|  |
| --- |
| Final Sotera Logo w ds 300dpi (2) |
| Datawake User’s Guide 1.0 |
| Memex |
| DW-USER-GUIDE |
| Brodie McDougald |
| 3/25/2015 |

Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Revision | Reviser | Description |
| 03/25/2015 | 1.0 | Brodie McDougald | Initial draft |
|  |  |  |  |
|  |  |  |  |

Table of Contents

[1 Datawake, The Basics 4](#_Toc415557572)

[2 The Plug-in 4](#_Toc415557573)

[3 Teams 5](#_Toc415557574)

[4 Domains 5](#_Toc415557575)

[5 Trails 6](#_Toc415557576)

[6 Start Building a Trail 7](#_Toc415557577)

[7 Extracted & Domain Features 8](#_Toc415557578)

[8 The Context Menu 10](#_Toc415557579)

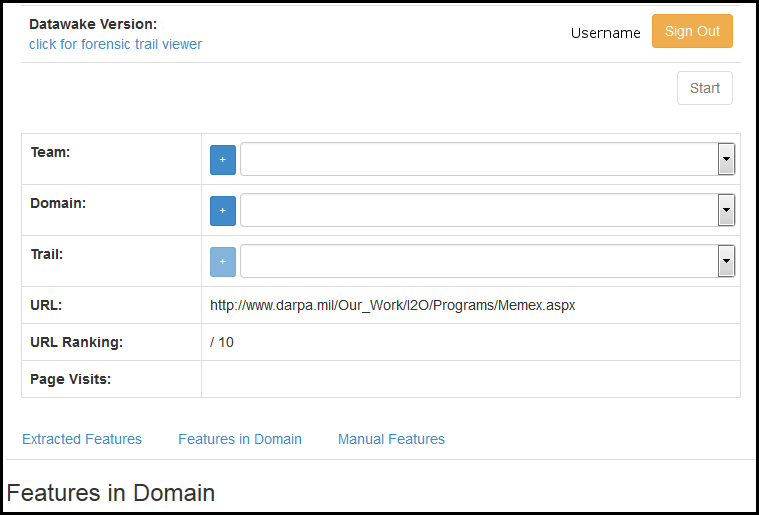
[9 Forensic View - Trail Visualization 10](#_Toc415557580)

# Datawake, The Basics

The Datawake project consists of various server and database technologies that aggregate user browsing data via a plug-in using domain-specific searches. This captured, or extracted, data is organized into browse paths and elements of interest. This information, in the form of Trails (more on Trails later), can be shared or expanded amongst teams of individuals. Elements of interest which are extracted either automatically, or manually by the user, are given weighted values. Extracted elements that are not of interest or might be confused with an element that is of interest (e.g. an Organization with a similar name but not associated in any meaningful way to the one being researched) can be manually removed from the extracted data list.

# The Plug-in

Datawake’s basic operation is handled through a Firefox/TOR browser plug-in. The configuration options of the plug-in itself is managed, like other add-ons, via the Firefox ‘*Add-ons > Extensions*’ page. Once installed, an icon  is placed in the browser toolbar. Clicking on the icon brings up the usage screen for the plug-in.



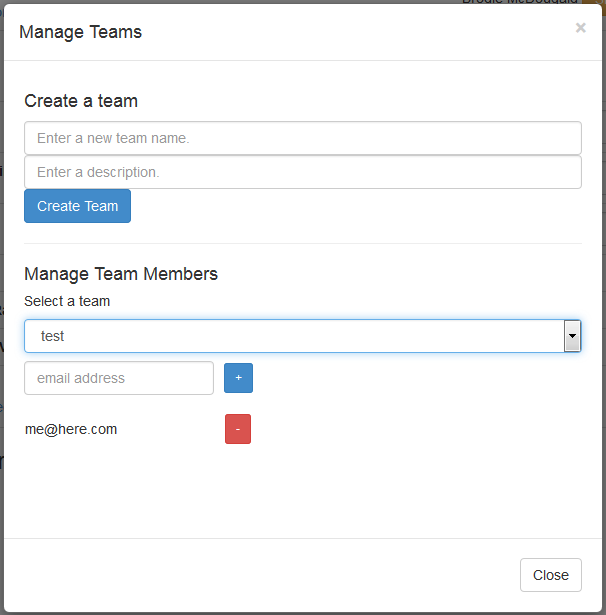
This screen controls the basic operation of the plug-in. The various components that go into starting a user’s information gathering process, and the managing of extracted items of interest can be found here. Additionally, this is where a user ‘Starts’ and ‘Stops’ the data collection process. Each of these areas of interest will be discussed in the sections that follow.

# Teams

Teams are groups of users that can share and collaborate together in their efforts to gather information.

They can be selected, if they already exist, from the pulldown list next to the ‘Team’ section in the plug-in window or can be created and managed from within the plug-in by clicking on the ‘+’ sign next to ‘Team’.  
  


This brings up the ‘Manage Teams’ form:



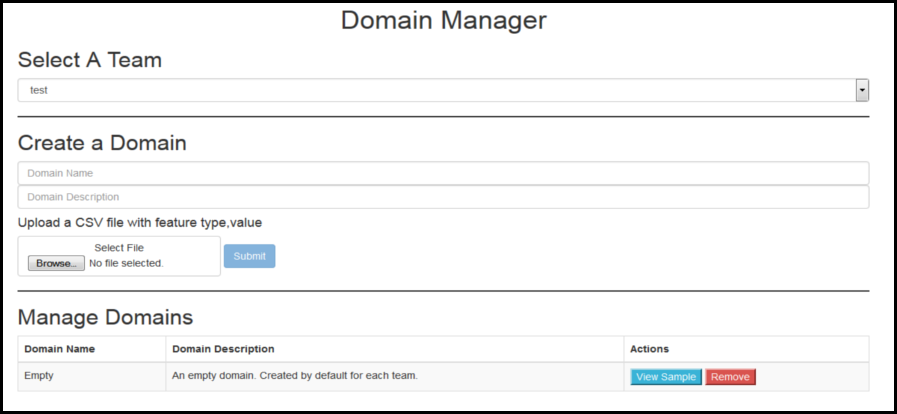
On this form, you can create a new team by entering a name and description then simply clicking on the ‘Create Team’ button. Once created, you may manage a team’s members by selecting the team from the ‘Select a Team’ pulldown list and using the text fields and controls to add or remove members from that team.

# Domains

Domains help Datawake ‘key-in’, or focus, on specific elements of interest during the information gathering process.

Similar to Teams, Domains can be selected from the pulldown list next to the ‘Domain’ section in the plug-in window if they already exist; or they can be created and managed from within the plug-in by clicking on the ‘+’ sign next to ‘Domain’.  
  


Clicking on the ‘+’ sign brings up the Domain Manager page.



From here, a user can select a Team to associate the new Domain with. The new Domain needs a name and description added in the text fields under the ‘Create a Domain’ section. Once a name and description are added, the user can import their Domain via submitting a CSV file which contains the information that will make up this Domain. The CSV file format for a Domain consists of lines of ‘type’ and ‘value’ pairs separated by a comma. For example, a CSV file might contain something similar to the following:

*PERSON,John Smith  
PERSON,Mary Jane  
ORGANIZATION,ABC Merchant Shipping  
EMAIL,”mary@abc-ship.com”  
WEBSITE,”http://abc-ship.com”*

Once added/submitted, the new Domain will show up under ‘Manage Domains’ for the selected Team along with any other Domains that are associated with that Team. You can see the terms contained in a Domain listed under ‘Manage Domains’ by clicking the ‘View Sample’ button, or remove it from the Team by clicking ‘Remove’.

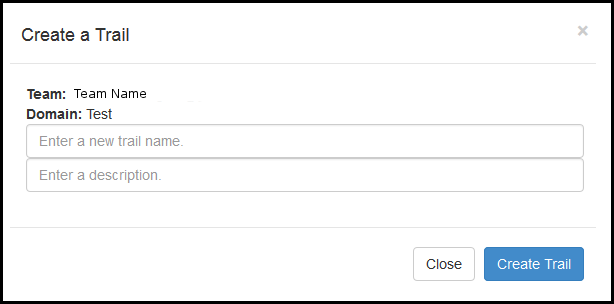
# Trails

The captured, or extracted, data is organized into ‘Trails’ which contain organized browse paths and elements of interest. These ‘Trails’ are initially focused to specific elements of interest through the use of ‘Domains’. Trails can be shared amongst, or expanded by, teams of individuals.

As with the other fields mentioned so far in the plug-in, Trails can be selected from the pulldown list next to the ‘Trail’ section in the plug-in window if they already exist; or they can be created and managed from within the plug-in by clicking on the ‘+’ sign next to ‘Trail’.



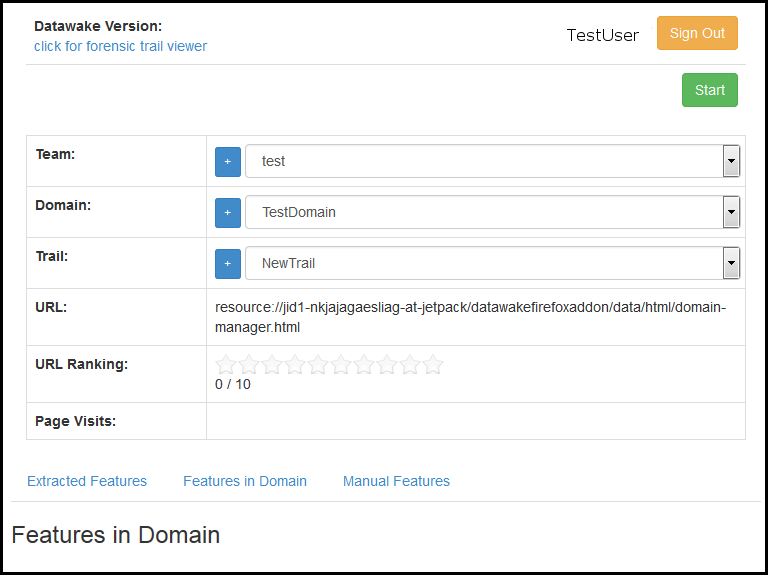
Once you have selected a ‘Team’ and ‘Domain’, clicking on the ‘+’ sign brings up the ‘Create a Trail’ page if the user desires to create a new Trail.



On this page, you can create a new Trail by entering a name and description, then clicking on ‘Create Trail’.

# Start Building a Trail

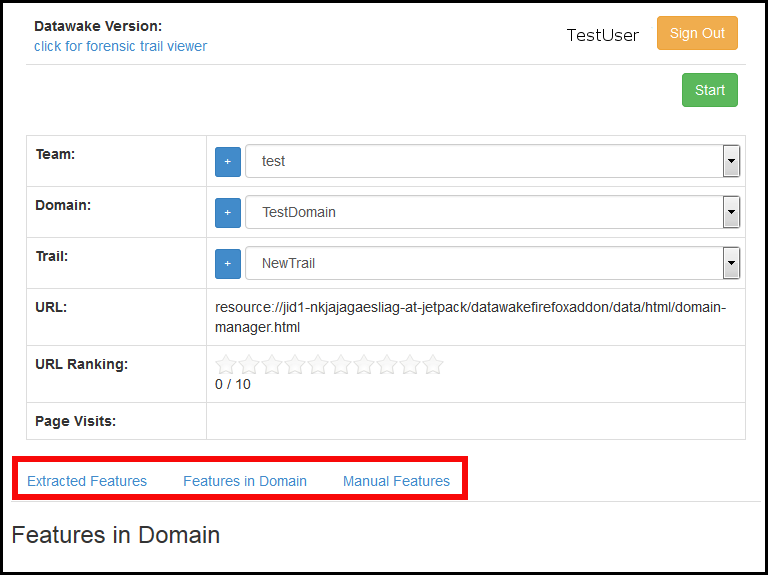
Once you have selected your Team, Domain, and the Trail to start working with, the ‘Start’ button in the plug-in window will light up. At this point, you can click the ‘Start’ button and the Datawake plug-in will start tracking your browse history and extracted elements of interest. This will occur on any new tabs or URLs that are opened in the browser after the ‘Start’ button is pressed.



When it is decided to stop contributing to the running Trail, the user can click the plug-in icon in the browser toolbar and select ‘Stop’ from the plug-in window.

# Extracted & Domain Features

The plug-in window also helps provide information about ‘Features’ contained in, or relevant to, the current page and Trail. There are three types of feature references visible here. They include ‘Extracted Features’, ‘Features in Domain’, and ‘Manual Features’ (shown in the following image outlined in red).



**Extracted Features**The section for ‘Extracted Features’ lists elements on the page currently being viewed that the extractor has deemed of interest. It then attempts to associate the extracted entities to various object types. A user can click on the Edit button at the top of the list and choose to remove extracted elements from the list that are not of interest by clicking the red ‘minus’ button next to the item. Clicking on ‘Done’ after the edits keeps only those elements the user has chosen to be relevant to the Trail. Extracted Elements that also match elements present in the currently loaded Domain will highlight automatically on the web page being viewed.

**Features in Domain**  
This section lists those Features that are part of the currently loaded Domain. They are not necessarily on the page currently being viewed, but are a reference to key elements loaded as part of the domain. This is the same list of items a user would see when clicking ‘View Sample’ from the ‘Manage Domains’ section of the Domain creation page.

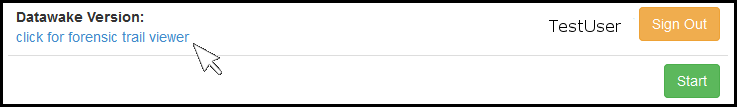
**Manual Features**Manual Features represents any items that were not extracted automatically or were not automatically extracted as being a match for Domain specific elements. Highlighting the element of text you want manually extracted you can tag it as a feature from the Datawake plug-in context menu. A popup will open confirming the text selected, and then you will be prompted to associate it with a ‘type’ of entity. For example, if the text on the web page for ‘ABC Company’ wasn’t extracted automatically; you can highlight it, choose ‘Tag a Feature’ from the context menu, confirm ‘ABC Company’ as the text selected, and call it ‘ORGANIZATION’ for the type. Opening back up the plug-in window and clicking on ‘Manual Features’ will show the item that was manually tagged.

# The Context Menu

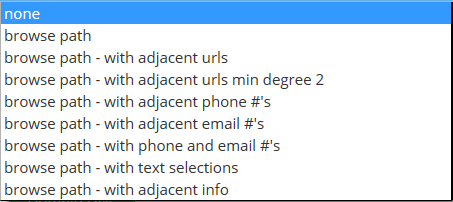
The plug-in also installs items to the right-click context menu within the browser. These include ‘Capture a Selection’ which will note items of interest that can be automatically highlighted or hidden for ease of viewing. The context menu also provides the ability to manually extract elements within the page, as described in the previous section, using ‘Tag a Feature’ option. The final two items are ‘Hide Selection’ and ‘Show Selection’. These are tied to the ‘Capture Selection’ feature. ‘Show Selection’ will highlight any elements that have previously been captured using ‘Capture Selection’ to make them more visible for analysis. ‘Hide Selection’ will make these same elements disappear from the page to clean up the view.

# Forensic View - Trail Visualization

Forensic View is a tool where a user can visualize a Trail they have been working with. It provides several built-in diagraming, or graphing, methods to help in this visualization. To access the Forensic View tool, click on the Datawake plug-in icon, and left-click the hyperlink ‘click for forensic trail viewer’ in the top left as shown here:

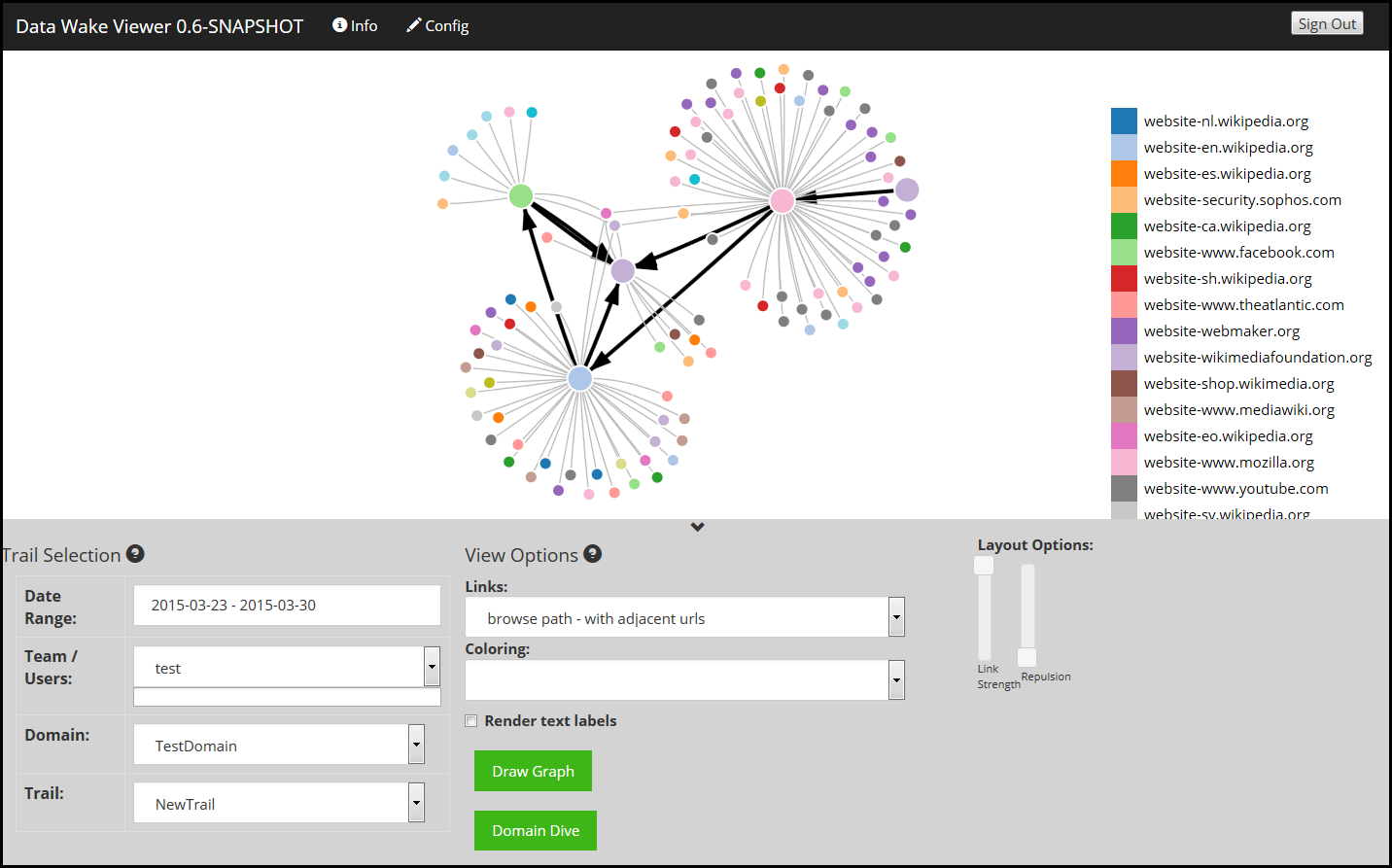


The Forensic View tool will come up in a browser tab. From this browser window, a user can select the Team/Users, Domain, and Trail you wish to Graph. The type of ‘browse path’ information you wish to display can also be chosen from the available presets seen in the following list:



Additionally, the tool also provides some controls for tweaking the way the graph is displayed and colored.

Here is an example of what a browse path graph in Forensic View might look like.



The large dots (or vertices) represent URLs that were visited (the browse path) during the building of the Trail. The smaller vertices represent adjacent pages that were not visited but relate to elements of the domain. Clicking on any of these vertices opens a popup that tells you what the node type is (e.g. webpage), and provides a link to open the page for viewing in a new browser tab. An analyst could investigate the related content for elements relevant to their information gathering tasks and build on the existing Trail if desired.