# DRS User Manual for Data Loading

**Version 5.004**

The Digital Repository Service (DRS) provides Harvard-affiliated owners of digital material with a storage and retrieval system for their collections.

This manual is designed to assist deposit agents with the batch deposit of materials to DRS. Included are procedures for batch depositing, descriptions of the data loading process, and descriptions of the base elements of the DRS batch DTD.

Supporting information about the batch DTD (sample batch.xml files, supplemental metadata dictionaries, etc.) are available in the DRS Documentation Center section of the OIS web site.

For more detailed descriptions of DRS services and policies regarding DRS use, consult the DRS section of the OIS web site.

**Need help with DRS deposits?**

- Consult support information on the Depositor iSite:

  http://isites.harvard.edu/icb/icb.do?keyword=k26186 (Harvard-only access)

- To report a problem or ask a question about DRS deposits, please use the DRS feedback form:

  http://nrs.harvard.edu/urn-3:hul.ois:drshelp

  If reporting a problem, please describe the activity leading up to the problem and any error messages you receive.

## Versions and Revision history

**12/1/09    Version 5.004**  Replaced links to old documentation, corrected some links

**7/16/09    Version 5.003**  Added PDF to the list of accepted DRS formats and Mimetypes

**3/6/09     Version 5.002**.  Removed references to unacceptable JPEG 2000 filename extensions, .jpx and .jpf.

**4/8/08     Version 5.0.**  Added description of new usageClass and new successMethod attribute and updated links to correspond with new OIS website release and new DRS storage architecture changes.

**3/21/07    Version 4.008.**  Changed hours that batch processing occurs

**8/1/06     Version 4.007.**  Changed max characters in batch directory name from 32 to 100

**7/11/05    Version 4.006.**  Republished in HTML and PDF versions.  Minor reorganization of contents for delivery as HTML.  Removed Chapter 4 section on space planning.  Added details to Chapter 5 section on correcting data already deposited.

**6/16/05**     Added IS_PRESERVATION_REPLACEMENT_OF relationship

**6/10/05**     Added mime type application/zip

**6/08/05**     Added new batch report summary

**6/01/05**     Revised for multiple batches per drop box loading enhancement

**1/11/05**     Corrected error in <access> element Required Attribute(s) section (p23)

**10/12/04**    Corrected error in drop box address (p8)

             Changes to DRS batch report format (p9)

             Changes to the DRS DTD: local file name now retained in DRS but not searchable; updates to these elements: <relationship>, <fileFormat>, <access>, <mimetype>, and <metadata>.

**7/1/2004**    Revised all FTP references to "SFTP". DRS batch deposit now requires a secure FTP client.

             Changed the SFTP drop box address to ldidropbox.hul.harvard.edu

**3/23/2004**   Updated JPEG2000 mimetypes

             Updated imagemetadata supplement address to: http://preserve.harvard.edu/resources/imagemetadata.pdf

             Updated Appendix: M5 and XML resources links

**10/07/2002**  Updated DRS XML validator address to: http://drstest.harvard.edu/cgi-bin/drs_validate.pl

# 1.0  Registration of Owners and Depositors

In order to deposit digital materials to the DRS, the Harvard organizational owner of the materials must be registered as a DRS object owner and the agent responsible for depositing these materials must be registered as a DRS batch depositor.

### Registration for DRS owners

The Harvard organization with financial and curatorial control over objects being prepared for deposit is called the object owner.  Object owners using the DRS for the first time must register as a digital object owner.  To register, submit the online registration form available from the OIS web site:

http://hul.harvard.edu/ois/systems/drs/f-drs-owner-reg.html

Owner registration is a one-time process.  More information about the registration process appears in the Digital object owner FAQ on the OIS web site.

Consult the List of Registered Owners for a current list of Harvard organizations that have registered as DRS object owners.

## Registration for deposit agents

A deposit agent is an individual or organization authorized to deposit batches of digital objects into the DRS.  An agent may be a reformatting/digitization vendor that deposits objects on behalf of the Harvard organization that owns the objects.  Or, an agent may be an individual within the owning organization.

New deposit agents must register and prepare for their first deposit by following the steps outlined in How to become a DRS deposit agent.  Upon registration, the deposit agent will receive an SFTP DRS drop box account and DRS loading instructions.

**Related topic:**

Maintenance of DRS Data

# 2.0 The Batch Deposit Process

This section describes the DRS batch deposit process, including SFTP drop boxes and batch directories as well as the actual deposit procedure.  To find out more about how the DRS processes batch deposits, see The Data Loading Process.

DRS consists of an Oracle database and a UNIX file system.  Digital objects are stored in the DRS using an OIS load program that reads instructions from a batch control file formatted in XML and named 'batch.xml'.  The file system provides physical storage for the objects while the database tracks file locations on disk, management information about the objects, and object-specific metadata.

Before deposits can begin, both the object owner and depositing agent must be registered with the DRS.  To make deposits, the depositing agent will transfer objects to a batch directory within a DRS SFTP drop box account.  Accompanying these objects will be a batch control file (batch.xml) formatted according to the DRS batch DTD.

For more information about the batch deposit process, see:

What is a deposit to the DRS?
About the SFTP drop box
About batch directories
Performing a batch deposit
Controlling loading order of batches
Best practices for depositors and owners
DRS deposit tools and resources

See also information about the Data Loading Process:

Checking batch queue status
Batch loader reports

## 2.1  What is a deposit to the DRS?

Digital objects come in two varieties – objects originally in digital form ("born digital") or reformatted objects (changed from analog to digital).  Both kinds of electronic objects are loaded into the DRS in the same way.

A deposit includes a batch directory containing one or more digital objects and a batch transaction file.  The batch control file (written in XML and called 'batch.xml') describes the objects and defines these actions:

- Add an object

- Create the URN (Uniform Resource Name) for this object

- Create a relationship between objects

## 2.2  About the SFTP drop box

Batch deposits are sent to OIS over the Harvard campus network.  Each depositing unit will be issued an SFTP drop box on an OIS secure server.  The depositor transfers the material by SFTP to one or more batch directories within his or her drop box. Depositing units may be issued multiple drop boxes to improve workflow.

The drop box issued to a depositing unit will contain the following directories: `usr`, `lib`, `incoming`, `etc` and `bin`. Deposits will be made by transferring batch directories and their contents to the `incoming` directory.

**Note:** Do not put deposit files directly into the `incoming` directory. The `batch.xml` file and digital object files should be within a sub-directory that is under the `incoming` directory. See Batch directories for more information.
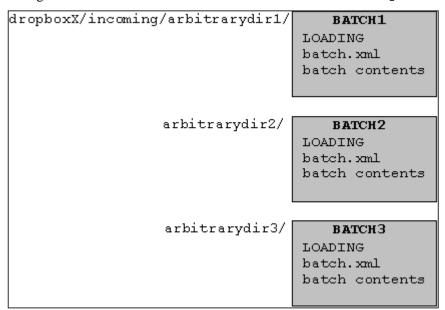
Closing a connection to an SFTP drop box will trigger the queuing of a batch if the drop box contains any batch directories containing a batch control file (`batch.xml`). A file named `LOADING` is automatically placed in a batch directory when an SFTP connection is closed and the batch directory contains a `batch.xml` file. The presence of the `LOADING` file is what triggers a batch to be put into the loading queue. Please do not transfer the `batch.xml` file to a batch directory until the batch is ready to be loaded and do not try to create or delete the `LOADING` file yourself.

SFTP connections can still be made to drop boxes while batches are processing so that additional batches can be dropped off. Please refrain from doing anything with batch directories that have already been fully transferred to a drop box to avoid conflicts with the DRS loading application.

## 2.3  About batch directories

All batch deposits to the DRS must be made within batch directories. A batch directory is a properly-named sub-directory within the incoming directory of a drop box. All batch contents, including the batch control file (`batch.xml`), must be located in a batch directory.

The figure below illustrates three batch directories within the `dropboxX` drop box.



In this example a depositor has transferred 3 batch directories (**arbitrarydir1**, **arbitrarydir2**, and **arbitrarydir3**) to a drop box called dropboxX.

The presence of a **LOADING** file indicates that these batches are already in the loading queue.

Batch directories can contain any number of sub-directories which in turn can contain any number of sub-directories.  Everything within a single batch directory is considered part of the same batch.

Batch directories can be named according to depositor preference as long as:

- There is no other directory with the same name in the incoming directory of that drop box.

- The batch directory name is less than 101 characters.

- The batch directory name contains only letters, digits, underscores ('_'), and hyphens ('-').

The Table below lists examples of valid and invalid batch directory names.

| Examples of VALID batch directory names | Examples of INVALID batch directory names |
| --- | --- |
| batch<br>1batch<br>batch1<br>_batch<br>-batch<br>batch_1<br>batch_2005_06_01<br>batch_2005_06_01-1<br>20050601_150502-ag | batch directory<br>batch!<br>batch@2005<br>a-batch-directory-name-which-exceeds-the-character-length-this-length-can-be-no-longer-than-100-characters |

A batch directory name must be less than 101 characters and only contain letters, digits, underscores ('_'), and hyphens ('-'). Note that the name can start with any of these valid characters but it is best practice not to start with a hyphen because of the impact on file name sorting.

## 2.4  Performing a batch deposit

Batch directory deposits to DRS can occur anytime to a specific SFTP drop box, even when a **different** batch directory load is already in progress for that drop box.  Because there is no locking mechanism on drop boxes or batch directories, care should be taken not to interfere with batches already in the queue.  If you see a LOADING file in the batch directory, loading of the batch is in progress.

In addition to the steps below, depositing agents should also consider information in the Best practice section of this manual.

DRS batch deposit requires a secure FTP client.  Information about SFTP client options is available on the DRS Documentation Center page.

The SFTP drop box address for DRS Production deposits is:  **drsrop.hul.harvard.edu**

The SFTP drop box address for DRS QA deposits is:  **drsrop-qa.hul.harvard.edu**

To deposit one or more batches, follow these steps:

1. Open an SFTP connection to **drsrop.hul.harvard.edu** (for production deposits) or to **drsrop-qa.hul.harvard.edu** (for QA deposits).

2. Transfer a uniquely-named batch directory of data to the incoming directory.  The batch directory can be named according to depositor preference as long as it meets the naming constraints described in About batch directories.

3. If you are ready to queue the batch, transfer the `batch.xml` file to the batch directory. If you are not ready to queue the batch, skip this step.

4. Close the SFTP connection. The act of closing the connection queues any batch directories with a `batch.xml` file. Batches will not be queued until the SFTP session is closed. Opening the SFTP session again is fine, as well as maintaining multiple SFTP sessions to the same drop box.

5. If you want to send another batch, go to step 1 and repeat these steps.

6. Once a batch is processed, the DRS loader will send an email message that reports the success or failure of the batch. See The data loading process for more information.

## 2.5 Controlling loading order of batches

Sometimes a batch is dependent on the contents of another batch and you want to make sure one loads before the other. There are two ways to control the queue order of your batches.

- **Using the SFTP client.** Follow deposit steps 1-4 described in Performing a batch deposit for the batch you want loaded first. Make sure that you close the SFTP connection to queue up the first batch. Then follow the same 4 steps again for the batch you want loaded next. By closing the SFTP connection between SFTPing batches, you will ensure that the batch SFTPed first will get loaded first.

- **Using batch directory names.** The second way to control the loading order of batches is by the names of batch directories. To utilize this method, you SFTP multiple batches in the same SFTP session. Name your batch directories so that the names of any batches you want loaded earlier would come earlier in an ASCII alphabetical sort order. That is: hyphens ('-'), then digits, then capital letters, then underscore ('_'), then lower-case letters.

  The table below shows the sorting order of 14 batch directories that had been transferred to a single drop box in a single SFTP session. When you close the SFTP connection, batches will be queued in the sort order of the batch directory names.

| Queue order sequence | Batch directory name |
| --- | --- |
| 1 | -1test |
| 2 | -_ |
| 3 | -adir |
| 4 | 123 |
| 5 | 1234 |
| 6 | 124 |
| 7 | Adir |
| 8 | Bdir |
| 9 | _1test |
| 10 | _1test1 |
| 11 | _1test1_ |
| 12 | _test |
| 13 | adirectory |
| 14 | bdirectory |

  Note that batches from other drop boxes might be queued between these batches but this order sequence would be maintained.

## *2.6  Best practices for depositors and owners*

- An MD5 signature is **required** for each object to ensure that the object has been successfully transmitted to DRS.  MD5 tool information is available on the DRS Documentation Center page.

- Object owners are **strongly encouraged** to retain the object on local servers until they have received notification by the DRS that the object has been successfully deposited and quality assurance procedures have been completed.

- Object owners are **strongly encouraged** to maintain a link between the local system and the DRS.  DRS permits users to associate local information (such as an accession number) with an object via the owner supplied identifier field.  Please see the Identifying deposited objects section of this manual.

## *2.7  DRS deposit tools and resources*

Consult the DRS Documentation Center page on the OIS web site for information and links on topics related to DRS batch deposits, including:

- the DRS batch DTD,

- sample `batch.xml` files,

- type-specific metadata supplements, and

- suggestions for xml validators, MD5 checksum utilities, and SFTP sources.

# 3.0 The Data Loading Process

Once a batch is deposited and the DRS loader detects the presence of a `batch.xml` file, the data loading process starts.

**Note:** While batch deposits can occur at any time, the DRS batch loading service processes batches only during business hours (**Monday to Saturday, 8am-8pm**).  Batches deposited after these hours will be processed the next business day.

The DRS loading process polls the SFTP drop boxes every few minutes for queued batch directories during times that the loader is running.  The data loading process includes these steps:

- All queued batch directories are sorted according to when they were put on the queue (when the `LOADING` file was put in the batch directory - corresponding to when the SFTP connection was closed and the sort order of batch directory names).

- Batches are processed and digital objects are deposited according to instructions provided in the `batch.xml` file.

  To check the status of your batch, consult the Batch queue status page.

- When a batch is completed, the loader sends an email report of the results.  See Batch loader reports for more information.

- After a successful load all files that have been loaded into the DRS will be deleted from the batch directory, along with the `batch.xml` file, the `LOADING` file, and any empty directories. Objects in the drop box that were not part of the load will not be removed. If the batch directory becomes empty after deleting these files, the batch directory will be deleted.

Note that depositors can specify file names and file name patterns in the `batch.xml` file that the DRS loader should delete after a successful load even though these files were not loaded into the DRS. See the **remove** attribute of the [<batch>](#) element for details.

After a failed load, all the digital objects associated with an error batch are left in the depositor's batch directory, while the `batch.xml` file and the `LOADING` file are deleted.

The depositor is responsible for all data until the DRS has sent a confirmation report.  Best practice requires depositors to keep a copy of all data until the project quality control is complete.

**Related topics:**

[Checking batch queue status](#)
[Batch loader reports](#)

## 3.1  *Checking batch queue status*

Depending on the size of the deposited data, the load may take multiple hours to process and report.  The status of batches in the queue can be seen at:

- For test drop
  boxes: [http://drstest.harvard.edu:9011/drs/servlet/WebAdminService?page=view_queue](http://drstest.harvard.edu:9011/drs/servlet/WebAdminService?page=view_queue)

- For production drop
  boxes: [http://drs.harvard.edu:8080/drs/servlet/WebAdminService?page=view_queue](http://drs.harvard.edu:8080/drs/servlet/WebAdminService?page=view_queue)

Although these batch queue status pages are part of the DRS Web Admin system, access to them requires only a valid Harvard ID (Web Admin registration is not required).

## 3.2  *Batch loader reports*

The DRS batch loader sends out an email message that reports on the success (or failure) of a processed batch.  Recipients of this message are identified within the batch.xml file.

For more information, see:

[Successful load report](#)
[Failed load report](#)

### 3.2.1  Successful load report

When a batch has been successfully processed, the DRS loader sends a confirmation report to the email addresses specified in the [<emailSuccess>](#) element of the `batch.xml` file.  Email contacts for a successful load should include the depositor and the digitizing project leader/manager.

It is important to save these reports, since the information included (DRS object IDs, owner-supplied identifiers, URNs) provides a link between deposited objects in DRS and your local information about these objects.

The successful load report has the following format:

```
Subject: DRS LOAD REPORT (owner:<owner>, batch
dir:<batch_directory_name>, batch:<batch_name>
[DB:<batch_id>])

Batch Summary
Digital Objects Added
Relationships Added
URNs Requested

Email Attachment (drsbatch_<batch_id>.txt):
     Digital Objects Added
     Relationships Added
     URNs Requested
```

Click here to view a sample successful load report.

The **Subject line** of the email message contains basic information about the batch.  The body of the report contains a **batch summary** followed by three sections that describe the **digital objects added**, **relationships added**, and **URNs requested**.  These same three sections are also included in a tab-delimited text file attached to the email message.

Within the **Subject line** of the report:

- owner is the DRS owner code, specified in the <owner> element of the batch.xml file.

- batch directory name is the name of the directory to which the batch was deposited.

- batch name is the name of the batch provided by the depositor in the <batch> element of the batch.xml file.

- batch id is an internal identifier generated by the DRS for each deposited batch.

The **Batch Summary section**, which is only in the body of the email, includes the following information about the batch:

- Batch directory name

- Batch name

- Batch id

- Owner(s)

- Batch drop off time

- Time waiting to start load

- Loading start time

- Loading end time

- Total load time

- Number of transactions

- Number of files deposited

- Batch size

- Number of files per mime type

The **digital object section** of the report lists information for each object deposited. The data in this section is tab-delimited with the above descriptors listed at the top:

- Filename

- Owner-supplied name

- URN (If no URN was requested in the current batch, a null value is returned here)

- Object ID

- Mimetype

- File Size

- Insertion Date

- Role

- Purpose

- Quality

- Owner

- Access Flag

- Usage Class

- MD5 Signature

The **relationship section** of the report lists the following information for each relationship added:

- DRS object id for 1st object

- Relationship

- DRS object id for 2nd object

- Owner

The **URNs requested section** of the report lists any URNs requested for objects outside the current batch. (The batch loader allows you to request a URN for an object already in the DRS.) In most cases, objects will be named during the deposit of that digital object. In that case, the URN is listed above in the digital object section. For URNs requested for objects outside the current batch, the report provides this information:

- DRS object id

- URN

- Owner

## 3.2.2 Failed load report

If any errors are detected during a load, the entire batch is rejected and an error report is sent to the email addresses in the <emailFailure> element of the batch.xml document. The error email will contain the name of the batch directory that failed along with a message describing the cause of the failure.

Here is a sample failure report:

To: xxxxx@fas.harvard.edu
To: drs-support@xxxxx.harvard.edu
Subject: DRS: Error processing current batch

There was an input data error while processing
your DRS batch.  The data files are still in
your batch directory and the batch.xml file has
been removed.  Please see the error text below.
Correct and upload the new batch.xml and
problematic object files (if any exist) to
restart the loading process.

For more information, contact drs-support@xxxxx.harvard.edu
Error Text:
--------------------------------------------------

Drop Box: fal6ftp
Batch Directory: NewacqAD309
Batch Name: NewacqAD309_3_10_2004
Context: validate
Transaction #: unknown
Top-Level Message: Following files not found:
U556253_1_smdl.jpg
U556254_1_smdl.jpg
U556253_1_prdwork.jpg
U531317_1.tif
U531315_1_smdl.jpg
U531315_1_prdwork.jpg
U556254_1_lgdl.jpg
U556254_1.tif

Embedded Exception Type: none

**Note:** If the failure is caused by errors in the header of the batch.xml file, processing may fail
before the loader can read the failure email addresses.  In this case, the loader will report the error
to DRS staff in OIS only.  If you submit a batch that does not go through and no email report is
received, contact DRS staff by submitting a DRS support request at: http://nrs.harvard.edu/urn-
3:hul.ois:drshelp.

All the digital objects associated with an error batch are left in the depositor's batch directory, while
the batch.xml file and the LOADING file are deleted.  To retry the batch, update any corrupt or
missing digital objects and upload the new batch.xml file.  After you disconnect from the SFTP
session, your batch will be queued for reloading.

# 4.0 Identifying and Relating Objects

This section describes the options supported by DRS for identifying deposited objects and defining relationships between deposited objects.  The process of assigning identifiers and defining relationships occurs when the batch is processed, based on instructions in the `batch.xml` file.

For more information, see:

Identifying deposited objects
Assigning a Uniform Resource Name (URN)
Defining relationships between objects

## *4.1  Identifying deposited objects*

There are three ways to identify an object in the DRS: DRS object id, a URN, and through owner-supplied metadata.  All of these identifiers should be tracked from the successful load report sent to the depositor.

### *DRS object ID*

The DRS object id is a numerical value generated automatically for every digital object deposited.  **Every** object in the DRS has a unique object ID.

### *URN*

A URN (Uniform Resource Name) can be requested for any object in the DRS.  URNs are persistent, location-independent identifiers for network-accessible resources.  The key value of a URN is its persistence: an object can be found by its URN even if its file name or physical location changes.  A URN is used just like a URL.

A URN is **required** for objects that will be delivered out of the DRS (for example, image files delivered to users of the VIA union catalog).  A URN is optional for objects intended for storage only (for example, archival versions of objects).

Consult the Assigning a Uniform Resource Name (URN) section for information about URN assignment options.

### *Owner-Supplied Name and Role/Purpose/Quality*

Each object stored in the DRS must be accompanied by an owner-supplied name. This name serves as a unique identifier that links deposited objects with local information about those objects.  The depositor must specify this name in the <ownerSuppliedName> element of the `batch.xml` file for each digital object they are depositing.  The name must be unique within that owner's collection in the DRS, although there is one exception.

More than one digital object may have the same owner-supplied name if the role/purpose/quality values are different.  It is common for many versions of the same logical object to exist in the DRS. For example, there may be a high-resolution archival version and a low quality deliverable of the same image.  One method for retaining this information is to have the same owner-supplied name for both images, but use different role/purpose/quality values to capture the differences between the physical objects.

This use of role/purpose/quality is **optional** for owners/depositors.  If the owner-supplied name is unique, there is no need to use role/purpose/quality at all.  If you are going to use these descriptors,

please consult the appropriate metadata supplement for best practice.  Metadata supplements are available from the DRS Documentation Center on the OIS web site.

## *4.2  Assigning a Uniform Resource Name (URN)*

A Uniform Resource Names (URN) is a persistent, location-independent identifier for a network-accessible resource.  The key value of a URN is its persistence: an object can be found by its URN even if its file name or physical location changes.  A URN is used just like a URL.

A URN can be specified for any object in the DRS.  A URN is **required** for objects that will be delivered out of the DRS (for example, image files delivered to users of the VIA union catalog).  A URN is optional for objects intended for storage only (for example, archival versions of objects).

For more information, see any of these topics:

URN syntax
URN assignment options

## 4.2.1 URN syntax

In the DRS a URN has the form:              urn-3:<authority-path>:<resource-name>
Example:                                     urn-3:FHCL:123456

where "urn-3" is the namespace identifier which indicates that the name is part of Harvard's NRS namespace, <authority-path> is the authority path, which identifies the Harvard organizational unit responsible for the name, and <resource-name> identifies the named object.  The <resource-name> portion must be unique relative to the specified <authority-path>.  A URN in this form (urn-3:FHCL:123456) is not actionable unless it is embedded within a URL.

Here is an example of an actionable URN as found in a catalog:

http://nrs.harvard.edu/urn-3:FHCL:123456

The domain name "nrs.harvard.edu" refers to the name resolution server for Harvard's NRS namespace.  URN assignments are supplied by the <urnMap> element in the batch.xml file that accompanies a deposit.

## 4.2.2 URN assignment options

To assign a URN to deposited objects, the depositing agent must have the appropriate authority path and a decision about style of resource-name assignment.  The object owner is responsible for providing the authority path.  The style of resource-name will be determined by the owner in consultation with depositing agent.  There are two options for resource-name style: request a specific URN or request that a URN be generated by DRS.

### *Requesting a specific URN*

Requesting a specific URN means to fully specify the URN as a literal string that will be assigned to the object in the deposit process.  In the DRS DTD, use the <urn> element to specify the URN. The literal string will include "urn-3" namespace identifier followed by the appropriate authority path and a unique local identifier (often an accession number) serving as the resource-name.  The resource name must be unique relative to the chosen authority path.

In the following sample URN:

urn-3:FHCL:ms12345

"FHCL" is the authority path for Harvard College Library and "ms12345" is the resource-name (in this case, the object's accession number).

### *Requesting that a URN be generated*

To request that a URN be generated for an object means specifying a URN that is a combination of literal values and auto-generated values. In the DRS DTD, use the <u>&lt;urnmask&gt;</u> element to request that a URN be generated.

The URN mask has components that are expanded by the DRS load program when a name is created, at which time the components are replaced in the URN with date and time stamps.

These components may be added anywhere in the string, and may be repeated, if desired. They are enclosed in braces ('{' , '}') to distinguish them from other parts of the URN string. This means that brace characters **may not** be part of a name (brace characters are not part of the valid NRS name character set in any event). The components are:

| Component | Replacement values | Meaning |
|---|---|---|
| {dd} | 01 – 31 | Current day of the month |
| {hh24} | 01 – 24 | Current hour, 24 hour clock |
| {mm} | 00 to 59 | Current minute of hour |
| {mo} | 01 – 12 | Current month |
| {n} | 0 – 1027 | Unique integer value |
| {ss} | 00 – 59 | Current second |
| {yyyy} | 1999 – 9999 | Current year |

Note that using the URN mask does not guarantee that the string created will be unique. Submitting URN masks that generate non-unique values will result in an error, and the rejection of the request to generate a name. To guarantee the generation of a unique name, include the unique integer value component, '{n}', in the mask.

Examples:

```
URN MASK                              CREATED VALUE
urn-3:dig:{n}                         urn-3:dig:75
urn-3:FAL:{yyyy}-{n}                  urn-3:FAL:1999-76
urn-3:HCL-DIG:{yyyy}{mo}{dd}          urn-3:HCL-DIG:20000103 (unique only for one
                                      object on any given day)
urn-3:HCL-DIG:{yyyy}{mo}{dd}{n}       urn-3:HCL-DIG:200001032 (always unique)
```

## *4.3  Defining relationships between objects*

The DRS provides a flexible and powerful mechanism for defining relationships between objects that have been placed into repository storage. The relationships may be one to one, many to one, or many to many.

The syntax for defining the relationship is as easy as the syntax of a sentence in English. The pattern follows the form <subject><verb><object>. The following pseudo-code constructions represent some possible object relationships:

• File 123          is derived from  File 345

• Object id 20           is target for        File 678

- Object id 25             is derived from  Object id 27

Relationships can only exist between objects in the DRS.  These relationships can be added through the batch xml transaction file during deposit of the related objects or any time after. Consult the <relationshipMap> element in the DTD element descriptions for a list of currently defined relationships.

If you plan to relate objects between batches, you must track the DRS object IDs returned from loading reports.  If digital object A is deposited in batch #1 and digital object B is deposited in batch #2, the relationship transaction in batch #2 must specify the object ID or URN of object A because object A is not in the current batch.  The object id is the preferred identifier for referencing something already in the DRS.  The second example above demonstrates this relationship transaction, where object A has id 20 and object B has the file name 678 in batch #2.

# 5.0  Maintenance of DRS Data

This section covers topics related to the maintenance of DRS data.  For more information, see:

Correcting data already in DRS
Adding values to controlled vocabulary lists

## 5.1  Correcting data already in DRS

The DRS batch deposit process supports the addition of data only.  Options at batch deposit include:

- Add digital objects

- Create a URN to an object (for an object in the batch or an object already deposited)

- Create a relationship between objects (for objects within the batch and also objects already deposited).

Once objects are deposited, you can use the DRS administrative system (called DRS Web Admin) to further manipulate the objects and related metadata.  DRS Web Admin allows authorized object owners and deposit agents to view, add, update, and delete their data in DRS using only a web browser.  Using DRS Web Admin, authorized users can:

- Deposit individual objects and related metadata.

- View and download deposited objects.

- Update existing objects and related metadata (replace an object, change its metadata, add/update relationships, create URNs).

- Delete deposited objects.

What a user can do from DRS Web Admin depends on the security role assigned to the user and the DRS owner code associated with that role.  A user representing a single organization will usually be authorized to manipulate only objects under that organization's DRS owner code.  A user working for more than one organization (such as a deposit agent), will be authorized to manipulate objects for multiple owners.

Consult the DRS Web Admin section of the OIS web site for more information about functions, security roles, registration and access.

## 5.2  Adding values to controlled vocabulary lists

Some elements in the batch loading DTD have a set of controlled terms to select from (examples include relationships, role, purpose, quality). If the list of terms in any particular element does not reflect your archival needs, contact OIS and ask for the OIS Metadata Analyst.

# 6.0 DTD Base Element Descriptions

The DRS batch DTD is available from the OIS web site:

http://hul.harvard.edu/ois/xml/xsd/drs/drs_batch.dtd

This section defines the base elements within the batch DTD. For definitions of type-specific elements in the DTD (e.g., still images, audio), consult the type-specific metadata supplements available in the DRS Documentation Center of the OIS public web site.

Click on an element name to move to its definition in the list below.

| | | | |
|---|---|---|---|
| <batch> | <add> | <object> | <fileFormat> |
| <contactInfo> | <relationshipMap> | <file> | <createDate> |
| <emailSuccess> | <objectID> | <objectData> | <access> |
| <emailFailure> | <id> | <ownerSuppliedName> | <usageClass> |
| <sucessMethod> | <relationship> | <billingCode> | <signature> |
| <transaction> | <urnMap> | <role> | <mimetype> |
| <owner> | <urnmask> | <purpose> | <metadata> |
| | <urn> | <quality> | |

**<batch>**

| | |
|---|---|
| Purpose | root node of the DRS batch loading document |
| Mandatory | yes |
| Required attribute(s) | **name** - DRS emails a report back to the depositor about successful loads. This report references the load by this name attribute. |
| Optional attribute(s) | **userval** - this attribute allows the depositor to associate text information of their choosing with that particular loading session.<br>**directive** - used to request special processing, see appendix<br>**remove** - a comma-delimited list of file patterns that should be deleted by the DRS loader after a successful load even if these files were not loaded into the DRS. Four types of file patterns are supported: (1) specifying the exact file title, e.g. data.out (2) specifying a file ending, e.g. *.bak (3) specifying a file start, e.g. temp* (4) specifying all files, e.g.: * |
| Elements contained | contactInfo, transaction |
| Sample use | <batch name="run17" userval="shooting with new hasselblad digital unit"><br><batch name="run17" remove="*.bak, temp*"><br><batch name="run17" remove="*"><br><batch name="run17" remove="data.copy"> |

**\<contactInfo\>**

| | |
|---|---|
| Purpose | Area where email addresses are listed to report batch load successes and failures. |
| Mandatory | yes |
| Attribute(s) | none |
| Elements contained | emailSuccess, emailFailure |

**\<emailSuccess\>**

| | |
|---|---|
| Purpose | Declare a single or comma delimited list of email addresses to send the result report. |
| Mandatory | Yes |
| Attribute(s) | none |
| Sample use | \<**emailSuccess**\>diglib@lab.harvard.edu, jane_doe@harvard.edu\</**emailSuccess**\> |

**\<emailFailure\>**

| | |
|---|---|
| Purpose | Declare a single or comma delimited list of email addresses to report errors about the load. |
| Mandatory | yes |
| Attribute(s) | none |
| Sample Use | \<**emailFailure**\>diglib@lab.harvard.edu, jane_doe@harvard.edu\</**emailFailure**\> |

**\<successMethod\>**

| | |
|---|---|
| Purpose | Request a method of delivery for DRS success reports. |
| Mandatory | Yes<br>Note that if this element is missing from batch.xml DRS loader will add it at deposit with the default value="EMAIL" |
| Attribute(s) | Value<br>Possible values are:<br>EMAIL \| DROPBOX \| ALL |

| | |
|---|---|
| Sample Use | **<successMethod value="EMAIL" / >** |

**<transaction>**

| | |
|---|---|
| Purpose | Marks the start and end of an individual operation in the DRS. |
| Mandatory | yes |
| Required attribute(s) | none |
| Optional attribute(s) | userval – this attribute is similar to the userval attribute provided to you in the opening **<batch>** element.  A unique comment can be provided for each transaction.  Currently, add is the only operation enabled through the batch loader. |
| Elements contained | owner, add |

**<owner>**

| | |
|---|---|
| Purpose | DRS owner code that specifies who owns the digital object being deposited.  This owner code should be given to you during the owner registration process.  Every transaction in the DRS must be owned. |
| Mandatory | yes |
| Attribute(s) | none |
| Elements contained | none |
| Sample use | **<owner>**HUL.OWNER**</owner>** |

**<add>**

| | |
|---|---|
| Purpose | Marks the start and end of an add operation within per transaction.  Currently, you can add digital objects, relationships between objects, or request URNs for digital objects in the DRS. |
| Mandatory | yes |
| Attribute(s) | none |
| Elements contained | object, relationshipMap, urnMap |

**<relationshipMap>**

| | |
|---|---|
| Purpose | Defines a DRS relationship.  A relationship map in the DRS consists of three items: |

- reference to a DRS object

- relationship type

- reference to a DRS object

The references can identify an object previously loaded in the DRS or in the current batch. The relationships are read left-to-right.

| | |
|---|---|
| Mandatory | no |
| Required attribute(s) | none |
| Optional attribute(s) | **constrained** - A relationship can be constraining or unconstrained. A DRS object cannot be removed if any constraining relationships to it exist. Typically, any relationship that exists between two objects with the same DRS owner is constrained. Unconstrained relationships are only used for specifying inter-owner relationships. |
| Elements contained | %objectID;, relationship, %objectID; |
| Sample use | `<transaction>`<br>`<owner>DRS.TEST</owner>`<br>`<add>`<br>`    <relationshipMap>`<br>`        <file>vase.jpg</file>`<br>`        <relationship value ="IS_DERIVATIVE_OF"/>`<br>`        <file>vase.tif</file>`<br>`    </relationshipMap>`<br>`</add>`<br>`</transaction>` |

**\<objectID\>**

| | |
|---|---|
| Purpose | XML entity is used to refer to a DRS object in the current batch or already in the DRS. There are four ways you can identify a DRS digital object during the loading process: |

- **file** - filename of the object in the current batch. This identifier can only be used locally within a batch.

- **urn** - the URN of an object already in the DRS. This identifier cannot be used for referring to objects in the current batch. Note that only objects that have had specific add URN transactions are named in the DRS. Not all objects have URNs.

- **id** - DRS object identifier. Each object in the DRS has a numerical identifier associated with it upon deposit. This is the key for digital objects in the DRS. Every object already in the DRS must have an object id. This identifier cannot be used for referring to objects in the current batch.

- **ownerSupplieName/role/purpose/quality/mimetype** - the combination of these can be used to identify an object either in the current batch or already in the DRS.

Note that these must be unique within the owner's space.

**<id>**

| | |
|---|---|
| Purpose | DRS object id is a numerical value that uniquely identifies a digital object in the DRS.  Note that this is different from a URN, which identifies a digital object in a globally unique namespace. |
| Mandatory | no |
| Attributes | none |
| Sample use | **<id>**123**</id>** |

**<relationship>**

| | |
|---|---|
| Purpose | To specify a relationship between two objects in the DRS. Relationships are used to describe how separate digital objects were produced, such as IS_DERIVATIVE_OF, or how they form a complex object, such as IS_PART_OF. |
| Mandatory | yes |
| Required Attribute(s) | **value -** the relationship text.  Here is a list of the relationships supported by the DRS and their meaning. |
| Sample use | **<relationship value**='IS_DERIVATIVE_OF'/> |

| Name | Metadata Supplement |
|---|---|
| | |

| HAS_DECLARATION | Text |
|---|---|
| HAS_DTD | Text |
| HAS_ENTITIES | Text |
| IS_AUXILIARY_OF | Audio |
| IS_DERIVATIVE_OF | Image |
| IS_ICC_FOR | Image |
| IS_ICC_OF | Image |
| IS_IMAGE_FOR | Image |
| IS_INDEXED_BY | Text |
| IS_OCR_FOR | Text |
| IS_OCR_OF (deprecated) | Text |
| IS_PART_OF | Audio, Text |
| IS_PRESERVATION_REPLACEMENT_OF | |
| IS_RELATED_CHANNEL_OF | Audio |
| IS_TARGET_OF | Image |
| IS_TDF_FOR | Image |
| IS_WAVEFORM_FOR | Audio |
| IS_WORLD_FILE_OF | Image |

<urnMap>

| | |
|---|---|
| Purpose | Element allows you to name a digital object in the current batch or one that is already in the DRS.  You may request a specific URN (using the <urn> element), or you may request one be generated for you (using the <urnmask> element). |
| Mandatory | no |
| Attribute(s) | none |
| Elements contained | %objectID, urnmask, urn |
| Sample use | <transaction> <owner>DRS.TEST</owner> <add>    <urnMap>       <file>vase.jpg</file>       <urnmask>urn-3:DRS.Guest:{n}</urnmask>    </urnMap> </add> </transaction> |

**<urnmask>**

| | |
|---|---|
| Purpose | To request URN generation for a particular digital object.  A URN mask specifies a format for generating a URN.  Please see the "URN Masks" part of this document that describes that. |

| | |
|---|---|
| Mandatory | no |
| Attribute(s) | none |
| Sample use | **<urnmask>**urn-3:DRS.Guest:{n} **</urnmask>** |

**<urn>**

| | |
|---|---|
| Purpose | To request a specific URN for a digital object.  This requested URN must have the correct authority path and format. |
| Mandatory | no |
| Attribute(s) | none |
| Sample use | **<urn>**urn-3:DRS.Guest:12345**</urn>** |

**<object>**

| | |
|---|---|
| Purpose | This element brackets base object data that is to be added to DRS. An object during the load process consists of a file name, base object data, and type-specific metadata.  Each of these items is captured as a sub-element of object. |
| Mandatory | no |
| Attribute(s) | none |
| Elements contained | file, objectData, metadata |

**<file>**

| | |
|---|---|
| Purpose | The depositor provides the name of the physical file that has been dropped into the SFTP drop box. The file names must be unique within the **batch.xml** driver file.  This file name is retained in the DRS but is not searchable- see ownerSuppliedName for defining a local identifier. |
| Mandatory | yes |
| Attribute(s | none |
| Sample use | **<file>**vase_aquatic.tif**</file>** |

**<objectData>**

| | |
|---|---|
| Purpose | Specifies crucial data about the digital object that you are depositing.  It has base metadata about the digital object that is used for billing, identification, validation and access.  Type-specific data about this object is stored in the metadata element. |
| Attributes | none |
| Elements contained | ownerSuppliedName, billingCode, role, purpose, quality, fileFormat, signature, createDate, mimetype, access |

**<ownerSuppliedName>**

| | |
|---|---|
| Purpose | Depositors must provide a name for each deposited object. Most often, this is a tracking number used by the depositor.  This name, in combination with (role/purpose/quality), must be unique within the owner's set of objects in the DRS.  Please see |

the "Identifiers" part of this document that explains this.

| | |
|---|---|
| Mandatory | yes |
| Attribute(s) | none |
| Sample use | **<ownerSuppliedName>**music123-abc**</ownerSuppliedName>** |

**<billingCode>**

| | |
|---|---|
| Purpose | These codes are supplied to the owning organizations by the Office for Information Systems (**OIS**).  They should be received upon owner registration. |
| Mandatory | yes |
| Attribute(s) | none |
| Sample Use | **<billingCode>**HUL.OWNER.XYZ_0001**</billingCode>** |

**<role>**

| | |
|---|---|
| Purpose | Defines the role of the object as compared to other digital objects that are the same logically but a different format.  See the "Identifiers" part of this document for explanation.<br><br>In many cases, there may only be one version of a particular digital object.  In this case, the ownerSuppliedName should be unique and this value is not needed. |
| Mandatory | yes |
| Required attribute(s) | **value -** value of the role.  Should be 'NA' if not-applicable. |
| Sample use | <role value='NA'/> |

| Name | Metadata Supplement |
|---|---|
| ARCHIVAL_MASTER | Image |
| DELIVERABLE | Image |
| PRODUCTION_MASTER | Image |
| CONTAINER | |
| NA | |

**<purpose>**

| | |
|---|---|
| Purpose | Defines the purpose of the object as compared to other digital objects that are the same logically but in a different format.  See the "Identifiers" part of this document for explanation.<br><br>In many cases, there may be only one version of a particular digital object.  In this case, ownerSuppliedName is unique and this value is not needed. |
| Mandatory | yes |
| Required attribute(s) | **value -** value of the purpose.  Should be 'NA' if not-applicable. |

Sample use                      <purpose value='NA'/>

| Name | Metadata Supplement |
|------|---------------------|
| BITONAL | Image |
| COLOR | Image |
| CROPPED | Image |
| GRAYSCALE | Image |
| PRINT | Image |
| PROCESSED | Image |
| RAW | Image |
| VIEW | Image |
| WHOLE | Image |
| NA | |

**\<quality\>**

Purpose                      Defines the quality of the object as compared to other digital objects that are the same logically but in a different format. See the "Identifiers" part of this document for explanation.

The owner or depositor decides metric for quality. It could be size, resolution, length or any other measurement that is meaningful.

In many cases, there may be only one version of a particular digital object. In this case, ownerSuppliedName is unique and this value is not needed.

Mandatory                      yes

Required attribute(s)           **value -** value of the quality. Should be 'NA' if not-applicable. These values range from 1 to 10, where 1 is the lowest quality and 10 is the greatest.

Sample use                      <quality value='NA'/>

**\<fileFormat\>**

Purpose                      The basic nature of the binary material on disk. It is more specific than the DRS metadata type in many cases.

Mandatory                      yes

List of valid                  ICC | GIF | JPEG | TIFF | TDF | TEXT | PCD | AIFF |
element values                RealAudio | APP | WAV | WFR | JP2 | ZIP | GZIP | PDF

**\<createDate\>**

Purpose                      The date this digital object was created. This field has the following valid date formats:

|  |  |  |
|---|---|---|
|  |  | YYYYMMDD<br>YYYY:MM:DD<br>YYYY:MM:DD HH:MM:SS |
|  |  | Leading zeros should be prepended if necessary. The hours are on 24-hour time. |
|  | Mandatory | no |
|  | Attribute(s) | none |
|  | Sample use | **<createDate>**2001:04:14 15:15:20**</createDate>** |

**<access>**

|  |  |  |
|---|---|---|
|  | Purpose | This element specifies what bodies can access this digital object if it is deliverable. There are three options for this: P for public, R for restricted and N for no access. A public object is available to the world. A restricted object is available to only Harvard. An object marked as having no access is only available to administrators via the DRS Web Admin. |
|  | Mandatory | yes |
|  | Required Attribute(s) | **values -** P = public, R = restricted, N = no access. Only capital letters are accepted. |
|  | Sample use | <access value= "P"/> |

**<usageClass>**

|  |  |  |
|---|---|---|
|  | Purpose | Each digital file is stored according to its usage classification as either "high use" or "low use." Deliverables used for public access should be classified as "highuse". Preservation and archival versions, masters that are not being delivered and other "dark objects" should be classified as "lowuse". |
|  | Mandatory | yes |
|  | Required Attribute(s) | **values** – HIGHUSE, LOWUSE |
|  | Sample use | **<usageClass value="HIGHUSE"/>** or **<usageClass value="LOWUSE"/>** |

**<signature>**

|  |  |  |
|---|---|---|
|  | Purpose | The signature validates the integrity of the digital object during transfer between the depositor's system and the DRS. It is also used for file validity within the DRS after it has been deposited.<br><br>MD5 signatures are always character strings of length 32 specifying a hexadecimal checksum. All letters [a-f] should be lowercase. |
|  | Mandatory | yes |
|  | Required Attribute(s) | **type** - the type of the signature. Currently, the DRS only |

Version 5.004
Revised: January 27, 2010

supports MD5 signatures

Sample use          **<signature type='MD5'>** 7c9b35da4f2ebd436f1cf88e5a39b3a2 **</signature>**

**<mimetype>**

Purpose             Specifies the mime type of the digital object you are depositing.

Mandatory           yes

Attribute(s)        None

Valid mimetypes

| | |
|---|---|
| application/x-esri-pyramid-file | application/x-esri-statistics-file |
| application/x-icc | application/x-sonic-waveform-reduction |
| application/x-wavelab-waveform-reduction | application/zip |
| audio/x-aiff | application/pdf |
| audio/x-wave | audio/x-pn-realaudio |
| image/jp2 | image/gif |
| image/jpx | image/jpeg |
| image/x-photo-cd | image/tiff |
| text/sgml | text/plain |
| | text/xml |

Sample use          <mimetype>**text/plain**</mimetype>

**<metadata>**

Purpose             This element holds the type-specific metadata for the digital object.  As opposed to the base-object metadata above, different metadata types (images, text, audio) have completely different sets of attributes required for sufficient description and archiving.

This tag also specifies the DRS metadata type.  Specifying a metadata type (IMAGE) and a metadata sub-element (imageMetadata) may seem redundant.  However, in some cases, more than one metadata sub-element may be used for the same metadata type element.  Here is a list of valid DRS metadata types and the metadata sub-elements they should use.

| Metadata type | Sub-Element | Supplement |
|---|---|---|
| APP | appMetadata | Application |
| AUDIO | audioMetadata | Audio |
| IMAGE | imageMetadata | Image |
| TARGET | imageMetadata | Image |
| TEXT | textMetadata | Text |

Version 5.004
Revised: January 27, 2010

| CONTAINER | containerMetadata | |
|-----------|-------------------|---|

*Mandatory*                    **yes**

Required Attribute(s)          type - metadata type.  See table above for appropriate values.

Sample use

```
<metadata type='IMAGE'>
   <imageMetadata>
       consult Image appendix
   </imageMetadata>
</metadata>
```

or,

```
<metadata type='TEXT'>
   <textMetadata>
       consult text appendix (not yet available)
   </textMetadata>
</metadata>
```

## 7.0  Requesting Assistance with Batch Loading

The HUL Office for Information Systems (OIS) provides administrative and technical support for DRS and its related systems.

- For questions about registration and setup to use DRS, please contact the Digital Projects Team.

- To report a problem or ask a question about batch deposits or other DRS technical issues use the OIS Support Center.

If you need to talk with someone in person, consult the OIS Support Center for the current DRS contact in OIS.