

# Digital Libraries Division Data Loss Escalation Procedures

**Date:** October 2015

**Version:** 1.0

**Contributors:**

**Mark Phillips** Assistant Dean for Digital Libraries

**Ana Krahmer** Supervisor, Digital Newspaper Unit

**Hannah Tarver** Department Head, Digital Projects Unit

**Daniel Alemneh** Supervisor, Digital Curation Unit

**Laura Waugh** Repository Librarian for Scholarly Works



This work is licensed under a Creative Commons Attribution 4.0 International License.

# Digital Libraries Division Data Loss Escalation Procedures

## Introduction

While data loss is a difficult subject in any repository environment, the UNT Libraries strives to make transparent information about data loss in either the Coda repository or the Aubrey access systems that act as underlying infrastructure for the UNT Libraries' Digital Collections.

This document is being created as a way of notating the procedures related to the escalation of events for a significant data loss event that adversely affects the UNT Libraries' Digital Collections.

For the purposes of this document we define data loss event as any occurrence that results in the loss of preservation or access digital objects, or the metadata related to these items.

## Definitions

We also define data loss into two categories, ***temporary data loss*** and ***permanent data loss***. Data loss can occur in the format of either loss of data or a corruption of data that renders a digital object or set of digital objects unusable.

**Temporary Data Loss:** Data loss that can be recovered either from a replica copy of the data, or from a traditional tape backup. Often times temporary data loss may not have adverse affects on the repository's designated community of users because of various redundancies in place.

**Permanent Data Loss:** Data loss that can not be recovered from in any traditional way. In this instance data is lost forever and will require regeneration. Usually this type of data loss will adversely affect the repository's designated community of users.

## Procedures

When a data loss of any kind has been identified the first step is to verify that the data loss is in fact a loss. This includes the verification from another device, and in most cases having another person verify that the data loss has occurred.

Once a data loss has been positively identified, the Assistant Dean for Digital Libraries is notified if there has been any issue related data corruption or loss in the system.

Options for how to handle the data loss is discussed with members of the Digital Libraries Division and a recovery plan is generated to attempt to correct the issue.

## **Recovery Plan**

A data recovery plan describes the data loss that occurred, when and how it was identified, how it was corroborated by a second person, and information about the plan to attempt to recover the data that was identified as missing or corrupt. This recovery plan includes the individuals who will work on the process, how we will proceed with the process of recovery. This recovery plan is distributed via email and includes as recipients the Assistant Dean for Digital Libraries, Head of the Digital Projects Unit, and the members of the DPU Software Development Team. It may be necessary to include the Libraries Network Manager or Director of Facilities and Systems Division in the email to keep them abreast of the situation.

## **Coda**

If data loss occurs in one of the Coda repository instances, the digital objects that are affected are identified and restored from the other Coda repository instance. The steps for recovering these items include the following:

- Validate digital object on Coda repository that is unaffected.
- Replicate digital object to Coda repository instance where it is missing.
- Perform a full validation of the item once replicated.
- Create a PREMIS Event for the validation and the successful replication in the PREMIS Event Service.
- Confirm that affected digital objects are complete and present.
- Alert those identified in the recovery plan of its completion.

## **Aubrey**

If the data loss occurs in the Aubrey access system then an inventory of the affected items is conducted and a process to re-populate the affected digital objects from the Archival Information Packages (AIPs) in the Coda repository begun. This workflow includes:

- Create a list of affected items in Aubrey
- Generate list of Coda identifiers that match the public access ARKs
- Retrieve digital objects from Coda
- Regenerate Dissemination Information Packages (DIPs) to be reloaded into the Aubrey access system.
- Verify that the items affected have been fully regenerated via the Aubrey user interface.

- Alert those identified in the recovery plan of its completion.

## **Metadata**

If the data loss occurs in the Metadata Storage Server, then the affected files are recovered either from the nightly rsync data or if that fails from the nightly tape backups made for the system. In this event it is likely that metadata edits that were made within the previous 24 hours would not be present. A list of the records that were edited during this period is available from the Metadata Edit Event Service. Once this list is generated, persons who edited metadata records can be notified as to which records will need to be re-edited.

## **Search Index**

If the data loss occurs in the Solr search index system that is not recoverable with the built-in replication and recovery procedures of Solr, the index will need to be rebuilt. This rebuilding will generally require a long indexing time so a notice to the end users of the Aubrey access systems will be put in place. During the index period it is possible to edit metadata records with the Edit system but all records may not show up as available even if they are. Once a full index has been completed and confirmed the notice to end users can be removed.

## **Extended Recovery Issues**

If the recovery plan is the recovery time will significantly impact services for more than 24 hours offered by the UNT Libraries, a message is added to the affected systems that communicates the service unavailability or degraded performance until the recovery can be completed.

If the recovery plan is either unsuccessful or if the recovery time period will significantly impact services for more than 72 hours, or if the data recovery is unsuccessful or considered a *permanent data loss* the following individuals are alerted by the Assistant Dean for Digital Libraries.

- Dean of Libraries
- Associate Dean of Libraries
- Director of Facilities and Systems
- Network Manager

If the data loss will adversely affect the public access to resources in the UNT Libraries Digital Collections a notice will be applied to appropriate systems within an expected period of time to remedy the problem and an email message will be sent to appropriate collection owners and maintainers in the UNT Libraries.

If the issue cannot be resolved within a seven day period a public message will be sent to External Partners with content managed by the UNT Libraries about the situation, steps being taken to remedy the problem, and expected timelines for issue to be resolved.

## **Notice of Recovery Plan Completion**

Once the data loss recovery plan has been fully executed and results in either a successful restoring of digital objects and services, or in an outcome of a *permanent data loss*, the Assistant Dean for Digital Libraries will communicate to those parties notified in previous steps an overview of the issue and steps that will be taken so that the same situation does not occur again.

This information will be documented on the Digital Projects Shared Storage Drive in a folder called "System Issues".

If the data loss occurred with the Coda repository system, a Major Event is logged with the Digital Libraries Division's Major Event Log application so that a record of the event is saved and associated with affected items in the Coda repository.