Issues:

1) Correct existing JPEG datastreams currently in the repository that were derived from a DjVu when a source TIFF is present
2) Correct method used for generating JPEG datastreams for the repository
3) Provide a thumbnail and other derivatives using a datastream of the object
4) Decide which content models will have thumbnails

1) Correct existing JPEG datastreams currently in the repository that were derived from a DjVu when source TIFF is present

Current JPEG datastreams were created in the file generation pipeline from a DjVu file that were created from a source TIFF. The fidelity of these JPEG's is in question. All objects with a JPEG datastream derived from a DjVu file that were derived from a source TIFF need to be corrected by purging the JPEG datastream and generating a new JPEG datastream from the source TIFF. This would be a one-time update of the identified objects with this issue in the repository. JPEG size will be a maximum of 1,600 pixels wide or high, whichever is greater, retaining the original aspect ratio.

2) Correct method used for generating JPEG datastreams for the repository

Current JPEG datastreams are created from a DjVu file which is derived from a source TIFF. This workflow needs to be corrected so that the JPEG datastreams are created from the source TIFF. This would require a change in the file generation pipeline in the WMS. JPEG size will be a maximum of 1,600 pixels wide or high, whichever is greater, retaining the original aspect ratio.

3) Provide a thumbnail and other derivatives using a datastream of the object

From a user interface perspective providing three different sized derivatives of the source object will provide a good amount of flexibility. After processing all images will retain their original aspect ratio.

Size 1: 80 pixels wide or high, whichever is greater. Used as a thumbnail when viewing a brief record display in a result list. Also can used in the future as a thumbnail when target screen resolution is small by nature, ie cell phone or PDA type device.

Size 2: 160 pixels wide or high, whichever is greater. Used as a thumbnail when viewing a full record display.

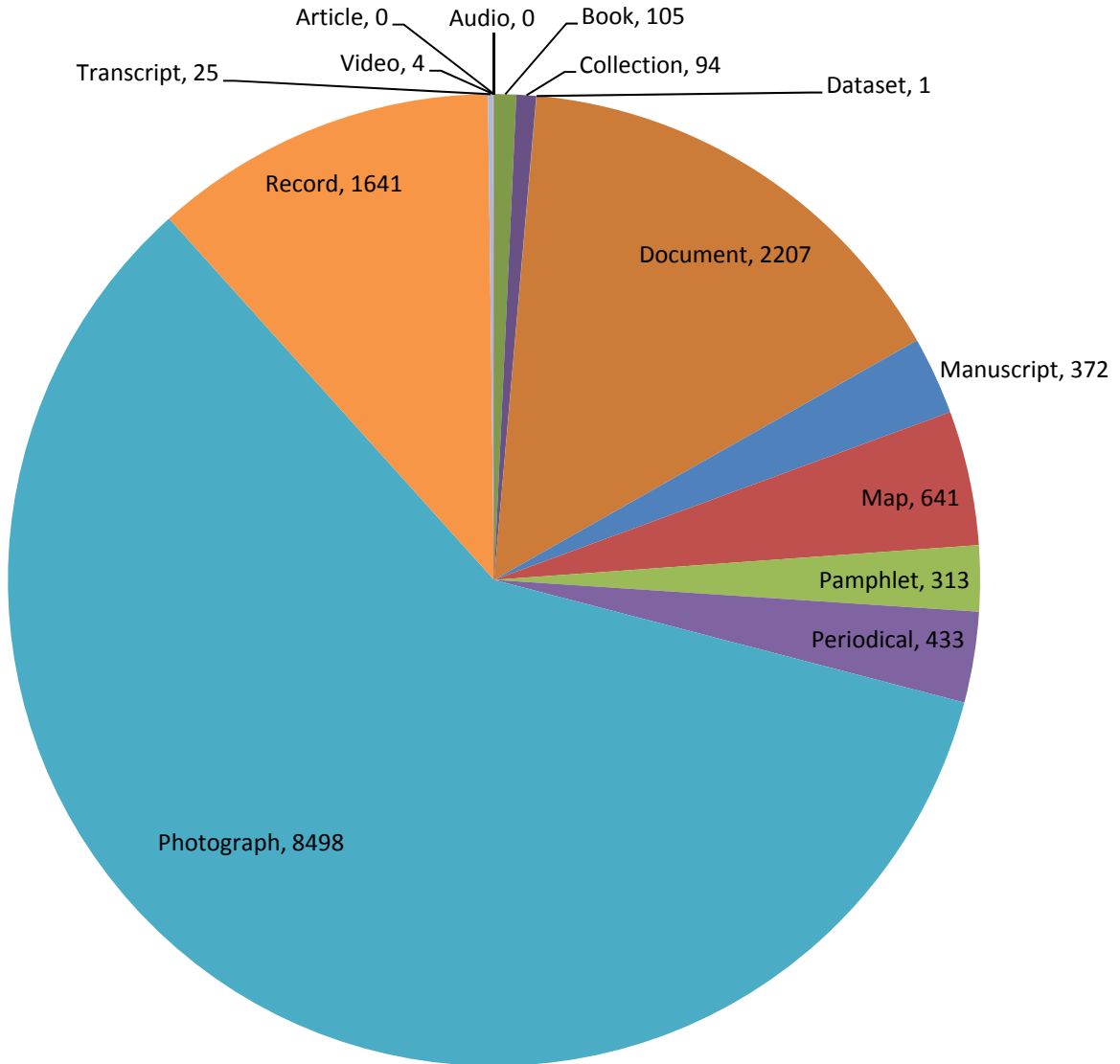**Size 3**: 500 pixels wide or high, whichever is greater. Used for medium sized presentation display.

Solution: Generate one thumbnail datastream and create other derivative sizes dynamically
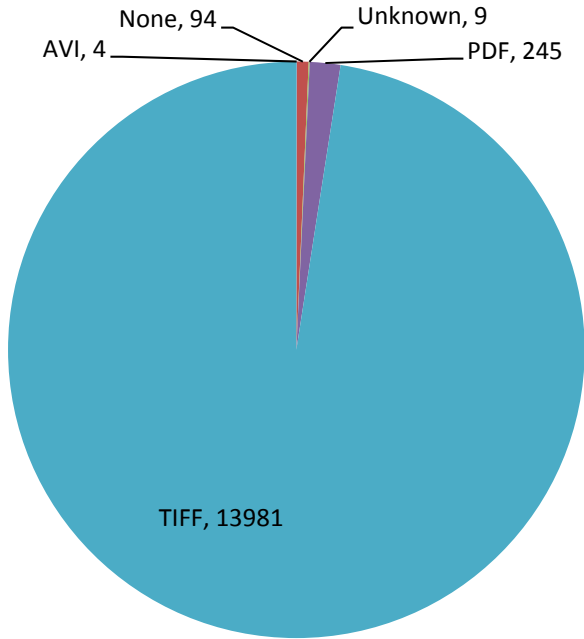
For a result list of 10 to 50 objects generating all the Size 1 thumbnails dynamically, on the fly, may prove to be to an intensive process. This solution would require a Size 1 thumbnail datastream to exist and other sizes needed would be generated dynamically using something like gdLibrary, http://www.libgd.org, from the JPEG datastream, when requested. Since these other sizes would be on a single request basis and not used in a result list of 10 to 50 objects the transformation process might not prove to be excessively intensive. This would also provide us added flexibility for

changing the resolutions of the Size 2 and Size 3 in the future, if we desire, and possibly providing even more derivative resolutions. We will count access of these JPEG's that are generated on the fly using the 'JPEG-1M' moniker in the statistics program.

4) Decide which content models will have thumbnails

Using content model information provided by Ron Jantz for R5.0 the following charts were created. This data is a snapshot in time. The first chart is the breakdown of the content models that will be created for R5.0 and the number of objects for each. The second chart is of what file format the objects in the analysis use as their master datastream.

None Collection content models

PDF Manuscript(ETD's-238) and Documents(Proceedings-7)

Unknown Document (Data – 5, Poll-1, Questionnaire-1, Schedules-1), Dataset (Poll-1)

After reviewing the charts the largest impact for R5.1 would be to start with the Photograph Content Model.   Also the Video Content Model will have a thumbnail datastream. Processes for producing thumbnails for both content models will be pursued. It is thought that for the Photograph Content Model an automated process can be used for both the initial and continual/ongoing creation.

Datastream Names:

The following is a recommendation for the name of the thumbnail datastream, THUMBJPEG-1

Example - THUMBJPEG-1: 80px

Display of Thumbnails in result list:

It is thought that having a list of results from a search with thumbnails not of equal dimensions might created a cluttered or confused look. There are possible solutions for this if it is deemed an issue. When display the 80px thumbnails the <img> tag used can change the dimensions of the image being displayed to make them square.

Example: Actual JPEG thumbnail is 80px by 64px.

Not altered	Altered
	

An alternative is to render 85px by 85px boxes around the thumbnails as well.


```
<div style="width:85px; height:85px; border:1px solid #999999;">
        <img src="thumbjpeg-1" width="80" height="64" border="0"/>
</div>
```

WMS Impact:

In R5.1 any new objects introduced will need to either create or manage thumbnails datastreams. Working with Kalaivani Ananthan and the under development R5.0 release of WMS we were able to determine assigning a user uploaded thumbnail datastream is an option easily configured. Also the generation of the JPEG from the source TIFF instead of the DjVu is a configure option as well. At this time in order to create a reduced sized JPEG from the source TIFF the function that allows for user defined parameters to be applied will need to be developed in the WMS.

Search Impact:

In R5.1 the search interface will need to use the Thumbnail and JPEG datastreams instead of using the external sources it has used in the past. In the event there is no thumbnail jpeg for an object the search interface will fall back to using icons, like it has in the past.

Statistics Impact:

In R5.1 access to the Thumbnail and JPEG datastreams will start or increase. It needs to be determined what access will count against the "download" count in the statistics package. The two main questions are:

1) Should thumbnail datastream access be counted?
   No, thumbnail access will not be counted in the statistics reporting.
2) Should JPEG datastream access be counted?
   Yes, this will be counted as a view in the statistics reporting. For JPEG datastream access that was rescaled this will be distinguishable in the statistics from full resolution JPEG datastream access.

Examples

Example 1: Source is a TIFF of dimensions: 1004px by 803px	
THUMBJPEG-1: 80px by 64px JPEG: 1004px by 803px	

Example 2: Source is a TIFF of dimensions: 2132px by 2708px	
THUMBJPEG-1: 63px by 80px JPEG: 1260px by 1600px	

Example 3: Source is a TIFF of dimensions: 482px by 213px	
THUMBJPEG-1: 80px by 35px JPEG: 482px by 213px	