

# Preserving NYARC's Web Archives

*A step towards long-term stewardship*

Karl-Rainer Blumenthal  
National Digital Stewardship Resident  
New York Art Resources Consortium (NYARC)

*Revised for publication: Date TBD*

**Draft: May 22, 2015**

## **Contents**

[I. Policy](#)

[II. Implementation](#)

[III. Management](#)

[IV. Opportunities and constraints](#)

[References](#)

## I. Policy

### A. Summary statement

Towards their long-term viability as critical resources in art historical scholarship, the New York Art Resources Consortium (NYARC) commits to sustaining the control, integrity, security, and accessibility of the born-digital source materials that it collects from the World Wide Web. NYARC implements tools and procedures proven to effectively preserve the [Web ARChival \(WARC\) File Format \(ISO 28500:2009\)](#) and further advocates for the development of greater preservation insurances for that international standard.

### B. Scope

NYARC's web archiving program, instituted in 2013, addresses the need to collect, store, and preserve web-native resources vital to art historical scholarship before such materials are lost to a 'digital black hole' [1]. The program employs several policies, procedures, and technological tools towards sustaining continuous and on-demand access to these resources among members of their primary user community, details of which may be reviewed in brief form on [NYARC's Web Archiving Program FAQ](#), and in increasing depth on [the consortium's documentation site](#). This documentation specifically addresses the accumulation and management of *preservation copies* [2] throughout and beyond the foreseeable lifespan of any *access copies* [3] of NYARC's web archives.

### C. Core responsibilities

Stewardship of preservation copies of NYARC's web archives beyond the lifespan of any current or future harvesting and/or access environment requires NYARC to ensure their viability in four principal areas:



#### 1. Control

NYARC must sustain control over complete copies of all WARC files collected in its program's archives by maintaining redundant copies among multiple geographically dispersed locations and within a storage system apart from any current or future harvesting or access system component.



#### 2. Integrity

All preservation copies of NYARC's WARCs must be systematically monitored for fixity ("the property of a digital file or object being fixed or unchanged" [4]) at the bit level to ensure that no data loss or corruption occurs upon ingest into or thereafter within their long term storage environment. To account for these files' provenance, or 'digital chain of custody,' the storage system must systematically generate and continuously provide access to fixity information in an openly accessible format.



### 3. Security

Access privileges to the storage environment for preservation copies of NYARC's web archives must be restricted to include only core web archiving program and storage system maintenance personnel. In order to ensure their accurate and complete provenance, all interactions with (ie. ingesting, copying, moving, deleting, etc.) these files must automatically be logged and made accessible in an open format by the storage system.



### 4. Accessibility

While NYARC maintains access copies of its web archives for the purpose of immediate online access, it must also ensure the longer term accessibility of its preservation copies. It must always ingest these into the storage environment in the most current international standard file format for web archives, and must always advocate for expanded tools to validate and characterize these files' contents with standard preservation metadata towards any required format migration and/or environment emulation.

## II. Implementation

### A. DuraCloud Sync

To meet its responsibilities with respect to their archival control, integrity, and security, and until a successor system is deemed necessary, NYARC will manage preservation copies of all of its current and future web archives through a storage system designed by the non-profit [DuraSpace Organization](#), called *DuraCloud*--"an open source platform and managed service that provides on-demand storage services for digital content in the cloud" [5]. DuraCloud will enable NYARC to automatically ingest preservation copies of its WARC files into redundant, geographically dispersed data storage centers, to manage and log all access to those files, and to monitor and report fixity information.

Upon the approval of a service agreement between executive representatives of the web archiving program and of the DuraSpace Organization, DuraCloud will immediately ingest all WARC files collected by NYARC from Archive-It's LOCKSS API and will by means of its REST API thereafter immediately and automatically synchronize this collection with any new acquisitions.

### 1. Terms of service

#### Duration

The initial service agreement between NYARC and DuraSpace will cover a period of one calendar year, July 2015 - July 2016.

#### Service level and pricing

The service level for this agreement will be "Preservation Plus," as per the most recent DuraCloud pricing schedule [6]. In addition to standard online access, web console management, fixity checks, and fixity information, this service level will enable automatic synchronization and recovery capabilities for

files between two data storage providers: Amazon S3 and Amazon Glacier. It will afford NYARC 1TB of data storage capacity (not to include synchronized duplicate files) in return for the subscription fee of \$2,000, to be invoiced immediately upon the execution of the agreement, and with the understanding that an additional \$825 will be paid by NYARC per additional TB of data it accrues over the initial 1TB in the first subscription year.

#### Technical support

Upon finalization of the agreement, DuraSpace will provide NYARC the same webinar training opportunities and routine email support and issue tracking made available to all of its subscribers.

#### Succession planning

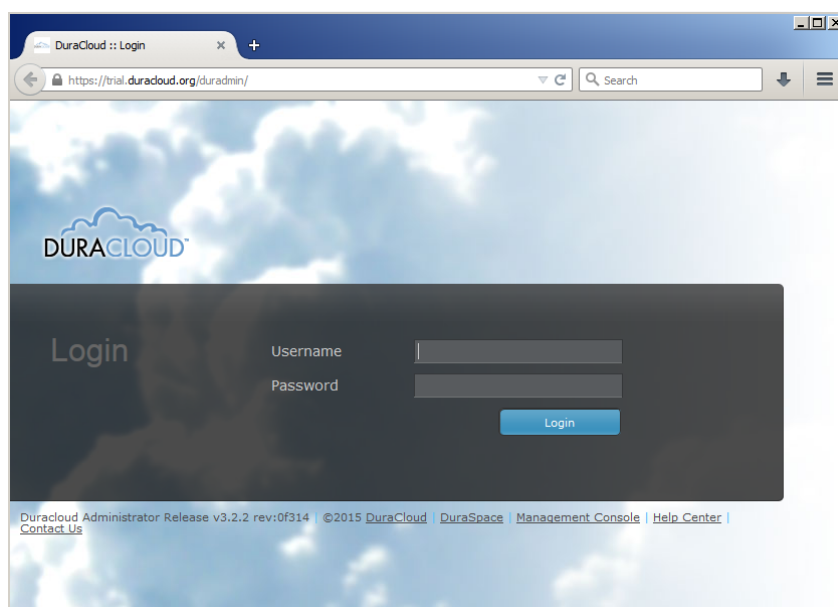
The agreement must include DuraSpace's assurance that an acceptable lead time and availability of data migration consultation will be provided to NYARC in the event of any termination of service provided by third party storage vendors named in the agreement/above.

In the event that they discontinue the DuraCloud service entirely, DuraSpace will deliver all necessary management credentials and instructions for third party data storage vendors directly to NYARC.

## III. Management

### **A. Access privileges**

Access to preservation copies of NYARC's web archives through the DuraCloud web console must be restricted to 1) the NYARC Web Archiving Program Coordinator and 2) the NYARC Coordinator & Systems Manager. No other person or institution is granted access to these files unless and until they are explicitly afforded that privilege in this document. Access credentials for all authorized personnel must be kept strictly confidential; DuraSpace reserves the right to suspend access by any user account that it deems compromised.



Once their user accounts and passwords have been created by DuraSpace, the personnel above may access preservation copies of NYARC's web archives for the purposes of monitoring and management through the DuraCloud web console here: <http://nyarc.duraspace.org>.

## B. Integrity checks

The authorized NYARC web archiving program personnel above will use their access to preservation copies of NYARC's web archives primarily for the purpose of verifying the success of or addressing potential errors detected within regular bit-level integrity--or *fixity*-- checks. DuraCloud runs and provides reports on these "health checks" of all of NYARC's files managed through its service on at least a semiannual basis, and NYARC supports the service's ambition to increase that frequency to a quarterly basis in 2015 [7].

The principal tool that DuraCloud uses to establish and monitor fixity, and to record fixity information, is the MD5 checksum value. MD5 is a cryptographic hashing algorithm that generates a 128-bit value equivalent to a 'digital fingerprint' for each file upon which it is run [8, 9]. A human user may read this value as a 32-digit hexadecimal string.

The screenshot displays the DuraCloud web console interface. At the top, there are navigation tabs: "Directory", "Files", and "Selection details". The "Files" tab is active, showing a list of files in a space. The "Selection details" tab is also visible, showing details for a selected file. A red box highlights the "Toggle storage environments" button in the top right corner. Another red box highlights the "Last Health Check" field in the "Space Detail" section, which shows a green status and a date/time stamp. A third red box highlights the "Access fixity information" button at the bottom right of the interface.

DuraCloud and each of the cloud storage services that its service overlays generate this value upon ingest and immediately thereafter compare those values among their services in order to verify that each file's unique value is consistent throughout the system. Each Amazon storage environment will thereafter re-generate the checksum value upon each regular fixity check [10]. Automatic verification of that value's precise match to the original checksum value recorded by DuraCloud at the time of ingest constitutes successful stewardship of the file's integrity at the bit level. Any failure of any file's

checksum value to match the original value recorded by DuraCloud at the time of ingest constitutes a critical failure and must immediately result in the automatic replacement of the corrupted file in one storage environment with an integral back-up from the redundant storage environment. Regardless of the success of any necessary file replacement, all errors encountered in fixity checking must be reported immediately and by the NYARC Web Archiving Program Coordinator directly to DuraSpace.

The screenshot displays the DuraCloud web interface. A red box highlights the 'Download Raw Report' button, with an arrow pointing to a label 'Download fixity information'. A blue box encompasses the entire report table, with an arrow pointing to a label 'Fixity information'. The report table contains the following data:

space-id	content-id	file-service-out/fix-integrity/fingerprints-carissa-images-2012-05-26T01:27.tsv	file-service-out/fix-integrity/fingerprints-carissa-images-2012-05-26T01:27.tsv	status
carissa-images	PrincetonProjective.gf	296475059964505512064a105a7	296475059964505512064a105a7	VALID
carissa-images	02125_wormingnet_1289x800.jpg	f185970718838910608a704094e1490	f185970718838910608a704094e1490	VALID
carissa-images	JaneAusten.gf	c661311846a9039a294d2373a66086	c661311846a9039a294d2373a66086	VALID
carissa-images	standing.p2	8b66a30d15616a3073ee4650bcbf	8b66a30d15616a3073ee4650bcbf	VALID
carissa-images	01465_wintermorning_1280x800.jpg	a7184937ad7946fad055833ac8172	a7184937ad7946fad055833ac8172	VALID
carissa-images	JaneAustenBiography.gf	3ef1e4c4237969ab056a30740c52a	3ef1e4c4237969ab056a30740c52a	VALID
carissa-images	tpar.txt	5bdc3f625d7a94659d9c23701799e	5bdc3f625d7a94659d9c23701799e	VALID
carissa-images	metad_EdwinBostonian.tif	8a24e106a3a8f603929423ac407a8	8a24e106a3a8f603929423ac407a8	VALID

All fixity information generated by ingest, by subsequent fixity checks, and by any other action taken by the DuraCloud service or by authorized personnel is recorded in the 'health reports' made available to NYARC through the DuraCloud web console, as illustrated above [11]. In consultation with DuraSpace and in alignment with its schedule for regular acquisitions of new web archival materials (in particular scheduled crawling of NYARC institutional web presences), the NYARC Web Archiving Program Coordinator may define the most effective schedule for auditing the results of these reports.

### C. Institutional archiving

While the system described above is its sole strategy for long term storage of preservation copies of its web archives, the NYARC web archiving program will facilitate the storage and/or preservation of web archive collections specific to each of its member institutions within the environment chosen by each institution.

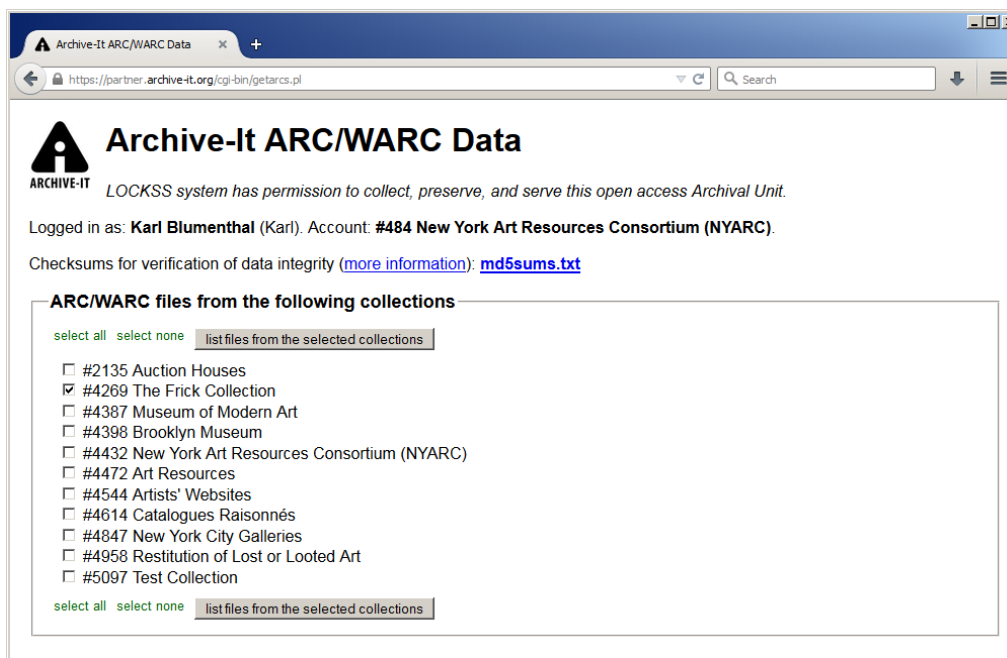
At the request of archives staff at the Brooklyn Museum Library & Archives, the Frick Art Reference Library, or the Museum of Modern Art Library, NYARC's Web Archiving Program Coordinator will deliver copies of all heretofore collected WARC files specific to the relevant institution and will coordinate a schedule for enriching that collection with the files acquired in regularly scheduled future institutional harvests.

Unless and until explicitly documented here, transmission of web archives in any form from the NYARC web archiving program to institutional archives must originate with files retrieved from access--not preservation--copy stores, and may only be effected by the NYARC Web Archiving Program Coordinator.

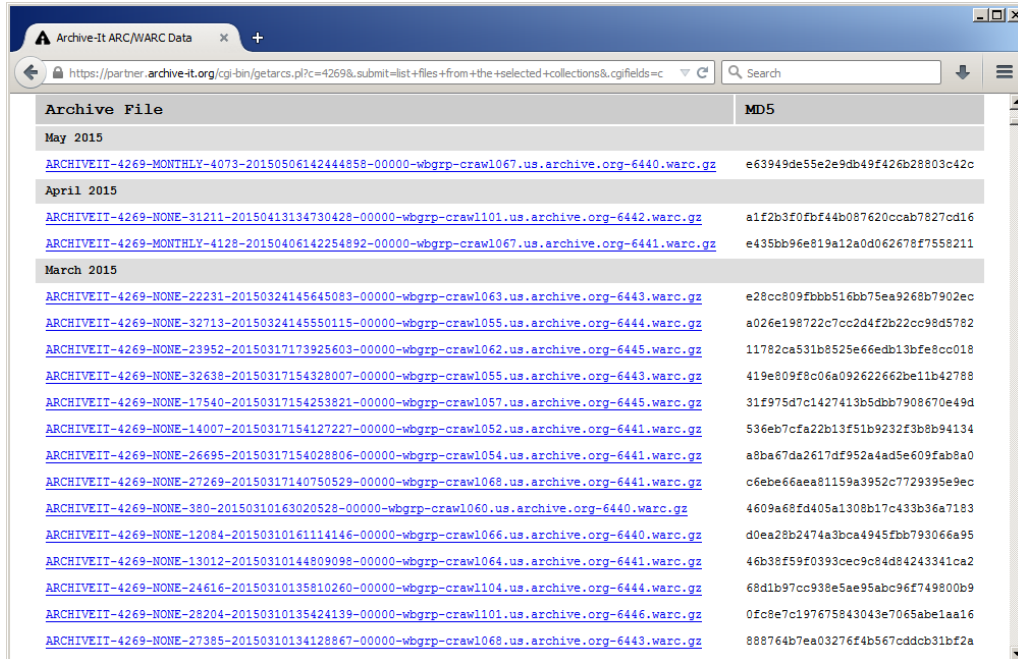
## 1. Preservica workflow

As of May 2015, the Frick Art Reference Library and the Museum of Modern Art Library have each selected to run [Preservica Cloud Edition](#) for the long term storage and preservation of their respective digitized and born-digital institutional collections. Upon the request of either institutional archives, the NYARC Web Archiving Program Coordinator is authorized to deliver copies of institutionally relevant WARC files for ingest into this storage and preservation environment by executing the following steps:

1. Navigate to access copies of NYARC's web archives here:  
<https://partner.archive-it.org/cgi-bin/getarcs.pl>

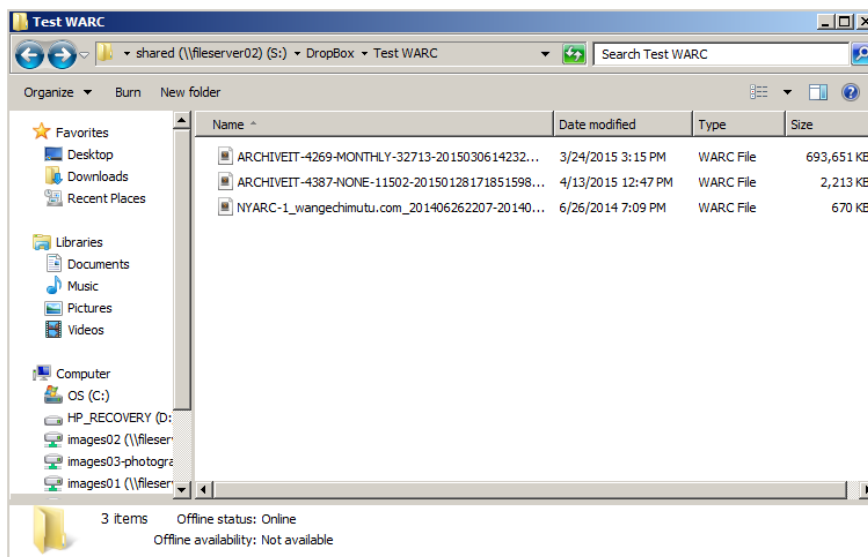


2. Select the relevant institutional collection and click on the "list files from the selected collections" button



Archive File	MD5
<b>May 2015</b>	
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-MONTHLY-4073-20150506142444858-00000-wbgrp-crawl067.us.archive.org-6440.warc.gz</a>	e63949de55e2e9db49f426b28803c42c
<b>April 2015</b>	
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-31211-20150413134730428-00000-wbgrp-crawl101.us.archive.org-6442.warc.gz</a>	a1f2b3f0fbf44b087620ccab7827cd16
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-MONTHLY-4128-20150406142254892-00000-wbgrp-crawl067.us.archive.org-6441.warc.gz</a>	e435bb96e819a12a0d062678f7558211
<b>March 2015</b>	
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-22231-20150324145645083-00000-wbgrp-crawl063.us.archive.org-6443.warc.gz</a>	e28cc809fbb516bb75ea9268b7902ec
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-32713-2015032414550115-00000-wbgrp-crawl055.us.archive.org-6444.warc.gz</a>	a026e198722c7cc2d4f2b22cc98d5782
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-23952-20150317173925603-00000-wbgrp-crawl062.us.archive.org-6445.warc.gz</a>	11782ca531b8525e66edb13bfe8cc018
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-32638-20150317154328007-00000-wbgrp-crawl055.us.archive.org-6443.warc.gz</a>	419e809f8c06a092622662be11b42788
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-17540-20150317154253821-00000-wbgrp-crawl057.us.archive.org-6445.warc.gz</a>	31f975d7c1427413b5d5b7908670e49d
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-14007-20150317154127227-00000-wbgrp-crawl052.us.archive.org-6441.warc.gz</a>	536eb7cfa22b13f51b9232f3b8b94134
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-26695-20150317154028806-00000-wbgrp-crawl054.us.archive.org-6441.warc.gz</a>	a8ba67da2617df952a4ad5e609fab8a0
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-27269-20150317140750529-00000-wbgrp-crawl068.us.archive.org-6441.warc.gz</a>	c6be66aea81159a3952c772395e9ec
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-380-20150310163020528-00000-wbgrp-crawl060.us.archive.org-6440.warc.gz</a>	4609a68fd405a1308b17c433b36a7183
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-12084-20150310161114146-00000-wbgrp-crawl066.us.archive.org-6440.warc.gz</a>	d0ea28b2474a3bca4945fbb793066a95
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-13012-20150310144809098-00000-wbgrp-crawl064.us.archive.org-6441.warc.gz</a>	46b385f9f0393cec9c84d424341ca2
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-24616-20150310135810260-00000-wbgrp-crawl104.us.archive.org-6444.warc.gz</a>	68d1b97cc938e5ae95abc96f74980b9
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-28204-20150310135424139-00000-wbgrp-crawl101.us.archive.org-6446.warc.gz</a>	0fc8e7c197675843043e7065abe1aa16
<a href="https://partner.archive-it.org/cgi-bin/getarcs.pl?c=4269&amp;.submit=list+files+from+the+selected+collections&amp;.cgfields=c">ARCHIVEIT-4269-NONE-27385-20150310134128867-00000-wbgrp-crawl068.us.archive.org-6443.warc.gz</a>	888764b7ea03276f4b567cddcb31bf2a

- Download all warc.gz files not yet ingested into the respective institutional archives' Preservica environment to a staging folder, or 'watch space,' created by archives staff for automated ingests into that environment.



Stewardship of the downloaded files into and within the Preservica storage and preservation environment is the responsibility of each institutional archives.

The process described above must initially deliver all WARC files contained within the relevant institutional collection, and thereafter only those files acquired after the date of this initial transmission. Coordinating a schedule for systematic transfer of successively acquired files is the responsibility of NYARC's Web Archiving Program Coordinator and the relevant staff at each



institutional archives; it is recommended, however, that these transmissions occur no more frequently than annually.

## D. Further and updated information

### 1. NYARC policy and procedures

It is the responsibility of the Web Archive Program Coordinator and/or the Coordinator & Systems Manager to update the above policy and procedures as implementation of NYARC's storage and preservation regime dictates. **Information regarding storage system selection, services, and pricing, must at the least be updated to reflect implementation status by no later than July 2016.**

### 2. DuraCloud implementation and management

For the most recent information on DuraCloud technical specifications, web console management tools, other add-ons and features, and for webinars and presentations on using this service, consult DuraCloud's wiki-based documentation site here:

<https://wiki.duraspace.org/display/DURACLOUD/DuraCloud>

## IV. Opportunities and constraints

### A. Accessibility and preservation metadata

NYARC's responsibility with respect to the long term accessibility of preservation copies of its web archives is bound to its ability to package those archives in their ultimate storage environment with standard and thorough preservation metadata in order to enable the eventual migration and/or emulation of their contents in obsolete digital file formats. As of May 2015, the under-development of tools, processes, and standards to reliably create and package standard preservation metadata for web archives precludes NYARC from fulfilling this responsibility.

Profiles for web archive metadata manifests that include the *de facto* preservation metadata standard, [PREMIS](#), have existed for nearly a decade [12], however the necessary tools to generate these manifests at the scale of a multi-institutional web archiving program have lagged far behind [13]. NYARC therefore advocates for the further development of preservation metadata generation for web archives in the following two areas:

#### 1. WARC metadata enrichment

Towards the long term viability of the WARC file format as an international standard for containing and enabling access to contents within web archives, NYARC supports updating the standard WARC metadata header record [14] in order to include or clarify vital information directly mappable to core PREMIS metadata fields of the *Object* semantic unit [15].

Specifically, this would include: 1) expanding the *format* field in the current WARC metadata header record into two fields--one for identification of the file format, and another for specifying the file format's version; 2) adding a metadata field to record the format registry from which information about that file format may be retrieved; and 3) adding a metadata field to record the identifier key for that format's entry within said format registry.

Such metadata enrichment would effectively make all subsequent WARC files collected by NYARC conformant to PREMIS Conformance Level 1A as per the latest PREMIS conformance statement [16].

## 2. Metadata extraction tools

Towards enabling the migration of contents internal to the WARC container file from obsolete formats and/or the emulation of access environments necessary to render those same contents in their original formats, NYARC supports the improvement of tools designed to identify and validate the file formats of those internal contents and to record their characteristics in PREMIS-conformant metadata manifests.

Tools presently standard to identification and validation services for digitized and born-digital records--[DROID](#) and [JHOVE](#)--cannot reliably perform these functions and output PREMIS-conformant records for files contained *within* a standard WARC [17]. NYARC therefore supports the development of next generation identification and validation tools, such as [JHOVE2](#) and [JHoNAS](#), to perform these functions and to subsequently be integrated into the repository ingest workflows of digital preservation service platforms like Preservica, [Archivematica](#), and [ArchivesDirect](#).

Such metadata extraction capabilities would enable the packaging of WARC files for long term storage with PREMIS-conformant metadata manifests at the container as well as contained file levels, as evidenced by the web archive workflow and metadata output achieved with JHOVE and JHoNAS by the Bibliothèque nationale de France's [Scalable Preservation and Archiving Repository \(SPAR\) project](#) [18, 19].

## B. Storage environment

### 1. Trustworthiness

Towards their viability in the longest possible term, NYARC ultimately aspires to store preservation copies of its web archives within a storage environment that achieves the highest level of certification for the long term stewardship of digital content: TRAC-certified "Trustworthy Digital Repository (TDR)" status [20].

As of May 2015, neither the Amazon S3 nor Amazon Glacier storage environment described in this documentation has achieved that level of certification. DuraCloud, however, does enable the integration of the non-profit [San Diego Supercomputer Center's \(SDSC\) Chronopolis repository](#), which achieved TDR status in March 2012 [21], as an alternative redundant storage environment to Amazon Glacier in its 'Preservation Plus' service level.

The SDSC alternative is provided to DuraCloud partners at an additional annual subscription fee [22] not payable under the terms of the Andrew W. Mellon Foundation grant that initiated NYARC's web archiving program and which facilitated the selection of DuraCloud as a storage service provider. If and when operational funds beyond the period of the current (2015) grant may be appropriated for this purpose, NYARC should with DuraSpace's facilitation migrate the redundant preservation copies of its web archives from Amazon Glacier to SDSC storage.

## References

- [1] Web archiving. (n.d.). in *NYARC Documentation*. Retrieve from <http://wayback.archive-it.org/4432/20150413191710/https://sites.google.com/site/nyarc3/web-archiving>
- [2] Pearce-Moses, R. (2005). "Preservation copy." In *A Glossary of Archival and Records Terminology*. Society of American Archivists. Retrieve from <http://web.archive.org/web/20150522022824/http://www2.archivists.org/glossary/terms/p/preservation-copy>
- [3] Pearce-Moses, R. (2005). "Access copy." In *A Glossary of Archival and Records Terminology*. Society of American Archivists. Retrieve from <http://web.archive.org/web/20150522022918/http://www2.archivists.org/glossary/terms/a/access-copy>
- [4] De Stefano, P., et al. (2014). *Checking Your Digital Content, An NDSA Publication: What is Fixity, and When Should I Be Checking It?* National Digital Stewardship Alliance Standards & Practices and Infrastructure Working Groups. Retrieve from [http://web.archive.org/web/20150426233228/http://www.digitalpreservation.gov/ndsa/working\\_groups/documents/NDSA-Fixity-Guidance-Report-final100214.pdf](http://web.archive.org/web/20150426233228/http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA-Fixity-Guidance-Report-final100214.pdf)
- [5] DuraCloud. (2013). "DuraCloud." DuraSpace Organization. Retrieve from [http://web.archive.org/web/20150519205627/http://duraspace.org/about\\_duracloud](http://web.archive.org/web/20150519205627/http://duraspace.org/about_duracloud)
- [6] DuraCloud. (n.d.). "Subscription plans." DuraSpace Organization. Retrieve from <http://web.archive.org/web/20150520174909/http://www.duracloud.org/pricing>
- [7] Smith, C. (2015, May 12). Personal communication. Retrieve from <https://basecamp.com/2199695/projects/3605648/attachments/156040215/>
- [8] MD5.net (2015). Retrieve from: <http://web.archive.org/web/20150315021053/http://www.md5.net/>
- [9] MD5. (n.d.). In *Wikipedia*. Retrieve from <http://en.wikipedia.org/wiki/MD5>
- [10] Smith, C. (2012, May 31). DuraCloud brown bag series: Content integrity and health checks [Video file]. Retrieve from <https://www.youtube.com/watch?v=u8KeRZSqCck>
- [11] Smith, 2015.
- [12] University of Illinois at Urbana-Champaign, Grainger Engineering Library Information Center (2006, October 18). "ECHO Dep METS Profile for Web Site Captures." In *METS: Metadata Encoding & Transmission Standard, Official Web Site*. Retrieve from <http://web.archive.org/web/20130305105354/http://www.loc.gov/standards/mets/profiles/00000016.html>
- [13] Blumenthal, K. (2015, March 6). Preserving (and storing) web archives. Presentation to NYARC Directors Group. Retrieve from

[https://asset1.basecamp.com/2199695/projects/3605648/attachments/140053392/e6847a1f05ee30327b321abfa060e62a0010/original/Preservation-presentation-slides\\_2015-0306.pdf](https://asset1.basecamp.com/2199695/projects/3605648/attachments/140053392/e6847a1f05ee30327b321abfa060e62a0010/original/Preservation-presentation-slides_2015-0306.pdf)

[14] International Organization for Standardization (2009). ISO 28500:2009: Information and documentation -- the WARC file format.

[15] PREMIS Editorial Committee. (2012, July). *PREMIS Data Dictionary for Preservation Metadata, Version 2.2*. Library of Congress. Retrieve from <http://web.archive.org/web/20150522030313/http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>

[16] PREMIS Editorial Committee. (2015, April). Conformant Implementation of the PREMIS Data Dictionary. Library of Congress. Retrieve from <http://web.archive.org/web/20150522134353/http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf>

[17] Blumenthal, K. (2015, May 12). Preserving NYARC's web archives: A roadmap for long term stewardship. Presentation to NYARC Directors Group.

[18] Oury, C., & Peyrard, S. (2011). "From the World Wide Web to the digital library stacks: Preserving the French web archives." Paper presented at the 8th International Conference on the Preservation of Digital Objects (iPRES). Retrieve from <http://web.archive.org/web/20140912161221/http://halshs.archives-ouvertes.fr/docs/00/86/87/29/PDF/ipres2011-paper31-oury-peyrard.pdf>

[19] Caron, B. (2015, February 27). Personal communication. Retrieve from <https://basecamp.com/2199695/projects/3605648/attachments/156047527/>

[20] Consultative Committee for Space Data Systems. (2011, September). *Audit and Certification of Trustworthy Digital Repositories*. Recommendation for Space Data System Practices: CCSDS 652.0-M-1 (Magenta Book). Retrieve from <http://web.archive.org/web/20140403013436/http://public.ccsds.org/publications/archive/652x0m1.pdf>

[21] Center for Research Libraries (2012, March). Certification report on Chronopolis. Retrieve from [http://web.archive.org/web/20150522043248/http://www.crl.edu/sites/default/files/d6/attachments/pages/Chron\\_Report\\_2012\\_final\\_0.pdf](http://web.archive.org/web/20150522043248/http://www.crl.edu/sites/default/files/d6/attachments/pages/Chron_Report_2012_final_0.pdf)

[22] DuraCloud, n.d.