|  |
| --- |
| Final Sotera Logo w ds 300dpi (2) |
| Datawake User’s Guide 2.0 |
| Memex |
| DW-USER-GUIDE |
| Brodie McDougald, Michael Frame |
| 7/31/2015 |

Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Revision | Reviser | Description |
| 03/25/2015 | 1.0 | Brodie McDougald | Initial draft |
| 07/31/2015 | 2.0 | Michael Frame | Various edits, Update to new UI |
|  |  |  |  |

Table of Contents

[1 An introduction to Datawake 4](#_Toc426111427)

[2 The Firefox Plugin 4](#_Toc426111428)

[3 Datawake Teams 5](#_Toc426111429)

[4 Datawake Domains 5](#_Toc426111430)

[5 Datawake Trails 6](#_Toc426111431)

[6 Start Building a Trail 7](#_Toc426111432)

[7 Extracted & Domain Features 8](#_Toc426111433)

[8 The Context Menu 9](#_Toc426111434)

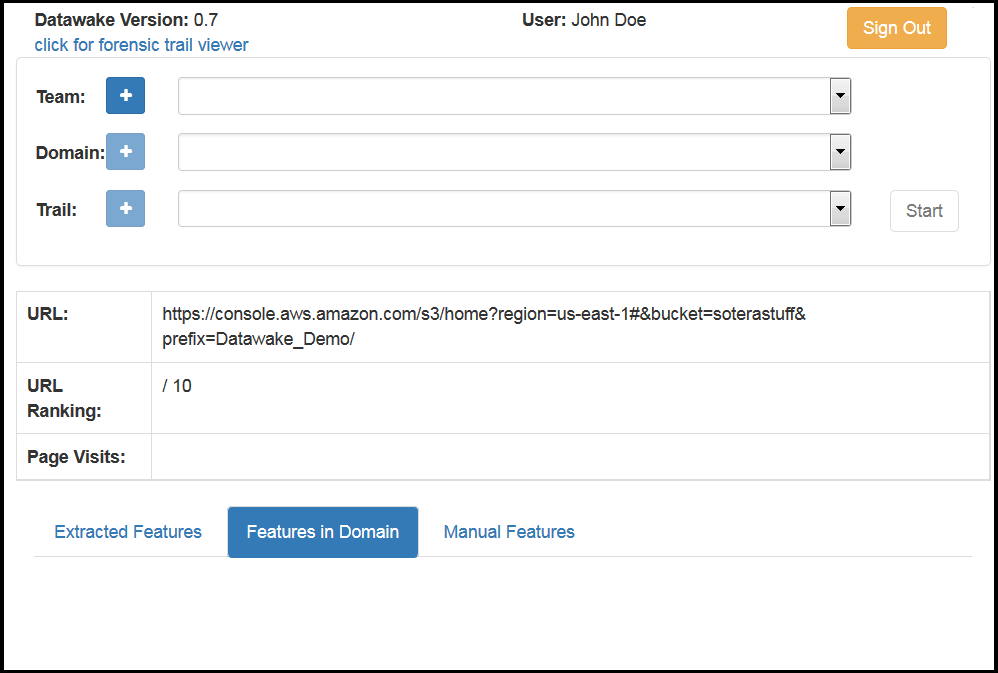
[9 Forensic View - Trail Visualization 9](#_Toc426111435)

# An introduction to Datawake

The Datawake project consists of various server and database technologies that aggregate user browsing data via a Firefox plugin using domain-specific searches. This captured, or extracted, data is organized into browse paths and elements of interest. This information is persisted in the form of Trails (more on Trails later) that can be shared or expanded among teams. Elements of interest are extracted automatically or manually added by the user, which are then given weighted values in various visual representations. 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 Firefox Plugin

Datawake’s basic operation is handled through a Firefox/TOR browser plugin. The configuration options of the plugin itself are managed via Firefox’s “*Add-ons > Extensions”* page. Once installed, an icon C:\Development\Memex\Datawake\FirefoxAddon\data\images\waveicon38.png is placed in the browser toolbar. Clicking on the icon brings up the usage screen for the plugin.



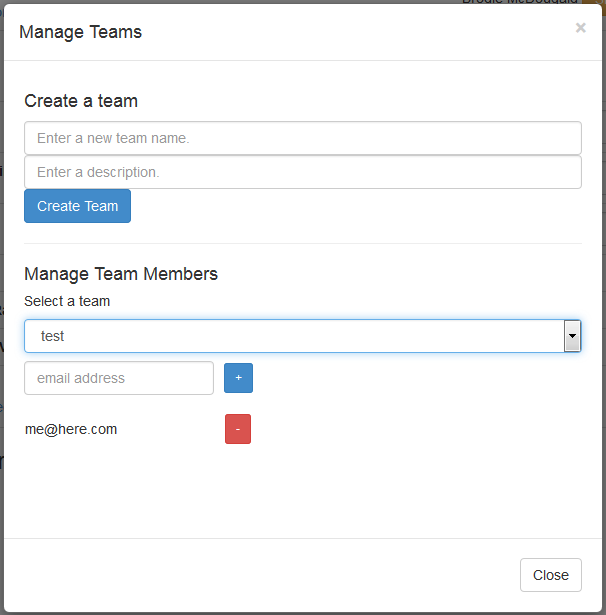
The plugin controls the basic functionality of Datawake, allowing the user to select where the session’s activity will be stored. The various components that go into starting a user’s information-gathering process, and the managing of extracted items of interest can be found here. Most importantly, this is where a user “Starts” and “Stops” the data collection process. It is important to note that all web browser activity and content is trailed once the plugin is “Started”, so care should be taken to “Stop” it before visiting any sites that should not be tracked. Each of these areas of interest will be discussed in the sections below.

# Datawake Teams

Teams are groups of users that can collaborate and share gathered information.

An available team can be selected from the “Team” pulldown list in the plugin window, or a new Team can be created and managed by clicking the “+” sign next to “Team”.  
  


Doing so brings up the “Manage Teams” form:



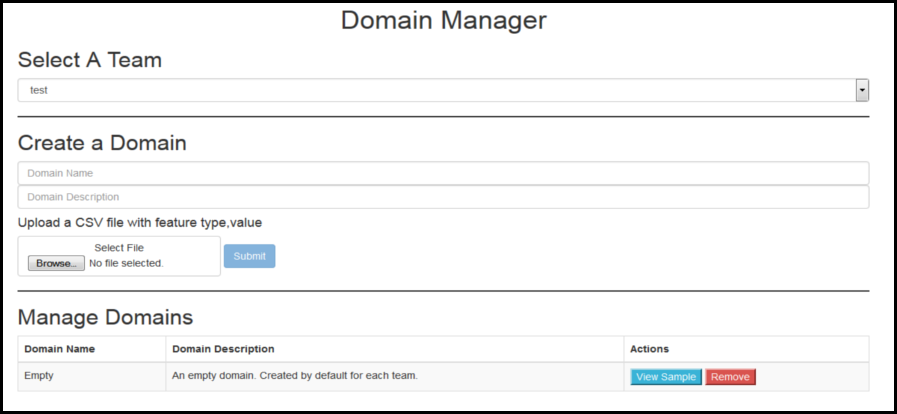
A new team is created by entering a name and description and then clicking on the “Create Team” button. Once created, a team’s members are managed by selecting the team from the “Select a Team” pulldown list and using the text fields and controls to add or remove members from that team.

# Datawake Domains

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

Similar to Teams, a existing Domain can be selected from the pulldown list next in the plugin window or one can be created and managed from within the plugin by clicking on the “+” sign next to “Domain”.  
  


Doing so brings up the Domain Manager page.



From here a user selects a Team to associate the new Domain with. A 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 comma-separated “type/value” pairs. 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 Submitted the new Domain will be displayed under “Manage Domains” for the selected Team along with any other Domains previously with that Team. Domain items for a given domain may be viewed by clicking the “View Sample” button. Domains may be removed from a Team by clicking the “Remove” button.

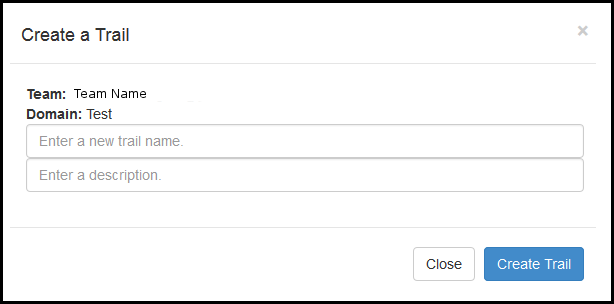
# Datawake Trails

Captured data is extracted and persisted 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 among, or expanded by, teams of individuals.

As with the other attributes mentioned so far in the plugin, Trails can be selected from the pulldown list next to the “Trail” section in the plugin window if they already exist; or they can be created from within the plugin by clicking on the “+” sign next to “Trail”.



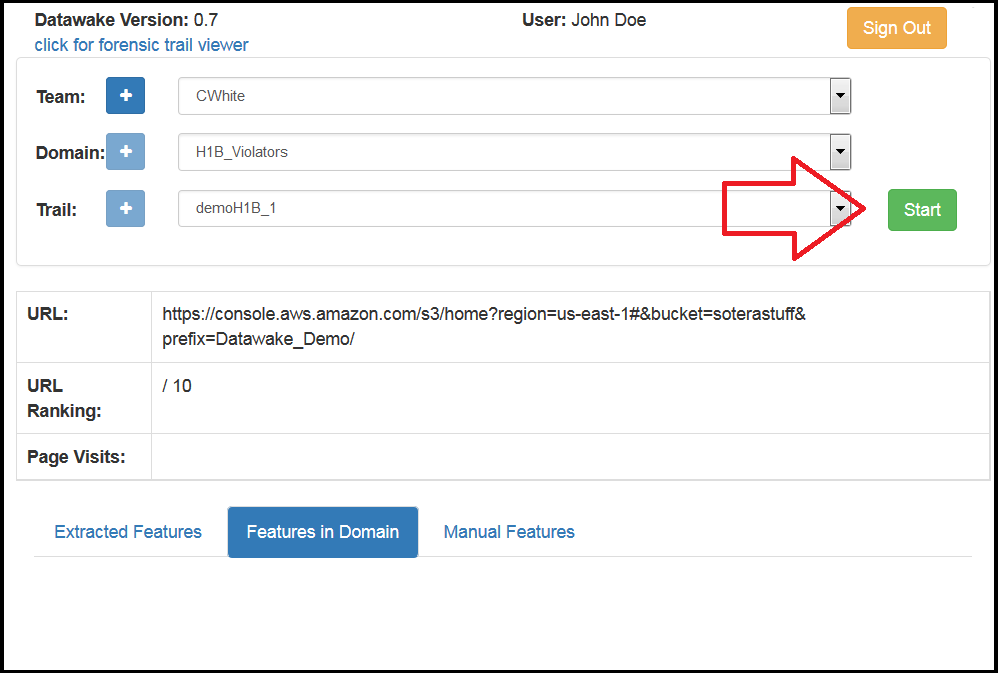
Once you have selected a “Team” and “Domain”, clicking on the “+” sign brings up the “Create a Trail”.



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

# Start Building a Trail

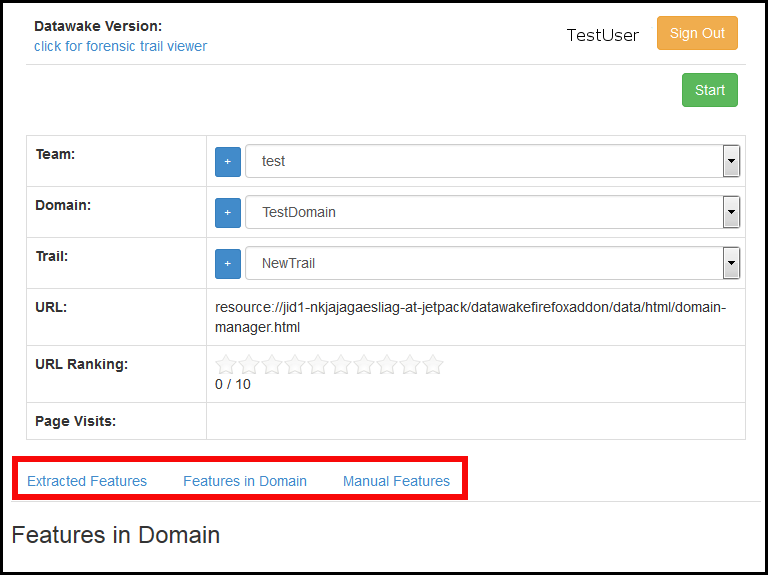
Once you have selected a Team, Domain, and the Trail to work with, the “Start” button in the plugin window will be enabled. You may click the “Start” button to have the Datawake plugin start tracking your browse history and extract elements of interest. Please note **that this will occur on all new tabs or URLs that are opened in the browser after the “Start” button is pressed**. Be sure to click the “Stop” button if you are browsing content you do not wish to be persisted.



When it is decided to stop contributing to the running Trail, the user can click the plugin icon in the browser toolbar and select “Stop” from the plugin window.

# Extracted & Domain Features

The plugin window also provides information about “Features” contained in current page that are relevant to the 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**“Extracted Features” lists elements on the current page that the extractor has deemed of interest. The extractor attempts to associate the extracted entities to the defined object types. A user can remove extracted elements from the list that are not of interest by clicking the “Edit” button and then clicking the red “minus” button next to the item to be removed. Clicking on “Done” will retain only those elements the user has chosen to be relevant to the Trail. Extracted Elements that also match values in the Trail’s Domain are automatically highlighted on the web page being viewed.

**Features in Domain**

“Features in Domain” lists features that are part of the Trail’s associated Domain. They are not necessarily present on the current page, but are a reference to key elements loaded as part of the domain and are a guide to help the user remain on “focus”. 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” represent candidate items, that is, items that were not automatically extracted as matching Domain specific elements. Highlighting the element of text allows you to manually tag it as a feature from the Datawake plugin 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 but is desired to be in the future, you can highlight it and choose “Tag a Feature” from the context menu. Once you confirm “ABC Company” as the text selected, and set “ORGANIZATION” for the type, re-opening the plugin would reveal the item that was manually tagged and now present in “Manual Features”.

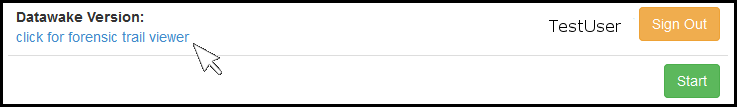
# The Context Menu

Installation of the Datawake Firefox plugin creates several right-click context menu options within the browser. These include:

* “Tag a Feature” – as described in the section above.
* “Capture a Selection” - notes items of interest that can be automatically highlighted or hidden for ease of viewing.
  + “Show Selection” highlights any elements previously captured using “Capture Selection” to make them more visible for analysis.
  + “Hide Selection” makes these same elements disappear from the page to clean up the view.

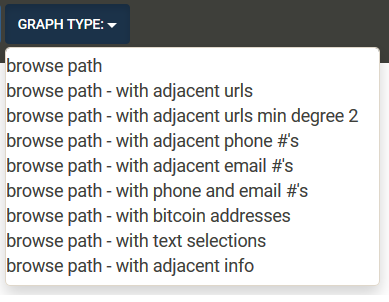
# Forensic View - Trail Visualization

The Datawake Forensic View allows a user to visualize a Trail they have been working with. It provides several built-in diagraming and graphing methods to aid this visualization. To access the Forensic View tool, click on the Datawake plugin icon, and left-click the hyperlink “click for forensic trail viewer” in the top left as shown here:

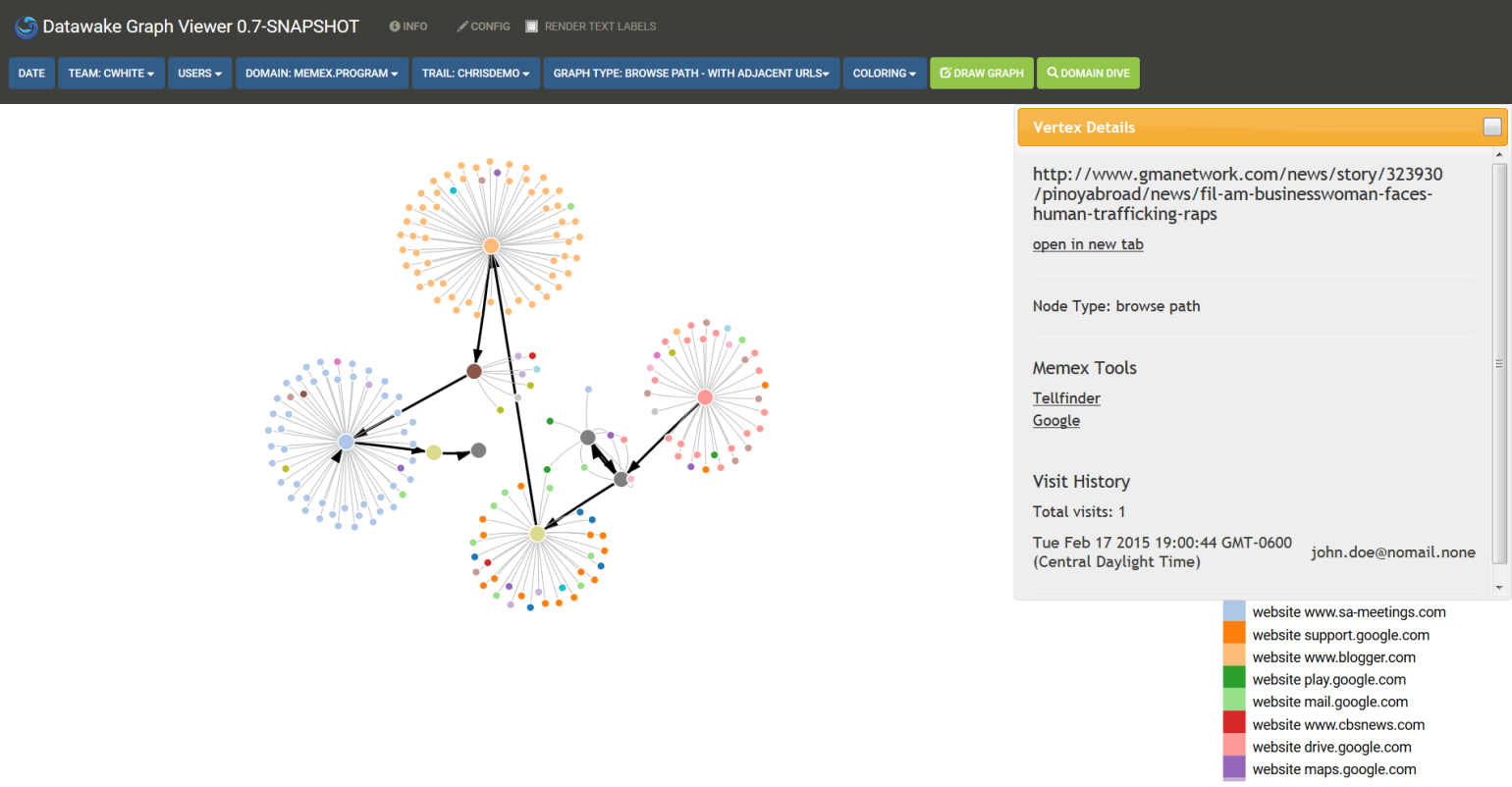


**The user must be logged into the Datawake plugin in order to use the Forensic Viewer**.

The Forensic Viewer will open in a new Firefox browser tab. From the browser tab the user first selects a Team, then one or more Team Members (i.e. “Users”) and the desired Domain, Trail. The Graph Type, or “browse path” information you wish to display, is chosen from the available presets seen in the following list:



The Forensic Viewer also provides 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.