

CollectiveAccess System Back-up Guide

CollectiveAccess System Back-up Guide

Last revision: 21 December 2009

For version 0.6+

Copyright © 2009 Whirl-i-Gig. All rights reserved.

Table of Contents

1. About this document	1
Introduction	1
Audience	1
2. Backing-up CollectiveAccess collections data	2
Types of data	2
Back-up tools	2
Backing up your database	2
Backing up your digital media	3
Backing up your configuration files	3
Summary	3
3. Storage media	5
Storage media	5
Optical disks (CD's, DVD's and DVD-DL's)	5
Tapes	5
Harddrives	6
Back-ing up data on a hosted server	6
Further resources	7

Chapter 1. About this document

Introduction

Your collections data is more than just a list of items and digitized media. For the typical museum or archive a collections database is the product of countless hours of work by teams working in concert for years. It may well contain information available nowhere else. All too often, however, the long-term security of valuable data is given scant notice until something happens. It might be a fire, a storm or a theft. More likely it is a common hardware failure - a dead harddrive or a power surge. No matter the cause, all organizations absolutely require an effective disaster recovery plan for their collections databases.

This document describes techniques for the safe and effective back-up of your CollectiveAccess collections management system in a variety of situations. The storage requirements – and required back-up capacity - for a CollectiveAccess installation can vary widely depending upon the size of your database and the types of media you are managing. In particular, systems with significant volumes of audio and video (especially uncompressed audio and video) can quickly grow to require more storage than common back-up systems can provide. While many CollectiveAccess installations can be easily back-ed up on a single DVD, it is also not uncommon for systems to be measured in terrabytes.

We have attempted to suggest workable solutions for all systems, small, large and in-between. If you wish to suggest additions or corrections to this document please contact us at <support@CollectiveAccess.org>. If you have questions that are not adequately addressed by this document please post your queries on the CollectiveAccess online support forum at <http://www.CollectiveAccess.org/forum>.

Audience

This document is intended primarily for system or database administrators and information technology managers. However anyone who wishes to understand how CollectiveAccess stores data may find this document useful as well.

Chapter 2. Backing-up CollectiveAccess collections data

Types of data

CollectiveAccess stores the various types of data that make up your collections database in different locations. It is important that you understand how the locations are selected, and that any back-up system you implement can archive data from *all* of the locations. If your CollectiveAccess system is incompletely backed-up then it may prove impossible to restore the system to full function should recovery be necessary due to disaster, hardware failure, theft or other such circumstances.

When designing a data back-up procedure for your CollectiveAccess installation, there are three types of data that must be taken into consideration:

1. *The CollectiveAccess database* containing your collections metadata and database-specific configuration information. This includes all of your object cataloguing, authority information, controlled vocabularies, configured "pick-lists" (object types, relationship types, etc.), user login information, etc. This data is stored in a MySQL relational database and, as discussed in detail later, is most often backed-up using a combination of MySQL's system back-up tools and your chosen file back-up tool.
2. *Uploaded digital media such as images, video, audio and documents.* All uploaded media is stored outside of the MySQL database, in designated directories on your server. Files in these directories can be backed-up using your standard file back-up tools. However care must be taken to back-up *all* of the directories, otherwise media referenced by collections metadata will be missing upon system restore.
3. *Installation-specific "setup" and configuration files.* There are many configuration files that are used to customize the behavior of an CollectiveAccess installation. The structure of the configuration files has been designed such that most common settings are made in a single file - `setup.php`. However, to ensure that your system can be restored to function precisely as it was *all* configuration files should be routinely backed-up.

Back-up tools

All of the procedures outlined in this document assume that you're using a tool that backs up *files* - something that copies files on your system to some other archival medium, and can restore them from that medium as needed. This could be a program such as BRU (<http://www.tolisgroup.com/>), AMANDA (<http://www.amanda.org/>) or Retrospect (<http://www.emcinsignia.com/>) archiving your files to digital tape, a CD/DVD burning program such as Nero, Toast or CDRecord writing your files to optical disk or even a command-line Unix program such as rsync mirroring files from your server to another server or an external harddrive.

Backing up your database

To back-up your CollectiveAccess database you need to have MySQL "dump" it to a file and then have your back-up tool archive that file. MySQL comes with a command-line program called `mysqldump` that can create a complete snapshot of a database in a single file. The snapshot file will contain SQL commands to restore both the structure of the database and all of the data.

A typical command-line invocation of `mysqldump` would look something like this:

```
mysqldump -udb_login_name -pdb_login_password database_name > /path/  
to/dumpfile/my_database_backup.dump
```

where the `db_login_name`, `db_login_password` and `database_name` reflect, of course, the settings on your system, and `my_database_backup.dump` is the name of the newly created file. You can learn more about `mysqldump` and its various options and capabilities at <http://dev.mysql.com/doc/refman/5.1/en/mysqldump.html>.

You can automate the execution of `mysqldump` by adding an invocation to your crontab (on Unix-like systems) or equivalent on Windows. A more featureful solution is a MySQL backup automation script such as `AutoMySQLBackup` (free software available at <http://sourceforge.net/projects/automysqlbackup/>) which can take care of naming and compression of snapshots, and can easily handle multiple databases.

Remember that "dumping" your database to disk only makes a copy on your server. You still must run your back-up tool to copy the resulting file to your back-up media. One common way to accomplish this is to schedule a `mysqldump` run at a specified time daily, and schedule a run of your back-up tool to follow with sufficient time for the dump to complete. The time required by `mysqldump` is dependent upon the size of your database and the speed of your system, but even large databases on slow servers rarely take longer than 15 minutes, and 5 minutes or less is typical.

Also note that `mysqldump` locks tables as it dumps them. This may cause certain `CollectiveAccess` functions to become unresponsive as `mysqldump` runs. If you have a large database this may have the effect of making the system unusable for several minutes. You should evaluate the effect of a running `mysqldump` process on your database before running it while users are logged in. Most users choose to run `mysqldump` daily at an hour when users are typically not logged in. For systems that cannot tolerate any period of unresponsiveness other backup possibilities are available. See <http://dev.mysql.com/doc/refman/5.1/en/backup-methods.html> for more information.

Backing up your digital media

As shipped `CollectiveAccess` stores all uploaded and derived digital media in a series of sub-directories under `media` in the root of your installation. Simply back-ing up the entire contents of `media` is sufficient in most cases.

Within `media` media is separated into a series of sub-folders based upon file type. This makes it easy, if needed, to serve specific media types (Flash Video, `WindowsMedia` or `RealMedia` video, for example) using proprietary server software. You can reconfigure where `CollectiveAccess` stores various types of media by modifying the directory paths in the `media_volumes.conf` configuration file. So, for example, if you wanted to have `CollectiveAccess` store all uploaded or derived Flash Video files in the root directory of a Flash Media Server installation, rather than in `media`, you would simply change the path in `media_volumes.conf` to point to the Flash Media Server. *It is important that the settings for your file back-up take into account any such modifications.*

Backing up your configuration files

As shipped all `CollectiveAccess` configuration files are stored in the non-web accessible ("home") directory in a sub-directory named `conf` under `app` (aka. `app/conf`). This entire directory should be back-ed up. Your `setup.php` file, located in the root of your installation and which contains some basic configuration information such as the locations of the application configuration file, should also be back-ed up.

Summary

For a typical `CollectiveAccess` installation where media and configuration files are stored in the typical (and pre-configured) locations, and assuming that you are writing database snapshots into a "dumps" directory in a location outside of the web server root, you should be backing-up, at a minimum, the following directories:

- `/path/to/mysql/dumps`

Backing-up CollectiveAccess collections data

- `/path/to/collectiveaccess/app/conf`
- `/path/to/collectiveaccess/media`
- `/path/to/collectiveaccess/setup.php`

Depending upon the setup and size of your CollectiveAccess installation, server and back-up system you may elect to simply back-up the entire CollectiveAccess directory structure including the application code and supporting directories, rather than specifically selecting the directories above. This has the advantage of providing a complete ready-to-run backup and is the preferred option if it is possible. If you cannot do this, you can always download the CollectiveAccess application code at CollectiveAccess.org.

Chapter 3. Storage media

The storage medium for your CollectiveAccess system back-ups will be largely determined by a few key factors:

- *The amount of storage required.* The overall storage requirements of an CollectiveAccess installation are largely determined by the size of the digital media being stored. Storage for the database and application files are almost always dwarfed by that of digital media. The overall storage needed to complete a back-up is often used as the first criteria for selecting a medium. For example, if your CollectiveAccess installation takes up 50 gigabytes of storage (not uncommon for a medium-size image database or a small video database) then CD media, which stores only 0.7 gigabytes per disk is clearly inappropriate. Each full back-up would require 72 disks! DVD media, which stores 4.3 gigabytes on a disk would require 12 disks - better, but still somewhat impractical. Dual-layer DVDs, which store nearly 8 gigabytes on a disk, would be tenable but inconvenient. LTO and DLT digital tape drives which, depending upon model, store upwards of 400 gigabytes per tape may be a better choice.
- *The longevity and stability of the medium.* How long the medium employed can be expected to survive is a key consideration and still the subject of much research and debate. Studies have found that high-quality CDs and DVDs can survive if properly cared for for decades. Other studies have found that poor quality optical media may last for as little as 18 months. You should do research on the quality and reliability of the medium you select, and follow the preservation guidelines set forth by authorities such as the US National Institute for Standards and Technology (NIST). You should also establish a long-term media migration policy, since periodically migrating digital from old media to new is an accepted requirement of long-term stewardship.
- *The equipment that you already have.* Your choice of medium will be influenced by any equipment you may already own. Since many back-up devices, particularly high capacity devices such as digital tape drives, are expensive often using what you already have will override other considerations.
- *Cost.* The initial cost of equipment is important, but that of consumable media is often even more so. With commodity media such as DVDs, even high-quality media is cheap, with a typical cost of less than \$1.00 per disk. However, for some tape drives the cost per-tape is well over \$100. Of course, if you need 100 DVDs to complete a backup that 1 tape can handle this may not be such a bad deal!

This list is not intended to be a comprehensive discussion of data back-up strategies and considerations (that could fill a book). You should consult with your local disaster recovery advisor when formulating your own plan.

Storage media

What follows is a brief overview of common back-up mediums and their applicability to various types of CollectiveAccess installations. This is not a comprehensive treatment and should only be used a starting point in the development of a data back-up strategy.

Optical disks (CD's, DVD's and DVD-DL's)

Optical disks have the great advantage of being cheap and ubiquitous. CD media, which only stores about 0.7 gigabytes per disk, is generally impractical for CollectiveAccess backups which tend to quickly grow larger than that capacity. DVD media, which stores about 4.3 gigabytes per disk, may be a useful backup medium for small systems and properly handled can have a reasonable lifespan. Dual-layer DVDs (DVD-DL) with their larger capacity can be a more convenient alternative to standard DVD media but their longevity is unproven.

Tapes

LTO and DLT tape drives are high-speed high-capacity back-up solutions suitable for larger CollectiveAccess installations. The latest iterations of these technologies can store 800 gigabytes of

data on a single tape, and planned enhancements will increase the capacity to 3.2 terabytes in the near-future. Both the drives and tapes can be expensive however, with drive prices of \$2000 or more and tape prices of \$100 or more per tape.

Harddrives

Standard harddrive mechanisms can be a cost-effective alternative to costly tape drives. With data densities of a terabyte or more per drive and prices of under \$300 per drive organizations with large installations and tight budgets may find that the purchase of several drives for rotating back-ups (with at least one drive stored off site at all times, of course) is a viable option.

Back-ing up data on a hosted server

If you are running CollectiveAccess on a hosted or shared server you likely won't have the option of plugging a tape drive or DVD into the machine for back-up. For installations of a relatively small size (< 20 gigabytes) it may be practical to simply mirror the CollectiveAccess files on your hosted server to a local machine over the Internet using a file transfer program or Unix utility such as `rsync`. You can then back-up the files on your local machine using standard tools. You may also want to investigate the data back-up services your hosting company provides. Many hosts bundle data back-up services with their higher-level service plans.

For larger installations where mirroring over the Internet may not be practical due to bandwidth constraints, you should consult your service provider about back-up services they may provide.

Further resources

MySQL reference page for database backups: <http://dev.mysql.com/doc/refman/5.1/en/backup.html>

MySQL reference page for mysqldump: <http://dev.mysql.com/doc/refman/5.1/en/mysqldump.html>

AutoMySQLBackup - a script for automating MySQL database backups: <http://sourceforge.net/projects/automysqlbackup/>

Care and Handling Guide for the Preservation of CDs and DVDs. NIST Special Publication 500-252, October 2003. <http://www.itl.nist.gov/iad/894.05/papers/CDandDVDCareandHandlingGuide.pdf>

Care and Handling Guide for the Preservation of CDs and DVDs (Quick One-Page Reference.) NIST Special Publication 500-252, October 2003. <http://www.itl.nist.gov/iad/894.05/papers/onepage.pdf>

Digital Preservation. North Carolina ECHO, 2007 Revised Edition. <http://www.ncecho.org/Guide/preservation.htm>