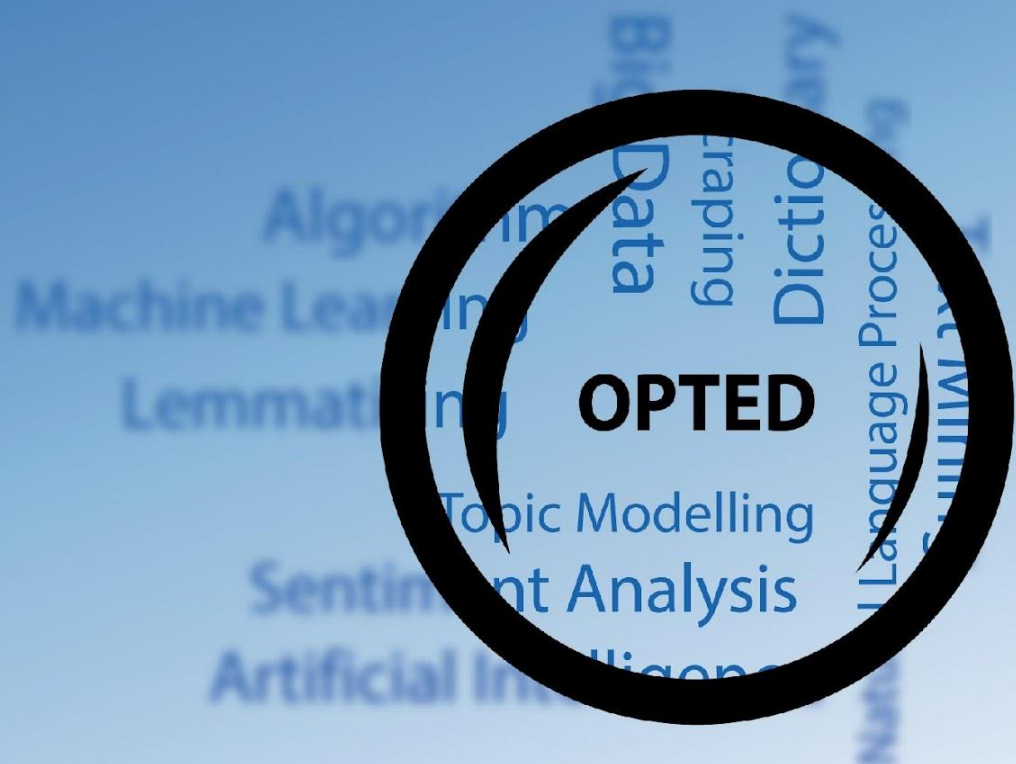


# OPTED

## Registry of Data Sources

Paul Balluff, Fabienne Lind, Hajo G. Boomgaarden & Annie Waldherr



## **Disclaimer**

This project has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement No 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

## **Dissemination level**

Public

## **Type**

Report



## **OPTED**

Observatory for Political Texts in European Democracies:  
A European research infrastructure

# **Registry of Data Sources**

## **Deliverable 3.2**

**Authors: Paul Balluff, Fabienne Lind,  
Hajo G. Boomgaarden & Annie Waldherr<sup>1</sup>**

<sup>1</sup> University of Vienna, Department of Communication

**Due date:** September 2022

## Executive Summary

Increasingly complex and hybrid media landscapes are putting high demands on researchers aiming to provide comprehensive and/or comparative analyses of news media and their coverage. This relates to the identification and classification of media sources and their availability as well as to the identification of archived or annotated media contents. Responding to these challenges of a very rich yet highly fragmented hybrid media landscape this deliverable introduces the implementation of D3.1: the OPTED Media Text Analysis Open Registry (Meteor). *Meteor* is a newly developed registry for European news sources, data archives, datasets, corpora, and text analysis tools. The registry is implemented as a web application in which users can not only query entries, but also make contributions to the registry, keeping it dynamic and adaptable. It is publicly available at <https://meteor.opted.eu/>. The registry aims to benefit research projects using methods that require content analyses of news texts, or meta-information about news sources and media organisations.

*Meteor* distinguishes between various entity types, including news sources, countries, organisations, data archives, corpora, datasets, and tools. All these entity types can be linked with each other. *Meteor* does contain a comprehensive list of news sources across European countries and in addition includes a set of meta information or properties per news source, potentially relevant to subfields of political communication and journalism studies, including media organisations behind the news sources. Furthermore, data archives, digital storages for full-text media text data, static text corpora sometimes containing annotations and news content analysis datasets typically containing only annotations are included in the registry. As of now, *Meteor* shows over 2,200 news sources from 17 countries, as well as almost 600 organisations, and 19 full text data archives. Meteor will help researchers to rapidly find (comparable) data sources, background information about news sources, links to related full text archives, as well as annotated media data sets if available. This will facilitate the selection, access and work not only with sources but also with textual data. Due to the flexible structure of the database, additional meta-information can be added upon request.

## 1 Introduction

News texts are nowadays produced by diverse groups of actors such as traditional media organisations, alternative journalism initiatives, or journalistic blogs. Traditional means for distributing news texts are complemented and partially replaced by information and communication technologies such as e-papers, online newspapers, social media, apps, and streaming platforms (Mattoni & Ceccobelli, 2018). The rise of digital media has caused a fragmentation of the media landscape and simultaneously an integration of “newer” and “older” media logics (Chadwick, 2013: 15; Klinger & Swanson, 2015; Mattoni & Ceccobelli, 2018: 541). With the digitalization of media, the interactions between audience and newsmakers are changing, and the relationship between politics and media is shifting as well (Deuze & Witschge, 2018; Ryfe, 2018). Relatedly, some scholars observe the boundaries of journalism to become increasingly “fuzzy” (Maares & Hanusch, 2020). Valuable insights into the hybrid media logics have been generated via surveys of media workers and audiences, as well as indicators at the system level (e.g., Brüggemann et al., 2014; Hallin & Mancini, 2017; Hanitzsch et al., 2019; Mattoni & Ceccobelli, 2018). Yet, comprehensive assessments of hybrid media systems and its dynamics should also be based on content characteristics, yet this is rather rarely done (e.g., Popescu et al., 2010).

However, investigating media landscapes based on content characteristics would require a encompassing registry of available data sources for journalistic texts. On the one hand, information and communication technologies enable to retrieve journalistic texts from a variety of data sources. On the other hand, it also poses a challenge to maintain a structured overview of available data. Hence, **we argue that it is beneficial to establish a comprehensive registry of news sources along a broad set of meta-information as well as their distribution channels.**

There are several good reasons for collecting news sources and related meta-information in a new registry. First, **journalism and the global media landscape are in constant transformation** (Deuze & Witschge, 2018; Lewis, 2012; Mythen, 2010). The same can be said for political communication (Negrine, 2008). Moreover, alternative media are increasingly taking up space in the media landscape. With an encompassing registry of news sources, we put scholars in a better position to study these transformations.

Second, a new registry is useful from a research practice perspective as it seeks to overcome the disadvantages of existing inventories, such as LexisNexis<sup>1</sup>, Factiva<sup>2</sup> or the Europe Media Monitor (Steinberger, 2013). What these lists have in common is that **they are driven by data availability and not by source availability** and that they do not include meta-information per source but are (mere) lists of source names. However, for research practice it is important to base the design on rich meta-information such as the average audience size of a news source. Wikipedia is another example of an existing registry: it provides several pages that list news sources for different countries<sup>3</sup>. While Wikipedia's lists are already useful for gaining an overview, they do not provide a standardised set of required meta-information and get sparse on the subnational level. Digital media channels (e.g., social media, or messengers) are also not part of Wikipedia lists. Both the source lists by media data archives and Wikipedia are not necessarily curated and reviewed by the research community, who are the experts for and targeted main users of *Meteor*. An example for an expert-curated list of media sources in Europe is euro|topics<sup>4</sup>. However, their collection only covers news sources with the largest audiences per country and also does not include alternative media, which have become an important boundary phenomenon in today's media systems.

Third, a **registry helps balance sample selection biases**. The available choices in archives typically depend on agreements of the archive proprietor and publishing houses. Not all publishers choose to make their content available, which limits the coverage of text archives. While the growing numbers of available data sources and text archives are welcomed, they lack organisation, curation, and a systematic overview (e.g., Esser & Vliegthart, 2017, p. 11).

Last but not least, country-comparative research projects will benefit from a registry. In fact, comparative communication scholars call for classifications and typologies to identify similarities and differences (Esser & Hanitzsch, 2012) and regret that there is no reliable international catalogue where one can look up media sources (Esser & Vliegthart, 2017: 11).

Responding to these empirical and analytical challenges of a very rich yet highly fragmented hybrid media landscape (Neumann, 2016; Prior, 2007), this deliverable introduces the implementation of D3.1: the OPTED Media Text Analysis Open Registry (*Meteor*). *Meteor* is a new inventory for European news sources, data archives, datasets, corpora, and text analysis tools (see D3.3).

The registry is implemented as a **web application** (see Figure 1) in which users can not only query entries, but also make contributions to the registry, keeping it dynamic and adaptable. The repository aims to benefit research projects using methods that require content analyses of news texts, or meta-information about news sources and media organisations. Therefore, it also includes a curated collection of available full-text archives and access to secondary data. *Meteor* shows which news sources are included in large archives, which in turn helps researchers to choose their sample based on other criteria than availability of data. **In sum, *Meteor* is tailored to the needs of researchers and will assist scholars to make well-informed research decisions.**

In the following sections, we outline the scope of *Meteor*, and introduce its implementation which can be accessed at <https://meteor.opted.eu/>. We then provide a summary of the data collection process as well as summary statistics of *Meteor*.

---

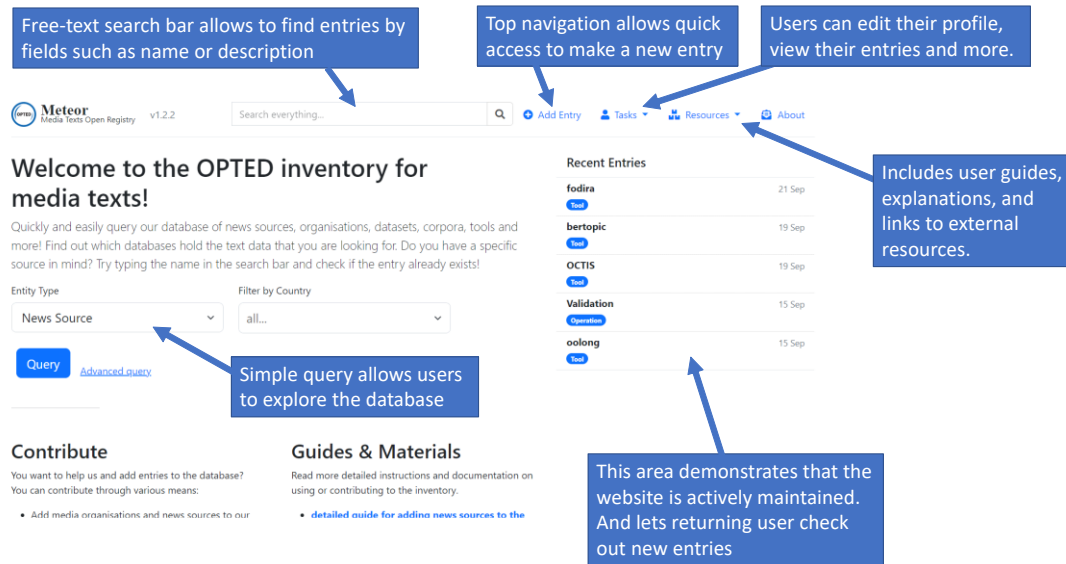
<sup>1</sup> <https://www.lexisnexis.com>

<sup>2</sup> <https://professional.dowjones.com/factiva/>

<sup>3</sup> E.g., for Ireland see [https://en.wikipedia.org/wiki/List\\_of\\_newspapers\\_in\\_the\\_Republic\\_of\\_Ireland](https://en.wikipedia.org/wiki/List_of_newspapers_in_the_Republic_of_Ireland)

<sup>4</sup> <https://www.eurotopics.net/>

**Figure 1:** Landing page of *Meteor* with user interface explanations.



## 2 Meteor’s Structure and Scope

As mentioned, “journalism” is blurry, ambiguous, and context dependent, yet there are ways which can account for fuzzy and inter-related phenomena and to “deal with the mess (we made)” (Witschge et al., 2018: 651). Therefore, **we implemented a knowledge graph structure for *Meteor***<sup>5</sup>. A knowledge graph is “a large semantic net” (Fensel et al., 2020: 6) with a nested and interlinked structure.

Based on the literature reviewed in D3.1, *Meteor* distinguishes between various entity types, including news sources, countries, organisations, data archives, corpora, datasets, and tools. All these entity types can be linked with each other with a subject-predicate-object relationship. For example:

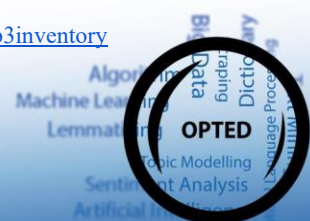
- “The Guardian Media Group” (organisation) publishes “The Guardian” (news source).
- “The Guardian Media Group” is located in “UK” (country).
- “The Guardian” is distributed as a “Print Publication” (channel).

This also allows linking available data sets to news sources, which enables researchers to query *Meteor* based on readily available secondary data, archives or corpora (e.g. the *European Parliament Election Study 2009* (Schuck et al, 2010) investigated *The Guardian*).

**News sources** are the foundation of *Meteor*. As described in D3.1, they have the **primary purpose to regularly report and comment on recent events and ideas** (see also Anderson, 2012). News sources are operated by collective actors (i.e., organisations) with a standardised set of practices (e.g., editorial control) and aim to maximise their reach within a defined target audience in order to maintain a habitual and repeated exposure (Potter, 2013). They produce output that claims to be non-fictional and the audience also generally accepts this claim (Jensen, 2002).

Within this general scope *Meteor* includes a wide range of news sources that are shown in Table 1 alongside an example. The registry also includes news blogs as long as they are published by a group of people (i.e., different authors), and appear to be created with editorial support. Finally, alternative news media sources that propagate that they provide an alternative view in contrast to mainstream media are also part of the registry. By highlighting the news producing activity of a source, we exclude news aggregation services. We also exclude other forms of news

<sup>5</sup> The development of *Meteor* is open-source and can be accessed at: <https://github.com/opted-eu/wp3inventory>



such as blogs by individuals (e.g., citizen journalism), or news by political actors, NGOs, think tanks, or corporations. More specifications are listed in the contribution guidelines in the Appendix.

A last general clarification relates to the geographical and temporal scope of *Meteor*. The aim is to cover all EU member states as well as the UK, Switzerland, Norway, and Israel. In addition, the classification should be suitable for sources actively published today (starting from 2021).

**Table 1:** Summary of included and excluded news sources in *Meteor*

<i>Included</i>	<i>Example</i>
Print news sources	<i>The Guardian</i>
Online news sources	<a href="http://www.theguardian.com">www.theguardian.com</a>
News agencies	AFP
Social media accounts distributing news	@guardian (Twitter account)
Transcripts for self-created audio-visual contents	Transcripts of <i>Zeit im Bild</i> (TV show) available at the APA-Archive ( <a href="http://aomlibrary.apa.at">aomlibrary.apa.at</a> )
News sources for multinational audiences	<a href="http://www.euronews.com">www.euronews.com</a>
News sources for national audiences	<a href="http://www.derstandard.at">www.derstandard.at</a>
News sources for subnational audiences	<a href="http://www.standard.co.uk">www.standard.co.uk</a>
News blogs published by collective actors and editorial support	<a href="http://www.globalvoices.org">www.globalvoices.org</a>
Alternative news sources	<a href="http://www.order-order.com">www.order-order.com</a>
<i>Excluded</i>	<i>Example</i>
News aggregators	Google news
Blogs by individuals	Individual journalists, individual citizens
News by NGOs	<a href="http://www.transparency.org/en/news">www.transparency.org/en/news</a>
News by think tanks	<a href="http://www.merics.org/en">www.merics.org/en</a>
News by corporations	<a href="http://www.bp.com/en_gb/united-kingdom/home/news.html">www.bp.com/en_gb/united-kingdom/home/news.html</a>

## 2.1 Collected information per entity type

### News sources

*Meteor* does not only consist of a list of sources that fall in the above-mentioned scope but in addition includes a set of meta information or properties per **news source**, relevant to subfields of political communication (McNair, 2017; Semetko & Scammell, 2012) and journalism studies (Carlson et al., 2018; Wahl-Jorgensen & Hanitzsch, 2009). Based on the literature review of D3.1, we distilled the key properties of news sources (see D3.1: 14) to inform the structure of the registry. In addition, the selected properties were refined within a feedback and test round with the OPTED consortium team as well as with two external experts, an expert for social media and automated content analysis and another for journalism studies<sup>6</sup>. Table 2 shows the properties covered for the different sources.

### Media organisations

The properties collected for **media organisations** are shown in Table 3. Most of the properties are publicly available through business registers<sup>7</sup>. Based on the previous work in D3.1, we highlight the importance of the legal nature of the organisation (e.g., state-owned vs privately owned).

<sup>6</sup> We thank Tobias Heidenreich and Phoebe Maares for their expertise.

<sup>7</sup> For example, the European e-Justice Portal contains a list of all available business registers in the EU: [https://e-justice.europa.eu/106/EN/business\\_registers\\_in\\_eu\\_countries](https://e-justice.europa.eu/106/EN/business_registers_in_eu_countries).

## Digital Resources: Data Archives, Corpora, and Datasets

As mentioned above, there are numerous ways of retrieving media text data. All of them have different qualities and serve different research designs and approaches. To facilitate the selection of readily available digital resources, we distinguish between **data archives**, **corpora** and **datasets**.

We define **data archives** as digital storages for full-text media text data that are regularly updated and can be queried on demand. Access is often limited in some form, as copyright protection is a typical concern of the providers for such kind of data. One of the most prominent archives for example is provided by *LexisNexis*.

**Corpora are static collections of full-text data.** Unlike data archives, **corpora not regularly updated.** They present a snapshot of text data and may contain annotations, such as the *One Million Posts Corpus* (Schabus et al., 2017) which is full text data from the Austrian news outlet *Der Standard*.

**Datasets are the result of content analysis** of media data and do not contain full-text data. They are made usually available for secondary analysis, such as the datasets of the AUTNES (Litvyak et al, 2022) or REMINDER (Lind et al, 2020) studies.

The properties that describe data archives, corpora, and datasets were developed in cooperation with WP9 and presented in Table 4. Many digital resources contain additional information for the text data that they provide. We distinguish these into **meta variables** and **conceptual variables**. Meta variables are exogenous, such as publication date, page number, or section. Concept variables are measurements of theoretical constructs, such as sentiment, populism, or toxic language. Unlike meta variables, concept variables may not be directly compatible with each other. For example, “sentiment” is a construct that is frequently measured in text analysis, but it also has various definitions (Hase, 2021).

We are using a bottom-up approach for meta and conceptual variables covered by resources. **The list of meta variables and concepts are dynamically added to the knowledge graph.** When a dataset or corpus is added to the collection that covers a “new” concept (i.e. not yet part of the knowledge graph) the concept is added as well. In the course of an internal review process the concept and meta variables are then verified by experts. The verification includes adding a short description to the variables and also adding other names to them. This step also prevents duplicated or very similar kinds of concepts from appearing in the collection. Another advantage of this approach is that the variables can continuously be evaluated and improved, based on feedback and input by the researcher community.



**Table 2:** Properties per news source

Property	Short description
name	Common brand name of source
other names	Some sources are known by several names. For example the “Mitteldeutsche Zeitung” is also called “MZ”.
channel	Technology used for publication of news items (print, online, transcripts of radio/tv/podcasts, Twitter, Facebook, Instagram, Telegram, VKontakte)
user comments	The online news source has user comments below individual news articles. (yes/no)
founding year	The ‘founding’ year of the source. If no founding year is available this can relate also to the year of the first edition (for print), the first post (social media) or the website launch (for websites).
publication kind	Self-ascribed publication kind label (e.g., newspaper, news agency, magazine, alternative media)
topical focus	Topical focus of the news source (e.g., General news, vs. Financial news)
publication cycle	Routine that determines the intervals for publication (e.g. daily, weekly, constantly)
main geographic scope of source (i.e., country)	Primary geographic scope of source (e.g., global, national, subnational). The information can be inferred from the content (actors, topics, events) At which geographical location do most of the reported news stories happen? From what geographical perspective is the news looked at?
language(s)	Written language or languages of a publication
payment model	Monetary requirements for accessing news items (e.g., free, freemium, hard paywall)
ads	The source contains advertisements. (yes/no)
ownership	Media organisation that owns the source
audience size	Number of users in a set time interval (e.g., daily site visitors, subscribers, followers)
e-paper	For print news sources: is there an e-paper version available?
archives	Names of archives and links to archives where full texts for a source are available
datasets and corpora	Name of a dataset and link to the dataset, which includes annotations for the (full) texts of the source
party affiliated	Remark whether the news source is affiliated with a political party
related sources	Other news sources that are related by brand

**Table 3:** Properties of organisations

Property	Short description
name	Legal name of organisation
other names	Abbreviations or former names
is person	Flag to indicate that the entity is an individual (in cases where an individual is the sole legal owner of a news source or another media organisation)
ownership kind	Ownership structure of the organisation to distinguish state-owned media organisations from privately owned ones. (public, private, political party, unknown, NA)
founding year	Date of incorporation
country	Country where the organisation is registered
publishes	News sources published by this organisation
owns	Media organisations owned by this organisation
party affiliated	Remark whether the organisation is affiliated with a political party
address	Address of the organisation. Automatically coded into geo-data.
employees	Number of employees of the organisation

**Table 4:** Properties of data archives, corpora, and datasets

Property	Data Archive	Corpus	Dataset	Short description
name	✓	✓	✓	Name of resource
other names	✓	✓	✓	Abbreviations or former names
authors		✓	✓	Creators of the resource
published date		✓	✓	Year when resource was released
last updated		✓	✓	Last update of resource
URL	✓	✓	✓	Link to the resource
DOI		✓	✓	DOI of the resource or its associated publication
arXiv		✓	✓	ID of the resource or its associated publication on www.arxiv.org
GitHub		✓	✓	Link to the GitHub repository of the resource
description	✓	✓	✓	Brief description of the resource
access	✓	✓	✓	Requirements to use the resource (free / registration / upon Request / purchase)
country	✓	✓	✓	Countries covered by resource
start date		✓	✓	Start date of coverage
end date		✓	✓	End date of coverage
file format	✓	✓	✓	Available file formats of resource
sources included	✓	✓	✓	News sources included in resource
materials	✓	✓	✓	Links to additional materials for the resource such as manuals or codebooks
initial source		✓	✓	If the resource is derivative work of another resource it can be linked here
meta variables	✓	✓	✓	Meta variables included in the resource (growing and curated list of meta variables such as date, or news outlet)
concept variables		✓	✓	Concepts that are annotated in the resource (growing and curated list of concepts)
text units	✓	✓	✓	Text units or units of analysis in the resource (e.g., article, paragraph, sentence)

## 2.2 Querying and Creating Entries

*Meteor* has two main functions. First, users can query the registry, second, they can contribute information to the registry. There are various filters available to query the registry based on the properties described above. Additionally, *Meteor* also offers a live full-text search that matches the users search terms against free text fields such as “other names”, “authors”, or “description”. Figure 2 shows which information is displayed to the user after a source has been successfully searched and retrieved. To visualize the complex ownership structure of news sources, *Meteor* also automatically renders an interactive graph element in the detail views of news sources and media organizations (Figure 3).

When a user contributes a new entry to *Meteor*, a questionnaire guides the user to enter the properties as listed in Table 2. After the new entry is submitted, *Meteor* leverages available APIs to enrich the information as much as possible<sup>8</sup>. Figure 4 displays the start of the questionnaire which users can complete to add a new source. The questionnaire includes various explanations and examples for each question to complete.

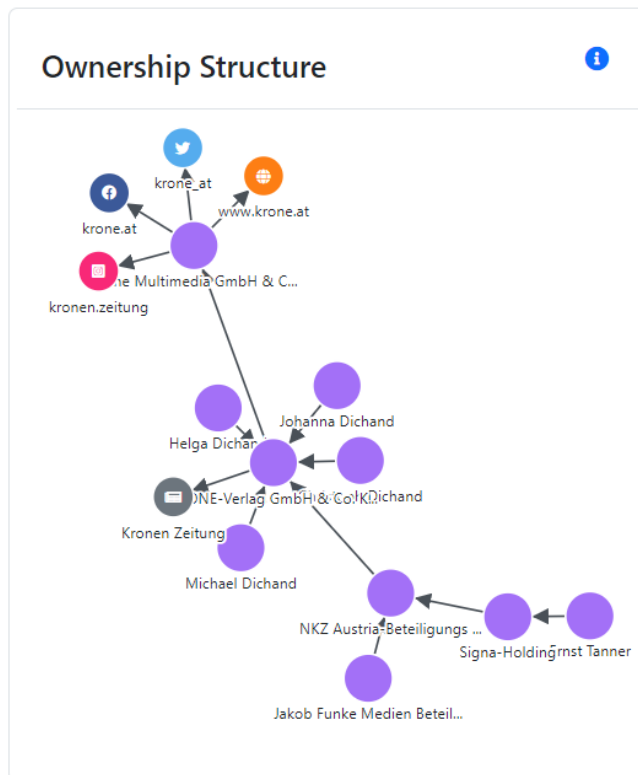
<sup>8</sup> For general information, such as geographic names or addresses, we query Wikidata and Openstreetmap. For channel-specific information we call various APIs, such as siterankdata.com to retrieve information about daily website visitor count, or the Twitter API to get the follower count of an account.

The new entry is not automatically displayed publicly, instead it enters a review process. Reviewers with expertise for the respective area (country or source type such as alternative media) double check the information before it becomes publicly visible. The reviewers are currently members of WP3.

**Figure 2: Query the registry (example)**

The screenshot shows the Meteor Media Texts Open Registry interface. At the top, there is a search bar and navigation links for 'Add Entry', 'Tasks', 'Resources', and 'About'. The main content area displays details for the source 'www.krone.at'. The 'General Information' section includes 'Also known as' (Krone), 'Channel' (Self-Hosted Website), and 'Link to Source' (https://www.krone.at/). The 'Routines' section lists 'Publication Kind' (Newspaper / News Site), 'Special Interest' (No), 'Publication' (Continuous), 'Geographic Scope' (National), 'Countries' (Austria), and 'Languages' (German). The 'Audience' section is currently empty. On the right side, there is a 'Related Sources by Brand' section listing various social media profiles for 'krone.at' and 'kronen.zeitung'. Below that is an 'Ownership Structure' section with a small graph showing relationships between 'krone.at', 'www.krone.at', and 'kronen.zeitung'.

**Figure 3: Close-up of automatically rendered graph.**



**Figure 4: Contribute a new source**

The screenshot shows the Meteor Media Texts Open Registry v1.0.3 interface. At the top left is the Meteor logo and version information. A search bar is located at the top center. On the top right, there are navigation links for 'Add Entry', 'Tasks', and 'Re'. The main heading is 'Add new entry'. Below this is a light blue box with the title 'First time doing this?' and a link to a 'detailed guide for adding entries'. Underneath is a 'Quick Tips' section with two bullet points: one about information icons and another about retrieval icons. Below the tips are two input fields: 'Name of New Entity' (an empty text box) and 'Entity Type' (a dropdown menu currently showing 'News Source'). At the bottom left of the form is a blue 'Add new entry' button.



### 3 Data Collection

We employed several strategies to add entries to *Meteor* depending on the entry type. For adding **data archives**, we relied on systematic literature reviews and the University of Vienna Library. When we reviewed the literature for D3.1 and later for D3.4, we also took note of every data archive that was mentioned in the texts. Additionally, we queried the Database Service of the University of Vienna Library<sup>9</sup>. It is a meta-database which can be queried according to research fields or data type. We searched for all databases listed in the categories “Communication” and “Social Sciences” that contain full-text data. We also queried for the free text search terms “newspaper”, and “media”. These approaches yielded a total of 19 data archives. We informed WP6 of our findings and requested feedback on the coverage of data archives.

For adding **news sources and media organisations** we employed a different strategy that is divided into four phases (see Table 5). In an initial testing phase (I) from July 2021 until October 2021 we entered all news sources that were analysed in the *European Parliament Election Study 2009* (Schuck et al, 2010). Because it allowed a wide coverage of 35 larger print news sources from all EU countries. In this phase, all other channels of each news source were also entered (e.g., Twitter accounts, Facebook pages, etc.) as well as the corresponding media organisations owning each news source. Through this process, the online platform was debugged and also the codebook (see Appendix) was refined.

**Table 5:** Phases for data collection

<i>Phase</i>	<i>Time Period</i>	<i>Results</i>
I: testing phase	July 2021 – Oct 2021	Debugging and refinement of codebook
II: closed beta with semi expert	Oct 2021 – Mar 2022	Added 1,100 news sources, 319 media organizations from 6 countries.
III: student assistant	Mar 2022 – Jul 2022	Added 1,092 news sources, 252 media organizations from 22 countries.
IV: public	Aug 2022 – cont.	Invitation to researcher community to contribute entries.

The next phase (II) was a closed beta with a semi-expert audience. We successfully conducted a seminar for master students from October 2021 until March 2022, where the participants compared media systems in Europe based on their contributions to the registry. The students learned methodological and theoretical foundations of comparative research, typologies of media systems, and the current development of hybrid media systems and new information ecologies. The course materials also include a curated collection of resources that are useful for gathering information on news media in Europe (e.g., databases for subscriber counts). After two training sessions, a total of 13 participants formed three research teams where each group chose two countries for comparison. They added the largest news sources from Austria, France, Germany, Greece, Ireland, and Spain. This process yielded a total of 1,100 news sources and 319 organisations.

For the third phase (III), we hired a student assistant<sup>10</sup> who added a diverse mix of entries. At this stage, several OPTED countries were underrepresented in *Meteor*. Therefore, the assistant also entered all relevant news sources that are listed by *LexisNexis* in Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Hungary, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, and Switzerland. Next, the assistant added German speaking alternative media as listed by Schwaiger (2022) in order to

<sup>9</sup> [https://usearch.univie.ac.at/primo-explore/dbsearch?vid=UWI&lang=en\\_US](https://usearch.univie.ac.at/primo-explore/dbsearch?vid=UWI&lang=en_US)

<sup>10</sup> We thank Celina Dinhopl for her assistance in this project.



improve the coverage of “non-mainstream” news sources. Finally, the assistant also added all major news agencies in Europe. From the period of March 2022 until July 2022, she added a total of 1,092 news sources and 252 organisations.

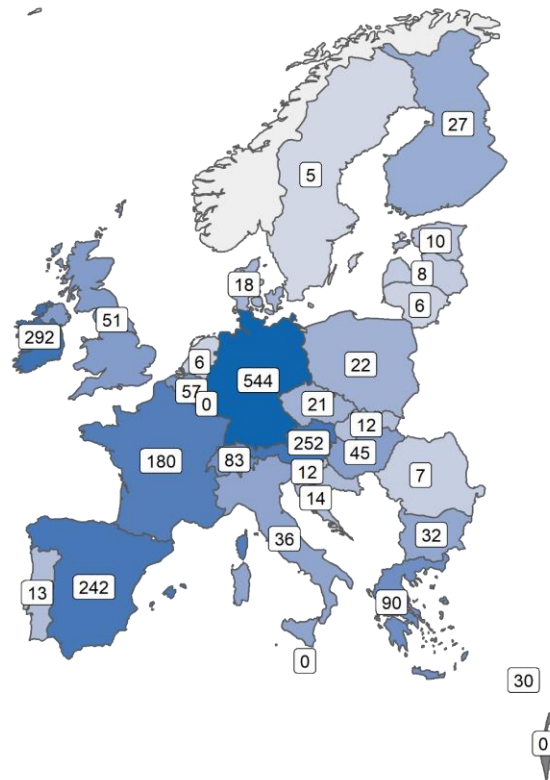
The fourth phase (IV) commenced in August 2022 and is open-ended. We published a research note in the *European Journal of Communication* where we introduced *Meteor* to the researcher community and invited researchers to participate in adding entries. For making *Meteor* a useful resource and keeping it updated, in the future we need to rely on community engagement as well as expert inputs and validation in an institutionalised setting.

## 4 Summary Statistics

As of now, *Meteor* shows over 2,200 news sources from 17 countries, as well as almost 600 organisations, and 19 full text data archives. Figure 5 shows that *Meteor* has already a wide coverage across many European countries. Especially, the German speaking countries, Ireland, as well as France and Spain have a great number of news sources. We plan to increase the number of entries even more by reaching out to more researchers throughout the project period. We especially aim to improve the coverage for the Eastern part of the EU and Israel in particular.

As of now, *Meteor* has the largest coverage for Websites, followed by Twitter, Facebook and Print (see Table 6). Telegram and VK appear to be niche channels. From our current coverage, it becomes apparent that these channels are popular among news sources that are labelled as alternative media. This highlights that they are important social media platforms for specific audiences that should not be missed in future research.

**Figure 5:** Number of news sources in *Meteor* by country



**Table 6:** Number of news sources per country and channel

Country	Print	Website	Twitter	Facebook	Instagram	VK	Telegram
Austria	41	61	42	57	43	0	8
Belgium	13	13	11	11	9	0	0
Bulgaria	3	9	8	8	4	0	0
Croatia	0	4	4	5	1	0	0
Cyprus	5	6	9	5	5	0	0
Czechia	4	5	4	4	4	0	0
Denmark	3	4	3	4	4	0	0
Estonia	2	3	2	2	1	0	0
Finland	4	5	6	7	5	0	0
France	47	53	34	33	13	0	0
Germany	50	161	99	99	55	11	69
Greece	24	21	19	18	8	0	0
Hungary	26	6	5	5	3	0	0
Ireland	62	60	65	67	38	0	0
Italy	10	11	5	6	4	0	0
Latvia	1	2	3	2	0	0	0
Lithuania	1	3	1	1	0	0	0
Netherlands	0	2	1	1	2	0	0
Norway	0	0	0	1	1	0	0
Poland	5	7	4	5	1	0	0
Portugal	2	4	3	3	1	0	0
Romania	0	3	2	2	0	0	0
Slovakia	2	3	4	2	1	0	0
Slovenia	2	3	4	3	0	0	0
Spain	42	48	47	48	47	0	10
Sweden	0	2	1	1	1	0	0
Switzerland	7	22	17	18	7	2	10
United Kingdom	15	9	11	11	5	0	0
<i>Total</i>	371	530	414	429	263	13	97

## 5 Summary and Outlook

*Meteor* is designed to be a platform for researchers investigating news sources in Europe. We believe that members of this diverse research community can greatly benefit from participation. By contributing news sources and media organisations, researchers gain a better overview over their area of expertise and identify potential blind spots. Researchers can also discover data archives where full text data is available. Furthermore, *Meteor* is also a utility for teaching courses related to news media in Europe.

*Meteor* will help researchers to rapidly find (comparable) data sources, background information about news sources, links to related full text archives, as well as annotated media data sets if available. This will facilitate the selection, access and work not only with sources but also with textual data. Due to the flexible structure of the database, additional meta-information can be added upon request. It is also possible to integrate newly emerged channels (e.g., Gettr).

Identifying the fundamental properties that capture the essence of news sources was a non-trivial task (see D3.1). “Journalism” is notoriously difficult to define, hence we prefer to have a fairly wide understanding of this concept. Further, media organisations have been affected by globalisation and are thus not anymore strictly separated by country borders. For example, online, audiences interact with mediated content from any country and in any language by using translation tools. Thus, at this stage, it remains to be seen how well a registry works that still categorises sources by country. Here, the advantages of using a knowledge graph become evident. They allow for multiple and nested properties and therefore make it possible to account for news sources that are, for instance, operating across borders and have international audiences.

Knowledge graphs can not only store complex structures, but also enable browsing and querying the data flexibly.

Achieving complete coverage of news sources in Europe with *Meteor* is going to be a **continuous task**. As described, we already have good coverage for six countries and a basic coverage across the EU. Of course, *Meteor* still lacks entries for Eastern European countries or Israel. The fact that the complete number of relevant news sources per country is hard to estimate, makes this endeavour challenging. We demonstrated that *Meteor* provides a solution for researchers to systematically study the media landscape in Europe. **An institutionalised setting of *Meteor* would ensure that the registry achieves wide coverage and remains a reliable resource in the future.**

Furthermore, some properties are not easily available for various reasons. For example, while determining the audience size is straightforward for news sources using a social media profile, it may remain unknown for printed newspapers that do not publish subscriber counts. Even though there will likely be gaps in the data, we believe that this is acceptable, because we are convinced that a registry that tries to include as much data as possible is useful for other researchers and also shows clearly where research is left to be done.

Another area where *Meteor* will receive more entries are datasets and corpora. So far, we have collected news sources from Europe including their social media accounts. While researchers can already leverage the abundance of these accounts for data collection, there are still a lot of data sources for media texts available.

Looking ahead, ultimately, we maintain a registry that is practically useful and caters to various research interests. *Meteor* is designed to benefit researchers who a) investigate the output of news sources through content analysis, b) track ownership structures of media organisations, c) analyse media production logics, d) study media systems, or e) are interested in boundary phenomena such as alternative media.



## References

- Anderson, C. W. (2012). Professionalization of journalism. In: Donsbach, W. (ed.) *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd.
- Balluff, P., Lind, F., Boomgaarden, H. G. and Waldherr, A. (2021). Towards an Inventory of Journalistic Mass-Mediated Political Texts (Deliverable 3.1). *Observatory for Political Texts in European Democracies (OPTED)*.  
[https://opted.eu/fileadmin/user\\_upload/k\\_opted/OPTED\\_Deliverable\\_D3.1.pdf](https://opted.eu/fileadmin/user_upload/k_opted/OPTED_Deliverable_D3.1.pdf)
- Carlson, M., Robinson, S., Lewis, S. C. and Berkowitz, D. A. (2018). Journalism studies and its core commitments: The making of a communication field. *Journal of Communication* 68(1): 6–25.
- Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford University Press.
- Deuze, M. and Witschge, T. (2018). Beyond journalism: Theorizing the transformation of journalism. *Journalism* 19(2): 165–181.
- Erdal, I. J. (2012). Bridging the gap: Toward a typology of cross-media news production processes. In Siapera, E. and Veglis, A. (eds) *The handbook of global online journalism*. Wiley, pp. 179–191.
- Esser, F. and Hanitzsch, T. (2012). *The handbook of comparative communication research*. Routledge.
- Esser, F. and Vliegenthart, R. (2017). Comparative research methods. In Matthes, J., Davis, C. S. and Potter, R. F. (eds) *The international encyclopedia of communication research methods*. Wiley, pp. 1–22.
- Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J. and Wahler, A. (2020). *Knowledge graphs*. Springer International Publishing.
- Fletcher, F. and Young, M. L. (2012). Political communication in a changing media environment. In Semetko, H. A. and Scammell, M. (eds) *The SAGE handbook of political communication*. SAGE Publications, pp. 36–48.
- Hase, V. (2021). Sentiment/tone (Automated Content Analysis). *DOCA - Database of Variables for Content Analysis*. DOI: [10.34778/1d](https://doi.org/10.34778/1d)
- Jensen, K. B. (2002). *A handbook of media and communication research: Qualitative and quantitative methodologies*. Routledge.
- Klinger, U., & Svensson, J. (2015). The emergence of network media logic in political communication: A theoretical approach. *New Media & Society* 17(8): 1241–1257.
- Litvyak, O., Balluff, P., Müller, W. C., Kritzinger, S., Boomgaarden, H. G. (2022). AUTNES Manual Content Analysis of the Media Coverage 2019 (SUF edition). *AUSSDA*. <https://doi.org/10.11587/FVTDXG>
- Lewis, S. C. (2012). From journalism to information: The transformation of the Knight Foundation and news innovation. *Mass Communication and Society* 15(3): 309–334.
- Lind, F., Heidenreich, T., Eberl, J.-M., Galyga, S., Edie, R., Herrero-Jiménez, B., Gómez Montero, E. L., Berganza, R., Boomgaarden, H. G. (2020). REMINDER: Historical Media Analysis on Migration 2003-2017 (OA edition). *AUSSDA*. <https://doi.org/10.11587/IEGQ1B>
- Maares, P. and Hanusch, F. (2020). Exploring the boundaries of journalism: Instagram micro-bloggers in the twilight zone of lifestyle journalism. *Journalism* 21(2): 262–278.
- McNair, B. (2017). *An introduction to political communication*. Taylor & Francis.

- Mattoni, A. and Ceccobelli, D. (2018). Comparing hybrid media systems in the digital age: A theoretical framework for analysis. *European Journal of Communication* 33(5): 540–557.
- Mythen, G. (2010). Reframing risk? Citizen journalism and the transformation of news. *Journal of Risk Research*. 13(1): 45–58.
- Neuman, W. R. (2016). *The digital difference: Media technology and the theory of communication effects*. Harvard University Press.
- Popescu, M., Toka, G., Gosselin, T., Pereira, J. S. (2010). European media systems survey 2010: Results and documentation. Department of Government, University of Essex. [www.mediasystemsineurope.org](http://www.mediasystemsineurope.org)
- Potter, W. J. (2013). Synthesizing a working definition of "mass" media. *Review of Communication Research* 1: 1–30.
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
- Robinson S. (2012). Experiencing journalism: A new model for online newspapers. In Siapera, E. and Veglis, A. (eds) *The handbook of global online journalism*. Wiley, pp. 59–76
- Ryfe, D. (2018). The ontology of journalism. *Journalism*, 20(1), 206–209.
- Schabus, D., Skowron, M., Trapp, M. (2017). One Million Posts: A Data Set of German Online Discussions. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 1241–1244. <https://doi.org/10.1145/3077136.3080711>
- Schuck, A., Xezonakis, G., Banducci, S., de Vreese, C.H. (2010). European Parliament Election Study 2009, Media Study. *GESIS Data Archive*, Cologne. ZA5056 Data file Version 1.0.0, <https://doi.org/10.4232/1.10203>.
- Schwaiger, L. (2022). *Gegen die Öffentlichkeit: Alternative Nachrichtenmedien im deutschsprachigen Raum*. Bielefeld: transcript Verlag. DOI: [10.1515/9783839461211](https://doi.org/10.1515/9783839461211)
- Semetko, H. A. and Scammell, M. (2012). *The SAGE handbook of political communication*. SAGE Publications.
- Steinberger, R. (2013). Multilingual and Cross-Lingual News Analysis in the Europe Media Monitor (EMM) (Extended Abstract). In: Lupu, M., Kanoulas, E. and Loizides, F. (eds) *Multidisciplinary Information Retrieval. IRFC 2013. Lecture Notes in Computer Science, vol 8201*. Springer, Berlin, Heidelberg.
- Veglis, A. (2012). Journalism and cross-media publishing: The case of Greece. In Siapera, E., Veglis, A. (eds) *The handbook of global online journalism*. Wiley, pp. 209–230.
- Wahl-Jorgensen, K. and Hanitzsch, T. (2019). *The handbook of journalism studies*. Routledge.
- Witschge, T., Anderson, C. W., Domingo, D. and Hermida, A. (2018). Dealing with the mess (we made): Unraveling hybridity, normativity, and complexity in journalism studies. *Journalism* 20(5): 651–659.

# Appendix

## Detailed Guide for adding News Sources

This guide is also available on the *Meteor* website which also includes a FAQ section as a curated list of third-party information: <https://meteor.opted.eu/guides/resources>

### What is included in this registry?

Before you start, please carefully read the following instructions to learn about the type of news sources we are looking for:

News texts are nowadays produced by diverse groups of actors and are accessible in various formats and via various channels. For the OPTED registry of European news sources, we focus on a specific type of news producing sources. Please carefully read the following definition.

By news sources, we mean sources that:

- have the primary purpose to regularly report and comment on recent events and ideas
- produce (self-create) textual content
- claim to be non-fictional
- can be regarded as a rather complex organisations and has standardized practises (not a one-man or one women operation)

Important, we look for:

- Sources mainly associated with EU countries + UK, Switzerland, Norway, and Israel
- Sources that are actively publishing today.
- The channel counts: e.g., Spiegel print and spiegel online are listed as separate sources
- Various regional print editions are included as separate sources if the individual region is as wide-reaching as a country (e.g., 'Zeit' and 'Zeit Österreich' are considered 2 separate sources). If the regions are subnational units, only the main source is included (e.g., 'Badische Zeitung' had 9 editions for small geographic units (e.g. Badische Zeitung Elztal', 'Badische Zeitung Freiburg'), it is only included once as 'Badische Zeitung' in the registry.
- Various social media accounts of a news source are included as separate sources
- Please provide information about the sources that are correct for the moment (to the best of your knowledge) in which you enter the information.

More specifically the registry includes:

- Print news sources
- Online news sources
- Websites that provide transcripts to self-created audiovisual contents (tv, radio, or podcasts)
- News agencies (e.g., Reuters, AFP, AP)
- Social media accounts providing news, self-created textual content
- Messaging apps providing news, self-created textual content
- Sources for multinational audiences (e.g., [www.euronews.com](http://www.euronews.com)), national audiences (e.g., [www.derstandard.at](http://www.derstandard.at)), and subnational audiences (i.e., for a region or city, e.g., [www.standard.co.uk](http://www.standard.co.uk))
- News blogs that are published by a group of people (i.e. different authors, appears to be created with editorial support, see for example [www.nachdenkseiten.de](http://www.nachdenkseiten.de), [globalvoices.org](http://globalvoices.org))
- Alternative news media sources (i.e., propagate that they provide an alternative view in contrast to mainstream media, e.g., [order-order.com](http://order-order.com))

Please note, the registry does not include:

- News aggregation sites (e.g. Google news, Apple News, Upday, Flipboard)
- Citizen journalism (i.e., news blogs by individuals, without any notion of editorial control)

- Sources clearly related to corporations, foundations, think tanks, universities, advocacy groups (often .org website urls)
- Sources with a main focus on celebrities, sport, travel, music, arts, movies, event announcements, highly specific industries
- Social media accounts only sharing news that were created by others (e.g., a private account sharing links to articles from a newspaper)
- Collective political actors (e.g., parties, governments) and (political) advertisement/public relations

#### List of credible resources

We provide a curated and growing list of resources that we found useful for adding entries to the registry: <https://meteor.opted.eu/guides/resources>

#### Audience Size for Print Newspapers

Getting detailed information about the reach of a printed news source can be challenging. However, there are interest groups that collect such data and sometimes make it available for the public.

Visit the [International Federation of Audit Bureaux of Circulations](#) to find an appropriate source for many European countries. E.g., for Austria the [OEAK](#) or Germany the [IVW](#).