

Identifying online display of Food Hygiene Rating Scheme ratings

Maes o ddi-ddordeb ymchwil: [Emerging challenges and opportunities](#)

Hyd yr astudiaeth: 2021-02-01

Cwblhau arfaethedig: 30 Ebrill 2021

Statws y prosiect: Wedi'i gwblhau

Cynhaliwyd gan: FSA Strategic Surveillance Team

Dyddiad cyhoeddi: 22 Mehefin 2023

DOI: <https://doi.org/10.46756/sci.fsa.rdn988>

Identifying online display of Food Hygiene Rating Scheme ratings: Executive summary

Results available: No results available

Maes o ddi-ddordeb ymchwil: [Emerging challenges and opportunities](#)

Cynhaliwyd gan: FSA Strategic Surveillance Team

DOI: <https://doi.org/10.46756/sci.fsa.rdn988>

Hyd yr astudiaeth: 2021-02-01

Cwblhau arfaethedig: 30 Ebrill 2021

Statws y prosiect: Wedi'i gwblhau

Dyddiad cyhoeddi: 22 Mehefin 2023

Background and aims

The Food Standards Agency (FSA) wish to make display of Food Hygiene Rating Scheme (FHRS) rating mandatory for businesses online, extending the current physical display requirements in Wales and Northern Ireland and making both compulsory in England. This project was undertaken to support the rollout and enforcement of mandatory online display, by providing insight into current practice.

Approach

We created an automated solution to finding business websites and establishing whether they are displaying an FHRS rating. Data from Google Places was matched against a sample of businesses from the FHRS open data. This allowed us to collect the businesses' websites where they exist. Only an establishment's own website was included; if the Google data returned a Facebook or other social media page, a chain website, or a presence on an aggregator or booking site (e.g. OpenTable, Trivago), these were excluded.

The websites were passed through a matching pipeline that fetches the website images and matches them against reference images of FHRS ratings. Each image received a score from 0 to 100, representing the confidence of the match. The highest scoring images from each website were compiled, and all images scoring higher than a particular threshold (in this case 30) were examined manually to establish whether they were ratings or not.

Findings

Just over half of the sample (54%) had a business website. We are confident that, where a business has a website, this is being correctly obtained in around 90% of cases. After passing all websites through the matching pipeline, we estimate that the prevalence of online display is around 3% of websites. Takeaways are more likely than other business types to display a rating, while pubs are less likely. We found twice as many FHRS images in England than in Wales or Northern Ireland, despite the initial sample sizes being very similar.

Standardised, high quality images are easy to detect, while variations or low image quality/blurring make it more difficult to establish a match. Our process worked in around 80% of cases; where it did not, this was usually due to unusual features in the source code of the websites that our method had not been able to take into account.

Policy implications and scope for future work

It is hoped that the analysis carried out here can support the drafting of legislation, by providing insight into the prevalence and nature of online display as it is now. The process can also be re-run quickly and easily, for example just prior to introducing a new policy, or in the future to gauge compliance across or for particular local authorities. The technology developed here could also form the basis of a more interactive tool to aid enforcement.

There is limited scope to improve the success rate of the process itself; there will always be variation in websites that cannot be accounted for in full, and some websites (up to 5%) are simply not accessible with programmatic methods such as these. However there is a possibility of increasing the success with which the scraper is able to grab images, and reduce the 20% failure rate by a few percent. A more standardised approach to displaying images on the part of the food businesses themselves would also improve the accuracy with which the image matching algorithm is able to identify ratings images and eliminate similar but non-matching ones.

Identifying online display of Food Hygiene Rating Scheme ratings: Introduction

Background

The present situation

The display of a Food Hygiene Ratings Scheme (FHRS) rating is currently mandatory at physical premises in Wales and Northern Ireland (NI).

Legislation for mandatory (online) display is in the process of being drafted. A version of the legislation was drafted in Northern Ireland in 2017, but this fell through due to the suspension of the parliament, and there is now a desire to revisit it to ensure it is fit for purpose. However there are still unresolved questions, such as what constitutes 'online display', and how best to introduce and enforce this requirement, without creating an undue burden on businesses or Local Authorities. An impact assessment has been conducted for rollout of mandatory display (including online) in England. This suggests a cost both to businesses of compliance, and to Local Authorities of enforcement. A [digital discovery](#) conducted last year concluded that mandatory display would be challenging for businesses, and recommended conducting a voluntary pilot first.

To inform policy on mandatory display in Wales and Northern Ireland, the FSA already commissions an annual audit of physical display, and it would be possible to conduct a similar audit of online display. However, a request was made to the Strategic Surveillance team to explore what can be done in house, including a more automated approach.

This project

The overarching aim of this project was to develop an automated solution to understanding the current state of play with respect to food businesses' online presence and the extent and nature of FHRs ratings display. We developed methods of finding websites and retrieving and assessing images, in order to cover a much larger sample than manual research could in the same timeframe, at a much lower cost.

The key question was: what proportion of businesses are displaying an FHRs rating on their online presence? And as supplementary questions, where online display is found, how is it split across nations and business types, what ratings are being displayed, and how are businesses typically displaying their rating?

The work involved in answering these questions can be divided into two main parts: obtaining the website addresses for the establishments (as this data is not collected by the FSA); and then finding and analysing the images on these websites to see if any of them are (or could be) an FHRs rating.

Developing automated methods of answering these questions allows us to do three things:

1. to explore these questions ahead of time, to help inform the drafting of legislation
2. to generate baseline online display figures before any policy intervention, but also an analysis that can be easily re-run at intervals going forward as policy is rolled out
3. to build a technological solution that could be developed into a tool in future to facilitate enforcement by identifying businesses that are (most likely to be) non-compliant

Timeline

The project grew out of an initial 'proof of concept' piece of work that was carried out in Summer 2019. This phase of development began at the start of February 2021, and ran for three months. Fortnightly checkpoints were carried out with key stakeholders to share developments, resolve questions and ensure development was on an appropriate path. A final 'show and tell' event was held at the start of May 2021.

Sample selection

Although there is not yet an agreed definition of 'online display', for the purposes of this project a working definition was required. It may be that any future legislation takes a different approach, for example prioritising businesses that sell food online. If this is the case, the analysis sample could be altered for future iterations of the work. Table 1 shows the decisions that were taken during this process.

Table 1: Characteristics included and excluded from the sample

Dimension	Scope
------------------	--------------

Include presence on an online aggregator such as Deliveroo?	No – we know these already display FHRS rating
Include social media pages?	No – business websites only
Does the website need to have an ordering facility ('point of sale')?	No – any website even if purely informational ('point of decision')
Include all business types?	No – only restaurants, hotels, takeaways and pubs. Other types would not be expected to have physical display.
What about branches of a chain?	Include if the branch has its own page, exclude if the website is just the overall chain's website

Once a definition had been agreed, a sample was taken from the [FHRS open data](#). The total sample size was set at 1500, which is the same as the physical audit. The sample was split equally across the three nations, but proportionally by business type (i.e. restaurants constitute around half of the in-scope businesses, so were around half of the sample). The other requirement in selecting the sample was that the establishment needed to have at least a postcode to identify its location, so that its business website could be retrieved (this process is detailed in the next section). However as the analysis was not considering mobile caterers, it is unlikely that this would have resulted in substantial loss. The composition of the final sample is shown in Table 2.

Table 2: Final sample composition

	England	Northern Ireland	Wales
Hotel/bed & breakfast/guest house	31	44	52
Pub/bar/nightclub?	108	81	122
Restaurant/Cafe/Canteen?	243	253	226
Takeaway/sandwich shop?	118	122	100
Total	500	500	500

This composition could be adjusted in any future iteration. It would be very straightforward to include a different subset of business types, or change the relative size and representation of the business types and nations. There would also be some scope to change other exclusion criteria, such as allowing chain websites or Facebook pages (although the latter would present both legal and technical barriers with respect to image retrieval and analysis).

Identifying online display of Food Hygiene Rating Scheme ratings: Finding online presence of food businesses

Methods and challenges

The FHRS data does not include an establishment's web presence, and we wanted to avoid having to search manually for this information, to speed up the process both in this project and in any future iterations or re-runs. Therefore this information needed to be brought in from an external source that holds information about businesses, and matched against our data.

There were two main services that could potentially have provided this information: Google Places, and TomTom. Google Places provides data on 'points of interest', including businesses; this is the information about a place that you would see if you searched for it on Google Maps, although the data is used in a variety of ways across different Google and third party applications. TomTom also collects this type of data on places for use in its satellite navigation systems, and also makes this data available to third party developers.

Of the two sources, Google provided by far the better coverage. In a random sample of 150 businesses, Google was able to retrieve details about 112, compared with just 19 for TomTom. Therefore Google Places was used as the method for finding business websites in this project. The cost of retrieving the information from Google Places for a sample of 1500, the size used in this project, is around £60.

Google provides an interface to its data that allows the user to submit a business name and a location (in the form of latitude/longitude co-ordinates), and it will search on that name within a given radius of that location. It then returns data for the corresponding business (or what it considers the best match, if there is more than one). Although these requests need to be submitted individually, a simple program can be written to do this across a sample of businesses, and return a dataset of information about those businesses.

Although the vast majority of establishments will have a presence on Google Places, there will be some that do not, and although Google's algorithm for searching for and retrieving data on the business is very good, and will return the correct data in the most cases, it is not perfect. However it is the best available method for retrieving this kind of information, and quality assurance procedures (detailed below) suggested that the results are of high quality.

This process was applied to our sample of 1500 businesses. Once the website addresses had been retrieved, a further step took place to exclude websites that were social media pages (Facebook, Instagram or Twitter), booking websites (booking.com, Trivago or hotels-247), or the main website, as opposed to a branch-specific page, for one of the 75 largest chains in the UK (e.g. Costa, Starbucks, Dominos).

Results

Full sample

This process of retrieving business URLs returned business websites for 803 businesses – 54% of the total sample. Below summarises the number of businesses that were eliminated at each stage.

- Full sample = 1500
- Matched to a place on Google = 1336 (89%)
- Has a business website = 927 (62%)
- Website is not Facebook, chain level or third party = 803 (54%)

Just under 90% of establishments were matched to a place on Google Places. Of these, 927 had a business website, but after the final exclusion step, there were 803, or just over half of the sample.

Quality assurance

It is not possible to go through the whole sample to check whether this process has worked, and would defeat the purpose of an automated solution. However, a quality check was carried out on a sample of 102 businesses to establish whether the websites were missing because they do not exist, or because they could not be found with our methods, and whether the websites that were returned were the correct ones.

Of these 102 businesses, the process had failed to find a Google Place for 7 of them. A manual check suggested that 4 truly did not exist as a Google Place, 1 was not returned because it was marked as closed, 1 was present but the name recorded in the FHRS data was too different to match, and the other non-retrieval was unexplained. Therefore we can say that coverage of the businesses in the Google Places data is 96%, and our process successfully matches establishments to this data in 98% of cases.

In 32 cases, an establishment had been matched to a place on Google, but there was no business website in the data returned. This was found to be correct in every case; none of these businesses had a website that we could find. Of the 54 websites that were found, 49 were correctly identified. Therefore it is highly unlikely that a website exists and we have not found it, but around 9% of the websites found will not be the correct web address.

This suggests that the process, while not failsafe, has returned a set of business websites that we can be broadly confident are the right ones, and that it is not systematically failing to uncover businesses' online presence. The distribution of business websites across business type and nation is shown in Table 3. Despite the total sample size reducing by nearly half, all business types in all nations are still represented in the sample.

The lack of a business website for almost half the sample raises the question of whether many food businesses are relying on social media for their online presence. As we have excluded such websites from this analysis, our process would be failing to capture a significant aspect of online display if this were the case. We examined a sub-sample of 100 businesses in our sample for which there was no business website. Of these, 37 were found to have a Facebook page. Interestingly, 7 of these Facebook pages were displaying a rating, which is a much higher prevalence of online display than was found among business websites. However the wrong rating was displayed in 4 of these cases. Restrictions on the scraping of social media pages mean that it is difficult to incorporate them in the image matching process developed for this project, however the proportion of 'Facebook only' businesses in the sample is relatively small; around 17% of the sample, or around 256 businesses.

Table 3: Composition of final sample of websites

	England	Northern Ireland	Wales
--	---------	------------------	-------

Hotel/bed & breakfast/guest house	22	30	43
Pub/bar/nightclub?	59	23	61
Restaurant/Cafe/Canteen?	144	127	124
Takeaway/sandwich shop?	71	49	53
Total	296	229	281

Options for future iterations

At the start of the project it was not known what proportion of the sample would have an online presence. As it transpires that this figure is quite low, a larger sample of businesses would ensure a final website sample size closer to the target. As around 8% of the links in the sample were broken/inaccessible, a larger sample would also compensate for this loss.

As the only restriction was on business type as it is recorded in the FHRS data, this sample includes establishments that may not be truly in scope, for example workplace canteens, and educational establishments (where the FHRS record has been matched to the company or institution rather than its café or catering facilities specifically). If mandatory display is restricted to consumer-facing businesses in future, then this sample should probably be restricted in a similar way, with manual fine-tuning of an initially random sample.

Implications for policy and practice

The difficulties encountered in finding food businesses' online presence would provide some support for requiring businesses to provide this information, at point of registration or inspection, and for local authorities to submit it with their FHRS return.

Identifying online display of Food Hygiene Rating Scheme ratings: Assessing online display

Methods and challenges

Once a sample of websites had been generated, these were passed to an image matching pipeline. This pipeline obtains the underlying code for the website, identifies the images within the code, extracts them, and applies image matching algorithms to determine whether each image is an FHRS rating. It also travels to each page that is one 'click' into the website, and evaluates the images it finds on each of those pages. All images are downloaded temporarily while the website is being analysed, and then deleted at the end of this process. Only the URL of the closest matching image is retained.

The matching algorithm compares each image against a set of reference images – i.e. those that are known to be FHRS ratings. Four versions of each rating were used, encompassing both the digital and ‘window sticker’ formats, and including a bilingual version.

Each website image received a score from 0 to 100, representing the confidence of the match, against each reference image. The highest scoring images from each website were compiled, and all images scoring higher than a particular threshold (in this case 30) were examined manually to establish whether they were ratings or not.

A key challenge in implementing this process across a large sample was the time taken by the image matching algorithm. These issues were addressed by:

- adding a quicker initial matching step to eliminate all but the most likely matches
- restricting the number of images that can be sent to the slower, more powerful matching algorithm (at a cost of potentially missing a rating image in a very large website – our quality checks suggested this is unlikely to affect many websites however)
- utilising the computer’s processing capacity more efficiently, allowing multiple matching processes to occur at the same time
- once the pipeline was fully developed, running the full sample through it on a slightly more expensive, but more powerful computer

These innovations resulted in a process that can be run end to end within a day.

Results

Full sample

Analysis of the full sample suggested that the prevalence of online display of FHRS ratings is around 3%. This is the proportion of businesses with an online presence that are displaying a rating. It represents around 1.7% of all in scope businesses (as only around half were found to have an online presence).

The matching process identified 21 websites in the sample that were displaying an FHRS rating. We believe based on our subsequent quality assurance checks (see next section) that this represents about 80% of what is truly present. Therefore the final total is likely to be around 26 websites, or 3% of the sample of 803.

Of the 21 websites that were uncovered, half were establishments in England, with a quarter each in Wales and Northern Ireland respectively. As the nations were roughly equally represented in the sample, this suggests that online display is disproportionately common in England.

Table 4 shows the proportion of each establishment in the original sample, compared to the 21 establishments that were found. Restaurants and hotels were represented in the same proportion as they were in the sample, with pubs under-represented, and takeaways over-represented.

Table 4: Online display by business type

	% in sample	% in found
Hotel/bed & breakfast/guest house?	12	10

	% in sample	% in found
Pub/bar/nightclub?	18	5
Restaurant/Cafe/Canteen	21	33
Takeaway/sandwich shop	12	10

All the websites were displaying a 5 rating, apart from one, which turned out not to be the business's own website. A comparison with the actual ratings on ratings.food.gov.uk found that two of these businesses were rated 4, despite their professed dedication to hygiene.

Quality assurance

There are two types of error in using a predictive algorithm; it predicts that an image is a rating when in fact it is not (a 'false positive'), or it predicts that an image is not a rating when in fact it is (a 'false negative'). The only way to truly know these numbers would be to go through every website in the sample manually; the outcome we are trying to avoid. However, the analysis of smaller sub-samples can help us to estimate the extent of these two error types, and therefore how much confidence we should have in our results.

False positives

The image matching process does not make a 'yes' or 'no' decision about whether the image is an FHRS rating; it assigns a score between 0 and 100, where 0 is definitely not a match and 100 is a definite match. Therefore a decision needs to be taken by the analyst of what threshold should be taken to imply a (possible) match. For this analysis, a threshold of 30 was used.

However, this does not mean that every image above this threshold was an FHRS rating. The analysis returned 48 unique images, across 66 websites, above this threshold. Each of these images was examined manually to determine whether it was a rating. 27 of these were not ratings; they were 'false positives' in this process. This is quite a high false positive rate of around two thirds.

The false positive rate can be reduced by increasing the threshold score. For example, raising this threshold to a matching score of 50 would reduce the number of false positives to just 5. However, the disadvantage is that it increases the false negative rate; in this case it would only have identified 18 of the true images instead of 21. Therefore, given the fairly low prevalence of high or even moderate scores (only 8%, or 66 websites in this case, scored over 30), it is better to set the threshold low and manually review. It takes considerably less time to run the analysis and review 48 images (1 day) than it does to review 800 websites (which could take several weeks), so a great deal of time could be saved.

It is difficult to say why an image matching algorithm has assigned the score it has. However, an examination of some of the false positive images suggests that circular features within an image – like the circles around the numbers in a rating image – may be tricking the algorithm.

False negatives

False negatives are harder to quantify, as you do not know the true extent to which images are displayed online. However, we used two approaches to try and quantify the false negative rate.

The first was to look at a sample of 100 sites to which the algorithm had assigned a very low score (below 15), or for which it had been unable to obtain images or calculate a score. Of these, 2 were in fact displaying a rating. In both cases, the scraper and image matching pipeline had run successfully, but had failed to identify the FHRS image. In one case this was because the image was bundled in with others, and in the other because the image was not of a type that the image matching pipeline could process. In 72 cases the pipeline had worked as expected, but the image was not present. Of the remaining 24 that did not work, 8 of the links were broken or inaccessible, and in the remaining cases, the scraper was unable to retrieve the images. Therefore we could say that the scraping and matching pipeline failed for 18 of the 92 working links, or about 20% of the time.

The matching pipeline was also run for an additional sample of 50 websites that had already been identified as displaying a rating; these were discovered through reverse Google image search or by chance. In 40 of these, the correct image was identified and scored above the threshold (it was detected in a further 2 cases, but with a score below the threshold). Again this suggests that the pipeline is failing around 20% of the time.

Taken together, these suggest that the matching process is likely to be detecting around 80% of the websites displaying FHRS images, and there is unlikely to be widespread online display beyond what we have been able to estimate here.

Factors affecting the score

The score assigned to an image is in part beyond our control; these algorithms are complex, externally developed libraries of code. There is therefore little that can be done about false positives; the algorithm has decided that the Instagram logo looks like the FHRS one, and it cannot be trained to believe otherwise. However, there are a number of factors around image quality that seem to result in the algorithm giving an undesirably low score to images that are actually ratings.

The highest scoring images received an almost-certain score of 98.9. They were universally high quality images of the digital version of the badge. The 'window sticker' style images are also capable of receiving a high score when they are good quality, which received a score of 98, but a poor quality version scored just shy of the threshold.

The matching algorithm also found it much easier to match standardised images. Images that did not pass the threshold included: one website that had made its own version; another that had a photo of a sticker in a window (although elsewhere these were picked up if they were clear, straight on photos); a website that took a screenshot from ratings.food.gov.uk (with too much surrounding noise); and a rating image that was bundled together with other graphics in a single image.

Identifying online display of Food Hygiene Rating Scheme ratings: Conclusion

Options for future iterations

There is some scope to reduce the proportion of websites that do not successfully pass through the matching process, which is currently around 20%, but it is limited. The scraper is looking for images in the website's source code, and expects to find these expressed in a particular way; however in around 7% of cases it could not find the images because they were represented in a slightly different way. Although it is not possible to incorporate every variation in this respect, some adjustment to the scraper could probably halve this. This would increase the success rate of the matching pipeline by a few percent.

However, beyond this, there is little from a technical point of view that could improve the scraping and matching process. There will always be quirks in websites that will fall foul of the assumptions that need to be made to enable this type of bulk, automated processing. Some websites use types of images that are not compatible with even the latest technology available to carry out image matching. And some websites will always be designed, intentionally or unintentionally, in a way that makes it impossible to retrieve their content using code.

The technology developed in this project to find and analyse food business websites could form the basis of a more interactive tool to reduce the amount of routine work involved in monitoring compliance. However, the legal implications of doing so would need to be thoroughly evaluated; they were not here, as any such tool was beyond the scope of this project. However, any future work of this nature would need to consider whether it would be in breach of legislation such as the Regulation of Investigatory Powers Act, which limits the use of online surveillance in regulatory enforcement.

Policy implications

The biggest gains in improving this image detection process would be found if the quality and standardisation of FHRS images were improved. None of the false positive images scored higher than 75, and most scored 50 or less. By comparison, the best quality FHRS images received match scores close to 100. Yet only two of the images in the full sample were found to be using this type of top quality FHRS image. More widespread use of a standardised image, of the correct type and of a minimum size, would make it much easier to detect which sites are displaying a rating.