

# THE NATION IN THE NETWORK: CO-LINK ANALYSIS OF ONLINE NETWORKS FOR THE BRITISH MUSEUM

Rebecca Kahn<sup>a</sup>,

<sup>a</sup>Department of Digital Humanities, King's College, London & Alexander von Humboldt Institut für Internet und Gesellschaft, Berlin  
Rebecca.kahn@hiig.de

**ABSTRACT:** This paper demonstrates how web research methods can be used to reveal the online networks which develop around museums, and the strength of the relationships between entities in those networks. These relationships can affect the way museums may be perceived by online visitors; they are also significant tools for museums to build identity and further their strategic aims and museological objectives online. Ultimately, this research aims to show how the web can be used to build networks and relationships that re-evaluate and reframe the narratives embedded in museum collections, and open opportunities for new voices to emerge as museum navigate the challenges of a global digital future.

## INTRODUCTION

For many museums, the benefit of being online means that anyone surfing the web is able enter and explore a collection with ease from anywhere in the world. Museums increasingly display clear digital consciousness through their descriptions of digital activity as distinct components of their core work. But how are museums, which see themselves as existing in at least two places (the physical institution and cyberspace) at one time, able to articulate an identity online? When the symbolic and stabilising markers of national identity, such as national flags, iconic architecture or geographical location are absent in the digital museum, institutions may struggle to articulate the national or regional identities they represent, or the audiences they see themselves as serving. This research explores how the digital version of the British Museum, when no longer physically located in London, makes use of other kinds of digital associations, in order to locate the institution within a certain national tradition.

In order to do this, I examined the Museum's online network using tools developed in the field of web studies. This allowed me to demarcate and locate the Museum it within a sphere of activity on the web which may be read as a proxy for national boundaries.

Once this virtual national web had been defined, I was able to examine the other institutions who are connected to the Museum, in order to establish what online networks the Museum is part of, how communications flow within these networks, and the significance of these relationships between the Museum and other actors.

## A DIGITAL APPROACH TO MUSEUM IDENTITY

By establishing and examining the digital Museum's relationships to the organisations around it, this research demonstrates how it fits into the socio-political establishment and its position in relation to the loci of cultural power and influence. Rather than locating these in London, or the United Kingdom, though, this examination looks to the see how the Museum interacts as an actor within the Web. The intellectual basis for this approach is rooted in Actor-Network Theory [3], and represents an attempt to better understand any impact that the Museum's behaviour might have within an established online network. By using this analytical approach it is possible to show that digital museums are sites of meaning and significance, rather than simply 'curated collection of artefacts for others to browse' [4].

On the whole, current trends in Museum Studies have tended to overlook the web as a subject of study itself, and have preferred to see it as a platform for the delivery of outreach-oriented content. By using this methodological approach the intention is to develop a way of looking at digitised museums and cultural heritage projects which is aligned with the perspectives of museology and cultural studies and applies them to an object which is both digital and pre-digital. Digitised museums and museum collections are in effect, new entities, with patterns of behaviour on the Web which have yet to be described and understood.

A digital approach enables the development of an emergent picture of the British Museum's online network and relationships, and prompts an investigation into whether it exists within a webspace which potentially could be described as 'British'. By reading the Museum's website as a publication which may be parsed for meaning, it becomes possible to build a study of how the Museum manifests Britishness online, in a space which is traditionally seen as multicultural, open and multi-vocal.

Contrary to the characterisation of the Web as a chaotic, and egalitarian space where the casual browser may hop between links and create their own pathway along the information superhighway, the reality is that the web is a much more directed, hierarchical and managed space than it may appear [2]. The algorithms used to power search and ranking of web pages are not neutral mechanisms, and the relationships between online entities such as the British Museum and other sites it links to are deliberately cultivated and filtered by both the institutions in question and the search engines through which online visitors access them. The working of these algorithms are at best opaque, and often completely invisible to the user, and run in the background of websites, subtly influencing how users find information, what information they are able to access, and how associations are made between institutions. These mechanisms also leave pathways and traces that web researchers are able to follow in order to create maps of certain sections of the web.

This research explores how the digital version of the British Museum, when no longer physically located in London, makes use of other kinds of digital associations, in order to locate the institution within a certain national tradition.

In order to do this, I examined the Museum's online network using tools developed in the field of web studies. This allowed me to demarcate and locate the Museum within a sphere of activity on the web which may be read as a proxy for national boundaries. Once this virtual national web had been defined, I was able to examine the other institutions who are connected to the Museum, in order to establish what online networks the Museum is part of, how communications flow within these networks, and the significance of these relationships between the Museum and other actors.

## **MAPPING THE MUSEUM'S NETWORK**

Mapping the Museum's online network required using a tool called the IssueCrawler, which was developed by the Digital Methods Initiative in Amsterdam [5]. This tool makes use of the hyperlinks, code and search algorithms which make up the web and has been used to map out minority webspaces [6], delineate the boundaries of the webspace of the UK [7] and to map portions of the web in Europe and South Korea before, during and after national elections [8]. It can also be used for explorations of web archives online, as it allows researchers to order, process and extrapolate significance from the vast volume of archived material on the web.

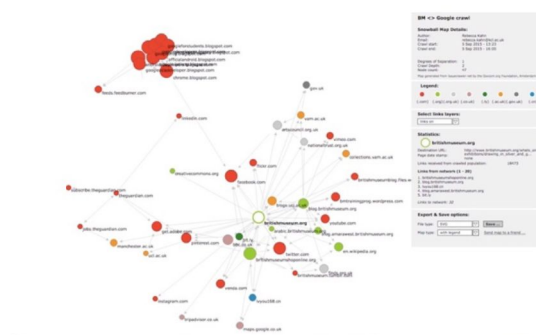
The IssueCrawler is usually used to discern patterns of communication and infer the standing of an issue by way of interpreting the search algorithm's authority [9], rather than relationships between the digital versions of organisations and institutions on the web. In this particular case, however, I use it to discern which institutions and organisations make up the Museum's network, and then visualise those links and relationships and draw conclusions about the Museum's associations and relationships.

This reality highlights the importance of cleaning the initial dataset in order to minimise the risk of including outlier links which might skew the result and of maintaining a critical examination of the resulting network map in order to be aware of any clusters of nodes and edges which may result from such existing links.

In order to do this, it was necessary to extract as many hyperlinks as possible from the entirety of the British Museum's site. To do this, I used a tool called Harvester, also one of the suite of analytical tools developed for web scholars by the DMI. Harvester extracts all the URLs from a webpage, and returns a list of outlinks, which can then be added into the list, which will be used as the basis of the Crawl. The British Museum website does not have a link- or blog-roll page, so that it was necessary to go through the site page by page, using the sitemap as a guide, harvesting all the URLs possible in order to come up with a comprehensive list. This list then needed to be cleaned, in order to strip out a great many duplicate URLs from final list. The reasons for this were pragmatic as well as technical - embedded in the foot of every page of the Museum's website, there are links to internal and external content, such as Accessibility information, the Museum's site map, their Terms of Use, policy for the use of Cookies, a list of FAQs, a link to the official Chinese and Arabic versions of the site, a link to the Department of Culture, Media and Sport's Portable Antiquities Scheme, the mobile version of the site and links to several social media accounts where the Museum has a presence, including Facebook, Tumblr, Twitter and Instagram. These links were removed from the final list, in order to give a true representation of the outlinks from the Museum. Had they remained in, they would have distorted the number of outlinks from the site. Once these had been stripped out, I was left with a list of 15931 unique URLs that formed the basis of the Crawl.

The next step was to set the parameters for the crawl itself. The IssueCrawler has four settings: the status of starting points, whether analysis is done by page or by site, the number of iterations and the depth of the crawl. By default these are set by the tool as follows: starting points are privileged, which means that they remain in the results which are rendered in the first iteration of the crawl, so that the basis for the network is the starting point plus any other entity with at

least two links from the starting points. Setting the analysis to be done by page rather than by site provides more specific results, since running analysis by site gives results for homepages only, even if several pages in a site contain in- and out-links to the starting point. The number of iterations for the search can be set to one, two or three. The DMI recommends using one iteration to uncover social networks, two to uncover an issue network and three to map an establishment network. By setting the depth for this particular crawl to 2, I instructed the crawler to include pages that have co-links to and from [www.britishmuseum.org](http://www.britishmuseum.org), but not pages that are at one more remove. These setting rendered the following visualisation:



*Figure 1: Visualisation of the IssueCrawler's initial map of the Museum's external network*

In this rendering, the nodes represent websites, which are differentiated by giving them different colours, depending on their domain name extensions - red for those ending with .com, pale green for .org, grey for .org.uk, pink for .co.uk, dark green for any URLs which have been shortened using the .ly shortener, orange for academic sites with the .ac.uk extension, dark grey for governmental sites with the .gov.uk extension and blue for .cn sites, which are domains registered in the People's Republic of China.

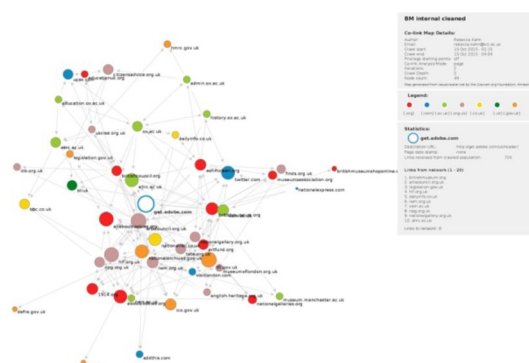
The arrows in-between the nodes represent the direction of the hyperlinks between the sites (inlinks or outlinks) and thicker arrows have more frequent links. The size of the node is determined by the number of inlinks the webpage has from the network, so the more inlinks, the larger the node. In this image, the British Museum's node is in the middle of the larger cluster of nodes, a white dot with a green circumference.

What is immediately noticeable in this rendering is that there are two distinct areas to the map - the central zone, with the British Museum in the middle, and another area of nodes in the top left-hand corner of the map, which consists primarily of websites with the .com ccTLD. These sites have no direct link to the BM but link to it at one level of remove, via Facebook, and have a large number of links between themselves, which explains their close proximity in the cluster map. This cluster of nodes contains links to developer blogs, blog feeds and other generic sites, and all link back to the BM through Facebook. This mapping provides a useful initial overview of the network of actors in the Museum's external network. The range of results, and the diversity of domains represented show the breadth of the Museum's network online. However, the fact that so many entities with very weak links to the Museum were included in the Crawl indicate that the dataset is actually too broad to give meaningful results. What was required, if a solid picture of the Museum's webspace and network are to be established is a cleaner dataset.

## LOOKING CLOSER: REVEALING THE NETWORK

The objective was to use the IssueCrawler to crawl and perform a co-link analysis of all the websites linked to the British Museum, as the central starting point, and interlinked to other sites at a two-page deep distance. The reason for using two pages as the depth setting was because this would render a more in-depth network. The resulting network map would give us an idea of which sites are a first-step link to and from the BM, and what sites link to and from the first level association, in order to establish what the Museum's online network looks like. This is the network represented in *Figure 1*. The algorithm which powers the IssueCrawler is designed to run a co-link analysis by establishing all the external links between the British Museum and any other sites included in the crawl. It sorts the links alphabetically, and then scans through the list and adds any sites which exist in the external links of two or more, thereby including them in the network. First a list is built of all external links (links not pointing to the host from which the links are extracted) found on each site. Then each link from each site is compared to all links of the other sites. If the same link is found it is put in a set of co-linkees with which the next crawl starts.

In order to establish the initial set of links which would form the second iteration of the crawl, I gathered a list of websites which are linked to *from* the BM website and cross-checked them against the set of links used in the first crawl, carefully removing any duplicates. Since IssueCrawler crawls this list of sites, and establishes the number of inlinks and outlinks between the BM and these sites, as well as the inlinks and outlinks which these sites may have between each other, I also took care to remove any websites which might skew the results of the crawl such as links to hosting sites. This included links to sites which were obviously advertising, such as hotels in the area, links to platform or operating system developer pages, or any other type of search engine optimisation content were removed. The end result was a list of 159 starting points, which were then fed into the IssueCrawler in order to render the following network map:

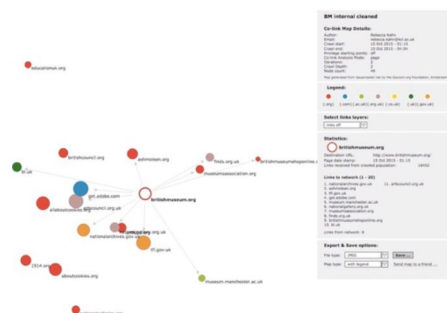


*Figure 2: Initial rendering of the British Museum's internal network, using the IssueCrawler's visualisation.*

In this map, the clusters of nodes are connected by arrows which indicate the direction of the links between the two nodes in the network. Nodes with different top-level domains are represented in different colours, and the nodes are different sizes, depending on the number of links which exist between the nodes. What is most significant is the variety of other types of organisations within the network. This map shows a several clusters of types of institutions, which can be identified by their top level domains, namely .org, .com, .ac.uk, .org.uk, .co.uk, .uk and .gov.uk. Natively digital objects such as IP addresses and URLs can be seen as digital border markers in the process of

mapping national webspaces, and therefore these markers offer us a useful guide for establishing which nationality a website is situated within, as well as the type of organisation it is. It is important to note, though, that there is not always consistency in the use of the tURL across organisations or institutions. The Victoria and Albert Museum, for example, uses the .ac.uk domain, while other museums in the network, such as the Ashmoelan and the British Museum itself use the .org domain.

What this map shows is that the British Museum shares its webspace with a range of different organisations. However, the density of the map, and the number of arrows within in make is difficult to see exactly which institutions and organisations are connected to each other. IssueCrawler allows users to create visualisations of isolated connections between particular nodes in the network, also known as components, and to see only the inlinks and the outlinks which connect these components. When this feature is activated it is possible to see a clearer, more detailed picture of the BM's network:



**Figure 3:** Isolated segment of the network, showing only those sites which have Level 1 links to the Museum

This image shows that, within its immediate network, the British Museum has outlinks to eleven other websites, and of those eleven, five are other museums, archives or libraries: the Ashmoelan, the Manchester Museum, the National Archives, the National Gallery and the British Library. The other six sites are Transport for London (the capital's hub for transport information and journey planning), Get Adobe, a site where users can download and install software (and where, presumably, much of the BM's content links to, in order to allow users to access this content with the appropriate software), the Museums Association - a professional body representing museums,

galleries and heritage organisations in the United Kingdom, the Portable Antiquities Scheme, which is the government's project to encourage members of the public to report any archaeological finds to the appropriate authorities, the BM's own shop, which has a different top-level URL (tURL), and finally, the Arts Council England, a division of the Department of Culture, Media and Sport which distributes funding to the arts and museums in England.

Links in to the museum come from six different sites: Visit London is the official tourism board for the city, the Art Fund is one of the larger fundraising bodies for cultural heritage institutions in the United Kingdom, English Heritage is a charity which looks after the buildings, monuments and sites which comprise the National Heritage Collection, National Express is a British coach service, Ox.ac.uk links to the University of Oxford, and the British Council is the UK's international organisation for cultural and educational exchange. It is important to remember that inclusion in the map is not analogous to having a single link between the British Museum's site and the corresponding node. The IssueCrawler's analytical approach is based on co-link analysis, therefore a site must receive at least two inlinks from the other sites in the network to be included. While this double-link requirement renders a much more robust visualisation of the network that the British Museum operates within, it also means that the network is significantly wider than the seventeen nodes included in the map presented above. In the map legend on the left, I can see that the British Museum's site received 16452 links from the population of websites crawled, and each of those links is a second-level connection between the museum and another site. In order to fully understand the scale of the network, I used the IssueCrawler's *Pages in the Network* feature to render a breakdown of all the starting point pages in the network, the number of deeplinks they received, and where those links came from. This breakdown helped to give some sense of the scale of the complexity within the larger nodes of the network. It also highlights the tension which results from large-scale network analysis; namely the need to balance the desire to reveal a network which has been thoroughly probed in order to render enough detail to provide a useful overview, and a dataset which is small and flexible enough to be meaningfully analysed.

## SUMMARY OF RESULTS AND FINDINGS

In an effort to draw out the significant details of geographic location and relational significance, I checked every link for a location. This consisted of visiting each site, and looking for evidence of nationality or location. Of the 89 starting points in the network, 84 of them are specifically British. This was ascertained through two mechanisms - firstly looking at the ccURLs, which were either .ac.uk, .org.uk, .gov.uk or .co.uk and secondly, in the case of the sites which used generic tURLs like .org or .com, by visiting the site. This process and the results it presented revealed a correlation with the research presented by made by Halavais and Baeza-Yates et al - that the majority of the entities in the British Museum's online network are other British sites, indicating that national similarities are mirrored in online networks, and conforming that 'the number of hyperlinks that cross international borders are significantly less than those that link to sites within the home country' [10]. The implications of this finding will be considered in the conclusion of this chapter.

After locating the British Museum's network in a predominantly British web space, the next step was to examine the entities within this network, in order to surface a profile of the kinds of institutions associated with the Museum as part of the network. In order to achieve this, the 89 sites in the network were coded by type, since the URL extensions are not always an accurate reflection of the nature of the site:

- Academic (universities and research institutes)
- Commercial (businesses or services)
- Funders (research councils and funding bodies)
- Governmental (Ministries and other government services)
- GLAM (Galleries, libraries, museums and archives)
- Non-Profit (registered charities with stated public service mandates)
- Statutory bodies (regulators, ombudsmen)

Seen as percentages of the entire network, the Museum's network consists of the following types of entities, ranked by percentage of the total:

- GLAM: 26 or 29%
- Statutory: 18 or 20%
- Commercial: 13 or 14%
- Government: 12 or 13%
- Funders: 8 or 8%
- Academic: 5 or 5%
- NonProfit: 5 or 5%
- Social media: 2 or 2%

These general results show that while there are more links to other British cultural heritage institutions, there is no preponderance of links to any of the other entities in the Museum's network, but rather a fairly even spread across the categories.

## CONCLUSION - THE NATION IS IN THE NETWORK

This chapter was concerned with looking for evidence of how a sense of national identity is established around British Museum's digital incarnation, and how this identity is bolstered and transmitted in the online space. Taking a methodological cue from current research into issue mapping and national webspace articulation, I used hyperlink and network analysis tools to reveal the constitution of the Museum's online network. Once this was established, and by using Actor Network Theory as a conceptual underpinning, I was able to draw several more conclusions about what the established sense of identity is, and how the network influences (and is in turn, influenced by) the formation and perpetuation of that identity.

What resulted was a series of network visualisations which revealed two significant characteristics of the Museum's network. Firstly, the webspace within which the Museum is located is overwhelmingly British and secondly the Museum's connections are to institutions which for the most part can be characterised as being part of the establishment of British society - government ministries, statutory bodies, charitable trusts and other museums and galleries make up a significant proportion of the British Museum's network. These two factors help to develop the argument that the national identity created and transmitted is actually to be found within the network itself. The network is the evidence of the nation, not only in terms of providing the structural framework also as a result of the origins of the comprising nodes. Unlike issue networks, which map the how disparate groups coalesce

around issues online, the Museum's network allows us to visualise a more static collection of institutions and organisations, which, while heterogenous, also share certain critical characteristics. These characteristics (a common language, orientation towards academia and funding of research and the arts, civic engagement, online proximity to governmental ministries) prompt the argument that the Museum's current network in the digital space is not that different in character from the network it has been part of since the Enlightenment, as explored in Chapter 2. Nationality is defined, represented and bounded by the Museum's location in a network rather than the language it uses to describe itself. While the Museum may publicly articulate itself as a museum for the world, the reality is that there is very little difference between the online and offline positioning of the British Museum. The digital contact zone has not yet been extended beyond the British webspace, and although the Museum might be embracing digital normativity in the descriptions and depictions of its activities, this transformation has yet to filter into the Museum's web-based existence.

## REFERENCES

References should be given as follows:

1. Aronsson, P. & Elgenius, G., 2014, *National Museums and Nation-building in Europe 1750-2010*, Taylor & Francis 2014, pp. 150.
2. Rogers, R., *Preferred placement: Knowledge politics on the Web*, Jan van Eyck Akademie Editions Rogers, 2000 pp.12
3. Latour, B., *Reassembling the Social-An Introduction to Actor-Network-Theory*, by Oxford University Press, 2005
4. Trettien, W. Blog post: "Digital Humanities Vs. The Digital Humanist." Accessed April 2010
5. For more about the DMI, see <https://wiki.digitalmethods.net/Dmi/DmiAbout>
6. Ben-David, Anat. "Mapping Minority Web-spaces: The Case of the Arabic Web-space in Israel", in: Caspi D. and Elias, N (Eds.), *Ethnic Minorities and Media in the Holy Land*. Vallentine-Mitchell Academic, 2014. pp.137-157.
7. Hale, S.A. et al, Proceedings of the 2014 ACM conference on Web Science, *Mapping the UK Webspace: Fifteen Years of British Universities on the web*, 2014. pp. 62-70
8. Woo Park, H. & Jankowski, N.W A hyperlink network analysis of citizen blogs in South Korean politics, *Javnost-The Public*, 15(2), 2008, pp. 57-74
9. Larsson, A. *Is Size is Everything? Comparing Parties on Social Media During the 2014 Swedish Election* Paper presented at ECPR Joint Sessions Workshop, 2015
10. Rogers, R., *Digital Methods*, Boston MIT Press 2013
11. Halavais, A., 2000, National Borders on the World Wide Web, *New Media & Society*, 2(1), 2000 pp. 7-28