



Digital Preservation Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

PRESERVING SOCIAL MEDIA

12 February 2016

Strategies for Managing Social Media Research Data

Cambridge

Our digital memory accessible tomorrow

Sara Day Thomson
sara.thomson@dpconline.org

www.dpconline.org



Digital **Preservation** Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Outline

Introduction
Strategies
Challenges
Conclusions and
Recommendations

Tweet

Share your thoughts
on Preserving Social
Media

Me @sdaythomson

DPC @dpc_chat



Machine-readable Data

Application Programming Interfaces (APIs)

Terms and Conditions

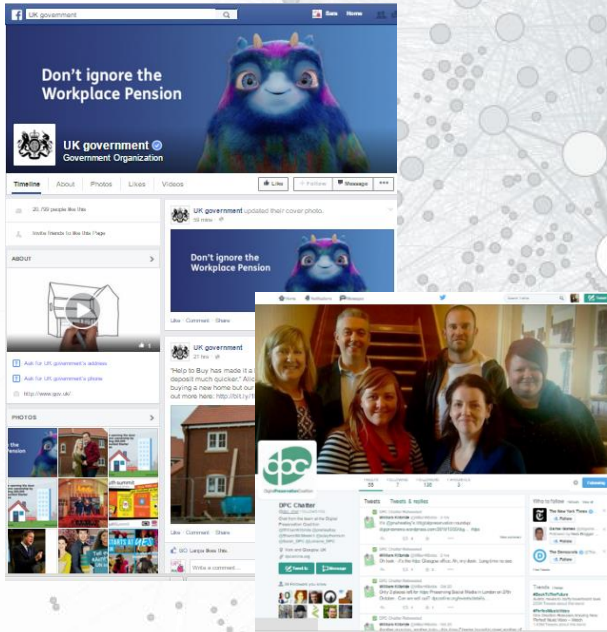




Social Media Data

Datasets

Institutional Accounts



Fare Basis	Airline	Booking Class	Trip Type	Fare	Cabin	Effective Date	Expiration Date	Min / Max Stay	Adv Purchase Req
TA14A0SP	DL	T	One-Way	189.00(USD)	E		02/16/11		14
TA03A0SG	DL	T	One-Way	209.00(USD)	E		02/28/11		03
UA10A0UY	DL	U	One-Way	236.00(USD)	E				10
LA10A0VY	DL	L	One-Way	251.00(USD)	E				10
LA07A0NY	DL	L	One-Way	286.00(USD)	E				07
KA07A0UY	DL	K	One-Way	326.00(USD)	E				07
QA03A0NP	DL	Q	One-Way	396.00(USD)	E				03
HA00A0NY	DL	H	One-Way	466.00(USD)	E				
UC14A0NJ	DL	U	Round-Trip	482.00(USD)	E			V / 30	14
HA00A0UY	DL	H	One-Way	616.00(USD)	E				
MA00A0RY	DL	M	One-Way	694.00(USD)	E				
MA00PNY	DL	N	One-Way	1076.00(USD)	B				
BA00UPRQ	DL	N	One-Way	1348.00(USD)	B				
Y0	DL	Y	One-Way	1435.00(USD)	E				
YUP	DL	N	One-Way	1527.00(USD)	B				
F0	DL	F	One-Way	1828.00(USD)	F				
Y	DL	Y	One-Way	1889.00(USD)	E				
C	DL	C	One-Way	2213.00(USD)	B				
F	DL	F	One-Way	2766.00(USD)	F				

Data Archive





Digital Preservation Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

- Platform APIs
- Data Resellers
- Third-party Services
- Platform Self-Archiving Services
- Research Dataset Identifiers

Strategies



Platform APIs

'Backdoor into platform'

Common Characteristics

- Access Tokens
- Rate Limits
- Samples Sizes
- JSON data
- XML data
- Metadata
(geolocation, user ID,
etc.)



'Walled garden' of social media platforms



Digital Preservation Coalition

UK Data Service



APIs



Test Demo @jondee_test · Apr 30

This is a #test tweet @LoveforTestingT with an image.



```

1 {
2   "created_at": "Thu Apr 30 21:53:11 +0000 2015",
3   "id": 593895901623496700,
4   "id_str": "593895901623496704",
5   "text": "This is a #test tweet @LoveforTestingT with an image. http://t.co/2vgHovKZq4",
6   "source": "<a href='\"http://twitter.com/\"' rel='\"nofollow/\">Twitter Web Client</a>",
7   "truncated": false,
8   "in_reply_to_status_id": null,
9   "in_reply_to_status_id_str": null,
10  "in_reply_to_user_id": null,
11  "in_reply_to_user_id_str": null,
12  "in_reply_to_screen_name": null,
13  "user": {
14    "id": 2993982541,
15    "id_str": "2993982541",
16    "name": "Test Demo",
17    "screen_name": "jondee_test",
18    "location": "Denver, CO",
19    "url": null,
20    "description": "this is a test account.",
21    "protected": false,
22    "verified": false,
23    "followers_count": 2,
24    "friends_count": 43,
25    "listed_count": 0,
26    "favourites_count": 0,

```

Rendered Tweet from Gnip 'Data Format'

JSON of tweet (lines 1- 26 of 201) from Gnip 'Data Format'



Data Resellers

- Commercial service
- Provide data on behalf of platforms
- May provide other services, e.g. analytics
- Cater to corporations
- Built for consumer analysis

Reseller: Gnip



Service: PowerTrack*

Delivers: 100% of tweets

Pricing: \$US 2000 per month plus \$US 0.10 per 1000 tweets delivered

* <https://gnip.com/company/news/press-releases/announcing-powertrack>

Twitter's official reseller : Gnip

Facebook Topic Data reseller: Datasift



Digital Preservation Coalition

UK Data Service



Third-party Services

Common Characteristics:

- Commercial service or partnership
- Deliver archiving services
- Customised Solutions
- Solution for highly specialised technical skills

e.g. ArchiveSocial Standard Pricing Plan



Archive Bundle	Standard	Econom
Trade and save	Our most popular plan	Compliance at any
\$99/month	\$399/month	\$199/month
3,000 new records/month	3,000 new records/month	1,000 new records/month
Unlimited social accounts	Unlimited social accounts	10 social accounts
Unlimited data storage	Unlimited data storage	Unlimited data storage
Standard Plan + RMA add-on (save 15%)	All archiving platform features	All archiving platform features
Start Archiving	Start Archiving	Start Archiving





Digital Preservation Coalition

UK Data Service



Platform Self-Archiving Services

- Service for account owner
- Structured and unstructured data
- Some provide JSON

Facebook

Temperature

Fahrenheit

[Download a copy of your Facebook data.](#)

Twitter

Manage the contacts imported from your address book.

Download your entire Tweet history.

Review the apps that you have given access to your Twitter account.

Review the apps

Google

Your account, your data.
[Download a copy.](#)

Create an archive with your data from Google products.

[Manage archives](#)

[Twitter Archive](#)

[Connected apps](#)



Research Dataset Identifiers

- IDs for entries in a curated dataset
- e.g. in database, in spreadsheet
- Researchers can request dataset with same identifiers
- Not 100% accurate

Tweet (JSON)



```
{  
  "created_at": "Thu Apr 30 21:53:11 +0000 2015",  
  "id": 593895901623496700,  
  "id_str": "593895901623496704",  
  "text": "This is a #test tweet @LoveforTestingT with"
```




Digital Preservation Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

- Platform Terms & Conditions
- Brokering Agreements with Platforms
- Ethics of User Privacy and Awareness
- Selection
- Long-term Preservation

Challenges



Platform Terms and Conditions

Common Policies

- Terms of Service
- User Agreement
- Privacy Policy
- Terms of Use
- Developer Agreement or Developer Policy

Developer Agreement or Policy: Common Characteristics

- Controls use of API
- Forbids sharing of data (incl. cloud storage)
- Forbids preservation of deleted data



Digital Preservation Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Brokering Agreements with Platforms: Twitter

Transfer of Data: Twitter Research Access at the Library of Congress

- Agreement 2010
- Entire archive of tweets from 2006 – 2010
- All on-going tweets
- Authenticated research access only
- Not yet available



Ethics of User Privacy and Awareness

- Accidental disclosure
 - Anonymization
 - De-identification
- User Awareness
 - Commercial Use (profit)
 - Research Use (public good)
- Unequal access
 - Technical Skills
 - Resources

‘... combined, these [conditions] provide a more powerful picture, and thus a more manipulable population.’

–Ralph Schroeder ‘Big Data and the brave new world of social media research’ in Big Data & Society (2014)



Selection

- The 'Conversation'
 - across users
 - across social media services
- Embedded Media
 - images
 - audio
 - moving images
 - URLs
- Uniformity of language, tags, or keywords?
- Spelling errors
- Obscenity or libel
- Shortened links (e.g. bitly, tinyurl)
- Relevance



Long-term Preservation

- Vulnerable to loss
- No legal or regulatory requirement for platform to preserve
- Changes in platform policy and ownership
- Historical data less commercially valuable
- Expensive to curate and store



- Image hosting service
- Shut down by Twitter in Sept 2014
- Data taken by Twitter in Oct 2014
- Archived by Archive Team



Collaboration

Between Collecting Institutions

- Share resources
- Share access
- Share methods and knowledge

Collecting Institutions with Researchers and End Users

- Understand end user needs
- Outreach
- Promote collections

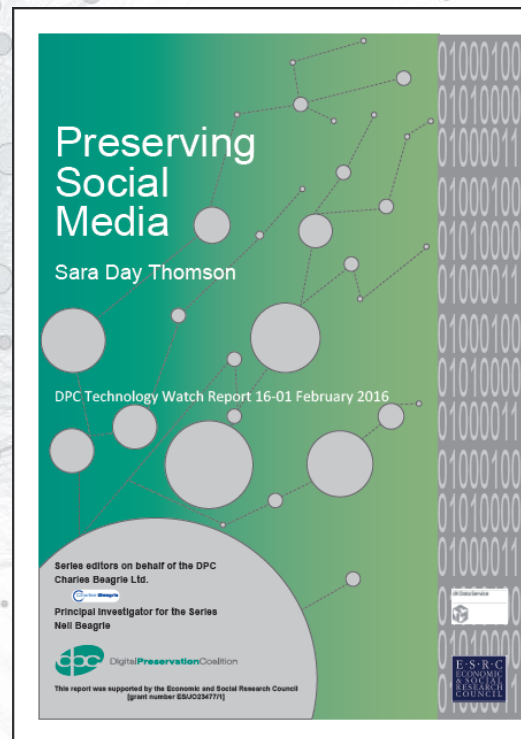


Preserving Social Media

DPC Technology Watch Report

Public Release: Monday 15 February

dpconline.org/publications/technology-watch-reports





Digital **Preservation** Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Look out for...

Preserving Transactional Data: a DPC Briefing Day

17 March 2016

London

*A Joint Workshop with the DPC
and the UK Data Service*

For more info and to register:

bit.ly/1Qx0ibn

Coming soon...

Preserving Transactional Data

*DPC Technology Watch
Report*

Summer 2016



Digital Preservation Coalition

UK Data Service



E · S · R · C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

#thanks!

