

An Open-Source Strategy for Documenting Events:

Our Case Study of the 42nd Canadian Federal Election

WADL Workshop @ JCDL

Ian Milligan
Assistant Professor
@ianmilligan1



UNIVERSITY OF WATERLOO
FACULTY OF ARTS
Department of History

Nick Ruest
Digital Assets Librarian
@ruebot



Why Twitter?

- Potential to reshape multiple avenues of historical research – **318,176 unique users on the #elxn42 hashtag**
- **Multiple use cases**
 - Political historians
 - Military historians
 - Social and cultural historians
- **Sheer scope of this data** – i.e. #IdleNoMore on 11 January 2013 = 1,800 pages!





Not a perfect source,
of course...

Leveraging the Twitter API

- **Time is of the essence**
 - 7 - 9 days: largely inaccessible after this point without bags of money (Gnip); could be shorter, ~ 5 days (#YMMFire example)
- But within the Twitter API window you can:
 - Create your own archives using twarc;
 - Analyze tweets using twarc-report & twarc-utilities;
 - Visualize material;
 - Use this as a seed for web archiving;
 - Share datasets



Concept of Hydration



We can collect – but should we?

Ethical considerations.

- Twitter Terms of Service & Copyright
- But legal ≠ Ethics
- Example of the “Black Twitter Project” at USC
- But if we do not collect, archives will be of powerful people...



**Is it fair to put the onus
on researchers rather
than collectors?**

The Case Study

- 2015 Canadian federal election
- Collection began on 3 August 2015, day before the formal writ dropped;
- Collection ended on 5 November 2015, the day after Justin Trudeau became 42nd Prime Minister of Canada
- #elxn42 hashtag

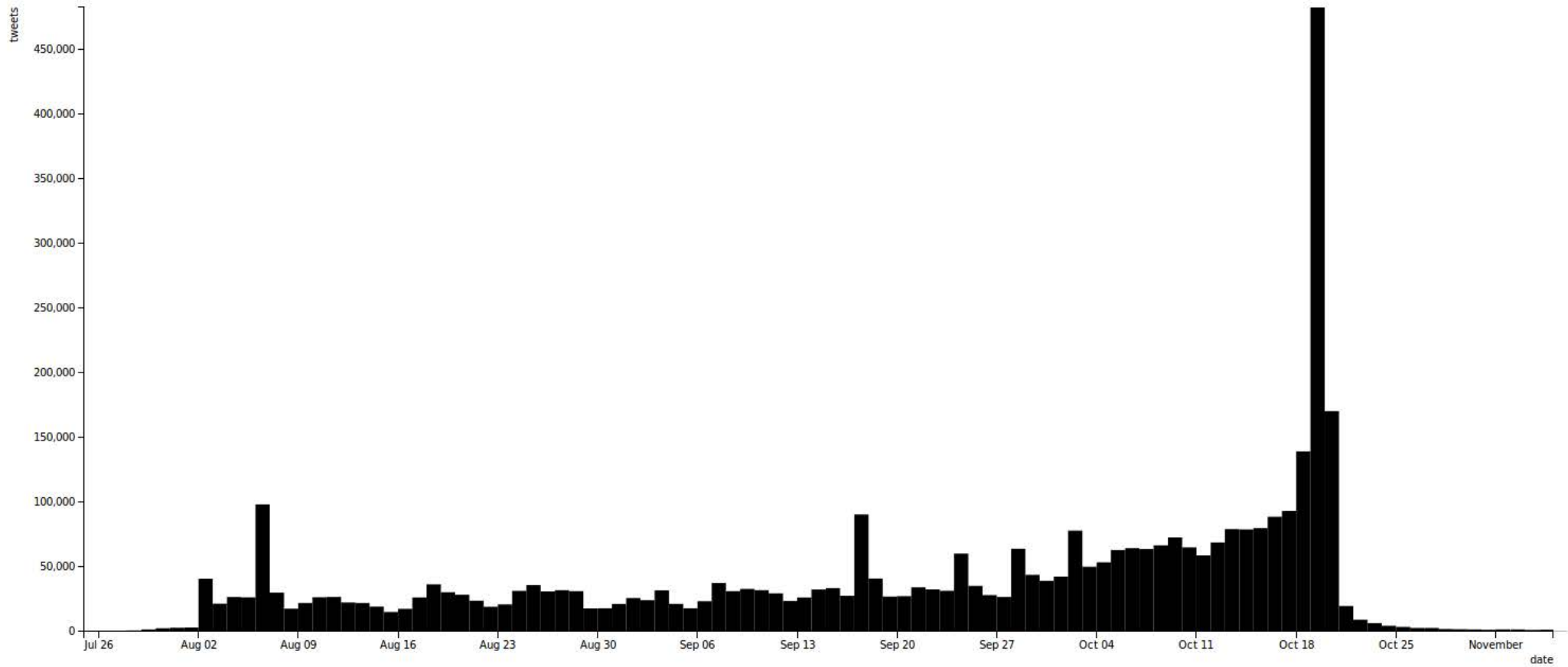


```
$ twarc.py --search "#elxn42" >  
    elxn42-search.json
```

```
$ twarc.py --track "#elxn42" >  
    elxn42-stream.json
```

Frequency on the #ELXN42 Hashtag

2015-07-25 17:56:45 EDT to 2015-11-05 06:46:45 EST



[illegible][illegible][illegible]

	Tweets	Username
1.	21423	DavidMorrison17
2.	15527	P_Wog
3.	10812	chuddles11
4.	10051	444_nal4b
5.	8871	JoanneCangal

6.	8346	littleshas
7.	8316	MadeInCanada
8.	8114	LucMatte
9.	7360	Frazzlin
10.	7019	StopHarper7

	Tweets	URL
1.	11956	http://www.cbc.ca/includes/federalection/dashboard/index.h
2.	9712	http://www.conservative.ca/
3.	4562	http://www.votetogether.ca/
4.	3983	http://www.cbc.ca/news/politics/macleans-debate-leaders-20
5.	3926	http://www.elections.ca/Scripts/vis/FindED?L=e&QID=-1&PA
6.	3104	http://www.elections.ca/home.aspx
7.	2812	http://www.theglobeandmail.com/try-it-now/?articleId=268753
8.	2808	https://www.facebook.com/abu.nawaf.581/posts/1020697771
9.	2757	http://dont-be-a-fucking-idiot.ca/
10.	2707	https://www.mypayingads.com/index.php?ref=51826





Liberal majority government

ALERTS 02:19 a.m.

CBC News projects Kent Hehr (LIB) elected in Calgary Centre

Show all alerts

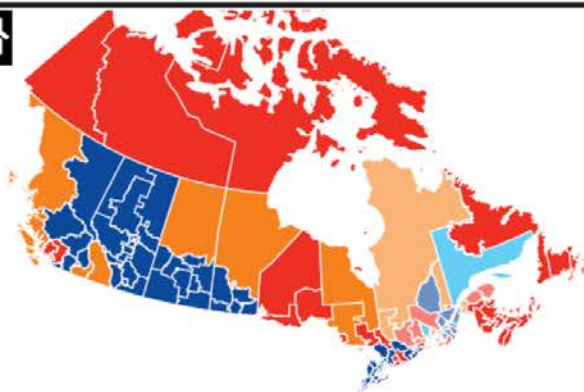
National Results

RIDING TOTALS		VOTE SHARE		
	MAJORITY ELECTED	LEADING	TOTAL	
		184	0	184
		99	0	99
		44	0	44
		10	0	10
		1	0	1

ELECTED LEADING TO COME: 0 RIDINGS

Show all parties

Riding Results



Search by postal code or riding name



MY FAVOURITES



RIDINGS TO WATCH

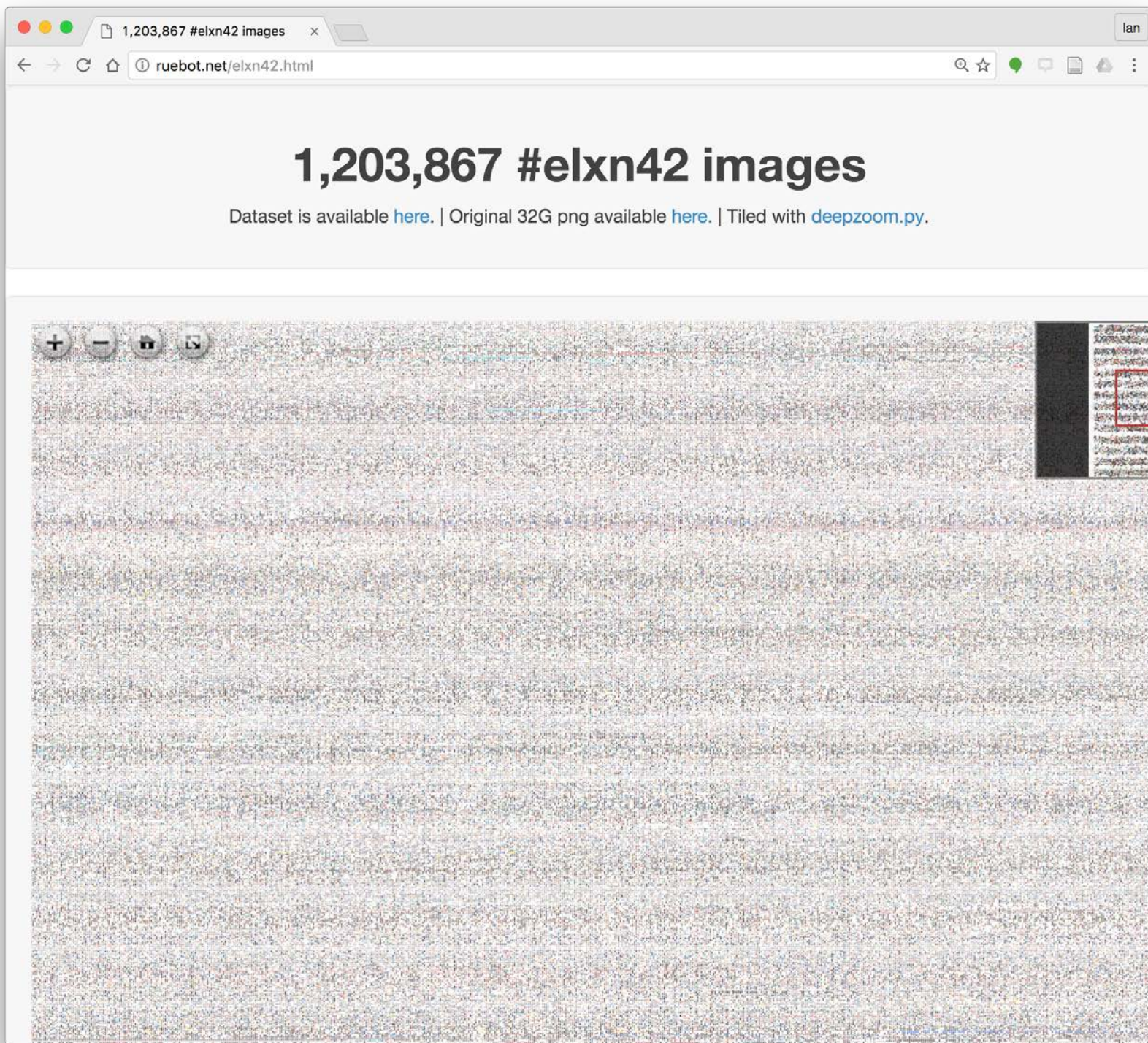
CANDIDATES TO WATCH

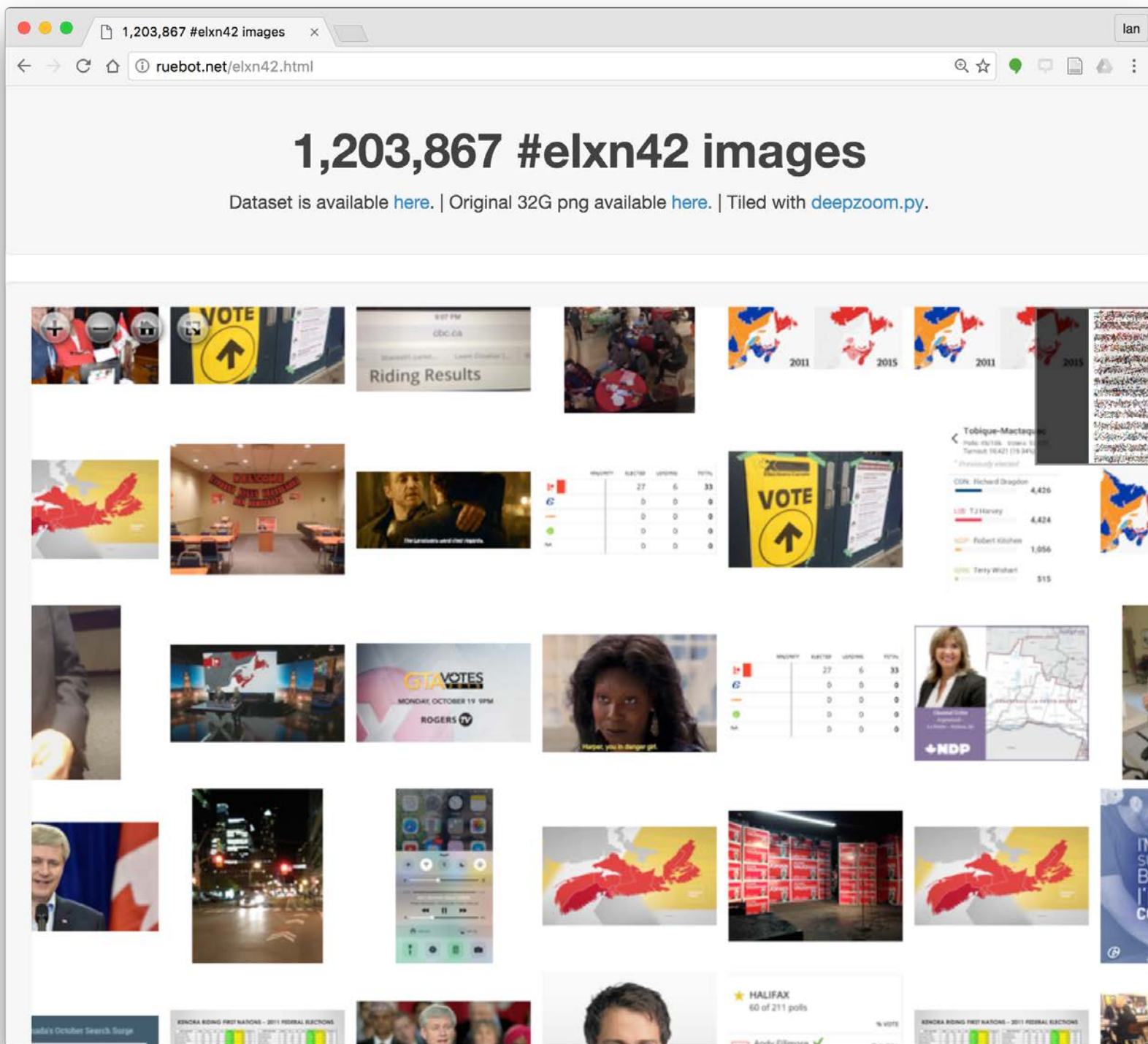
Top Ridings to Watch

☐ Show why these ridings matter

☆ Ajax	✓	LIB
☆ Beauport-Côte-de-Beaupré-Île d'Orléans-Charlevoix	✓	CON
☆ Berthier-Maskinongé	✓	NDP
☆ Brampton Centre	✓	LIB
☆ Brome-Missisquoi	✓	LIB
☆ Burnaby North-Seymour	✓	LIB
☆ Calgary Centre	✓	LIB
☆ Edmonton-Griffith	✓	CON

11,956 tweets





[http://ruebot.net/
elxn42.html](http://ruebot.net/elxn42.html)

#elxn42 tweets (42nd Canadian Federal Election)
lan

← → ↺
dataverse.scholarsportal.info/dvn/dv/wahr/faces/study/StudyPage.xhtml?globalId=hdl:10864/11311
☆

SP Dataverse Network >

POWERED BY THE
 Dataverse Network
PROJECT v. 3.6.2

Create Account
Log In

Web Archives for Historical Research Group Dataverse

#ELXN42 TWEETS (42ND CANADIAN FEDERAL ELECTION)

hdl:10864/11311

Version: 2 – Released: Tue Jan 26 17:34:45 EST 2016

CATALOGING INFORMATION
Data & Analysis
Comments (0)
Versions

Data Citation

i If you use these data, please add the following citation to your scholarly references. [Why cite?](#)

Ruest, Nick; Library and Archives Canada, 2015, "#elxn42 tweets (42nd Canadian Federal Election)", <http://hdl.handle.net/10864/11311> V2 [Version]

Citation Format Print

Data Citation Details	
Title	#elxn42 tweets (42nd Canadian Federal Election)
Study Global ID	hdl:10864/11311
Authors	Ruest, Nick (York University); Library and Archives Canada (Library and Archives Canada, Government of Canada)
Production Date	December 07, 2015
Software	twarc
Funding Agency	Social Sciences and Humanities Research Council
Grant Number	435-2015-0011
Contact	Nick Ruest (York University), ruestn@yorku.ca
Deposit Date	December 07, 2015
Original Dataverse	Web Archives for Historical Research Group Dataverse

Description and Scope

Description

Tweet ids for #elxn42 tweets. Tweets can be "hydrated" with Ed Summers' twarc (<https://github.com/edsu/twarc>). `twarc.py -hydrate elxn42-tweet-ids.txt > elxn42-tweets.json`. Hydrating will recreate the original tweet(s) in json format, provided the content is still available on Twitter. This dataset is the combination of hydrated <http://hdl.handle.net/10864/11310> tweet ids, and <http://hdl.handle.net/10864/11270>.

Related Material	http://hdl.handle.net/10864/11310 http://hdl.handle.net/10864/11270
Date of Collection	August 03, 2015 - November 05, 2015

Deleted Tweets

Twitter Development Agreement & Policy

Spam tweets lost... an invaluable part of the Twitter experience?

Ultimately: **archives are always full of large gaps and omissions; at least here we know that people could make decisions to be removed.**



A cartoon illustration of Scrooge McDuck and his nephews, Huey, Dewey, and Louie, sliding down a large, curved slide made of money. Scrooge is at the top, wearing a red vest and a black hat, holding a steering wheel. The nephews are below him, also sliding down. The slide is covered in green banknotes and gold coins, and the background is a bright blue sky with more money falling around them.

**To do twitter event archiving, you
don't need to be Scrooge McDuck..**



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada



compute | calcul
canada | canada



Thanks!!



UNIVERSITY OF
WATERLOO

Questions?

Ian Milligan
Assistant Professor
@ianmilligan1



UNIVERSITY OF WATERLOO
FACULTY OF ARTS
Department of History

Nick Ruest
Digital Assets Librarian
@ruebot

