

Measuring Social Media Activities of Dutch Museums: Developing a Social Media Monitor

Thijs Waardenburg, Rogier Brussee, Erik Hekman

HU University of Applied Sciences Utrecht

Postbus 8611 – 3503 RP Utrecht, the Netherlands

thijs.waardenburg@hu.nl, roger.brussee@hu.nl, erik.hekman@hu.nl

ABSTRACT

In the research project “Museum Compass” we have developed a prototype of a social media monitor, which contains data of current and historic online activities on Facebook, Twitter, YouTube, Foursquare and Flickr of all registered Dutch museums. We discuss – mostly in a practical sense – our approach for developing the monitor and give a few examples result of its usage.

Author Keywords

Social media; analytics; metrics; cultural heritage; museums

ACM Classification Keywords

H.5.m. Information interfaces and presentation: Miscellaneous

INTRODUCTION

Social media have enabled easy and inexpensive interaction between millions of individuals and communities, and this has not gone unnoticed by the popular press, businesses and by scholars. Thus the online strategy of an organization has evolved from having a web site, to include a presence on social media. The question arises what all these social media activities actually bring. To answer that question, one has to start with measuring social media activities. However, current metrics solutions often consist of a confusing accumulation of statistics, across several systems, which reveal “little about online user behaviour, engagement and satisfaction” [3].

In this paper we discuss – mostly in a practical sense – the general approach and choices we made in developing a prototype of a social media monitor. The main goal of the museum monitor is to offer museum professionals and researchers better insight in the effects of their own social media usage and compare this with others in the cultural heritage sector. For researchers it gives the opportunity to consider communication within the sector as whole. The monitor, however, was developed in such a way that it

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author(s).

Published in: van Leeuwen, JP, Stappers, PJ, Lamers, MH, Thissen, MJMR (Eds.) *Creating the Difference: Proceedings of the Chi Sparks 2014 Conference*, April 3, 2014, The Hague, The Netherlands.

could also be used for other sectors. In most of the paper, the word ‘museums’ and cultural heritage sector can therefore be substituted with companies in an industry, or a group of organisations that have some cooperation but also compete for a roughly similar audience.

The prototype was developed in the context of “Museum Compass”¹ a project that helps Dutch museums deploying ‘crossmedial’ strategies more effectively. In the project we developed several other services and tools to support Dutch museum professionals such as the “Museum Guide”², an online questionnaire that helps museums determining their current and desired cultural and societal identity (its ‘archetype’).

The organization of this paper is as follows. The next section is focused on related work, after that the approach and choices that we made are described. The conclusions and discussion are described in the remainder of this paper. A paper in preparation (part II) will focus on results and data-analysis.

RELATED WORK

Many articles, books, papers, etc. are written about social media, and various definitions are proposed. For instance, Kaplan and Haenlein [4] define social media as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content”. Brussee and Hekman [2] propose a higher-level definition. According to them social media are “highly accessible media”: social media characterized by the accessibility of the whole media supply chain to the general public. Where most definitions agree that current social media are digital by nature, they pose that the internet and web technologies are just very well suited for providing this accessibility.

The public and digital nature of social media activities allow for automatically monitoring and collecting data. Commonly used terms for such data collection are ‘social monitoring’, ‘social media analytics’, and ‘social media metrics’. Many commercial off-the-shelf offerings can be found. They range from free ‘one-size-fits-all’ web ‘dashboards’, which collect data of one, or a limited number of platforms and present simple statistics, to extensive software that mines both structured and unstructured data and data from social and traditional media, and integrates quantitative data, qualitative data in the analysis often with

¹ <http://www.museumkompas.nl> (in Dutch, checked 20-01-2014)

² <http://www.museumkompas.nl/museumwijzer/> (checked 20-01-2014).

accompanying consultancy writing reports. For an overview with over 230 solutions see <http://wiki.kenburbary.com>³. These solutions range from relatively simple and free of charge, to extensive and costly. Examples of social media reports are the Dutch social media monitor for healthcare (“Social Media Monitor Zorg”)⁴ and the British Culture24 report [3], titled “Let’s Get Real: How to Evaluate Online Success?”⁵.

According to Kaplan and Haenlein [4], integration is important because “what is true for different types of Social Media also holds for the relationship between Social Media and traditional media: Integration is key!” Murdough [5], on the other hand, stresses “the important rule is to focus on just a few metrics for each objective so that program evaluation remains simple and one does not end up in “analysis paralysis””. Bruns and Liang [1], state “for more sophisticated research programmes, and for the tracking and study of larger-scale datasets over longer time periods, more advanced and usually custom-made tools and methods are required.” This is the approach we have taken.

APPROACH AND CHOICES

This research project is similar to, and partly inspired by the Culture24 project. An important difference is that the latter is based on social media data from *a selection* of museums, and *at a given moment*. The goal of the monitor that we are developing is to *continually track* social media activities of the ‘*whole*’ museum sector, and mine its history. As mentioned in the previous section, there are many commercial off-the-shelf solutions available to monitor social media. However, we chose to develop our own because we believe that it offers better opportunities to experiment, customize, and learn. It allows us to directly access the detailed data required to get a better understanding of the subject by tracking individual posts, post-likes, retweets, etc.

Requirements

We identified three basic requirements for the monitor, which can also be considered as development phases:

Measuring social media activities of museums on a daily basis. To answer “How much effort do museums put in social media?”

Measuring the activities of the public related to a museum and its social media activity on a daily basis to answer “To what extent does ‘the public’ respond on the social media activities?”

Serving as a social media benchmark, so that museums can evaluate their activities and responses as well as compare

themselves with their peers to answer “How do museums relate to each other, with respect to their social media activities?”

On the basis of popularity (or ‘maturity’) and type we chose the following five social media platforms: Facebook (‘social networking’), Twitter (‘micro blogging’), Flickr (‘photo sharing’), YouTube (‘video sharing’), and Foursquare (‘location sharing’).

Regarding the handling of data of social media platforms, we identified the following requirements for our monitor:

Data needs to be collected and stored in a database (‘back-end’).

Data needs to be interpreted and combined where possible.

Data needs to be presented (‘front-end’, in the form of a ‘dashboard’) for easy use by museum professionals and decision makers.

DATA COLLECTION

To collect data from any social media account, one needs to have their account identifiers (ID’s) on the platforms they use. The official Netherlands Museum Register⁶ does register the URL’s of the websites maintained by museums, but does not register their account-ID’s on social media platforms. We therefore needed to collect and verify this information ourselves. Collecting social media account-ID’s is in essence a simple task, but it proved to be very time-consuming. Since it is difficult to automate the verification of authenticity of accounts, this was done manually. The protocol that we used to collect account-ID’s was:

```
if account is linked on official museum website with platform logo:
```

```
accept account
```

```
else if search on social media platform for at most three variants of name of museum name returns account:
```

```
if account has official museum website as associated website:
```

```
accept account
```

```
else if account comes up when googling museum name + platform name AND looks real by value judgement:
```

```
accept account
```

```
else: reject account
```

³ <http://wiki.kenburbary.com/social-media-monitoring-wiki> (checked 20-01-2014).

⁴ <http://www.socialmediamonitorzorg.nl/> (in Dutch, checked 20-01-2014).

⁵ <http://weareculture24.org.uk/projects/action-research/how-to-evaluate-success-online/> (checked 20-01-2014).

⁶ The Netherlands Museum Register, <http://www.museumregisternederland.nl/> (in Dutch, checked 20-01-2014).

Rejected accounts were mostly pages automatically generated from Wikipedia and tourist information and easy to recognise. The rest almost all belonged to well-meaning individuals that write about large, well-known museums.

To collect the data we used the application programmers interfaces (API's) offered by the social media platforms. The earlier mentioned social media platforms all offer an API that enables basically anyone to develop an application that exchanges data with the platform. Using the scripting language PHP we extracted most data fields from the JavaScript Object Notation (JSON) returned by the platforms and stored the data fields in a MySQL database. We used PHP because it is widely used, supported by all platform API's, and because it made it easier to find freelance programmers. Wherever the platform allowed, we carefully kept track of individual actions of the organisations (posts, tweets, etc.) and the public (comments, likes, retweets) storing the ID's and creation timestamps provided by the platform and our own retrieval timestamps.

Each API (i.e., platform) has its specific set of data-elements. Which data-elements can be accessed depends on authentication levels. We distinguish three levels:

No authentication: 'Access token' and account-approval are not needed. Only basic account data can be received (e.g. account name, account-ID, profile image, etc.).

One-sided authentication. 'Access token' is needed from the platform provider, but account-approval is not needed. Publicly available / visible account-data can be received (e.g. messages, number of fans, etc.).

Two-sided authentication. Application needs to be registered at the platform; 'access token' and account-approval are needed. Extensive account-data can be received (e.g. friend list, private messages, etc.).

We generally focussed on the second level, as it allowed us to track the whole Dutch museum sector without having to make agreements with every individual museum. Fortunately, the data-elements that can be retrieved on the second level are fairly comprehensive. A Facebook 'post', for instance, is retrieved as a JSON document with 26 data-elements like 'id', 'from', 'to', 'message', 'likes', etc. next to the text of the post.

For practical reasons, we chose to make a selection of data-elements available from the platform, as it kept the size of the SQL schemas and queries in check. In addition some potentially available information required lots of additional queries to the platform. This is a problem as the number of requests per hour that can be made on a platform is fairly limited. For example, the Twitter (REST) API has a rate limit of 15 or 180 data requests per 15 minutes, depending on the type of data request. Flickr has a limit of 3,600 requests per hour. Facebook does not offer a clear insight in request limits at all but seems to be limited around 3,600 requests per hour per token, per IP-address. We found that data request limits are not always very 'strict', and can be apparently lowered for no obvious reasons. We dealt with this problem by daily scheduling the allowed number of

data-requests by using 'cronjobs' until we had queried the activity of all museums. It would have been easier, but much costlier to simply buy the data⁷. This has the advantage that one actually receives *all* historical data: for instance, it turns out that the free API of Twitter 'only' provides 3,200 historical tweets per account. However only three museums were reported as having more than 3,200 tweets when we started and for financial reasons, buying data was not an option.

THE FRONT-END INTERFACE

After approximately one month, we collected enough data to make a number of basic data views for the front-end of the monitor, like tag-clouds, tables and graphs using Google Chart Tools⁸ and other open-source tools (e.g. D3.js).

The front-end of the monitor is primarily intended for museum professionals. Data is presented from the point-of-view of a single museum, and museums can compare themselves with other Dutch Museums. For example, it allows them to view the change in the number of Facebook followers during a specific period and compare that with several other museums (e.g. of the same type or in the same city). This data is visualized in a graph, as in Figure 1.

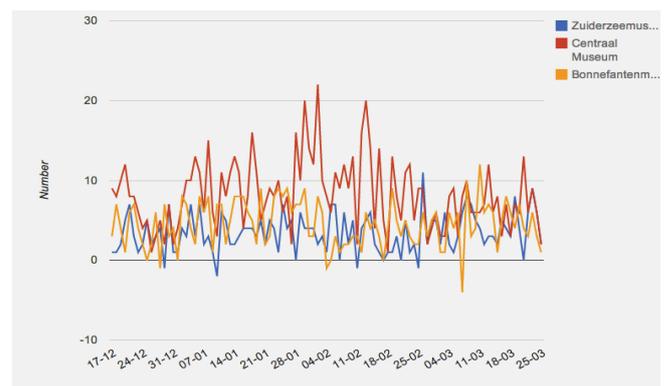


Figure 1: Screenshot of part of the interface for museum professionals

EXAMPLE RESULTS

A detailed data analysis will appear elsewhere ('Part II'). Here we give two examples of the kind of results that can be obtained by querying the database directly. This interface is only intended for researchers.

Sometimes museums react on other museums and thus show up as a 'public' reaction. Selecting these gives us a small social network in the sense of social network analysis (Figure 2), with avg. degree weighed by number of reactions = 4.85, and modularity = 0.605 representing the interactions of museums. It shows that museums are largely clustered by city (e.g. The Hague or Utrecht), and

⁷ For instance: Twitter offers a service called 'Firehose' (accessible through 3rd parties like Gnip or DataSift).

⁸ <https://developers.google.com/chart/> (checked 20-01-2014).

secondarily by subsector (e.g. art museums or botanical gardens).



Figure 2: A fragment of the social network based on Facebook likes amongst Dutch museums.

Another example result is distribution over clock hours of Facebook posts of museums, and likewise, the distribution of the responses of the public. These are easily determined, since every Facebook post each post, comment and like etc. is tracked individually and is given a creation date stamp by the platform. This histogram (Figure 3) can help museums plan their Facebook posts for optimal impact.

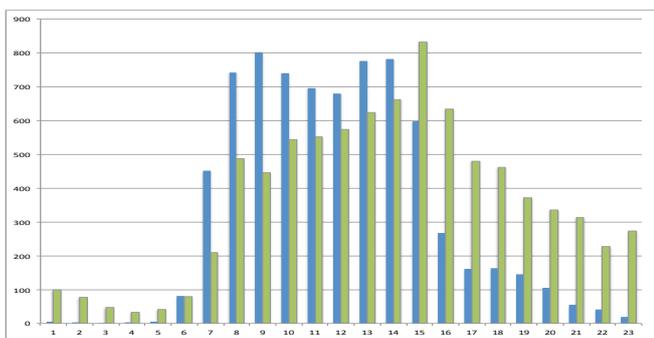


Figure 3: Histogram with the number of Facebook posts of Museums (blue) and ‘the public’ (green) over the 24 hours of the day. Museums post mostly during office hours peaking at 9h, the public peaks at 15h continuing during the evening and night.

DISCUSSION

We chose to develop our own tool. However, we do not exclude that, in the future, other (commercial) solutions may also fulfil our requirements. The development of this tool has proved a useful way to get a better understanding of the subject.

The first version has a number of basic and separate data representations. We found that Facebook and Twitter provide by far the most interesting data and therefore intend to focus more on these platforms in the near future. This will also reduce the maintenance burden.

We chose to store only a *selection* of data, instead of collecting all available data. Long-term this could be a decision that we may come to regret as we may come to the conclusion that we need specific data-elements that we left out for analysis. Whereas rate limits remain a problem, a solution for the excessive schema problem is to collect and store all data in an alternative way, e.g., ‘NoSQL’ or ‘NewSQL’ database.

The protocol for verifying account-IDs should be refined however, as it is too subjective in its current form.

Ethical questions arise when we collect social media data. We only collected public data that anyone can collect and the organisations that we tracked are public and visible organisations. However we will take care to anonymize data of individuals when publishing results.

CONCLUSION

We developed a prototype social media monitor to collect social media activities on Facebook, Twitter, Flickr, YouTube and Foursquare of all Dutch registered museums. The monitor should be useful in other contexts as well. Practical considerations, in particular the use of platform API’s and their limitations are discussed. We give an example of the results that can be obtained. Results of the data collection and analysis will be presented elsewhere.

ACKNOWLEDGMENTS

We thank the Foundation Innovation Alliance (SIA – Stichting Innovatie Alliantie) for funding RAAK-Museumkompas and all partnering museums.

REFERENCES

1. Bruns, A., Liang, Y.E.: Tools and Methods for Capturing Twitter Data during Natural Disasters. *First Monday* 17, 4 (2012).
2. Brussee, R., Hekman, E.: Social Media are Highly Accessible Media. *Proc. of IADIS international conference on WWW/Internet 2009*.
3. Finnis, J., Chan, S., Clements, R.: Let’s Get Real: How to Evaluate Online Success (2011). *Culture24*. <http://weareculture24.org.uk/projects/action-research/how-to-evaluate-success-online/>.
4. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53, 1 (2009), 59-68.
5. Murdough, C.: Social Media Measurement: It’s Not Impossible. *J. of Interactive Advertising* 10, 1 (2009), 94-99.