RESET AUSTRALIA POLICY MEMO

Data Access Mandate for a Better COVID-19 Response in Australia

A 'LIVE LIST' RESOURCE FOR RESPONDERS AND PRACTITIONERS

Reset. AUSTRALIA

Matthew Nguyen,
Alexandra McIntosh,
September 2020

Executive Summary

Reset Australia is an independent research and advocacy organisation driving awareness of and solutions to address the digital threats to democracy.

The Issue

Our information landscape is dictated by digital platforms such as YouTube, Facebook and Twitter, and the COVID-19 pandemic has highlighted how misinformation on these platforms hampers our public health response. And there are now growing levels of concern on the impact of misinformation on the roll out of the vaccine.

An AVAAZ Report has highlighted the extent of this problem on Facebook, citing that over the past year, content from a global health misinformation network spanning 5 countries has been viewed an estimated 3.8 billion times[1]. Another recent study revealed how showing people false COVID-19 information leads people to reconsider a potential vaccine[2].

This is illustrative of the scale of this issue, as well as the real world implications that misinformation has on the current management of the crisis and any future vaccination efforts. In order to counter this, we must be able to assess and where appropriate, address misinformation. Currently, this data is guarded by the digital platform companies, and even though many of these companies have acted in good faith, this has been solely at the discretion of these private entities while the problem persists.

By mandating transparent data access via a public regulator, we can guarantee the appropriate governance, enforcement and oversight safeguards to ensure that public interest is prioritised during this crisis.

The Solution

The Data Access Mandate for a Better COVID-19 Response in Australia is a proposal that achieves the necessary data access to bolster our COVID-19 response whilst ensuring user privacy rights and balancing intellectual property issues. The objective of this Mandate is to provide relevant stakeholders access to information held by the digital platforms which will aid in the COVID-19 response and benefit all Australians.

This Mandate will do this by providing access to de-identified data on COVID-19 related URLs circulating on the platforms. URLs which are deemed eligible to be included must meet certain criteria. Once deemed eligible, this Mandate compels the digital platforms to share certain demographic (such as age, gender, location) and impact (such as views, shares, engagement) data. This 'Live List' of circulating misinformation will be a resource for Australia's COVID-19 response and allow for the timely identification, better analysis and more effective communication and community engagement strategies to respond to misleading and harmful information.

[1] AVAAZ (2020), 'Facebook's Algorithm A Major Threat to Public Health'. Found online here.

[2] Loomba et. al. (2020) 'Measuring the Impact of Exposure to COVID-19 Vaccine Misinformation on Vaccine Intent in the UK and US'. Found online here

The Mandate

For URLs to be deemed eligible to be included, they must meet certain criteria, which are set out here. Please see the full Mandate for further information:

A. Appearance on a digital platform with significant impact to Australians

B. Relevance to an Australian audience

C. If it is related to COVID-19 and/or related issues

D. If it has cumulatively amassed over 1000 impressions

Once deemed eligible, this Mandate compels the digital platforms to share certain demographic and impact data which include:

Reach and Engagement

- Number of times they have been shared on public and private accounts/groups
- Number of link clicks
- Number of engagements
- Number of views / impressions / reach
- Number of flags / reports
- Date first shared
- Date if it was fact-checked
- Date if removed
- Which public accounts have shared the URL (e.g. pages, brands, celebrities, verified accounts)

Content theme (COVID related)

- List of keywords identified in URL (i.e. in the string, metadata, post content) that match those in the Keyword database

Societal Impact

- Across shares, link clicks, engagement and impressions, a % breakdown across gender, age, location (Local Government Area) and language.*
- Number of flags / reports by users
- Number of shares by suspected bot account

**Demographic % breakdown data will only be provided for shares, link clicks and engagement if these respective engagement metrics surpass 100 unique users*

Implementation

We propose that the governance and administration of this Mandate lie with the Australian Communications and Media Authority, with an additional Independent Oversight Board made up of relevant stakeholders and industry representatives in order to ensure all concerns are incorporated and ethical use is guaranteed.

Specific governance, enforcement, implementation and privacy mitigation measures have been incorporated and are detailed in the full mandate accessible on our site: au.reset.tech/livelist.

This Live List will be a resource for Australia's COVID-19 response and allow for the timely identification, better analysis and more effective communication and community engagement strategies to respond to misleading and harmful information.

To learn more please contact Chris Cooper at chris@au.reset.tech

Contents

01. Objectives

The goal of the ‘Data Access Mandate for a better COVID-19 Response in Australia’ (the ‘Mandate’) is to encourage a proactive and transparent information and data sharing partnership between the intended users (see Section 4.1) and the digital platform companies that would enable a more efficient and effective response to the ongoing COVID-19 crisis in Australia.

This Mandate proposes an enforced regulatory instrument that compels the digital platform companies to share relevant impact and demographic data of COVID-19 related URLs that appear on their respective platforms.

The intention for this data access is to provide a ‘Live List’ resource for responders and practitioners to enable timely identification, effective analysis and better communication and community engagement responses to misleading and harmful information.

02. Context

The development of this Mandate has come as a response to the growing levels of concern on the impact misinformation has had on the COVID-19 response. As the COVID-19 pandemic has tested the limits of our public health, medical response and scientific discovery systems, we also need to galvanise the same proportionate tools, expediency and transparent systems to respond to the information pandemic that has resulted.

At a global level, there is a growing body of evidence for how misinformation and disinformation has negatively impacted the response, including communities and individuals. A recent AVAAZ Report has highlighted the extent of this problem on Facebook, citing that over the past year, content from a global health misinformation network spanning 5 countries has been viewed an estimated 3.8 billion times[1]. Looking specifically at Australia, narratives around face masks[3], plandemics[4] hydroxychloroquine[5], 5g[6], vaccination[7], Asian people[8] and other conspiracy theories have regularly surfaced since the start of the pandemic.

As this crisis continues, the gargantuan level of false information must be assessed and where appropriate, addressed. This is especially important as a new vaccine is discovered and inoculation to end lockdowns begin. The first step in understanding, analysing and intervening on circulating information is ensuring transparent data access. Responders and practitioners must be able to understand which marginalised and/or niche communities are being targeted and by what messages, in order to decide whether to intervene and if so, which strategic communication tools to use.

Currently, this information is hidden, guarded by the digital platform companies. And even though many of these companies have acted in good faith during this crisis, proactively working with some authorities and providing data, this is all still at the behest of these private, international entities. This Mandate seeks to provide a pathway to achieving the necessary data access, while balancing Australia's public health interest, user privacy rights and intellectual property issues.

[1] ibid (Page 2)

[3] Iorio K (24 July 2020), 'There are many myths about using face masks in the coronavirus pandemic. Here's what we know about them', ABC News. Found online here.

[4] Bedo S (21 Aug 2020), 'Plandemic documentary: Wild theory virus trail goes back to 1999', news.com.au. Found online here.

[5] McGowan M (27 Aug 2020), 'Coalition shields firebrand MP Craig Kelly from censure for spreading Covd misinformation', The Guardian. Found online here.

[6] Australian Government (2020), '5G misinformation and COVID-19'. Found online here.

[7] Davidson J (24 Aug 2020), 'Facebook, Twitter must do more to stop anti-vaxxer COVID-19 lies: GPs', Australian Financial Review. Found online here.

[8] South China Morning Post (13 Feb 2020), 'Australians avoid Chinese restaurants amid coronavirus fears, fake news'. Found online here.

03. Policy Development and Methodology

3.1 Methodology

To develop this Mandate, a consultation process involving a series of qualitative interviews with seventeen individuals was conducted. Participants included government officials, public health officials, public health academics, political science scholars, digital communications/ platforms academics, civil society representatives and journalists, chosen for their alignment with our intended relevant stakeholders.

Two of the academics interviewed were involved in Social Science One[9], a partnership between Facebook and academic institutions that intended to provide researchers access to URLs and associated data shared on the platform in order to study issues of social and political significance - such as the impact of social media on elections. These interviews were of particular significance as the Social Science One project had a similar scope to the Mandate in terms of the data which was to be made available to researchers, however a number privacy related challenges stopped Facebook from delivering the dataset that was initially agreed on.

The interviews with participants were semi-structured and split into two parts:

- **Problem validation** - a set of questions intended to surface the impacts of COVID-19 misinformation during the pandemic, and to validate whether a lack of transparency on COVID-19 information circulating on social media during the pandemic is impacting the ability of government, public health officials, civil society and academics to respond
- **Solution validation** - the Mandate was introduced as a topline 1-page draft scope (see appendix: Live list draft prototype for consultation). This was used to gauge if this policy proposal would fulfill its objectives and obtain feedback for further development

Following the consultation process, insights were incorporated into the final Mandate.

[9] Social Science One program website - https://socialscience.one/

3.2 Research Findings

Our interviews validated that there is a lack of transparency on circulating COVID-19 online content and its impacts, and the urgent need for better solutions. Input from the interviewees informed amendments the scope of the Mandate as well as the recommendations for the implementation. Additionally, a number of limitations of the Mandate were identified and captured in Section 5.1.

1 MISINFORMATION ON SOCIAL MEDIA DURING THE COVID-19 PANDEMIC HAS IMPACTED PUBLIC HEALTH AND SAFETY AND SOCIAL COHESION.

Misinformation during the COVID-19 pandemic has been widespread. False information and conspiracies have emerged related to the origin of the disease, how it spreads, prevention measures and possible cures. Existing conspiracies, such as the anti-5G and anti-vaccination movement have co-opted the confusion surrounding the pandemic, resulting in these fringe narratives spilling over to the mainstream. This misinformation has resulted in real-world impacts, such as:

- Members of the Australian public ignoring public health advice to attend protests against lockdown measures and 5G rollout
- Contributing toward the increasing opposition toward a potential Covid-19 vaccine[10]
- Despite no scientific evidence, public opposition to the rollout of 5G is widespread which is evident in public submissions to a 5G committee, with the submissions reflective of the language used in Facebook conspiracy groups[11]
- Misinformation spread on social media has pointed toward minority communities in Australia for both spreading misinformation and for spreading the disease.

2 TRANSPARENCY AROUND WHAT IS COVID-19 RELATED INFORMATION CIRCULATING ON SOCIAL MEDIA IS CRITICAL TO COMBATING MISINFORMATION.

As medical treatments and vaccines are still in development, Australia's frontline defence against the virus is public health information. Understanding what misinformation related to COVID-19 is circulating on social media is key in developing an appropriate public health response.

Communication will also be vital moving forward, as we plan messaging around a potential vaccine and address vaccine hesitancy. It is critical for an effective response to understand what information is out there, how many people have seen it and who these people are, yet this information is not readily available.

For civil society organisations, transparency around content targeting their communities is important to understand what harmful content needs to be addressed. For example, it was noted by a participant that doctors and nurses of Asian backgrounds in Australia have been asked by patients not to treat them and have been accused of spreading the 'Chinese Virus'. To begin to address this harmful content it is critical to understanding how it is being shared and who is being exposed.

[10] "Rhodes, Hoq, Measey, Danchin (14 September 2020) "Intention to vaccinate against COVID-19 in Australia", The Lancet. Found online here.
[11] "How misinformation about 5G is spreading within our government institutions – and who's responsible" by Michael Jensen, The

3 METHODS TO UNDERSTAND WHAT COVID-19 INFORMATION IS CIRCULATING ON SOCIAL MEDIA ARE LARGELY INADEQUATE

One of the biggest challenges in combating misinformation is around accessing what is circulating privately in closed groups, on private profiles and within direct messages. Many interviewees noted that it is within these channels that much of the current misinformation originates and spreads

Whilst Twitter data is the more readily available, the platform is designed to prioritise content for a public audience, rather than private. Facebook provides the CrowdTangle platform to understand what is circulating on the platform it is again limited to what is in the public domain i.e. content posted on public profiles and pages. Additionally, access to this platform is currently restricted to researchers and several other limited use cases.

Currently there is no efficient or ethically sound way to understand the impact of COVID-19 related content spread in the private domain.

4 MANDATING DIGITAL PLATFORMS MAKE AVAILABLE DATA ON URLS RELATED TO COVID-19 WILL AID IN THE RESPONSE.

The intended data access provided by this Mandate was noted to be directly useful in the current work of 15 out of 17 participants. The two respondents who indiated it wouldn't be directly useful stated that whilst not directly relevant, it would likely be useful to the overall pandemic response. A number of academic participants noted it connected well with their own research which has become increasingly computationally intensive as the pandemic has progressed. It was noted that the centralised collection of this data would make it more consistent, useful, accurate and efficient.

The real-time nature of the Live List was identified as critical for responders. The rationale for this is the rapid emergence of misinformation and hence, need for rapid assessment and response.

The accompanying demographic data (age, gender, location and language) was seen as highly valuable, as it allows users to understand who is being impacted by harmful content to enable a tailored response. It was noted location data would be most useful at a granular level, such as a local government area, to allow for a targeted public health response. Language was also identified as a key data point as some participants noted the current ineffectual response in relation to non-english speaking communities.

5 ENSURING RELEVANT COVID-19 RELATED URLS ARE SURFACED MAY PROVE TECHNICALLY CHALLENGING

Websites peddling misinformation have been known to deliberately attempt to evade detection and obscure the nature of the content through the URL. This could involve using a numerical URL or spelling keywords wrong e.g. c0v1d name.

The solution proposed to identify COVID-19 related URLs in the draft scope - the development of a database of COVID-19 related keywords - was deemed as appropriate by participants. It was noted to detect relevant URLs the keywords should be searched in not only the URL string, but also the page metadata and associated user generated post copy. Additionally, the keyword list would need to evolve to include new keywords as they become associated with the virus, with suggestions this could be partly automated with human supervision. These measures would allow for the identification of URLs which may be trying to evade detection. Some participants noted the digital platforms would already have this type of measure in place as part of their efforts to combat misinformation.

There were concerns around having a proposed 'reach threshold' for URL inclusion, whereby URLs would only be included if a certain number of platform users were exposed to the content. The concerns were due to the potential for misinformation to bubble away under the surface while still reaching enough people to have negative consequences. However, some participants did note that either a low or no threshold would raise privacy concerns, with the potential for URLs of a more private nature to be surfaced.

6 DIGITAL PLATFORMS WILL LIKELY BE RELUCTANT TO PROVIDE ACCESS TO THE RECOMMENDED DATA

Social Science One[12] was a collaboration initiative between Facebook and academic institutions which sought to make available to researchers the URLs shared in order to study and address societal issues. Throughout the project a number of privacy related concerns became roadblocks to researchers accessing the originally agreed upon data set. This led to a number of funders pulling out of the project[13].

While the Social Science One initiative is Facebook specific, similar concerns are likely to emerge in relation to the live list data access policy across Facebook and potentially other platforms. A number of these were raised by academics:

- Sharing demographic data points relating to platform users - Data fields in the live list data access policy include the demographic details of age, location, gender and language. The sharing of demographic data in relation to URLs comes with concerns of re-identification.
- Compliance with the EU General Data Protection Regulation (GDPR)[14] - Under GDPR, the processing of personal data of EU individuals is prohibited unless it is expressly allowed by law or the user has consented. While the draft scope was limited to URLs which Australian users have seen, interacted with or shared, it may be that GDPR regulations are a roadblock due to the global nature of these companies and their user base.
- External liability for privacy breaches - Another concern raised by researchers was around the need for digital platforms to pass on liability. If the data was made publicly available, no other party would assume any level of liability with this falling squarely on the digital platforms.
- The potential for identified URLs to contain material of a private nature - Concerns around the possibility that URLs which meet the URL keyword criteria contain personal information. For example, a personal wedding album.

7 CONCERNS WERE RAISED IN RELATION TO THE DRAFT PROPOSAL TO MAKE THE DATA PUBLICLY AVAILABLE.

In the draft scope shared with participants in the interview process, our recommendation was that the resource would be publicly available. Concerns were raised around the potential for misuse. These concerns included:

- Unintentionally increasing the reach of misinformation - in the creation of a public resource, false COVID-19 information would be surfaced, giving more air to misinformation
- The potential for the resource to be gamed - a common action of those who intentionally spread misinformation is to attempt to get a conspiracy to trend online, with concerns this would be the case if URL data on shares and impressions was made publicly available in real-time

[12] ProSocial Science One programme website - https://socialscience.one/

[13] "Frustrated funders exit Facebook's election transparency project" by Alex Pasternack, Fast Company (Oct 28th, 2019)

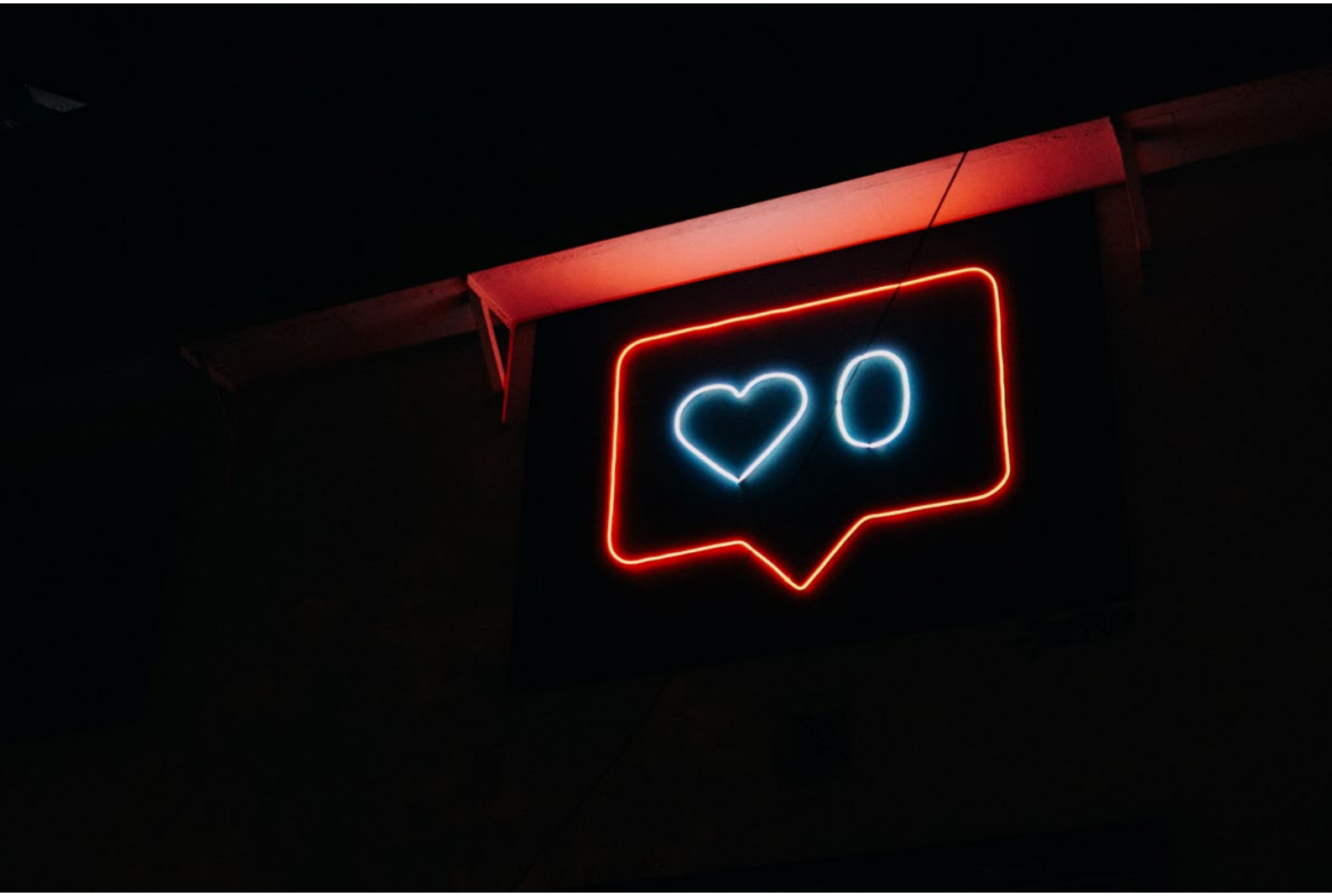
[14] EU GDPR - https://gdpr-info.eu

- Misuse by amateur researchers - there is potential for the misrepresentation of findings, as an example there could be an accusation that minority communities are responsible for spreading misinformation
- Becoming a resource for conspiracy theorists - the resource could potentially be used to source and push out false COVID-19 misinformation

8 KNOWLEDGE GAPS MAY HAMPER THE ABILITY OF THE GOVERNMENT TO RESPOND TO MISINFORMATION SURFACED THROUGH THE RESOURCE.

A number of participants noted a lack of technical skills and social media expertise may impact the ability for government and public health officials to use the resource to respond to misinformation. Concerns raised included:

- When do decision-makers decide to respond to false information
- How decision-makers choose to respond to false information
- The level of understanding of policymakers o the impact of false information



04. The Mandate

4.1 Intended Users

The intended users of this resource will be termed ‘relevant stakeholders’. We define ‘relevant stakeholders’ as individuals and/or organisations that have a significant role in the public health response to COVID-19 in Australia. This includes but is not limited to:

GOVERNMENT OFFICIALS

PUBLIC HEALTH OFFICIALS

MEDICAL PRACTITIONERS

ACADEMIC RESEARCHERS

JOURNALISTS + OTHER MEDIA PERSONNEL

COMMUNITY GROUPS + OTHER CIVIL SOCIETY ORGANISATIONS

4.2 Scope

4.2.1 Criteria for URL Inclusion

We define a URL as the web address that uniquely identifies a web page, this includes but isn’t limited to pages, websites, tweets, videos and/or posts. This Mandate applies to URLs on relevant digital platforms (see below for criteria) that meet the following criteria:

- Is a URL link shared on the platform to either an internal (i.e. another Facebook post, another tweet) and/or external (website, blog) destination e.g. a URL is shared in a private Facebook group, a URL shared in a Twitter post
- Is user-generated content with a unique URL link on a platform which allows users to share / repost the content e.g. a Youtube or TikTok video, a Twitter post

A APPEARANCE ON A DIGITAL PLATFORM WITH SIGNIFICANT AUSTRALIAN IMPACT

This Mandate should apply to digital platforms that have a significant impact on the information ecosystem for Australians. This is defined as having at least one million active users.

Relevant digital platforms are those that allow users to share content as well as view and engage with content shared by other users. These platforms have served as both formal (e.g. government to citizen) and informal (peer to peer) information sharing networks for COVID-19 and related content.

We believe this will likely include

FACEBOOK
YOUTUBE
TWITTER
INSTAGRAM
TIKTOK
LINKEDIN
SNAPCHAT

Messaging platforms have been excluded from this Mandate due to the often encrypted nature of these services.

B RELEVANCE TO AN AUSTRALIAN AUDIENCE

This is defined as:

URLS that are shared from Australian accounts or reach Australian users

C RELATED TO COVID-19 AND/OR RELATED INFORMATION

This is defined as:

URLs must be relevant to COVID-19 and/or related matters. This may also include vaccine hesitancy, 5G and/or face mask information as it relates to the COVID-19 pandemic.

Our recommended process to determine whether a URL relates to COVID-19 or related matters is as follows*:

- **Construction of an evolving database of keywords known to be related to COVID-19** Work with the digital platforms and relevant stakeholders to define an initial keyword database that's based on current knowledge and best practice.
- **Develop a system for flagging and identifying new keywords that are related to COVID-19** As the language and online discussion around COVID-19 evolves, new words, terms and/or phrases will arise to reference COVID-19.

  E.g. China virus, wuhan virus.

  In order to capture these emerging keywords, an iterative and regular review must be undertaken in order to understand emerging trends and keep the keyword database up to date. This might take the form of:

  - » **Working with relevant stakeholders** As new keywords emerge and are picked up by relevant stakeholders, they are flagged and assessed to determine if they are added to the database
  - » **Working with the digital platforms** As new keywords emerge and are flagged by the digital platforms, they are flagged and assessed to determine if they are added onto the database
  - » **Automated detection.** Recognising that the digital platforms might have internal automated detection processes
- **Linking Keywords to URLs**. In order for the URL to be included, one or more of these keywords must appear within the URL and/or post.

This is defined as if the keyword appears in the:

- » URL string,
- » URL meta description,
- » other aggregate or post level metadata (not user level data),
- » any associated descriptor and/or
- » user generated post text (either as hashtags or natural text) itself.

This Mandate is intended to capture URLs shared in both public (e.g. Twitter, Youtube) as well as private (Facebook Groups) online spaces.

In the case that keywords are only detected in user generated post text, it is recommended the URL is only included if the keywords detected are present in a predetermined proportion of user generated post text (e.g. 25% threshold). This is to avoid URLs being included when a single person includes a COVID-19 keyword in their post in relation to a URL which has nothing to do with COVID-19.

D ENSURING USER PRIVACY

Ensuring that the captured data adds utility to the intended outcome of this Mandate whilst ensuring that the highest levels of digital rights protections is paramount. Whilst implementing an impression threshold for URLs to be included would limit the resource's ability to detect emerging, fringe or niche content, it mitigates privacy concerns such as re-identification.

This is defined as:

URLs with over 1000 cumulative impressions

*This identification process may be supplemented or superseded by relying on an internal digital platform classifier, however measures to ensure transparency and reliability must be ensured.

This comes from an assumption that most of the major platforms would have an internal classification system for identifying COVID-19 related content circulating on their products (e.g. the Facebook COVID-19 CrowdTangle Dashboard). In instances where an adoption of an internal system is more effective, efficient and/or reliable, this internal classifier could be used however the following challenges must be addressed.

Firstly, using an internal digital platform classifier will highlight the differences in how each platform tags, classifies and determines COVID-19 related information and thus limits the universal accessibility of the data. In order to mitigate some of these concerns, respective companies must:

- Make accessible the processes, criteria and definitions of how they are identifying relevant information
- Work in good faith with relevant stakeholders and users of the Mandate to ensure that this capture is up to date and as comprehensive as possible
- Work in good faith with other digital platforms to ensure that data is provided in a format that can be harmonised and usable

URL eligibility for Live List

(must answer YES to all questions to be included)

A. DOES IT APPEAR ON A DIGITAL PLATFORM WITH SIGNIFICANT AUSTRALIAN IMPACT?

Digital platforms that have at least one million monthly active users in Australia.

B. IS IT RELEVANT TO AN AUSTRALIAN AUDIENCE?

URLS that are shared from Australian accounts or reach Australian users

C. IS IT RELATED TO COVID-19 AND/OR RELATED ISSUES?

URLs must be relevant to COVID-10 and/or related matters. This may also include vaccine hesitancy, 5G and/or face mask information as it relates to the pandemic.

D. DOES IT SURPASS AN IMPRESSION THRESHOLD THAT ENSURES USER PRIVACY?

URLs must have over 1000 cumulative impressions.

4.22 Data Access Requirements

The following data points must be provided once a URL is deemed eligible for inclusion. Recognising that there will be differences across digital platforms, respective companies must provide the following points to the extent that they capture them.

REACH AND ENGAGEMENT

- Number of times they have been shared on the platform
  - » From Public accounts (e.g. pages, brands, celebrities, verified accounts)
  - » From Private accounts
  - » Total amount of shares
- Number of link clicks
- Number of engagements
  - » Variable by platform but might include - likes, reacts etc.
- Number of views / impressions / reach
- Number of flags / reports
- Date first shared
- Date if it was fact-checked
- Date if removed
- Which public accounts have shared the URL (e.g. pages, brands, celebrities, verified accounts)

CONTENT THEME

- List of keywords identified in URL (i.e. in the string, metadata, post content) that match those in the Keyword database

SOCIETAL IMPACT

- Across shares, link clicks, engagement and impressions, a % breakdown across
  - » Gender
  - » Age
  - » Location (Local Government Area)
  - » Language
- Number of flags / reports by users
- Number of shares by suspected bot account

*Demographic % breakdown data will only be provided for shares, link clicks and engagement if these respective engagement metrics surpass 100 unique users

This data should be provided both historically as relevant and as close to real-time as possible.

4.23 Format for Data Access

Our recommendation is for relevant stakeholders to be granted access to the data set via an Application Programming Interface (API). Digital platforms which meet the inclusion criteria would be required to become an API provider and grant data access to authenticated users via an API standardised across participating platforms. In each instance a URL is shared, this would be recorded and pushed into a live stream with accompanying data fields if it meets the inclusion criteria.

The keyword list should be made available in full to API users, enabling them to capture only the data they need if they have a specific search parameters e.g. only 5G related content.

Given their comparable role as Data Standards Body (DSB) in the development of the Consumer Data Right (CDR)[15], we recommend Data61 be appointed as DSB to lead the development of the common technical standards in collaboration with the ACMA and participating digital platforms.

Due to the potential for misuse (see Research Findings 3.2) a level of gatekeeping to grant users rate-limited API access is recommended. API access should be granted by the regulator on users in accordance to our Governance recommendations found in Section 4.31.

The recommendation for an API solution for this Mandate has been made for the following reasons:

1. Enables standards to be set for data provision by digital platforms, ensuring data is reliable and consistent and that there is a centralised means to access the mandated data across multiple platforms
2. An API solution promotes innovation. Providing a standardised interface for data access enables layers of services / tools to be built on this API e.g. Fact-checking services, analysis of diffusion of URLs across platforms, tracking the impact of misinformation interventions
3. It is the preferred format for research institutions which will ensure our longer term understanding of mis/ disinformation and harmful content on social media

While we believe an API solution is the most appropriate, we recognise this may be inhibitive to smaller institutions who may not have the technical capabilities or financial resources to analyse this data.

To address this we recommend either the regulator or an independent body is assigned to develop and host a light-weight web-based dashboard to provide a second tier of access. This interface would provide a centralised dashboard where users can read but not export the data. This dashboard should enable users to perform some analysis on the data set, such as sorting and filtering by the various data points accompanying each URL. It may be that this is a limited dataset, such as the 1000 URLs each day with the highest reach, collated at the end of each 24 hour period. Daily lists should be stored as a web-page and be available retrospectively. Access to the web interface would be based on the same criteria as the API access.

[15] Data 61 Consumer data right

Public view

While there are concerns around potential misuse, we believe a level of public transparency around how misinformation is propagating on social media is in the national interest.

We recommend:

- The eSafety commissioner is assigned the role of determining how this information might be shared publicly with the goal being to inform and educate consumers
- A promotion is run to encourages those with ideas for innovative interventions which address the impact of misinformation on the public to apply for API access

4.3 Implementation

4.31 Governance

Our recommendation is that the administration and governance of this Mandate be housed within Australian Communications and Media Authority (ACMA) given:

- Their existing responsibilities regarding regulation of media content, including news and current affairs
- Their existing efforts on misinformation regulation, in particular the development of the 'Australian voluntary code(s) of practice for online misinformation'

Our recommendation is the development of this Mandate be conducted in parallel to the development of the voluntary code, with ACMA overseeing the process and working with digital platforms, stakeholder groups and academics to align on and design the technical solution.

While the purpose of the voluntary code is to provide users with the appropriate projections and remedies for misinformation, the purpose of this Mandate is to empower COVID-19 responders, enabling them to design their responses with the understanding of what is circulating in the COVID-19 information ecosystem. Because of this distinction, we strongly recommend this Mandate be fast-tracked to be developed and implemented before the end of the year. This rapid response is proportionate to the threat posed by COVID-19 misinformation to public health and safety.

ENSURING TRANSPARENCY AND INDEPENDENCE

The management of this Mandate and resource must be both independent and transparent. An added layer of governance would be achieved by establishing an Independent Oversight Board. The composition of this Board should include representatives from all included digital platforms and/or industry representatives, diverse members from the intended stakeholders of this resource, members of the regulating authority and members of the general public (through application) as the ultimate beneficiaries of this resource.

The objective of this Board would be to ensure that the right level of autonomy and and sector knowledge be considered, when addressing issues of:

- Data governance
- Ethical management
- Specific case arbitration

DETERMINING ACCESS TO THE LIVE LIST RESOURCE

Access to this Live List resource must be regulated to ensure that only people who are actively working on Australia's COVID-19 response. This is to further mitigate the privacy risks associated with the release of this dataset.

We recommend that users intending to access this resource must apply, with their application being assessed by both the Independent Oversight Board and the ACMA. This application should provide information that makes a clear case for their usage, including information such as:

PERSONAL INFORMATION

- Name
- Title/Role
- Institution
- Contact Information

USAGE

- Intended use for this data and this resource
- Measures
- Timeframe of use
- Data storage and deletion structures

4.32 Enforcement

The ACMA must ensure that compliance with this Mandate is enforced. This reflects the public interest and critical nature of this data in supporting the COVID-19 response. Specifically, digital platform companies must:

1. Ensure that once a URL is deemed eligible, the required data points must be released in a timely manner, within 24 hours of the inclusion assessment
2. Respond to requests for clarification from relevant stakeholders within 7 days

Punitive measures for non-compliance should be proportionate to the level and intention of the transgression and determined at the discretion of the ACMA and Independent Oversight Board.

4.33 Mitigating Privacy Concerns

A number of privacy related concerns digital platforms were identified during the consultation process (see 3.2: Research findings). Suggestions for mitigating these privacy concerns have been outlined below.

1 CONCERNS RELATED TO THE SHARING OF DEMOGRAPHIC DATA RELATED TO PLATFORM USERS

MITIGATION:
Demographic data (gender, age, location, language) required by this Mandate to be shared will be provided in aggregate for each respective URL, rather than all four data points in relation with a specific individual user who shared, engaged with or viewed a URL. Our recommendation is demographic data is provided as percentages, rather than raw numbers, reducing the possibility of a platform user being re-identified on the basis of piecing together those four demographic data points. This aggregate data view is comparable to:

- Facebook Ad Library[16]: displays aggregate demographic data associated with ads currently running on Facebook. This is accessible through both a web interface and through the Facebook Ad Library API. The data accessible through the API's Audience Distribution endpoint includes age, gender and region, similar to what is proposed in this Mandate.

Facebook Advertising platform[17]: Facebook advertisers use the Ads Manager platform setup and monitor their campaign performance. Within this platform, advertisers are able to view anonymised, aggregate demographic breakdowns of users who are viewing or engaging with their ads, including gender, age bracket, location and language.

[16] Facebook ad library
- Platform - https://www.facebook.com/ads/library
- Documentation- https://www.facebook.com/help/259468828226154
- Facebook ad library API - https://www.facebook.com/ads/library/api
- Graph API Audience Distribution fields - https://developers.facebook.com/docs/graph-api/reference/audience-distribution/

[17] Facebook Ad Manager platform - https://www.facebook.com/business/measurement

2 CONCERNS RELATED TO COMPLIANCE WITH THE EU GENERAL DATA PROTECTION REGULATION (GDPR)[18]

MITIGATION:
If workable, our recommendation is to collect this aggregate data within the existing scope of what is collected by the digital platforms in their privacy policy. This data is similar to that which is collected by Facebook for the Facebook Ad Library and Ads Manager tools. While a Facebook specific example, this Mandate appears to fit within the existing Facebook privacy policy[19] based on the following:

- We use the information we have (including from research partners we collaborate with) to conduct and support research and innovation on topics of general social welfare, technological advancement, public interest, health and well-being.
- We also provide information and content to research partners and academics to conduct research that advances scholarship and innovation that support our business or mission, and enhances discovery and innovation on topics of general social welfare, technological advancement, public interest, health and well-being.

If this is not workable, we recommend the relevant platforms are required to update this in their terms of use for Australian residents.

3 CONCERNS RELATED TO REDUCING PLATFORM LIABILITY FOR PRIVACY BREACHES:

MITIGATION:
When applying for API access, Live List users should be required to adhere to a code of conduct, with breaches of this code and misuse of the data resulting in liability falling on the user rather than for the digital platforms.

4 CONCERNS RELATED TO URLS CONTAINING MATERIAL OF A PRIVATE NATURE

MITIGATION:
As stated in the criteria for URL inclusion, a lower end volume for either impressions or shares should be set whereby any URL below this is not made available through the API. In addition, consultation with digital platforms should aim to uncover how else this concern could be mitigated, taking lessons from the platforms own efforts to identify and combat misinformation.

[18] EU GDPR - https://gdpr-info.eu
[19] Facebook Data Policy - https://www.facebook.com/policy.php

05. Challenges and Limitations

A number of challenges and limitations were identified during the consultation process:

1 IDENTIFYING SPECIFIC URLS SHARED ON SOCIAL MEDIA IS BOTH TECHNICALLY COMPLEX AND LEGALLY AMBIGUOUS

There are several challenges in identifying and aggregating URLs related to COVID-19 in Australia. These challenges will need to be addressed in the legal and technical development of this Mandate:

- The varied format of URLs include shortened, personalised and internal links. Concerns include:
  - » Challenges related to aggregating content associated with an individual web page but has multiple URLs
  - » The potential for personal information to be exposed through the use of URL shorteners or personal URLs

- Currently, there is a level of ambiguity of whether URLs are classified as content or non-content (metadata). It is not clear whether URLs shared in a comment or message would be regarded as the content or substance of a communication under Australian law. The ambiguity under the Telecommunications (Interception and Access) Act 1979 has implications on whether URL data can be collected without informed consent[20].

2 SOCIAL MEDIA USAGE IS MUCH MORE DIVERSE THAN ONLY URL POSTING/ SHARING

Social media users communicate via many ways beyond sharing URLs, including status updates, commenting on posts, private messages, images and memes. Tracking URLs will not accurately capture how diffusion of misinformation spreads on and across platforms, taking on many different forms.

An example of a piece of content which would be excluded from this Mandate was an image that went viral early in the pandemic of a fake QLD Department of Health media release warning against visiting certain areas with high populations of Chinese Australians. Other types of COVID-19 related content on social media have also been excluded due to:

- The infringement on user privacy of gathering this data when it is shared through private posts, messages and profiles
- Lack of analysis capabilities required to detect, analyse, catalogue this content

Whilst this is a known limitation, URLs still represent a significant type of content that is shared on these platforms, particularly with regards to the spread of false information. As such, the information that will be obtained just from URLs will be valuable.

[20] "Australia's 2020 Cyber Security Strategy – A call for views" by Dr. Stanley Shanapinda, Research Fellow LaTrobe University (Nov 1st, 2019)

3 TRANSPARENCY DOESN'T ENSURE ADEQUATE AND IMPACTFUL PUBLIC HEALTH RESPONSES

Access to this data alone does not guarantee responders have the knowledge and capacity to adequately act on this information, even if it is an important first step. COVID-19 responders, in particular Federal and State government bodies, must ensure that this resource is supported by the appropriate systems, protocols and structures to ensure utility.

4 THIS MANDATE DOES NOT HAVE IN-BUILT ANALYSIS MECHANISMS

This Mandate is focused on securing access to data from the digital platforms. As such it does not require either the regulator or the digital platforms to perform analysis such as fact-checking.

While data points which are indicators of misinformation or harmful content have been included ('reports by users', 'date it was fact checked', 'shares by suspected bot accounts') there is an assumption users would take it on themselves to further investigate the URLs as relevant to them in their work and/or research.

5 LIMITED UNDERSTANDING HOW TO BEST RESPOND TO MISINFORMATION AND HARMFUL CONTENT ON SOCIAL MEDIA

There is currently no best practice for the design of interventions to address misinformation. Little data exists on which interventions are most effective and public communication responses can be seen to do more harm than good. While we are starting to see practical and ethical guidelines emerge in newsrooms on responsible reporting, there needs to be more investment into understanding how to craft initiatives that will work to combat misinformation. A key concern with a public response is that this action will serve to further amplify the false information, potentially to new audiences or embolden conspiratorial paranoia.

This Mandate presents an opportunity to begin to measure the effectiveness of different interventions. A focus for government, academics and civil society should be to establish best practice around responding to misinformation and harmful content on social media.

6 THE ANONYMITY OF WHERE MISINFORMATION IS COMING FROM IN THE LIVE LIST SOLUTION MAY REDUCE UTILITY FOR SOME USERS

Some civil society representatives from racial justice/multicultural NGOs noted that the ability to identify perpetrators of hate speech or misinformation would be particularly useful. However, this level of identification is antithetical to our right to privacy.

Collecing data on specific individuals is out of scope of this Mandate as the potential risks associated with this breach in privacy outweigh potential benefits that might be gained from access.

Concluding Remarks

In order to have an optimal response to any crisis, it's critical to have clear and up-to-date data so that the situation may be responded to. In the face of the first global pandemic in 100 years, it's more important than ever that we have all the tools at our disposal. It is our hope that this Mandate will provide this resource and enable better analysis and a more efficient response.

06. Contributors

The following individuals were interviewed in the policy consultation process and / or participated in a round table event. A few individuals / organisations who participated in the process were not able to or did not wish to be listed.

Adam Dunn
School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney

Adel Salman
Islamic Council of Victoria

Derek Wilding
Centre for Media Transition, UTS

Erin Wen Ai Chew, Australian Asian Alliance

Francesco Bailo
School of Communication, University of Technology Sydney

James Meese
Media and Communication, RMIT University

Jennifer Stiffe
Australian Radiation Protection and Nuclear Safety Agency

Katie Attwell
School of Social Sciences, University of Western Australia

Markye Susanne Steffens
Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University

Michael Jenson
Institute for Governance and Policy Analysis, University of Canberra

Nathan Wahl
Australian Radiation Protection and Nuclear Safety Agency

Rebekah Tromble
Institute for Data, Democracy & Politics
School of Media & Public Affairs
The George Washington University

Rita Jabri-Markwell
Australian Muslim Advocacy Network (AMAN)

Stanley Shanapinda
Optus Cyber Security Research Hub, La Trobe University

07. The Appendix

‘Live list’ requirements - lo-fi prototype

OVERVIEW	The COVID-19 live list is a publicly available list of trending URLs related to COVID-19 collated across a number of social media platforms. The intention of the list is to allow for the timely identification of harmful and misleading content circulating on social media, enabling the relevant parties to respond effectively to public health and safety threats as they arise. While the list will be available publicly, the intended users are health and government officials, civil society, academics and the media
URLS SHOULD BE MADE PUBLICLY AVAILABLE IF THEY MEET THE FOLLOWING CRITERIA	» Where impressions is above XXXXX in Australia » Where the post contains COVID-19 keywords from a database (covid, coronavirus, etc.
URLS SHOULD SHARED WITH THE FOLLOWING DATA POINTS:	**REACH AND ENGAGEMENT** » Number of times they have been shared on the platform (public or private) » Number of impressions / reach (public or private) » Number of link clicks » Date first shared » Date it was fact-checked (if fact checked) **SOCIETAL IMPACT** » **Shares by public accounts e.g. brands, pages, celebrities** » **% of impressions across:** » Gender » Age » Location » Language » **% of shares across** » Gender » Age » Location (State and city level) » Language » **Number of flags / reports by users** Associated keywords identified from database
URLS SHOULD BE PROVIDED IN THE FOLLOWING FORMAT:	» Provide hourly updates to the above fields
THE FOLLOWING PLATFORMS NEED TO COMPLY	Social media platforms including » Facebook (Instagram, Whatsapp, Facebook) » Twitter » Google (Youtube) » Snapchat » Tiktok » Linkedin » Reddit

Reset.
AUSTRALIA

Reset Australia is an independent organisation raising awareness and advocating for better policy to address the digital threats to Australian democracy.

hello@au.reset.tech

au.reset.tech