

Portico Disaster Recovery Plan

1. Goal

Portico's goal for disaster recovery is to ensure that all content currently preserved in the archive or destined for preservation will be recoverable from backup or an archive replica.

The only system with any external time sensitivity is the delivery interface for triggered content and content fulfilling post-cancellation access requests. There is no external need for any of the other systems to be recovered quickly.

2. The Archive

In general, there is no time sensitivity to recovering the archive and thus no concerns with the time it might take to establish a new data center or purchase new hardware. Per the Portico Replication and Backup Policy, the master copy of the archive is located in Princeton, NJ, a second on-line copy of the archive is located in Ann Arbor, MI, a third on-line copy of the archive is located in commercial cloud storage. There is one off-line replica of the archive in the Netherlands.

Due to the size of the content and the performance limitations of traditional tape backup, recovery of the archive involves reinstalling the software (the operating system, commercial off the shelf software such as Documentum and Oracle, and custom built software) and then ingesting content from a replica into the new, master archive. The plan outlined below details the recovery of the archive by ingesting content from the on-line replica located in Ann Arbor, MI.

The steps to recover the archive in case of disastrous failure in the Princeton data center are:

1. Rebuild hosts as per the shared IT disaster recovery plan – this may involve contracting with a new data center and purchasing new hardware
2. Ensure that the CVS repository (a version control system) has been recovered from the off-site backup tapes located on the host pr2ptcdev01.ithaka.org. The CVS repository is in the directory /export/home/cvs/cvsroot.
3. Build and deploy the Portico software as per the details in the Technical Operating Environment (TOE).
4. Install Documentum server v5.3 SP6 on the docbase host.
5. Configure a new docbase named Archive.
6. Install the docapp that is in the CVS repository at Portico\scripts\docapp\ArchiveDocApp.zip.
7. Create a new schema archive_md in the Oracle server.
8. Run the DDL script located in CVS at Portico/Database/scripts/ ArchiveMD_DDL_Scripts.sql.
9. Run the archive recovery application on the host running the archive management server.

3. The Content Storage Area

Content that has been received by Portico from the content providers, but which has not yet been processed through the content processing system and deposited in the archive resides in a file store.

The steps to recover the content storage area in case of disastrous failure in the Princeton data center are:

1. Rebuild hosts as per the shared IT disaster recovery plan – this may involve contracting with a new data center and purchasing new hardware.

2. Ensure the host pr2ptcthumper01.ithaka.org is accessible.
3. Restore the mount points /pcontent/ongoing, /pcontent/bulk_content and /pcontent/elsevier_content from off-site backup tapes.

4. The Content Processing System (ConPrep)

ConPrep is a pre-ingest system and has significant amounts of data flowing through it on a daily basis. Traditional tape-backup of this system is not feasible given the size of the content (4-6 TB) and the amount of change that takes place on a minute-by-minute basis. The strategy used in the recovery of this system involves reinstalling the software (the operating system, commercial off the shelf software such as Documentum and Oracle, and custom built software), restoring the database files from backup, and then using the information in the database to load content into the system from the content storage area. Note that we will not try to capture the work that might be lost, rather we will reload and rerun content that was in ConPrep at the time of the disaster.

The steps to recover ConPrep in case of disastrous failure in the Princeton data center are:

1. Rebuild hosts as per the shared IT disaster recovery plan – this may involve contracting with a new data center and purchasing new hardware.
2. Ensure that the CVS repository has been recovered from off-site backup tapes on the host pr2ptcdev01.ithaka.org. The CVS repository is in the directory /export/home/cvs/cvsroot.
3. Build and deploy Portico software as per the details in the TOE.
4. Install Documentum server v5.3 SP6 on the ConPrep docbase host.
5. Configure a new docbase named ConPrep.
6. Install the docapp that is in the CVS repository at Portico\scripts\docapp\ConprepDocApp.zip.
7. Create a new database schema named conprep_md on the Oracle server.
8. Restore the database from backup files.
9. Reload content into the conprep docbase based on information in the metadata database (conprep_md). We currently do not have an application that does this but it would be trivial since the original file names are recorded in the conprep metadata DB.

5. The Delivery Site

The delivery site is the one Portico system where the time to recovery matters, as it is the system used by faculty, staff, and students at participating institutions to access triggered and PCA content.

The steps to recover the delivery site in the case of disastrous failure in the Princeton data center are:

1. Rebuild hosts as per the shared IT disaster recovery plan – this may involve setting up the system at one of the alternate Ithaka data centers. If necessary, the Portico delivery site could share a machine with other Ithaka services.
2. Set up the A&A software:
 - a. Install and configure WebSphere, per the Portico delivery and audit installation documentation.
 - b. Install and configure IBM's Apache Web Server, per the Portico delivery and audit installation documentation.

- c. Install and configure the Ithaka A&A security application, per the Portico delivery and audit installation documentation.
 - d. Rebuild the A&A database.
3. Set up the delivery software:
- a. Install and configure WebSphere, per the Portico delivery and audit installation documentation.
 - b. Install and configure IBM's Apache Web Server, per the Portico delivery and audit installation documentation.
 - c. Install and configure the Ithaka A&A security application, per the Portico delivery and audit installation documentation.
 - d. Install and configure the Portico application.
 - e. Retrieve the triggered and PCA content from the Archive or replica and process it through the pipeline.

6. The Audit Interface

Recovery of the audit interface is not time sensitive. The steps to recover the audit interface in the case of disastrous failure in the Princeton data center are:

- 1. Rebuild hosts as per the shared IT disaster recovery plan – this may involve contracting with a new data center and purchasing new hardware.
- 2. Set up the A&A software:
 - a. Install and configure WebSphere, per the Portico delivery and audit installation documentation.
 - b. Install and configure IBM's Apache Web Server, per the Portico delivery and audit installation documentation.
 - c. Install and configure the Ithaka A&A security application, per the Portico delivery and audit installation documentation.
 - d. Rebuild the A&A database.
- 3. Set up the audit interface software:
 - e. Install and configure WebSphere, per the Portico delivery and audit installation documentation.
 - f. Install and configure IBM's Apache Web Server, per the Portico delivery and audit installation documentation.
 - g. Install and configure the Ithaka A&A security application, per the Portico delivery and audit installation documentation.
 - h. Install and configure the Portico application.
 - i. Retrieve the content from the Archive or replica and process it through the pipeline.

7. Document History

Version	Date	Change	Author
0.1	7/28/2009	Drafted.	Vinay Cheruku
1.0	7/28/2009	Polished and glossed. Added in the delivery site.	Amy Kirchhoff
1.1	8/5/2009	Tightened up the delivery bits.	Amy Kirchhoff

* An approved version of this document.