Module: DSpace statistics and Google Analytics

Module overview:
This module will introduce the statistics available within DSpace, and the facilities that are available if you use Google Analytics with DSpace.

Module objectives:
By the end of this module you will:

1. Know what statistics DSpace can provide
2. Know how the statistics are collected
3. How to make the statistics publicly available
4. Understand what Google Analytics is and some of the statistics it can provide
DSpace statistics

- DSpace statistics:
  - Collated from DSpace log files
  - Reports generated daily (daily and monthly reports)
  - http://dspace.example.com/dspace/statistics
    - Or via the Administer menu
  - Can be private (must be logged in) or public
    - In dspace.cfg:
      - report.public = [true | false]
Statistics collected

The following statistics are collected

- General overview (e.g. number of items archived / number of item views / user logins)
- Archive Information (numbers of each type of item)
- Item view counts
- Actions performed
- Search terms used

Statistics collected
The following statistics are collected and shown in the reports:

- General overview
  - This section gives an overview of metrics such as how many items are in your repository, how many times your items have been viewed in total, and how many times users have logged in.
- Archive information
  - A list of how many of each type of item you have archived
- Item view counts
  - A list of all items and how many times they have been downloaded
- Actions performed
  - A list of how many times each type of action (e.g. browse a collection) has been completed
- Search terms used
  - A list of the search terms that users of your repository have used
Statistics Configuration

- Configuration for the stats system is in
  - [dspace]/config/dstat.cfg
- For regular generation set up the following cron jobs:
  - # Run stat analyses
    - 0 1 ** * [dspace]/bin/stat-general
    - 0 1 ** * [dspace]/bin/stat-monthly
    - 0 2 ** * [dspace]/bin/stat-report-general
    - 0 2 ** * [dspace]/bin/stat-report-monthly
- Report scripts need to be run AFTER other scripts

In order that statistical reports are generated regularly and thus kept up to date you should set up the following cron jobs:

# Run stat analyses
0 1 ** * [dspace]/bin/stat-general
0 1 ** * [dspace]/bin/stat-monthly
0 2 ** * [dspace]/bin/stat-report-general
0 2 ** * [dspace]/bin/stat-report-monthly

Obviously, you should choose execution times which are most useful to you, and you should ensure that the -report- scripts run a short while after the analysis scripts to give them time to complete (a run of around 8 months worth of logs can take around 25 seconds to complete); the resulting reports will let you know how long analysis took and you can adjust your cron times accordingly.
Practical: Generate Statistics

- Ensure the report.dir is set correctly (by default it will):
  
  ![dspace]/config/dspace.cfg file

- Set the start_year and start_month in ![dspace]/config/dstat.cfg

- Run the statistics scripts
  1. cd ![dspace]/bin
  2. ./stat-initial
  3. ./stat-general
  4. ./stat-monthly

- Run the report scripts

- Open a browser and navigate to ![dspace.url]/jspui/statistics

Practical: Generate Statistics

Please see below for practical
Google Analytics

Google Analytics allow a richer and more detailed suite of statistics such as:

- Time visitors spent on the site
- Where they came from
- Terms they used in search engines to find items
- The geographic location of visitors
- How many pages they looked at
- Which pages they started and ended their visit on

Google Analytics requires a small code change, Manakin has a configurable option.

Google Analytics

Google Analytics allow a richer and more detailed suite of statistics such as:

- Time visitors spent on the site
- Where they came from
- Terms they used in search engines to find items
- The geographic location of visitors
- How many pages they looked at
- Which pages they started and ended their visit on

Google Analytics collects data via a small bit of JavaScript that needs inserting in the footer of all your DSpace pages. The JSPUI requires a small code change in order to use Google Analytics. Manakin has a configurable option.
Practical: Configuring Statistics

1) Ensure the report.dir is set correctly (by default it will):

```
[dspace]/config/dspace.cfg file
```

2) Set the `start.year` and `start.month` in `[dspace]/config/dstat.cfg`

3) For this practical ONLY we’ll need to update an error in the Java code.

Open the file below and remove line 1216:

```
Gedit
[dspace-src]/dspace-api/src/main/java/org/dspace/app/statistics/LogAnalyser.java
```

Save and exit gedit

```
4) DSpace will need to be repackaged:
   `cd [dspace-src]/dspace`
   `mvn package`
```

5) Then DSpace will need to be redeployed:

   `cd [dspace-src]/dspace/target/[version]/dspace-[version].build.dir`
   `Ant -Dconfig=[dspace]/config/dspace.cfg update`

6) Finally, tomcat will need to be restarted

   `sudo /etc/init.d/tomcat5.5 restart`
7) Run the statistics scripts
   
   cd [dspace]/bin
   
   ./stat-initial
   ./stat-general
   ./stat-monthly

8) Run the report scripts
   
   ./stat-report-initial
   ./stat-report-general
   ./stat-report-monthly

9) Open a browser and navigate to [dspace.url]/jspui/statistics
Credits

- These notes have been produced by:
  - Stuart Lewis & Chris Yates
  - Repository Support Project
    - http://www.rsp.ac.uk/
  - Part of the RepositoryNet
  - Funded by JISC
    - http://www.jisc.ac.uk/