

Identifying and Mapping UK Digital Content Streaming Platform Firms. A Machine Learning Analysis.

Pei-Yu Yuan, Raquel Ortega-Argiles, Xiaoxiao Yu and
Stephen Roper

May 2026

1. Purpose and scope

This annex documents the identification protocol used to construct an empirical dataset of UK-based digital content streaming businesses, and to classify a subset of these as platform owners, representatives, and related organisational forms. It is designed as a methodological note and a structured mapping output, rather than an evaluation of impacts.

The annex focuses on how candidate firms were surfaced using The Data City's machine learning (ML) list-building tool, how the outputs were validated and refined through manual review and supplementary checks, and how the final set of firms was categorised. The annex does not attempt to estimate the economic impacts of streaming platforms, nor does it assess value retention in the UK.

1.1 Project context and initial scoping

At the start of the project, UKRI provided an initial list of exemplar platform firms. The list was used as scoping input to support early internal discussions on the breadth of the platform economy and the feasibility of sector-specific mapping. The exemplar firms were grouped into six broad categories.

Table 1 UKRI exemplar list

Category	Firm
Online marketplaces	eBay; Etsy
Transportation	Uber; Bolt
Food delivery	Deliveroo; Just Eat
Digital content streaming service	Spotify; Amazon Video
Social media and networking	Twitter (X); Instagram
Accommodation and travel	AirBnB

The UKRI exemplar list covered several platform segments. For the purposes of this technical annex, one segment was selected to document the identification and classification protocol in a structured way. The segment examined here is digital content streaming services. The selection should be understood as a scoping decision for methodological transparency rather than as an assessment of relative importance across platform types. The same identification principles could, in principle, be applied to other segments.

2. Data source and identification strategy

2.1 Data source

This mapping exercise uses company-level data from The Data City, a UK-based data-as-a-service provider that offers extensive coverage of UK companies. The Data City tracks how companies describe their activities online, which is particularly relevant for identifying firms in fast-moving sectors where conventional industrial classifications may be less informative.

2.2 Machine Learning list-building workflow

We use The Data City's machine learning list-building tool to generate a tailored dataset of UK digital content streaming businesses. The tool trains an underlying model to classify firms based on website content. The process began by feeding the model a set of known examples of streaming companies, covering video, audio and live streaming, together with negative examples to help the model distinguish non-relevant firms.

Next, keyword filters such as “on-demand service,” “digital marketplace,” and “streaming platform” were applied to surface additional candidate firms. The resulting list was reviewed and refined through multiple iterations, with adjustments made to the training inputs and retraining the model as needed. The overall workflow is summarised below.

An important boundary condition is that the training and filtering process does not include platforms that are primarily regarded as other types of platforms even if they have streaming features (for example, social media and networking platforms).

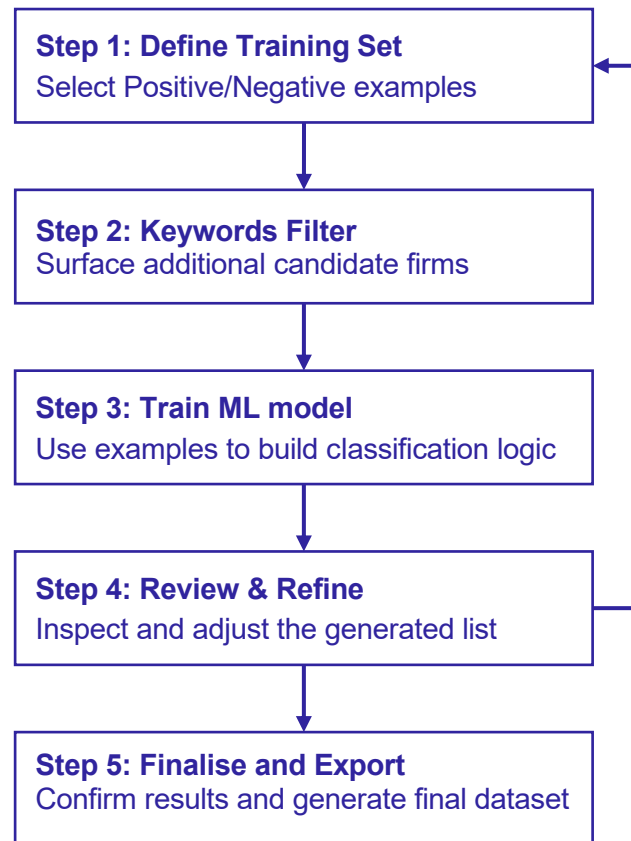


Figure A1 Machine Learning Process to generate the dataset. Own elaboration.

3. Classification and validation

3.1 Initial model output and keyword inspection

The final machine learning model generated a ranked list of 743 UK-based companies based on how closely they matched the profile of known digital content streaming platforms. This ranked output formed the starting point for the validation and selection process reported in this annex.

To assess model performance and interpret why firms were flagged as relevant, we examined the keywords used in classification. High-scoring terms such as “streaming”, “subscription”, “live”, and “platform”, alongside familiar names such as “Twitch”, frequently appeared on the websites of known streaming platforms. Table A2 reports the top positive and negative indicators identified by the model, together with their scores.

Table A2 Model-Derived Indicator Terms for the Digital Content Streaming Classification

Top 20 Positive Indicators	Score	Top 20 Negative Indicators	Score
streaming	11.13	social media	-2.59
subscription	8.43	sat	-2.6
stv	6.82	marketing	-2.62
twitch	6.22	sport	-2.66
platform	5.8	training	-2.73
universal	5.56	cheese	-2.81
sky	5.44	night	-2.89
live	5.08	ynab	-2.89
help	4.86	contact	-3.01
content	4.82	data	-3.04
stream	4.6	life	-3.06
gift	4.43	project	-3.42
watch	4.41	events	-3.56
failed	4.35	cart	-3.59
device	4.2	services	-3.64
karaoke	4.02	make	-3.65
browser	3.93	team	-3.89
netflix	3.87	course	-4.07
alive	3.82	design	-4.24
supported	3.76	view	-4.81

3.2 Review and cleaning

The model output reflects common limitations of text-based classification using online self-descriptions. Firms may adopt platform-related terminology without operating a platform business, and some results may not correspond to unique, verifiable companies. For this reason, we conducted a two-stage review to ensure basic verifiability and to align the sample with the operational boundary used in this annex.

First, we applied mechanical checks to remove clear mismatches. We excluded cases with duplicate or incorrect website links, ensuring each entry corresponded to a unique, verifiable company. In total, 189 firms were excluded at this stage. We also removed cases that represented creator pages hosted on larger platforms (for example, individual pages on SoundCloud or Vimeo), rather than independent firms operating streaming services. A further 68 cases were excluded on this basis. Following these exclusions, the remaining firms formed the pool for detailed review.

Second, we conducted a substantive review of the remaining firms by examining the retrieved websites and the nature of the services offered. This stage focused on confirming relevance to the segment boundary and resolving borderline cases. Where necessary, we drew on Companies House, Orbis, and other publicly available sources to verify company status, principal activities, and ownership structure. Following this extended validation, the number of relevant companies was narrowed to 94. These include UK-based digital content streaming businesses and UK-based parent companies of streaming businesses. The definitional boundary and the validation criteria used to make these inclusion decisions are set out in Sections A3.3 and A3.4.

3.3 Definitional boundary: platforms versus streaming “platforms”

In common usage, many streaming services are described as “platforms”. In this annex, **platform** is used as an operational classification label with a specific meaning. More specifically, the annex is concerned with streaming services that operate as platforms, rather than streaming services in general. This boundary is defined to ensure that subsequent validation decisions are interpreted consistently.

We therefore distinguish between streaming services that operate as platforms and those that function primarily as content distributors. Services in the latter category typically deliver licensed or produced-in-house content to viewers but show little to no observable evidence that they mediate interactions between distinct participant groups. Such services fall outside the operational boundary applied in this annex.

3.4 Validation criteria and final inclusion

The validation criteria in this section operationalise the boundary described in Section A3.3 and were applied to confirm whether each remaining case fell within the operational scope of this annex. In applying these criteria, we aimed to ensure that each included case corresponds to a distinct, verifiable company and that the firm’s primary activity aligns with the segment definition of UK-based digital content streaming services.

Inclusion criteria

A case was retained where the available evidence indicated that:

- It corresponds to a unique, verifiable company with an independent web presence (rather than a hosted creator profile or a non-firm page on a larger platform).
- The firm is UK-based, or a UK-based parent company of a streaming business.
- The core service involves digital content streaming, and there is observable evidence that the service operates as a platform by mediating interactions between distinct participant groups (for example, creators and audiences, or multiple market sides).

- The website and related public sources provide sufficient information to validate the firm's status and principal activities.

Exclusion criteria

A case was excluded where evidence suggested that:

- The entry could not be linked to a distinct, verifiable company (for example, due to missing, incorrect, or duplicate website links).
- The page represents an individual creator profile hosted on a larger service (for example, a channel or user page on SoundCloud, Vimeo, or similar services) rather than an independent firm.
- The service functions primarily as a content distributor, delivering licensed or produced-in-house content to viewers without observable evidence of intermediation between distinct participant groups.
- The firm's principal activity is outside the scope of digital content streaming services, or the relevance is only incidental.

Where necessary, we drew on Companies House, Orbis, and other publicly available sources to verify company status, principal activities, and ownership structure. Applying these criteria yielded a final sample of 94 relevant companies, comprising UK-based digital content streaming businesses and UK-based parent companies of streaming businesses.

3.5 Composition of the final dataset

Applying the validation process described in Sections A3.2 to A3.4 yielded a final dataset of 94 relevant companies. These include UK-based streaming businesses, UK-based parent companies of streaming businesses, UK representatives of international services, and a small number of related organisational entities identified through ownership and service relationships.

The final dataset is therefore not limited to firms that directly own and operate streaming services. It also includes parent companies, UK subsidiaries or representatives of international services, service providers distinct from the owner, and one UK establishment of an international platform owner. This structure reflects the organisational diversity observed in the review process and the need to map how streaming activities are represented in the UK company landscape.

Table A3 summarises the composition of the final dataset across these organisational categories.

Group	Category	Description and key features	Frequency	Example
Streaming service owners	Platform owner	A company that directly owns and operates one or more streaming platforms. In these cases, the available evidence indicates that the service mediates interactions between distinct participant groups , rather than functioning solely as a distributor of licensed or in-house content.	13	MIXCLOUD LTD (Mixcloud)
	“Platform” owner	A company that owns and operates a streaming service commonly described as a “platform”. However, the available evidence suggests that the service primarily distributes licensed or produced content to viewers, with little observable evidence that it mediates interactions between distinct participant groups .	45	THE COMMUNITY CHANNEL (Together TV)
UK operating entities	“Platform” service provider	A company that operates or provides the streaming service but is distinct from the entity identified as the owner . These cases capture situations where the service brand and the operating company are not the same legal entity.	3	TOGETHER TV LTD (Together TV)
	UK representative (subsidiary) of an international platform	A UK-based subsidiary or representative company that represents the foreign owner of a streaming platform in relation to its UK activities. These cases indicate that the platform is foreign-owned but has an identifiable UK corporate presence.	7	TWITCH UK LTD (Twitch)
	UK representative (subsidiary) of an international “platform”	A UK-based subsidiary or representative company that represents the foreign owner of a streaming service commonly described as a “platform”. As above, the service itself shows limited observable evidence that it mediates interactions between distinct participant groups .	7	STINGRAY DIGITAL INTERNATIONAL LTD (Stingray Music)

Group	Category	Description and key features	Frequency	Example
UK parent and related entities	UK parent company of a “platform” owner	A UK parent company that sits above a “platform” owner in the ownership structure . These cases were included to reflect the corporate chain through which ownership and control of streaming services can be traced.	14	UNIVERSAL STUDIOS LTD (Hayu)
	UK parent company of a representative of an international platform	A UK parent company sitting above the UK representative of an international platform. These cases reflect additional layers in the UK corporate ownership structure linked to platform activities.	1	SONY INTERACTIVE ENTERTAINMENT EUROPE LTD (PlayStation Plus Cloud)
	UK parent company of a representative of an international “platform”	A UK parent company sitting above the UK representative of an international streaming service commonly described as a “platform”. This category captures further ownership links within the UK corporate structure .	3	PARAMOUNT PICTURES INTERNATIONAL LTD (Paramount+)
	UK establishment of an international platform owner	A UK establishment or branch of an international platform-owning entity identified during the review. Unlike subsidiaries, these cases reflect a UK operational presence without a separately incorporated UK company .	1	DEEZER S.A. (Deezer)

Note: The grouping shown here is an analytical presentation of the final sample composition and is consistent with the report’s classification of the 94 identified cases. The categories shown here refer to different organisational roles represented in the UK final dataset. In this annex, *platform* refers to streaming services for which there is observable evidence that the service mediates interactions between distinct participant groups. “*Platform*” refers to streaming services commonly described as platforms but primarily functioning as distributors of licensed or produced content.

4. Descriptive profile of the final dataset

This section presents a descriptive profile of the 94 companies included in the final validated dataset. It summarises the main characteristics of these cases, including their geographic distribution, firm age, organisational size, and ownership structure.

4.1 Geographic distribution

Table A4 Geographic distribution of companies in the final dataset

Region	Extended Validation	
	Freq.	Percent
London	69	75.82
South East (England)	7	7.69
South West (England)	2	2.20
West Midlands (England)	3	3.30
East Midlands (England)	2	2.20
Scotland	1	1.10
North West (England)	2	2.20
Yorkshire and The Humber	1	1.10
East (England)	2	2.20
North East (England)	1	1.10
Wales	1	1.10
Northern Ireland	0	75.82
Total	91	100

4.2 Firm age

Table A5 Age distribution of companies in the final dataset

Firm Age	Extended Validation	
	Freq.	Percent
1	4	4.3
2	5	5.38
3	2	2.15
4	3	3.23
5	6	6.45
6	9	9.68
7	5	5.38
8	4	4.3
9	4	4.3

Firm Age	Extended Validation	
	Freq.	Percent
10	6	6.45
11 - 20	24	25.81
21 - 30	12	12.90
> 30	9	9.68
Total	93	100

4.3 Organisational size

Table A6 Firm size by employment band

Employee	Extended Validation	
	Freq.	Percent
No employee	18	19.35
Micro	37	39.78
Small	9	9.68
Medium	13	13.98
Large	16	17.20
Total	93	100.00

Note: Company size categories are based on the variable of best estimate current employees, with firms classified as follows: no employees (0), micro (1–9), small (10–49), medium (50–249), and large (250 or more).

Table A7 Firm size by turnover band

Turnover	Extended Validation	
	Freq.	Percent
No turnover	16	17.20
Micro	38	40.86
Small	11	11.83
Medium	2	2.15
Large	26	27.96
Total	93	100.00

Note: Companies are classified as follows: micro (turnover of £1 million or less), small (turnover between £1 million and £15 million), medium (turnover between £15 million and £54 million), and large (turnover exceeding £54 million).

4.4 Ownership structure

Table A8 Parent company nation

Parent Company Nation	Extended Validation	
	Freq.	Percent
No Parent company	35	37.63
UK	36	38.71
Unknown	4	4.30
USA	6	6.45
Netherlands	1	1.08
Denmark	1	1.08
Finland	1	1.08
Luxembourg	2	2.15
St. Vincent	1	1.08
Sweden	1	1.08
Barbados	1	1.08
Jersey	1	1.08
Total	93	100.00

Table A9 Ultimate parent company nation

Ultimate Parent Company Nation	Extended Validation	
	Freq.	Percent
No Parent company	35	37.63
UK	25	26.88
USA	19	20.43
Unknown	5	5.38
Canada	1	1.08
Japan	2	2.15
St. Vincent	2	2.15
Denmark	1	1.08
Finland	1	1.08
Jersey	1	1.08
Luxembourg	1	1.08
Total	93	100.00

4.5 UK streaming platform businesses

Within the final dataset, a smaller subset of firms corresponds to companies classified as platform owners or UK representatives of international streaming platforms. The identified platform owners consist primarily of independent UK businesses, several of which are registered in London and operate with relatively small organisational teams.

In contrast, the UK representatives of international streaming platforms are typically subsidiaries of larger multinational organisations. These entities tend to form part of broader corporate structures and are ultimately owned by parent companies registered outside the United Kingdom.

Together, these cases illustrate the range of organisational arrangements through which streaming-related platform activities are represented in the UK company landscape.

5. Measurement considerations

The approach documented in this annex provides a structured way to identify and profile companies associated with digital content streaming services using web-derived firm descriptions, manual validation, and supplementary checks. It supports a systematic mapping of relevant companies in the UK company landscape, but several interpretation boundaries should be noted.

First, the identification process relies primarily on publicly available online information. Firms differ in how clearly and how fully they describe their activities on company websites and related sources, and some may not maintain detailed or up-to-date online profiles. This means that some streaming-related activities may not be visible through web signals alone.

Second, streaming services may be represented through complex organisational structures. Some international services appear in the UK through subsidiaries, representatives, parent companies, or other related entities rather than through the ultimate platform owner itself, while others may operate in the UK market without a clearly identifiable UK corporate entity. The final dataset should therefore be interpreted as a structured mapping of companies that show observable involvement in digital content streaming services, rather than as a complete census of all streaming services operating in the UK.

Authors

The core members of the research team for this project were as follows:

- o Dr Pei-Yu Yuan – University of Manchester
- o Prof Raquel Ortega-Argiles – University of Manchester
- o Dr Xiaoxiao Yu – University of Manchester
- o Prof Stephen Roper – Warwick Business School

This document relates to IRC Project IRCP0022: Mapping the unknowns: understanding digital platforms and the search for UK evidence

Acknowledgements

This work was supported by ESRC grant ES/X010759/1 to the Innovation and Research Caucus and was commissioned by Innovate UK. We are very grateful to the project sponsors at UKRI for their input into this research. The interpretations and opinions within this report are those of the authors and may not reflect the policy positions of Innovate UK.

About the Innovation and Research Caucus

The Innovation and Research Caucus supports the use of robust evidence and insights in UKRI's strategies and investments, as well as undertaking a co-produced programme of research. Our members are leading academics from across the social sciences, other disciplines, and sectors who are engaged in various aspects of innovation and research systems. We connect academic experts, UKRI, Innovate UK and the Economic and Social Research Council (ESRC), by providing research insights to inform policy and practice. Professor Tim Vorley and Professor Stephen Roper are Co-Directors. The Innovation and Research Caucus is funded by UKRI via the Economic and Social Research Council (ESRC) and Innovate UK, grant number ES/X010759/1. The funders' support is acknowledged. The views expressed in this piece are those of the authors and do not necessarily represent those of the funders.

Find out more

Contact: info@ircaucus.ac.uk

Website: <https://ircaucus.ac.uk/>

Now that you have read our report, we would love to know if our research has provided you with new insights, improved your processes, or inspired innovative solutions.

Please let us know how our research is making a difference by completing our short feedback form [via this link](#).

You are also welcome to email us if you have any questions about this report or the work of the IRC generally: info@ircaucus.ac.uk

Thank you

The Innovation & Research Caucus