

SPONSORED BY:

MOBILE



Transform Your Business With Data

[ABOUT US](#) [REGISTER](#) [LOGIN](#)

[HOME](#) [BLOGGERS](#) [MESSAGES](#) [POLLS](#) [WEBINARS](#) [RESOURCES](#) [VIDEOS](#) | [STRATEGY](#) [ANALYTICS](#) [TECHNOLOGY](#) [FB](#) [TW](#) [LI](#) [G+](#) [RSS](#)

[FINANCIAL SERVICES](#) [HEALTHCARE](#) [GOVERNMENT](#) [EDUCATION](#) [RETAIL](#)

**BIG DATA ANALYTICS**

**Twitter: Big Data & IPO 140 Characters at a Time?**



**Robert Plant**, Associate Professor, School of Business Administration, University of Miami  
 9/25/2013  
 Comment  
 21 comments

[Login](#)  
 50% 50%

Twitter is not just a mechanism for sending 140-character messages. It is a community, a culture, and a social network where less can be more, and more is often a bad idea. On Sept. 12, the company filed an S-1 document with the SEC as a first step toward going public.

The S-1 is a mechanism for registration of a company's securities with the SEC, detailing its financials so that a prospectus can be created for investors. In this case, the investors are most likely to be investment bankers, who will underwrite the offering. The company has decided to maintain secrecy in its filings, so it can claim emerging-growth company status. Revenue is stated at less than \$1 billion, and the full details and potential risks and rewards have not been revealed to anyone beyond a select few insiders and bankers.

**Whatever the numbers, it is big data**

The lack of clear information on Twitter is troubling for non-insider investors. Even the number of tweeters is vague; figures from different sources vary from 200 million to more than 500 million.

However, one thing is clear; the company is gathering significant amounts of big data. One public consequence of Twitter data streams is that the word "trending" has entered the popular lexicon of the pundit media. The populist view of trending is not without substance. As a real-time indicator of sentiment, it holds value, but it should be understood in statistical context. For example, the outrage that tweeters show about chemical weapons in Syria is not a statistically valid sample of the US population. As such it would be a really bad idea to run the country based on tweets.



This being said, Twitter's management is fully aware that the data in its vaults does hold the potential for a rich analysis of many topics, and companies and governments will be only too willing to pay for it.

- Politicians could see which topics interest their constituents and whether there is a reaction when they or their opponents raise an issue. A longitudinal study of such data could be extremely valuable, and with meta-data such as geospatial tracking, ZIP code analysis could be possible. [Mashable delves into the topic](#) of using Twitter to maximize party donations.
- More complex interactions could be extracted. [Wall Street hedge funds](#) would pay for any early indicators that could be used in the market, such as tweets about using a drug or product for off-label conditions. Sales at a store chain could be identified before the chain relays the sales data to analysts. Even energy consumption generated by utilities could be judged by weather tweets within a community.
- Social media big data was recently highlighted in the Edward Snowden case, but the use of Twitter by security services has been highlighted before. In January 2012, the US Department of Homeland Security detained two British tourists for using the keywords "Destroy America" in a tweet. Governments and surveillance firms will also be willing to pay for access to such data -- if they are not already doing so.

**Privacy vs. revenue**

Clearly, the revenue potential of providing data feeds for third parties such as governments and hedge funds is high. Nevertheless, there is also a risk factor for investors: concerns of defections due to privacy violations.

**The Twitter big data technology stack**

From a technology perspective, since its founding seven years ago, Twitter has built a company on big data. Its platform is built on common open-source software, including Ruby on rails, MySQL, and Memcached. It has developed a service-oriented architecture as it has grown. It has integrated OpenJDK (JVM), Netty, Apache Lucene, Apache Thrift, Apache Hadoop, Redis, and other technologies

**MORE BLOGS FROM ROBERT PLANT**

**A Review: Big Data in 2013**

39 comments  
 Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

**What Would Seymour Cray Think of Big Data?**

13 comments  
 How would the tech legend respond to the 4 Vs?

**The 360 Degree View of a Customer**

56 comments  
 Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

**Lessons for Big Data From President Obama's Healthcare Implementation**

24 comments  
 Three big data lessons for CEOs from HealthCare.gov's bumpy start.

**Is Big Data Ruining Sports?**

29 comments  
 From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

[More from Robert Plant](#)



**FLASH POLL**

All Polls

**DIGITAL AUDIO**



**Big Data Analytics Radio**

Latest Archived Broadcast  
**Your Big Data Initiative Starts With a Great Team**  
 The Big Data Conference and Nemertes Research have teamed up to provide expert advice on building a big data team and finding the right leader. Join us on Tuesday, Aug. 20.  
[Listen](#)

[Full schedule](#) [Archived shows](#) [Download to calendar](#)

**DATA VISUALIZATION SHOWCASE**

**Visualizing Sporting Value: Football Player Transfers**



A great visualization of transfer fees leads us to think of deeper sporting analytics. Explore this data [here](#).

[More Data Visualization Showcase](#)

while contributing significantly to the big data software community through projects such as Apache Hadoop, HBase, BookKeeper Mahout, and Pig.

A significant number of its internal projects have been [released as open-source tools](#), including:

- Flight (an event-driven web framework to map behaviors to DOM nodes)
- Mesos (a cluster manager)
- Bower (a packet manager for the web)
- Scalding (Scala library for specifying Hadoop MapReduce jobs)
- hRaven (a collector of runtime data and statistics from MapReduce jobs)
- Twemproxy (a fast and light proxy for Memcached)
- Typeahead.js (a fast and full featured autocomplete library)
- Zipkin (a distributed tracing system)

Twitter has also a keen eye for technology acquisitions that can bind into its platform and drive its model. Notable acquisitions include:

- BackType, which developed the real-time analytics engine [Storm](#)
- Bluefin Labs, a social analytics technology firm (acquired for [\\$90 million](#))
- MoPub, a technology for helping mobile publishers manage ad inventory (for [\\$350 million](#))
- Trendrr, a [social tracking TV service](#).

#### Developing the ad model

Each of these pieces of the Twitter business model indicates that the company is intent on developing an ad model for a variety of platforms and industry sectors.

Twitter has already established a unique place in the cultural landscape and a ubiquity that goes beyond 140 characters. The IPO -- should it happen -- will undoubtedly be an interesting event tweeted around the world. I have two tweets in my Tweet Deck ready to go "*Caveat emptor*" (buyer beware) and "*Comprehenderent tempus*" (seize the opportunity). I'm just not sure which one to prioritize at this point.

#### Related posts:

- [Telling the Story of Your Data](#)
- [Can Social Media Help the Struggling Airline Industry?](#)
- [Big Data Explained: What Is Entity Disambiguation?](#)
- [Using Twitter to Analyze Linguistic Standards in the US](#)

[Email This](#) [Print](#) [Comment](#)

#### COMMENTS

[Newest First](#) | [Oldest First](#) | [Threaded View](#)

PAGE 1 / 3 > >>



Saul Sherry, User Rank: Blogger  
10/4/2013 | 4:45:58 AM

#### Re: Figures at last!

It's clear the huge user base will contribute to that hopeful 1Bn valuation - but (and I might not be privy to all the info here) I think the monetization is still a bit wonky - that's where the risk in investing is. There's great potential (even if, and I'm just plucking a number out of the air here) 50% of that 218 Million are doubled, fake or bot accounts, that's still a lot of insight to turn in to money. But it needs to prove it can do it.

[Login](#)



[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



netcrawl, User Rank: Petabyte Pathfinder  
10/4/2013 | 4:38:47 AM

#### Re: Figures at last!

The biggest challenge for Twitter is monetization, they still in the process of figuring out how to monetize Twitter- that a massive userbase 218 million! But I think they're still plenty of room to fill here. We can do market research on Twitter, we got have huge opportunities here, they got now the data I think its about time that the company think and focus more on business. How to effectively monetize the site.

[Login](#)



[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



netcrawl, User Rank: Petabyte Pathfinder  
10/4/2013 | 4:28:44 AM

#### Re: Figures at last!

@Saul there's a lot of talks and buzz about Twitter, they have been the subject of endless speculation and talks. They getting bigger, with 218 million users and 500 million tweets a day, well I think they need a much bigger infrastructure, that a lot of BIG DATA! The IPO will provide the company with a war chest to use on strategic acquisitions and international expansion. IT will also help the company forge stronger ties with advertisers and media businesses who want to reach Twitter's massive userbase.

[Login](#)



[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



Saul Sherry, User Rank: Blogger  
10/4/2013 | 3:59:47 AM

#### Figures at last!

So, on the BBC this morning - <http://www.bbc.co.uk/news/business-24392336>

[Login](#)



News on the IPO (or expectation of it) brings some new figures to the thoughts around Twitter.

Big fact 1 - they are looking to raise 1 Billion from the IPO (queue mental images of Dr. Evil demanding money)

Big Fact 2 - they claim to have 218 million users and 500 million tweets are sent a day.

To get a better picture, I'd like to see some stats on registration attrition and especially bots...

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

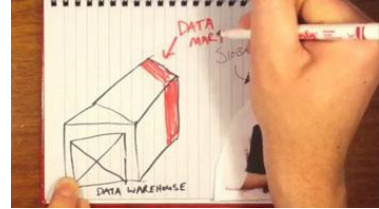
Saul Sherry, User Rank: Blogger  
10/1/2013 | 5:01:42 AM

#### BDR IN YOUR INBOX

Enter email

#### FEATURED VIDEO

### Big Data Explained: What is a Data Mart? 18



Use this video to explain quickly to someone the difference between a data warehouse and a data mart.

[Watch This Video](#)

[More Video Blogs](#)

#### FOLLOW US ON TWITTER

**Big Data Republic**  
@BigDataRepublic

15 Apr

AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... [ow.ly/2FHbY7](http://ow.ly/2FHbY7)

Retweeted by Yakamoto

Expand

---

**Hajime Ozaki**  
@hajozaki

21h

BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic [wp.me/p4Aavi-2](http://wp.me/p4Aavi-2)

Show Summary

---

**Marcus Borba**  
@marcusborba

28 Apr

#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

#### LIKE US ON FACEBOOK

**Big Data Republic**

Like

4,168 people like Big Data Republic.

Facebook social plugin

#### ACCOLADES

Media Pioneer Award Winner



**Re: Another caveat emptor**

I wouldn't worry about noise in that sense @Alphaedge - as you say NLP will fill that gap. Of course when we talk about Twitter, we are talking multi-lingual, different kinds of sentiment, different kinds of meta-sentiment (ironic etc.) so it's a huge problem - but with enough power and brilliant minds behind it, it can be done.

[Login](#)



50% 50%

That maybe forms part of a service the Twitter teams can provide - a really deep drill down on meanings in messages etc. Even expanding to offer the service to content pulled from FB etc.

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



AlphaEdge, User Rank: Exabyte Executive  
9/30/2013 | 9:25:18 PM

**Re: Another caveat emptor**

On the other hand, I think it is important to extract main topics from twitter data given some of them might be just noise. I guess that's why NLP in discussion for big data analytics.

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



smkinoshita, User Rank: Exabyte Executive  
9/29/2013 | 3:07:17 PM

**Re: Another caveat emptor**

@James Connolly: Quite right about Twitter and the media.

[Login](#)



50% 50%

However, there's a whole lot of great data within Twitter, and many of its users even help organize this data using hashtags.

The problem is there's a whole lot more garbage data within Twitter, and a good portion *looks* like good data -- and that's not even touching the spammers.

Still, with enough analysis Twitter is a goldmine and it's a great tool for business if it's used properly. It's powerful because it's small, but because it's small it will need context to be recorded along with it.

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



technetronic, User Rank: Blogger  
9/28/2013 | 5:21:16 PM

**Re: Another caveat emptor**

i think they'd make a hell of a lot of money from select "premium" users -- the content consumers or the content generators, but someone has to pay, and someone needs to incorporate micropayments into it

[Login](#)



50% 50%

it's the same way i feel about newspapers...make a system where people can actually pay for what they use, and certainly for something like twitter, where you gain "clout" by participating, perhaps you could charge content creators

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



dcawrey, User Rank: Exabyte Executive  
9/27/2013 | 3:50:09 PM

**Re: Another caveat emptor**

I honestly used to wonder just how Twitter would make money, but now I can see the value that it has as a communications medium. Actually, I am now wondering if its value should be higher than that of Facebook. Anyone agree?

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



Saul Sherry, User Rank: Blogger  
9/27/2013 | 5:05:54 AM

**Re: Another caveat emptor**

I'm imagining when all that is formalised it will be behind closed doors, for the investors eyes only.

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

PAGE 1 / 3 > >>

**BIG DATA REPUBLIC**

[ABOUT US](#) | [CONTACT US](#) | [HELP](#) | [REGISTER](#)  
[TWITTER](#) | [FACEBOOK](#) | [LINKED IN](#) | [GOOGLE+](#) | [RSS](#)



bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/268135



UBM TECH

OUR MARKETS: [Business Technology](#) | [Electronics](#) | [Game & App Development](#)

Working With Us: [Advertising Contacts](#) | [Event Calendar](#) | [Tech Marketing Solutions](#) | [Corporate Site](#) | [Contact Us](#) / [Feedback](#)

[Terms of Service](#) | [Privacy Statement](#) | Copyright © 2014 UBM Tech, All rights reserved