

Social Media Metrics for the Cultural Heritage Sector

Developing a Prototype

Thijs Waardenburg, Erik Hekman

Crossmedia Business Research Group, University of Applied Sciences Utrecht, the Netherlands
[thijs.waardenburg@hu.nl, erik.hekman@hu.nl]

Abstract: The online presence of organizations is long gone from being just a web page. Social media have enabled easy and inexpensive interaction between millions of individuals and communities. This has not gone unnoticed by cultural heritage institutes. The question is what all these social media activities bring. Even if an institute knows what it tries to achieve online, the metrics often consist of confusing accumulation of statistics, across several systems and reveal little about online user behaviour, engagement and satisfaction. In the research project Museum Compass a prototype of a social media monitor is developed, which will contain data of current and historic online activities on Facebook, Twitter, YouTube, Foursquare and Flickr of all registered Dutch museums. The first version of this monitor has been developed, and we believe that this is a good moment to discuss – mostly in a practical sense – our general approach and preliminary results.

Keywords: social media, analytics, metrics, cultural heritage, museums

1 Introduction

Social media have enabled easy, inexpensive interaction between millions of individuals and communities. This has not gone unnoticed by archives, museums, libraries and cultural heritage institutes and is visible in a broad array of initiatives and experiments with social media, in and outside these institutes (Hekman and Van Vliet, 2011). The question is what all these social media activities bring. To answer that question, one has to start with measuring these activities. However, current metrics solutions often consist of confusing accumulation of statistics, across several systems, that reveal “little about online user behaviour, engagement and satisfaction” (Finnis et al, 2011).

The Crossmedia Business research group of the University of Applied Sciences Utrecht runs a project called “Museum Compass”. Within this project, several services and tools are being developed to support Dutch museum professionals to use crossmedia solutions more effectively and efficiently. One of the tools is a prototype of a monitor for social media activities of Dutch registered museums¹. The main goal of this monitor is to offer museum professionals better insight in the effects of social

¹ The Netherlands Museum Register, <http://www.museumregister nederland.nl/>, checked 11-04-2012.

media usage within the museum sector. It is being developed in the context of the Museum Compass project, and is therefore focused on museums, or more generally the ‘cultural heritage sector’. However, it is developed in such a way that it can also be used for other sectors. In that sense, the word ‘museums’ in this paper can also be substituted with, for instance, ‘festivals’ or ‘retailers’.

The first version of the monitor has been developed, and we believe that this is a good moment to discuss – mostly in a practical sense – our general approach and the choices we made for the development of this prototype. In this paper, the actual data that is being collected with the monitor is not discussed. The organization of this paper is as follows. Section two is focused on related work, in section three the approach and choices that we made are described. The conclusions and discussion are described in the remainder of this paper.

2 Related Work

Many articles, books, papers, etc. are written about social media, and various definitions are proposed. For instance, Kaplan and Haenlein (2009) 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 (2010) propose a more philosophical definition. According to them social media are “highly accessible media”. They say that social media are not characterized by the role (digital) technology takes *an sich*, but by the accessibility of the media supply chain to the general public. However, most definitions agree that current social media are digital by nature.

The digital nature allows automatically collecting and viewing data of social media activities. Commonly used terms for this are ‘social monitoring’, ‘social media analytics’, and ‘social media metrics’. Many commercial of-the-shelf solutions can be found for this. For instance, an overview of more than 220 solutions can be found at the website wiki.kenburbary.com². These solutions range from relatively simple and free of charge, to extensive and costly. Hofer-Shall (2011) recognizes three categories:

- (1) “Social Dashboards: web-based tools that focus primarily on managing and analysing social media data”.
- (2) “Multichannel Analytics Providers: analytics infrastructures that mine social media data along with other structured and unstructured data sources”.
- (3) “Listening Service Partners: vendors with proprietary social analytics tools and professional consulting teams that write custom research reports based on social media data.”

Most of the solutions of the first category are ‘one-size-fits-all’ tools, and they collect data of one, or a limited number of platforms. The solutions of the second category, as Hofer-Shall describes, mine both structured and unstructured data, meaning that besides quantitative data, also qualitative data and data from traditional media is collected and analysed. According to Kaplan and Haenlein this is important, because “what is true for different types of Social Media also holds for the relationship

² <http://wiki.kenburbary.com/>, checked 11-04-2012.

between Social Media and traditional media: Integration is key!” Murdough (2009), on the other hand, stresses that “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.””. The third category implies the need for customized research reports. The fact that social media research reports are written for specific sectors supports this observation. Two examples of these kind of reports are the Dutch social media monitor for healthcare (“Social Media Monitor Zorg”)³ and the British Culture24 report, titled “Let’s Get Real: How to Evaluate Online Success?”⁴.

3 Approach and Choices

The Museum Compass research project is similar to, and partly inspired by the Culture24 project. An important difference is that it 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 a ‘*whole*’ 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 offers better opportunities to experiment, customize, and learn, in order to get a better understanding of the subject. This is also recognized by, for instance, Bruns and Liang (2012), as they explain that “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.”

As a research group we do not have the intention to develop production software. The monitor that we develop will therefore be a prototype that, in the end, may serve as a basis to develop production software, which is more reliable, maintainable, scalable, etc..

3.1 Requirements

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

- (1) Measuring social media activities of museums. What are museums sending?
- (2) Measuring impact of social media activities of museums. How does ‘the public’ respond on these activities?
- (3) Serving as a social media benchmark, so that museums can compare themselves to one another on this area. How do museums relate to one another, with regard to social media activities?

On the basis of popularity (‘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’).

³ <http://www.socialmediamonitorzorg.nl/>, only available in Dutch, checked 18-04-2012.

⁴ <http://weareculture24.org.uk/projects/action-research/how-to-evaluate-success-online/>, checked at 18-04-2012.

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

- (1) Data needs to be collected and stored in a database ('back-end').
- (2) Data needs to be interpreted and combined where possible.
- (3) Data needs to be presented ('front-end', in the form of a 'dashboard').

We started with collecting and storing publicly available data of social media activities of museums.

3.2 Data Collection

The above-mentioned platforms all offer an application programmers interface (API) that enables basically anyone to develop an application that exchanges data with the platform. Each API (i.e. platform) has its specific set of data-elements. Which data-elements can be accessed depends on authentication levels. We recognize three general levels of data access:

- (1) 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.).
- (2) One-sided authentication. 'Access token' is needed, but account-approval is not needed. Publicly available / visible account-data can be received (e.g. messages, number of fans, etc.).
- (3) 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.).

The number of data-elements can be fairly comprehensive. A Facebook 'post', for instance, consists of 26 data-elements like 'id', 'from', 'to', 'message', 'likes', etc. For practical reasons, we chose to collect a selection of data-elements. The data-transfer rate varies per platform. For example, the Twitter API⁵ has a limit of 350 data requests an hour (called 'rate limit'). Flickr has a limit of 500 requests. However, we established that data request limits are not always very clear, and for no obvious reason can be turned down. Facebook does not offer a clear insight in request limits at all. We anticipated on this by carefully planning our 'cronjobs' for data-requests. An easier, but much costlier solution for this is to buy the data⁶, which has the advantage that one actually receives *all* historical data (for instance, it turns out that the free API of Twitter 'only' provides 3200 historical tweets per account⁷). Due to the financial consequence, buying data was not an option for us at this moment.

To actually collect and structure the data we used the scripting language PHP in combination with the database software of MySQL. PHP is widely used and is therefore supported by most (if not all) API's. Another advantage of PHP is that it is a language that many freelance programmers are familiar with and that allows hiring programmers more easily when needed.

To collect data from any social network account, one needs to have account identifiers (IDs). The Netherlands Museum Register does register official websites of

⁵ Twitter has several types of API's. In this case the REST API is meant.

⁶ For instance: Twitter offers a service called 'Firehose' (accessible through 3rd parties).

⁷ 3 of the 146 museum Twitter accounts had more than 3200 tweets.

museums, but does not register account-IDs of social media networks. We therefore needed to collect and verify this information ourselves. Collecting social media account-IDs is in essence a simple task, but it proved to be time-consuming. Because it is difficult to automate the verification of authenticity of accounts, this was basically done manually. The protocol that we used to collect account-IDs was: ‘*search for the museum in question on the social media websites*’. If account(s) are found: ‘*look for reference to official museum website*’ and/or ‘*make a selection*’.

After a set of account-IDs was collected, we started collecting as much historical data as possible. Because collecting historical data involved large amounts of data, we had to be careful with the request-limits, mentioned earlier in this section. We therefore decided to execute these scripts by hand. When the historical data was collected, we created ‘cronjobs’ to periodically run the scripts to collect new data. After approximately one month, we collected enough data to make a number of basic data views, like tag clouds, tables and graphs using Google Chart Tools⁸.

4 Lessons Learned and Discussion

One of the main questions that we try to answer within the Museum Compass research project is what social media activities bring for the cultural heritage sector. The monitor that is being developed is meant to *help* answer that question. It is not *the* answer to the question. Besides this monitor, other tools are being developed, specifically designed for the (Dutch) cultural heritage sector, and these tools will be linked to the monitor. For instance, we develop the so-called Museum Match, an online questionnaire that helps museums making their current and desired cultural and societal identity, its ‘archetype’, explicit.

The first version collects data from Twitter, Facebook, and Flickr and has a number of basic and separate data representations. The data is not yet being interpreted or combined. In that sense, the monitor cannot demonstrate its actual potential at this moment. However, as mentioned in the introduction, this is not discussed in this paper. What *is* discussed is our approach and choices we made during the development this prototype.

To collect social media accounts, one can consider these two options: (1) create an inventory of social media platforms that are currently used by museums, or (2) select social networks with a large user base and look which museums use them. In retrospect, we believe that former choice is better, because it gives a complete overview of the social media platforms that are used. We choose the latter, because this is less time-consuming. The question is if this is a defensible argument. Besides that, the protocol for verifying account-IDs should be refined, as it is too subjective.

The final point of discussion here is our choice to collect a *selection* of data, instead of collecting all available data. On the long-term this could be an impractical, inefficient decision, as we may come to the conclusion that we need specific data-elements that we left out in the first place. A solution for this is to collect and store all

⁸ <https://developers.google.com/chart/>, checked 20-04-2011.

data in a non-relational, or 'NoSQL' database⁹. This makes the monitor also easier scalable.

We chose develop our own tool, for which we described our reasons. However, we do not exclude that, in the near future, other (commercial) solutions may also fulfil the requirements. Still, we believe that the development of this tool is useful way to get a better understanding of the subject.

The next version of the monitor will contain data of the three other social media platforms: Flickr, YouTube and Foursquare. At this moment we are researching the possibilities for visualizations and for combining data-elements.

References

1. Hekman, E., Van Vliet, H.: Bringing the Past to the Present. The Use of Tagging and Storytelling for the Enrichment of Digital Cultural Heritage. Utrecht (2011)
2. Finnis, J., Chan, S., Clements, R.: Let's Get Real: How to Evaluate Online Success. Brighton (2011)
3. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53 (1), 59 - 68 (2009)
4. Brussee, R., Hekman, E.: Social Media are Highly Accessible Media. Utrecht (2009)
5. Hofer-Shall, Z.: The 2011 Listening Platform Landscape. Forrester Research (2011)
6. Murdough, C.: Social Media Measurement: It's Not Impossible. *Journal of Interactive Advertising* (2009)
7. Bruns, A., Liang, Y.E.: Tools and Methods for Capturing Twitter Data during Natural Disasters. *First Monday* (2012)

⁹ For an explanation, see, for instance
<http://www.couchbase.com/sites/default/files/uploads/all/whitepapers/NoSQL-Whitepaper.pdf>, checked 24-04-12.