CUA Admin Documentation

Last modified: 2018/12/04 12:47

Terms Of Use

BY ACCESSING AND/OR USING THIS DOCUMENTATION, YOU ACKNOWLEDGE THAT YOU HAVE READ, UNDERSTOOD, AND AGREE TO BE BOUND BY THESE TERMS AND TO COMPLY WITH ALL APPLICABLE LAWS AND REGULATIONS. IF YOU DO NOT AGREE TO THESE TERMS, DO NOT USE THE @MIRE NV DOCUMENTATION.

INFORMATION IN THE DOCUMENTATION IS PROVIDED "AS-IS". SUBJECT TO STATUTORY WARRANTIES, IF ANY, THAT CANNOT BE EXCLUDED. @MIRE NV, ITS DEVELOPERS AND SUPPLIERS MAKE NO WARRANTIES, REPRESENTATIONS, GUARANTEES OR CONDITIONS OF ANY KIND, EITHER EXPRESS, IMPLIED OR STATUTORY, INCLUDING, BUT NOT LIMITED TO, THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT.

IN NO EVENT WILL @MIRE NV BE LIABLE TO ANY PARTY FOR ANY DIRECT, INDIRECT, SPECIAL CONSEQUENTIAL OR OTHER DAMAGES OF ANY KIND FOR ANY USE OF THE DOCUMENTATION, WEB SITE OR CONTENT, OR FOR ANY OTHER HYPER LINKED WEB SITE AND ITS CONTENT, INCLUDING, WITHOUT LIMITATION, ANY LOST PROFITS, BUSINESS INTERRUPTION, LOSS OF PROGRAMS OR OTHER DATA ON YOUR INFORMATION HANDLING SYSTEM OR OTHERWISE, EVEN IF WE ARE EXPRESSLY ADVISED OF OR AWARE OF THE POSSIBILITY OF SUCH DAMAGES OR LOSSES. THIS EXCLUSION AND WAIVER OF LIABILITY APPLIES TO ALL CAUSES OF ACTION, WHETHER BASED ON CONTRACT, WARRANTY, TORT, OR ANY OTHER LEGAL THEORIES.

@mire NV cannot guarantee the validity of the information found here at all times because:

- The content may contain technical inaccuracies or typographical errors and may be changed, improved or updated without notice.
- The software described in this documentation might have been customized on your request, to add or change specific functionality. In this case, please refer to custom made documentation or directly to your @mire NV representative.

Please be aware that @mire NV does not grant any rights or licenses to any copyrights, patents, trademarks, or any other intellectual property rights, except as otherwise expressly stated for the content. All rights not explicitly granted under this agreement are reserved.

@mire NV grants you non-exclusive, non-transferable, limited permission to access and display the documentation for use within your organization. If you use or reference content from the documentation, note that such content is not provided as "public domain" and is copyrighted @mire NV. You may not mirror any of the content from this documentation in any other media outside of your organization. You may not copy or distribute in whole or in part the contents of this documentation outside of your organization. Your failure to comply with the terms, conditions and notices in the documentation will result in automatic termination of any rights granted to you, without prior notice.

@mire NV provides no assurances that any reported problems may be resolved with the use of the documentation.

To request permission to reproduce materials, call +32 2 888 29 56, or +1 (858) 368-4681 email info@atmire.com, or write to @mire NV, Technologielaan 9, 3001 Heverlee, Belgium or 2888 Loker Avenue East, Suite 305, Carlsbad, CA 92010, U.S.A.

Table of Contents

Introduction

This document contains the administrator documentation required to use and configure @mire's Content & Usage Analysis module. It contains three major sections describing the different aspects of the CUA module. The first section will explain how to use the *Content Statistics* component of the module and how to configure this component to suit your requirements. The second section describes how to use and configure the *Usage Statistics* component. The final section describes the configuration and creation of different *Statlets* which are the statistical components that can be integrated into several pages of a DSpace repository.

Content Statistics

This component of the Content & Usage Analysis module allows you to analyse the the number of items in your repository on a number of criteria (called datasets). For example the default analysis is to analyse the growth of your repository content over time, the two default datasets are Time at one hand, and Communities & Collections at the other hand.

- **Time** allows you to select the items you want to include in your analysis, based on the value in their `dc.date.issued` metadata field.
- **Communities & Collections** allows you to limit your selection to items in a specific collection or community. When selecting a community, you are actually selecting the whole set of sub-communities and collections underneath this community. When you select a specific collections, items that are mapped from other collections are included in the counts. When the counts are aggregated on the level of a community, items mapped into one of the contained collections are excluded.

The 2 datasets above are the default datasets, for other types of analyses there are more datasets selectable and configurable for repository content analysis, the other available datasets (without any further configuration) are: Type, Language, and Date Accessioned.

Configuring additional datasets

A powerful way to further customize the possible ways of analyzing your repository content, is by defining additional metadata fields you want to be able to select as datasets.
For example, if you have a metadata value indicating the type of a document, this might be a relevant property for a dataset, because it will allow you to analyse the growth over time, for all the distinct document types.

To configure additional datasets, please refer to the Client Customization Documentation for this module.

The results

After the data source selection, you will be able to generate a Data Table and different sets of graphs, that both visually represent the data you selected.

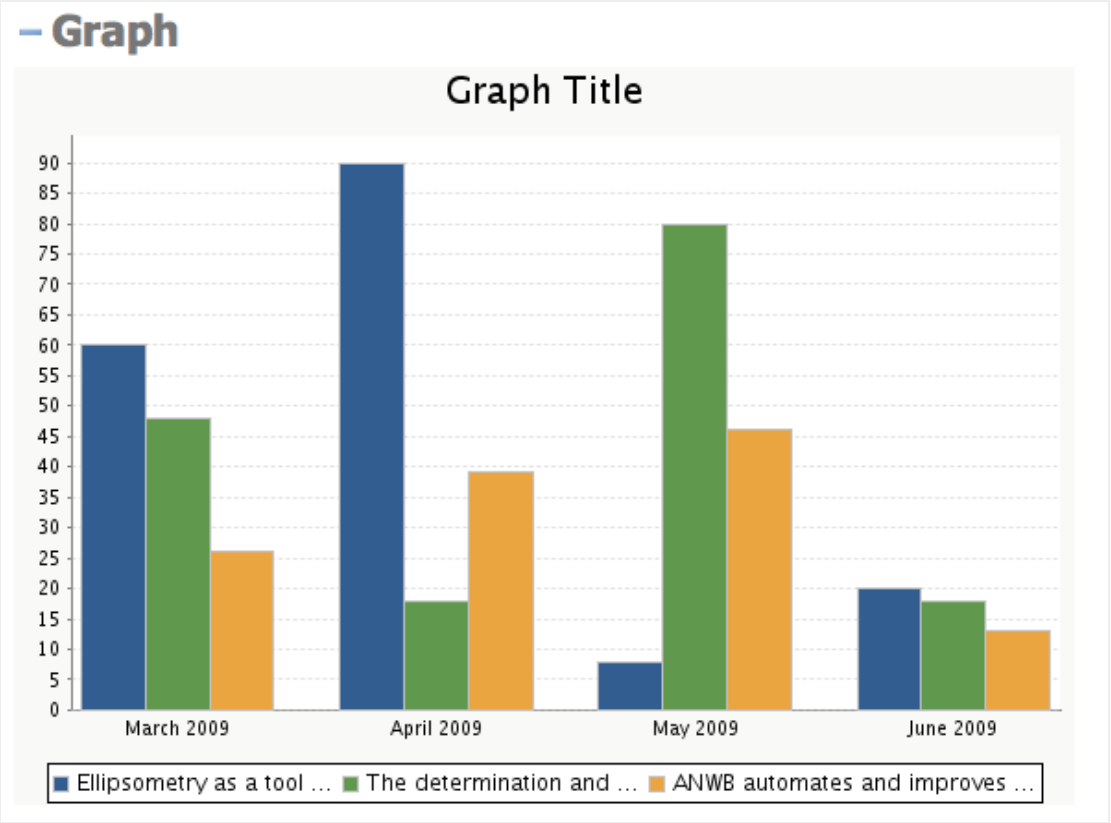
- **Data Table**: A table containing the data retrieved.
- **Graph**: A graphical representation of the data table.

The Data Table

This view shows the results of the analysis in a table. You can click on any cell to change the name or value. By clicking the '*Swap rows & columns*' button you can flip the rows & columns so that in this example the months will be seen as a rows and the items will be shown as columns. #
You can also export your data table as a spreadsheet in XLS format (that can be viewed in Excel and other spreadsheet editors), this can be done by either clicking on the '*Export table as XLS file*' link or the Icon.

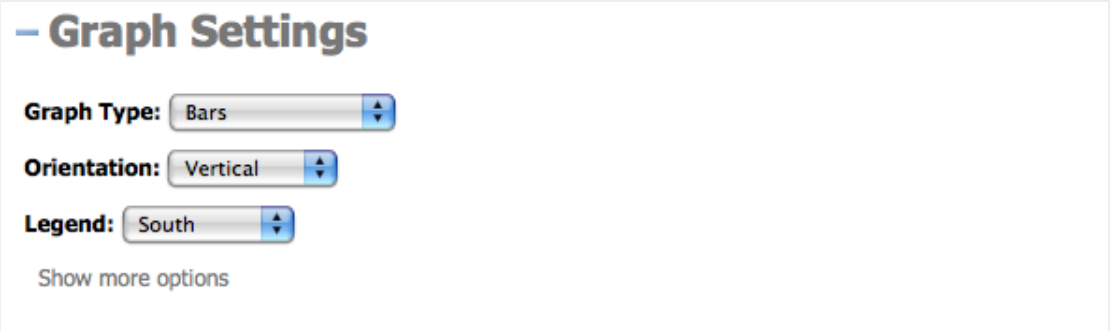
The Graph

The graph is a graphical representation of the data table. When the data table is modified, the graph gets updated asynchronously.



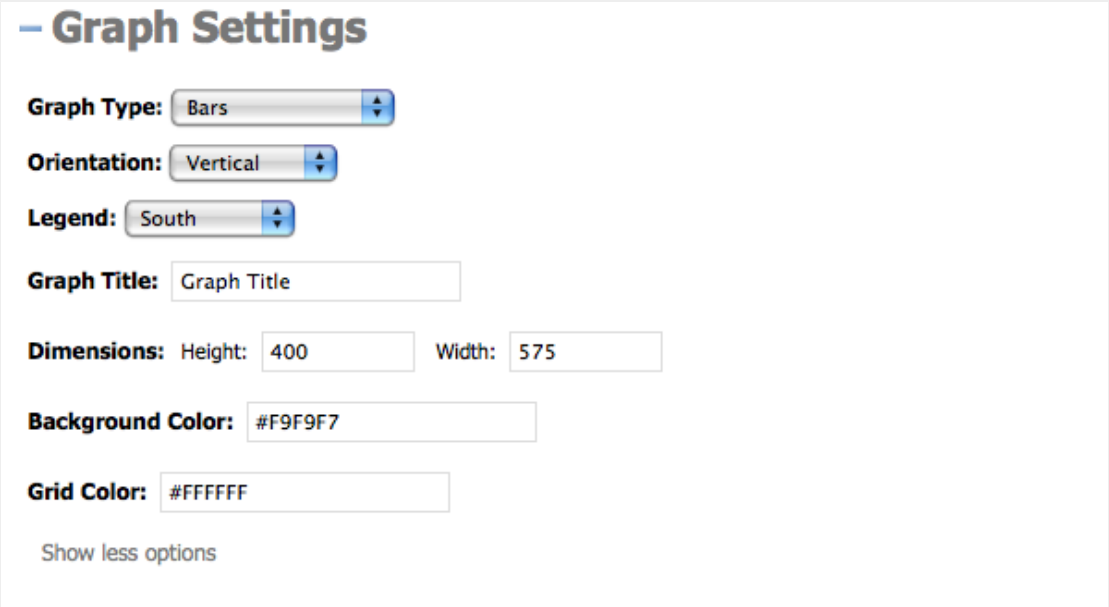
Graphical representation of the data table

By altering the “Graph Settings”, you can alter the different visual graph properties.



Graph settings

#By clicking on show more options you will be able to view all the options (as shown below).



Advanced graph settings

- **#Graph Type**: Here you can select the type of graph, the available types are Bars, Stacked Bars, Pie, Line, ... Keep in mind that not every graph is suitable for your data. Based on the data you want to analyse and visualize, some graph representations will be more relevant than other ones. For example, a pie chart is not very useful to represent the growth of the repository contents over time, although it's useful to illustrate for instance the number of items per type.
- **Orientation**: The orientation of your graph can either be vertical or horizontal.
- **Legend**: The location of the graph legend. By default this is South but this can also be: North, East, South or West.
- **Graph Title**: The title of the graph.
- **Dimensions**: The height & width of the graph.
- **Background Color**: The background color of the graph. Please be aware that the color has to be entered in RGB hexadecimal format.
- **Grid Color**: The color of the grid. Please be aware that just as the background color this has to be entered in RGB hexadecimal format.

Usage Statistics

The usage statistics allow an administrator to analyze the usage of the repository. The Usage Statistics component of the Content & Usage module captures data about the following usage events:

- Repository homepage visits
- Community homepage visits
- Collection homepage visits
- Item display visits
- Bitstream downloads

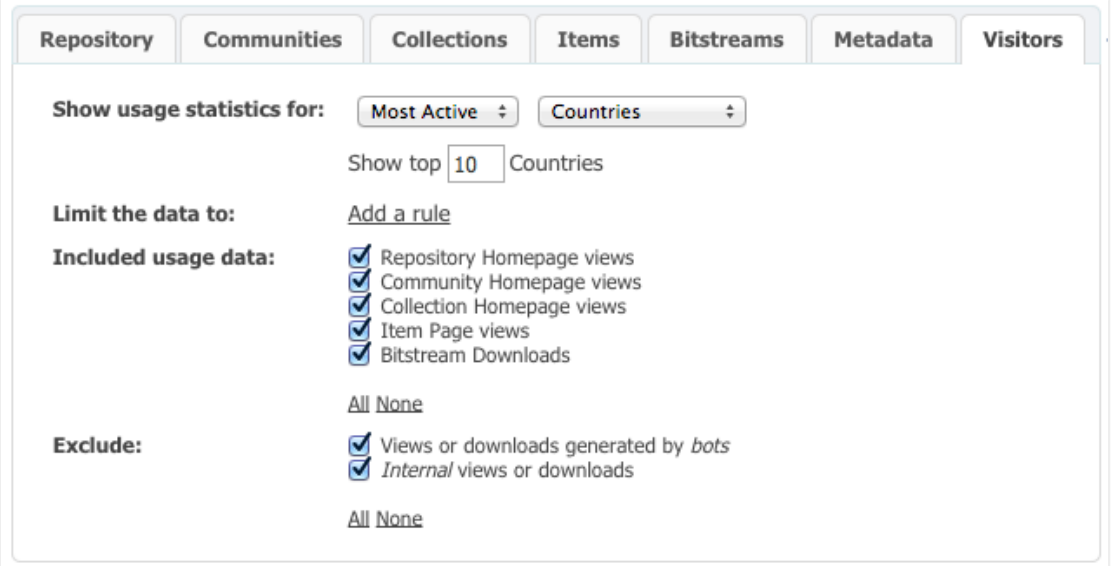
The Usage Statistics component allows you (as a DSpace administrator) to generate both data tables and graphs about the events described above. The administrator has the ability to select a number of different dataset options about which the data tables and graphs should be generated.

Scope

When creating data tables and graphs about usage data, the first choice that needs to be made is the selection of the DSpace content about which a usage analysis is requested. The administrator can choose between 5 different DSpace objects or by metadata, as depicted as tabs in the screenshot below, about which statistics can be generated:

- The repository
- Communities
- Collections
- Items
- Bitstreams
- Metadata

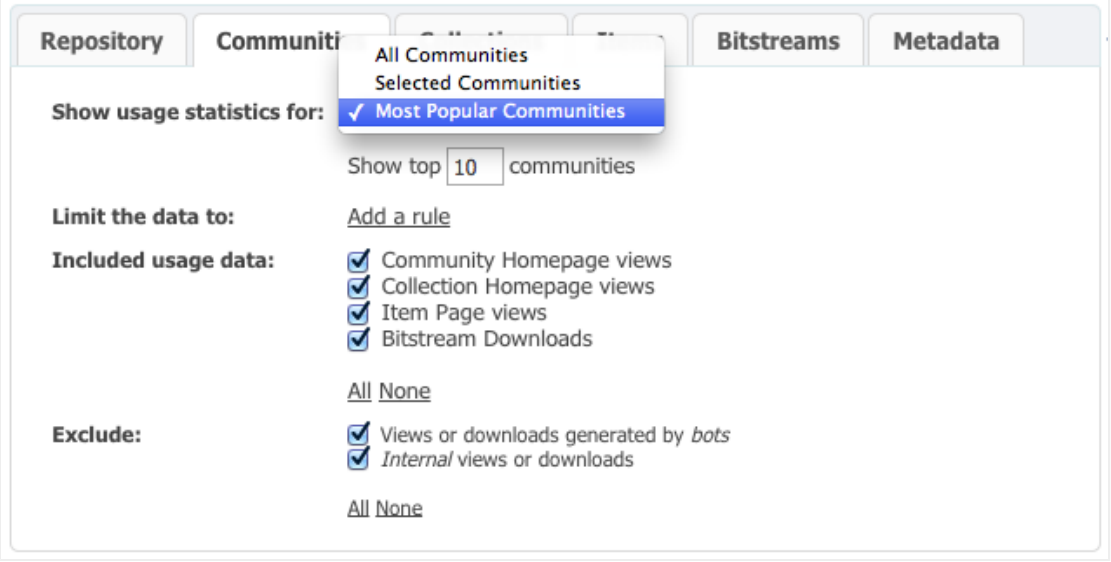
- Visitors



DSpace object type selection

For each scope, except for the DSpace repository, the administrator is able to further refine the selection to specific DSpace objects about which the statistics will be generated. For each of these types, for example the options for a Community object are the following:

- All instances of a certain DSpace object: e.g. all the communities in the repository.
- A selection of instances of a certain DSpace object: e.g. certain selected communities.
- The top objects of a certain type: e.g. the top 10 communities according to the selected usage events, which will be discussed in the next section.



DSpace object type selection - object selection

Limit the data to

For each scope, the administrator can specify filters to limit the data of the analysis based on usage event properties. These properties include the DSpace object, the DSpace object's metadata and visitor based properties. These filters are independent of the scope:

DSpace Objects

The analysis will only contain data from the selected DSpace objects. For each type of object only one row can be added. Possible DSpace types are:

- Communities

- Collections
- Items
- Bitstreams

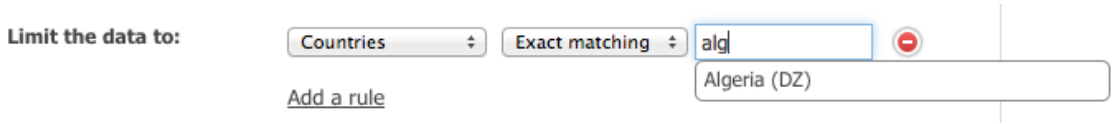
Metadata

The analysis will only contain data from DSpace objects that match the specified metadata rule. This rule can either be an *exactly matching* rule or a less strict *containing* rule.
When *exactly matching* is selected the search box will show an autocompletion menu. Multiple filter rules can be added, even for the same metadata field.
The metadata fields to limit data on are configurable.

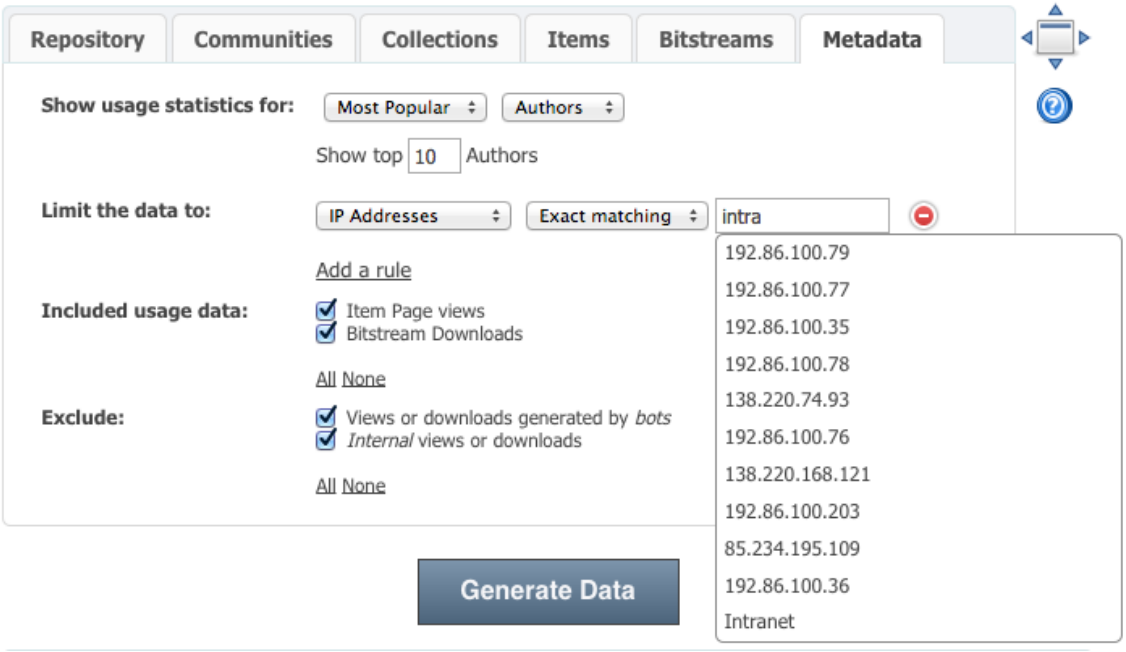
Visitors

The analysis will only contain data from DSpace objects that match the specified visitor's property rule. This rule can either be an *exactly matching* rule or a less strict *containing* rule.
When *exactly matching* is selected the search box will show an autocompletion menu. Multiple filter rules can be added, even for the same property.
Filtering is possible on the following visitor properties:

- Countries : The origin country of the visitor from the usage event. This rule expects a country code, but when *exactly matching* is selected complete names can be entered in the search box and the autocompletion menu will provide matching country codes by selecting one.



- IP Addresses : The IP address of the visitor from the usage event. Apart from normal IP addresses, descriptions from configured internal IP ranges or addresses are also supported.



- User Agents : The User Agent of the visitor from the usage event.
- HTTP Referrers : The HTTP Referrer from the usage event.

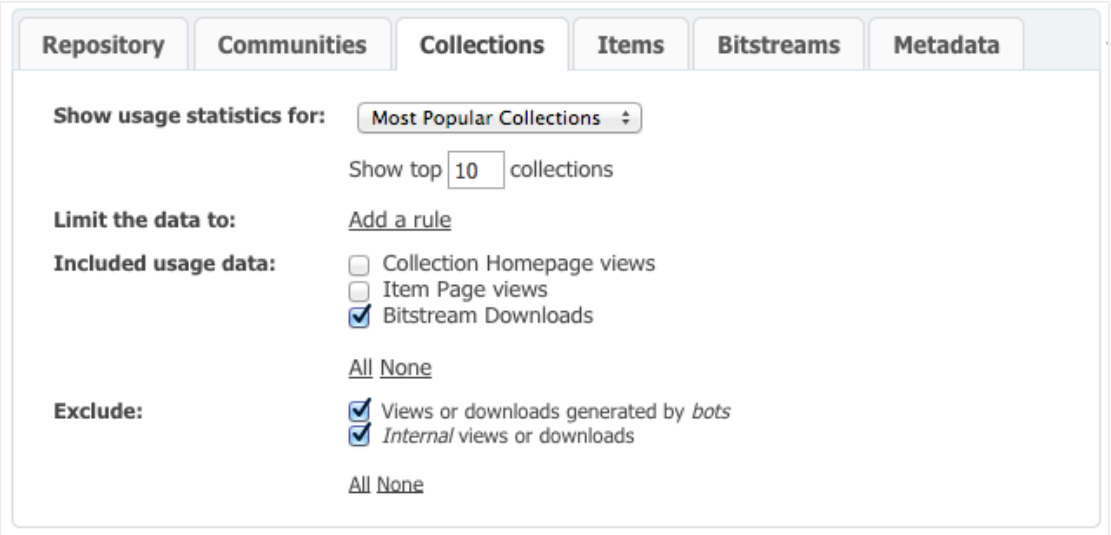
Included usage data

For each scope, the administrator can also select which types of usage events have to be taken into account when performing the usage analysis. Depending on the scope, the following types of usage events can be selected to be counted:

Repository	Community	Collection	Item	Bitstream	Metadata
Repository homepage visits	Community homepage visits	Collection homepage visits	Item page visits	Bitstream downloads	Item page visits

Community homepage visits	Collection homepage visits	Item page visits	Bitstream downloads		Bitstream downloads
Collection homepage visits	Item page visits	Bitstream downloads			
Item page visits	Bitstream downloads				
Bitstream downloads					

For example, if you want to analyze what are the top 10 collections according to the number of bitstream downloads for items in those collections, the settings need to be as depicted below:



Top 10 collections according to the number of bitstream downloads

Excludes

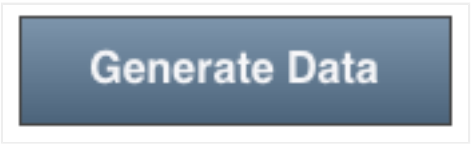
For each scope, the administrator can also select a number of exclude options based on the usage event properties to exclude from the analysis. These exclude options are independent of the scope:

- Views or downloads generated by bots: exclude usage event generated by bots from the analysis.
- Internal views or downloads: exclude usage event generated by internal IP ranges or addresses from the analysis. Internal IP ranges and addresses can be configured.

Exclude options

Generating graphs and data tables

In order to generate graphs and data tables about the different selected properties, the Generate button must be clicked. In case there already some graphs and data tables available, these will be updated according to the current requirements.



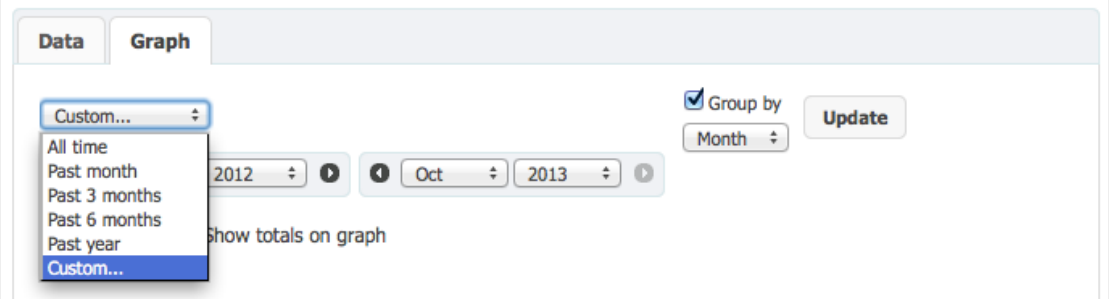
Generate button

Time filter

The time filter allows the administrator to select the time range for which the statistics should be generated. By default the time filter is set to show the number of selected usage events over the period of one year to date, grouped per month, as depicted in the image below.

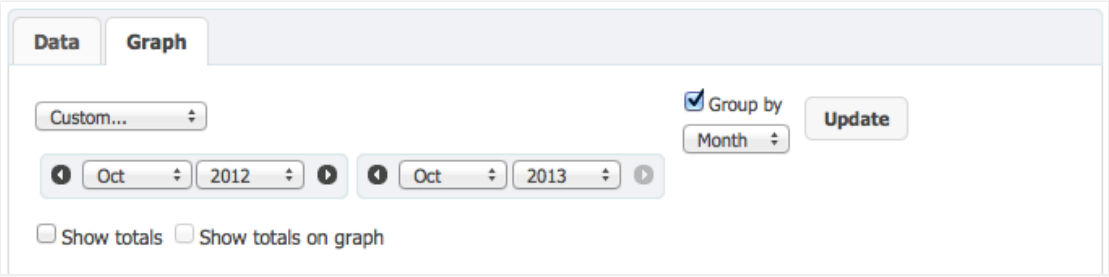
The user has the option of selecting a simple time filter or a custom time filter. By default the options *All time*, *1 month*, *3 months*, *6 months*, *one year* and custom are available. All time disables the time filter entirely. The other options limit the statistics for the specified period in the past starting from today. The custom option selects a precise timeframe for which statistics are generated.

The group by option will group the options by the specified unit so the evolution over time of the requested statistic can be seen.



Time filter options

Note that a three month period means the previous two months and the period between the first of this month and the current date (inclusive). E.G. when the page is viewed on Oct 23 the *3 months* period will include statistics from Aug 1 to Oct 23 and not Jul 23 to Oct 23. If the latter behaviour is required a custom date filter can be used as shown on the image below.



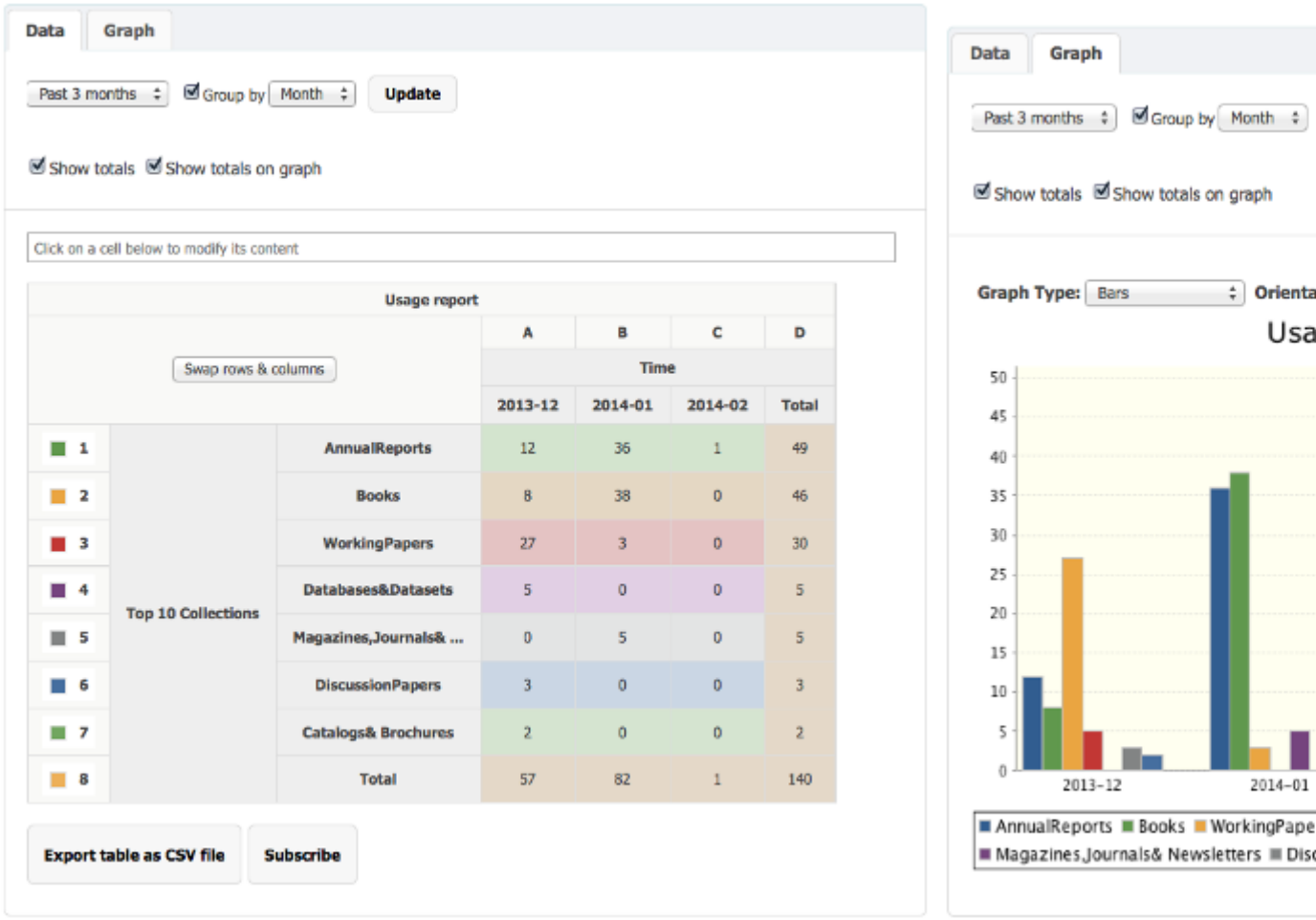
Custom time filter options

Totals

The check-boxes *Show Totals* and *Show totals on graph* control if the group totals will be shown in the data table and on the graph.

Show Totals will add the totals to the datatable view.
Show totals on graph requires *Show totals* to be active and will also show the row totals on the graph.

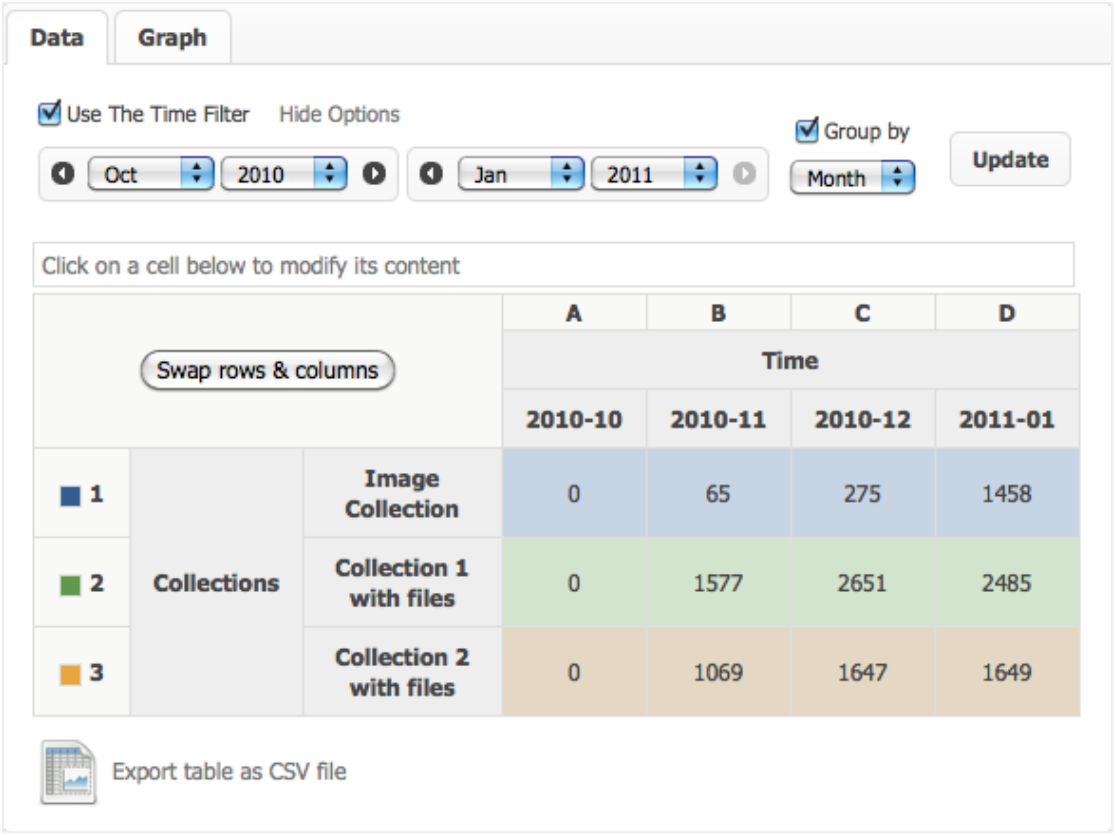
For example:



The table on the right shows usage statistics grouped by month for the last three months. The total block in the image on the right show the graph from that same table. The totals on the graph correspond with the rightmost column on the table.

Data tables

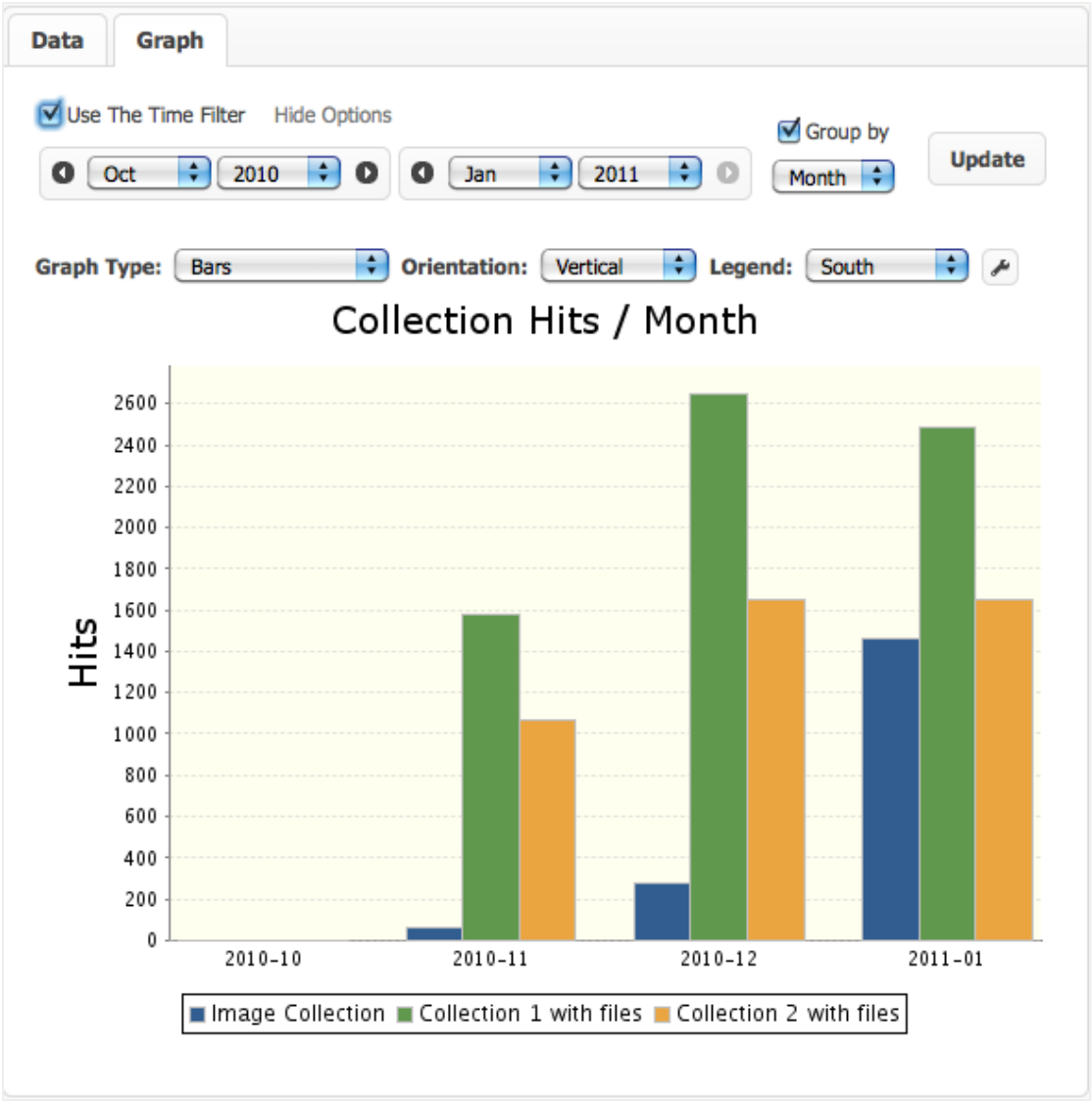
This view shows the results in table form. The administrator can click on any cell to change the name or value. By clicking the “Swap rows & columns” button it is possible to flip the rows & columns. You can also export your data table as a spreadsheet in XLS format (that can be viewed in Excel and other spreadsheet editors), this can be done by either clicking on the '*Export table as XLS file*' link or the Icon.



A data table representation of the data

Graphs

The graph is a graphical representation of the data table. When the data table is modified, the graph gets updated automatically.



The graphical representation of the data

By altering the “Graph Settings”, it is possible to alter the different visual graph properties, similar to the Content Statistics component:

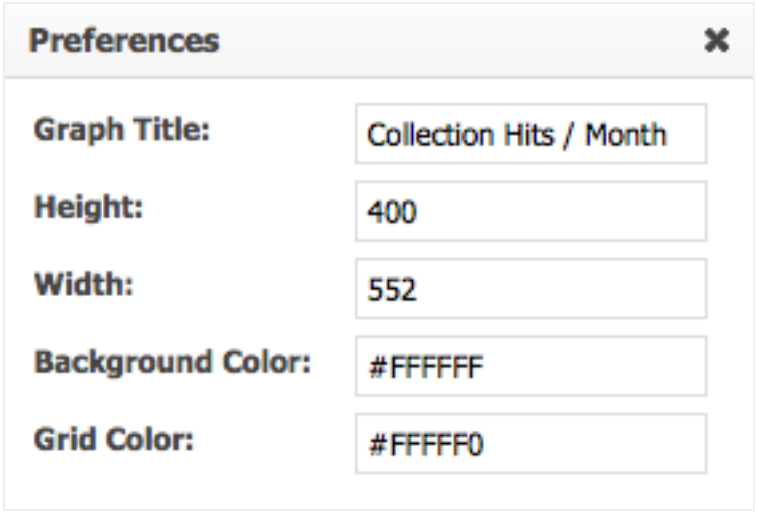
- **The graph type**: This property includes different type of graphs such as bars, pie charts, 3D chars, etc. Keep in mind that not every graph is suitable for your data. Based on the data you want to analyse and visualize, some graph representations will be more relevant than other ones.
- **Orientation**: The orientation makes it possible to create horizontal or vertical graphs.
- **Legend**: The legend option defines the position of the legend on the page.
- **Preferences**: By clicking the wrench icon, further graph preferences can be selected

Graph settings

The remaining graph preferences make it possible to further manipulate the appearance of the graph by offering the following additional properties:

- **Graph title**: The title of the graph.
- **Height**: The height of the graph.
- **Width**: The width of the graph.
- **Background color**: The background color of the graph. Please be aware that the color has to be entered in the RGB hexadecimal format.

- **Grid color**: The color of the grid. Please be aware that just as the background color this has to be entered in the RGB hexadecimal format.



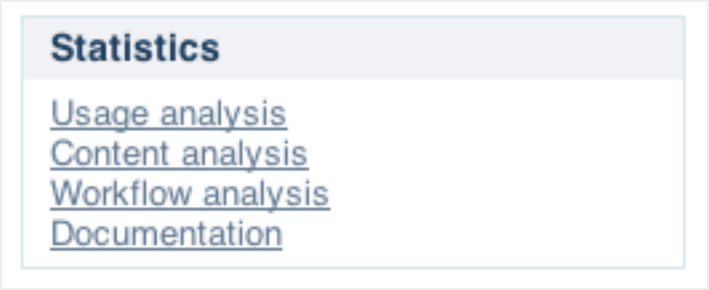
Additional graph preferences

Workflow Statistics

The workflow statistics component gives administrators an overview of the number of workflow actions in a given timespan or scope. A workflow action is anything a user can do to make an Item advance through the workflow: submitting an item, reviewing an item, etc.

Like the content and the usage statistics, the workflow statistics component allows administrators to generate tables and graphs using the data gathered about the workflow.

Administrators can open the workflow statistics component by clicking the link labeled 'Workflow statistics' in the Statistics section of the menu bar.

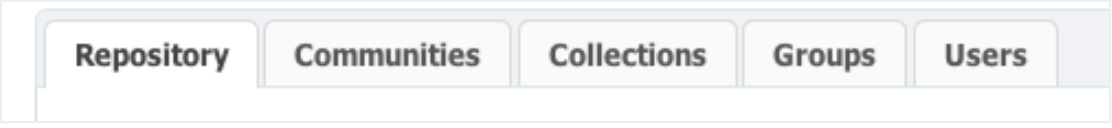


The statistics menu

Granularity & Scope

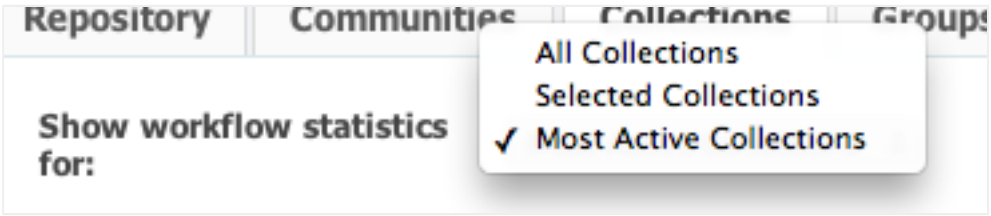
The first thing you need to do is decide on the level of detail of the report you want to generate. This is represented by the tabs in the top box. The name of the tab decides what the rows in your report will represent:

- **Repository**: there will be only one row in the report, with the data for the entire repository added up.
- **Communities**: rows represent the Communities in your repository.
- **Collections**: rows represent the Collections in your repository.
- **Groups**: rows represent groups of users.
- **Users**: rows represent individual users.



The tab decides the granularity of the report

For the categories 'Collections', 'Communities' and 'Users' you can choose between including all objects in that category, the *n* most active, or a custom selection. For 'Groups', you have to make a custom selection. See the section on configuring groups to determine which groups you can select and how they are presented. The 'Repository' category has no extra options

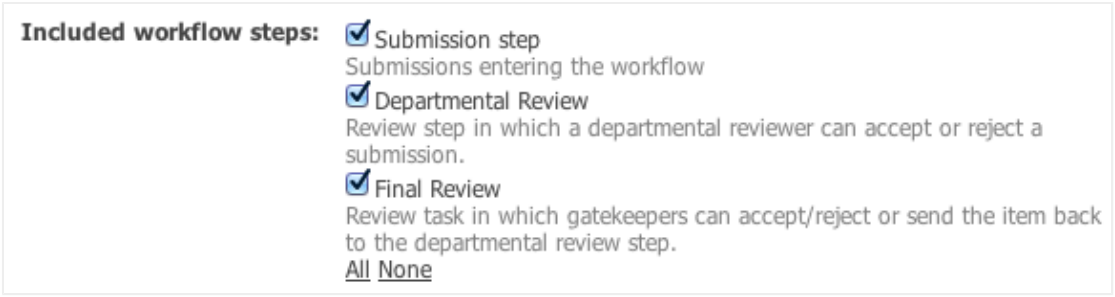


The scope options in the collection tab

Included Workflow Steps

After you've selected the granularity and the scope of your report, you can select which workflow steps you'd like to include. For every step you select here, the number of times it was executed will be included in the results.

So if, for example, you have two types of review steps, and you want to know how many times both combined were performed, you'd select both. But if you want to know how many submissions your reviewers have performed, you'd only select the submission step here.

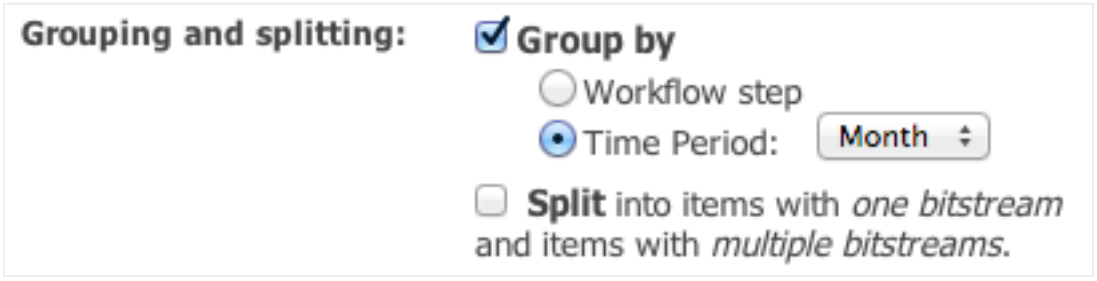


Select which workflow steps should be included in the report

See the section on configuring visible workflow steps for more information on how to configure which steps are shown.

Grouping and Splitting

Groups are the columns in your report. If you disable grouping there will be only one column. You can group the data by either workflow step or time period.



The grouping and splitting options

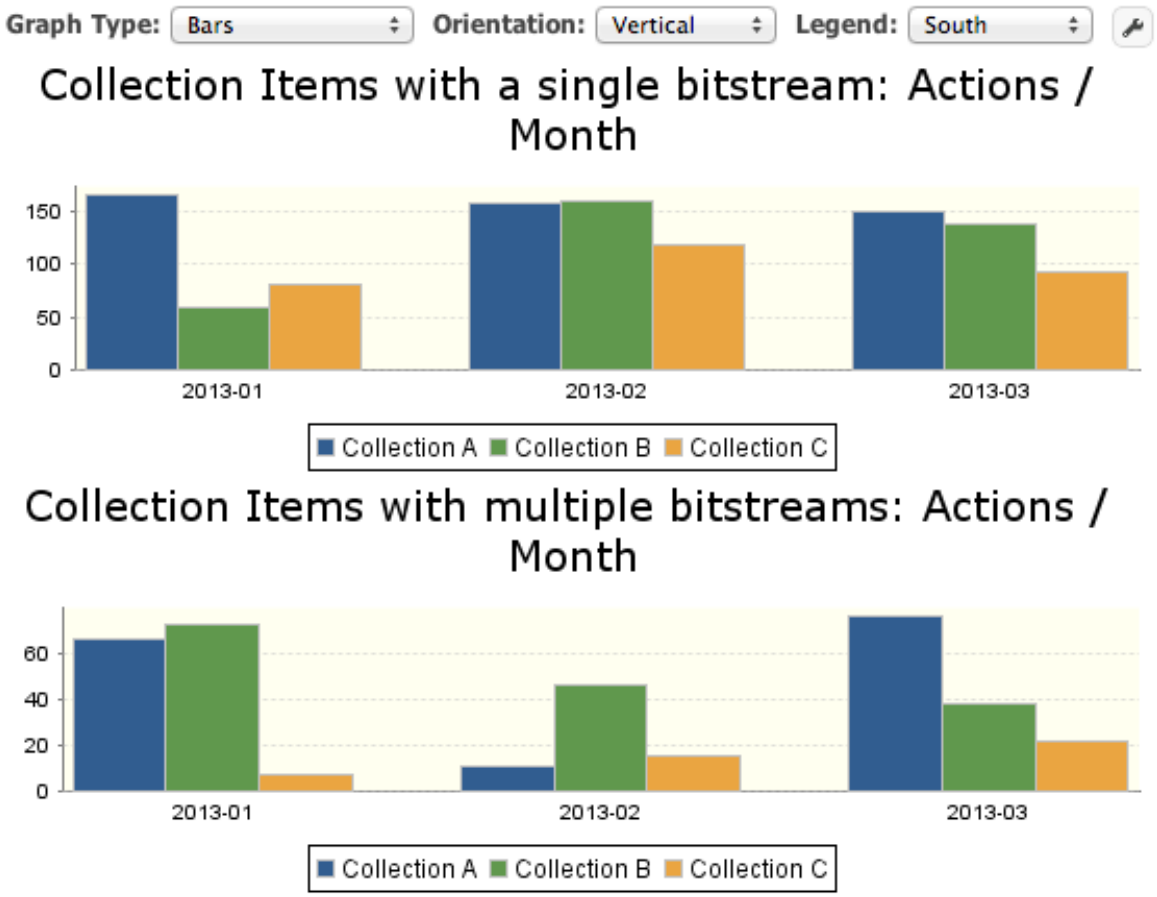
You can also choose to split the report into two tables or graphs, one for steps involving items with a single bitstream, and one for steps involving items with multiple bitstreams.

If you activate splitting the resulting data tables will look like this:

Collection Items with a single bitstream: Actions / Month					
Swap rows & columns			A	B	C
			Time		
			2013-01	2013-02	2013-03
1	All Collections	Collection A	165	158	149
2		Collection B	58	160	138
3		Collection C	81	118	92

Collection Items with multiple bitstreams: Actions / Month					
Swap rows & columns			A	B	C
			Time		
			2013-01	2013-02	2013-03
1	All Collections	Collection A	66	11	76
2		Collection B	73	46	38
3		Collection C	7	15	22

And the graphs will look like this:



Generating the Report

Generating reports, and working with the data-tables and graphs works almost exactly the same way as in the Usage Statistics component, so for more information take a look at that section

Configuring Visible Workflow Steps

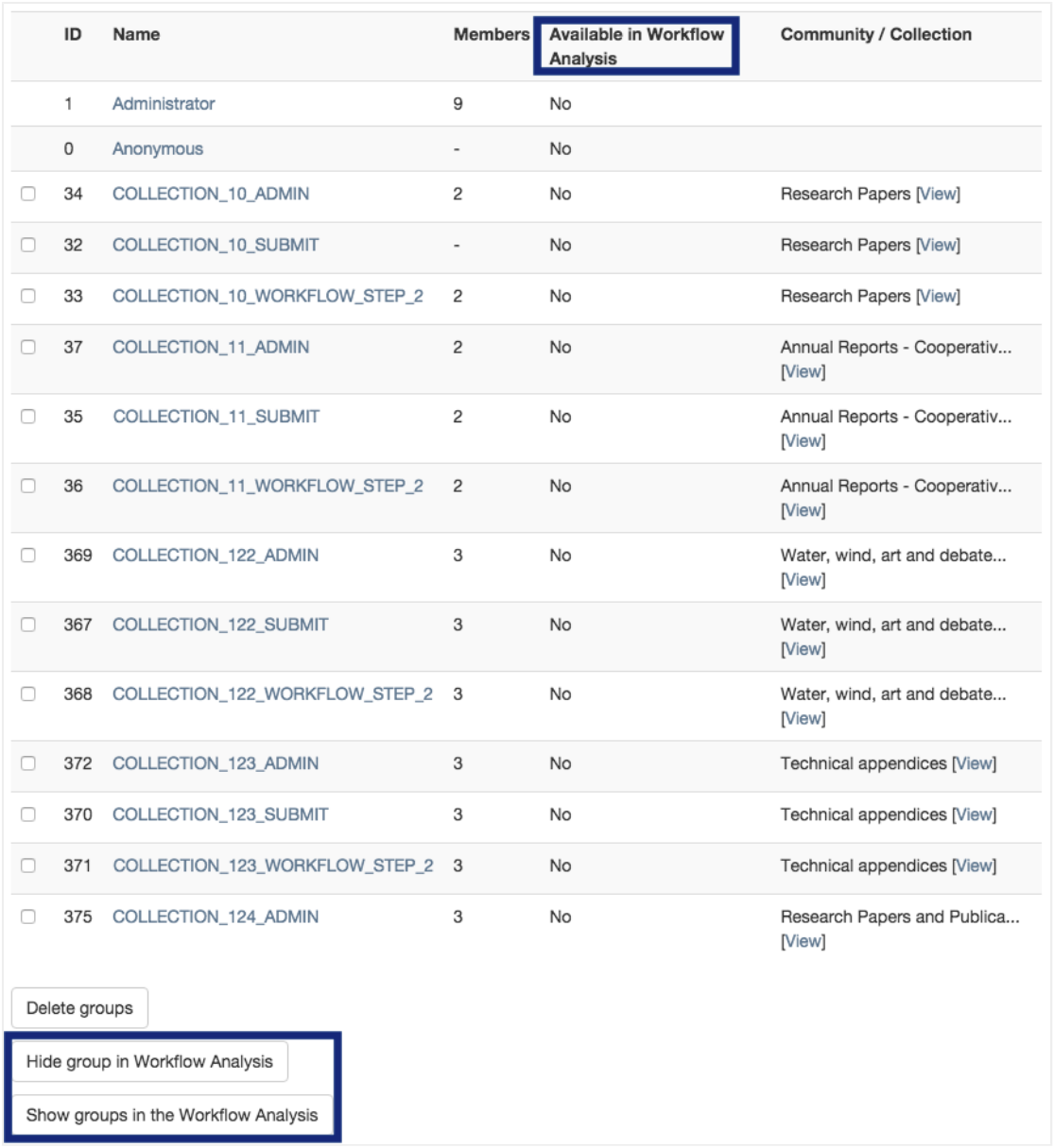
Because DSpace workflows can be complex, have hidden steps and differ from Collection to Collection, not all workflow steps are shown in the Workflow Statistics. Some aren't useful in a statistics report, and others need to be grouped together and shown in a report as one.

if you would like to alter the Workflow steps shown, please refer to the Client Customization Documentation for more information.

Configuring Groups

Because groups of users are used for more than just the workflow in DSpace and they often have confusing names, we don't show all groups in the Workflow Statistics. The groups that are shown are the ones that were manually selected by an administrator in the group management page.

After the cua was installed, some extra options were added to the group management page: A new Column 'Available in Workflow Analysis' was added, and two buttons 'Hide group in Workflow Analysis' and 'Show group in Workflow Analysis'.



The new group management page

If you check a checkbox in front of one or more groups and click 'Show groups in Workflow Analysis', they will become usable in the Workflow Analysis' Groups tab.

We advise you to create a bit of structure in your groups to make them easy to use. For instance, if you create a new group called 'All Submitters' and add all other automatically generated submission groups as members, you can make only this new group visible in the Workflow Analysis. If you do the same for reviewers it becomes very easy to compare the number of items being submitted to the number of items being reviewed within a given period of time.

That is just one suggestion. You should configure the groups in a way that best represents the structure of your workflow.

Clean up usage statistics

This functionality is only available from CUA 2.10.3

When communities/collection/items/bitstreams are deleted, the usage stats remain. To overcome this issue a script can be run to remove the statistics for all deleted objects. This script can be run by using the following command:

```
[dspace.dir]/bin/dspace dsrun org.dspace.statistics.util.CleanUpStatistics
```

Command line export

The CUA offers support to export statistics based on the usage. The command line script can be run by going to the `[dspace.dir]` directory and running `dsrun org.dspace.statistics.export.csv.StatisticsCSVExport`, the available parameters are the following (mandatory params are indicated in bold:

Parameter short	Parameter long	Explanation
-u	**--type**	The types of usage statistics (comma separated list), this property is required and can be only contain the following values: repository: 5, communities: 4, collections: 3, items: 2, bitstreams: 0
-l	--handle	Export the handle of the requested object in a separate column
-p	--parents	Export the parent handles as well as the object handle (true by default)
-m	--metadata	The metadata fields to be exported (a comma separated list) using the following format: {schema}.{element}.{qualifier} (only available for item & bitstream export
-c	--scope	The scope (handle of the parent object)
-s	--start	Start date, must be in the yyyy-MM-dd format (optional, if not configured, the first date found will be used)
-e	--end	End date, must be in the yyyy-MM-dd format (optional, if not configured, the current date is used)
-g	--group	Group by timeframe, can either be DAY/ MONTH/YEAR
-t	--total	Include the total counts true/false (true by default)
-o	**--output**	Output file
-h	--help	Display help

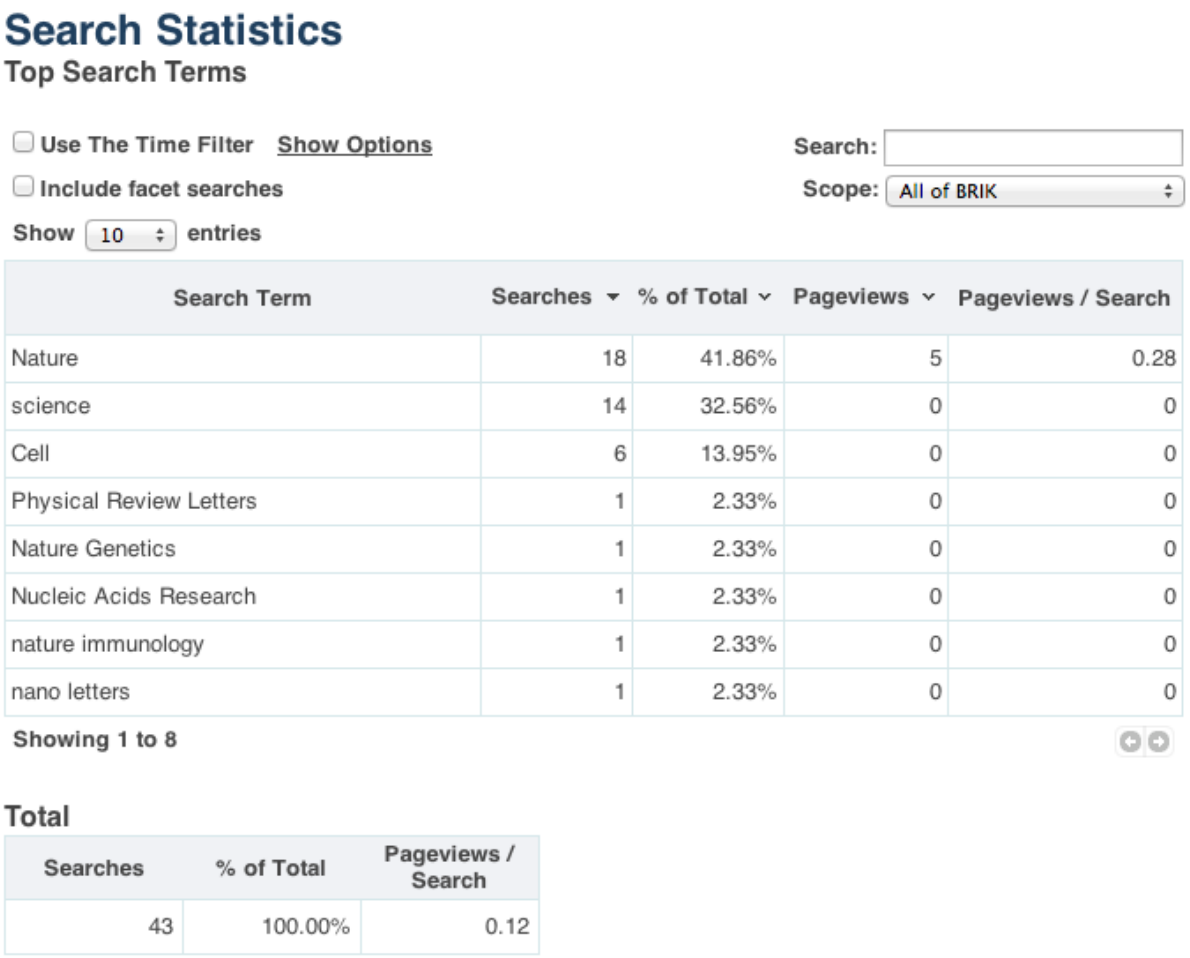
The -u option will always use the top object found when exporting, for example when exporting `-u 0,2`, item title and handle will be used and the bitstream downloads will be grouped by item. If for example `0,2,3` is exported the collections will take precedence and the item views & file downloads will be grouped by collection and so on.

An example of an export command can be found below:

```
dsrun org.dspace.statistics.export.csv.StatisticsCSVExport -u 2,0 -m
dc.contributor.author,dc.identifier.uri -t true -o {outputh.path}/output.csv -p true
-l true -g MONTH
```

Search Statistics

The Search Statistics page allows an administrator to analyze the used search queries in the repository.



Search Statistics

Top Search Terms

Use The Time Filter Show Options

Include facet searches

Search:

Scope: All of BRIK

Show 10 entries

Search Term	Searches	% of Total	Pageviews	Pageviews / Search
Nature	18	41.86%	5	0.28
science	14	32.56%	0	0
Cell	6	13.95%	0	0
Physical Review Letters	1	2.33%	0	0
Nature Genetics	1	2.33%	0	0
Nucleic Acids Research	1	2.33%	0	0
nature immunology	1	2.33%	0	0
nano letters	1	2.33%	0	0

Showing 1 to 8

Total

Searches	% of Total	Pageviews / Search
43	100.00%	0.12

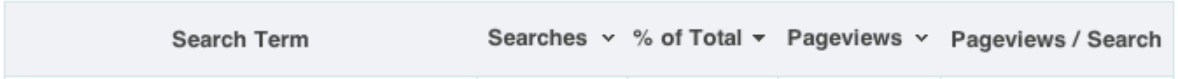
Facet Searches

Include facet searches

Facet queries can optionally be included in the results. If excluded, two combinations of identical queries with different facet queries will not influence the result. The query will get the sum of the amounts of searches and pageviews of both combinations. When facet searches are included, they will be displayed as two different result rows. Since a search can consist out of normal queries or facet queries, this option will also influence the totals table.

Sorting

The statistics can be sorted in three different ways : by number of searches, by percentage according to total amount of searches and by amount of resulting pageviews. Sorting by number of searches and by percentage will of course result in the same result.



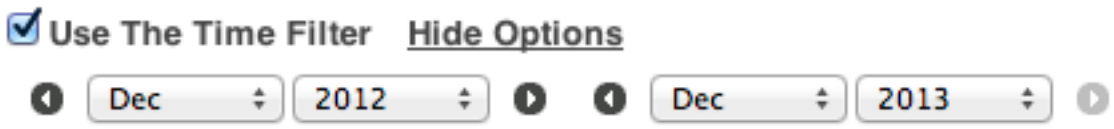
Search Term	Searches	% of Total	Pageviews	Pageviews / Search

Filter

To limit the number of displayed search statistics different filters can be applied. Note that these filters will influence the table that contains the totals summary.

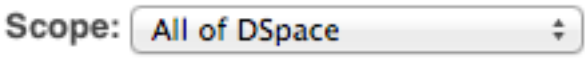
Time Filter

The time filter allows the administrator to select the time range for which the statistics should be generated. In case the "use time filter" checkbox is unchecked no time filter will be used, and all the available search statistics will be used. Otherwise, the selected range will be used to generate the statistics.



Scope Filter

Searches can be made within a certain community or collection. With the scope selector the returned statistics can be limited to searches made within that community or collection.



Keyword Filter

By entering a keyword in the search field, the returned statistics can be limited to searches that contained these keywords.

Storage Statistics

Update Storage Statistics

For performance reasons, the data that is used for the storage statistics isn't calculated in real time, but cached. To ensure that the cache is updated regularly, a script is configured to automatically run on a regular basis using a cron job.

Basic command

The command requires no parameters.

Basic command:
`[dspace]/bin/dspace update-storage-reports`

If this does not work, try running the CLI script directly:
`[dspace]/bin/dspace dsrun com.atmire.statistics.util.StorageReportsUpdateCLI`

MultipleBrowseCreateDAO

Since the script uses a database table that is only updated if the RDBMS BrowseCreateDAO implementation is used as browse system. A new BrowseCreateDAO, `org.dspace.browse.MultipleBrowseCreateDAO`, was implemented that can use multiple *BrowseCreateDAO* classes.

If an other class than `org.dspace.browse.BrowseCreateDAOPostgres` is used as BrowseCreateDAO implementation for the Browse Configuration, this configuration needs some modifications. By using `org.dspace.browse.MultipleBrowseCreateDAO` as BrowseCreateDAO implementation, multiple BrowseCreateDAO classes can be used. The configuration below is the one needed when the Solr Browse implementation is used. Notice that property `MultipleBrowseCreateDAOClass.1` should contain the class reference to the RDBMS BrowseCreateDAO implementation for your db.

Since DSpace 4.0, the Solr implementation is used by default.

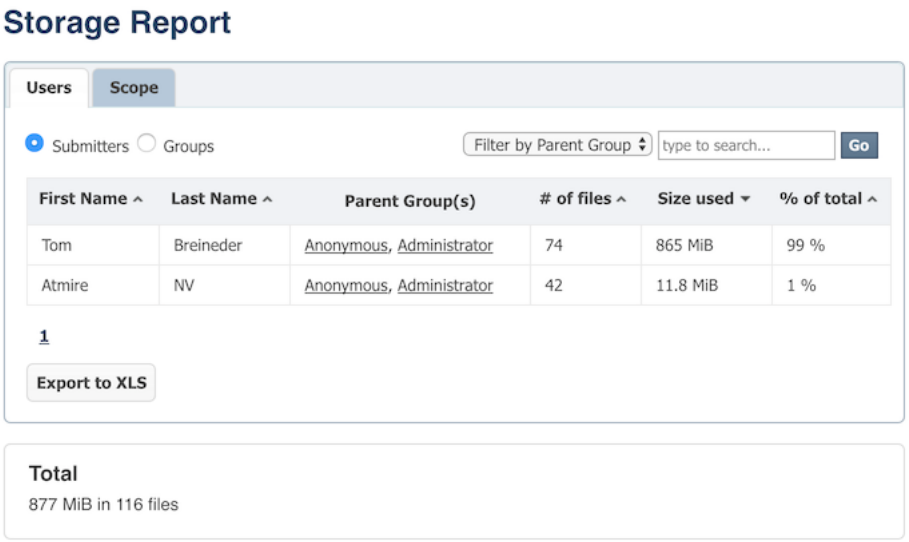
```
# Use Multiple DAO classes:
browseCreateDAO.class = org.dspace.browse.MultipleBrowseCreateDAO
# org.dspace.browse.BrowseCreateDAOPostgres is required for CUA.
```

```
MultipleBrowseCreateDAOClass.1 = org.dspace.browse.BrowseCreateDAOPostgres
MultipleBrowseCreateDAOClass.2 = org.dspace.browse.SolrBrowseCreateDAO
```

User Interface

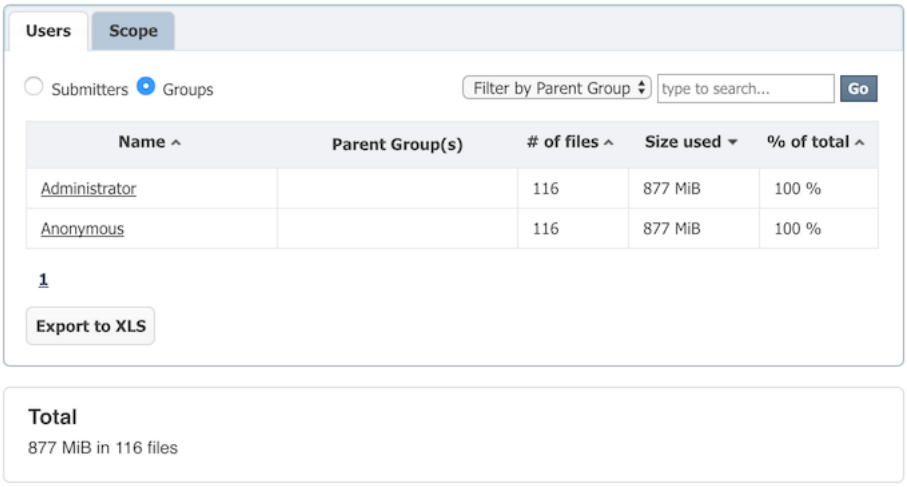
The storage report page supports four different views. The total box, and the '% of total' column are always about the entire repository, they don't take into account the filters or the search query.

Submitters View



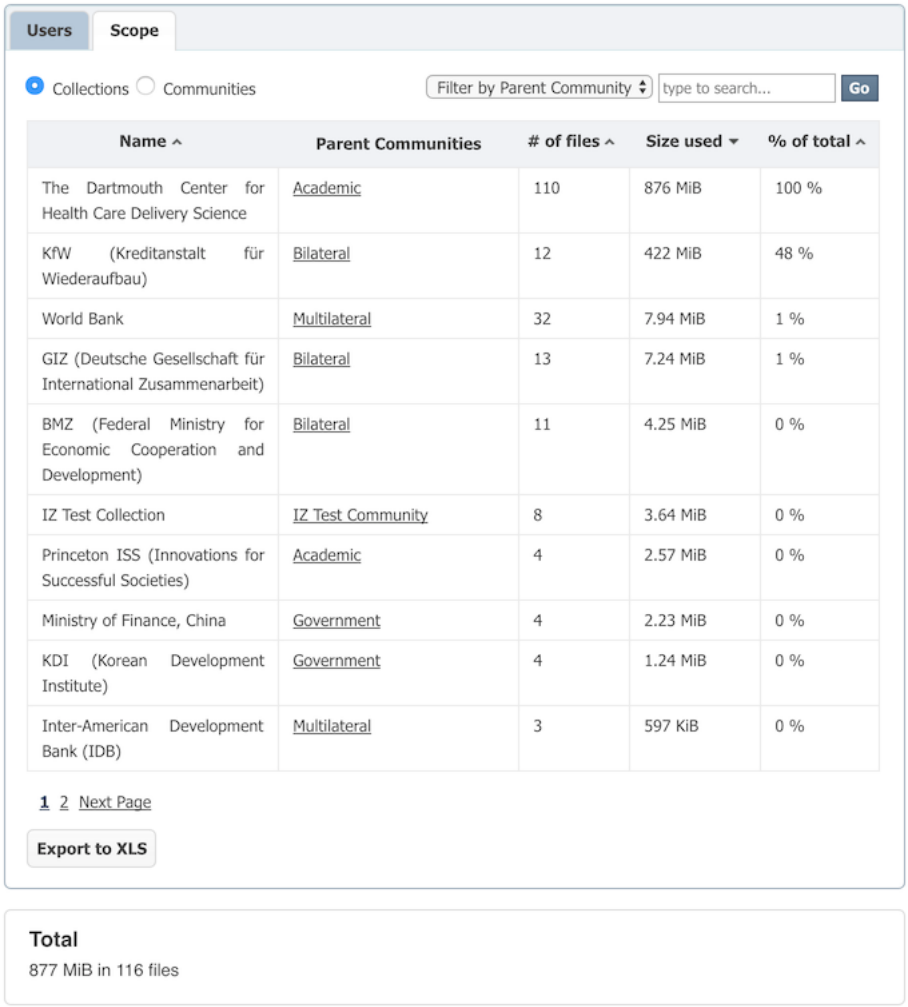
The search box searches for both submitter and group names take 'id:123' queries for user ids. All columns, except for the parent column are sortable. Clicking a link in the parent column switches to the groups view and searches for that group by id.

Groups View



The search box searches for both submitter and group names as well as take 'id:123' queries for group ids. Clicking a link in the parent column will do a search for that group by id, clicking a link in the name column will switch to submitter mode and filter by that group. Notice that groups may contain other groups and users can be part of multiple groups. In DSpace, every user is at least part of the Anonymous group and thus this group will always contain all files.

Collection View



Users | Scope

Collections | Communities | Filter by Parent Community | type to search... | Go

Name	Parent Communities	# of files	Size used	% of total
The Dartmouth Center for Health Care Delivery Science	Academic	110	876 MiB	100 %
KfW (Kreditanstalt für Wiederaufbau)	Bilateral	12	422 MiB	48 %
World Bank	Multilateral	32	7.94 MiB	1 %
GIZ (Deutsche Gesellschaft für International Zusammenarbeit)	Bilateral	13	7.24 MiB	1 %
BMZ (Federal Ministry for Economic Cooperation and Development)	Bilateral	11	4.25 MiB	0 %
IZ Test Collection	IZ Test Community	8	3.64 MiB	0 %
Princeton ISS (Innovations for Successful Societies)	Academic	4	2.57 MiB	0 %
Ministry of Finance, China	Government	4	2.23 MiB	0 %
KDI (Korean Development Institute)	Government	4	1.24 MiB	0 %
Inter-American Development Bank (IDB)	Multilateral	3	597 KiB	0 %

1 2 Next Page

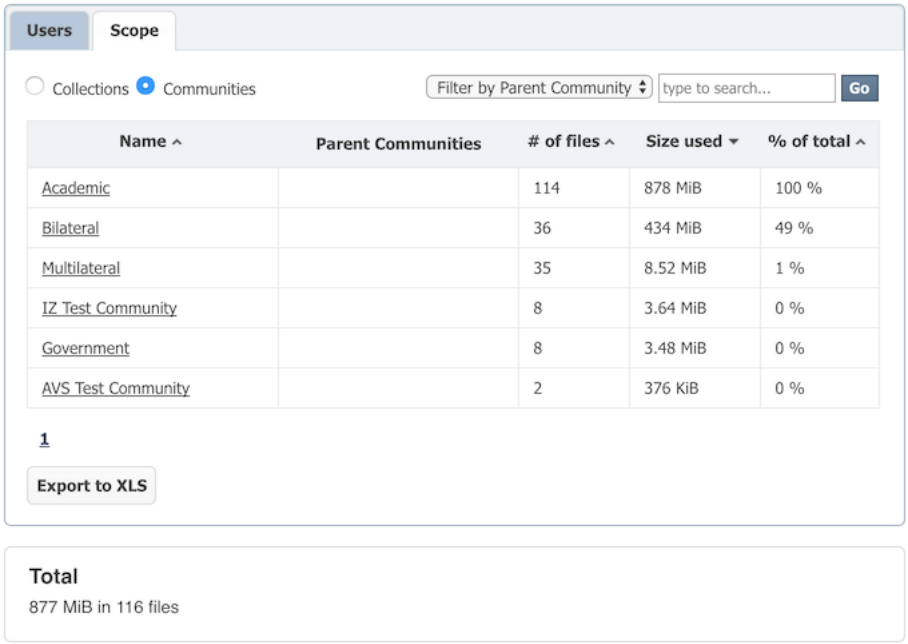
Export to XLS

Total

877 MiB in 116 files

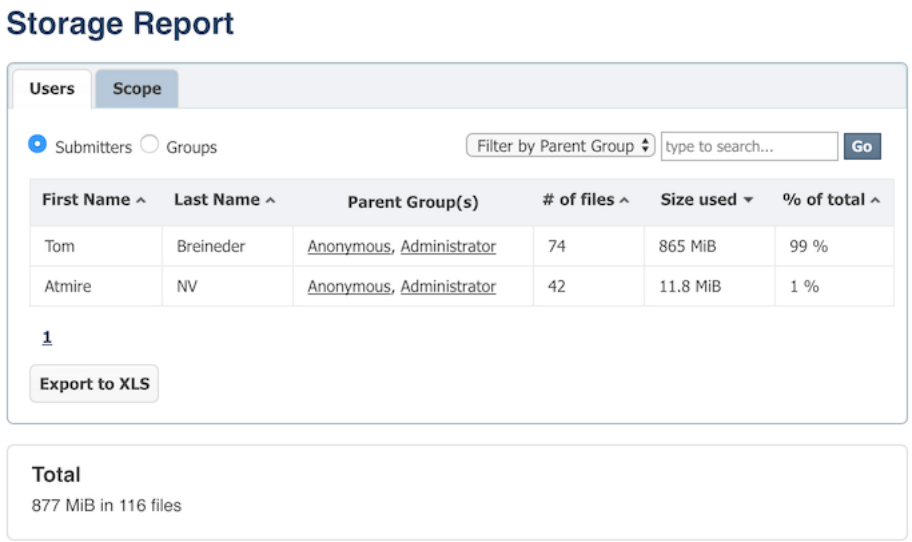
The search box searches for both collection and community names as well as take 'id:123' queries for collection ids. Clicking a link in the parent column will switch to community mode and do a search for that community by id. Notice that files from items that appear in multiple collections will be counted in both collections. The total summary will not be influenced by this.

Communities View



The search box searches for both collection and community names take 'id:123' queries for community ids. Clicking a link in the parent column will do a search for that community by id, clicking a link in the name column will switch to collection mode and filter by that community. Notice that files from items that appear in multiple collections from a community will only be counted once. Files from items that appear in multiple collections from different communities will be counted in both communities. The total summary will not be influenced by this.

XLS Export



Every storage report can be exported to an XLS file. The XLS will contain the same data as the webpage, but is not limited to a certain number of rows i.e. it does not take the pagination into account.

Config

The config can be found in `[dspace]/config/modules/atmire-cua.cfg`

```
storage.count.workspace.items=true
```

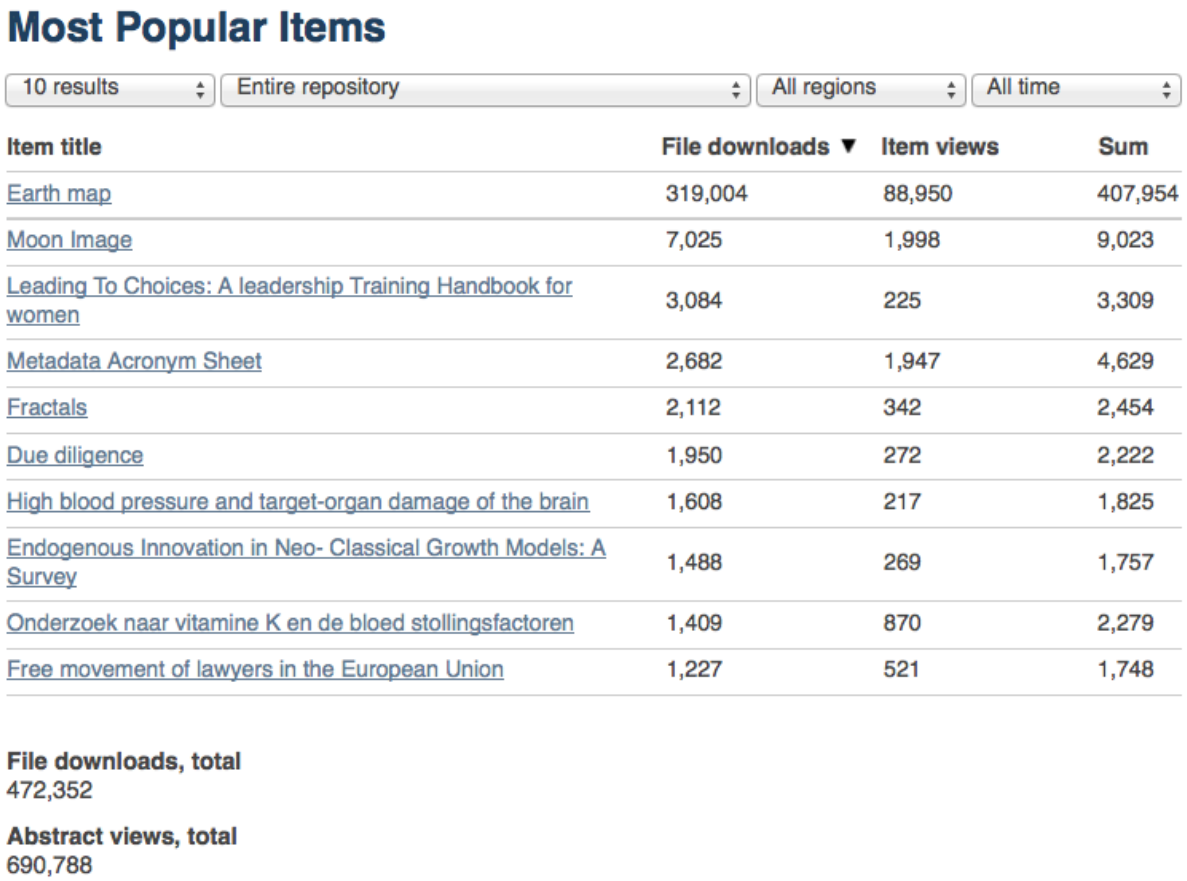
Defines whether the bitstreams from workspace items, i.e. unfinished submissions, should be taken into account or not.

```
storage.number.of.results.per.page=10
```

The number of results that are displayed.

Interactive statistics

The interactive statistics are a public feature of the CUA. They show the most popular items of the repository, the geographical break-down of repository visits or the breakdown in the most popular values of a specified metadata field.
Each of these tables can be further filtered on community or collection, on time span and geographical region. The links listed on these pages refer to the corresponding item pages.



Most Popular Items

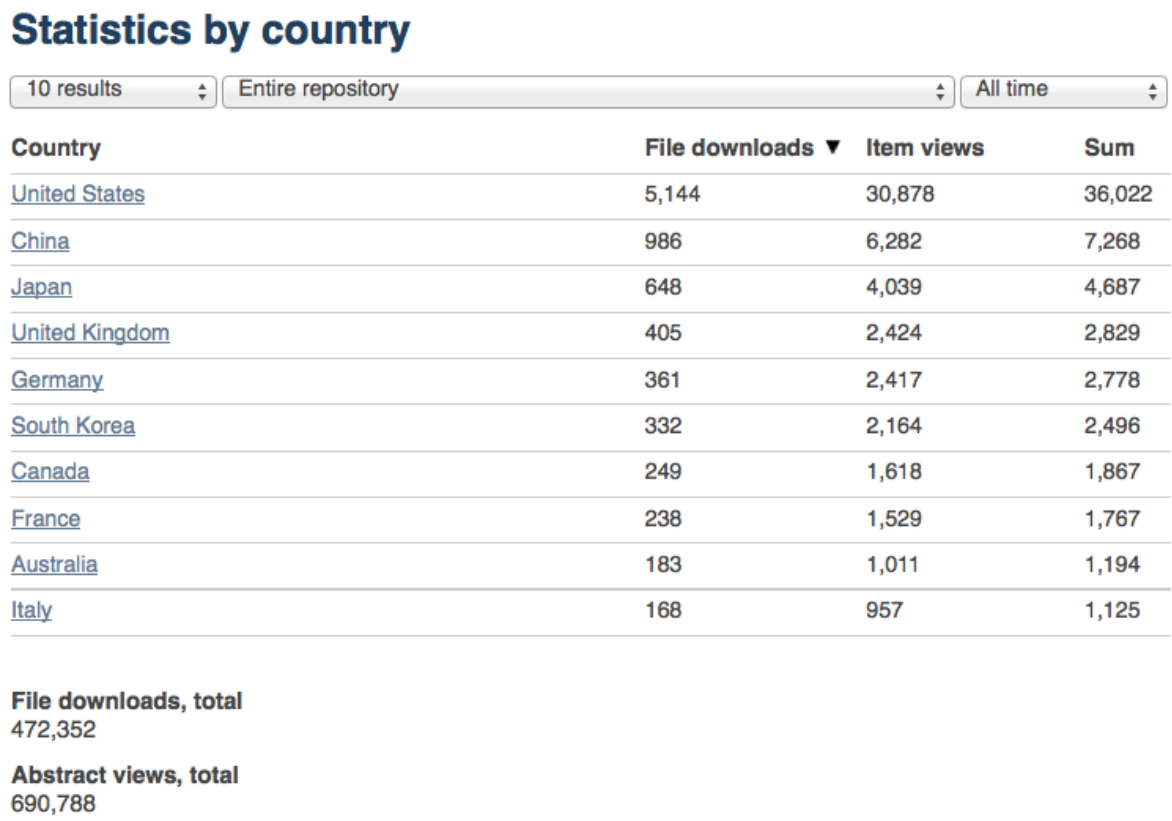
10 results | Entire repository | All regions | All time

Item title	File downloads ▼	Item views	Sum
Earth map	319,004	88,950	407,954
Moon Image	7,025	1,998	9,023
Leading To Choices: A leadership Training Handbook for women	3,084	225	3,309
Metadata Acronym Sheet	2,682	1,947	4,629
Fractals	2,112	342	2,454
Due diligence	1,950	272	2,222
High blood pressure and target-organ damage of the brain	1,608	217	1,825
Endogenous Innovation in Neo- Classical Growth Models: A Survey	1,488	269	1,757
Onderzoek naar vitamine K en de bloed stollingsfactoren	1,409	870	2,279
Free movement of lawyers in the European Union	1,227	521	1,748

File downloads, total
472,352

Abstract views, total
690,788

The above image shows the most popular items for the entire repository. The popularity is ranked by file downloads. The other figures in the table are the item's page views and the sum of both file downloads and page views for the item. Clicking any of the other columns will rank the popularity by that property. E.g. clicking on the 'item views' column header will show the most popular items by page views. It is entirely possible that a completely different set of items is shown.



Statistics by country

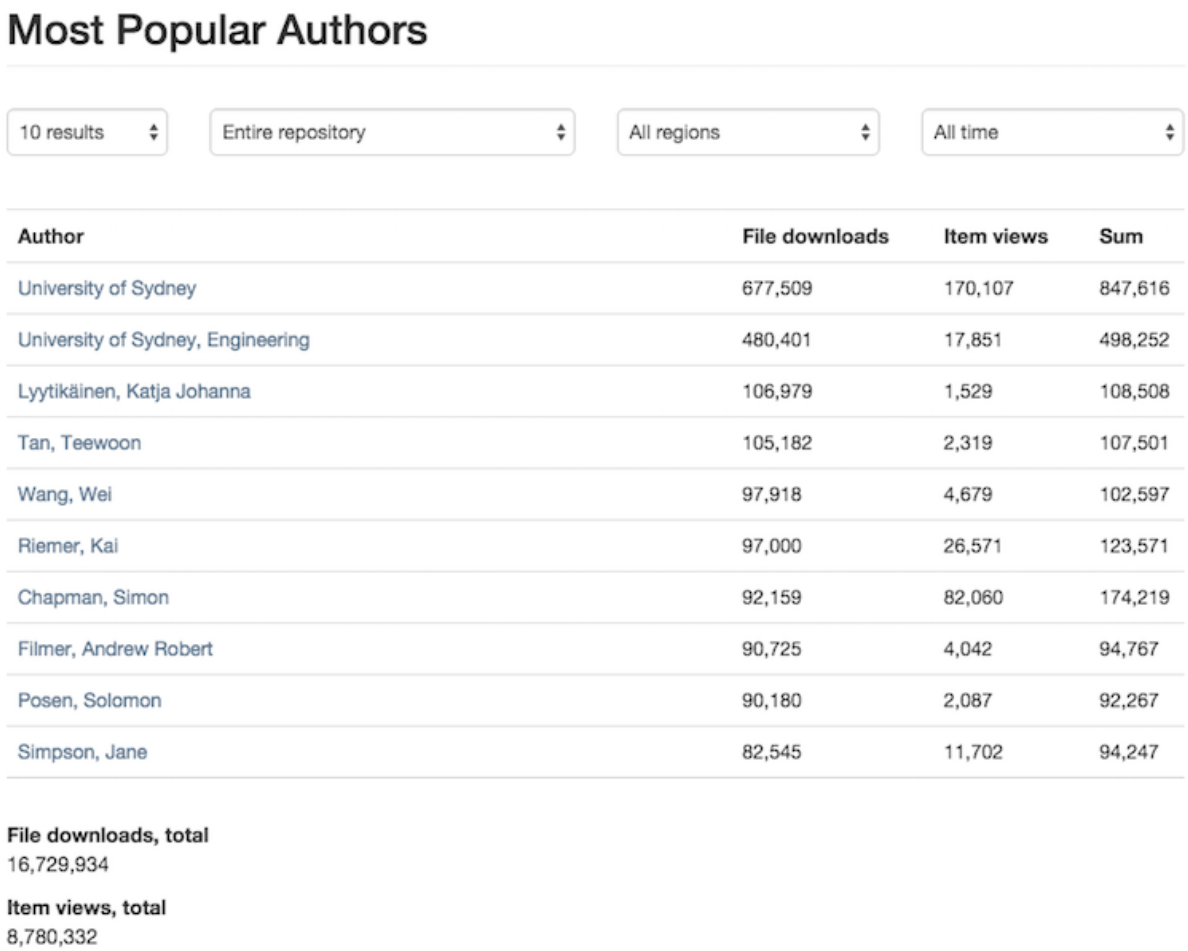
10 results | Entire repository | All time

Country	File downloads ▼	Item views	Sum
United States	5,144	30,878	36,022
China	986	6,282	7,268
Japan	648	4,039	4,687
United Kingdom	405	2,424	2,829
Germany	361	2,417	2,778
South Korea	332	2,164	2,496
Canada	249	1,618	1,867
France	238	1,529	1,767
Australia	183	1,011	1,194
Italy	168	957	1,125

File downloads, total
472,352

Abstract views, total
690,788

The above image shows the geographical breakdown of the statistics. The ranking mechanism is identical to the most popular item page. Clicking a country will refer to the most popular item page with the clicked country enabled as a filter.



The above image shows the most popular values of a certain metadata field, in this case the dc.contributor.author field. This kind of interactive page can be created for each indexed metadata field. It has the same sorting and filtering options as the 'Most Popular Items' page. Clicking the links in the list will show the browse page corresponding to that particular metadata value.

Scheduled Email reports

Introduction

Each of the aforementioned tools for Content, Usage and Workflow statistics allow an administrator to subscribe email addresses to reports that your DSpace will send out at regular intervals. After designing a suitable report that contains data for the past month, week, quarter or year, an administrator can subscribe email addresses by clicking “Subscribe” below the data table screen. This brings up a dialog where the administrator can define the frequency of the report, a title and the list of recipients. These email addresses do not need to correspond with DSpace user
accounts.

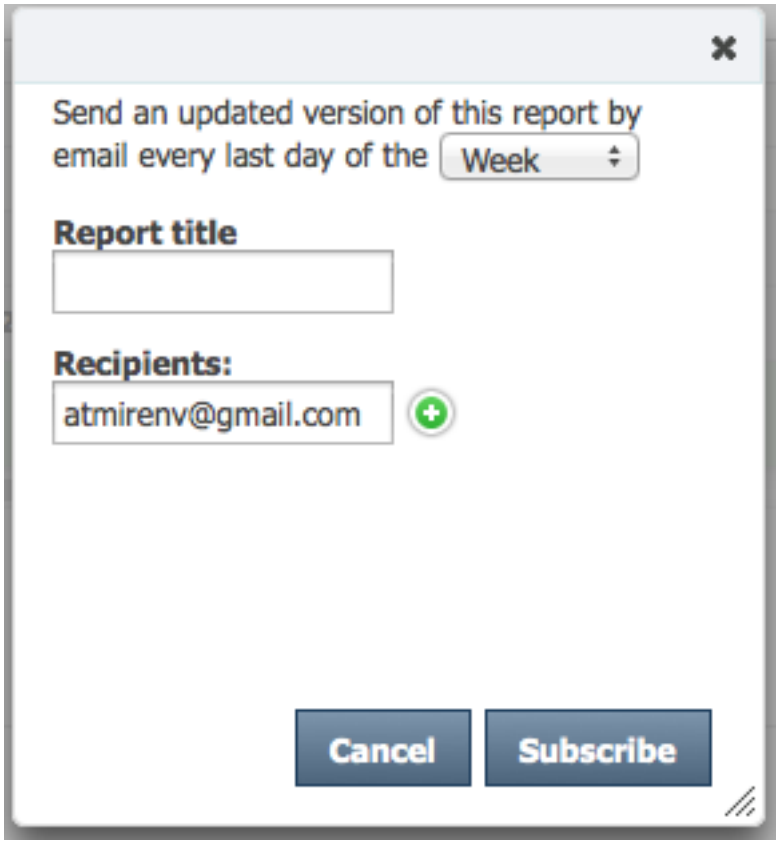
The report emails that are sent out based on this configuration will contain both the PNG of the configured graph and the data table represented in a XLS file. Recipients can unsubscribe through a link in the report email.

The template text of an email report is treated like any other email template in DSpace: the text is isolated in a single file that can customized.

Configuration

When a relative time filter is selected, one can subscribe to a periodic report of that period in the future. In the data tab a subscribe button is shown.

Clicking this button will open the following form.



Submitting the form will create a subscription specifying how often the report is sent: daily, weekly, monthly, quarterly or yearly. The report is always sent on the last day of the specified period. Note that it is possibly to receive a yearly report going back one day: the statistics of Dec 31 of that year.

Any email address can be entered. By default the current user's e-mail address is entered. It is also possible to give the report a name. A sample of the report is immediately sent to all addresses to allow the recipient to unsubscribe immediately.

Sending the reports

A cron job running each night needs to be set to send today's reports. The script to run is *com.atmire.utils.ReportSender*.

Statlets

The graphs and tables that can be displayed on the following DSpace pages are called *statlets*:

- The repository homepage
- The community homepages
- The collection homepages
- The item display pages

After a default installation of the module, following statlets are available:

Repository homepage

- GRAPH A donut chart depicting the total amount of all time page views and bitstream downloads for the entire repository.
- TABLE The most popular communities of the past six months based on page views.
- TABLE The most popular communities of the past six months based on bitstream downloads.
- GRAPH A donut chart depicting the 5 countries generating the most item views in the past six months.
- GRAPH A donut chart depicting the 5 countries generating the most bitstream downloads in the past six months.
- TABLE The most popular authors of the past six months based on page views.
- TABLE The most popular authors of the past six months based on bitstream downloads.
- TABLE The most popular items of the past six months based on page views.
- TABLE The most popular items of the past six months based on bitstream downloads.
- GRAPH A line graph depicting the item views by month for the past 6 months.
- GRAPH A line graph depicting the bitstream downloads by month for the past 6 months.

Community homepage

- GRAPH A donut chart depicting the total amount of all time page views and bitstream downloads for that community.
- TABLE The most popular authors of that community for the past six months based on page views.
- TABLE The most popular authors of that community for the past six months based on bitstream downloads.
- GRAPH A donut chart depicting the 5 countries generating the most item views in the past six months for that community.
- GRAPH A donut chart depicting the 5 countries generating the most bitstream downloads in the past six months for that community.
- TABLE The most popular items of that community for the past six months based on page views.
- TABLE The most popular items of that community for the past six months based on bitstream downloads.
- GRAPH A line graph depicting the item views of that community by month for the past 6 months.
- GRAPH A line graph depicting the bitstream downloads of that community by month for the past 6 months.

Collection homepage

- GRAPH A donut chart depicting the total amount of all time page views and bitstream downloads for that collection.
- TABLE The most popular authors of that collection for the past six months based on page views.
- TABLE The most popular authors of that collection for the past six months based on bitstream downloads.
- GRAPH A donut chart depicting the 5 countries generating the most item views in the past six months for that collection.
- GRAPH A donut chart depicting the 5 countries generating the most bitstream downloads in the past six months for that collection.
- TABLE The most popular items of that collection for the past six months based on page views.
- TABLE The most popular items of that collection for the past six months based on bitstream downloads.
- GRAPH A line graph depicting the item views of that collection by month for the past 6 months.
- GRAPH A line graph depicting the bitstream downloads of that collection by month for the past 6 months.

Item page

- GRAPH A donut chart depicting the total amount of all time page views and bitstream downloads for that item.
- TABLE The 5 countries generating most item page views for that item in the past 6 months.
- TABLE The 5 countries generating most bitstream downloads for that item in the past 6 months.
- GRAPH A line graph depicting the item views of that item by month for the past 6 months.
- GRAPH A line graph depicting the bitstream downloads of that item by month for the past 6 months.
- GRAPH The 5 most popular works by the item's first author based on item views.
- GRAPH The 5 most popular works by the item's first author based on bitstream downloads.

Statlets are only shown when there is actual statistics data to present. For a fresh installation of DSpace, you need to wait until the first usage hits are registered before the statlets appear.

Configuring statlets

The different statlets that can appear on these pages can be completely configured by altering the `statistics.xml` configuration file which you can find in the DSpace config directory.

For more information on configuring statlets, please refer to the Client Customization Documentation for this module.

atmire-cua.cfg

```
#------------------------------------------------------------------#
#-----------------Content & Usage Analysis-----------------------#
#------------------------------------------------------------------#
#From which date do we want our datepicker to start from ?
statistics.default.start.datepick = 01/01/2000
```

Some fields in the content statistics use a datepicker.
The timeframe of that datepicker goes from a start date configured here untill the current date.
E.g. on the *Content statistics* page, when *Date accessioned* is selected as Primary Dataset.

```
#Which dc fields shall be used from the input-forms.xml to generate item-statistics
statistics.items.dc.1=dc.type
statistics.items.dc.2=dc.language.iso
statistics.items.dc.3=dc.date.accessioned
#Type can either be dcinput or date
statistics.items.type.1=dcinput
```

```
statistics.items.type.2=dcinput
statistics.items.type.3=date
```

The list of items configured here are added to the dropdown list of the *Primary Dataset* on the *Content statistics* page.
It's necessary to specify for each item whether the input needs to be a date or dcinput.
statistics.items.type.1 specifies this for statistics.items.dc.1, type.2 for dc.2 and so on.
The items which are defined as "dcinput" have to be present as a `field` in the `form-definitions` in [DSpace.dir]/config/input-forms.xml else they won't be added.

```
#The starting year if none given it will start from the year 1985
statistics.items.startyear=1985
```

Whenever dropdown selects are present to choose a year, the timeframe for these dropdowns goes from the start date configured here to the current year.

```
# restrict permissions access to the admin editor to specific group.
# Enter the group id in the field below to specify the designated group.
# The user must always be logged in prior to seeing the admin editor.
# By default, only the general administrators can view the admin editor
#statistics.permissions.group=11
statistics.permissions.group=1
```

Only one group can be set to have these permissions.
A global administrator is always allowed to access the module.

Solr Caching

Statlets are using a caching mechanism since some statlets may take a long time to load.

```
#---------------------------------------------------------#
#--------------------- Solr Caching ----------------------#
#---------------------------------------------------------#
solr.caching.server = ${solr.server}/cached_statistics
#Max age of the cached data to be valid in seconds.
solr.cache.max.age = 86400
#Ignore the age of the cached data to be valid.
solr.cache.always.use.cached.data=false
```

`solr.cache.max.age` defines the time before a cached version of the statlet is invalidated. By setting `solr.cache.always.use.cached.data` on true, if available, the cached version will always be used, even if it's invalidated. However, if it's invalidated, a new cached version will be generated for the next request.

atmire-cua.xml

To configure custom metadata fields for the usage analysis or to define custom internal IP ranges, a spring file should be placed in `[dspace]/config/spring/api/atmire-cua.xml`. Please refer to the Client Customization Documentation for more information.

Public CUA parts

The public parts of the Content and Usage analysis can be made admin only by setting the properties

disable.admin.only.interactive.stats=true
disable.admin.only.statlets=true

to false.

Browser Support

The table below shows the browser support for this module

- Browsers that aren't listed are **unsupported**.

- In any browser **JavaScript must be enabled** for the module to work.

	Windows	OS X	Linux
Chrome	Recommended	Recommended	Supported
Firefox	Recommended	Recommended	Supported
Safari	Supported	Supported	n/a
Opera	Supported	Supported	Supported
Internet Explorer 11.x	Recommended	n/a	n/a
Internet Explorer 10.x	Recommended	n/a	n/a
Internet Explorer 9.x	Supported	n/a	n/a
Internet Explorer 8.x	Unsupported	n/a	n/a

Recommended	The module has been tested and works in this browser
Supported	The module should work in this browser
Unsupported	The module probably won't behave as it should in this browser.