



# DATA SIFT

## DataSift Training Module 1: Basic Concepts

The Real-Time Social Data Mining Platform

# DataSift Training Module 1: Basic Concepts

## What is DataSift?

DataSift is a real-time content curation platform that aggregates inputs from across a wide range of social and web input feeds, augmenting them with sentiment analysis, storage and analytics to offer an unrivaled service.

The world is moving to streams, and consumers will consume and curate their own news. DataSift follows this paradigm shift and seeks to become the platform of choice for Stream Curation, Consumption, and ultimately Monetisation.

# DataSift Training Module 1: Basic Concepts

## Advantages of DataSift

Companies and marketers are demanding better ways to listen to their customers on Twitter, and **understand conversations** about their **brands and products**. With so much news and data being created across multiple social networks, businesses need a way to cost-effectively and **efficiently** filter it all down to find the information that is valuable and meaningful to them.

DataSift lets companies use extensive search queries to access only the data they **value**.

# DataSift Training Module 1: Basic Concepts

## Use Cases

Crisis management

Reputation Management

Risk Analysis

Research

Viral Campaigns

Qualitative Feedback

Augmented Research

Sports

Banking and Financial Services

Media

Transportation

Technology

Government

Retail

Insurance

Property

# DataSift Training Module 1: Basic Concepts

## CSDL: Curated Stream Definition Language

The integral part of the DataSift content curation is the underlying Curated Stream Definition Language.

CSDL is **domain specific language** designed to **filter data** from several sources and create streams. You can filter data based on **location, target type, number of re-tweets, social influence** through Klout / PeerIndex score and numerous other parameters.

# DataSift Training Module 1: Basic Concepts

## Streams and Stream Definitions

The data is curated by the specified set of filters and stored in a “Stream”.

The filters are defined using the **CSDL**. The combination of a filter along with a logical operator is called a Stream Definition.

Example:

```
interaction.content contains "apple" AND  
interaction.content contains "orange"
```

A Stream Definition can have any number of **filter + logical operator** combinations and can also include an Existing Stream Definition.

Example:

```
Rule "2393j83f38e2jk23j89eok9ek2495k9" AND  
Rule "3455nijsoer0934ktjrow090ekko93kf"
```

# DataSift Training Module 1: Basic Concepts

## Targets

Targets are the data sources specified by the user in a filter and DataSift uses this information to get a matching result.

**There are three types of arguments that are supported by DataSift:**

### Common Targets

Common to all types of data.

### Sources

Enables to specify more refined target .

### Augmentations

Uses third party sources to provide additional information to the curated data.

# DataSift Training Module 1: Basic Concepts

## Operators

Operators specify an operation to find the matching content in the filter.

Some of the common operators are:

Exists

Contains

Contains Any

Equals

In

Greater Than

Less Than

# DataSift Training Module 1: Basic Concepts

## Geo Operators

DataSift also supports Geo based operators so users can match content from a specific location or an area.

**There are three types of Geo Operators that are supported in DataSift:**

### Specific Location

By providing the co-ordinates.

### Geo Polygon

Specify a geographic area with 32 co-ordinates (max).

### Geo Radius

Specify a circular area from a given location.



### Toolbox:

The geo co-ordinates can be easily inserted using the DataSift Toolbox on the website.

# DataSift Training Module 1: Basic Concepts

## Arguments

The arguments are the values specified in the filter to find a matching content.

The argument is optional for some operators, for example:

```
twitter.content exists
```

This example just checks if any there is any data available on Twitter.

# DataSift Training Module 1: Basic Concepts

## DataSift Users and API keys

Every user has a profile on the DataSift Website and each user is allotted an API key. The API key is required to access any of the DataSift API. The API key is used for licensing and authentication.

The API key is 32 characters long.

# DataSift Training Module 1: Basic Concepts

## DataSift Website

Create User Profiles

Accept Licenses

Create, update Streams and Recordings

Use Toolbox

Quick Reference

Browse other Streams

Import other Streams

Insert Geo Operators with ease

Check for comments

# DataSift Training Module 1: Basic Concepts

The screenshot shows the DataSift user interface for editing a stream definition. The top navigation bar includes the DataSift logo, links for Home, Browse, Platform, About, and Support, a user profile for 'gemma\_skeats', and a search bar. Below this is a secondary navigation bar with 'Dashboard', 'My Streams', 'My Recordings', 'My Licenses', 'Settings', and a '+ Create Stream' button. The main content area is titled 'Edit Test's Definition' and includes a breadcrumb trail 'My Streams > Test > Edit'. A text box with a line number editor (1-15) is labeled 'Editor'. To the right is a 'Code Toolbox' with items: 'CSDL Reference', 'CSDL Operators', 'Insert Stream Link', 'Keyword List', and 'Map Location'. Two callout boxes, 'Quick Reference' and 'Operator Reference', have arrows pointing to 'CSDL Reference' and 'CSDL Operators' respectively. A 'Tool Box' label has an arrow pointing to the entire toolbox. At the bottom, there are 'Save', 'Save & Preview', and 'Cancel' buttons. A yellow status bar at the bottom reads 'Current Status: DataSift is not currently filtering all data. Twitter input stream offline'. The footer contains '© 2011 MediaSift Ltd' and links for 'Contact | Blog | Status | Terms | Privacy'.

# DataSift Training Module 1: Basic Concepts

## Licensing

DataSift uses many third party sources to provide an enhanced information to the users. Some of the services require you to sign a license agreement.

At the moment only Twitter requires a license agreement but in the future other partners may also introduce a licensing scheme.

The license agreement can be found under My Licenses section of the DataSift Website.



# DataSift Training Module 1: Basic Concepts

The screenshot displays the DataSift web application interface. At the top, there is a navigation bar with the DataSift logo and links for Home, Browse, Platform, About, and Support. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a secondary menu with links for Dashboard, My Streams, My Recordings, My Licenses (which is the active page), Settings, and a + Create Stream button. The main content area is titled "Licenses" and contains the following text: "DataSift uses a variety of third party sources in order to provide you with the best collection of data. Each of these sources may require you to agree to their licensing conditions before you can use the service." and "Failure to sign a required license for a particular service could lead to a intermittent or no interactions for that service." Below this text, there are two columns: "Accepted Licenses" and "Unaccepted Licenses". The "Accepted Licenses" column contains the text: "You currently do not have any accepted licenses. You may receive a reduced service because of this." The "Unaccepted Licenses" column contains three items, each in a box with a right-pointing arrow: "Daniel - Daniel's License", "Buzz - Google Buzz", and "Twitter - Twitter". At the bottom of the page, there is a yellow banner with the text: "Current Status: DataSift is not currently filtering all data. Twitter input stream offline". The footer contains the copyright notice "© 2011 MediaSift Ltd" and a list of links: "Contact | Blog | Status | Terms | Privacy".

**DATA SIFT** Home Browse Platform About Support DS Search

Dashboard My Streams My Recordings My Licenses Settings + Create Stream

## Licenses

DataSift uses a variety of third party sources in order to provide you with the best collection of data. Each of these sources may require you to agree to their licensing conditions before you can use the service.

Failure to sign a required license for a particular service could lead to a intermittent or no interactions for that service.

### Accepted Licenses

You currently do not have any accepted licenses. You may receive a reduced service because of this.

### Unaccepted Licenses

- Daniel - Daniel's License
- Buzz - Google Buzz
- Twitter - Twitter

**Current Status:** DataSift is not currently filtering all data. Twitter input stream offline

© 2011 MediaSift Ltd [Contact](#) | [Blog](#) | [Status](#) | [Terms](#) | [Privacy](#)

# DataSift Training Module 1: Basic Concepts

## How it works

Create user profile.

Define filters using the DataSift Website or DataSift APIs.

DataSift finds the matching data and create Streams.

Access your streams using DataSift APIs or Website.

Output format: JSON and XML.

# DataSift Training Module 1: Basic Concepts

## Pricing

We have a interim pricing model during the Beta.



Pay as you use

Flat Monthly Fee

Contact DataSift sales team for more details: [sales@datasift.net](mailto:sales@datasift.net)

# Questions?