

# Apache Roles Dataset



Dr. Megan Squire  
Elon University, NC  
[FLOSSmole.org](http://FLOSSmole.org)

# Goal

Collect a giant list of all the Apache participants, what their role(s) are within Apache, and on what project.

# Goal

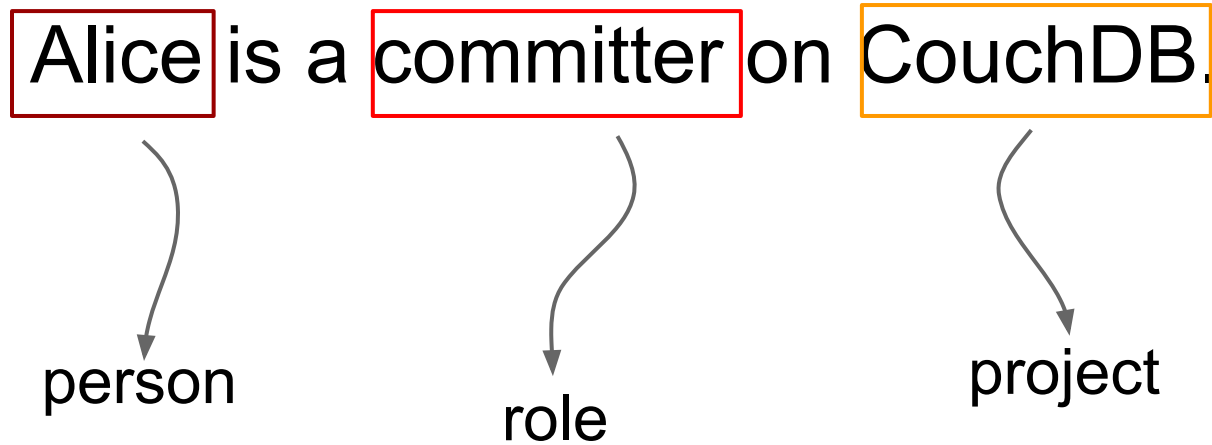
Collect a giant list of all the Apache participants, what their role(s) are within Apache, and on what project.

Timestamp everything so when I collect this data again, I can see changes.

# Why Apache?

- Big (200+ subprojects)
- Transparent
- Popular to study

# What I'm looking for



# Complete record

date collected  
+ source =  
datasource\_id

On May 19, 2013 I learned from the  
Minutes of the April Board Meeting that  
Alice is a committer on CouchDB.

person

role

project

# Why is this useful?

Discussions that require knowledge of roles:

- Power relations
- Leadership structure
- Group dynamics
- Promotion and retention
- Decisionmaking
- Workload, responsibilities

# Data sources

- List of Committers (CLA signers)
- List of Signers (not yet committed)
- List of Committers (self-submitted)
- List of ASF members
- Minutes of Board Meetings
- Individual Project Team Listings



# List of CLA-signed committers

[A](#)[B](#)[C](#)[D](#)[E](#)[F](#)[G](#)[H](#)[I](#)[J](#)[K](#)[L](#)[M](#)[N](#)[O](#)[P](#)[Q](#)[R](#)[S](#)[T](#)[U](#)[V](#)[W](#)[X](#)[Y](#)[Z](#)

## svn-committers

SVN id	Name	SVN Projects
a_horuzhenko	Artyom Horuzhenko	<a href="#">openmeetings</a>
aadamchik	<b>Andrus Adamchik</b>	<a href="#">apsite</a> , <a href="#">cayenne</a> , <a href="#">cayenne-pmc</a> , <a href="#">click</a> , <a href="#">incubator</a> , <a href="#">incubator-pmc</a> , <a href="#">member</a> , <a href="#">pmc-chairs</a> , <a href="#">wave</a>
aadamowski	Aleksander Adamowski	<a href="#">directory</a>
aadrian	Adrian Antal	<a href="#">click</a>
aaf	Alexei Fedotov	<a href="#">incubator</a> , <a href="#">openmeetings</a> , <a href="#">openmeetings-pmc</a>
aanderson	Aaron Anderson	<a href="#">ode</a>
aaron	<a href="#">Aaron Bannert</a>	<a href="#">apr</a> , <a href="#">apr-pmc</a> , <a href="#">apsite</a> , <a href="#">httpd</a> , <a href="#">httpd-pmc</a> , <a href="#">incubator</a> , <a href="#">incubator-pmc</a> , <a href="#">infrastructure</a> , <a href="#">member</a>
ab	<b>Andrzej Bialecki</b>	<a href="#">apsite</a> , <a href="#">gora</a> , <a href="#">gora-pmc</a> , <a href="#">incubator</a> , <a href="#">incubator-pmc</a> , <a href="#">lucene</a> , <a href="#">lucene-orp</a> , <a href="#">lucene-pmc</a> , <a href="#">member</a> , <a href="#">nutch</a> , <a href="#">nutch-pmc</a>

<http://people.apache.org/committer-index.html>

# List of CLA Signers, not yet committers

## Unlisted CLAs

Persons with signed CLAs but are not committers:			
Hari A V	Software AG	Abdullah Ali AL-Dokkan	Cristina Lucia Abad
Steve Abatangle	Reda Abdi	solid	Walied Amer Abdullah
Jalud Abdulmenan	Abhilash	Bryce Acer	Jin Adachi
Henry Adams	David Adcox	Rolf Ade	Igor Afanasyev
Neeti Agarwal	Vinay Kumar Agarwal	Kapish Aggarwal	Tulika Agrawal
Hasan Ahmad	Onur Aktas	Daniel Albers	Nick Alcock
Rail Aliev	Patrick Allaert	Alan Allegret	Jon Allen
Miguel Almeida	Luis Alves	Guido Amarilla	Louis Amodeo





<http://people.apache.org/committer-index.html#unlistedclas>

# Committers by ID (self-submitted)

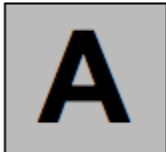
he  
tion  
rg /

## Committers List




[Homepage](#)  [PGP Key](#)  [Photo](#)  [External FOAF File](#)  [Geographical Location](#)

The following list is created **ONLY** from the information provided by committers. The information is voluntary so the following list is not comprehensive.



[Shinichiro Abe](#)  

[Erik Abele](#) 

[Saminda Abeyruwan](#)  

[Andrus Adamchik](#)  

[Alessandro Adamou](#)

[Glenn Adams](#)  

[Rahul Akolkar](#)  

[Jim Ancona](#) 

# Committers by ID (self-submitted)

**Stefan Bodewig**



Projects:

- [Apache Ant](#)
- [Apache Commons](#)
- [Apache Gump](#)
- [Apache Jakarta](#)
- [XMLUnit](#)
- [Apache Logging](#)
- [Apache Lucene.NET](#)

PGP Keys:

- ID: 5F6B8B72 Fingerprint: CE80 75A2 5154 7BEE 249B C151 A211 5AE1 5F6B 8B72

Weblogs

- [Stefan Bodewig's Weblog - Mixed Content](#) 
- [Stefan Bodewigs Weblog - Gemischtwarenladen](#) 

**Bernd Bohmann**



Projects:

- [Apache MyFaces](#)

PGP Keys:

- ID: C4FC9A65 Fingerprint: C4D5 1950 987B E71C B9C5 1E35 ABD6 32EC C4FC 9A65

# Members of the ASF

## Members of The Apache Software Foundation

Id	Name	Projects
erikabele	<a href="#">Erik Abele</a>	<a href="#">HttpComponents</a> , <a href="#">httpd</a> , <a href="#">httpd-docs</a>
samisa	<a href="#">Samisa Abeysinghe</a>	<a href="#">Axis2/C</a> , <a href="#">Axis C++</a> , <a href="#">Axis2</a>
aadamchik	Andrei Adamchik	
rahul	Rahul Akolkar	
gregames	Greg Ames	<a href="#">httpd</a> , <a href="#">APR</a>
jta	Jean T. Anderson	<a href="#">Derby</a>
cezar	<a href="#">Cezar Cristian Andrei</a>	<a href="#">XMLBeans</a> , <a href="#">VXQuery</a>
danny	<a href="#">Danny Angus</a>	<a href="#">James</a> , <a href="#">Jakarta</a>
assaf	Assaf Arkin	

<http://www.apache.org/foundation/members.html>

# Board Meeting Minutes

RESOLVED, that the persons listed immediately below be and hereby are appointed to serve as the initial members of the Apache CloudStack Project:

- \* Joe Brockmeier <jzb@apache.org>
- \* Will Chan <willchan@apache.org>
- \* Chip Childers <chipchilders@apache.org>
- \* Mohammad Nour El-Din <mnour@apache.org>
- \* Sebastien Goasguen <sebgoa@apache.org>
- \* Matt Richard Hogstrom <hogstrom@apache.org>
- \* Wido den Hollander <widodh@apache.org>
- \* Alex Huang <ahuang@apache.org>
- \* Jim Jagielski <jim@apache.org>
- \* Alex Karasulu <akarasulu@apache.org>
- \* John Kinsella <jlk@apache.org>
- \* Kevin Kluge <kluge@apache.org>
- \* Olivier Lamy <olamy@apache.org>
- \* David Nalley <ke4qqq@apache.org>
- \* Brett Porter <brett@apache.org>
- \* Noah Slater <nslater@apache.org>
- \* Marcus Sorensen <mlsorensen@apache.org>
- \* Edison Su <edison@apache.org>
- \* Hugo Trippaers <hugo@apache.org>
- \* Chiradeep Vittal <chiradeep@apache.org>

"initial members"

apache "svn id" given

NOW, THEREFORE, BE IT FURTHER RESOLVED, that Chip Childers be appointed to the office of Vice President, Apache CloudStack, to serve in accordance with and subject to the direction of the Board of Directors and the Bylaws of the Foundation until death, resignation, retirement, removal or disqualification, or until a successor is appointed; and be it further

"Vice President"

# Board Meeting Minutes

location within the data source

Attachment V: Report from the Apache Karaf Project [Jean-Baptiste Onofré]

Apache Karaf provides higher level features and services specifically designed for creating OSGi-based servers.

Community

=====

Indicators during the last 3 months:

- Download: 31,490 downloads (all distributions, all branches) between December, 1 and February, 28.
- User mailing list: 557 messages between December, 1 and February, 28.
- Dev mailing list: 382 messages between December, 1 and February, 28.
- Commits: 621 commits between December, 1 and February, 28.

We are pleased to count two new committers elected in February:

- Filippo Balicchia
- Christoph Grtischenberger

VP/PMC chair

"committers"

# Individual Project "Team" Pages

## PROJECT

[Home](#)  
[Downloads](#)  
[Features](#)  
[License](#)

## COMMUNITY

[Get Involved](#)  
[Mailing Lists](#)  
[People](#)

## DEVELOPMENT

[Source & Guide](#)  
[Issues](#)  
[Builds](#)

## DOCUMENTATION

[Manual 1.4](#)  
[Javadoc 1.4](#)  
[Examples 1.4](#)  
[Screenshots](#)



## APACHE ACCUMULO CONTRIBUTORS


### PMC AND COMMITTERS

username	name	organization
acordova	Aaron Cordova	
afuchs	Adam Fuchs	<a href="#">sqrrl</a>
bimargulies	Benson Margulies	
billie	Billie Rinaldi	<a href="#">Hortonworks</a>
brianloss	Brian Loss	<a href="#">Praxis Engineering</a>

company  
and URL



# Team Pages, another version



Apache  
**CouchDB**  
*"Time to relax"*

[Home](#) [Downloads](#) [Screenshots](#) [License](#)






## Committers

- [Damien Katz, damien@apache.org](#)  
Original developer and remains the project lead.
- [Jan Lehnardt, jan@apache.org](#)  
Worked on the original UNIX port and now works on al consultant and gives presentations around the world.
- [Noah Slater, nslater@apache.org](#)  
Developed and maintains the Autotools build system an manager and maintains a number of related packages f

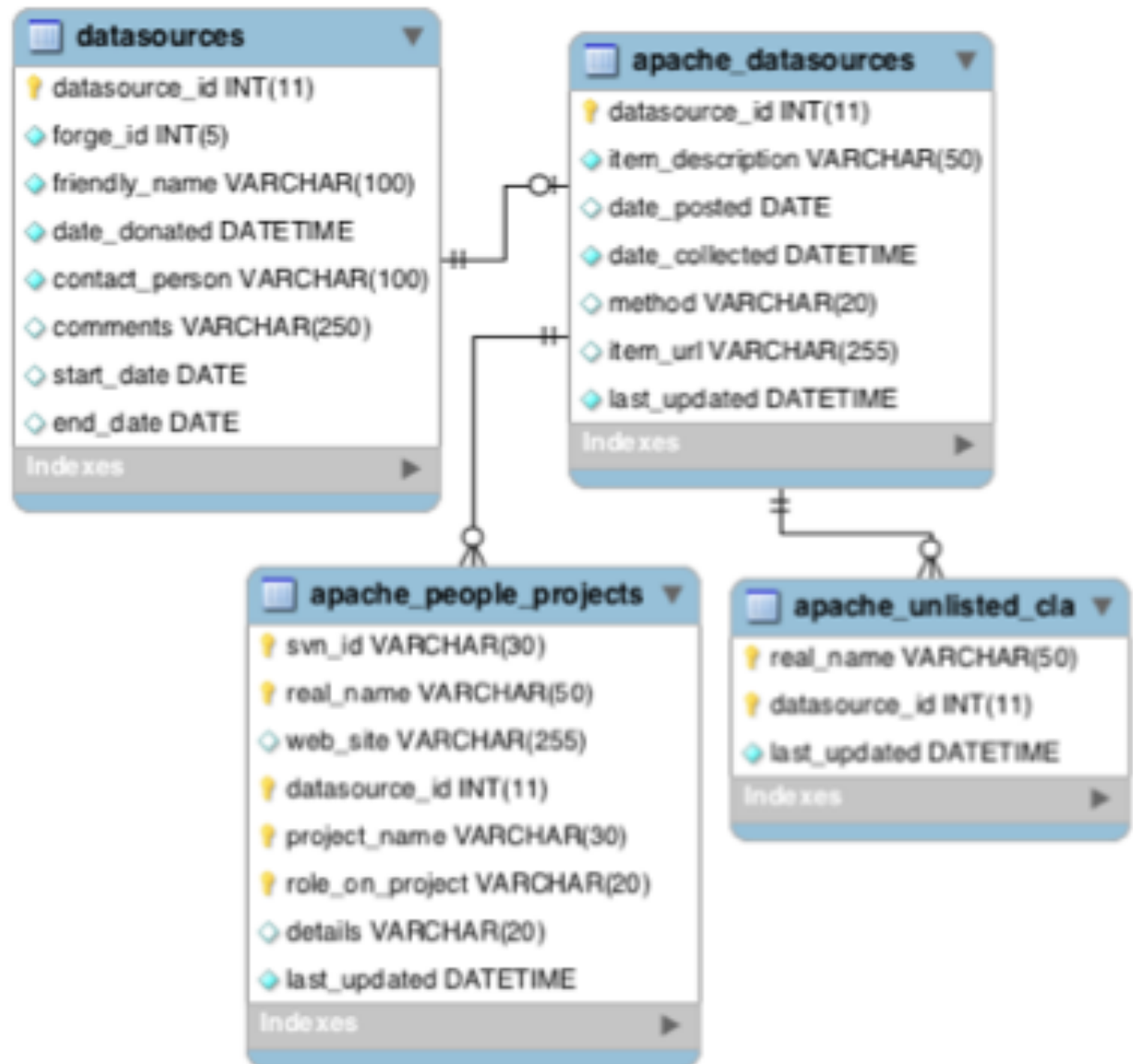
# Team Pages, yet another version

## Members

The following is a list of developers with commit privileges that have directly contributed to the project in one way or another.

Image	Id	Name	Email	Organization	Roles	Time Zone	Actual Time (GMT)
	olamy	Olivier Lamy	<a href="mailto:olamy@apache.org">olamy@apache.org</a>	-	PMC Chair	Australia/Melbourne	Wed May 15 2013 05:14:40 GMT-0400 (EDT)
	aheritier	Arnaud Héritier	<a href="mailto:aheritier@apache.org">aheritier@apache.org</a>	-	PMC Member	+1	Tue May 14 2013 20:14:40 GMT-0400 (EDT)
	baerrach	Barrie Treloar	<a href="mailto:baerrach@apache.org">baerrach@apache.org</a>	-	PMC Member	Australia/Adelaide	Wed May 15 2013 04:14:40 GMT-0400 (EDT)
	bimargulies	Benson Margulies	<a href="mailto:bimargulies@apache.org">bimargulies@apache.org</a>	-	PMC Member	America/New_York	Tue May 14 2013 14:14:40 GMT-0400 (EDT)
	brett	Brett Porter	<a href="mailto:brett@apache.org">brett@apache.org</a>	ASF	PMC Member	+10	Wed May 15 2013 05:14:40 GMT-0400 (EDT)

# End result



everything on FLOSSmole:  
-- in MySQL, and  
-- at our Google Code site

# Sample data

Column Name	Key	Nullable?	Sample Row
svn_id	PK		
real_name	PK		Luis Bernardo
web_site		Yes	
datasource_id	PK, FK		375
project_name	PK		Apache XML Graphics
role_on_project	PK		Committer
details		Yes	Attachment BC
email		Yes	
organization		Yes	
last_updated			2013-04-14 20:25:08

# Dig in...

```
SELECT *  
FROM `apache_people_projects`  
WHERE real_name LIKE "Alice%"  
AND role_on_project = "PMC Chair" OR "Vice President";
```

more inspiring examples in the paper...

# Got a better idea?

Make a dataset and send it to FLOSSmole.



# Apache Twitter Dataset



Dr. Megan Squire  
Elon University, NC  
[FLOSSmole.org](http://FLOSSmole.org)

# Goal

Collect a giant list of Twitter handles for Apache-affiliated people and projects.



# Goal

Collect a giant list of Twitter handles for Apache-affiliated people and projects.

Timestamp this so I can see new entries.

# Why is Twitter useful?

- Just another interesting dev artifact
- Useful for social mining
- Everybody's doing it

# Data sources

Directly from Apache

- project/team web pages

- listed personal blogs

# Data sources

Directly from Apache

--project/team web pages

--listed personal blogs

Indirectly

--Search on Twitter using items from my own


Apache Roles dataset

- names
- emails
- svn
- project names

# From team pages

## Core Committers

Name	Mail	Webpage	PMC	A short summary of yourself
Isabel Drost	isabel@...	<a href="#">Homepage,</a> <a href="#">Blog</a>	Yes	Passion for free software (development, but to some extend also the political and economic implications), interested in agile development and project management, lives in Germany. Follow me on Twitter <a href="#">@MaineC</a>
Ted Dunning	tdunning@...	<a href="#">MapR</a> <a href="#">Technologies</a>	Yes	




no standard way of writing this  
on any team page; lots of  
unstructured data

# From individual blogs

## Core Committers

Name	Mail	Webpage
Isabel Drost	isabel@...	<a href="#">Homepage</a> , <a href="#">Blog</a>
Ted Dunning	tdunning@...	<a href="#">MapR Technologies</a>
Jeff Eastman	jeastman@...	<a href="#">Windward Solutions</a>
Drew Farris	drew@...	
Grant Ingersoll	gsingers@...	<a href="#">Homepage</a>
Sean Owen	srowen@...	<a href="#">Myrrix Ltd</a>
Robin Anil	robinanil@...	<a href="#">Homepage</a>

write programs to follow the link, try to discern the Twitter name, if any



ROBIN ANIL  
/BIN/RAW


HOME ABOUT

Type your search here... GO

Featured Post  
**My first book**

Its been an amazing year, finished two summer of codes, got out of college, got voted as a committer at Apache Mahout, landed a job at Google. Meanwhile, I got an opportunity to write a book about Mahout. Its called Mahout in Action and its available here.

[Read More](#)




Google Summer of Code 2010 @ Mahout

Posted by **Robin Anil** | Posted in **Google, Mahout** | Posted on 20-03-2010

0

MY BOOK



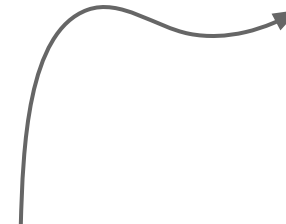
Its a new year and a new Google Summer of Code. I would be mentoring this year at Mahout along with **Grant** and **Ted**. If you are a student interested in participating, this is what you should do:

1. Instead of a wiki like last year this year we are putting up ideas and proposals as JIRA issues here. If you are looking for ideas for your proposal take a look at the JIRA link and explore things which you want to implement. If you have an idea of your own proceed to step 2

# Using Twitter itself: API

```
def get_info_by_id(t, ids):  
    id_to_info = {}  
  
    while len(ids) > 0:  
        # Process 100 ids at a time...  
  
        ids_str = ','.join([str(_id) for _id in ids[:100]])  
        ids = ids[100:]  
  
        response = make_twitter_request(t,  
                                       getattr(getattr(t, "users"), "lookup"),  
                                       user_id=ids_str)  
  
        if response is None:  
            break  
  
        if type(response) is dict: # Handle Twitter API quirk  
            response = [response]  
  
        for user_info in response:  
            id_to_info[user_info['id']] = user_info  
  
    return id_to_info
```

Let program try to guess  
Twitter handle using known  
Apache data



# Using Twitter itself: manual

The image shows a Twitter search interface. At the top, a search bar contains the text "ted dunning". Below the search bar, a dropdown menu displays a search result for "ted\_dunning @ted\_dunning". A red arrow points from this result to the full profile page shown on the right. The profile page for "ted\_dunning @ted\_dunning" features a profile picture of an older man with glasses. His bio is circled in red and reads: "Committer on Apache Mahout, Apache Drill, PMC on Mahout, Zookeeper and Drill, Product Architect at MapR. ex-Chief Scientist at Veeva, Musicmatch. Northern California". Below the bio, statistics show 712 TWEETS, 40 FOLLOWING, and 2,257 FOLLOWERS. A "Follow" button is visible. The "Tweets" section shows a tweet from "ted\_dunning @ted\_dunning" dated "13 May". The tweet text is "download sources in IntelliJ works for me." and includes mentions of "@MaineC", "@alansad", and "@ApacheMahout". The mention "@ApacheMahout" is circled in red. A "View conversation" link is at the bottom of the tweet.

ted dunning

ted\_dunning @ted\_dunning

Search all people for ted dunning

ted\_dunning  
@ted\_dunning

Committer on Apache Mahout, Apache Drill, PMC on Mahout, Zookeeper and Drill, Product Architect at MapR. ex-Chief Scientist at Veeva, Musicmatch. Northern California

712 TWEETS 40 FOLLOWING 2,257 FOLLOWERS

Follow

Tweets

ted\_dunning @ted\_dunning

@MaineC @alansad @ApacheMahout @scdotopen Yeah... download sources in IntelliJ works for me.

View conversation

13 May

Curation!

Always confirm that this is really an Apache person.

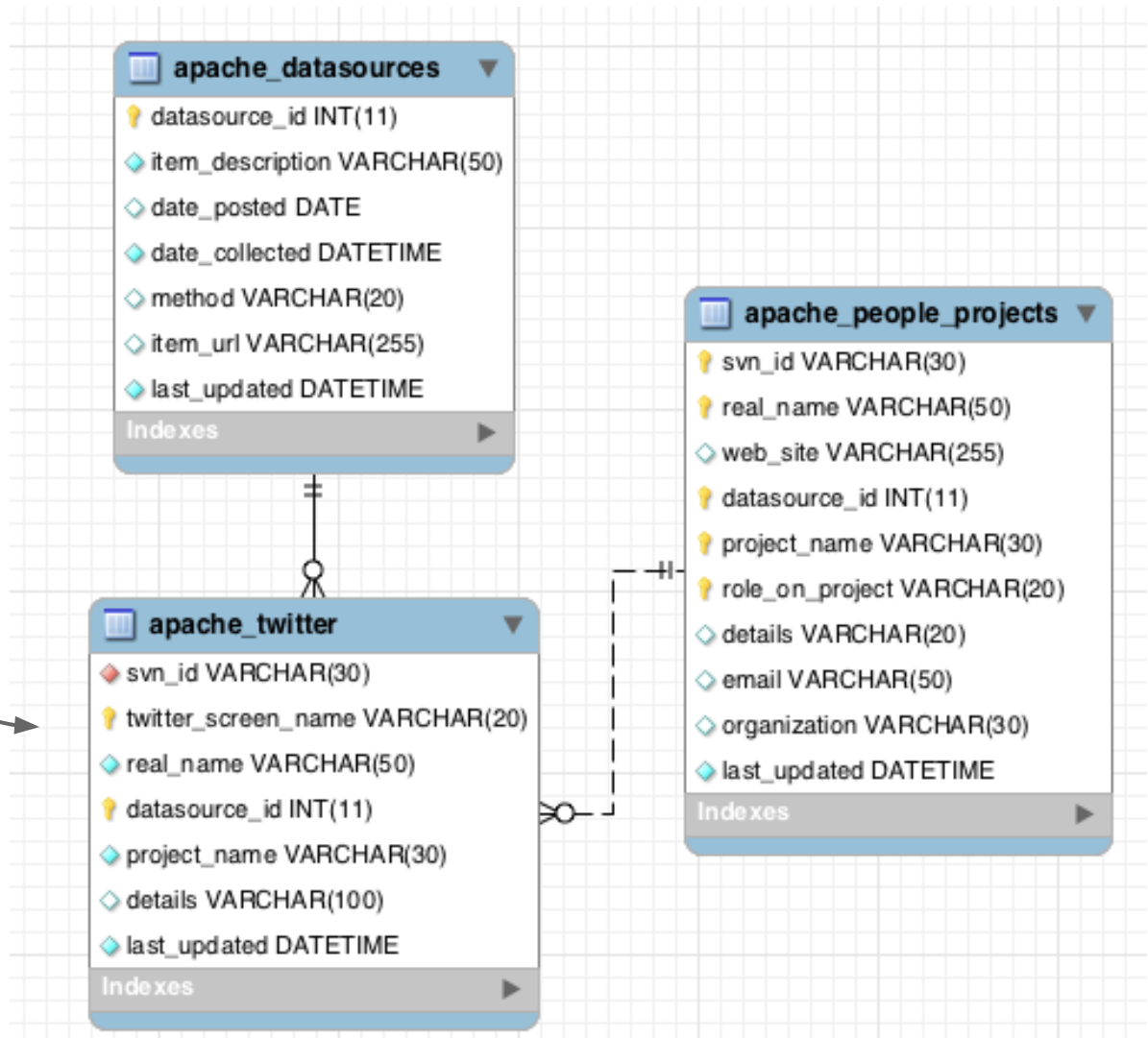


# End Result

about 500 of these



everything on FLOSSmole:  
-- in MySQL, and  
-- at our Google Code site



# Sample data

Column Name	Key	Nullable?	Sample Row 1	Sample Row 2
svn_id	FK	Yes	claudio	
twitter_screen_name	PK		claudiomartella	ApacheShiro
real_name		Yes	Claudio Martella	
datasource_id	PK, FK		372	370
project_name		Yes		Apache Shiro
details		Yes		<a href="http://shiro.apache.org/">http://shiro.apache.org/</a>
last_updated			2013-03-27 00:00:00	2013-01-24 14:01:56

# How to help

- Anyone can contribute data to FLOSSmole
- We'll help you store it (even if it's very large)
- We'll get you a DOI for it

