Space Policy 32 (2015) 11–16

Contents lists available at ScienceDirect

Space Policy

journal homepage: www.elsevier.com/locate/spacepol

Trending now: Using big data to examine public opinion of space policy

Wendy N. Whitman Cobb

Cameron University, Department of History and Government, 2800 W. Gore Blvd., Lawton, OK 73543, United States

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form
10 February 2015
Accepted 24 February 2015
Available online 12 March 2015

Keywords:
Public opinion
US space policy
Twitter
Google Trends
Big data

ABSTRACT

This article addresses a timely and widespread issue, that of public opinion and the rise of "big data." Analysts of US space policy have consistently noted the role that public opinion plays in setting the directions for US space exploration. However, the tools that have been used to measure public opinion suffer from serious shortcomings in terms of timing and lack of available data. This paper introduces two new measures of public opinion, Google Trends and Twitter, and details how they can be used to assist in measuring interest in space policy in the American public.

The role that public opinion plays in space policy has been widely acknowledged by scholars, practitioners, and reporters.[1] This is also evidenced in the number of studies, many published here, which have endeavored to examine the contours of public sentiment for space activities [1–5]. However, new options in tracking public opinion of space policy are emerging out of the trend towards "big data," including the use of social media and search engines to track public sentiment for space. This article will outline how scholars have traditionally tracked and measured public opinion for space and introduce two new measures, Twitter and Google Trends. Finally, I will explore how these new tools may be better able to capture public opinion of space exploration in the United States.

In total, this article sets out to do a number of things, primarily among them introduce a new measure of public opinion that could be better used to track interest in space exploration. In introducing such a concept, this article lays out potential uses, not all of them, and merely begins to scratch the surface of what could be possible. While this article is rather critical of the public opinion studies that have been done to this point, it is not a criticism or an indictment of public polling methodology. To a great extent, the methodology that polling firms use to undertake opinion measurement is quite advanced and sufficient. However, when it comes to secondary policy areas like space exploration, questions are simply not asked consistently enough. Without consistent measuring across the public or even across the type of question being asked, how can we be sure that the polls utilized thus far are consistent in their results? In other words, I am not criticizing the use of such results, merely noting that secondary policy areas often are at a public polling disadvantage and suggesting another data source that could be used to fill in the gaps.

Before deciding whether it is even worth the time to discover new ways to measure public interest in space activities, one must show that it *matters* [for an excellent review of this, [6]]. Of note, public opinion has been shown to be important in areas such as defense spending [7,8] and the war in Vietnam [9]. With respect to multiple policy areas, both Page and Shapiro [10] and Monroe [11] have shown that there are significant correlations between American public opinion and policy responsiveness. And to a larger extent, Stimson, Mackuen, and Erikson [12] present evidence that the public opinion-policy link is not limited to institutions such as Congress but exists across government.

With respect to US space policy in particular, public opinion has always held a particular fascination. Gabriel Almond, in "Public Opinion and the Development of Space Technology" in 1960 [13], identified the roll for public opinion by writing "Popular opinion may be viewed as 'latent policy' and 'latent politics.' It not only

E-mail address: wwhitman@cameron.edu.

[1] A bibliography on this topic is available from NASA Headquarters Library at http://www.hq.nasa.gov/office/hqlibrary/pathfinders/opinion.htm.

http://dx.doi.org/10.1016/j.spacepol.2015.02.008
0265-9646/

indicates potential changes in public policy and the political elite, it is a most significant component of that public policy and must be understood and appreciated if a proper estimate of the meaning of that policy is to be made." Launius [2] introduced his research (detailed further below) into public opinion on space policy by detailing the many times he had heard the phrase "if NASA just had the popular support that it enjoyed during the 1960s all would be well." Regardless of these platitudes, research of the kind undertaken for other policies noted above has been sparing with Whitman Cobb [4] and Nadeau [5] identifying particular issue publics for space policy. Interestingly, Steinberg [3] demonstrates that NASA's funding is particularly responsive to public opinion especially when public opinion that spending for NASA is too high.

For some time, political science and government officials have recognized the link between public opinion and public policy. The rather explicit assumption is that when public concern for an issue is high, policy responsiveness occurs. If this is the case, it is important to know the public's opinion on policies such as space exploration.[2] However, given the traditionally used tools to do so, the task is much tougher for an area such as space.

1. Traditional measures of public opinion

The field of public opinion polling has come a long way since *Readers' Digest* polls in the 1930s. While polling experts will still quibble over the finer points of sampling methodology, most policy experts can be relatively certain in the results gathered from organizations like Gallup or Rasmussen.[3] The data gathered from these periodic polls has become important not only for politicians and policymakers but policy analysts and political scientists alike. Unfortunately for most political scientists, undertaking regular polling operations is far out of theirs or their university's budgets. As such, we must rely on other polling organizations to poll for us, hoping that they ask the questions we're interested in at an often or steady enough interval. For most policy areas like economics or international affairs, this is the rule and not the exception but for secondary and tertiary policy areas like space [14], it is the exception and not the rule. The result is that even though the methods are excellent, the data resulting from the polling can make analyzing public opinion on space a difficult prospect.

Previous studies of public opinion on space have utilized various different polling organizations. Launius [1,2] uses data compiled from Yankelovich, Gallup, ABC/Washington Post, CBS/New York Times, and Media General to detail importance of the space program to the American public from the 1960s to late 1990s. While the data demonstrate lackluster American support that is rather dependent on which way the pop cultural winds blow, there are several potential drawbacks to data of this sort. First, the data is not consistently generated using the same question over multiple years. Rather, certain questions are asked at irregular intervals; for example, figure 5 in Launius's 2003 paper, displays the percentage of Americans who favor or oppose government funding for human trips to Mars. This question was asked in June 1961, February 1965, October 1965, July 1967, April 1970, July 1979, July 1994, and July 1995.

This specific problem is representative of the more general issue that space policy analysts have when examining public opinion data on space: space as a policy area is simply not salient or relevant enough to cause major polling organizations to ask questions on the topic regularly. If we truly wish to get a handle on the contours of public opinion on space, questions must be asked at a regular interval, regardless of whether it is salient or not.

A second major problem is Launius' need to compile data from several polling organizations. For example, in exploring whether Americans support human or robotic spaceflight, Launius utilizes data from Yankelovich, ABC/Washington Post, and Gallup (figure 9 and footnote 13 in the Launius paper). It is highly unlikely that all three of those groups asked the exact same question regarding support for human or robotic spaceflight. Further, polling techniques used by the organizations were also likely different with varying margins of error perhaps leading to flawed conclusions on what Americans are saying. However, despite these pitfalls and because of the holes in public opinion tracking for space, Launius and others are forced to put together questions asked by different organizations at different times. This potentially leads to questions of internal validity, let alone questions of external validity.

Another option for public opinion on NASA and US space policy is time series data provided by the General Social Survey (GSS). Utilized by Steinberg [3], Whitman Cobb [4], and Nadeau [5], the GSS asks respondents whether they believe we are spending too little, about right, or too much on the space exploration program. This question was asked every year from 1973 to 1978, and then in 1980, 1982 to 1991, 1993 to 1994, and every two years from 1996 to 2012. This yields a fairly consistent measure of support for space exploration.

Even this question and its answers suffer from the timing in which it is administered. Given the relatively short periods of time in which a policy like space may come to be salient and then recede, asking the question once every two years may simply not be often enough to capture small scale, yet important changes in support for the program. For example, imagine if the question had been asked in the spring of 2003, shortly after the *Columbia* accident. It's possible that a larger number of respondents would have said that the US was not spending enough on space exploration. However, the question was not asked then, but only in 2002 and 2004. By the time the question was asked in 2004, it's likely that the salience of space policy following *Columbia* had declined substantially.

In sum, traditional measures of salience or support for US space policy suffer from a number of problems. Questions are not asked regularly enough by the same polling organizations to give reliable data over a large period of time. Second, even with the GSS, small (or even major) changes in salience are not detectable at the time the issue is at the top of the political agenda. Given these drawbacks, how can newer measures of salience assist analysts and policymakers in gauging the public's interest in space policy?

2. New measures of public opinion

With the rise of "big data" [15], political scientists and others are encountering new sources of data to consider using when it comes to representing public opinion. Two of these, Google Trends and social media (Twitter and Facebook, for example), have already made waves, being used in economic studies tracking macroeconomic indicators such as inflation [16], labor and housing markets [17], consumer behavior [18] and private consumption forecasting [19].

Google Trends offers a panoply of potential data resources broken down by country, state, region, and time. When using the Google Trends tool, researchers can enter a search term of interest (for example, "economy") and Google will report back, on a scale of

[2] While budgets for space activities related to NOAA, national security, Earth-sensing, and other science related activities are often much higher, these types of activities generally have a far lower profile towards the American public. As such, when space policy is usually discussed, it is with a heavy leaning towards the exploration endeavors that are most publicly accessible.

[3] Following the 2012 presidential elections, Gallup conducted a review of their polling methodology in response to the rather wide difference between its poll results predicting presidential votes and the actual votes on Election Day.

0–100, how frequently that term has been searched for relative to the total number of searches performed on Google. This data can be refined on a time horizon and also globally, by region, by country, and by state. Google Trends, then, offers a timely, flexible, and easy-to-use tool by which political scientists can look at salience.

A methodological issue with Google Trends, however, is how the index is measured and for what time period. Because the trend number is calculated relative to the total number of searches, the number can change depending on the time period specified. Take, for example, the Google Trend for "government" from September 3–December 3, 2013 as shown in Fig. 1. The scores for all of the dates in this period are measured against the total number of searches. On October 1, the score is 100 whereas on November 15, it is 4. If we change the time period from 90 days to November 1–30, 2013 as seen in Fig. 2, November 15 now has a value of 85. This difference is caused by the different time periods which are being used to capture the information.

This example also demonstrates the susceptibility of the Google Trends measure to high visibility events such as the government shutdown that began October 1 and lasted 16 days. Because of public interest in the goings on of Washington, D.C., Google searches on government spiked on October 1 meaning that all of the trend data during that search period would be measured against the high yardstick.

While this quirk in Google Trends can be tricky, it is not insurmountable. Instead of tracking or capturing the data on a daily or weekly basis, researchers should capture all of the requisite data at one point in time. While this has the effect of flattening the data, particularly in this case, it eliminates the difference in trend numbers that can be caused by taking the measure at different points in time.

Social media sources such as Twitter and Facebook offer other sources of knowing what people are interested in at particular points in time. Gayo-Avello [20] offers an overview of these sorts of measures and notes some successes in using Tweets to predict box office revenue for movies, among other things. Like Google Trends, Twitter data and Facebook updates offer yet another up-to-the-minute source of what people are interested in and doing, except with greater context. Where Google Trends may offer a search term of phrase, Facebook and Twitter offer a platform where individuals can state exactly why they're interested in such a phrase. Consider the topic of the economy. Where user A may type "economy" into Google to perform a search, they may write "Heard the stories about how bad the economy is lately? Well, I'm still on the job hunt" or "The economy's great! Just got a job!" on Twitter or Facebook. Both posts offer context that researchers would not have been aware of in Google Trends. Researchers should also keep in mind that with any social media or Google measure, there is no way to account for individuals who may search for a term or tweet about a topic multiple times. While on the whole, these people are likely to account for a small proportion of tweets or searches, one still cannot rule out this multiplicity as you could with measures like Gallup.

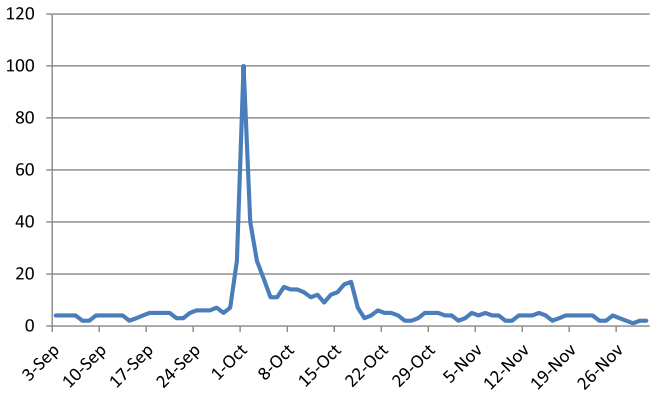

Fig. 1. Google Trends for "government," September 3–December 3, 2013.

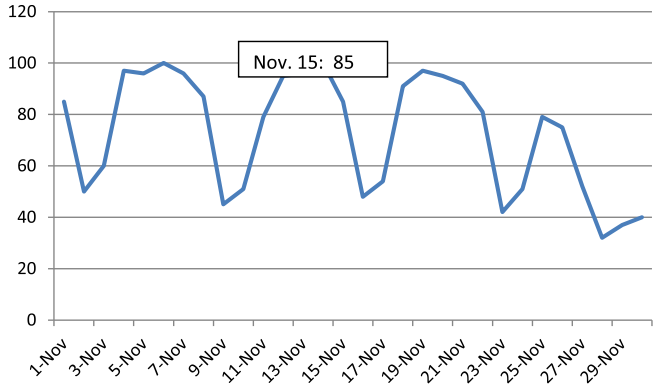

Fig. 2. Google Trends for "government," November 1–30, 2013.

While Facebook and Twitter offer the opportunity to view a bit more context to what individuals are interested in, Laura Granka, a Google researcher, argues that when comparing "relative query volume," the "noise," "will be evenly distributed across all states such that the relative differences will stand out and be most salient to analyses" [21]. In other words, given the large sample, those who search for "economy" but may be searching for "economy cars" will balance out across distributions.

Similarly, these new measures also take advantage of their large numbers in possibly bypassing the difficulties of sampling. This is the argument of Cukier and Mayer-Schoenberger [15] as they explain that prior to having this large amount of data, pollsters have relied on sampling that can break down when you get to smaller and smaller categories, such as "university-educated, single Asian American women under 30" (ibid, 30). Today, however, Facebook, social media, and Google Trends allows researchers to approach "N = all" negating those difficulties. They argue, "Tapping vastly more data means that we can now allow some inaccuracies to slip in (provided the data set is not completely incorrect), in return for benefitting from the insights that a massive body of data provides" [15]. This argument taps into statistical theory that when the N (or number of subjects) is smaller, the margin of error will tend to be higher; the larger the N is, the lower the margin of error will be. As a result, approaching the point where N would include all possible subjects means that the margin of error will be substantially decreased. Thus, Google Trends offers yet another advantage over Gallup in the larger number of "respondents" that are used in a salience indicator.

Despite these benefits to using Google Trends and social media sources as measures of public opinion, researchers still must be careful for a number of reasons. First, there is a question of how representative Google, Facebook, and Twitter users are of the population as a whole. While Google is the ubiquitous search leader in the United States, we cannot be sure that there are not significant differences between those who use Google and those who do not. A 2012 Pew Research found that 67% of internet users utilize Facebook which primarily appeals to women ages 18–29 and that only 16% of internet users are on Twitter which appeals to adults age 18–29, African Americans and urban residents [22]. With respect to Twitter itself, a later 2013 Pew Research survey finds that only 16% of US adults utilize Twitter, meaning that the vast majority of

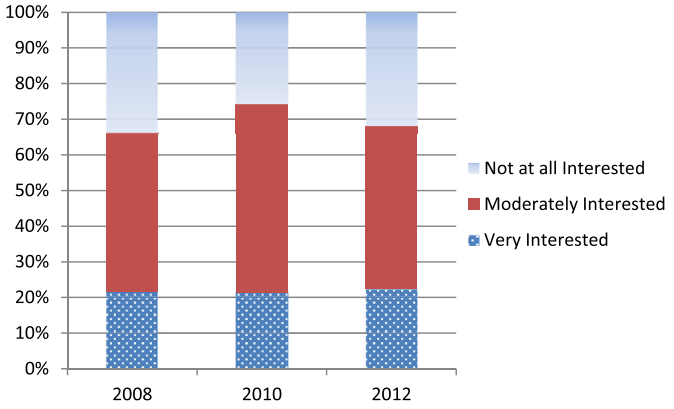

Fig. 3. GSS responses to the question "Are you interested in space?"

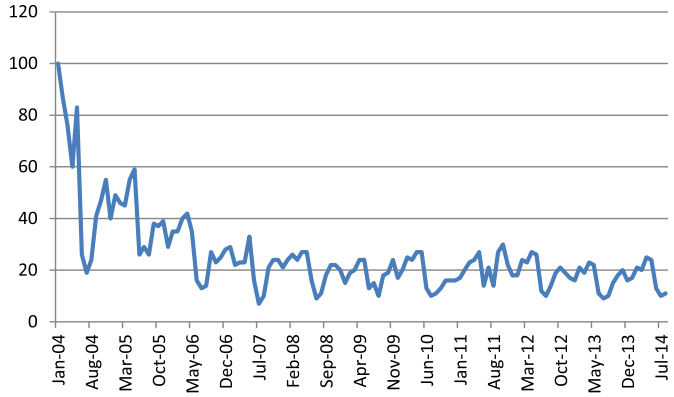

Fig. 5. Google Trends for "Space Exploration," 2004–2014.

Americans, 84%, do not use the social media tool at all. Given these findings, we must be very careful to ensure the representativeness of these measures.

A more interesting finding emerging from the 2013 Pew study where researchers looked at and analyzed Twitter conversations around big news stories and found three major trends: "(1) A core function of Twitter is passing along pieces of information as the story develops. (2) The Twitter conversation about big news events can shift and evolve, both in terms of sentiment and topic. (3) Although sentiment on Twitter can sometimes match that of the general population, it is not a reliable proxy or public opinion" [23]. This last finding is particularly important to note for our purposes here and stems from the fact that Twitter users, because they are a small subset of the American population, may in fact have significant differences between them and the 84% of Americans who do not use Twitter.

Three final concerns have to do with privacy, ethics, and the availability of data. Because of the growing awareness of privacy concerns, Facebook users in particular, have been given greater options in limiting who accesses their accounts [24]. Researchers will not have full access to all Facebook profiles leading again to another question of representativeness. Second, because of the newness of these internet tools, data for Google Trends is only available back to 2004 with Facebook and Twitter data being even more limited. This is compared to both Gallup and the *New York Times* which can be used for salience over a longer period of time. Finally, there is a question of ethics relating to searches of social media in particular and peripherally to Google searches. In Gallup surveys, respondents are aware that their answers to survey questions are going to be recorded and used later on. Social media users are not always informed that this is possible or that it is ongoing; thus the ethically accepted premise of "informed consent" is not always present [25]. While we can search their profiles and record data, the question then becomes whether we should.

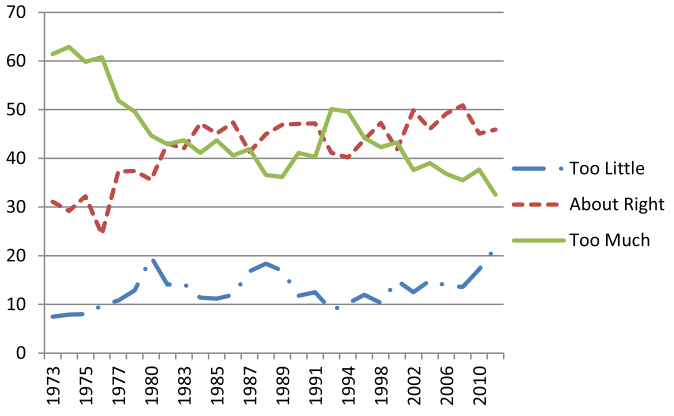

Fig. 4. Attitudes toward spending on space, 1973–2012.

3. Comparisons

To truly understand the different dynamics of these different measures and how newer measures such as Twitter and Google Trends might aid the space community in measuring public opinion, it is helpful to compare their relative performance. Figs 3 and 4 display data regarding the two space exploration related questions on the General Social Survey (GSS), the extent to which people are interested in space and whether people believe the US government is spending too much or too little on space exploration, respectively. Importantly, the first question on interest in space is a relatively new question, with data only for 2008, 2010, and 2012.

The data in Fig. 4 is particularly helpful in demonstrating overall patterns in public perceptions of spending on space from 1973 to 2012. However, while the survey used to be given on a yearly basis, since 1994, the survey has been given biannually. As such, up to date information of the kind that might be particularly influential to policymakers and agenda setters is absent.[4]

On the other hand, newer measures of opinion can show up-to-date and relatively up-to-the-minute information on what the public is thinking regarding space exploration recently. Figs 5 and 6 display the Google Trend index for "space exploration" and the number of tweets with "space exploration" in them, respectively. Two things should be noted. First, because social media and Google are relatively new, data for Google only goes back to 2004 while Twitter is only available since 2006. Two, with respect to Twitter, the data is a bit harder to gather. Because Twitter itself does not aggregate its tweets into a comprehensive search function, people searching for this kind of data must use outside analytics services. For these purposes, I utilized the search engine Topsy which only allows a limited Twitter search for free.

[4] While budgets are constructed on an annual basis, they are done so not once during the year, but continuously. For example, public interest in space exploration might be low at the beginning of the year only to rise throughout it in response to some sort of trigger or event. If this is the case, measuring public opinion only once a year may not give an adequate picture of true public sentiment towards exploration policy. Even if change throughout a year may be taken into account in subsequent budgets, many long term projects of the kind inherent in space policy may suffer in not being approved in the first year or in receiving lower budgets than they might have otherwise.

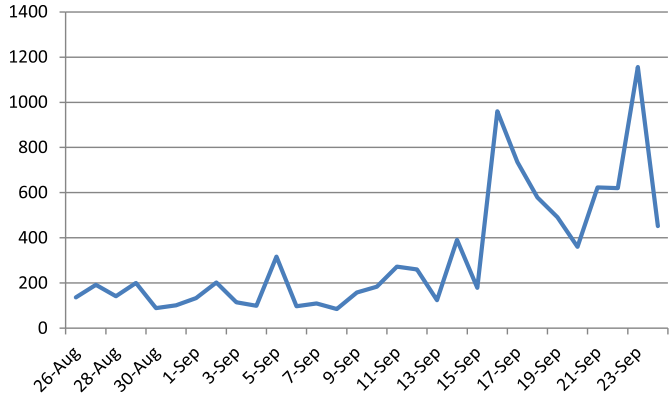

Fig. 6. Number of tweets containing the phrase "space exploration," August 26–September 25, 2014.

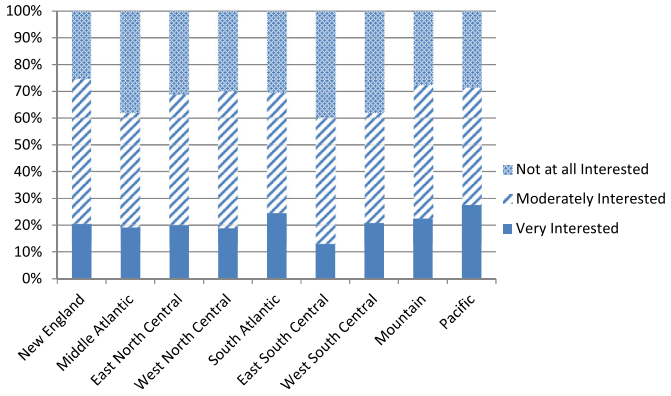

Fig. 8. Interest in space exploration by region, 2012.

Admittedly, searching Twitter for the rather general term "space exploration" could leave out a host of other space-related activities that could be talking about via social media. This would include things such as "ISS" or "International Space Station" or even references to the recent comet lander, "Philae." Given the limited scope of this paper is to demonstrate the applicability of this methodology, I am restricting the demonstration to the one generic term although using multiple search terms would also be appropriate.

Looking at the data in Fig. 5 in particular, it is apparent that since January of 2004, the overall number of searches on Google has declined from the index high of 100 to an average of 19 from January 2006 through August 2014. This trend is confirmed when looking at the number of tweets containing the term "space exploration" during the period of August 26–September 24, 2014. The highest number of tweets in any given day was only 1156. This is especially low considering that the average number of tweets on Twitter during any given day is more than 500 million [26].

To be fair, most people don't think of the term "space exploration" when they are thinking about space policy in the US. A fairer test of the usefulness of Twitter might be to search for a term that is more familiarly associated with space in America and that is NASA. Fig. 7 displays the total number of tweets containing NASA per day between August 26 and September 24, 2014. On average during that time period, there were 38,048 tweets, far higher than the average of 318 in Fig. 6. Even though not every tweet containing NASA was directly related to human spaceflight or space exploration, since that is the agency most directly connected with space exploration, it is arguably a better gauge of public opinion.

As a final demonstration of the usefulness of social media measures, consider the search tools that Google Trends provides. Not only can researchers specify time frames in which to garner results, but you can also specify country, region, and even state to compare search frequency among different populations over time. This feature could be of particular importance for local politicians and policymakers whose local constituencies will be of greater import than national ones. The only traditional data available to look at interest in space is through the GSS which leads to a number of problems if you want to know about interest by state. First, the latest GSS data available is 2012 meaning that should policymakers require up-to-date information, none is available. Secondly, GSS only codes their respondents by region in the US (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific). While the regions are fairly broken down, they still do not offer more specific locational data.

Fig. 8 displays the results from the 2012 GSS question asking if the respondent was interested in space exploration. Fig. 9, on the other hand, displays the Google Trends data for searches containing "space exploration" throughout 2012 (the darker the color, the higher the search index). At the top of the ranking is North Carolina and Texas followed by Florida (99), Pennsylvania (89), Ohio (88), Washington (83), and Virginia (83). While Fig. 9 provides some more in depth data, it also has some surprising findings as well. Of the top states, North Carolina, Pennsylvania, and Washington do not have major NASA installations located within their borders. In other words, in states with NASA installations, we might expect searches for space exploration to be rather high; more people live there who would be interested. But the fact that people in these

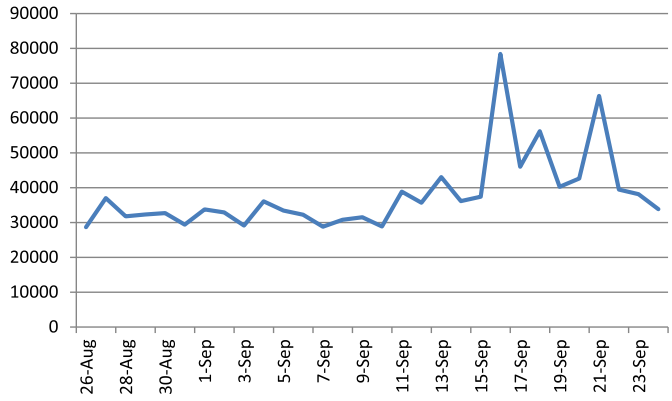

Fig. 7. Number of tweets containing "NASA," August 26–September 25, 2014.

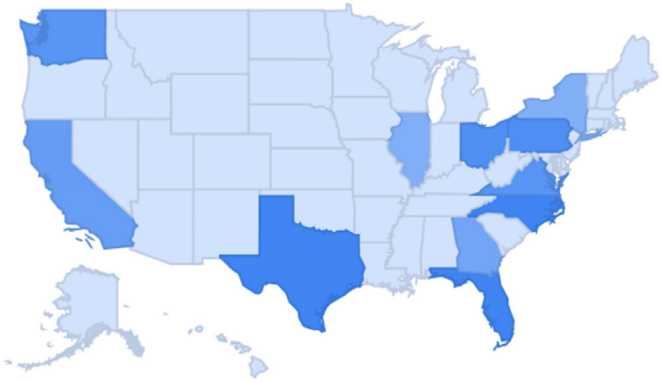

Fig. 9. Google Trends by state for "space exploration," 2012.

other states are also interested could be of importance to policymakers and officials alike.

4. Conclusions

What is useful looking at these comparisons is interest in space exploration in the macro and micro. Google Trends, with its data set going back to 2004, provides a longer term lens through which to view public interest in space exploration while Twitter gives analysts a larger window into the minute patterns of support. Taken together, social media and Google allow space policy analysts to look at more than simple support or interest in funding and provides a flexible tool through which to measure public interest from space exploration in general to missions to Mars or the outer planets more specifically. Moreover, we do not have to wait for Gallup or other polling organizations to put a question into the field; Google and Twitter allow for immediate searches and returns of data that make examining public opinion more flexible, immediate, and helpful.

If we accept the premise that public opinion and public support are important determinates of the outcome of space policy in America, then we have to know what public salience and support looks like. Policy entrepreneurs in both the space community and political community could find this type of data valuable as they endeavor to lobby Congress and the executive branch to support further activities or spending for NASA. The scientific community could also find this a valuable tool in helping to demonstrate the interest that the public may have in their work. Demonstrating an increased level of public support could help scientists to draw greater public and private support for the work that they undertake. While there are certainly those in Washington, D.C. and NASA centers across the country that regularly look to gauge public support for their various activities, having information of the such derived from Google Trends or Twitter provides data that is accessible and understandable for even the least tech-savvy among politicians.

As demonstrated above, the tools analysts have traditionally used to do such a thing have significant problems in the frequency of questions and the lack of questions being asked in the first place. Newer measures of public salience such as Google Trends and Twitter help to fill in those gaps and allow policy analysts a quicker and perhaps better lens through which to view public interest. While the tools presented here have drawbacks as well, their use can better inform policymakers and politicians about what the public is thinking concerning space exploration.

References

[1] Launius RD. Evolving public perceptions of spaceflight in American culture. Acta Astronaut 2003:823–31.
[2] Launius RD. Public opinion polls and perceptions of US human spaceflight. Space Policy August 2003;19:163–75.
[3] Steinberg A. Space policy responsiveness: the relationship between public opinion and NASA funding. Space Policy 2011;27:240–6.
[4] Whitman Cobb WN. Who's supporting space activities? An 'issue public' for space policy. Space Policy 2011;27:234–9.
[5] Nadeau F. Explaining public support for space exploration funding in America: a multivariate analysis. Acta Astronaut 2013;86:158–66.
[6] Burstein P. The impact of public opinion on public policy: a review and an agenda. Polit Res Q 2003;56:29–40.
[7] Bartels LM. Constituency opinion and congressional policy making: the Reagan defense buildup. APSR 1991;85:457–74.
[8] Hartley T, Russett B. Public opinion and the common defense: who governs military spending in the United States. APSR 1992;86:905–15.
[9] Burstein P, Freudenberg W. Changing public policy: the impact of public opinion, antiwar demonstrations, and war costs on Senate voting on Vietnam war motions. Amer J Sociol 1978;84:92–122.
[10] Page BI, Shapiro RY. Effects of public opinion on policy. APSR 1983;77:175–90.
[11] Monroe AD. Public opinion and public policy 1980–1993. Public Opin Q 1998;62:6–28.
[12] Stimson JA, Mackuen MB, Erikson RS. Dynamic representation. APSR 1995;89:543–65.
[13] Almond G. Public opinion and the development of space technology. Public Opin Q 1960;24:553–72.
[14] Handberg R. The fluidity of presidential policy choice: the space station, the Russian card, and US foreign policy. Tech Soc 1998;20:421–39.
[15] Cukier KN, Mayer-Schoenberger V. The rise of big data. Foreign Aff 2013;92:28–40.
[16] Aruoba SB, Diebold FX. Real-time macroeconomic monitoring: real activity, inflation, and interactions. Amer Econ Rev 2010;100:20–4.
[17] McLaren N. Using internet search data as economic indicators. Q Bull 2011;Q2:134–40.
[18] Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with web search. Proc Natl Acad Sci U S A 2010;107:17486–90.
[19] Vosen S, Schmidt T. Forecasting private consumption: survey-based indicators vs. Google Trends. J Forecast 2011;30:565–78.
[20] Gayo-Avello D. Don't turn social media into another 'Literary Digest' poll. Commun ACM 2011;54:121–8.
[21] Granka L. Using online search traffic to predict US presidential elections. PS 2013;46:271–9.
[22] Dugan M, Brenner J. The demographics of social media users-2012." Pew Research Center 2013, Retrieved from: http://pewinternet.org/Reports/2013/Social-media-users.aspx.
[23] Pew Research Center. Twitter news consumers: young, mobile, and educated. 2013.
[24] Couper MP. The future modes of data collection. Public Opin Q 2011;75:889–908.
[25] Vayena E, Mastroianni A, Kahn J. Ethical issues in health research with novel online sources. Am J Public Health 2012;102:2225–30.
[26] "New tweets per second record, and how!" https://blog.twitter.com/2013/new-tweets-per-second-record-and-how.