digital cultural heritage: FUTURE VISIONS

Edited by Kelly Greenop and Chris Landorf

Papers presented at the digital cultural heritage: FUTURE VISIONS Conference
19–20 April 2017 in Brisbane, Australia
http://digitalculturalheritageconference.com/

Garduño Freeman, C and Aguayo, MAC (2018)
Google Keyword Planner, Google Trends and Google Images: digital methods for investigating the social value of heritage.

The Conference Convenors received a total of 44 abstracts. Abstracts underwent a double-blind peer review by two members of the Conference Organising Committee. Authors of accepted abstracts (32) were invited to submit a full paper. All submitted full papers (18) were again double-blind peer reviewed by two reviewers. Papers were matched as closely as possible to referees in a related field and with similar interests to the authors. Sixteen full papers were accepted for presentation at the conference and a further 6 papers were invited to present based on submitted abstracts and work-in-progress. Revised papers underwent a final post-conference review before notification of acceptance for publication in these conference proceedings.

Please note that papers displayed as abstracts only in the proceedings are currently being developed for submission to a digital cultural heritage special edition of an academic journal.
Abstract

Many World Heritage monuments stand as symbols for their cities. In Sydney, the Opera House is an icon, in Paris the Eiffel Tower, in Agra, the Taj Mahal. This concurrence suggests that such iconic sites are deeply rooted within their communities. However, in Guadalajara, Mexico, this does not seem to be the case. Hospicio Cabañas is a World Heritage property located in the centre of Guadalajara. The Neoclassical architectural complex is a unique example of an architecture for public assistance and houses frescos by Jose Clemente Orozco, one of the great Mexican muralists. While this place is of international significance, our observation was that neither the architecture nor the art, both iconic in their own right, appears to be prominently associated with its city as an icon. Instead, it is the twin spires of the Catedral Basílica de la Asunción de María Santísima, the main cathedral located a kilometre away, that perform this role. Every day we search for information online; searching has become a cultural phenomenon, much of it mediated by Google. In this paper, we explore how the data collected as part of our culture of search can offer an emergent methodology for a broad assessment of the social value of heritage. Prompted by an anecdotal observation about the value of two heritage sites in Guadalajara, we investigate how each is publicly perceived, using Google’s free analytical tools, Google Keyword Planner, Google Trends and its image based service, Google Images. The paper adopts the theoretical stance proposed by Richard Rogers, which conceptualises Google as a socio-technical phenomenon driven by society and algorithms. In this way, it speculates and makes an initial investigation and comparison between the social connections with World Heritage and National Heritage, with a view to future development.

Keywords: Google; search culture; World Heritage; National Heritage; Guadalajara; cities; images; methodologies.
Introduction
The focus of this paper arose out joint observations about the social value of heritage sites in Guadalajara, Mexico and the proliferation of contemporary online communication technologies. We were familiar with Guadalajara’s significant buildings and their prominence in everyday culture and were interested in how Google’s search data and free analytics tools could validate our observations about two particular buildings.

Hospicio Cabañas is one of largest and oldest historic hospital complexes in the Americas. The property was the first site of modern heritage in Mexico; inscribed as World Heritage in 1997. Designed by Spanish architect Manuel Tolsá, the hospice was founded in early 19th century to provide shelter for the needy, including orphans, the elderly, disabled and chronic invalids. In addition, the architectural complex houses murals by José Clemente Orozco that were added in 1937. The murals are considered masterpieces of Mexican art and depict a fusion of Spanish and indigenous Mexican culture. Today, Hospicio Cabañas operates as a cultural institute and museum and while an important place in the City of Guadalajara, is infrequently depicted in its visual culture (Figure 1).

In contrast, Guadalajara’s main cathedral, La Catedral Basílica de la Asunción de María Santísima
keywords into Google's search engine which currently has the largest global market share (Armstrong 2016; Netmarketshare 2017b, 2017a). Accustomed to having information at our finger tips (Sparrow, Liu, and Wegner 2011), we Google in all manner of social situations; so much so, that ‘googling’ has become an everyday activity generating masses of data encapsulating our interests, concerns and needs. Google’s mission has become ‘to organize the world’s information and make it universally accessible and useful’ (Google 2016) which has enabled the evolution of a search culture, where ‘googling’ has become an embedded but inconspicuous practice (Hillis, Petit and Jarret 2013). We search on mobile devices in casual conversation, to answer questions, to find places, satisfy boredom or quench curiosity. Hillis, Petit and Jarret observe that, ‘to search has become such a natural and obvious condition of using the web, and the web such a natural and obvious feature of the internet, that the specific contingency of these everyday practices has become obscured’ (2013, 2). They emphasize that it is crucial to reflect on the way in which searching through Google, structures knowledge and information as part of a growing search culture curated through computational algorithms. Google is powered by a closely guarded algorithm known as ‘PageRank™’(Lee 2016, 4). The search service enables users to find webpages, images videos and other forms of information by sorting millions of pages and returning relevant results. Google (now owned by parent company Alphabet Inc) is also a corporate behemoth; by 2016, Google, had overtaken Apple to become the most valuable company worldwide (La Monica 2016). Since incorporation in 1998, Google has tailored its search engine service to multiple languages and cultural settings and to specific types of content (Google Scholar, Google Books, Google Image) and developed game changing services (Google Street View, Google Translate). However, keywords are essential to the way that Google’s search algorithm crawls and indexes webpages. These webpages are then ranked by the
algorithm dependent on the quantity and quality links pointing to the webpage. (Google 2017). While not neutral, Google can be loosely conceived of as an ‘echo of our online conversations’ (Rupiah 2016) as it is dependent on the interplay of all the elements of the system: the algorithm; the information on each webpage; the search queries and keywords entered by people. Google is a dynamic and complex socio-technical phenomenon structured on one hand by the searches performed by a ‘community’ of global users and on the other by a constantly evolving engineered algorithm.

Google collects indexing and search query data and offers several free analytical tools: Google Trends; Google Keyword Planner; and Google Display. These, along with more tailored services such as Google Analytics, help users to refine website keywords, advertising and search engine optimization to increase search rankings. Stephens-Davidowitz proposes that search data can reflect people’s true opinions more accurately than other social science methods (2017).

Search data is a valuable resource for understanding human perception because of two key factors: first, simply because of its size and reach (human beings produce approx. 2.5 million trillion bytes of data every day); and second because search queries are (usually) anonymous and therefore offer a sense of privacy to express oneself without judgment. Google search data has been used to predict the likelihood of diseases, or the relationship between mood and weather (Stephens-Davidowitz 2017). Here, we explore how Google’s data and tools can offer quantitative and qualitative ways to evidence social value around heritage places.

**Social value**

The concept of social value is well established in Australia and is part of legislation and best practice guides for natural and built forms of heritage (Australian Government 1999; ICOMOS Australia 2013). In contrast, social values for buildings are not considered in the Mexican context; heritage remains defined as monuments and zones designated for their archaeological, artistic or historic value (Mexican Government 1972). Nonetheless, there is increasing international interest in recognising social and community values evidenced in the inclusion of communication and communities within World Heritage Convention’s strategic objectives (UNESCO 2017). Mexico has been part of this shift, at least at an academic level (Johnston 2016), but also at an international level through their ratification of UNESCO’s *Convention for the Safeguarding of the Intangible Cultural Heritage* (2003). Broadly, social values acknowledge the associations communities have with significant places and their role in providing a sense of identity and belonging. But conceptions of well-defined and geographically co-located communities are being challenged by internet and communication technologies in an increasingly networked world (Garduño Freeman 2016; Waterton 2010; Lewi et al. 2010). Harnessing Google’s search data to assess the social value of places can contribute a great deal to an area of assessment that is complex and often resource intensive. Social values are dynamic, fluid and subjective (Johnston 1992; Johnston et al. 2003; Johnston 2012, 2016; Byrne, Brayshaw and Ireland 2003; Byrne 2013; Smith 2006) and are therefore distinct from more stable values such as aesthetic, historic or scientific kinds of significance. Using real-time data has the potential to offer methods to understand how social values change over time or in response to specific events. At the same time, we acknowledge that such an approach is not without its limitations: access to online services is not uniform and such a methodology relies on data currently controlled by a global corporation. Recognising these conditions, the paper’s approach is speculative and the discussion here recognises such limitations and acknowledges that reliability on such a method requires further in-depth research.

**Google Keyword Planner and Google Trends**

Our comparison of the social value of Hospicio Cabañas and La Catedral using Google’s search data uses three tools: Google Keyword Planner, Google
Trends and Google Images. First, we make a direct quantitative comparison of the volume of search data for keywords for each place, and a qualitative analysis of associated keywords using Google Keyword Planner. We then consider these results over time and in terms of their geographic origin using Google Trends. Second, we exploit keywords in relation to photographs using Google Images where what is represented in the image can be distinct from its text-based filename.

Google’s analytic tools are intended to assist in the development of advertising campaigns and search engine optimization. The goal is increasing website audiences and traffic. While there are many factors that affect the way Google’s algorithm ranks pages, one of the central variables is keywords. Both Google Keyword Planner and Google Trends analyse search volume in relation to specific keywords. Google Keyword planner focuses on the average monthly search volume over the past 12 months and provides a list of associated keywords, while Google Trends provides a longitudinal analysis of search volume from 2004. Google Trends also allows for comparisons between search terms. Because the analysis is dependent on keywords, language must also be accounted for.

Our Google Keyword Planner search was centred on place names in both English and Spanish: Hospicio Cabañas (same in both languages), Catedral (Spanish) and Cathedral (English). To differentiate from other similar named places, we added Guadalajara to our search terms. The results confirmed our initial observation that La Catedral was more prominent by tenfold than Hospicio Cabañas, but only when the search is carried out in Spanish (Table 1), which suggests that La Catedral is far more significant within the Spanish speaking (and presumably Mexican) communities. In English, the results were the same between the two places. Other searches were performed using other keywords including ‘heritage’, ‘heritage guadalajara’ and ‘heritage mexico’ in both languages to contextualise these results (Table 2).

The search volume for ‘heritage’ was one thousand times greater than the search volumes of Hospicio Cabañas, while one tenth the volume for ‘heritage guadalajara’ and ‘heritage mexico’. Such differences could indicate that Guadalajara is not perceived as an important location for heritage in Mexico or indeed internationally. Further, the volume of searches for ‘heritage’ was tenfold that of its Spanish translation ‘patrimonio’. Again, this could be indicative of the dominance of Western English-speaking nations in the conceptualisation and use of the term. This could also offer insights into Mexico’s slower adoption of such all-encompassing terms, reflected within their heritage legislation that is still founded on terms such as ‘monuments’ and ‘zones’ (Mexican Government 1972).

Table 1: Results for average monthly search volume for place keywords over the period from May 2016 to May 2017; retrieved from Google Keyword Planner.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Search Volume</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospicio Cabañas Guadalajara (English)</td>
<td>100-1000</td>
<td>1</td>
</tr>
<tr>
<td>Hospicio Cabañas Guadalajara (Spanish)</td>
<td>100-1000</td>
<td>1</td>
</tr>
<tr>
<td>Cathedral Guadalajara (English)</td>
<td>100-1000</td>
<td>1</td>
</tr>
<tr>
<td>Catedral Guadalajara (Spanish)</td>
<td>1000-10,000</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2: Results for average monthly search volume for broad keywords over the period from May 2016 to May 2017; retrieved from Google Keyword Planner.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Search Volume</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heritage Guadalajara (English)</td>
<td>10-100</td>
<td>0.1</td>
</tr>
<tr>
<td>Patrimonio Guadalajara (Spanish)</td>
<td>10-100</td>
<td>0.1</td>
</tr>
<tr>
<td>Heritage Mexico (English)</td>
<td>10-100</td>
<td>0.1</td>
</tr>
<tr>
<td>Patrimonio Mexico (Spanish)</td>
<td>10-100</td>
<td>0.1</td>
</tr>
<tr>
<td>Heritage (English)</td>
<td>100,000-1,000,000</td>
<td>1000</td>
</tr>
<tr>
<td>Patrimonio (Spanish)</td>
<td>10,000-100,000</td>
<td>100</td>
</tr>
</tbody>
</table>

Google Keyword Planner also returns a list of keywords associated with the primary keyword entered. These associated keywords are retrieved in order of relevance alongside monthly search volumes. These keywords can be analysed using word frequency analysis using Wordle.com, (a website that creates word clouds, where font size indicates higher
also included references to other World Heritage sites in Mexico such as the archaeological ruins of Monte Albán, Uxmal and Chichén Itzá (Figure 4). Interpreting these themes, frames Hospicio Cabañas as a significant site of heritage in English, but in Spanish its value is as a cultural institute. In contrast, the associated keywords in Spanish for La Catedral indicated its role in the historic centre of the city, with references to tourism and places of accommodation (Figure 5). In English, La Catedral was clearly associated with Mexico, with the state of Jalisco, and with attractions and places to visit (Figure 6).

The results from Google Keyword Planner are interesting: because while La Catedral has a higher search volume, the associated keywords indicate it is for its value as an urban icon within the historic centre of Guadalajara rather than as a religious space or as a place of heritage significance. In contrast, Hospicio Cabañas, while having a lower search volume, seems to be a significant heritage site for English language audiences. For Spanish language audiences, its value is as a cultural institute and museum. Insight into the longitudinal patterns of the search volumes and their predominant country of origin was gathered using Google Trends (Figure 7). The relationship observed between the prominent position of La Catedral in relation to that of Hospicio Cabañas, is longstanding.

Figure 3. Wordle diagram of associated keywords from Google Keyword Planner results for ‘hospicio cabañas guadalajara’ in English.

Figure 4. Wordle diagram of associated keywords from Google Keyword Planner results for ‘hospicio cabañas guadalajara’ in Spanish.

Figure 5. Wordle diagram of associated keywords from Google Keyword Planner results for ‘catedral guadalajara’ in Spanish.

Figure 6. Wordle diagram of associated keywords from Google Keyword Planner results for ‘catedral guadalajara’ in Spanish.

Figure 7. Google Trends graph of keywords in Spanish and English.
Google Trends also reveals that most of the searches for La Catedral originate from Mexico and the USA, reinforcing our observations that this building was more locally significant than Guadalajara’s World Heritage site, Hospicio Cabañas. This longstanding differentiation is significant because both La Catedral and Hospicio Cabañas could be urban icons; they are located less than one kilometre away from each other. The two buildings are part of an urban cruciform arrangement and form an axis from west to east through the historic urban centre of the city. Both have heritage significance, albeit within different contexts and of different levels of significance. The results appear to offer some preliminary confirmation of our observation that La Catedral is an urban icon for Guadalajara that far surpasses the perceived importance of a World Heritage site from a global perspective.

Google Images
Images are one of the most popular types of content being searched for on Google (Smith and Google 2010). Google Images, as the name indicates, retrieves images sourced from websites, social media platforms and archives. Like the standard text based web search, Google Images uses over 200 parameters to rank results within its algorithms (Lee 2016, 197). In 2001 when it launched, the service gave access to the first 250 million images indexed (Google 2014); by 2005 this had increased to 1 billion, and by 2010 it had reached 10 billion (Smith and Google 2010). Like Google Search, keywords is one of the key variables used to index and rank images. Title tags, size, textual richness and the content of the webpage also contribute to Google’s ranking (Rose 2016, 301). However, unlike text based searches where the data being indexed is the actual content, with images, the data being indexed is the file name and its alternative text rather than the visual content of the image itself.5

The second exploration using Google was through its image search service, Google Images. Searches using the same keywords from the first part of our investigation were carried out. When the keywords based on the names of the places were used (in English and Spanish), Google Images returned an accurate series of photographs of each site (Figure 8). This confirmed that the keywords used were sufficiently accurate. However, the value of the Google Images search is the ability to use broader search terms that are common to both places based on the intentional association created between the filename and the content of the image. For example, broader searches using the English and Spanish variations of ‘heritage guadalajara’, ‘heritage mexico’ and then even more broadly ‘guadalajara’ can reveal how Hospicio Cabañas and La Catedral rank against each other in terms of the number of times featured and their order of appearance on the page.

Taking the keyword ‘heritage’ as an example, Google Images’ results feature several images of Stonehenge and other ‘old’ buildings as well as logos containing this word (Figure 9). Notably, no images of modern,
intangible or digital heritage were returned. This reinforces the traditional heritage perspective as ‘old, grand, monumental and aesthetically pleasing’, and as evidencing part of the ‘authorised heritage discourse’ (Smith 2006, 11). While these results may simply confirm accepted notions within the field at present, over time such results could track changing perceptions of how heritage is understood.

The Google Image search using the keywords ‘heritage guadalajara’ returned images of both Hospicio Cabañas and La Catedral, as well as of other local monuments such as the former gate to the city known as Los Arcos (Figure 10). Yet, only Hospicio Cabañas appears in the search for ‘patrimonio mexico’; it appears as the twenty-fifth image in the Spanish search, which was dominated by images of several of Mexico’s World Heritage sites (Figure 11). The English search, ‘heritage mexico’ returned images of archaeological sites, such as Chichén Itzá, traditional costumes, local cuisine and icons of Mexico City. Such results suggest that the term ‘patrimonio’ in Spanish is more readily associated with World Heritage rather than with national or local examples, and that while both places feature as heritage in relation to Guadalajara, only Hospicio Cabañas is significant at a national or international level, regardless of their actual legal inscriptions.
Richard Rogers argues for a more integrated definition of digital methods for research, where methodologies must incorporate both qualitative and quantitative aspects in order to understand the circulation of information online in its fullest sense. Google’s search data and tools open up new ways to explore both dimensions. Google also allows for the analysis of visual data; at the same time visual and digitally mediated participation methods for understanding heritage is gaining interest in the field. Waterton and Watson’s edited collection *Culture, Heritage and Representation*, positions representations and visuality as significant components of heritage (2010). Giaccardi’s *Heritage and Social Media: Understanding Heritage in a Participatory Culture* recognizes participatory forms of communication as evidence of the connection between people and heritage expressed in their everyday experiences (2012).

When we performed one final search, using ‘guadalajara’, our observations about La Catedral’s iconic role in Guadalajara’s city-image were also confirmed (Figure 12). The results from the search, like those from the analysis of the associated keywords, indicated that La Catedral is an urban icon for the city; of the images returned, ten of the first eighteen were of La Catedral. In contrast, Hospicio Cabañas appears in eleventh place, just after Los Arcos. These results were dependent on the term ‘guadalajara’ being present in most of the file names of the images, providing some evidence of the strong association of these buildings with the city (Figure 13).

**Discussion on the potential use of Google as a methodology**

The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions by means of the Internet.

(Rogers 2013, 21)

The formal inscription of heritage is not neutral; rather it is now accepted that it is a political and subjective act, situated within a larger social and dynamic cultural processes. Monuments and heritage sites are ‘a potent resource for place branding. They accentuate the history of a place, and thus assert the place’s uniqueness’ (Lai and Ooi 2015, 276). World Heritage inscriptions have also been implicated in economic results associated with place branding such as tourism, city rankings and global status (Kirschenblatt-Gimblett 1998, 2006; Di Giovine 2009; Hall and Piggin 2003). Place branding is connected with place identity and sense of place; it is implicated in the narratives, representations and popular discourses that circulate and draw together citizens and visitors alike (Mayes 2008). In Guadalajara, the inscription of Hospicio Cabañas as a site of World Heritage has not enabled it to contribute greatly to the city’s branding at a global level. The local icon of La Catedral is an urban icon and plays this role.
associations and results are seen as negotiations between people on a mass scale and technological algorithms seeking to understand accurately what we are looking for. Rogers argues that digital methods should embrace the entangled nature of results alongside computational components.

Conclusion

In a world where information is embedded as a cultural phenomenon in our everyday experiences, understanding its implications is crucial. At present, Google is the primary search engine platform. It is simultaneously structuring knowledge through its algorithm as well as being structured by the inputs of society. Google is a real-time representation and a way to track changing perceptions over time. In analysing sites of heritage using Google’s free tools, we seek to develop methods that could be adopted by professionals and heritage managers. There are other services in Google’s suite which could also add further information: Google Display Planner gives demographic information; Google Maps is a site for digitally mediated experiences. While emergent, further research is needed to comprehend the full potential for heritage of Google’s data. Our methodology, explored here, suggests that Google search data and Google Images can evidence broad perceptions and associations, confirm well-grounded observations and therefore enable new ways to compare the social value of a range of heritage places.

Notes

1. For this analysis keyword means one or more words.
2. Google Keywords gives results in broad bands unless there is an advertising campaign running.
3. Note: Google collects data from searches, maps, YouTube and other services unless users block or opt out of this.
4. We acknowledge that the data only goes back as far as 2004.
5. In 2009 Google added a feature to its image search service to browse similar images (Murphy-Chutorian, Rosenberg, and Google 2009). This developed into a new service called Search by Image, rolled out in 2011, which harnesses image recognition technologies and allows users to perform a search by uploading or providing the url of a specific image rather than a keyword (Wright and Google 2011). Interestingly, Search by Image was (apparently) instigated to enable users to take a picture of a famous landmark and access information about it (Wright and Google 2011).

Acknowledgement

We would like to acknowledge Sonia González Velázquez, Universidad de Guadalajara, México who contributed as a research assistant to parts of this paper.
References


