The Greek Online Media Dataset (2024)

Annotated corpus of article titles in 106 Greek websites between 01/01/2024-31/12/2024.

YANI KARTALIS

Center for Social Data Science (SODAS), University of Copenhagen
kartalisie@gmail.com

Data Collection/Sources

The data in this report has been collected in the context of the project titled “From Ostensible to Actual Media Pluralism. An examination of content diversity in Greece's fragmented online media system”. The research was possible with funding by the A.C. Laskaridis Charitable Foundation and Dr Vassili G. Apostolopoulos at the Hellenic Observatory of the London School of Economics and Political Sciences.

The dataset contains 3,301,004 articles titles from 106 websites. All data has been collected via public access to each website’s newsroom tab which allowed for historical access through the year of 2024. Figure 1 shows the publication rates of articles for all websites pooled and in 7-day rolling averages. Table 1, contains article title counts for all websites included in the dataset. Below you can find a description of each variable included in the dataset.

Variables

custom_id: A unique ID variable counting articles per website.

URL: A string variable containing the article’s url address

Title: A string variable containing the article’s title.

Website: A string variable containing the website hosting the article.

Date: A datetime variable containing the article’s publishing date. This is almost always the ‘last-updated’ date.

Category: A string variable containing the article’s news category if available. These categories come as designated by the website’s internal category structure and are not unifrom across websites.

Summary: A string variable containing a summary of the article’s content. Not availble in all websites.

Author: A string variable containing the article’s author as designated by each website.

Body: A string variable containing the article’s main body. This text is available for a mere 5% of articles.

Figure 1. 7-Day Rolling Average of Articles Per Day (01.01.2024-31.12.2024)

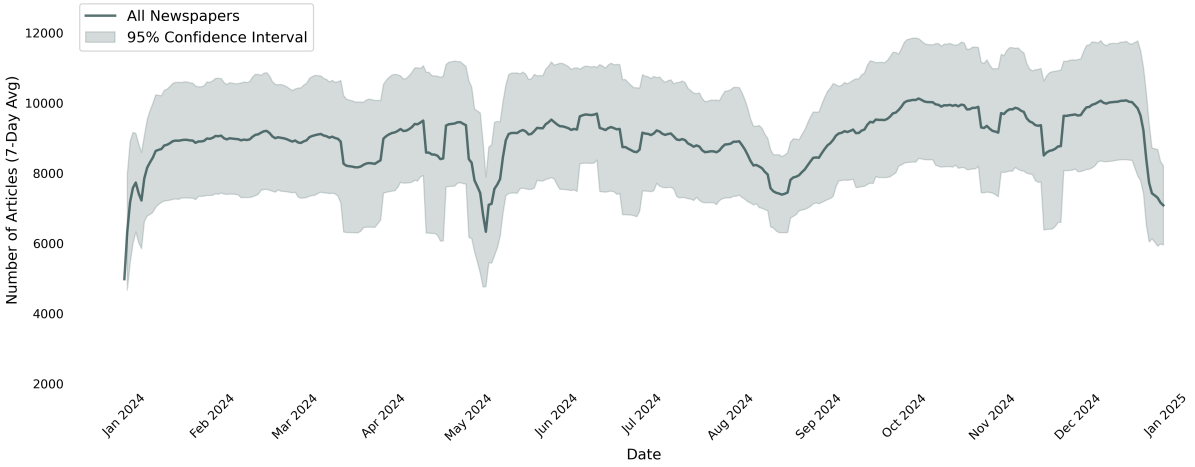

Table 1. Total Article Counts per Website

1	iefimerida	94753	**28**	news247	42626	**55**	megatv	25913	**82**	iellada	10028
2	in	90742	**29**	ieidiseis	41597	**56**	alphatv	25893	**83**	1077statusfm	9971
3	sdna	89740	**30**	powergame	41493	**57**	paraskhnio	25287	**84**	fimotro	9842
4	newsbeast	85107	**31**	vradini	41019	**58**	grtimes	23753	**85**	popaganda	9778
5	newsit	81950	**32**	thetoc	40091	**59**	moneyreview	23169	**86**	afteroffice	8770
6	politic	80729	**33**	efsyn	39863	**60**	makeleio	22368	**87**	thepressproject	8596
7	capital	75675	**34**	insider	39740	**61**	businessnews	21840	**88**	neolaia	7861
8	zougla	71273	**35**	liberal	38973	**62**	avgi	21468	**89**	dealnews	7727
9	sport-fm	70637	**36**	reader	38686	**63**	rp	21095	**90**	moneypress	7685
10	pentapostagma	70448	**37**	thepresident	37802	**64**	myportal	20688	**91**	koutipandoras	6827
11	ertnews	70341	**38**	athensvoice	36434	**65**	govnews	19963	**92**	parapolitika	6720
12	enikos	69515	**39**	newsbreak	35380	**66**	reporter	19855	**93**	agrocapital	5718
13	flash	62095	**40**	ot	35208	**67**	imerisia	17244	**94**	newstok	5644
14	skai	58938	**41**	hellasjournal	35199	**68**	todaypress	17110	**95**	fpress	5586
15	voria	57769	**42**	theopinion	34112	**69**	parallaximag	16899	**96**	madamefigaro	5497
16	tanea	56619	**43**	monobala	33678	**70**	fortunegreece	16570	**97**	ratpack	4494
17	eleftherostypos	56307	**44**	documentonews	33572	**71**	protagon	16369	**98**	dimokratia	3281
18	naftemporiki	56042	**45**	tovima	33159	**72**	rosa	15887	**99**	ecopress	2086
19	kathimerini	55647	**46**	iapopsi	32279	**73**	tvxs	15686	**100**	apotis4stis5	2027
20	sportdog	51766	**47**	lykavitos	30553	**74**	sofokleousin	14896	**101**	onmed	1930
21	sportal	50231	**48**	ereportaz	29475	**75**	ekathimerini	13680	**102**	ikivotos	1644
22	newpost	49540	**49**	lifo	29054	**76**	missbloom	13491	**103**	adesmeuti	935
23	dnews	48411	**50**	aftodioikisi	28674	**77**	media	13289	**104**	ota365	933
24	sport24	48382	**51**	businessdaily	28170	**78**	economistas	12858	**105**	paron	876
25	karfitsa	46934	**52**	bankingnews	27950	**79**	onsports	12751	**106**	estianews	360
26	mononews	43989	**53**	star	27277	**80**	euronews	11922			
27	newsbomb	43535	**54**	euro2day	26203	**81**	flight	10862			