

# The Social Media Project in the National Archives Arrangements, objections, obstacles

2019-09-25

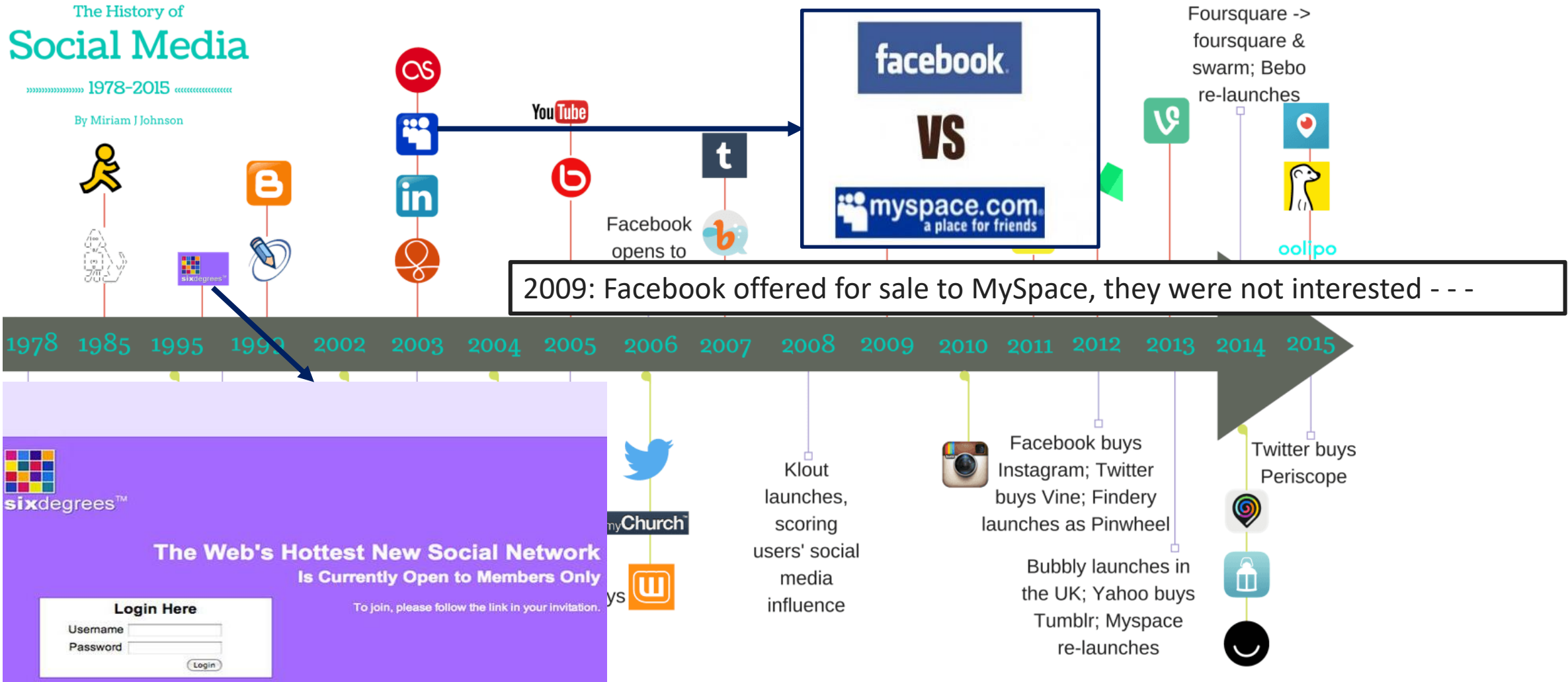
*Espen Sjøvoll & Erik Aaberg, Norwegian National Archives*

# The History of Social Media

## The History of Social Media

1978-2015

By Miriam J Johnson







# Another kind of endangered species

## The 'Bit List' of Digitally Endangered Species

The DPC's 'Bit List' of Digitally Endangered Species is a crowd-sourcing exercise to discover which digital materials our community thinks are most at risk, as well as those which are responsible for compiling and maintaining this list over the coming years, the DPC aims to celebrate great digital preservation endeavors as entries become less of a 'concern,' whilst still highlighting considered 'critically endangered.'


Click on each of the Risk Classifications to see the entries for each category, as assessed in 2017. Following the collation of 2019 nominations, the new edition of the Bit List will be put out in November 2019.

<p><b>LOWER RISK</b></p>  <p>Digital materials are listed as <i>Lower Risk</i> when it does not meet the requirements for other categories but where there is a distinct preservation requirement. Failure or removal of the preservation function would result in re-classification to one of the threatened categories.</p>	<p><b>VULNERABLE</b></p>  <p>Digital materials are listed as <i>Vulnerable</i> when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs. This classification includes <i>Lower Risk</i> materials in the presence of aggravating conditions.</p>	<p>Digital materials are listed as <i>Endangered</i> when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not equipped to meet preservation needs. This classification includes <i>Vulnerable</i> materials in the presence of aggravating conditions.</p>
<p><b>CRITICALLY ENDANGERED</b></p>  <p>Digital materials are listed as <i>Critically Endangered</i> when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs. This classification includes <i>Endangered</i> materials in the presence of aggravating conditions.</p>	<p><b>PRACTICALLY EXTINCT</b></p>  <p>Digital materials are listed as <i>Practically Extinct</i> when the few known examples are inaccessible by most practical means and methods. This classification includes <i>Critically Endangered</i> materials in the presence of aggravating conditions.</p>	<p>Digital materials are listed as <i>Practically Extinct</i> when the few known examples are inaccessible by most practical means and methods. This classification includes <i>Critically Endangered</i> materials in the presence of aggravating conditions.</p>

### Vulnerable

Digital materials are listed as *Vulnerable* when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs. This classification includes *Lower Risk* materials in the presence of aggravating conditions.

In alphabetical order, Digital Species found to be *Vulnerable* are:

<h4>Materials Posted to Current Web-based Social Media Platforms or Equivalents</h4>		
<p>Data generated on web-based social interaction services with asymmetrical terms and conditions and depending on the business model of a single corporate service provider which is in good health but no requirement to preserve. Often provided at no cost to users.</p>		
<p><b>Action:</b> Assessment is Urgent (Within 2 years)</p> <p><b>Assessment Completed:</b> November 2017</p>	<p><b>Examples:</b></p> <ul style="list-style-type: none"> <li>Facebook posts</li> <li>Twitter</li> <li>WhatsApp Messages</li> <li>SnapChat Messages</li> <li>Slack Channels</li> <li>Online political campaigns</li> <li>Blogs and blog comments</li> <li>Online discussion fora</li> <li>Online communities</li> </ul>	<p><b>Aggravating conditions:</b></p> <ul style="list-style-type: none"> <li>lack of preservation capacity in supplier;</li> <li>lack of preservation commitment at or incentive for social media companies;</li> <li>dependence on proprietary products or formats;</li> <li>poor management of data protection;</li> <li>inaccessibility to web archiving services;</li> <li>political or commercial interference;</li> <li>lack of offline equivalent;</li> <li>encryption;</li> <li>infrequent access;</li> <li>conflation of preservation and access;</li> <li>political or commercial interference</li> </ul>
		
<p><b>Endangered</b> in the presence of Aggravating Conditions.</p> <p><b>Lower Risk</b> where good practice can be demonstrated.</p>		

# Background for our efforts

- › Spring 2017 – there is a media discussion:  
“No one preserves our Prime Minister’s tweets”
- › Social media belongs to a “grey area”
  - › Is it worth preserving?
  - › Is the publishing private or “role based”?
  - › Who is responsible?
- › The National Archives initiated a pilot before the 2017 Parliamentary Election

## Kultur

Langlesing fra NRK Kultur Litteratur Musikk Film og serie >

### Solbergs Facebook-meldinger kan gå tapt

Toppolitikeres egne innlegg og kontakt med velgerne på sosiale medier blir ikke dokumentert for ettertiden. Direktør for Nasjonalbiblioteket Aslak Sira Myhre mener det er svært viktig at dette tas vare på for å dokumentere den offentlige samtalen i vår tid.



# Challenges: Volume, Volatility – an example from the Library of Congress



GUT Digital Media Research Centre

Follow

Sign in

Get started

## The Library of Congress Twitter Archive: A Failure of Historic Proportions



Axel Bruns

Follow

Jan 2 · 5 min read

It's dead: the U.S. Library of Congress has officially pulled the plug on its project to create a full, complete archive of all of Twitter—past, present, and future. The Library will now only archive tweets “on a selective basis”.

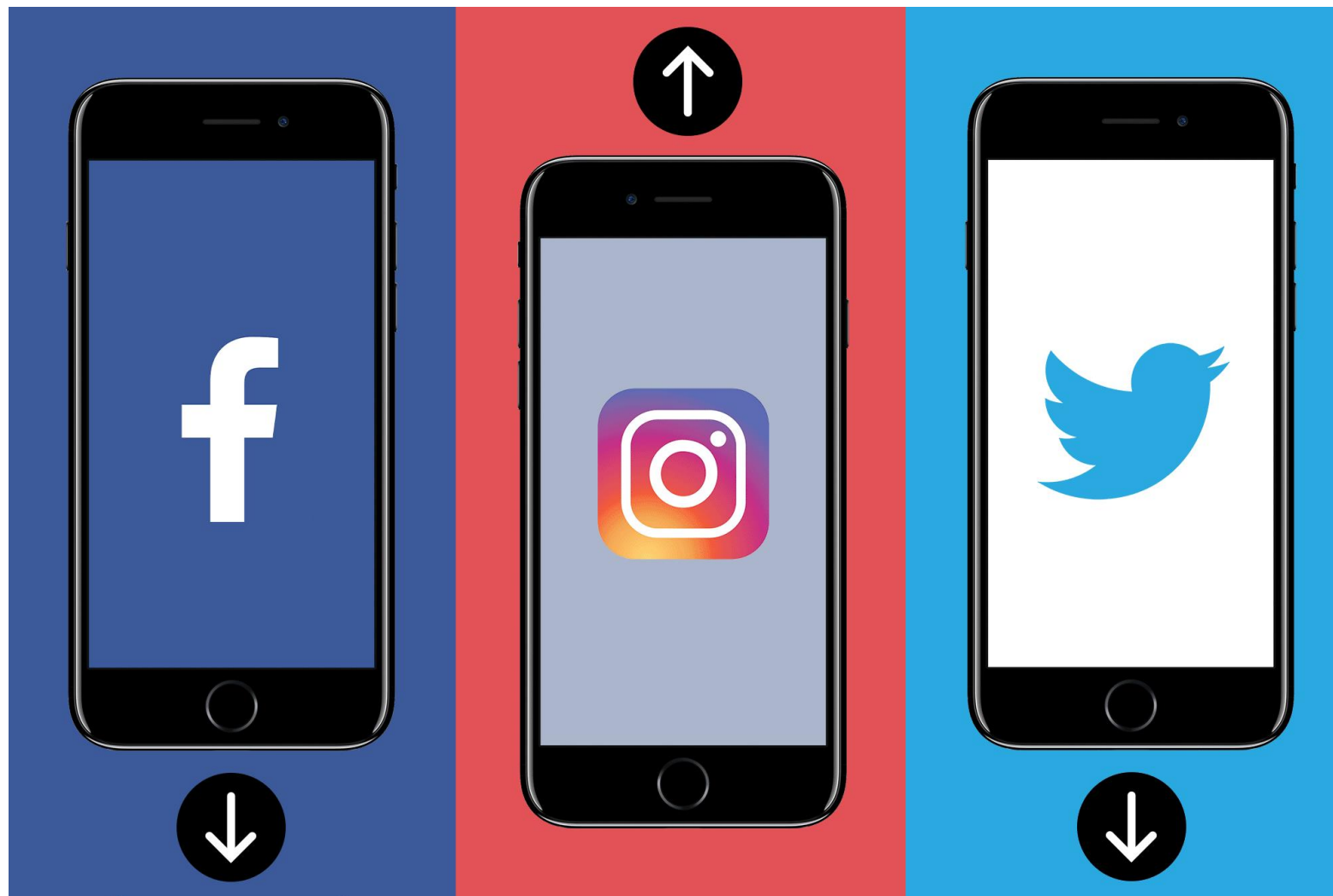
The aborted Twitter Archive project was established after Twitter, in a well-publicised move, gifted the Library access to its tweet archive and live feed in 2010. But unfortunately, beyond the fanfare, the Library never provided the project with the support it required.

### Tweet Preservation

By Biz Stone

Wednesday, 14 April 2010    

## Our (first) choice of channels



# How to: «Web scraping» or use Programming Interfaces?

## Web crawl versus API crawl

- Our main goals
  - Build **representative** collections  $\neq$  all the tweets
    - Choice of accounts and hashtags according to a collection policy
  - Render them **as they were published** on the live web
    - The links between the accounts and hashtags
    - The media (images, videos...)
- Our constraints
  - No funding for an API service subscription
  - Less development on crawl, access and preservation

depth = Page +1



**HERITRIX**

**WayBackMachine**



## Two complementary approaches

	Web crawl	API crawl
Tools for crawl, access and preservation	BCWeb Heritrix 3 OpenWayback SPAR	Open source tools and libraries Specific access and preservation development
Advantages	Render tweets as online Keep the links between accounts Keep the existing workflow (homogeneity and interoperability of collections, preservation)	Developer friendly and easy to use Reliable (until now) Scalability Full metadata
Limits	Number of accounts (3000?) Number of tweets Missing videos Missing short links Few metadata	Public API limits Develop and maintain specific access solution

### API Crawl?

```

[{"contributors": null, "truncated": false, "text": "@NZQandA @nikkikaye @CorinDann They didn't support ACT's charter school ideology either until they needed their vote in Parliament.", "in_reply_to_status_id": 904129509426774017, "in_reply_to_user_id": 904129509426774017, "source": "a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone(a)", "retweeted": false, "coordinates": null, "entities": {"symbols": [], "user_mentions": [{"id": 268005830, "indices": [0, 8], "id_str": "268005830", "screen_name": "NZQandA", "name": "Q+A"}, {"id": 60550677, "indices": [9, 19], "id_str": "60550677", "screen_name": "nikkikaye", "name": "Hon Nikki Kaye"}, {"id": 115541419, "indices": [20, 30], "id_str": "115541419", "screen_name": "CorinDann", "name": "Corin Dann"}], "hashtags": [{"text": "act", "indices": [31, 35], "in_reply_to_screen_name": "NZQandA", "in_reply_to_user_id": 268005830, "retweet_count": 6, "id_str": "904129509426774017", "favorited": false, "user": {"follow_request_sent": false, "has_extended_profile": false, "profile_use_background_image": true, "default_profile_image": false, "id": 25754056, "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/262098666/Twitter_BG.PNG", "verified": true, "translator_type": "none", "profile_text_color": "3c3940", "profile_image_url_https": "https://pbs.twimg.com/profile_images/8675831566471072/M9VWuGw_normal.jpg", "profile_sidebar_fill_color": "FFC2C2", "entities": {"url": {"urls": [{"url": "https://t.co/IeS37dreVV", "indices": [0, 23], "expanded_url": "http://www.chrishipkins.org.nz", "display_url": "chrishipkins.org.nz"}]}, "description": {"text": "I'm a former member of the Labour Party Education Spokesperson. All tweets authorized by me, 216 Main St, Upper Hut.", "followers_count": 7407, "profile_sidebar_border_color": "FF0000", "id_str": "25754056", "profile_background_color": "E8E8E8", "listed_count": 162, "is_translation_enabled": false, "utc_offset": 43200, "statuses_count": 5316, "description": "MP for Rimutaka. Shadow Leader of the House. Labour Party Education Spokesperson. All tweets authorized by me, 216 Main St, Upper Hut.", "friends_count": 1609, "location": "New Zealand", "profile_link_color": "B80000", "profile_image_url": "http://pbs.twimg.com/profile_images/8675831566471072/M9VWuGw_normal.jpg", "following": false, "geo_enabled": false, "profile_banner_url": "https://pbs.twimg.com/profile_banners/25754056/1495682921", "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/262098666/Twitter_BG.PNG", "screen_name": "chrishipkins", "lang": "en", "profile_background_tile": false, "favorites_count": 445, "name": "Chris Hipkins", "notifications": false, "url": "https://t.co/IeS37dreVV", "created_at": "Sun Mar 22 00:35:27 +0000 2009", "contributors_enabled": false, "time zone": "Wellington", "protected": false,
  
```

## What we get – (is what you will see) – and what we store - I

- Profile (who/what)
- Channel
- Time and date (“post time” and our “created time”)
- Party affiliation (if any)



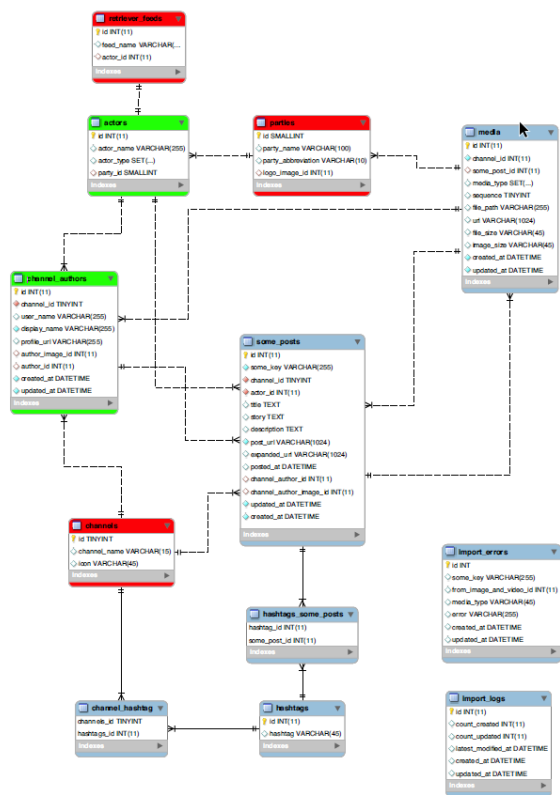
## What we get – (is what you will see) – and what we store - II

WYSIWYG

- Title, description, content
- Author (link) – as appearing in the post
- Screen Dumps and Thumbnails
- Link to original post (might be a “dead” link)

# And - - - visualised

## (Text) Tables



## Post images



## Screen Dumps



## Author images



## Thumbnails

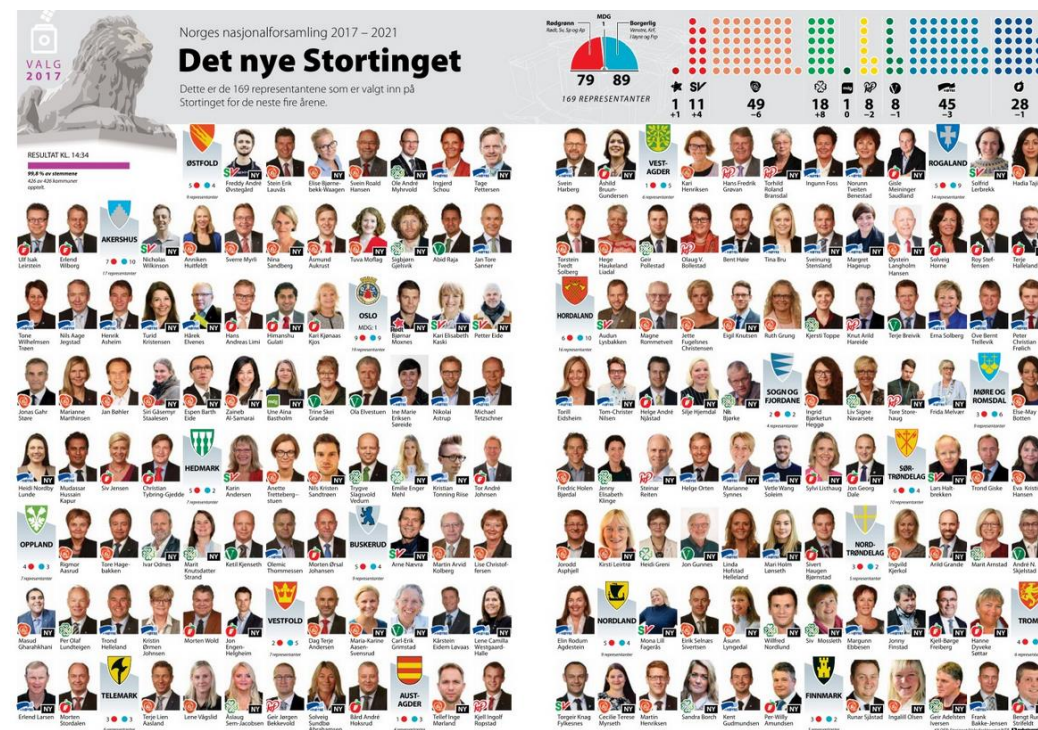


## «Feeds» (Profiles) harvested

- Activity for all political parties represented in our Parliament
  - - - and their party leaders



- - - - and all representatives in the Parliament



## Statistics reflect political activity

Activity - Party/Leaders - 14/8-11/9 ("Election kick off to Election").				
	Facebook	Twitter	Instagram	Sum
Statsministerens kontor	0	12	2	14
Høyre	199	64	10	273
Erna Solberg	130	64	1	195
Arbeiderpartiet	185	112	907	1204
Jonas Gahr Støre	66	21	0	87
FRP	122	38	205	365
Siv Jensen	108	1	55	164
SP	56	2	6	64
Trygve Slagsvold Vedum	16	0	0	16
KRF	131	26	645	802
Knut Hareide	56	13	20	89
Venstre	88	374	398	0
Trine Skei Grande	61	246	3	310
SV	88	536	80	704
Audun Lysbakken	61	222	26	309
MDG	107	529	26	662
Rasmus Hansson	0	114	0	114
Rødt	140	47	16	203
Bjørnar Moxnes	71	33	0	104
<b>Sum</b>	<b>1685</b>	<b>2454</b>	<b>2400</b>	<b>5679</b>



# An infamous post from Sylvi Listhaug (Minister of Justice) - - -

beta.some.arkivverket.no 80%

statsborgerskap til fremmedkrigere LEGG TIL FILTER SØK

**KILDE**

alle

facebook  twitter  instagram

**PERSON / AKTØR**

Sylvi Listhaug

**TIDSPERIODE**

FRA: 01.01.2017 TIL: 07.12.2018

Nullstill filter

Søket ga 2 treff

Arbeiderpartiets standpunkt ble klart sent torsdag kveld, og jeg la ut mitt innlegg morgenvisten dagen etter. Vi lever i en usikker tid hvor vi trenger nødvendige verktøy for å bekjempe terror og trygge landet vårt. Hvis Ap ikke liker å få kritikk for at de stemmer ned dette forslaget, er de velkommen til å snu. For oss handler dette om å raskt kunne frata statsborgerskap og pass fra terrorister og fremmedkrigere. Andre land har lignende bestemmelser, og det er viktig at vi gjør det vi kan for å bekjempe terror. Grunnen til at vi ønsker å utfordre Arbeiderpartiet i denne...

[LES MER](#)

**Sylvi Listhaug**  
2018-03-10  
13:38:00

Facebook

Vi vil inndra pass og statsborgerskap til fremmedkrigere og terrorister raskt og effektivt! Det vil Arbeiderpartiet stemme ned. I kampen mot terror kan vi ikke sitte og toe hendene våre!

[LES MER](#)

**Sylvi Listhaug**  
2018-03-09  
08:35:04

Facebook


beta.some.arkivverket.no/post/563

Hjem / Innlegg fra Sylvi Listhaug

facebook Registrer deg

Sylvi Listhaug

Vi vil inndra pass og statsborgerskap til fremmedkrigere og terrorister raskt og effektivt! Det vil Arbeiderpartiet stemme ned. I kampen mot terror kan vi ikke sitte og toe hendene våre!



PERSON / AKTØR:

Sylvi Listhaug  
[listhaugfrp på Facebook](#)  
Parti: FRP

INNLEGG:

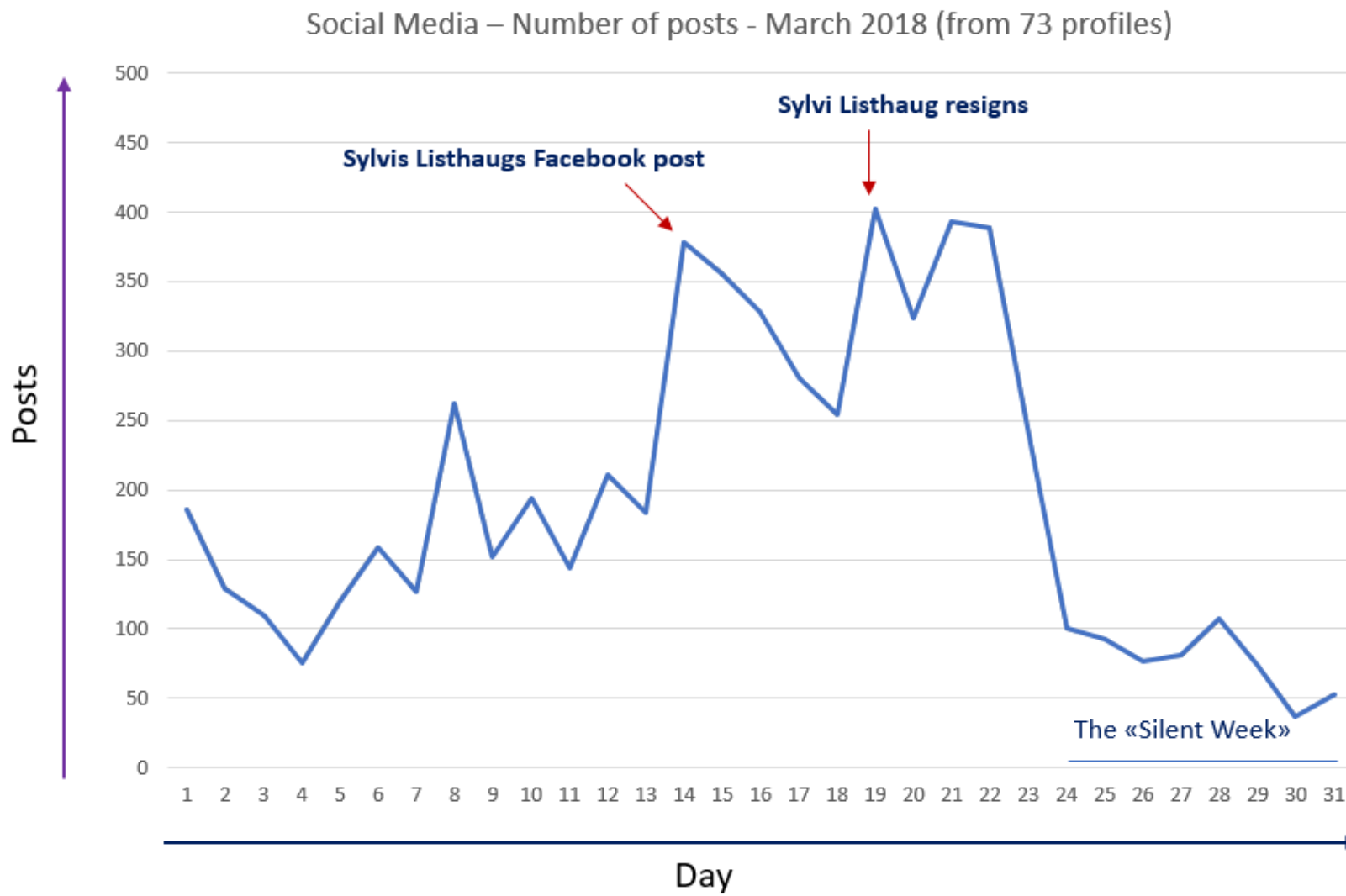
Vi vil inndra pass og statsborgerskap til fremmedkrigere og terrorister raskt og effektivt! Det vil Arbeiderpartiet stemme ned. I kampen mot terror kan vi ikke sitte og toe hendene våre!

POSTET:

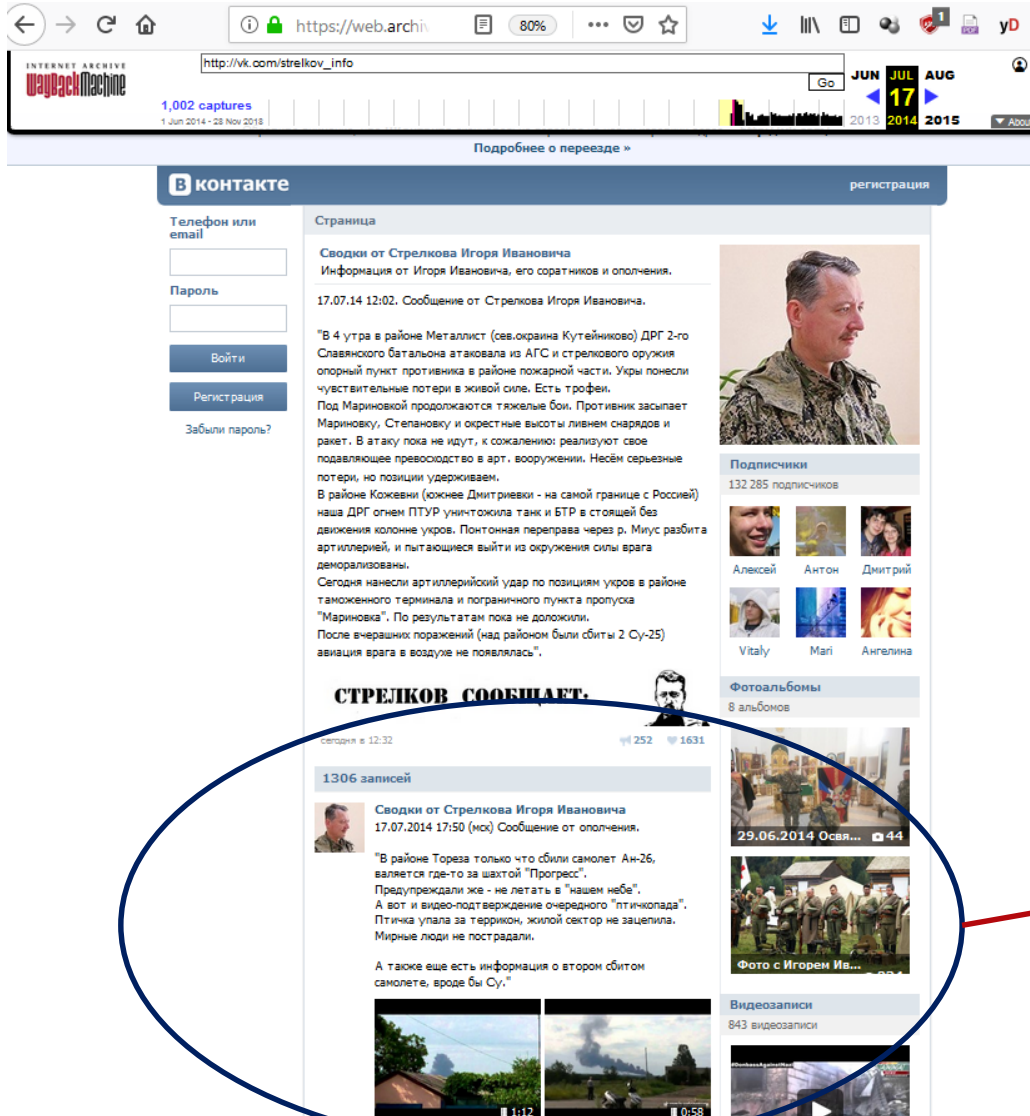
2018-03-09 08:35:04  
[Originalt innlegg på Facebook](#)

- - but – after political pressure removes the post
- - and after a few days resigns





# Another important documentation - Malaysia Airlines MH17



Wayback Machine: http://vk.com/strelkov\_info

**В контакте** регистрация

Страница

**Сводки от Стрелкова Игоря Ивановича**  
Информация от Игоря Ивановича, его соратников и ополчения.

17.07.14 12:02. Сообщение от Стрелкова Игоря Ивановича.

"В 4 утра в районе Металлист (сев. окраина Кутейкиво) ДРГ 2-го Славянского батальона атаковала из АГС и стрелкового оружия опорный пункт противника в районе пожарной части. Укры понесли чувствительные потери в живой силе. Есть трофеи. Под Мариновкой продолжаются тяжелые бои. Противник засыпает Мариновку, Степановку и окрестные высоты ливнем снарядов и ракет. В атаку пока не идут, к сожалению; реализуют свое подавляющее превосходство в арт. вооружении. Несём серьезные потери, но позиции удерживаем. В районе Кожевни (около Дмитровка - на самой границе с Россией) наша ДРГ огнем ПТУР уничтожила танк и БТР в стоящей без движения колонне укров. Понтонная переправа через р. Миус разбита артиллерией, и пытающиеся выйти из окружения силы врага деморализованы. Сегодня нанесли артиллерийский удар по позициям укров в районе таможенного терминала и пограничного пункта пропуска "Мариновка". По результатам пока не доложили. После вчерашних поражений (над районом были сбиты 2 Су-25) авиация врага в воздухе не появлялась".

**СТРЕЛКОВ СООБЩАЕТ.**

сегодня в 12:32 252 1631

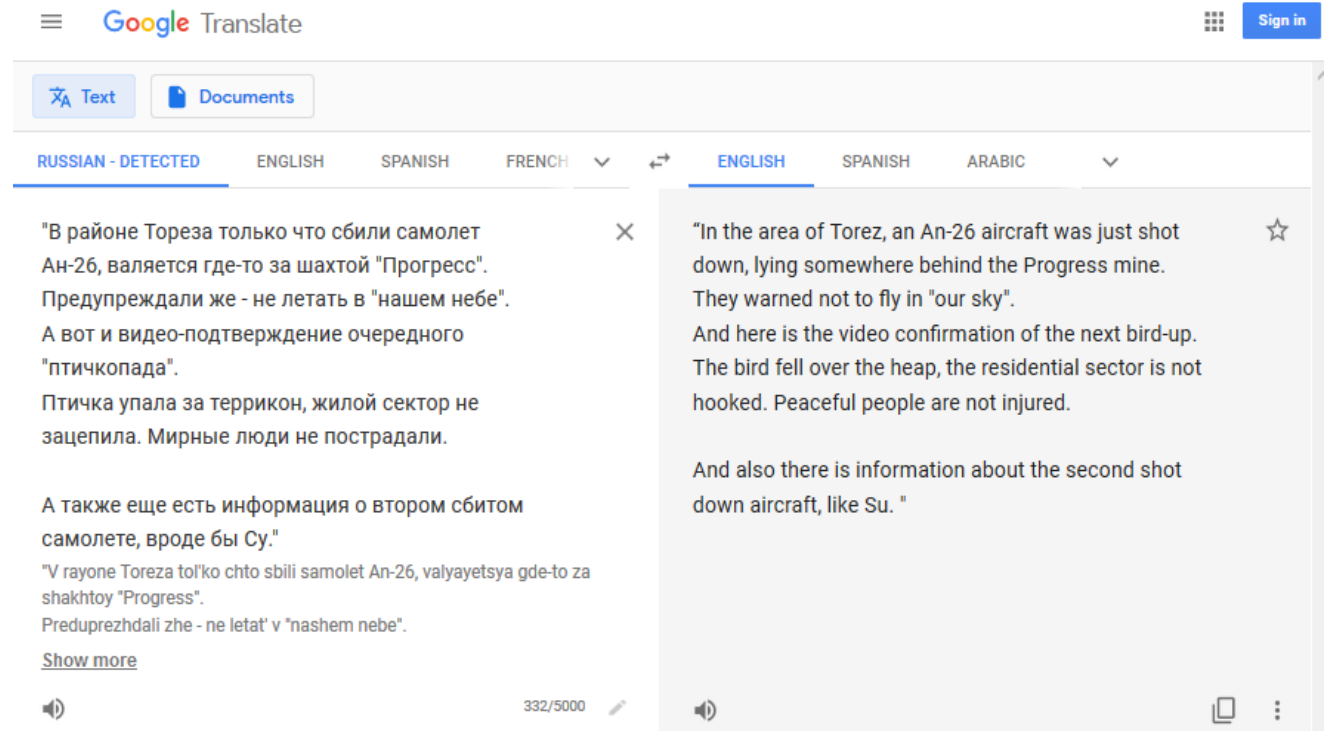
1306 записей

**Сводки от Стрелкова Игоря Ивановича**  
17.07.2014 17:50 (мск) Сообщение от ополчения.

"В районе Тореза только что сбили самолет Ан-26, валяется где-то за шахтой "Прогресс". Предупреждали же - не летать в "нашем небе". А вот и видео-подтверждение очередного "птичкопада". Птичка упала за террикон, жилой сектор не зацепила. Мирные люди не пострадали.

А также еще есть информация о втором сбитом самолете, вроде бы Су."

1:12 0:58



Google Translate

Text Documents

RUSSIAN - DETECTED ENGLISH SPANISH FRENCH ENGLISH SPANISH ARABIC

В районе Тореза только что сбили самолет Ан-26, валяется где-то за шахтой "Прогресс". Предупреждали же - не летать в "нашем небе". А вот и видео-подтверждение очередного "птичкопада". Птичка упала за террикон, жилой сектор не зацепила. Мирные люди не пострадали.

А также еще есть информация о втором сбитом самолете, вроде бы Су."

В районе Тореза, An-26 aircraft was just shot down, lying somewhere behind the Progress mine. They warned not to fly in "our sky". And here is the video confirmation of the next bird-up. The bird fell over the heap, the residential sector is not hooked. Peaceful people are not injured.

And also there is information about the second shot down aircraft, like Su."



# Privacy - GDPR

---

- › What is the rationale for keeping data within our archives?
- › What is the purpose of preservation?
- › Is processing of information necessary?
  
- › Who is formally responsible for processing?
- › Who is doing the actual processing of the data?
- › Who/where is the «legal authority»?
  
- › Preservation vs. publishing

## FAKTA



- › General Data Protection Regulation
- › Strengthens and harmonises the rules for privacy rights when processing personal data within EU/EEA

# A «thread» on Twitter – can it be considered as «Record Management»?

Tja, jeg tenker politisk debatt har godt av det formatet Politico ga for nettopp å reflektere bredt over hvorfor vi står i en situasjon med usikkerhet rundt samarbeid i Europa og rundt Brexit. Det bør jo være rom for refleksjon og ikke bare slagord i politisk debatt.

00:10 - 8. jun. 2018

2 liker

1 1 2

**Andreas Myge** @andmyg · 8. jun. 2018  
Svar til @erna\_solberg

Så for å være helt presis; du vil garantere at @Hoyre vil la spørsmålet om norsk EU-medlemskap bli avgjort via en bindende folkeavstemning? Ja eller nei? #folkeavstemning

3 1 1

**Erna Solberg** @erna\_solberg · 8. jun. 2018  
Folkeavstemninger i Norge er rådgivende, men Høyre er helt tydelig på at vi følger rådet.

1 1 1

**Andreas Myge** @andmyg · 8. jun. 2018  
Takk, det var i grunnen det jeg ville ha svar på. Men i samme slengen; hva tror du folk i Finnmark tenker når de leser svaret ditt her? Var det 86% som sa nei til sammenslåing med Troms?

1 1 4

**Erna Solberg** @erna\_solberg · 8. jun. 2018  
Folkeavstemning for å overprøve et vedtak fattet i Stortinget? Og sammenligner du betydningen av EU medlemskap med sammenslåing av 2 Fylkeskommuner?

1 1

**Andreas Myge** @andmyg · 8. jun. 2018  
Var vel mest opptatt av betydningen en folkeavstemning har for de politiske partienes beslutninger. Kan forstås som et generelt standpunkt når du skriver «folkeavstemninger».

**Erna Solberg** @erna\_solberg Følg

Svar til @SaysPhilippe @Trinesg

Rettstilstanden endres med sivilombudsmannenstolkning, vi mener at så omfattende endringer bør gjøres av lovgiver. Vi foreslår dette i 2 trinn, en unntakshjemmel nå og en mer prinsipiell gjennomgang av hvordan digitalisering og nye hjelpemidler skal reguleres i og. Lov.

13:47 - 9. jul. 2018

1 retweet 2 liker

2 1 2

**Gullfisktanken Philippe Schjelderup** @SaysPhilippe · 9. jul. 2018  
Svar til @erna\_solberg @Trinesg

Jeg skjønner at det er fornytt å få noen som helst gode svar på mine spørsmål.

1 1 4

**Erna Solberg** @erna\_solberg · 9. jul. 2018  
Hva er det du ikke får svar på ?

1 1

**Gullfisktanken Philippe Schjelderup** @SaysPhilippe · 9. jul. 2018  
Hvorfor Høyre har endret sitt prinsipielle standpunkt om mer innsyn i offentlig sektor.

1 1 3

**Erna Solberg** @erna\_solberg · 9. jul. 2018  
Det mener jeg vi ikke har, vi praktiserer meroffentlighet. Denne saken mener jeg dreier som om andre sider som bør belyses godt.

1 1 2

**Gullfisktanken Philippe Schjelderup** @SaysPhilippe · 9. jul. 2018  
Vi må jo kunne være enige om at tilfeldighetene står dere bi i og med at denne ønskede endringen kommer i direkte kjølvann av ønsket innsyn i din kalender. Jeg er stor tilhenger av tilfeldigheter.

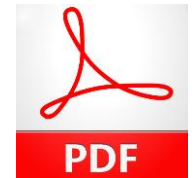
1 1 9

## Long term preservation – how do we keep/store this for the future?

- Keep it as a data base (SQL dump or similar)



- And/or as «flat» PDF files?



- And/or according to the «WARC» standard?  
(Web ARChive – the most common format for storing web pages)



# What Now?



- Expanding the scope?
  - More (types of) profiles (“influencers”, agencies, NGO’s, celebrities - - )
  - Expand social media channels (Flickr, Pinterest, Tumblr - - -)
- Judicial issues – GDPR – how to avoid violation
  - We believe this is OK – as long as we stick to public profiles
- How to pack this for long term preservation and understanding
  - With the OAIS model in mind
- Strengthen relations to partners?

# The Social Media Search is live -

<https://beta.some.arkivverket.no/>

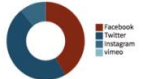
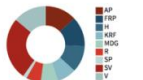
Hva sier stortingsrepresentantene på sosiale medier?

SØK ETTER #TAG @USER ELLER TEKST  FILTER

RT @GirlsNotBrides: Want to learn more about #childmarriage around the world? Check out this fantastic interactive guide from @CFR\_org... <https://t.co/UUpeg3dyrD>

Karin Andersen  
2018-05-10  
19:17:54

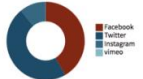
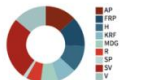
LES MER

RT @Kulturdep: I dag møttes de nordiske kulturministrene i Nordisk ministerråd for første gang siden Sverige overtok formannskapet... <https://t.co/9y4lWak6jH>

Trine Skei Grande  
2018-05-10  
19:10:39

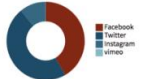
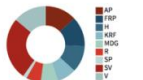
LES MER


RT @efsakariassen: Biskop emeritus Erling Pettersen og jeg er enige om mye, blant annet å forby atomvåpen! @SVpartiet fremmet grunnlovsv... <https://t.co/dLq5ZOehhg>

Karin Andersen  
2018-05-10  
19:10:07

LES MER

Hjem / Innlegg fra Knut Arild Hareide




Jonas Andersen Sayed (@SayedJonas) 8 mar 2018  
Rogaland KrFU har blitt invitert til møte med kunnskapsminister Gibb i Storbritannia! Ting vi skal diskutere er blant annet elevmedvirkning og læreplaner! #studietur

2 retweetet 18 liker

Jonas Andersen Sayed (@SayedJonas) 8 mar 2018  
@SayedJonas @GibbUK @GibbUK @GibbUK

Promotert Tweet  
Facebook (@Facebook) 23 mar 2017  
Tjen opp til 1200 kr per uke på spørreundersøkelser på nett. Bli med gratis i dag! <https://t.co/3k3y72C21g>

AKTØR:

 Knut Arild Hareide  
[KAHareide på Twitter](#)  
Parti: KRF

---

INNLEGG:

RT @SayedJonas: Rogaland KrFU har blitt invitert til møte med kunnskapsminister Gibb i Storbritannia! Ting vi skal diskutere er bla... <https://t.co/ch0ZTO5yu2>

---

POSTET:

2018-03-09 22:33:56 på [Twitter](#)