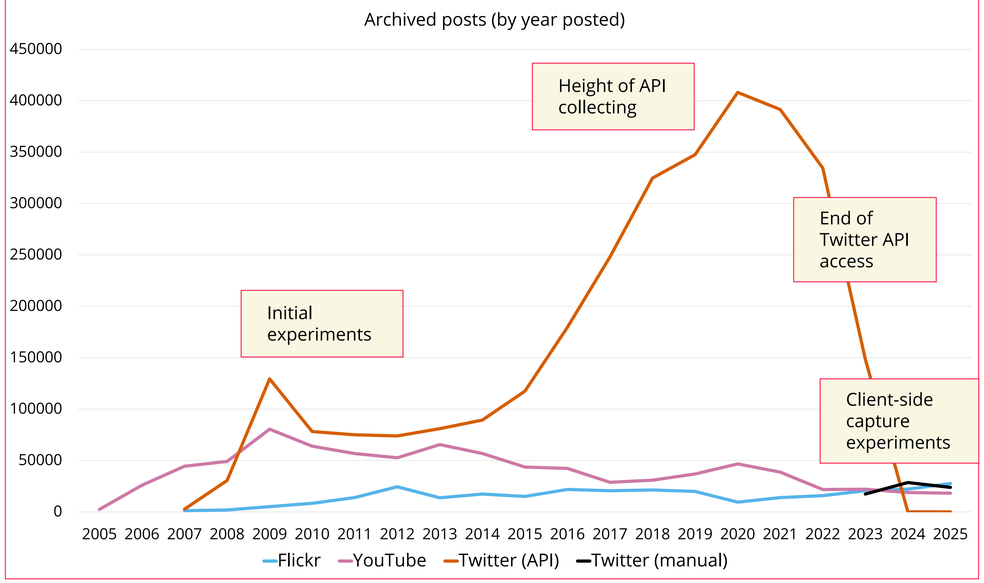
Building a National Archive of Government Social Media

- The National Archives realised in the early 2010s that social media had become an important communication method for government.
- We identified that YouTube, Twitter and Flickr were the most commonly used platforms.
- We agreed to prioritise content over look and feel, this enabled us to avoid copyright concerns.
- Following a research project with our suppliers, we began production archiving of YouTube and Twitter in 2014.
- Content was scraped directly from platform APIs by our suppliers and replayed through a custom interface.
- We began capturing more channels and started capturing Flickr in 2018.
- Also in 2018 launched full text search for archived social media – which we believe was the first in the world.
- We continue capturing YouTube and Flickr via the API method, but were forced to stop using it for Twitter/X in July 2023.

Captures over time (selected platforms)

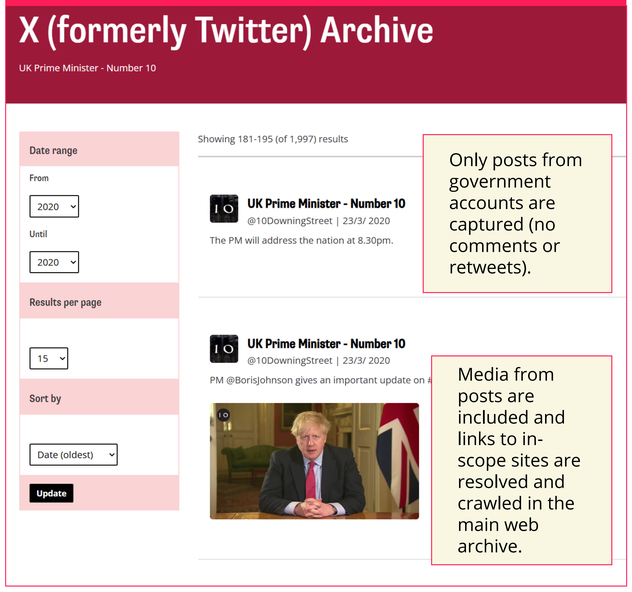


Capturing social media is an ongoing challenge. Other platforms don't allow us to use our established method so we research and productionise alternative technologies for capture. We recently began asking government departments to download and transfer archives of high priority channels – this is a big departure from our usual web archiving practices which involve little work for site owners/managers. There is work ongoing with our suppliers, Mirrorweb, to develop an access system for content captured using the alternative technologies or via direct transfer.

Platforms and Tools

Platform	No of Channels	Capture Technology	Current Status
YouTube	612	API capture	Automated production
Flickr	67	API capture	Automated production
Twitter/X (to July 2023)	931	API capture	Automated production up to 2023
Twitter/X (July 2023 onwards)	195	Gallery DL	Manual production – access to be developed
Twitter/X (July 2023 onwards)	34	Account download by owner	Manual production – access to be developed
Instagram (to June 2026)	467	Instaloader	Manual production up to 2026 – access to be developed, switching to user submission
Facebook (to June 2026)	83	ESUIT	Manual production up to 2026 – access to be developed, switching to user submission
~~Vimeo~~	~~1~~	~~API capture by MirrorWeb~~	~~Experimental~~ – API approach failed
TikTok	3	Gallery DL	Experimental
Other (eg Bluesky)	0	To be decided	Experiments in progress

Access model

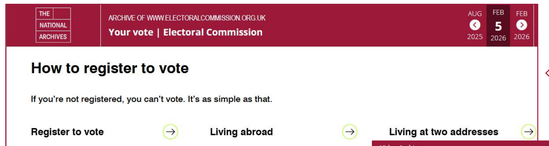


Only posts from government accounts are captured (no comments or retweets).

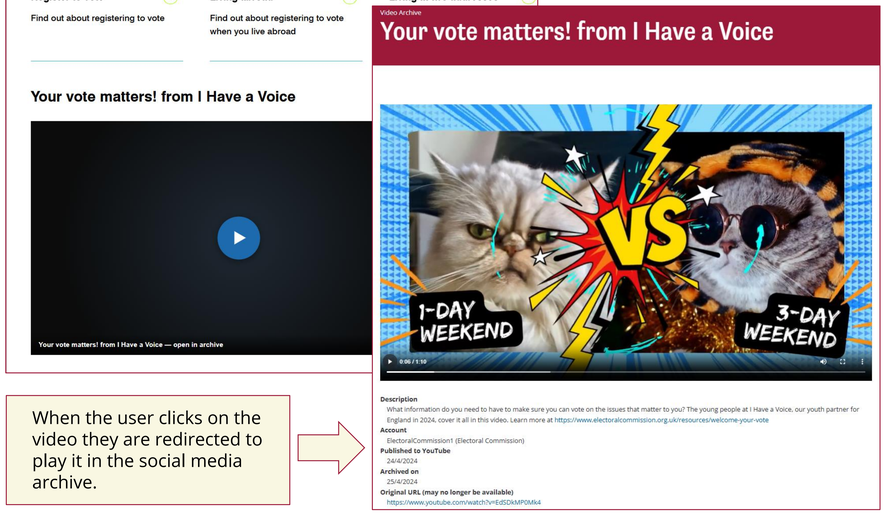
Media from posts are included and links to in-scope sites are resolved and crawled in the main web archive.

Linking the wider web archive to archived YouTube content

In May of this year we worked with our suppliers to implement a script that attempts to automatically link embedded YouTube videos found in the main web archive to the archived version in the social media archive.



Embedded video as it appears in the main web archive. If a thumbnail has been captured that will be shown initially.



When the user clicks on the video they are redirected to play it in the social media archive.

Further reading & similar projects

- British Library blogs on Archiving Social Media with Browsertrix and Webrecorder.
- Cannelli, Beatrice, *Archiving Social Media: A comparative study of the practices, obstacles, and opportunities related to the development of social media archives*.
- Digital Preservation Coalition blog on archiving Facebook, Instagram & LinkedIn.
- National Archives of the Netherlands and National Library of Luxembourg (BnL) IIPC blog series.
- National Archives, UK journal article on the rationale for the social media archive and blog on the history of the archive.
- Royal Belgian Library (KBR) & partners: BESOCIAL project.
- Thièvre, Jérôme; Ben-David, Anat; Cannelli, Beatrice; Langlais, Pierre-Carl & Oogh-Tabanou, Benjamin, *Archiving social media in an age of APIcalypse,* talk at the IIPC Conference 2024 at the BnF

Access the UK Government Social Media Archive

Claire Newing & Jake Bickford, The National Archives (UK)
webarchive@nationalarchives.gov.uk

THE NATIONAL ARCHIVES