Improving the relationships between Code of practice signatories and researchers (ERGA Report)



Table of Content

| Table of Content2 |
|--|
| 1 – Introduction and structure of the Report3 |
| 2 - Background |
| 2.1 - The Code of Practice on Disinformation: Pillar E3 |
| 2.2 - ERGA Report on disinformation: Assessment of the implementation of the Code of Practice4 |
| 2.3 - VVA independent study for the Assessment of the Implementation of the Code of Practice on Disinformation |
| 2.4 - EU Commission Staff Working Document Assessment of the Code of Practice in Disinformation – Achievements and areas for further improvement7 |
| 3 – Activities of the Workstream7 |
| 3.1 - Interviews with stakeholders7 |
| 3.2 – The most recent online platforms initiatives8 |
| 3.3 – Conclusions from the WS assessment11 |
| 4 – Final recommendations |

1 – Introduction and structure of the Report

The findings and the conclusions of the *Report on disinformation: Assessment of the implementation of the Code of Practice*¹, published in May 2020 by the European Regulators Group for Audiovisual Media Services (ERGA), showed enormous difficulties for the researchers to get access to relevant data from the Code of Practice signatories, although Pillar E of the mentioned Code contains specific commitments aimed at "*Empowering the research community*".

Wishing to investigate this issue more in detail, in 2020 ERGA created a specific Workstream (WS) within the Subgroup dedicated to the monitoring of the implementation of the Code of Practice's Pillar E and to study the relationship between online platforms and researchers. The goal of this Workstream, coordinated by the Italian national regulatory authority (NRA) AGCOM, was to analyse the current relationships between the signatory platforms and the researchers and to identify a set of recommendations for online platforms aimed at improving such a relationship.

In Section 2, the Report summarises the commitments of the Code of Practice on Disinformation aimed at empowering researchers and describes the findings of the Reports (ERGA Disinformation Report, EU Commission Staff Working Document, independent study from VVA) assessing the implementation of these commitments by the Code's signatories.

Section 3 describes the activities carried out by the WS members and the main findings.

Finally, Section 4 identifies some policy options and provides (i) specific recommendations addressed to online platforms and (ii) a list of KPIs aimed at *Empowering the research community* that could be considered for the revision of the Code of Practice on Disinformation.

2 - Background

2.1 - The Code of Practice on Disinformation: Pillar E

In September 2018, representatives of online platforms, leading social networks, advertisers, and the advertising industry agreed to draft a self-regulatory Code of Practice to counter the spread of online disinformation and fake news. The *Code of Practice on Disinformation*² aims at achieving the objectives set out by the Commission's Communication *Tackling online disinformation: a European approach*³ presented in April 2018 by setting a wide range of commitments, from transparency in political advertising to the closure of fake accounts and demonetization of purveyors of disinformation.

The Code of Practice signatories' commitments were organised under five pillars:

- A. Scrutiny of ad placements
- B. Political advertising and issue-based advertising
- C. Integrity of services
- D. Empowering consumers
- E. Empowering the research community

¹ <u>https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.</u>

² <u>https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.</u>

³ https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach.

The main commitments of Pillar E and the relative information useful for the monitoring activities.

| E. Empowering the research community | |
|---|---|
| 12. Support good faith independent efforts to | • Information on collaborations with fact- |
| track Disinformation and understand its | checkers and researchers, including records |
| impact | shared |
| 13. Not to prohibit or discourage good-faith | • Information on policies implementing this |
| research into Disinformation and political | commitment |
| advertising on their platforms | |
| 14. Encourage research into Disinformation | • Information on policies implementing this |
| and political advertising | commitment |
| 15. Convene an annual event to foster | • Report on the annual event |
| discussions within academia, the fact- | |
| checking community, and members of the | |
| value chain | |

ERGA has been entrusted with the task of assisting the EU Commission in monitoring the implementation of the Code's commitments by the Joint Communication *Action Plan against Disinformation*⁴ and by the Report from the EU Commission on the implementation of the Communication *Tackling online disinformation: a European Approach.*

2.2 - ERGA Report on disinformation: Assessment of the implementation of the Code of Practice

The conclusions of the ERGA Report on disinformation underlined that the contacts made by the WS with the universities and the researchers have clearly shown that the platforms provided very little (if any) access to data for independent investigations. The research community underlined some critical issues such as the problem of lack of useful, measurable and researchable data including data on *ad targeting* and *user engagement with disinformation*, the inadequateness of the ad libraries provided by online platforms in supporting in-depth systematic research into the spread and impacts of disinformation in Europe. Furthermore, the scholars interviewed by NRAs underline that: (i) not all projects need the same data; (ii) accessible data should be defined by the specific research interest and not by a company granting access on its own terms; (iii) there is no mechanism to effectively assess the quality of the data.

To their defence, the platforms have argued that they cannot provide free access to data because of privacy and data security reasons⁵.

⁴ <u>https://ec.europa.eu/commission/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en.</u>

⁵ Looking at the initiatives deployed by the Code's signatories in supporting the research community, only Facebook launched an *ad hoc* program aimed at partnering with academics and sharing privacy protected datasets. In particular, in April 2018, Facebook launched Social Science One, a very ambitious programme involving a commission of 83 academic researchers and a group of funders, with the goal of building a fair and transparent procedure to share the platform's data with academic research community. One year later, in April 2019, Facebook announced a new set of research projects that will look into social media's impact on democracy. The projects provided access to "privacy-protected Facebook data" to more than 60 researchers from 30 academic institutions across 11 Countries, in an attempt to help conduct research into a range of topics related to election campaign in Europe. To support these projects, Facebook built a first-of-its-kind data sharing infrastructure to provide researchers access to Facebook data in a secure manner that protects people's privacy. On December 11, 2019, the members of the European Advisory Committee of Social Science One

Therefore, the ERGA Report stressed the need for a set of recommendations aimed at improving the relationships between online platforms and researchers.

In a report drafted for an NRA of the WS, Prof. Rebekah Tromble, from the George Washington University in Washington, DC, examined recent scholarly research on two issues at the heart of the Code of Practice – online political advertising, micro-targeting, and disinformation – and sought to assess the extent to which this research has been enabled and supported by Google, Facebook, and Twitter. The report stated that very little scholarly research on online political advertising, micro-targeting, and disinformation has been based on data found in Facebook's, Google's, and Twitter's respective ad archives and that even the more advanced academic-platform partnership, Facebook's Social Science One, has been not so decisive in improving research on disinformation. Prof. Tromble provided several recommendations:

- As part of their public ad archives, the platforms should provide more precise data on ad spending and impressions.
- The platforms should also provide more precise targeting data in the ad archives. This should include direct targeting data, as well as information about categories targeted indirectly through the custom audience and lookalike features.
- For sensitive categories (e. g., race or political ideology), audience reach data might be substituted for targeting data. Alternatively, sensitive targeting data could be reported to regulatory authorities, with researchers given the opportunity to access the data under controlled conditions.
- The platforms should preserve deleted ad content, including content removed for violation of ad policies, for analysis by researchers. The platforms should provide formal analyses identifying their specific concerns regarding data sharing for independent academic research under General Data Protection Regulation (GDPR). Such analysis will provide a starting point for resolving areas of ambiguity and uncertainty.
- In turn, Data Protection Authorities should offer formal guidance on permissible data sharing practices under GDPR⁶.
- Regulatory authorities should begin to require that the platforms share data for research purposes. The types and amounts of data should remain flexible, with priorities set based on public interest as defined by the regulatory authorities, in consultation with both the platforms and scholars. The platforms' proprietary interests should not be neglected, but these should be balanced against the public's interest in platform transparency.

issued a public statement complaining about the lack of an adequate data access from Facebook. Surprisingly, on February 2020, Facebook has provided Social Science One with a large dataset, resulting from processing approximately an exabyte of raw data from the platform. According to Social Science One itself, this dataset will enable social scientists to study some of the most important questions of our time about the effects of social media on democracy and elections with information to which they have never before had access.

Google reported several efforts aimed at allowing researchers to access data: the researchers consulted by the NRAs stated that it is still difficult for any academic or researcher to get access from Google to useful raw data for his researches in disinformation field.

Twitter has been one of the few online platforms which made available APIs to researchers and developers. Twitter's APIs are a unique data source for academics that is used around the world in a wide range of fields, from disaster management to political science, every day.

⁶ In November 2020 the European Digital Media Observatory (EDMO) has introduced a working group on Access to Data Held by Digital Platforms for the Purposes of Social Scientific Research.' The working group's specific task is to develop a Code of Conduct under Article 40 of the General Data Protection Regulation. <u>https://edmo.eu/2020/11/24/call-for-comment-on-gdpr-article-40-working-group/</u>

• The establishment of experimental "sandbox" that would deeply representative of the ecosystem and freely available to researchers should be promoted, to the aim of supporting independent scholarly research carried out on platform data. Models from the health and medical sectors, as well as the government statistics offices, could be consulted.

Furthermore, the ERGA Report final recommendation regards the opportunity for ERGA to build cooperation with the European Digital Media Observatory (EDMO) that serves as a hub for fact-checkers, academics, and researchers to collaborate and actively link with media organisations and media literacy experts, and provide support to policymakers.

2.3 - VVA independent study for the Assessment of the Implementation of the Code of Practice on Disinformation

To support its evaluation about the Code of Practice on Disinformation, the EU Commission commissioned a study by the independent consultancy Valdani, Vicari and Associates (VVA). According to the stakeholders involved in VVA independent study, the pillar of the Code aimed at empowering the research community is the one which has proven to be the least advanced⁷.

Most of them noted that there is limited engagement with the research community and that the tools set up by platforms are still too weak. Even though some stakeholders agreed that *Crowdtangle*⁸ is an example of a good research tool, they noted that the tool is also owned by Facebook, and they hence see a conflict of interest (see next chapter). The lack of transparency regarding the access to data is a common concern raised by the researchers interviewed. The Ad Archives of Google and Facebook in particular are not seen as fit for purpose by some stakeholders. The data that could be extracted were deemed unreliable and it was noted that the archives were damaged with bugs, which ultimately made these tools effectively useless as a transparency tool for researchers, journalists, or stakeholders for whom this data was intended. Many stakeholders hence denounced the lack of user-friendliness of the platforms' databases.

Furthermore, many researchers report that access to data of platforms has not improved after the establishment of the Code. Some academics also raised the issue that the choice of platforms to grant access to researchers sometimes seemed arbitrary and that this has increased distrust between platforms and the research community.

According to the authors of the study, some service-level KPIs adequate for Pillar E, aimed at investigating how effective the cooperation is, and respectively regarding investment of platforms into research supporting fight against disinformation, could be:

- efficiency, in terms of benefits achieved versus administrative burdens;
- relevance in spurring stakeholder action to address disinformation;
- coherence with other EU interventions in adjacent areas;
- EU added value, in relation to initiatives taken at Member State level;
- sustainability as regards the longevity of outcomes produced.

⁷ The VVA Study collected evidence based on structured interviews with key stakeholder groups (Code signatories, nonsignatory platforms, national audio-visual regulatory authorities, academia, civil society organisations) identifying points of consensus and areas of disagreement with respect to the Code's:

[•] effectiveness, with each of the Code's five pillars reviewed individually;

⁸ CrowdTangle is a content discovery and social monitoring tool regarding Facebook and Instagram, owned by Facebook.

- The ratio of the number of academic/research organisations that enter into relevant arrangements with platforms against the number of data requests received;
- The ratio of donations made to academic/research organisations for research projects against total investment into combating disinformation.

2.4 - EU Commission Staff Working Document *Assessment of the Code of Practice in Disinformation* – Achievements and areas for further improvement

The EU Commission published a *Staff Working Document* on the assessment of the Code of Practice on Disinformation, that takes into account the observations stemming from the ERGA Report and the VVA Study. According to the Commission's document, it is a shared opinion amongst European researchers that the provision of data and search tools required to detect and analyse disinformation cases is still episodic and arbitrary, and does not respond to the full range of research needs.

This issue concerns firstly the quality of the datasets and APIs that should be made available to the research community at large in order to acquire a better understanding of sources, vectors, methods, and propagation patterns of false narratives having the potential to affect democratic debates and processes in the EU. Secondly, it also concerns the collaborative models developed so far by certain platforms with the academic community, which are based on discretionary, multi-bilateral arrangements, rather than on open and non-discriminatory approaches empowering a larger, multi-disciplinary community of researchers to carry out the appropriate detection and analysis activities.

The European Digital Media Observatory (EDMO)⁹ established in June 2020 is an important initiative taken by the European Commission to tackle these shortcomings: EDMO, with the support of the academic community, is expected to contribute to defining the necessary requirements to enable access to anonymised or disaggregated datasets and APIs for research purposes. EDMO should ensure that access to platform users' data is granted in compliance with the GDPR to avoid the identification of users and limit the purposes for which datasets may be used. In line with the GDPR, if, in exceptional circumstances, the processing of personal data is unavoidable, access to such data should only be made available pursuant to an appropriate legal basis for processing and safeguarded through appropriate technical and organisational security measures, including purpose limitation and data minimisation.

3 – Activities of the Workstream

3.1 - Interviews with stakeholders

In October, WS members had two conference calls with Professor Rebekah Tromble and Professor Fabio Giglietto (University of Urbino "Carlo Bo"), two scholars already involved in ERGA or NRAs activities¹⁰.

During these conference calls, WS members asked for updated information about their relationships with platforms, data needed for research purposes, and investigated their preferences about the

⁹ https://edmo.eu/

¹⁰ As said before, Prof. Tromble was one the experts contacted by a WS NRA, and the conclusions of her report on the relationships between platforms and researchers have been adopted by the ERGA Report. Prof. Tromble is one of the Social Science One European Committee. Prof. Giglietto is also a Social Science One European Committee member, and was involved in AGCOM monitoring activities related to the 2019 assessment on the implementation of the Code.

opportunity to build a framework with platforms (*ad hoc* programs such as Social Science One¹¹, open data repositories intended for researchers and scholars, data and information exchange platforms, etc.) and about the management of possible disputes between researchers and platforms¹².

Through these interviews, WS members were informed that:

- Facebook has released different kinds of datasets, data repositories, or open platforms for research purposes;
- Google has still not yet deployed a systematic approach aimed at satisfying scholars needs;
- Twitter has done a good job in sharing public APIs;
- TikTok is in an embryonal situation but it seems they would like partnering with scholars.

The WS members learned also that the Social Science One European Committee, headed by Prof. Claes De Vreese, seemed to be going to terminate its activities and its members to be trying to figure out possible alternatives.

At the same time, the EDMO Advisory Board members started to discuss with platforms how to support and facilitate the relationships between scholars and online platforms.

The interviewed researchers concluded that the best options for managing the relationships between platforms and scholars would be a Committee mainly made by scholars themselves (not far from the composition of the Social Science One European Committee) supported by EDMO which should play the role of being the organiser of the process with the platforms. In any case, there is a need for more precise rules, that could be contained in in a strengthened Code of practice on Disinformation.

3.2 – The most recent online platforms initiatives

In the latest months the issues regarding the relationship with the research community and the accessto-data for scientific purposes has been faced by all the signatory platforms, with different perspectives, solutions and degrees of development.

¹² The interview guidelines are reported below:

• a short introduction to their research activities about online platforms, and their eventual role in research institutions or committees (e. g. Social Science One Committee) dealing with access to (online platforms) data;

• some suggestions for the improvement of the relationships between online platforms and researchers/scholars:

• other privacy and ethical questions.

¹¹ Social Science One (<u>https://socialscience.one/</u>) is a consortium of leading social science research centers around the world, formed to share experiences about and develop models for collaborating with companies.

[•] the current state-of-the-art, with particular reference to updated information about relationships with online platforms (e.g. Google and Facebook political ads archives; Facebook tool Crowdtangle; ...);

[•] data needed (which data? with which temporal dimension? in which file format?...);

[•] their preferences about a framework with online platforms (*ad hoc* programs such as Social Science One, open data repositories intended for researchers and scholars, data and information exchange platforms, ...), with particular reference to a specific role to be assigned to EDMO;

[•] their preferences about the management of eventual disputes between researchers and platforms, with particular reference to a specific role to be assigned to EDMO;

[•] a discussion about the need for a template file that could be useful to NRAs in their information requests to online platforms and another template file regarding the more common data needed by any scholar for research purposes.

While it is too early to describe and assess the approach and the tools proposed by Google, TikTok, Microsoft and Mozilla (even if TikTok has expressly declared its willingness to partner with researchers in the near future) some insights about the other signatories platforms can be underlined.

Facebook

In the last months, Facebook has opened the access to Crowdtangle to scholars and researchers affiliated to public and private universities¹³. According to the Crowdtangle team, the access is currently prioritizing university researchers (faculty, PhD students, post-docs) focused on misinformation, elections, COVID-19, racial justice, well-being¹⁴.

Furthermore, Facebook has released different kinds of datasets for research purposes¹⁵, including a dataset about US elections held in November 2020¹⁶, and has deployed a specific platform, accessible only to accredited scholars, with the aim of providing researchers with the tools and the data they need to study Facebook's impact on the world, with a focus on elections, democracy and well-being, through specific datasets on users attitudes and behaviour.

In particular, through this initiative, called *Facebook Open Research and Transparency* (FORT), scholars who have applied and have been accredited, can access *Ad Targeting Transparency Data Sets*, including targeting information for more than 1.65 million social issues, electoral, and political Facebook ads that ran during the three-month period prior to Election Day in the United States, from August 3 to November 3, 2020¹⁷, and to URL Shares Data Set, including differentially private individual-level counts of the number of people who viewed, clicked, liked, commented, shared, or reacted to any URL (for any URL with at least 100 public shares) on Facebook between January 2017 and July 2019¹⁸.

Data for Good Program is another Facebook initiative specifically directed to researchers, which includes tools built from privacy-protected data on the Facebook platform, as well as tools developed using commercially and publicly available sources like satellite imagery and census data. With specific reference to COVID-19 health emergency, Facebook launched, in partnership with Carnegie Mellon University (CMU) and University of Maryland (UMD), the COVID-19 Symptom Survey, asking their users about how they are feeling, including any symptoms they or members of their household have experienced and their risk factors for contracting COVID-19¹⁹. Country and region-

¹³ Even the Social Science One team has worked closely with the CrowdTangle team over the past few months with the aim of making its data widely available to university researchers (more information at https://www.facebook.com/formedia/blog/crowdtangle-for-academics-and-researchers).

¹⁴ See https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-andresearchers#:~:text=CrowdTangle%20started%20a%20pilot%20program,and%20abuse%20of%20social%20platforms

¹⁵ Here we can find an official update: <u>https://research.fb.com/data/.</u>

¹⁶ More information at https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election/.

¹⁷ Specifically, this includes an Ad Targeting data set (the targeting information selected by advertisers running social issues, election, and political ads but as a privacy protective measure, excluding ads with fewer than 100 impressions) and an Ad Library data set (social issues, election, and political ads that are part of the Ad Library product) so that researchers can analyze the ads and targeting information in the same environment (more information at https://research.fb.com/blog/2021/02/introducing-new-election-related-ad-data-sets-for-researchers/).

¹⁸ The URL Shares Data Set has been the first dataset construced in the scope of the partnership between Facebook and Social Science One, and made accessibile to any researcher through a specific request for proposal process (more information about the launch of the URL Shares Data Set at https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one).

¹⁹ The survey is available in 56 languages. A representative sample of Facebook users is invited on a daily basis to report on symptoms, social distancing behavior, mental health issues, and financial constraints. Sampled users receive the

level statistics are published daily via public API and dashboards, and microdata are available for researchers via data use agreements. While the CMU and the UMD have made the aggregated data from these surveys publicly available, Facebook and partnering universities created a portal to provide eligible academic and nonprofit institution researchers with information about how to request access to non-aggregated survey data for research purposes. The sharing of non-aggregated data is intended to help facilitate more advanced modelling and forecasting efforts by researchers aiding public health responses around the world. Finally, in December 2020, Facebook created four datasets dedicated to economic recovery during the COVID-19 emergence, with the aim of helping researchers, nonprofits and local officials identify which areas and businesses may need the most support:

- Business Activity Trends in partnership with the University of Bristol, aggregating information from Facebook Business Pages to estimate the change in activity among local businesses around the world and how they respond and recover from crises over time;
- Commuting Zones a dataset aimed at giving visibility to geographical areas in which the commuters spend most of their time between home and work, regardless of administrative boundaries, so becoming a crucial tool for providing input to economic analysis;
- Economic Insights from the Symptom Survey, including new insights about whether people in different occupations are worried about their household finances, as well as if they have experienced disruptions in employment.
- New Waves of the Future of Business Survey, including data from monthly surveys on small businesses on Facebook, built with the World Bank and Organisation for Economic Cooperation and Development, with the aim of determining the effects of the global pandemic on their operating status, their employees and their business needs²⁰.

Twitter

Since 2006, Twitter's APIs have become a relevant data source for academics and have been used around the world in a wide range of fields, from disaster management to political science. Academic researchers have used data from public conversations to study topics as diverse as the conversation topics on Twitter itself (e.g state-backed efforts to disrupt the public conversation, floods and climate change, attitudes and perceptions about COVID-19, efforts to promote healthy conversation online and so on). Nowadays, according to Twitter itself, academic researchers are one of the largest groups of people using the Twitter API.

Twitter has constantly tried to help academic researchers use Twitter data for their research purposes, for example launching, in April 2020, the COVID-19 stream endpoint, the first free, topic-based stream built solely for researchers to use data from the global conversation for the public good²¹.

invitation at the top of their News Feed, but the surveys are conducted and collected off the Facebook app by our partners. Facebook does not collect, store, or receive survey responses, and university partners do not know who took the survey. The surveys may be used to generate new insights on how to respond to the crisis, including forecasting and modelling efforts Facebook provides weights to reduce nonresponse and coverage bias.. More information at https://research.fb.com/blog/2020/05/expanding-support-for-covid-19-research-through-the-symptom-surveys/

 $^{^{20}}$ More information on the economic recovery datasets included in Facebook Data for Good Program at https://about.fb.com/news/2020/12/data-for-good-new-tools-to-help-small-businesses-and-communities-during-the-covid-19-pandemic/

²¹ The COVID-19 stream endpoint is a unique dataset that covers many tens of millions of Tweets daily and offers insight into the evolving global public conversation surrounding the health emergency. The dataset has been made available for

At the beginning of 2021, Twitter made available to scholars and academic research community members a new Academic Research product track, allowing qualified researchers to access to all v2 endpoints released to date, with particular reference to:

- free access to the full history of public conversation via the full-archive search endpoint, which was previously limited to paid premium or enterprise customers;
- higher levels of access to the Twitter developer platform for free, including a significantly higher monthly Tweet volume cap of 10 million (20x higher than what's available on the Standard product track today)
- more precise filtering capabilities across all v2 endpoints to limit data collection to what is relevant for your study and minimize data cleaning requirements
- new technical and methodological guides.

The access to this new product track is available to researchers by applying for access with the Academic Research application, and using a new developer portal, an additional application step needed to protect the security and privacy of people using Twitter.

In particular, a manual review process was provided by Twitter to allow access to the Academic Research Product Track. First of all, applicants should meet three requirements:

- being either a master's student, doctoral candidate, post-doc, faculty, or research-focused employee at an academic institution or university.
- presenting a clearly defined research objective, and specific plans for how Twitter data are going to be used, analysed, and shared for research purposes.
- using the product track only for non-commercial purposes.

Twitter has outlined how this product track is a significant shift in the type of data it makes available for free to third-party academic researchers interested in studying user behaviours and trends related to online discourse²².

3.3 – Conclusions from the WS assessment

The meetings of the ERGA Workstream with researchers showed in the last months there have been some initiatives set forth by the online platforms, establishing partnerships with some researchers or groups of researchers or launching funding schemes aimed to scholars, to improve access to their data sets.

However, these initiatives seem to be addressing only few researchers, notably the most famous ones and they do not look like the definitive solutions to the issues the research community is facing. What the Code requires is systematic and generalised access to data owned by online platforms, and such

free from 29 April to 15 October, 2020. According to Twitter itself, this dataset should help academic researchers in analysing the spread of the disease, understanding the spread of misinformation, helping crisis management, emergency response, and communication within communities, and developing machine learning and data tools that can help the scientific community answer key questions about COVID-19. The COVID-19 stream endpoint provides access to COVID-19 and Coronavirus related public Tweets in real-time as defined by the criteria used to power this topic on Twitter. More information at https://blog.twitter.com/developer/en_us/topics/tools/2020/covid19_public_conversation_data.html

²² More information at <u>https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-</u>research-with-the-twitter-api.html

type of access is clearly still not available: the WS members analysed the problems that young scholars (not academics or independent researchers) and NGOs are facing in accessing the data owned by the online platforms, too. Some of new tools deployed by the online platforms have been specifically addressed to academic researchers, in particular to scholars affiliated to public and private universities, and independent researchers, even working for NGOs, are still facing enormous difficulties in accessing to these tools. Some of the NGOs representatives, that have been consulted by ERGA during webinars and private meeting, have complained about the different treatment carried out by online platforms in this regard.

ERGA keeps stressing the importance of laying down the conditions for improving the cooperation between scholars and platforms in a way that does not depend on the different initiatives launched by the different online platforms and addressed only to few (usually well-known) researchers. Only few of the initiatives set forth by the platforms seem to be going in this direction (giving access to specific data owned by the platforms to vetted researchers) and many more efforts will have to be done by the platforms before the commitments of the Code's pillar E may be considered complied with.

At the same time, researchers are fully convinced that only the access to APIs or raw data could be useful to academic and independent research activities, proprietary datasets being a solution convenient only to online platforms themselves and to the so-called secondary research activity. Researchers need indeed to analyse unstructured data, study the ways these data have been generated and therefore have access to an amount as wide as possible of data to being managed and scrutinised.

The role of EDMO to that end will be of utmost importance. Designing a framework to ensure secure and privacy-protected access to platforms' data for academic researchers is one of EDMO's main goals²³. The cooperation between EDMO and ERGA, each of them in respect of their respective roles, will be crucial to strike the balance between the various interests at stake and to provide clear indications on how to properly implement Pillar E of the Code of Practice.

Finally, since the issue of access to the data has been expressly mentioned by the EU Commission Communication *Tackling COVID-19 disinformation - Getting the facts right*²⁴ and by the *Staff Working Document*, where it is underlined how "the lack of access to data allowing for an independent evaluation of emerging trends and threats posed by online disinformation, as well as the absence of meaningful KPIs to assess the effectiveness of platforms' policies to counter the phenomenon, is a fundamental shortcoming of the current Code [of Practice on Disinformation]", the WG members have discussed the potential connections between the need for access to online platforms data by the research community and the need for specific data on the implementation on specific policies about disinformation, fact-checking, political advertising, and similar issues by the NRAs²⁵.

²³ More information at https://ec.europa.eu/digital-single-market/en/european-digital-media-observatory

²⁴ In the European Union Joint Communication to the European Parliament, the European Council, The Council, The European Economic and Social Committee and The Committee of the Regions "*Tackling COVID-19 disinformation - Getting the facts right*", social media platforms have been called for additional efforts and information-sharing, as well as increased transparency and greater accountability. With specific regard to the data on flows of advertising linked to COVID-19 disinformation, the signatories of the Code of Practice have been invited to provide data, broken down by Member State where possible, on policies undertaken to limit advertising placements related to disinformation on COVID-19 on their own services.

 $^{^{25}}$ Even during the conference call with Proff. Tromble and Giglietto, this issue has been specifically addressed (see footnote n. 9).

4 – Final recommendations

The analysis carried out by this WS highlighted that, in order to comply effectively with the commitments of the Code's Pillar E, platforms should build a research ecosystem based on:

- access to application programming interface (APIs) for research purposes, or availability of a tool allowing researchers to access to raw data (even regarding deleted pieces of content), and free access to ad archives (or similar archives) APIs²⁶;
- 2) identification of access requirements not penalising young and independent (NGOs) researchers even not affiliated to universities;
- 3) institution of a systematic cooperation between relevant stakeholders e.g. involving EDMO representatives, representatives from research community and independent regulatory authorities working on dataset and APIs jointly with online platforms, in charge of addressing specific issues about doing research with platforms and solving eventual disputes between platforms and researchers. The involvement of NRAs in such cooperation is key as they have rich expertise and experience in assessing data needed to evaluate and understand the impact of policies at the national level, as was highlighted in the ERGA report on disinformation.

With the aim of supporting the implementation of these measures, the access to data issues should be promptly covered by the Code of Practice on Disinformation 2.0, mentioned in the European Democracy Action Plan, and some KPIs regarding this issue should be considered in the monitoring activities carried out by ERGA or other relevant actors. Since access to data seems to be an important way by which NRAs could make the online platforms more and more accountable²⁷, some specific measures aimed at individuating specific ways by which public sector stakeholders could access to data needed for their monitoring and supervising activities (e.g. the provision of a basic set of row data regarding specific issues such as content moderation, partnership with fact-checkers, tackling hate speech and cyberbullying) should be taken in account as well.

In this regard, the Code of Practice on Disinformation 2.0 should include some specific KPIs relating to the Pillar E *Empowering the research community* which should refer to the European Union territory as a whole and to the EU Member States:

| Structural qualitative indicators | |
|---|--|
| 1) Availability of policies ensuring the connection between platforms and research community, even trough the provision of specific access-to-data tools (or even allowing the access through APIs) | |
| | |
| Service-level quantitative indicators | |
| 1) Amount of raw data made available to academic/research organisations through specific tools (or even allowing the access through APIs) | |
| 2) Number of specific requests for data received | |
| 3) Number of specific requests for data followed up | |

²⁶ In this sense, the EDMO initiative regarding the drafting of a code of conduct under Article 40 of GDPR would reduce any potential legal uncertainties and risks for the platforms, and offer researchers a clearer route to data access, including sensitive data..

²⁷ Even some NGOs dealing with platform society issues state that meaningful research access is a pre-condition fo informed and effective platform governance (see, among others, Ausloos, J., Leerssen, P., & ten Thije, P. (2020). *Operationalizing Research Access in Platform Governance*, Algorithm Watch, <u>https://algorithmwatch.org/wp-content/uploads/2020/06/GoverningPlatforms IViR study June2020-AlgorithmWatch-2020-06-24.pdf</u>).