



## CUNY Television Preliminary NDSA Levels Assessment

Andrew Weaver

September, 2016

### **Summary:**

After conducting an assessment using the National Digital Stewardship Alliance (NDSA) levels of digital preservation, I assigned CUNY Television a score of thirteen out of a possible twenty points. (For scoring system see methodology section). The Levels of preservation are a tiered matrix developed by the NDSA to aid institutions both in evaluating the current state of their digital preservation systems and with planning for future enhancements. Documentation for the NDSA levels is available online at: [http://www.digitalpreservation.gov/documents/NDSA\\_Levels\\_Archiving\\_2013.pdf](http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf).

CUNY Television is most compliant in the areas of 'File Formats' and 'Metadata', with scores of three and a half out of a possible four in each. It has the most room for growth in the areas of 'Storage and Geographic Location' (one and a half points) and 'Information Security' (two points). The final area, 'File Fixity and Data Integrity' has a relatively high score (two and a half points) that could be further enhanced with minor changes to current systems.

### **NDSR Project Information**

This assessment was conducted as a component of the 2016-17 CUNY TV National Digital Stewardship Residency Project (NDSR) overseen by the American Archive of Public Broadcasting (AAPB). The NDSR program is built around creating cohorts of professionals in the digital preservation field and placing them at project sites to aid in the development of digital stewardship models. Information about the CUNY Television project can be found at: <https://aapbndsr.files.wordpress.com/2016/01/cunytelevisionaapbndsrprojectproposal-projectplanonly.pdf>.

## **Methodology:**

This preliminary assessment was prepared both through observation of CUNY Television workflows and by direct querying of archivists at the CUNY Television Library. Points of observation and inquiry were drawn from the NDSA Levels of Digital Preservation matrix. Although the matrix does not directly support a point based evaluation, it was decided to assign point values to matrix levels to facilitate comparison of current levels, anticipated post-NDSR levels and actual post-NDSR levels.

For NDSA levels that CUNY Television was found to be fully compliant with, a value of one was assigned. For partially compliant levels a value of one-half was assigned, with a value of zero being assigned for non-compliant levels. As the matrix contains five categories with four levels each, a total of twenty points was possible.

## **Results by Category:**

### **Storage and Geographic Location:**

CUNY Television uses LTO tapes for archival storage of content. Digitized and born-digital content ultimately is ingested using a system of micro-services and written to LTO. CUNY TV is currently involved in a migration from LTO 5 to LTO 7 and maintains documentation about this process. For data written to tape at least two copies exist, with more in some cases.

CUNY TV's score in this area is limited primarily by lack of geographic diversity for copies. Storing one or more backup copies of LTO tapes in an offsite location (preferably in a region with different disaster risks) would greatly enhance CUNY TV's compliance in this category. **Score: 1.5**

### **File Fixity and Data Integrity:**

CUNY Television actively creates and validates fixity data for ingested files. All LTO tapes contain an MD5 manifest for all files. Currently these MD5s are verified on LTO tape creation and migration. As backup tapes exist, CUNY TV is able to replace corrupted content in the event a damaged file is detected.

Areas that CUNY TV is able to improve in this category are virus scanning of files and

more frequent verification of fixity information for files. Currently information is not scanned for viruses before being preserved on LTO tape. There also is no confirmation of checksum validity when information is accessed on tapes (except for migration purposes). Though migrations can be anticipated to occur roughly every four to five years, this does not seem to be frequent enough to satisfy the NDSA Level's requirement to "check fixity in response to specific events or activities." It is suggested that possible additional fixity verification points be investigated. **Score: 2.5**

### **Information Security:**

Access to files at CUNY TV is gained through physical access to the LTO tapes. Tapes are stored both in the CUNY TV Library and in the station's Master Control Room. The Master Control Room has restricted access using a keycard system. While this means access is limited to Archives and MCR staff, there are many actions that could theoretically be performed on files without logging/detection.

CUNY TV could become more compliant in this area with increased/more centralized logging of events and actors as well as more explicit documentation of content access policies. **Score: 2**

### **Metadata:**

CUNY TV currently collects and maintains an enormous variety of metadata about files ranging the full spectrum of administrative, descriptive, technical and preservation. If evaluated by the letter of the NDSA levels this makes CUNY TV extremely compliant, although ease of use could benefit from some restructuring and centralization of aspects of this metadata. This would make CUNY TV more compliant with the presumed intent of the NDSA levels.

It is suggested that fixity data and information corresponding to PREMIS events be stored in a centralized location (along with any other preservation metadata deemed to be useful by stakeholders). **Score: 3.5**

### **File Formats:**

File format obsolescence does not appear to present a significant problem for the CUNY TV collection. The archives have input into filetypes being used by content creators as well as an effective ingest system for normalization. Technical expertise and tools exist within the CUNY TV archive that appear sufficient to ensure playback of all preserved information stored for the foreseeable future.

As a complete inventory of file formats is explicitly specified in the NDSA matrix, it is suggested that a survey of file types held by the library be conducted to attain maximum compliance. **Score: 3.5**

## **Conclusion and Suggestions for Initial Steps:**

With a total score of thirteen points out of a possible twenty, the CUNY TV Library has a relatively high level of compliance with NDSA recommendations. In the areas where there is room for growth, I found that many could be improved by making modifications to existing systems for more centralization of data.

In addition to an NDSA levels assessment, the CUNY TV NDSR project plan calls for an evaluation of recommendations to be implemented within the scope of the AAPB NDSR project period. Accordingly, I have ranked suggestions by actions that will most efficiently achieve progress towards raising CUNY TV's overall NDSA score.

I suggest that initial efforts focus on building an infrastructure for automatically centralizing data that is generated both about and by the microservices that form the core of CUNY TV workflows. Following completion of that infrastructure, enhancement of workflows to include actions such as virus scanning should be considered, with a survey of collection file types being conducted at this point.

As CUNY TV is currently involved in a migration from LTO 5 to LTO 7, this is also an ideal opportunity to consider offsite storage for LTO tapes. There should be an ongoing discussion during the migration process with the practicability and effectiveness of various options being considered.

## Appendix A: NDSA Levels coded for findings

**Table 1: Version 1 of the Levels of Digital Preservation**

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	<ul style="list-style-type: none"> <li>- Two complete copies that are not collocated</li> <li>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system</li> </ul>	<ul style="list-style-type: none"> <li>- At least three complete copies</li> <li>- At least one copy in a different geographic location</li> <li>- Document your storage system(s) and storage media and what you need to use them</li> </ul>	<ul style="list-style-type: none"> <li>- At least one copy in a geographic location with a different disaster threat</li> <li>- Obsolescence monitoring process for your storage system(s) and media</li> </ul>	<ul style="list-style-type: none"> <li>- At least three copies in geographic locations with different disaster threats</li> <li>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems</li> </ul>
File Fixity and Data Integrity	<ul style="list-style-type: none"> <li>- Check file fixity on ingest if it has been provided with the content</li> <li>- Create fixity info if it wasn't provided with the content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity on all ingests</li> <li>- Use write-blockers when working with original media</li> <li>- Virus-check high risk content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of content at fixed intervals</li> <li>- Maintain logs of fixity info; supply audit on demand</li> <li>- Ability to detect corrupt data</li> <li>- Virus-check all content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of all content in response to specific events or activities</li> <li>- Ability to replace/repair corrupted data</li> <li>- Ensure no one person has write access to all copies</li> </ul>
Information Security	<ul style="list-style-type: none"> <li>- Identify who has read, write, move and delete authorization to individual files</li> <li>- Restrict who has those authorizations to individual files</li> </ul>	<ul style="list-style-type: none"> <li>- Document access restrictions for content</li> </ul>	<ul style="list-style-type: none"> <li>- Maintain logs of who performed what actions on files, including deletions and preservation actions</li> </ul>	<ul style="list-style-type: none"> <li>- Perform audit of logs</li> </ul>
Metadata	<ul style="list-style-type: none"> <li>- Inventory of content and its storage location</li> <li>- Ensure backup and non-collocation of inventory</li> </ul>	<ul style="list-style-type: none"> <li>- Store administrative metadata</li> <li>- Store transformative metadata and log events</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard technical and descriptive metadata</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard preservation metadata</li> </ul>
File Formats	<ul style="list-style-type: none"> <li>- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs</li> </ul>	<ul style="list-style-type: none"> <li>- Inventory of file formats in use</li> </ul>	<ul style="list-style-type: none"> <li>- Monitor file format obsolescence issues</li> </ul>	<ul style="list-style-type: none"> <li>- Perform format migrations, emulation and similar activities as needed</li> </ul>

Green=Compliant  
 Yellow=Partially Compliant  
 Red=Non-Compliant