

Can a Library Management System (LMS) be used to archive and preserve your social media tweets?

Marlize Palmer, Head of Information and Archive Services, Welsh Government

James Dawes-Hughes, Information and Archive Services, Welsh Government

marlize.palmer@gov.wales

Working with our LMS provider, Soutron¹, the Welsh Government's Information (Library) and Archive Service undertook a proof of concept and/or mini-pilot to make the Welsh Government's "Organ Donation" tweets available via our Library Management System (LMS).

Twitter is a 'micro blogging' platform that allows users to post short text messages up to 280 characters in length and converse with other users via their phones or web browsers. Unlike email or text messaging on mobile phones, these conversations take place publicly, unless a user's tweets are protected. Twitter is experiencing a phenomenal adoption curve in the UK and is being used increasingly by government departments, Members of Parliament, a number of our stakeholders, as well as millions of businesses, non-government organisations and individuals. It is free to use with a relatively low impact on resources and has the potential to deliver many benefits in support of the Welsh Government's communications objectives.

The Welsh Government uses a significant number of separate Twitters feeds on a day-to-day basis, but the three main feeds used are:

- @Welsh Government The English language account
- @LlywodraethCym The Welsh language account
- @FMWales The First Minister of Wales' account

¹ <https://www.soutron.com/>
<https://www.soutron.com/flexible-database-expanded-reach/>

These three accounts contain over twenty thousand separate tweets and provide an overview of the work undertaken by the Welsh Government since the organisation started tweeting in May 2009.

Capturing and archiving social media for re-use is a major challenge. Making the data (tweets) searchable and interpretable is challenging, but it's a necessity, especially when considering that most people use Twitter as one of their main sources of news, information and communication. It is important to find ways of incorporating social media into knowledge structures and archives, as they are valuable resources. Social media records can be used for research, especially when used in context with other digital resources such as email and Word documents (i.e. "linked information").

Tweets could also enrich our collection of [consultation documents](#) and provide additional context to the collections.

Background

On the 1st of December 2015, the Human Transplantation (Wales) Act came into full effect. It introduced a soft opt-out system for consent to organ and tissue donation¹. The Welsh Government (WG) used Twitter extensively to promote its PR campaign to support Organ Donation Wales during a six-month period (1st of June - 30th of November 2016) to increase organ donation registrations, attitude and awareness.

The campaign comprised of weekly schedules of social content (two Facebook posts and one tweet per day) to support monthly themes including: Living Donors; 18-34 year olds; Time to Talk / Organ Donation Week (over 520 posts & tweets in total). It is estimated that the Welsh Government managed to reach a total audience of around 1.5 million via its "Organ Donation Wales" social media campaign.

Twitter



The collection

We decided to trial hosting approximately one hundred and seventy eight tweets and their associated metadata on our LMS. The tweets that we identified for capturing on our LMS were held in the following file formats:

- 1) A PDF screenshot of the tweet in question.
- 2) A txt file containing just the text from the tweet in question.
- 3) An HTML file including the metadata of the tweet in question.

In addition to this, the team used The National Archives' (TNA) file profiling tool DROID² to help profile the file formats in the collection, and to create a simple CSV metadata spread sheet consisting of a file name, short description and type of tweet - see example image.

Complexities and Challenges

The data itself, privacy concerns and the need for us to develop appropriate access and usage policies were some of the complexities and challenges we faced. We explored these issues and the following as part of our mini-pilot:

- How to process and organise these tweets as well as how to physically store them.
- How to provide useful means of access and retrieval.
- Challenges involving policy, such as the creation of appropriate access controls to the collection, whether any information should be censored or restricted (we currently envisage that only WG staff will have access to these tweets).
- What is the minimum/maximum amount of metadata that accompanies each tweet that should be captured? ³
- Is our LMS, Soutron, capable of providing access to these tweets - what are the constraints regarding providing direct access to the data elements within the Twitter Archive, or with any restrictions?
- How much work is involved "processing" tweets as part of a broader collection and what do these changes mean to our current understanding of a record or library item? (i.e. What type of indexing and processing is required by information professionals and/or technology experts to make the collection accessible and re-usable.)
- Moving to second generation digital archiving, what type of sophisticated access tools might be required to provide a "basic level of access" for researchers and users of the collection?
- Broader ethical considerations of the very existence of such a collection (i.e. should there be any access and/or content restrictions and if so would a time limited "take-down" policy be sufficient?)
- Are there any privacy concerns about creating a permanent archive of government tweets and are there any GDPR related issues?
- Should we do this at all (i.e. will we be preserving a mountain of worthless information?)

File Name	Description	Type	Transfer
ESIV-F-4-00000001.pdf	6 mis tan Rhoi Organau	Retweet	Yes
ESIV-F-4-00000002.pdf	100 diwrnod tan Rhoi Organau	Tweet	Yes
ESIV-F-4-00000003.pdf	Datganiad gan Gweinidog Iechyd	Retweet	Yes
ESIV-F-4-00000004.pdf	Rhoi Organau wedi achub bywydau	Tweet	Yes
ESIV-F-4-00000005.pdf	74% yn gwybod am y system Rhoi Organau	Tweet	Yes
ESIV-F-4-00000006.pdf	Oes gennych chi gwestiwn am Rhoi Organau?	Retweet	Yes
ESIV-F-4-00000007.pdf	Statement from Health Minister	Retweet	Yes
ESIV-F-4-00000008.pdf	Paratoadau terfynol ar gyfer Rhoi Organau	Tweet	Yes
ESIV-F-4-00000009.pdf	6 mis tan Rhoi Organau	Retweet	Yes
ESIV-F-4-00000010.pdf	Y system Rhoi Organau Newydd	Tweet	Yes
ESIV-F-4-00000012.pdf	Full scrape of @LlywodraethCym	Feed	No
ESIV-F-4-00000013.pdf	Three tweets with #RhoddwrOrganau	Hashtag	No
ESIV-F-4-00000014.pdf	Bywydau wedi'u hachub	Tweet	Yes
ESIV-F-4-00000015.pdf	Paratoadau terfynol ar gyfer Rhoi Organau	Tweet	Yes
ESIV-F-4-00000016.pdf	Full scrape of @fmwales	Feed	No

Copyright issues

Copyright protection exists at the moment of creation and it is fixed in a tangible form. But copyright does not protect facts, ideas, systems, or methods of operation, although it may protect the way these things are expressed.

The general consensus seems to be that most tweets are falling outside copyright because very few are copyrightable or may rise to the level of copyright protection. Nevertheless, we decided not to capture other people's reply tweets to Welsh Government's original tweets for the pilot.

The Data

We worked with [Hanzo](#) to extract our previously harvested and archived Twitter presence between 2014-2016. This involved technicians at Hanzo writing a code to extract all the relevant tweets.

Four separate searches were undertaken using the following search terms.

- organ donation
- rhoi organau (Welsh language term)
- #organdonation
- #rhoiorganau

Once identified, Hanzo provided the data via download URLs that would allow us to transfer the collection to our systems using Winzip. The data consisted of five different types.

Data Types	
Tweets	Tweets produced by a Welsh Government account regarding organ donations
Retweets	Tweets retweeted by a Welsh Government account regarding organ donations
Feeds	These were screenshots of entire Twitter accounts
Hashtags	These were screenshots of the search results from Twitter using a particular hashtag relevant to organ donation
Replies	Replies from other accounts to Welsh Government Tweets regarding organ donation

After analysing the data (the five separate types shown in the image), we decided that the majority of Tweets and Retweets were worthy of retention, whilst the Feeds, Hash tags and Replies were less likely to have any research or re-use value.

Storage of tweets for research and re-use on LMS

The next step was to work with our Library Management System (LMS) provider, Soutron to upload the collection of tweets onto our LMS to enable us to preserve and re-use them for research purposes. We required our LMS to store and provide access to tweets irrespective of whether they were saved as attachments, images or links and to store the metadata accompanying each tweet.

Soutron started by examining the different variants of metadata and PDFs that were being harvested from the selected Welsh Government's Twitter accounts. The first thing that stood out was that the PDF contained only an image. It didn't include any metadata which, if it had, they then would have been able to automatically index the metadata on

upload of the image. Soutron's index is immediately updated and the search would find the image's content. However, a "text" file was also provided with the PDF file, using a common name, which meant that they could load this alongside the PDF as part of an automated load function. The idea of manually typing the words of each tweet into the LMS would have made the project unworkable.

There was additional metadata that accompanied the Twitter text file and image file, but it was in html format, not ideal to work with. They were able to quickly clean this metadata and load it into Excel so that this could be done. This extra bit of data was important as it contained the URLs of the tweet as well as the referral URL. Critically and very usefully, it contained the name of the PDF file containing the image of the tweet, which meant that they were able to link the metadata with the PDF using the standard Soutron importer tool.

In addition, the simple CSV created by the Welsh Government's Information and Archive Services with the language and description of the tweet provided Soutron with further metadata that could be used for the collection.

We jointly determined the fields that were required and created the required record and field structure using the standard menu driven facilities in Soutron that control the database structure. This included customising a new dedicated Search Result template for tweets.

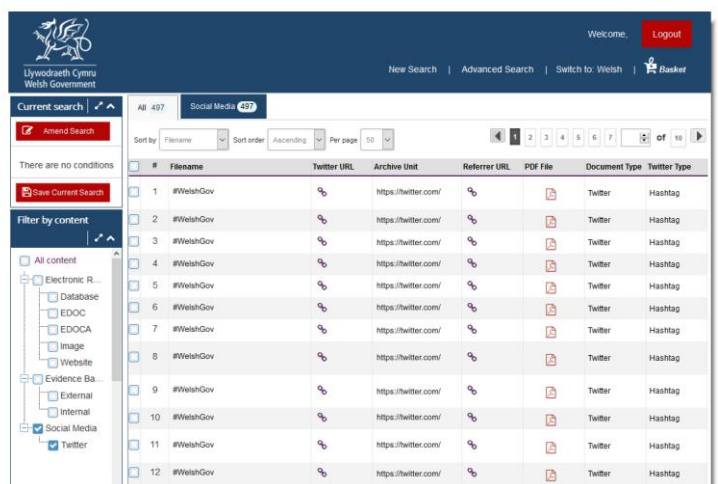
Testing the Results

A test database was used to set up these tasks and to load initial test data sets. The URL to the test system was provided to us to evaluate. The result was very positive and the original objectives were all achieved.

It became clear by experimenting and working with this data set that there is yet further potential. Using Twitter's API's it may be possible to automatically extract and index data from specific

Twitter accounts, to archive important data that otherwise may get lost.

It is exciting to capture and preserve data in this new medium. Librarians and Information Professionals have always been at the forefront of technologies and play a pivotal role to manage vital information, more and more of which is outside of traditional print material. *To explore how Soutron can use their LMS to archive your social media content, get in touch with them today at www.soutron.com.*



#	Filename	Twitter URL	Archive Unit	Referrer URL	PDF File	Document Type	Twitter Type
1	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
2	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
3	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
4	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
5	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
6	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
7	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
8	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
9	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
10	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
11	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag
12	#WelshGov	https://twitter.com/	https://twitter.com/			Twitter	Hashtag

Welsh Government LMS Screenshot of the Soutron LMS Application

Notes:

¹ The Human Transplantation (Wales) Act 2013 (anaw 5) (Welsh: Deddf Trawsblannu Dynol (Cymru) 2013) is an act of the National Assembly for Wales, passed in July 2013. It permits an opt-out system of organ donation, known as presumed consent, or deemed consent. The Act allows hospitals to presume that people aged 18 or over, who have been resident in Wales for over 12 months, want to donate their organs at their death, unless they have specifically objected. The Act varies the Law of England and Wales in Wales (still applicable in England), which relied on an opt-in system; whereby only those who have signed the NHS organ donation register, or whose families agreed, were considered to have consented to be organ donors.

² <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/>

³ More than just the plain text that a user types into the Twitter interface, each tweet contains 150 pieces of metadata, such as a unique numerical ID, a timestamp, a location stamp, IDs for any replies, favourites and retweets that the tweet gets, the language, the date the account was created, the URL of the author if a Web site is referenced, the number of followers, and numerous other technical specifications.