

Digitization and Dissemination

A Reverse Image Lookup study to assess the reuse of images of paintings from the National Gallery's website

Isabel Kirton, Master's Student in Digital Humanities, kirton134@gmail.com

Melissa Terras, Reader in Electronic Communication. m.terras@ucl.ac.uk

UCL Centre for Digital Humanities, UCL Department of Information Studies, Gower Street, London, WC1E 6BT



Reverse Image Lookup

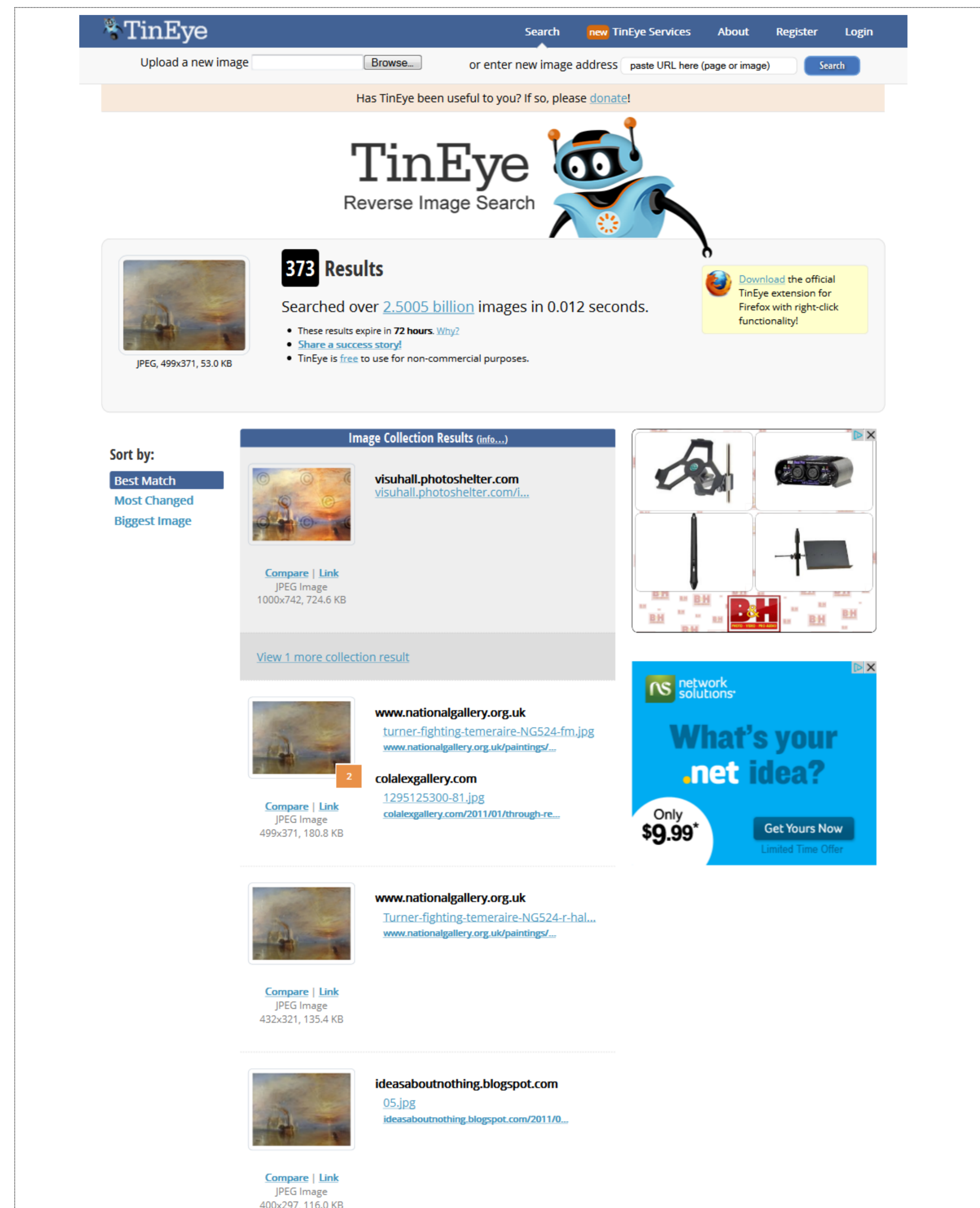
Once digital images of cultural and heritage material are digitized and placed online, how can we tell if they are copied, disseminated, and reused? We used Reverse Image Lookup (RIL) – usually used to identify unlicensed reuse of commercial photography – to ascertain whether it is a useful method in assessing the impact of digitized content.

RIL technologies are those which allow you to track and trace image reuse online, identifying webpages using images that have been copied from an image elsewhere. We believe ours is the first systematic study to use RIL to look at digitized heritage content to ascertain reuse of image content: do people take copies of images of paintings, and use them again? Can we say anything about why and how they use them?

We choose two samples of paintings from the National Gallery, London: all paintings held in Room 34 entitled 'Great Britain 1750-1850', containing 26 paintings by 9 artists, just over 1% of their total holdings. We also created a sample of 6 NG paintings, chosen from different artistic periods and of varying levels of fame. We analysed the dissemination of these images using TinEye (www.tineye.com) and Google Image Search, using a technique called Content Analysis to analyse the contexts for image reuse by identifying the most common categories of reuse, such as commercial art publishers, blogs, reviews, tourism, image collections, encyclopaedias, other museum websites, DVD cover images, and beyond. We demonstrate that type and volume of image reuse is both subject and artist specific.

We then triangulated findings using web access statistics from the National Gallery's Google Analytics account, and from the commercial ISP analysis firm Hitwise. Our results show that the most popular paintings (by access) are the ones most commonly reused elsewhere, but we also uncover a feedback loop which proves dissemination of images elsewhere online provides direct traffic back to the host institutions' website.

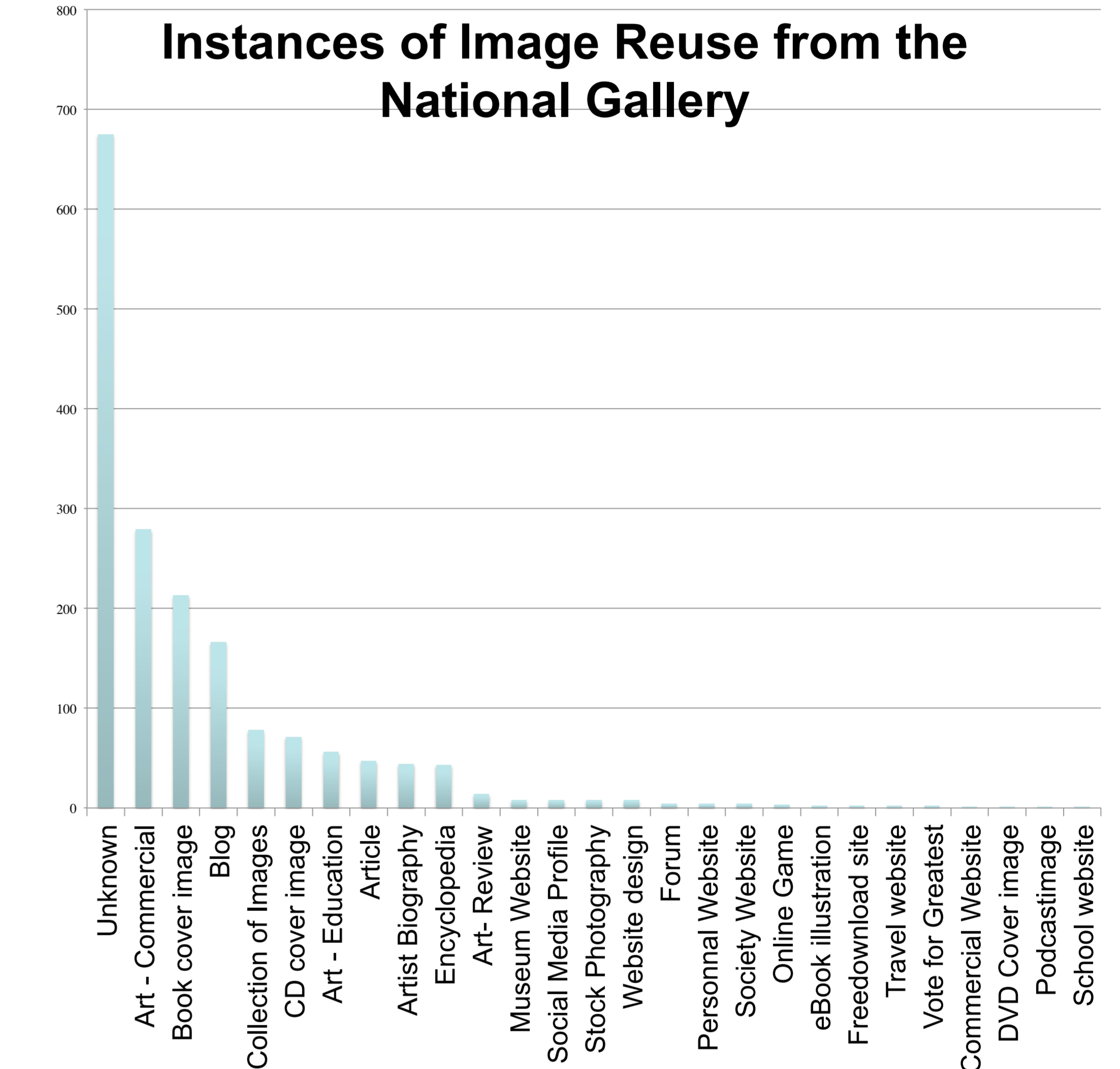
This study has allowed us to establish what motivates image reuse in a digital environment. We recommend a framework for data collection that could be used by other organisations. However, we also show that there are limitations to the information that can be gleaned from a study of this kind, due to the problematic implementation of the RIL tools which were not designed for this sector.



We looked up each image from the National Gallery website in TinEye and then categorised where it had been reused elsewhere online by combing through the results manually. For example, the exact image from the NG website of the Fighting Temeraire, by Turner, was reused 373 times elsewhere online. We then looked at each reuse case to identify in what context the image had been reused.

We also looked up a variety of images in Google Image Search, although the reporting mechanisms are not as easy to use for this type of quantitative study, so would advise anyone attempting to do RIL to use TinEye rather than Google. It took 15 working days to gather the data: RIL is therefore still prohibitive to undertake from a time management point of view, given the nature of the tools available.

Results



In total, we analysed the reuse of 32 paintings, investigating and classifying over 3000 reuse cases of images from the NG. The mean number of reuse instances per painting was 65 whilst the median number of results was 27: images of certain paintings were disseminated a lot more widely than others. Many uses are "unknown" as the URLs are no longer live – indicating the dated datasets currently available for RIL. Art- commercial is the most common reuse – covering posters and reproductions, but there was a high reuse of images on blogs, showing people want to use the content in their own way, reframing it themselves.

We were given access to the National Gallery's Google Analytics account, and Hitwise reports showing ISP traffic, and found that the most visited web pages correlated with the most widely disseminated images. More interestingly, we found a feedback loop which showed that reuse of images from the NG website drove traffic back to the original website – even when the images were not hyperlinked. This is strong evidence that collections should be making their image collections available for reuse on other websites, as the increased dissemination of their digitized content actually encourages people to visit the original website.