

# REPORT OF A BREACH OF THE CODE OF PRACTICE FOR STATISTICS

## 1. Core Information [\[guidance\]](#)

|   |   |
|---|---|
| <b>Title and link to statistical output</b>                       | <b>Unistats data set</b><br><a href="https://www.hesa.ac.uk/support/tools-and-downloads/unistats">https://www.hesa.ac.uk/support/tools-and-downloads/unistats</a>                                   |
| <b>Name of producer organisation</b>                              | <b>Higher Education Statistics Agency (HESA)</b>  |
| <b>Name and contact details of person dealing with report</b>     | <b>Jonathan Waller</b> <a href="mailto:jonathan.waller@hesa.ac.uk">jonathan.waller@hesa.ac.uk</a><br><b>Rebecca Mantle</b> <a href="mailto:Rebecca.mantle@hesa.ac.uk">Rebecca.mantle@hesa.ac.uk</a> |
| <b>Link to published statement about the breach (if relevant)</b> | <b>See above URL to Unistats web-page</b>   |
| <b>Date of breach report</b>                                      | <b>12 September 2019</b>  |

## 2. Circumstances of breach [\[guidance\]](#)

|  |  |
|--|--|
| <b>Relevant principle(s) and practice(s)</b>   | T3.6 Statistics should be released to all users at 9.30am on a weekday |
| <b>Date of occurrence of breach</b>  | 11 September 2019  |
| <p>The Unistats dataset contains information about courses offered by higher education providers. It is published by HESA on behalf of the Office for Students. It was scheduled to be published at 09:30 in accordance with the pre-announced publication schedule and principle T3.6 of the Code of Practice for Statistics. In the event the data set was actually published at 10:04, some 34 minutes late. Although the host web-page had been updated on schedule, the data set users were able to download between 09:30 and 10:04 was actually the previous edition. The edition of each downloadable data set is expressed as a date within the file naming convention, so any files downloaded during this period would have indicated the previous edition date.</p> <p>The cause was that, although our update process ran at 9:30 as planned (and finished at 9:32), there was an issue with a global caching rule which had to be manually overridden in order to fix it, which took rather longer than we would have expected. In more technical terms, content is served via a CDN (Content Delivery Network) service hosted in Azure. When a user clicks to download they access the file from the CDN from a server that's closest to them by utilising caching. This reduces the time it takes for them to receive the file. Each file is cached which means a copy of the file is stored on each server in the CDN, this is updated periodically which is why the older file was available for a short period. HESA makes use of "Caching Rules" to configure how long each file is cached before the new copy is pushed to the edge servers. This is why the issue occurred because the rules didn't function as we would expect, requiring manual intervention to resolve.</p> |  |

## 3. Impact of the breach [\[guidance\]](#)

|  |
|--|
| <p>The website usage metrics reveal that three visitors viewed the Unistats web-page between 09:30 and 10:04 on 11 September, one of whom was the HESA staff member monitoring the release. Of the other two we know that two OfS staff members were also monitoring the release and accessed the web-page. This leads us to conclude with a high level of confidence that no other external user would have attempted to download the Unistats data set within this period.</p> |
|--|

#### 4. Corrective actions (taken or planned) to prevent re-occurrence[\[guidance\]](#)

A message was posted on the Unistats web-page to notify users of the delay and apologise. An email was sent out to a group of users of the data set who have chosen to be notified of updates. This email also explained the delay and offered HESA's apologies.

In terms of corrective actions, we are working with Microsoft to develop an understanding of why the caching rules did not work as expected. Once a technical solution has been identified it will be implemented and thoroughly tested.

In the meantime any further releases of the data set will be subject to enhanced monitoring and manual intervention to ensure timely release.