

# Statistics Requirements for Release 5.1

## New Features in Release 5.1

Statistics website will add new features in release 5.1. It will have three different interfaces for RUCore Admin, Collection Managers/Collection Owners, and Faculty Members. New features include addition of Partner Portals, addition of ARL statistics, and addition of searches by domain section. It will retain object ID for RUCORE admin user and Collection managers, and replace object ID with object title for Faculty members. It will also change database architecture and generate HTML report on ARL statistics. In addition to the selection of single collection and all collections, selection of multiple collections will be also available.

For database scalability, we will always keep three years window for current Statistics website. With release 5.1, the statistics will not include the data earlier than July 2005. We will create an archival database every year as backup. We will be able to generate the report for all backup data as the user requests.

ARL statistics will have four major areas, i.e., RUCore, NJDH, NJVid, and NJEDL. We will not break out the statistics by collections for these four major areas. ARL statistics will have a selected time frame for academic year. It will include total number of collections, total number of objects, and total storage for cumulative statistics. It will include total numbers of search, total objects downloaded, total objects ingested for academic year. An HTML report will be generated for ARL statistics.

Statistics features will be extended to the Partner Portals. Faculty deposit portal is already in Release 4.5. In Release 5.1 we are going to provide more general information for all other Partner Portals as well as faculty deposit portal. It will only show the statistics of collections which the partner owns. User interface will include Searching statistics and Downloads Statistics.

Total objects will break into content models (Articles, Maps, and Books etc.) based on the techMD field "contentModel". Valid content models are taken from WMS. All others are considered as "Genre Unknown" and are candidates for correction. The Statistics report will use the word "class", in our case it is equivalent to content model. This requires a database structure change and data pre-population.

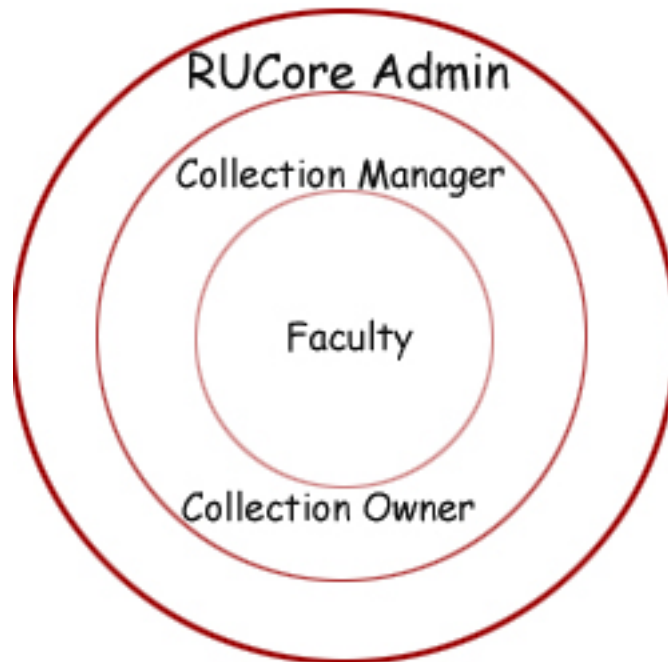
IP addresses and domain names will be collected in the statistics identity table when a search is performed. Searches by domains will be counted starting from release 5.1.

It will count the thumbnail image view toward downloaded statistics when the thumbnail is being treated as a fedora datastream.

The full record view count toward to the statistics will be considered for a future release (i.e. after R5.1).

## User Interfaces

There will be three different user interfaces. See below image for three different types of users, with the capability extended in outer circles.



The user will be able to select the timeframe for reporting statistics. It allows users to view a summary of all statistical data based on selected collections during a selected timeframe. You will have the flexibility to select from one user selected time to another user selected time.

Attachment 1 at end is the screen shot for RUCORE admin and collection manager to access the statistics page. Authentication and authorization will be performed through the DLR login. If you are a RUCORE admin, you will be able to select one collection, all collections, and some collections to view the summary. You will see a dropdown list of ROOT, RUCORE, NJEDL, NJVID and NJDH on the Management system page for ARL statistics access as well. If you are a collection manager, you will only see the collection statistics you own.

RUCore Admin User (see attachment 2, Figure 1 at end):

- Will be able to view all the collections, selected collections and single collection's statistics;

- Will be able to access all the statistics features (Downloads, Total number of Objects by Class, Searches, Searches by Domain); and
- Will be able to access ARL statistics.

Collection Manager/Collection Owner:

- Will be able to view statistics of single collection he or she manages. Authentication and authorization will be performed through the DLR login; and
- Will be able to access all the statistics features (Downloads, Total number of Objects by Class, Searches, and Searches by Domain).

Faculty Member (see attachment 3 – Figure 1 at end):

- Will be able to view statistics of collection he or she owns, and to access statistics through the faculty deposit webpage; and
- Will be able to access fewer of the statistics features (Total number downloads -- see attachment 3 – Figure 2 at end, Total number of Objects, Total number of Searches, and Searches by Domain).

Design Overview for RUCORE admin and collection manager:

## 1. Downloads

**Total objects downloaded:** It allows user to view the total number and percentage of downloads during selected timeframe VS total objects downloaded. It will draw a table of view from preservation datastreams, jpeg, pdf, djvu, etc. (see attachment 2, Figure 2 at end) For RUCORE admin and collection managers, the download list will show the Object id with mouse rollover to see the object title. Faculty members will only see the object title.

**Top 25 downloads:** It allows user to view top 25 objects downloaded. Object id will be changed to Object Title for faculty member and object ID will retain for RUCORE admin. It will draw a table of view from preservation datastream, jpeg, pdf, djvu, etc.

**Unique documents viewed:** It allows user to view the total number and percentage of unique documents downloads during selected timeframe VS total objects downloaded.

## 2. Total number of Objects by Class

**Total number of objects in database:** It allows user to view the total number and percentage of objects in database. It will not count XACML objects, Disseminators, and content model. It will count annotation objects.

**Article:** It allows user to view the total number and percentage of Article objects in database.

**Audio:** It allows user to view the total number and percentage of audio objects in database.

**Map:** It allows user to view the total number and percentage of map objects in database.

...

(Note: Class of the objects will be listed based on the object's contentModel)

**Number of objects added:** It allows user to view the total number and percentage of objects added. It will link to the ingest list. (See attachment 2, Figure 3 at end)

**Number of objects deleted:** It allows user to view the total number and percentage of objects deleted. It will link to the purge list.

**Number of objects edited:** It allows user to view the total number and percentage of objects edited. It will link to the edit list.

**Integrity Check:** It allows user to view the total number of checksum failure. Link to the checksum logs.

### 3. Searches

**Total Number of Searches:** It allows user to view the total number and percentage of searches for collections.

**Searches with more than (number#) hits:** User should be able to select from a list of numbers (25, 50, 75, 100, and 1000). It allows the user to view the total number and percentage of searches with more than user selected number hits. It will link to a list to include the detail of search term, number of hits, etc. (see attachment 2, Figure 4 at end)

**Searches with zero hits:** It allows the user to view the total number and percentage of searches with zero hits. It will link to a list to include the detail of search term, number of hits, etc.

**Search Results [Last 10 Days]:** It allows the user to view a summary of all relevant search activity in the last ten days.

**Top 25 Searches:** It allows the user to view a listing of the top twenty-five most popular search strings.

#### 4. Searches by Domain

**Searches from Management Domain (TSB, SCC etc.):** It allows the user to view the total number and percentage of searches from management domain.

**Searches from Rutgers:** It allows the user to view the total number and percentage of searches from Rutgers except management domain.

**Searches outside Rutgers:** It allows the user to view the total number and percentage of searches outside Rutgers.

#### 5. ARL Statistics (see attachment 4 at end):

If you are RUCORE Admin, a dropdown list of ROOT, RUCORE, NJEDL, NJVID and NJDH will show on the Management system page. Admin user shall pick the list first. A HTML report will be displayed and it will display a drop down menu for academic year on the report, with default current academic year. If another academic year selection is made, then an Html report will be generated. The Report shall include below:

**Cumulative statistics for** total collections, total number of objects, and total storage (both derivative and archival storage is included).

##### **Statistics for academic year:**

- **Total Objects added:** It allows user to view total number of objects added in academic year July 1 – June 30
- **Total objects downloaded:** It allows user to view total objects downloaded in academic year July 1 – June 30
- **Total objects searched:** It allows user to view total objects ingested in academic year July 1 – June 30

## Architecture Changes

#### 1. Database Modification

In order to collect the new statistics information and to avoid the large number of calculation on the fly, we shall change the database structure and pre-populate the data to accommodate the requirement for statistics and custom portals.

Domain name, IP address (How long do we keep the IP address?), collection classes (e.g. Article, Audio, Book, Diary, Manuscript, Map, Newsletter, OralHistory, Pamphlet, Periodical, Photograph, Poll, Record, Video etc.), object title, persistent URL, file size, etc. will be collected to the statistics database.

## 2. DLR impact

DLR authentication and authorization shall be set an additional access level for collection manager and collection owner. When collection managers or collection owners login to DLR, DLR shall pass the collection IDs to Statistics.

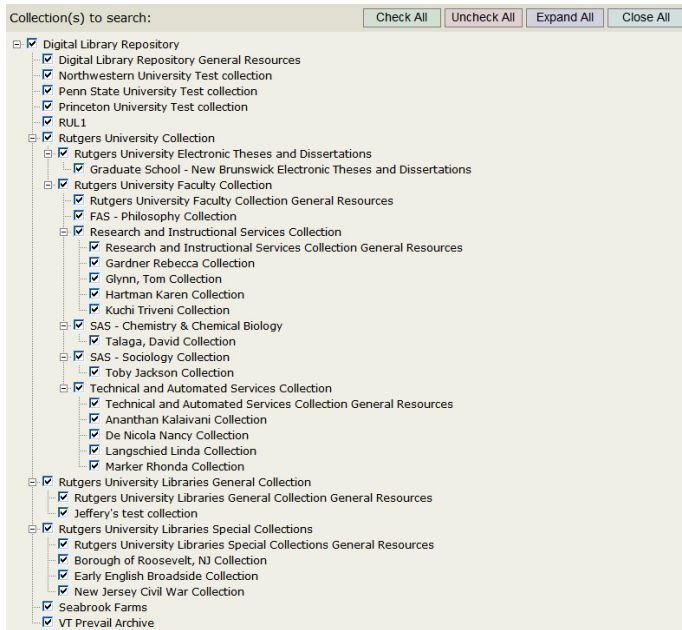
Showfed will call the statistics update function to be counted towards the download statistics when the audio or the video are downloaded.

## 3. Collection Display

For RUCORE admin, currently, statistics selection of single collection or all collections is under the DLR management system. We shall modify the selection in DLR management system. In addition to the selection of single collection and all collections, selection of multiple collections shall be also available.

The overall collection hierarchy of the repository is currently fragmented. Both NJDH and DLR/RUcore are starting collection nodes, and are not related to each other by a super/root collection. It has been discussed and agreed that a root collection node with NJDH and DLR/RUcore as its children should be added. This will create the ability to completely render the entire repository collection hierarchy. This root collection will be created and called "root" with the DLR and NJDH collections becoming children of it and having the corresponding mods:relatedItem fields in the collection objects updated.

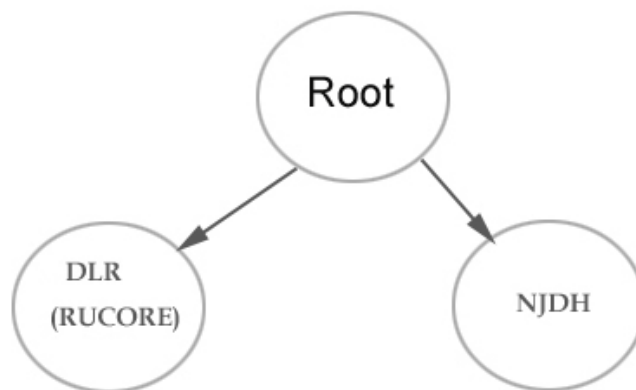
See below the tree display for RUCORE.



DLR (which is RUCORE) and NJDH, they are parallel top nodes before release 5.1. See below:



We will add a top parent node ROOT and put RUCORE and NJDH under it in release 5.1.



#### 4. WMS Changes

WMS will call the statistics update function to be counted towards the statistics when there is an update occurring for object edit and object purge using the WMS Fedora edit.

## 5. NJEDL scan service

We will capture the scan service requirement for statistics in the future release.

## 6. Modify Alerting System

Currently, Alerting is doing two functions. One is for sending email to collection manager while there is object ingested, edited or purged. Another function is to insert the information to the statistics database. We will separate those two functions after Release 5.1.

Attachment 1(Management System): RUCORE admin and collection manager to access the statistics through DLR. ARL statistics report drop down list shall be added to RUCORE admin user.





## Rutgers University Community Repository Management System

Searching and Editing | Collection Management  
Handle Creation | Indexing and Ingest



### Collection Database Functions

### Alerting

### Fedora Object Management

### Statistics

Select Collection(s):

- ☒ Rutgers University Faculty Collection
  - ☒ Rutgers University Faculty Collection General Resources
  - ☒ FAS - Philosophy Collection
  - ☒ Research and Instructional Services Collection
  - ☒ SAS - Chemistry & Chemical Biology
  - ☒ SAS - Sociology Collection
  - ☒ Technical and Automated Services Collection

Attachment 2 – Figure 1 (RUCORE Admin and collection manager):

Collection Management Statistics Report for All Collections



Timeframe: 1 January 2008 to 1 August 2008 Update

Downloads	Number	Percentage
Total Objects downloaded	4115	-
Unique Objects downloaded	4085	99%
Top 25 downloads	-	-

Total # of Objects by class	Number	Percentage
Total Number of Objects in database	1755	-
> Article	155	9%
> Audio	7	0%
> Book	2	0%
> Diary	19	1%
> Manuscript	257	15%
> Map	3	0%
> Newsletter	3	0%
> OralHistory	57	4%
> Pamphlet	9	1%
> Photograph	1	0%
> Poll	17	1%
> Record	11	1%
> Video	12	1%
Number of Objects added	211	12%
Number of Objects deleted	21	1%
Number of Objects edited	72	4%
Integrity Check	210	12%

Searching	Number	Percentage
Total Number of Searches	10559	-
Searches with more than 25 results	5197	50%
Searches with zero results	514	4%

Searching by Domain	Number	Percentage
Searches from Rutgers Institution	10101	10%
Searches from Management Domain (TSB, SCC etc.)	9101	10%
Searches Outside of Rutgers Institution	10101	10%

Attachment 2 – Figure 2 (RUCORE Admin and collection manager): Mouse rollover to see the object's title.

### Object Download Statistics

1 2 3 ...						
NEXT>>						
Object				File Types		
				pdf	txt	Total
rutgers-lib:659	The WPA Guide to 1930s New Jersey			0	0	1
rutgers-lib:486				0	0	1
rutgers-lib:4291				1	0	2
rutgers-lib:4260	1	1	0	2	0	4
rutgers-lib:4259	1	0	0	1	0	2
rutgers-lib:4258	2	0	0	2	0	4
rutgers-lib:4255	1	0	0	1	0	2
rutgers-lib:4254	2	0	0	1	0	3
rutgers-lib:4223	2	0	0	2	0	4
rutgers-lib:4201	1	1	0	1	0	3
rutgers-lib:4196	1	0	0	1	0	2
rutgers-lib:4193	2	0	0	2	0	4
rutgers-lib:4192	1	0	0	1	0	2
rutgers-lib:4191	2	0	0	2	0	4
rutgers-lib:4190	2	0	0	2	0	4
rutgers-lib:4144	1	0	0	1	0	2
rutgers-lib:4143	1	0	0	1	0	2
rutgers-lib:4142	1	0	0	1	0	2
rutgers-lib:4141	1	0	0	2	0	3
rutgers-lib:4138	1	0	0	1	0	2
rutgers-lib:4137	1	0	0	1	0	2
rutgers-lib:4135	1	0	0	1	0	2
rutgers-lib:4133	2	0	0	2	0	4
rutgers-lib:4131	1	0	0	1	0	2
rutgers-lib:4130	1	0	0	1	0	2
<b>TOTAL[Year 2008]</b>	32	2	0	31	0	65
<b>TOTAL[All Objects]</b>	2189	818	2	2478	570	6057

## Attachment 2 – Figure 3 (RUCORE Admin and collection manager): objects added log

Marker Rhonda Collection > Home

FEDORA Management Page

### Collection Management Statistics Report

SELECT A MONTH

2007 July

View Log

Specify Day

2008 January

View Log

Specify Day

Click on the View Log to see the detail

Marker Rhonda Collection > Home

FEDORA Management Page

DATE	TIME	OBJECT	USER
2007-07-10	10:30:15	<a href="#">rutgers-lib:201507</a>	Marker, Rhonda
2007-07-10	10:29:58	<a href="#">rutgers-lib:201506</a>	Marker, Rhonda
2007-07-10	10:29:46	<a href="#">rutgers-lib:201505</a>	Marker, Rhonda
2007-07-10	10:26:58	<a href="#">rutgers-lib:201504</a>	Marker, Rhonda

# Attachment 2 – Figure 4 (RUCORE Admin and collection manager): Searches with # of results

## Collection Management Statistics Report

1 2 3 ... NEXT>>

STRING	# Results	Search Field	Format	Date
Civil War	79	mods:subject	all	2008-06-18
black-and-white photographs	166	mods:genre	all	2008-06-18
EMPTY	49	fullrec	all	2008-06-18
James Seabrook	113	mods:name	all	2008-06-18
Seabrook Farms	151	mods:name	all	2008-06-18
Seabrook James	113	mods:name	all	2008-06-18
Japanese Americans	227	mods:subject	all	2008-06-18
Seabrook Charles F	110	mods:subject	all	2008-06-18
black-and-white photographs	166	mods:genre	all	2008-06-18
Seabrook Charles F	110	mods:subject	all	2008-06-18
Men	52	mods:subject	all	2008-06-18
War	207	mods:subject	all	2008-06-18
Jewish Americans	34	mods:subject	all	2008-06-18
EMPTY	118	mods:subject	all	2008-06-18
Seabrook Charles F	110	mods:subject	all	2008-06-18
Work	42	mods:subject	all	2008-06-18
EMPTY	81	mods:name	all	2008-06-18
Moravian Church	33	mods:subject	all	2008-06-18
EMPTY	230	fullrec	all	2008-06-18
migrant workers	37	mods:subject	all	2008-06-18
Moravian Church	33	mods:subject	all	2008-06-18
History	57	mods:subject	all	2008-06-18
EMPTY	81	fullrec	all	2008-06-18
Seabrook Charles F	110	mods:subject	all	2008-06-18
EMPTY	94	fullrec	all	2008-06-18
Agriculture	114	mods:subject	all	2008-06-18
Seabrook Educational and Cultural Center	98	mods:name	all	2008-06-18
Agriculture	114	mods:subject	all	2008-06-18
Seabrook Educational and Cultural Center	98	mods:name	all	2008-06-18
Women Japanese	36	mods:subject	all	2008-06-18
Seabrook Charles	124	mods:name	all	2008-06-18
EMPTY	43	mods:name	all	2008-06-18
EMPTY	118	mods:subject	all	2008-06-18
Seabrook James	113	mods:name	all	2008-06-18
Immigrants	123	mods:subject	all	2008-06-18
Men	52	mods:subject	all	2008-06-18
EMPTY	43	fullrec	all	2008-06-18
World War	203	mods:subject	all	2008-06-18
EMPTY	118	fullrec	all	2008-06-18
Civil War	79	mods:subject	all	2008-06-18

Attachment 3 – Figure 1(Faculty members): It shall have a link back to faculty deposit page.

#### Collection Management Statistics Report for Marker Rhonda Collection



Timeframe: 1 January 2008 to 1 August 2008 [Update](#)

Downloads	Number	Percentage
<a href="#">Total Objects downloaded</a>	30	-
Unique Objects downloaded	29	97%

Total # of Objects	Number	Percentage
Total Number of Objects in database	9	-

Searching	Number	Percentage
Total Number of Searches	4193	-

Attachment 3 – Figure 2(Faculty members): by click on total objects downloaded, only show the object title.

#### Detailed Download Statistics

Title	File Types					
	djvu	jpeg	mp3	pdf	txt	Total
<a href="#">The Estonian community of Seabrook, New Jersey</a>	2	0	0	2	0	4
<a href="#">Seabrook Farms staff group portrait</a>	1	1	0	1	0	3
<a href="#">Aerial view of Seabrook Farms</a>	1	1	0	1	0	3
<a href="#">I remember my father at Seabrook</a>	2	0	0	2	0	4
<a href="#">Army Discharge Paper for Kazuto Taniguchi</a>	1	1	0	1	0	3
<a href="#">1106 Parsonage Road, Seabrook, NJ</a>	1	2	0	1	0	4
<a href="#">A truck of asparagus</a>	2	2	0	2	0	6
<b>TOTAL(Year 2008)</b>	10	7	0	10	0	27

Attachment 4(ARL statistics by selection of the pick list):

Statistics Report for Assoc. of Research Libraries (ARL)



Cumulative ARL Statistics for - rucore	
Total collections:	<b>131</b> in rucore
Total storage of archive master:	<b>1301</b> GB in rucore
Total number of Objects:	<b>1768</b> in rucore
Academic 2005/2006 (July 1 - June 30) Statistics for rucore	
Year of	2005 - 2006 <input type="button" value="Update"/>
Total number of Objects added:	<b>78</b> in the selected academic year of 2006
Total number of Searches:	<b>361</b> in the selected academic year of 2006
Total number of downloads	<b>34</b> in the selected academic year of 2006

Jg – 2008/09/17