REPORT OF A BREACH OF THE CODE OF PRACTICE FOR STATISTICS

Government Statistical Service

1. Core Information [guidance]

Title and link to statistical output	**Business enterprise research and development 2018**
Name of producer organisation	**Office for National Statistics**
Name and contact details of person dealing with report	**Debra Prestwood** **Statistics Head of Profession,** **Office for National Statistics,** **Government Buildings,** **Cardiff Road,** **Newport,** **NP10 8XG**
Link to published statement about the breach (if relevant)	**The Business enterprise research and development 2018 publication was delayed due to disclosive data found in the time series dataset. The updated time series data will be published as soon as it is available.**
Date of breach report	**22/11/2019**

2. Circumstances of breach [guidance]

Relevant principle(s) and practice(s)	T3.6 Orderly release. Statistics should be released to all users at 9.30am on a weekday. T6.4 Data governance Appropriate disclosure control methods should be applied before releasing statistics and data.
Date of occurrence of breach	**21/11/2019**

The Business Enterprise Research and Development (BERD) survey is published annually by the Office for National Statistics. The publication was scheduled to be released at 9.30am, Thursday 21 November.

The publication is comprised of three elements; statistical bulletin, dataset and time series.

The statistical bulletin and dataset elements were signed off by the Deputy Director (DD) in the afternoon of 20 November in line with normal procedure. The time series is produced via CSDB (a legacy system) as part of an overnight run prior to publication. Therefore, the DD did not have sight of this when signing off the other components of the release the day before publication. The BERD team checked the time series when they were notified it was ready to be reviewed by publishing at 8:45am on 21st November (after the overnight CSDB run), and noted that disclosive data was present in the time series file. The publishing team was immediately told. The publishing team advised that, due to the timing of the notification, it was not possible to remove the time series from the release, so at 9:25am the decision was made for the entire release to be delayed, and the release was not published as scheduled at 9:30am.

The time series table was removed, and the statistical bulletin and dataset were published at 10:15am. A notice was also published informing users why the publication was delayed and that the time series issue would be investigated and a non-disclosive time series would be published at the earliest opportunity. The development team in publishing had to manually

remove the timeseries from the publication collection before publishing the rest of the release at 10:15am.

The manual removal of data is a complex process which involves the editing, moving and deleting of hundreds of timeseries data files. This is not a quick process, as the system we use is not designed for this action, it is specifically designed to prevent the rollback of data, to ensure that ONS is transparent when publishing updates and corrections (to provide a history of changes to data published).

However, after publication at approximately 10:45am, the development team noticed that the disclosive timeseries data had been published alongside the bulletin and dataset. This was due to human error while manually cleaning up the collection. In addition to the visible data in the collection there is also a 'zip' archive of the time series that is created immediately prior to publication (when in the publishing queue) to optimise the publishing performance. This zip file was missed while sanitising the collection. This meant that although the data wasn't visible at preview/sign off time, the system already had the data queued up to publish to the website.

The development team then proceeded to manually clean up the disclosive timeseries published to the website. This involved very similar steps to cleaning the collection but with additional manipulation of versions and repeating the changes across different servers. These changes were completed by 11:20am, and the disclosive timeseries was removed. It should be noted, that the route to navigate to the timeseries was not available when the release was published at 10:15am, therefore it would have been very difficult for users to find this timeseries data – but it was possible through the timeseries explorer on the ONS website - which is not typically used.

3. Impact of the breach [guidance]

Although there was a 45-minute delay in publication, we had no formal complaints from customers. We spoke to the Head of Profession in BEIS who is the main policy department that uses these statistics, and they were unaware of any difficulty this had caused.

Disclosive data was released. However, Google Analytics has shown that while 40 people looked at the release page, they would not have been able to access the time series data via that route. Only a user who used the time series explorer (which is not typically used), would have been able to find the data in the 65 minutes it was available before being pulled. Google Analytics suggests that no-one viewed the disclosive data through this route either.

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

Secondary data disclosure amendments failed in CSDB. This was identified by the team shortly before publication (9.20am). Amending secondary disclosure within the CSDB system is an established process that has been carried out successfully by the BERD team over a number of years. Therefore, we have introduced a new step to the process that negates the need to make data changes in CSDB in future.

The error was identified shortly before the scheduled release, and the following remedial action was taken immediately to prevent disclosive data from being published on the ONS website:

- We informed the publishing team who were unable to remove the affected tables in the short time before publication as they too are dependent on legacy systems
- The publishing team worked to remove the affected time series table from the bulletin, resulting in the bulletin publication being delayed by 45 minutes, however, as

highlighted, the disclosive timeseries was still published due to data being already in the publishing queue
- Notified the Bank of the issue and ensured they did not access the time series data that was automatically delivered to them via CSDB – which they confirmed they did not
- Work on the correction of the time series data in order to include in the bulletin the next day
- Contacted the Head of Profession in BEIS which is the main policy department to check that the delay did not cause them any issues
- We are working with other areas in ONS to understand where the process might have gone wrong so that we can reduce the risks in future publications

In the longer term, we recognise that it is important to review and update our processes. In order to prevent this happening again:

- We have introduced an additional step (implemented immediately) in the process carried out by the business area before the data is delivered to the publishing team. A new RADWRK status has been introduced after advice from CSDB experts, which allows the sign-off of CSDB data on the day before, rather than the day of publication.
- Secondary disclosure control will be implemented in Excel prior to upload to RADWRK, rather than editing within CSDB.
- The additional step will enable sign-off of the data prior to it being moved to RADHLD preventing need for wholesale checks on morning of publication.
- A programme of training and refresher courses will be put in place for the BERD team and other users in collaboration with the CSDB team.
- Through the Economic Statistics Quality Group, we will check if other areas of Economic Statistics have experienced these issues and share good practice with all members of the group to avoid these problems in future.

From a publishing perspective the steps in place are:
- The digital publishing development team will create documentation outlining which files need to be deleted in order to delete the most common content types. Although it is not possible to document all cases as each situation is different, this should decrease the cognitive overhead for some known situations.
- The Publishing team will support business areas using the Publishing Hub to ensure releases are correctly prepared for publishing.