Financial Times Limited

Test of accuracy of the FT's calculation and reporting of qualifying content ('Metric V'), for the period 1 January 2019 to 31 December 2019

pwc

Contents

Distribution list

For information: **Tom Betts**
                     (Chief Data Officer, FT)

                     Cait O'Riordan
                     (Chief Product and Information Officer, FT)

                     Caspar de Bono
                     (Managing Director, B2B, FT)

Introduction

Introduction

Objectives

The Financial Times Limited ('the FT') collects data on its customers' digital content consumption through its dedicated website 'FT.com', the FT mobile application, its online 'ePaper' and email-distributed newsletters to subscribed users. This data is used to calculate different metrics which measure the reader activity of individual users, and monitor the collective engagement levels of organisations with which the FT holds Business-to-Business ('B2B') contracts.

'Metric V', a measurement of the volume of 'counted' content consumed, is a key calculated metric used as an independent measurement and as a parameter in the calculation of other derived metrics. For the purposes of our reporting, we will refer to the collective as 'Metric V.'

PricewaterhouseCoopers LLP ('PwC' or 'we') have been engaged by the FT to test the accuracy of the calculation and reporting of Metric V.

The scope of our work is set out in our engagement letter dated 14 January 2020. This report sets our test procedures and results.

PwC test procedures

PwC has completed testing in five key areas:

1. User completion of digital subscription forms
2. Collection of user data on different digital platforms
3. Tagging of 'counted' content by the FT's editorial team
4. Secure maintenance of content usage data
5. Reporting of Metric V data in Enterprise Tools

Detailed results are included in the body of this report.

Our respective responsibilities

The FT is responsible for:
- developing the definition of Metric V;
- developing a methodology for calculating Metric V;
- maintaining the underlying systems from which data is extracted to calculate Metric V;
- accurately reporting Metric V; and
- ensuring appropriate controls are maintained over the methodology and systems used in the calculation and reporting of Metric V.

Our responsibility was to test the FT's calculation and reporting of Metric V, for the period 1 January 2019 to 31 December 2019. Our work was not designed to consider any changes that may occur to how Metric V is calculated subsequent to the date of our testing or the validity of how Metric V is defined.

This report has been prepared for and only for the directors of the FT as a body for management purposes. We permit the publication of our report for the sole purpose of assisting the directors in demonstrating how they have satisfied themselves as to the accuracy of the calculation and reporting of Metric V. We do not accept or assume responsibility for any other purpose or to anyone other than the directors as a body and the FT for our work or for this report.

PricewaterhouseCoopers LLP
London
5 October 2020

Executive summary

Executive summary

Counted content across digital platforms

The FT publishes content across four digital platforms:

- FT.com
- Mobile app (including both iOS and Android versions)
- ePaper
- Email-distributed newsletters

Metric V is defined as the total volume of 'counted' content consumed across each of these digital platforms, and determines a number of metrics, including the pricing for B2B subscribers. The FT Editorial team are responsible for assessing whether each article published should be tagged as 'counted' or not - this is based on the nature of the content and refers to content behind a paywall. Only articles which are tagged as 'counted' are assessed for Metric V.

Testing performed

We have performed procedures to assess:

- the accuracy of registration data for users (Procedure 1);
- the accuracy of the application of the methodology in identifying and counting Metric V and capture of counted content for each of the four digital platforms in scope (Procedure 2-5);
- the design of controls relating to program development and user access to the Redshift database where data is captured and stored (Procedure 6-9); and
- the accuracy of reporting of data in the reporting module Enterprise Tools (Procedure 10).

Executive summary

Results of testing

No exceptions were noted for procedures 1 to 9.

On procedure 10, (the reconciliation of the underlying content data usage recorded in Redshift to the value of Metric V that had been reported in Enterprise Tools for a sample of 35 users across 7 different corporate accounts), two exceptions were identified across five users on three different B2B contracts. The exceptions were a result of two separate issues that caused over counting of Metric V during the period 1 January 2019 to 31 December 2019.

Exception 1 was three cases of overcounting caused by an incorrect change to a paywall barrier in June 2019.

Exception 2 was two cases of overcounting caused by a code error that allowed multiple email opens to be counted if they occurred within the same timestamp, when only one email should be counted.

More details of our test procedures and results can be found on page 14-16.

Management comments

FT has responded to the exceptions noted with the code errors being fixed and the historical data corrected, although this has not been verified within this report. Management has identified that 36 additional B2B readers were charged for across 29 corporate customers. This is less than 0.01% of all B2B subscribers. The 29 corporate customers have been contacted by the FT. See management comments on pages 16.

Detailed findings

Detailed findings

Background

The Financial Times Limited ('the FT') is an international daily newspaper with a focus on reporting business, economic and financial news. The FT published its first newspaper, a four page journal, in 1888 in the UK, and is now part of Japanese media organization Nikkei following its acquisition in 2015. Until 1996, the FT operated solely as a print-based publication, circulated throughout Europe, the Americas and Asia. Since the late 1990s, the FT has operated through digital channels on multiple platforms as digital media has developed. The FT now reaches audiences through its dedicated website 'FT.com', the FT mobile application, its online 'ePaper' and email-distributed newsletters to subscribed users.

The FT collects data on its customers' digital content consumption across the multiple aforementioned media platforms. This data is used to calculate different metrics which measure the reader activity of individual users, and monitor the collective engagement levels of organisations with which the FT holds Business-to-Business ('B2B') contracts. 'Metric V', a measurement of the volume of 'counted' content consumed, is one of the key calculated metrics and is used both as an independent measurement and as a parameter in the calculation of other derived metrics including 'RFV' and the 'engaged reader metric'.

The process by which digital content consumption is captured, counted and reported is detailed in the following six sections:

1. Defining Metric V and the methodology for calculation
2. Tagging of 'counted' content by the FT's Editorial Team
3. User completion of the online registration form to subscribe to digital content
4. Collection of granular data on customer usage for four digital platforms
5. Secure maintenance of content usage data
6. Reporting and use of Metric V via in-house system 'Enterprise Tools' for commercial purposes

Our test procedures, and the results, are listed on pages 14 and 15.

Detailed findings

Definition and methodology for calculation Metric V

Metric V is defined as the total volume of 'counted' content consumed by subscribed users of the FT's four digital platforms, FT.com, mobile app, epaper and email distributed newsletters . This consumed content is recorded within a RedShift database.

How content is tagged as 'counted' by the FT

Only viewing counted content contributes to the calculation of Metric V. The Editorial Team use a method for tracking articles where content usage data is enriched and articles are tagged through the use of a set of 'Content Application Program Interfaces ('APIs')'. Any changes and additions are controlled through developments of these APIs. A formal change management process is in place to direct this, and change logs are captured within GitHub to record an audit trail of all amendments.

To be counted a content barrier must not be viewed (i.e. paywall), the Content API must state an article belongs to the correct classification defined below, be on a particular list of brands, or part of Markets Data, and the page must not be navigational (viewed for five seconds or less) on the FT mobile application. The FT's Editorial Team classifies articles into two groups which impact if they are 'counted' or not:

- Conditional articles are all potentially countable content and are split into four subcategories; Registered, Standard, Premium, and Weekend. Based on subscription type a user will be able to access different articles in these subcategories.
- Unconditional articles are available to all customers (including anonymous users) and are not counted.

User completion of the online registration form to subscribe to digital content

The FT's users sign up for digital content by completing the FT.com registration form. This provides the user with single sign on ('SSO') login details which can be used to access content via the website, mobile applications, ePaper and through the FT's email-distributed newsletters.

Detailed findings

User completion of the online registration form to subscribe to digital content (continued)

The content available to the user is dependent on their type of subscription:

1. Registered (non-paying): Users can read up to three standard-access articles each month for free.
2. Standard subscription: Users have non-restricted access to all standard-access articles and blogs via digital and mobile platforms with the exception of the ePaper.
3. Premium subscription: Users have non-restricted access to all standard-access articles, as well as exclusive content, via digital and mobile platforms, including the ePaper.
4. Weekend subscription: Users to have non-restricted access to all Weekend-access articles as well as up to three standard-access articles each month.

Collection of content usage data on four digital platforms

The Financial Times collects user activity data from multiple digital channels, four of which contribute to Metric V; the Financial Times' dedicated website 'FT.com', the FT mobile application, the online 'ePaper' (a digital replica of the printed daily publication), and the email-distributed newsletters. The FT records detailed user activity on the domain FT.com via the website, also known as 'Next', as well as on the App using the FT's in-house 'JavaScript' tracking program.

The four digital platforms consider content consumption on an individual basis, each following separate logic:

(a)   FT.com

The FT records detailed user activity on the domain FT.com, also known as 'Next', using the FT's in-house 'JavaScript' tracking program. All individual page views, including page refreshes and backwards navigation, are recorded. 'Counted content' excludes the FT.com home page. Markets Data also has specific areas on the page that qualify as 'Counted content'.

(b)   FT mobile application ('app')

A mobile app is available for both the iOS and Android operating systems. All individual page views are recorded, however only pages viewed for more than five seconds are counted, as views of five seconds or less are considered to be 'navigational'. All other rules over what is counted are the same as for FT.com and the same JavaScript tracking library is used. Offline activity is also captured on the App which can contribute to 'Metric V.

Detailed findings

Collection of content usage data on four digital platforms (continued)

(c)   ePaper

The FT's ePaper is a digital edition of the standard printed newspaper. Content usage is tracked and recorded by third party software 'Olive' before integration into the FT's main database. Page views are recorded when articles are clicked on.

(d)   Email-distributed Newsletters

The FT sends out regular emails to subscribed users. This includes general administrative correspondence, weekly summaries, promotional emails and the recurring distribution of newsletters (e.g. 'Best of Lex') which may contribute to Metric V.

Every email opened gets recorded, but only newsletters with identifiers on a designated 'tagged' list defined by the Editorial Team are counted for Metric V. If multiple emails with the same list identifier are opened with less than a thirty minute lag between them, only the first opening will be counted. Subsequent openings are recorded but are not counted. Clicking on links within emails does not contribute towards Metric V, but may navigate the user to counted content on the FT's website which will then be counted.

Content consumed across each of these digital platforms is assessed for each B2B customer, through the relationship between individual accounts and the customer to which they are mapped.

Secure maintenance of content usage data

The FT stores its content usage data in its Amazon Redshift PostgreSQL database. The integrity of this data is safeguarded by the following controls:

1. User access to the database is managed by segregating access rights to different groups based on role;
2. Hourly snapshots of the database are recorded in AWS and retained for 30 days;
3. Audit logging from the Redshift database is in place through Splunk and held for six months.

Detailed findings

Reporting and use of Metric V through 'Enterprise Tools' for commercial purposes

The external reporting platform 'Enterprise Tools' is used to monitor, manage and report the engagement of users, for which Metric V is one measure, in companies with B2B contracts with the FT. Enterprise Tools provides highly granular information on an individual and contract-aggregated level, as well as a visual representation of user engagement. The 'Active Core Reader' metric is also visible on the tool, defined as a binary metric stating whether any thirty day rolling window contained nine or more pieces of 'counted' content for a user. This metric has been assessed in the reconciliation between Redshift and Enterprise Tools for the sample of 35 users in addition to Metric V.

The content usage data is loaded from the Redshift database into a staging table that is segmented by Corporate ID before being transferred into Enterprise Tools. Enterprise Tools reporting filters data based on contract start and end dates. B2B contract owners can view this data through the Enterprise Tools website which can be found at: https://enterprise.ft.com/.

Detailed findings

Testing performed

We performed 10 procedures in total. For the reconciliation of Metric V to Enterprise, we selected a sample of thirty-five users across seven B2B contracts. Testing was performed for the period 1 January 2019 to 31 December 2019.

	Summary of procedure	Results
1	Test the capture of user registration data by completing the user registration process.	For PwC test entry, underlying data captured in the registration process was recorded accurately in the FT database.
2	Test the accuracy of the recorded number of counted pages in Redshift for sample users accessing counted content on FT.com and Markets Data, through re-performance of navigation to different counted FT.com and Markets Data pages.	PwC tested a sample of no less than 45 user actions which triggered 130 events of PwC test usage. All pages were counted and recorded accurately.
3	Test the accuracy of the recorded number of counted pages in Redshift for sample users accessing counted content on FT mobile app, through re-performance of navigation to different counted FT mobile app pages. The iOS and Android version of the mobile app were tested independently.	PwC tested a sample of no less than 25 user actions which triggered 60 events of PwC test usage. All pages were counted and recorded accurately.
4	Test the accuracy of the recorded number of counted pages in Redshift for sample users accessing counted content on ePaper, through re-performance of navigation to different counted FT ePaper pages.	PwC tested a sample of no less than 45 user actions which triggered 234 events of PwC test usage. All pages were counted and recorded accurately.
5	Test the accuracy of the recorded number of counted pages in Redshift for sample users accessing counted content via email-distributed newsletters, through re-performance of navigation to different counted email-distributed newsletters.	PwC tested a sample of no less than 25 user actions which triggered 32 events of PwC test usage. All pages were counted and recorded accurately.

Detailed findings

Testing performed

	Summary of procedure	Results
6	Review the GitHub repository for changes related to how counted content was measured during the period.	Changes noted were documented and inline with expected routine upgrades to the code, no changes fundamentally altered how counted content was measured per FT's methodology.
7	Review user access controls to the Redshift database to ensure access rights were restricted to appropriate personnel.	Access controls to Redshift were restricted to personnel within the data, customer, editorial, products, and database admin teams. Only users in the database admin team had write privileges to Redshift.
8	Review process for audit logging of database queries and data modification in the Redshift database.	Audit logging practices were confirmed to be in place and audit logs retained for six months (a significant improvement on prior year).
9	Review the process for database backup and remediation in case of outages or data loss.	Backups of the database are saved hourly in Amazon Web Services and retained for a period of 30 days and audit logging is tracked for the previous six months in Splunk.
10	Assess the accuracy of calculation and reporting of Metric V and Active Core Readers in Enterprise Tools through reconciliation with underlying content usage data recorded in Redshift for a sample of 35 users on 7 B2B contracts.	Calculation and reporting of Metric V in Enterprise Tools was reconciled to the underlying content usage data for 30 out of 35 tested users. Two exceptions were identified leading to overcounting of content affecting a total of 5 of users. These exceptions did not cause a difference in the Active Core Reader metric for each user which reconciled for all 35 users. See page 16 for review of the exceptions and management comment.

Detailed findings

Exceptions summary for procedure 10

Users affected	Description of Exception	Management comment
3	Overcounting of counted content for three users across two different B2B contracts was identified. Users content was over counted by: - A user with 322 events of qualifying content was overcounted by 2 events - A user with 1526 events of qualifying content was overcounted by 9 events - A user with 179 events of qualifying content was overcounted by 1 event The cause for the overcounting was a website change on June 10th, 2019 to the paywall barriers that prevented the barrier type from being added to the page view events. Thereby allowing content to be counted that should have been outside of a paywall barrier and therefore not eligible for counting. PwC identified this discrepancy due to a difference between the data in Redshift and Enterprise Tools for these users, with overcounting displaying in Enterprise Tools. Following FT discovery of this on June 28th, 2019, there was a manual correction of the barrier attribute on the page view table on July 1st, 2019. The event data was recovered from separate barrier view events that are also triggered when a user views content, and then used to updated the underlying data in Redshift on August 1st, 2019.	We are grateful to PWC for carrying out this data assurance audit for the third year. The quality and integrity of our subscription data is a vital part of our organisation. PWC identified two data quality issues in this report. These issues have been fixed and the historical data corrected in the underlying database (Redshift) and Enterprise tools. The consequence of these errors are that 36 additional B2B readers were charged for across 29 corporate customers. This is less than 0.01% of all B2B subscribers. The 29 corporate customers have been contacted by the FT. We are grateful for the support of all our customers and we remain committed to ensuring a high level of data quality
2	Overcounting of counted content for two users across one B2B contract was identified. Users content was over counted by: - A user with 133 events of qualifying content was overcounted by 10 events - A user with 2627 events of qualifying content was overcounted by 1 event A code error was identified in the counting of email usage that had been active since 2017 when email usage tracking was first introduced. The error is triggered if a user has multiple email opens with same timestamp and different event ids. A code fix was deployed by FT on March 6th, 2020 to restrict this double counting of email opens.	