#Datasaveslives: mixed methods analysis of a social media campaign to promote the benefits of using health data for research purposes

Original paper

Authors

Lamiece Hassan[1], PhD.
Goran Nenadic[2,*], PhD.
Mary Tully[3,*], BSc (Hons), MSc, PhD, FRPharmS, FFRPS.

[*]These authors contributed equally.

[1]Division of Informatics, Imaging and Data Sciences, School of Health Sciences, The University of Manchester, Manchester, UK.
[2] Department of Computer Science, The University of Manchester, Manchester, UK.
[3]Division of Pharmacy and Optometry, School of Health Sciences, The University of Manchester, Manchester, UK.

Corresponding author:
Lamiece Hassan, PhD.
Address: Vaughan House, The University of Manchester, Oxford Road, Manchester, England, M13 9PL.
Email: lamiece.hassan@manchester.ac.uk
Telephone: +44 0161 275 1160

Abstract

Background: Social media provides the potential to engage a wide audience about scientific research, including the public. However little empirical research exists to guide scientific organizations regarding what works and how to optimize impact. We examined the social media campaign #datasaveslives, which was established in 2014 to highlight positive examples of the reuse of health data in research.
Objective: The study aimed to examine how the #datasaveslives hashtag was used on social media, how often and by whom; thus, the study aimed to provide insights into the impact of a major social media campaign in the UK health informatics research community and further afield.
Methods: We analyzed all publicly available posts (tweets) between 1 September 2016 and 31 August 2017 on the microblogging platform Twitter that included the hashtag #datasaveslives (n=13,897). Using a combination of qualitative and quantitative analyses, we determined the frequency and purpose of tweets. Social network analysis was used to analyze and visualize the flow of information between hashtag users.
Results: Overall, we found 4,175 original tweets and 9,720 shared tweets ('retweets') featuring #datasaveslives by 3,649 unique Twitter users. In total, 2,756 (66.0%) of original posts were retweeted at least once. Higher frequencies of tweets were observed during the weeks of prominent policy publications, popular conferences and public engagement events. Cluster analysis based on retweet relationships revealed an interconnected series of groups of #datasaveslives users in academia, health services and policy, and charities and patient networks. Thematic analysis of tweets showed that #datasaveslives was used for a broader range of purposes than indexing information, including event reporting, encouraging participation and action, and showing personal support for data sharing.
Conclusions: This study shows that a hashtag-based social media campaign was effective in encouraging a wide audience of stakeholders to disseminate positive examples of health research. Furthermore, the findings suggest the campaign supported community-building and bridging practices within and between the interdisciplinary sectors related to the field of health data science and encouraged individuals to demonstrate personal support for sharing health data.

Keywords: social media; public engagement; social network analysis; medical research;

Introduction

Social media use by academics

Social media platforms such as Twitter, LinkedIn and Facebook have changed the way in which scientists interact with others, both socially and professionally. Though the specifics may vary between individuals, platforms and scientific disciplines[1], common scholarly purposes for using social media among academics include the following: to discover peers and enhance collaboration; to share links or

citations to their own, or others, work; to raise their own profiles; to engage in discussions and keep up to date with scholarly work; to solve problems to track metrics; and to discover job opportunities[2–8].

Numerous studies have examined how scientific communities use social media at conferences and meetings, reporting on outcomes such as promoting attendance[9], disseminating scientific outputs[9], and engaging delegates and wider audiences[10]. Outside of conferences, however, there has been less empirical research into the use and impact of social media as part of the broader communications strategies adopted by scientific organizations. This represents a missed opportunity, given the growing interest in using social media to engage a wide audience about scientific research, including the public[4,11,12].

A social media campaign for health informatics research: #datasaveslives

The social media campaign #datasaveslives[13] was established in 2014 by the Northern England branch of the Farr Institute for Health Informatics Research; a publicly funded, UK-wide research collaboration involving academic institutions and health partners. The campaign started with a simple goal: to promote the positive use of data in health research on social media. A select group of academic organizations belonging to, or affiliated with, the Farr Institute subsequently formally adopted #datasaveslives as part of their communications and public engagement strategies[14]. These supporters then encouraged a wider audience of people who supported health data research to use the hashtag #datasaveslives on social media sites (primarily Twitter) to index and share examples that demonstrated how health data from patient records and other sources could be used to create public health benefits. A second objective was to spark interest and dialogue about using health data for research purposes among wider audiences including patients, members of the public, healthcare professionals and policy makers.

About Twitter

Our study was focused on the microblogging platform Twitter[15]. Twitter allows users to post short messages (previously 140 characters, more recently extended to 280), known as 'tweets', which may include URL links, multimedia content (e.g. images or videos) and/or references to other users (signified using the '@' symbol, plus a username). Hashtags may also be used by assigning the '#' character to a term of their choice; this is a useful way of indexing and finding tweets on a similar topic. Users can view and engage with tweets in a number of ways, including liking, replying to and sharing ('retweeting') others' posts. They can also follow others to automatically see their tweets. Tweets are public by default although users can change their settings at any time to protect their tweets (historical and future) and to restrict visibility to their Twitter followers. Such private accounts were not considered to contain 'publicly available' tweets for the purpose of this study. Users can also choose to write a short description about themselves (known as a 'bio') and add their location.

Study aim

The aim of this study is to examine how the #datasaveslives hashtag has been used in the context of the use of data in health research and by whom. The analysis will determine how often the hashtag has been used and shared and examine the content posted alongside the hashtag to determine the range of purposes for its use. This will provide insights into the strategic use of social media campaigns by academics and explore their potential for encouraging wider dialogues within and between scientific communities and wider audiences.

Methods

Design and objectives

We used a mixed methods design, combining elements of descriptive statistics, social network analysis and qualitative research. This approach, which used qualitative and quantitative analysis in combination, was adopted to allow a richer analysis of Twitter posts, over and above that which could be achieved by commercially available social media analytics tools.

The following objectives were defined:

1. To determine the frequency of tweets and retweets featuring the hashtag #datasaveslives, including the types of tweets that were most frequently shared.
2. To characterize the range and structures of stakeholder groups that use and share #datasaveslives.
3. To identify and explore the different purposes that people used #datasaveslives for when tweeting.

Dataset, variables and definitions

The dataset comprised all publicly available tweets (n=13,897) that included the hashtag #datasaveslives posted over a one year period, between 1 September 2016 and 31 August 2017. This period was selected because it was perceived to represent a peak in campaign activity, thereby providing a sufficiently large and diverse sample of tweets for analysis. These were procured from Twitter's historical data service in January 2018.

The following variables pertaining to the tweet text and metadata associated with the tweet were retained for the analysis: tweet ID; tweet text ('body'); a list of hashtags included in tweets, number of retweets and date posted (recoded into day, month and year).

Each tweet was classified by Twitter as either an original 'post' or a 'share'. Posts referred to tweets where the user either created a new tweet with their own original text or where a user shared another user's tweet and added new text to

accompany it ('quote' tweets). Shares (more commonly referred to as 'retweets') referred to cases when the user had shared a post created by another user with their followers, without changing or adding new text. In all cases, tweets were only included if they referenced #datasaveslives somewhere in the body of the tweet, whether in the shared text or the text newly added by the user.

Where available, we also retained the following data pertaining to individual users who posted tweets, specifically: username; bio (optional self-written text about the user in 140 characters or less); friends count (users they had elected to follow) and follower count (users who had elected to follow them). For analysis purposes, the six user accounts belonging to the sites of the Farr Institute and the Connected Health Cities (CHC) programme, all of whom adopted #datasaveslives as part of their formal strategies, were classified as official supporters (namely @FarrInstitute, @CHCNorth, @HeRC_Farr, @FarrScotland, @FarrCIPHER, @FarrLondon).

Data pre-processing and analysis

Historical Twitter data were pre-processed using Python (version 3.7.2). Briefly, the 'pandas' Python library was used to convert data from JSON format into a two-dimensional dataframe for cleaning, recoding and validation tasks in preparation for data analysis.

Statistical analyses were completed in RStudio (version 1.1.456). Descriptive statistics were used to determine weekly and monthly frequencies of tweets featuring #datasaveslives, and percentages of the most commonly shared tweets (retweets). For the most commonly shared retweets, total reach was estimated by summing the follower count for every user who shared the tweet. Users were also grouped according to tweet frequency and their characteristics were analyzed in terms of median counts for followers, friends and posts. Pearson's R was used to determine associations between weekly counts for posts and retweets.

NodeXL (version 1.0.1.418)[16], a tool for social network analysis, was used to analyze and visualise the flow of information between users of #datasaveslives. We focused on the retweet network as a way of understanding sharing practices between connected users. Statistics about the overall network and individual vertices were generated based on who retweeted whom, including clustering coefficients and betweenness centrality. These were used to produce a network graph visualising the connections (edges) between users (vertices), grouped by cluster using the Clauset-Newman-Moore cluster algorithm. The graph was laid out using the Fruchterman-Reingold layout algorithm. Common hashtags, words and word pairs were also identified for each cluster.

Thematic analysis[17] was used to qualitatively analyze textual content of tweets featuring #datasaveslives. All original tweets accompanying the hashtag were imported into NVivo 12[18] for analysis. Briefly, after reading a sub-sample of tweets to gain familiarization with the data, we defined an initial coding structure,

covering the range of purposes tweets appeared to be used for. Tweets were then coded and recoded into key themes in an iterative fashion by LH. Due to the large size of the dataset, a convenience sample of approximately 1000 tweets were manually reviewed in total, by selecting tweets in no particular order. Tweets were coded until saturation occurred; that is, until no substantially new themes were found. The final set of themes was decided upon on following a discussion between authors.

Ethics and governance

Data were collected and processed in line with Twitter's terms and conditions. As this information was non-sensitive and already in the public domain, formal ethical approvals were not required to complete the project.

On advice from our University's research ethics office and in line with wider social media research guidelines[19], we took several measures to protect the privacy and confidentiality expectations of Twitter users. Firstly, only tweets of users with accounts set to 'public' were included in the analysis. Secondly, with the exception of clearly identifiable public groups, bodies or public officials tweeting in an official capacity (e.g. government organizations, University departments, heads of department) we have refrained from quoting usernames or verbatim text that might render individual users identifiable; rather, their identities have been masked (i.e. indicated via @individualuser) and/or textual content has been paraphrased (indicated by 'paraphrased') to protect anonymity. During the course of identifying popular tweets we discovered that some posts or accounts had subsequently been deleted by users following the time of data collection; in such cases, tweets were not quoted though they were retained for the purposes of aggregated quantitative and qualitative analyses.

Results

Tweet frequency over time

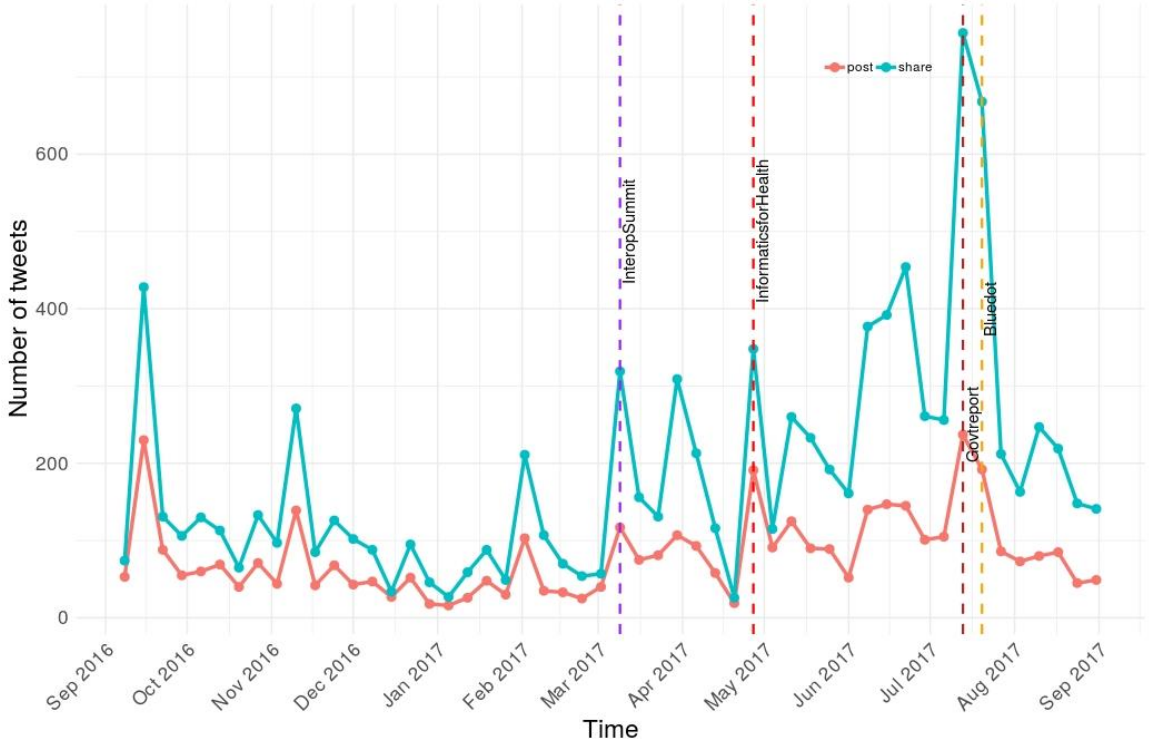
During the 12-month period 1st September 2016 to 31st August 2017, there were a total of 13,895 tweets containing #datasaveslives (Figure 1). Overall, 4,175 (30%) were original posts and 9720 (70%) were retweets. Among original posts, 1,453 (34.8%) were quote tweets.

The mean number of weekly total tweets was 267.3, although this varied substantially (range 43 – 994). There was a strong positive correlation between weekly counts for posts and retweets ($r = .927$, $df=50$, $p<.001$).

The highest number of tweets were observed during the week commencing 6th July 2017 (237 posts and 757 retweets), during which the UK government published a response[20] to a national review of security, consent processes and opt outs relevant to health data[21]. During the same week there were also tweets about

public engagement activities at high-profile cultural festivals in Cheshire (Bluedot Festival, England) and Edinburgh (Edinburgh Festival Fringe, Scotland). There were also high frequencies of tweets from official supporters during the week commencing 20th April (week 34), when there was a health informatics conference (Informatics for Health) hosted in Manchester in England.

Figure 1. Tweet frequency over time by tweet type



Who tweets #datasaveslives?

During the period 1st September 2016 to 31st August 2017, there were 3,649 unique Twitter users who posted or shared content including #datasaveslives (Table 1). In total, 2,756 (66.0%) of the original posts over the year were retweeted at least once. Approximately one in ten (11.3%, n=1,573) of all #datasaveslives tweets, and one in six original tweets (15.4%, n = 421), were by official supporters. Tweet type was significantly associated with official supporter status; official supporters used original posts relatively more often compared to others (26.8% vs. 18.7%;x=57.5, df=1, p<.001).

Table 1. Tweet frequency by tweet type and user type

	Tweet frequency by user type, n(%)		Total no of tweets, n(%)	Total no of unique users[a], n(%)
	Official supporter	Other		
Tweet Type				

Posts					
	Original	421 (26.8%)	2301 (18.7%)	2722 (19.6%)	613 (16.8%)
	Quote	243 (15.4%)	1210 (9.8%)	1453 (10.5%)	551 (15.1%)
Shares					
	Retweet	909 (57.8%)	8811 (71.5%)	9720 (70.0%)	3157 (86.5%)
Total					
		1573 (100%)	12322 (100%)	13895 (100%)	3649 (100%)

[a]Due to the overlap between users who use posts and shares this column does not add up to 100%.

Among the 3,649 users who posted or shared #datasaveslives at any time during the time window observed, 64.9% (n=2,367) did so only once (range: 1-455). Users who tweeted 10 times or more accounted for just 4.9% of users (n=178), yet produced over 50% of tweets. There were 16 users who tweeted 100 times or more. This included five of the six official supporters, plus the accounts of affiliated organizations and projects. Thirteen of the 16 accounts were associated with groups; in addition to official supporter organizations, these included health charities, professional membership organizations, event organizers and projects. Notably, one of these frequent tweeters was a patient advocate/campaigner (n=102 tweets).

How were tweets shared between users?

Six of the ten most frequently shared tweets were from accounts associated with organizations, networks or events (Table 2); only one originated from the account of an official supporter (@HerC_Farr). There was a modest, though significant, positive correlation between retweet count and follower count ($r = .214$, $df=4173$, $p<.001$).

Table 2. The top ten most frequently shared tweets[a]

	Tweet	Username and bio	No of retweets	Group[b]	Maximum follower reach[c]
Rank					
1	Without data, this wouldn't be possible. We welcome the Govt's response to @NDGoffice review #DataSavesLives	@NHSDigital Information and technology for better health and care.	58	G4	274,311
2	#DataSavesLives Our open letter from charities following the Government's response to the Caldicott Review	@wellcometrust We're a charitable foundation that exists to improve health for everyone. We support thousands of scientists & researchers, spark debate & take on big problems.	51	G4	206,171
3	Not available. Tweet deleted by user.	@individualuser1 Bio not available.	50	-	-
4	Remembering Alan Turing today, on his anniversary. An incredible scientist and human being, and an	@HeRC_Farr An academic, NHS & Industry Partnership: Harnessing health data for	41	G1	52,544

	original believer in #datasaveslives	patient and public benefit. #datasaveslives			
5	Better use of data means you don't have to tell your story again and again to doctors and nurses #DataSavesLives	@NHSEngland Health and high quality care for all, now and for future generations.	40	G7	212,478
6	Using patient data is vital to improve health + care for us all #datasaveslives	@NMRPerrin Leading new Understanding Patient Data initiative. Interested in all things data, with a bit of science policy on the side.	38	G4	35,532
7	Come + work with me! Understanding Patient Data team is recruiting a new policy/comms officer #datasaveslives	@NMRPerrin As above.	37	G4	138,791
8	Register now for our Annual Scientific Meeting-Research in the Digital Age #DataSavesLives	@SMHRN1 Scottish MH Research Network-supporting excellence in mental health studies as part of NHS Research Scotland.	36	G3	62,589
9	New #INTEROPen board: an open collaboration of #interoperability networks to drive #OpenStandards in #health & #socialcare #DataSavesLives	@INTEROPenAPI Leading organizations supporting patients clinicians & new care models. Accelerating the delivery of #Interoperability #OpenStandards in health & social care.	33	G2	37,204
10	Not available. Tweet deleted by user.	@individualuser 2 Bio not available.	31	-	-

[a] As of 31st August 2017.
[b] Group numbers cross-referenced with Tables 3 and 4.
[c] Calculated as the sum of followers across all users who retweeted the original post. This method overestimates total reach as it cannot account for the overlap of followers between users.

We visualised retweet relationships between Twitter users in the #datasaveslives topic network (Figure 2). Retweet connections were created when a user shared content by another user that included the hashtag. The analysis of retweets (n=9720) generated a network of 3419 users and 4606 unique connections between pairs of users.

Cluster analysis revealed an interconnected series of groups (n=118). These were arranged in 'hub and spoke' structures, with smaller numbers of relatively more tightly connected users at the center of each group. The four largest clusters or groups (G1-4) contained 65.9% (n=2253) of users and 62.4% (n=2873) of

relationships in the network. Eighty-five groups were very small, containing five or fewer users. Newman's measure of modularity was 0.439, suggesting a moderate level of distinction between clusters.

Figure 2. Retweet network showing relationships between users who tweet and retweet #datasaveslives.

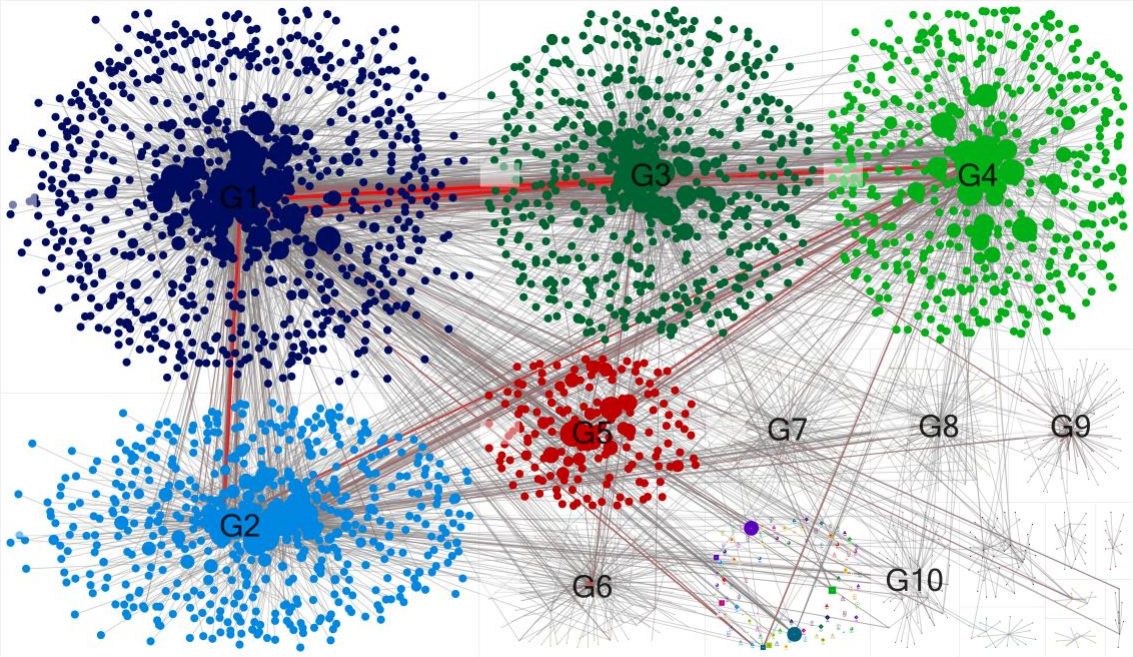


Figure 2 legend: This graph illustrates users among the largest groups that emerged as a result of the cluster analysis, and the connections between them. Twitter users are represented by circular nodes; for nodes in groups 1 to 5 these have been weighted by betweenness centrality, indicated by varying size. Retweet relationships are represented by lines between users, weighted by the number of retweets per pair; thicker, colored lines indicate stronger connections. For brevity and ease of interpretation, we have only used weighted nodes only for the largest five groups (G1-G5, also see Table 3). Smaller groups, including G6 to G10, and isolates (unlabeled) are pictured in the lower right hand side.]

We examined the size, users, hashtags and words associated with largest 5 groups yielded by the cluster analysis (Tables 3 and 4). The largest group (G1) contained five of the six users among the accounts we deemed official supporters (Table 3). Tweets frequently featured hashtags and words related to two major conferences held in Manchester during the period of observation (Table 4). The sixth official supporter (@FarrScotland) was assigned to the third largest group (G3); closer examination indicated these users shared affiliations with places, organizations and events located in Scotland. Group two (G2) included delegates at a major healthcare IT conference (indicated by the event hashtag #interopsummit) and users connected with a major medical records project based in North East England (@GNCR). Group 4 (G4) included users with connections to the NHS, healthcare

policy and major charities (e.g. Wellcome Trust, @NHSDigital), with tweet content suggesting a more applied focus on the goal of using data to improve health. The fifth largest group (G5) was notable for its focus on cancer and on patients; this featured, for example, calls for using data by people who self-identified as patients or carers. This contrasted with other groups which appeared to include professionals connected via geography and/or organizational purpose.

Table 3. Users and unique links in the #datasaveslives retweet user network

	Users n(%)	Unique links n(%)	Top 5 influential users[a]
Group			
G1			
	779 (22.6%)	1071 (23.3%)	@FarrInstitute
			@CHCNorth
			@HeRC_Farr
			@Individualuser3[c]
			@FarrCIPHER
G2			
	526 (15.4%)	653 (14.1%)	@GreatNorthCare
			@Individualuser4
			@InteropSummit
			@INTEROPenAPI
			@CompareSoftware
G3			
	497 (14.5%)	638 (13.9%)	@FarrScotland
			@cancerchallscot
			@ProductForge
			@ndividualuser5
			@IHDPscot
G4			
	451 (13.2%)	511 (11.1%)	@AMRC
			@NMRPerrin
			@Patient_Data
			@NHSDigital
			@Individualuser6
G5			
	156 (4.6%)	142 (3.1%)	@useMYdata
			@abcdiagnosis
			@Individualuser7
			@ICR_London
			@Individualuser8

[a]According to highest betweenness centrality score.

Table 4. Top 5 hashtags, words and word pairs for the #datasaveslives retweet user network

	Top 5 hashtags[a]	n	Top 5 words[a]	n	Top word pairs	n
Group						
G1						
	iforh2017	205	data	1021	health data	181
	data	154	health	808	find out	148
	manchester	76	chcnorth	556	connected health	114
	expo16NHS	74	herc_farr	493	data research	108
	health	60	research	357	health cities	105
G2						
	interopsummit	344	#interopsummit	344	looking forward	82
	interoperability	218	#interoperability	218	#interopsummit #interopwarrior	78
	openstandards	160	care	205	#interoperability #networks	55
	nhs	120	interopenapi	186	#health #socialcare	54
	health	97	together	174	#socialcare together	54
G3						
	cancerdatadive	341	data	455	#cancerdatadive #hackathon	146
	hackathon	151	cancerchallscot	386	cancer data	106
	data	74	cancer	378	innovation challenge	101
	datafest17	60	#cancerdatadive	341	cancerchallscot june	84
	iforh2017	47	June	153	data science	82
G4						
	kfdigital17	80	data	670	patient data	277
	healthinformation	11	health	367	health information	116
	updlaunch	10	patient	314	improve health	103
	sumsch17/ longitudinal/ nhs	8	response	177	data vital	90
			use	176	vital improve	86
G5						
	breastcancer	66	data	181	#datasaveslives #breastcancer	31
	bccww	35	patient	84	patient data	25
	britainagainstcancer	24	#breastcancer	66	patient advocate	25
	caldicott	20	Cancer	52	data gift	21
	data/ usemydata/ research / nhs	8	#bccww	35	#breastcancer #bccww	20

[a]Excluding #datasaveslives.

What did people use #datasaveslives to tweet about?

The thematic analysis of tweet content yielded four key ways in which #datasaveslives was used: to share information and updates; for reporting and discussion at events; to show support for data sharing and; as a call to action. While themes have been described separately for clarity, in practice there was substantial overlap, with the same tweets often being classified under multiple themes (Table 5).

Table 5. Examples of tweets with overlapping themes

				Example tweet(s)
Theme[a]				
A	B	C	D	
✓	✓			Today is #WorldHealthDay - Find out how we work to improve health & care for patients & public here: [link to website] #datasaveslives (@FarrInstitute)
✓		✓		Interesting paper from @[usernames] calls for clarity on conflicting data sharing guidance [link to website] #datasaveslives (@Patient_Data)
✓			✓	We are using patient data to implement learning health systems across the #North. Find out more: [link to website] #datasaveslives (@AMRC)
	✓	✓		The Farr Institute discusses importance of patient data at House of Commons event #APPGMedResearch #datasaveslives [link to website] (@FarrInstitute)
	✓		✓	Thank you to all of our speakers today, to find out more about their work follow @UoLCardioEpi #datasaveslives #LIDASeminar (@LIDA_UK)
		✓	✓	#datasaveslives spread the word, it's no secret and we should all be promoting this [link to website] (@Individualuser9, paraphrased)
✓	✓	✓		We believe #DataSavesLives! As do #interopsummit lecturers VIDEOS of Day 2 lectures on @YouTube [link to website] #interoperability (@InteropSummit)
	✓	✓	✓	If you're at #IforH2017 don't forget to take a selfie with #datasaveslives at our stall (12) - just like [first name] from @[username] [photo] (@FarrScotland)
✓		✓	✓	Help contribute to the latest inquiry by @LordsSTCom into the #LifeSciences #IndustrialStrategy and highlight that #datasaveslives [link to website] (@AMRC)

[a]Qualitative theme descriptions: A. to index and share information; B. for reporting and discussion at events; C. to show support for data sharing; D. as a call to action.

To index and share information

The most common types of posts featuring #datasaveslives, particularly by official supporters and members of groups 1, 2 and 3 (see Figure 2 and Table 3), were

tweets sharing information about users' own projects, research findings and news. These included announcements about new projects or funding, updates on progress and sharing results from research. While some tweets directly referenced peer-reviewed scientific literature by linking to journal publications, more often they were linked to less formal sources, including project websites, case studies, blogs and videos.

> "Thanks to data we know that the smoking ban in Scotland has been a success [link to case study on website] #datasaveslives" (@FarrScotland, Group 3)
> "Highlights from Informatics for Health 2017 by @HeRC_Farr: Watch the video at [website link] #IforH2017 #datasaveslives" (@FarrInstitute, Group 1)

Twitter users also used #datasaveslives to highlight the work of others and signpost wider news and policy developments in areas relevant to health data science. These included news stories published by universities, health service organizations and professional bodies, as well as reports in the popular media, including the local and national press and television and radio programmes.

> "BBC News - Artificial intelligence predicts when heart will fail [link to news report] #DataSavesLives" (@EmpowerD4H, Group 13)

In the vast majority of cases references towards data sharing were positive, or at least neutral; however, there were occasional exceptions.

> "See how @ukhomeoffice uses health information that denies patients healthcare [link to external news site] #DataSavesLives until it doesn't" (paraphrased,@individualuser10, Group 1)

Among tweets in this category, hyperlinks to other websites were very common; indeed, a subgroup of tweets were identified which included simply a hyperlink and the hashtag, indicating use of #datasaveslives as purely an index function. This was mainly used by official supporters.

For reporting and discussion at events

Frequently #datasaveslives was used to tag tweets related to events, including conferences, meetings and public engagement activities. Tweets included promotion of forthcoming events, discussion of past events, or even live reporting and commentary about events, talks and discussions that were currently underway. In the case of larger events, such as conferences, #datasaveslives frequently appeared alongside other official event hashtags (e.g. #iforh2017, #interopsummit). Images of slides, presenters, delegates, visitors and stalls were commonly included alongside the text.

> "Looking forward to workshops meetings and exciting stuff at @ExpoNHS tomorrow #datasaveslives #nhs" (paraphrased, @individualuser11, Group 1)

To show support for data sharing

One further use of #datasaveslives was to demonstrate personal support for sharing health data in general or backing the #datasaveslives campaign itself. Twenty-six users included the text #datasaveslives within their Twitter bio. Many tweets of this type included images of individuals or groups at events pointedly posing with placards, badges or clothing featuring the hashtag.

> "Thanks @GreatNorthCare for the #datasaveslives pin - wearing with pride! [photo]" (paraphrased, @individualuser12, Group 2)

Some tweets included a positive statement about reasons for supporting data sharing, either within the tweet or written on placards pictured in the tweet. The reasons referenced included sharing health data for research, sharing data as part of routine healthcare, or as part of larger projects that combined elements of both. Some tweets within this category signposted wider evidence in supporting data sharing, such as collections of case studies where health data had been used for patient benefit. These were especially common among groups 4 and 5. Some drew on first-hand experiences and opinions.

> "For more examples of how #datasaveslives in mental health read this @MQmentalhealth blog. See our case studies here [link to website] "(@Patient_Data, Group 4)
> ""The type of treatment that I had depended so much on the data of patients who went before me" - patient advocate - #datasaveslives" (@useMYdata, Group 5)

As a call to action

We identified a category of tweets that were used to make requests for others to act, participate or respond in some manner. Commonly, these included advertisements to register for or submit papers to future events, participate in research studies, visit exhibition stands at conferences or to apply for jobs. There were also requests to provide feedback, opinions or information.

> "We're inviting applications for a 2yr Clinical Research Fellow to study for an MD. Cardiology trainees please. #heartattack #datasaveslives" (@UoLCardioEpi, Group 6)
> "Help guide our consent modelling framework: happy to share a copy of your care org's consent forms? TY/please DM #datasaveslives #ontology" (@GreatNorthCare, Group 2)

Discussion

Principal Results

In analyzing tweets over the course of a year, this study has investigated the use of a dedicated hashtag to promote the re-use of health data for research purposes and public benefit. This shows that #datasaveslives came to be adopted by several distinct, diverse, yet interconnected groups in the UK with shared interests in health informatics, policy and research. We found evidence that reasons for tweeting #datasaveslives had evolved beyond the original objective of indexing information to a broader range of purposes including event reporting, encouraging participation and action, and showing personal support for data sharing.

Interpretation and Comparison with Prior Work

We found that users of #datasaveslives included individuals, organizations and communities outside of academia with large numbers of followers and high levels of influence, including government departments, policy makers, patient leaders and groups, NHS organizations, and major charities. However, one stakeholder group that had a weak presence in the networks and discussions we captured was industry. At the time, concerns had been raised about access to patient data by commercial companies, especially where these uses were perceived to be primarily motivated by profit rather than public benefit [22–24]. A major policy paper reviewing data security, consent and opt-outs had been released immediately prior to the period of observation for this study[21], shortly followed by the government's response[20]; indeed, this topic accounted for the two most frequently shared tweets in our analysis (Table 2). It is possible that lack of industry engagement in conversations tagged with #datasaveslives may have reflected hesitance to engage or discuss industry partners during a time of uncertain policy context and public trust. This may have changed over time and could be a fruitful area for future research to explore. Indeed, the government in England has since introduced measures to increase transparency and patient choice over uses of health data, whilst simultaneously calling for increased collaboration between industry and the NHS as part of a broader strategy to enable service transformation, and continued improvement in patient outcomes[25].

One of the principal uses of the hashtag was to share information about the use of health data as part of research and innovation. In using #datasaveslives, the hashtag enabled users to index content and make it more readily retrievable to a wider, not exclusively scientific, audience. Whilst there were many examples of references to the formal scientific literature – both directly (e.g. peer-reviewed scientific papers) and indirectly (e.g. via online case studies) - #datasaveslives was also used to reference a much broader source of information regarding data re-use, in a wide range of formats. This included website materials, news articles, opinion pieces and multimedia content, such as images and videos. Our findings fit with the wider literature which indicates scientists use Twitter not only to disseminate formal

scientific work and engage with other scientists, but also to communicate with broader audiences, including policy makers and the public[3,4].

Analysis of hashtags and themes showed that #datasaveslives was frequently used at events, generating commentary about the proceedings of meetings and events in real-time. So-called 'live tweeting' has become more common at scientific conferences and has the advantage of increasing transparency and rapidly disseminating information among a far larger audience over and above those who physically attend[26,27]. Cluster analysis suggested that particular communities associated with particular events. Using #datasaveslives, either alone or in addition to more specific conference hashtags, might have amplified the reach of information beyond that immediate audience, building links between communities by tapping into a wider conversation about health data science. Also, as a hashtag with continued popularity, using #datasaveslives might have offered greater longevity, avoiding the typically short shelf-life of dedicated conference hashtags.

Beyond indexing and dissemination related purposes, we found people used the hashtag to publicly demonstrate support for data sharing, request help and interact with each other. This is compatible with the wider literature, which suggests that hashtags are used by organizations for encouraging interaction and community-building, both online and offline[28–30]. In the tweets we reviewed, it was clear that some organizations actively fostered opportunities for creating online content, supported through the use of #datasaveslives branded conference stands, placards and products. This particular use of the hashtag seems pertinent given that our period of observation followed the high-profile failure of a major government initiative in England to share patient data[23]. It seems possible that adoption of #datasaveslives could have been used as part of community-building strategies to demonstrate solidarity, develop support networks and increase transparency in the wake of a substantial setback to the field.

Implications and further work

This study has shown how a hashtag campaign can be used to disseminate information, increase transparency of research activities and engage diverse communities of stakeholders. An open and flexible remit arguably allowed #datasaveslives to evolve beyond its original primary purpose of indexing positive examples of health data research. Our findings indicate that it would appear sensible for those wishing to implement similar social media campaigns in the health sector to attend to the strategies recommended in other sectors, such as considering the platform user base, identifying and gaining support from highly connected users, developing engaging multi-media content, tailoring content to target particular population segments and encouraging participation via small concrete actions[31,32]. In addition, we would recommend identifying and attending to the needs of different subgroups of stakeholders and considering how to increase their engagement via bespoke activities and events (online and offline).

Future research would benefit from examining how the use and users of #datasaveslives have changed over time, as well as suitable ways of determining the overall impact of varying strategies on key audiences, such as members of the public to enhance public engagement and scientific outreach.

Limitations

This research benefits from analysis of a near complete sample of #datasaveslives public tweets for an active year during the campaign. Nonetheless, we could not have captured all mentions and uses of #datasaveslives during this period. Private tweets and previously deleted tweets were not part of the sample. We did not attempt to capture content tagged with #datasaveslives on social media sites other than Twitter or other hashtags related to health informatics; these were not the focus of our study. Furthermore our dataset did not capture engagement unless an individual retweeted or used the hashtag themselves. Consequently, we accept that we were unable to quantify, much less categorize, the much wider audience who saw, read or otherwise engaged with tweets, in particular patients and members of the wider public not connected to organizations. These are limitations of the study.

Conclusions

The rise of social media has provided unprecedented opportunities for academic organizations and individual scientists to communicate with a much wider range of stakeholders than ever before, including the public. This analysis has shown that use of a simple hashtag campaign promoted by a core group of academic organizations was adopted by a much wider audience of stakeholders to disseminate and demonstrate support for sharing health data, with evidence to suggest this supported community-building and bridging practices within and between the interdisciplinary sectors related to the field of health data science. Future leaders within the health data science community have stated a vision to be team-based, transparent and inclusive, seeking involvement from a wide range of interdisciplinary stakeholders, including patients and the public[33]. Using such opportunities for social media to contribute towards building networks and engaging in dialogue in open forums would seem eminently compatible with this vision.

Acknowledgements

Lamiece Hassan is funded via a UKRI Innovation Fellowship at Health Data Research UK (ref: MR/S004025/1). Access to the data was funded by Connected Health Cities, which was funded by the Department of Health and Social Care in England. The views expressed are those of the authors and not necessarily those of Health Data Research UK, Connected Health Cities or the Department of Health and Social Care. We thank James Cunningham and Miguel Belmonte for their assistance with data preparation and visualisation. We acknowledge the work of Stephen Melia, University of Manchester, in tirelessly promoting #datasaveslives. He may not have been the first person to coin the phrase ‘data saves lives’, but was (probably) the first to put a hashtag in front of it.

Conflicts of Interest

Lamiece Hassan and Mary Tully were previously funded by the Connected Health Cities programme and the Health e-Research Centre, which founded the #datasaveslives campaign reported on in this paper.

References

1. Holmberg K, Thelwall M. Disciplinary differences in Twitter scholarly communication. Scientometrics 2014 Nov;101(2):1027–1042. doi: 10.1007/s11192-014-1229-3
2. Holmberg K, Bowman TD, Haustein S, Peters I. Astrophysicists' Conversational Connections on Twitter. Bornmann L, editor. PLoS One 2014 Aug 25;9(8):e106086. [Medline]
3. Priem J, Costello KL. How and why scholars cite on Twitter. Proc Am Soc Inf Sci Technol 2010;47(1):1–4. doi: 10.1002/meet.14504701201
4. Côté IM, Darling ES. Scientists on Twitter: Preaching to the choir or singing from the rooftops? Facets 2018;3(1):682–694. doi: 10.1139/facets-2018-0002
5. Bonetta L. Should You Be Tweeting? Cell 2009 Oct 30;139(3):452–453. [Medline]
6. Van Noorden R. Online collaboration: Scientists and the social network. Nature 2014 Aug 13;512(7513):126–129. [Medline]
7. Collins K, Shiffman D, Rock J. How Are Scientists Using Social Media in the Workplace? Goffredo S, editor. PLoS One; 2016 Oct 12;11(10):e0162680. [Medline]
8. Hanrahan B V., Convertino G, Nelson L. Modeling problem difficulty and expertise in stackoverflow. Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion; 2012 Feb 11-15; Seattle, USA. New York: ACM Press; 2012. doi: 10.1145/2141512.2141550
9. Winandy M, Kostkova P, de Quincey E, St Louis C, Szomszor M. Follow #eHealth2011: Measuring the Role and Effectiveness of Online and Social Media in Increasing the Outreach of a Scientific Conference. J Med Internet Res; 2016 Jul 19;18(7):e191. [Medline]
10. Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 2015 May 1;115(5):839–846. [Medline]
11. National Co-ordinating Centre for Public Engagement. 2018. What works: Engaging the public through social media. URL: https://www.publicengagement.ac.uk/sites/default/files/publication/what_works_engaging_the_public_through_social_media_november_2018.pdf [accessed 2019-09-20]
12. Choo EK, Ranney ML, Chan TM, Trueger NS, Walsh AE, Tegtmeyer K, McNamara SO, Choi RY, Carroll CL. Twitter as a tool for communication and

knowledge exchange in academic medicine: A guide for skeptics and novices. Med Teach 2015 May 4;37(5):411–416. [Medline]
13. #datasaveslives. 2019. URL: https://www.datasaveslives.info/ [accessed 2019-09-20]
14. The Farr Institute of Health Informatics Research. 2016. The Farr Institute of Health Informatics Research: Annual Report 2015-2016 URL: http://farrinstitute.org/wp-content/uploads/2018/09/The-Farr-Institute-Annual-Report-2015-2016.pdf [accessed 2019-09-20]
15. Twitter. 2019. Available from: https://twitter.com [accessed 2019-09-20]
16. Smith, M, Ceni A, Milic-Frayling, N, Shneiderman, B, Mendes Rodrigues, E, Leskovec, J, Dunne C. NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016, from the Social Media Research Foundation: https://www.smrfoundation.org. 2010
17. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77–101. doi: 10.1191/1478088706qp063oa
18. NVivo qualitative data analysis software. 2018. QSR International Pty Ltd.
19. Townsend L, Wallace C. 2016. Social Media Research: A Guide to Ethics. URL: https://www.gla.ac.uk/media/media_487729_en.pdf [accessed 2019-09-20]
20. Department of Health, Data Sharing and Cyber Security Team. 2017. Your Data: Better Security, Better Choice, Better Care. URL: www.gov.uk/dh [accessed 2019-09-20]
21. The National Data Guardian. National Data Guardian for Health and Care Review of Data Security, Consent and Opt-Outs. 2016. URL: https://www.gov.uk/government [accessed 2019-09-20]
22. Pollock AM, Roderick P. Trust in the time of markets: Protecting patient information. Lancet 2014;383(9928):1523–1524. [Medline]
23. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May 1;41(5):404–9. [Medline]
24. Wellcome Trust. The One-Way Mirror : Public attitudes to commercial access to health data. 2016. URL: https://www.ipsos.com/sites/default/files/publication/5200-03/sri-wellcome-trust-commercial-access-to-health-data.pdf [accessed 2019-09-20]
25. Office for Life Sciences. Life sciences: industrial strategy. 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf [accessed 2019-09-20]
26. Ekins S, Perlstein EO. Ten Simple Rules of Live Tweeting at Scientific Conferences. Bourne PE, editor. PLoS Comput Biol; 2014 Aug 2014;10(8):e1003789. [Medline]
27. Lister AL, Datta RS, Hofmann O, Krause R, Kuhn M, Roth B, Schneider R. Live Coverage of Scientific Conferences Using Web Technologies. Bourne PE, editor. PLoS Comput Biol [Internet] Public Library of Science; 2010 Jan 29 [cited 2019 Apr 8];6(1):e1000563. [doi: 10.1371/journal.pcbi.1000563]
28. Su LY-F, Scheufele DA, Bell L, Brossard D, Xenos MA. Information-Sharing and Community-Building: Exploring the Use of Twitter in Science Public

Relations. Sci Commun 2017 Oct 4;39(5):569–597. doi: 10.1177/1075547017734226
29. Lovejoy K, Saxton GD. Information, Community, and Action: How Nonprofit Organizations Use Social Media. J Comput Commun 2012 Apr; 17(3):337–353. doi: 10.1111/j.1083-6101.2012.01576.x
30. Liu X, He Q, Tian Y, Lee W, McPherson J, Han J. Event-based social networks: linking the online and offline social worlds. Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining; 2012 August 12-16; Beijing, China. New York: ACM; 2012.
31. Freeman B, Potente S, Rock V, Mciver J. Social media campaigns that make a difference : what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015;25(March, 2):1–8. [Medline]
32. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018; 20(12):e10911. [Medline]
33. Ford E, Boyd A, Bowles JKF, Havard A, Aldridge RW, Curcin V, Greiver M, Harron K, Katikireddi V, Rodgers SE, Sperrin M. Our data, our society, our health: A vision for inclusive and transparent health data science in the United Kingdom and beyond. Learn Heal Syst 2019 Mar 25;e10191. [Medline]