 00 4C 00                               2B  . .L.e.i.m.e.r.+
00073C80  00 4A 00 61 00                               33  .J.a.y.'.s. .G.3
00073C90  00 3A 00 54 00                               61  .:.T.e.m.p.o.r.a
00073CA0  00 72 00 79 00                               73  .r.y. .I.t.e.m.s
00073CB0  00 3A 00 57 00                               6F  .:.W.o.r.d. .W.o
00073CC0  00 72 00 6B 00                               20  .r.k. .F.i.l.e.
00073CD0  00 41 00 20 00 33 00 0C 00 4B 00 65 00 72 00 72  .A. .3...K.e.r.r
00073CE0  00 79 00 20 00 4C 00 65 00 69 00 6D 00 65 00 72  .y. .L.e.i.m.e.r
00073CF0  00 20 00 4A 00 61 00 79 00 27 00 73 00 20 00 47  . .J.a.y.'.s. .G
00073D00  00 33 00 3A 00 44 00 65 00 73 00 6B 00 74 00 6F  .3.:.D.e.s.k.t.o
00073D10  00 70 00 20 00 46 00 6F 00 6C 00 64 00 65 00 72  .p. .F.o.l.d.e.r
00073D20  00 3A 00 61 00 72 00 39 00 39 00 2E 00 64 00 6F  .:.a.r.9.9...d.o
```

Goto

Offset: 73B90

● hex ○ dec ○ oct

Offset relative to

● begin

○ current offset

○ end (backwards)

OK Cancel

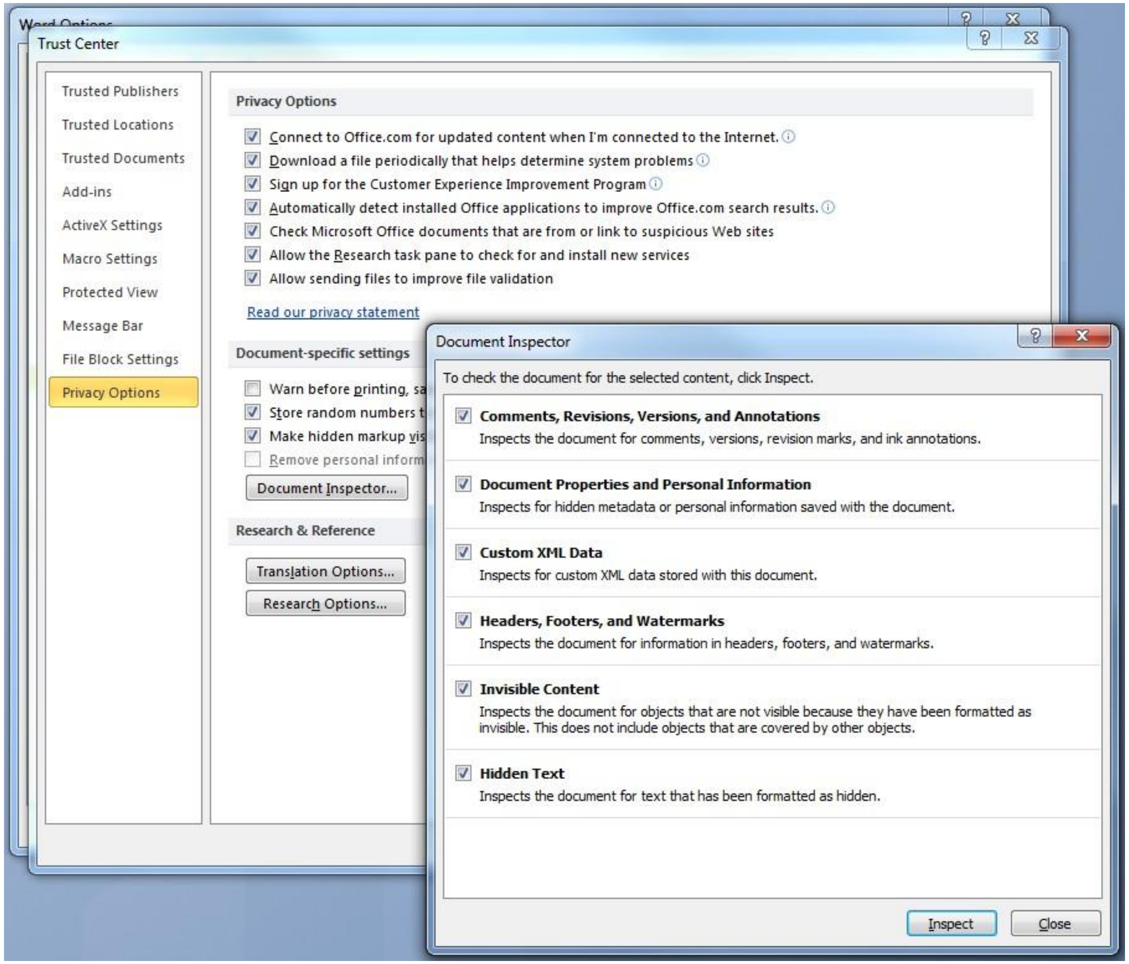
Offset: 73B90 Overwrite

Hidden Data Exercise – Inspection in MS Word

- Do the following:
  - Open it in Word – what is it?
  - If prompted to do so at the top, select “Enable Editing”
  - Select: File > Options > Trust Center > Trust Center Settings…
  - Then Privacy Options >

    Document Inspector >

    Inspect



Privacy Options

☑ Connect to Office.com for updated content when I'm connected to the Internet.
☑ Download a file periodically that helps determine system problems
☑ Sign up for the Customer Experience Improvement Program
☑ Automatically detect installed Office applications to improve Office.com search results.
☑ Check Microsoft Office documents that are from or link to suspicious Web sites
☑ Allow the Research task pane to check for and install new services
☑ Allow sending files to improve file validation

Read our privacy statement

Document-specific settings

☐ Warn before printing, sa
☑ Store random numbers t
☑ Make hidden markup vis
☐ Remove personal inform

Document Inspector

To check the document for the selected content, click Inspect.

☑ **Comments, Revisions, Versions, and Annotations**
Inspects the document for comments, versions, revision marks, and ink annotations.

☑ **Document Properties and Personal Information**
Inspects for hidden metadata or personal information saved with the document.

☑ **Custom XML Data**
Inspects for custom XML data stored with this document.

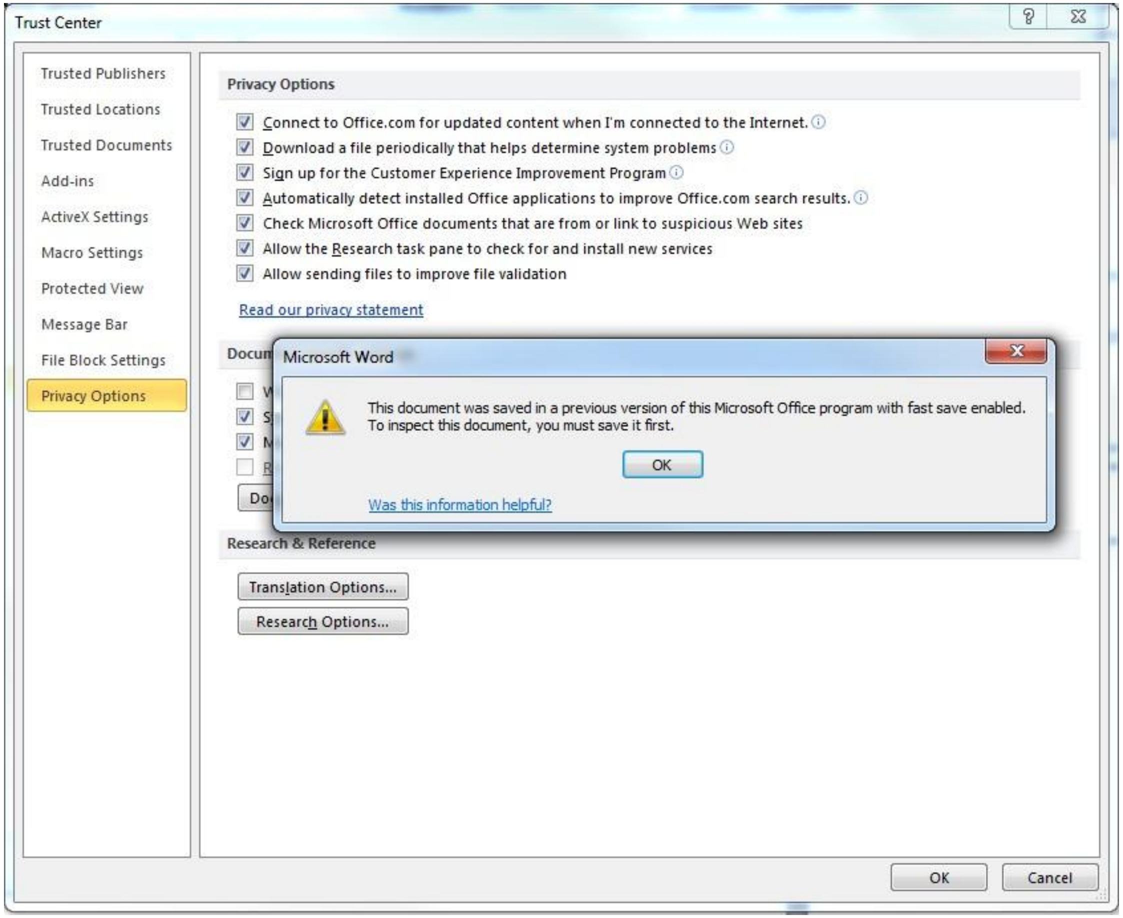
☑ **Headers, Footers, and Watermarks**
Inspects the document for information in headers, footers, and watermarks.

☑ **Invisible Content**
Inspects the document for objects that are not visible because they have been formatted as invisible. This does not include objects that are covered by other objects.

☑ **Hidden Text**
Inspects the document for text that has been formatted as hidden.

Hidden Data Exercise – Inspection in MS Word

- Are you prompted with this?
- Why do you think this is?
- If you see this, click OK, then save the document
- Run Document Inspector again
- What does it tell you?



Privacy Options

☑ Connect to Office.com for updated content when I'm connected to the Internet.

☑ Download a file periodically that helps determine system problems

☑ Sign up for the Customer Experience Improvement Program

☑ Automatically detect installed Office applications to improve Office.com search results.

☑ Check Microsoft Office documents that are from or link to suspicious Web sites

☑ Allow the Research task pane to check for and install new services

☑ Allow sending files to improve file validation

Microsoft Word

This document was saved in a previous version of this Microsoft Office program with fast save enabled. To inspect this document, you must save it first.

OK

Was this information helpful?

imation
2HD IBM
12/11/03
BitCuratorEdu
Advancing the adoption of digital forensics tools and methods in libraries and archives through professional education efforts
EDUCOPIA INSTITUTE
Community Cultivators
INSTITUTE of Museum and Library SERVICES

The BitCuratorEdu project is a three-year effort 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 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.