New potentials in the communication of open science with non-scientific publics: The case of the anti-vaccination movement

By François Barend van Schalkwyk

Dissertation presented for the degree
PhD in Science and Technology Studies
in the Faculty of Arts and Social Sciences at Stellenbosch University

Supervisor: Prof. Peter Weingart, Research Chair: Science Communication, Stellenbosch University
Co-supervisor: Prof. Johan Muller, Extraordinary Professor, Centre for Research on Evaluation, Science and Technology, Stellenbosch University

April 2019

DECLARATION

By submitting this research assignment electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: April 2019

Abstract

There is persistent pressure on science to be more open. But for all the fervour, scant attention has been paid to the full gamut of the potentials of openness, both positive and negative. These potentials are, in many cases, linked to open access to the formal communications of science made possible by digitisation, the internet and developments in information and communication technologies. A consequence of direct access to the formal communications of science is that traditional channels of communication are no longer the gatekeepers to the public's understanding of science. Instead, new and different types of channels for the communication of science are proliferating in a society that is increasingly online and networked, and it is therefore reasonable to expect attentive non-scientific publics to access the communications of science. If this is the case, then open science introduces new trajectories in its communication that are best understood with reference to flows of information in the communication networks that define the network society.

It is the direct access to the communications of open science by non-scientists that this thesis examines in order to answer the question 'What are the potentials of open science in the communication of science?'. It does so by investigating the presence of two products of science – open research data and open access journal articles – in the online communications of a specific non-scientific community: the anti-vaccination movement. Specifically, it determines (1) whether the product is being accessed by the anti-vaccination movement as indicated by references in three online spheres (Twitter, Facebook and the web); (2) whether the product is being used by the anti-vaccination movement as indicated by the movement's level of engagement in each online sphere; and (3) whether there are intermediaries in the online communication networks of the anti-vaccination movement as indicated by mapping the movement's online communication networks centred around the products of open science.

Findings show that the anti-vaccination movement is not accessing open research data. In the case of open access journal articles, findings show that online social networks allow the anti-vaccination movement to amplify its minority position by being selective in terms of the vaccine science it feeds into its online communication networks, and by being highly active without engaging closely with the scientific knowledge at its disposal. In part, the amplification was found to be attributable to the presence of different types and a disproportionate number of intermediaries.

The consequences of the anti-vaccination movement's use of open access journal articles in its online communications is the production and amplification of uncertainty around the safety of vaccinations. Science communicators will need to develop new strategies to counter the potentially detrimental health outcomes of increases in uncertainty and vaccine refusal in the broader population. This first foray into the potentials of open science shows that the development of such communication strategies will require further research to understand better how attention, influence and power function in a society increasingly defined by its global communication networks.

Opsomming
Daar is aanhoudende druk op die wetenskap om meer toeganklik te wees. Tog vir al die entoesiasme is daar min aandag gegee aan die volle omvang van oop wetenskap se potensiaal, beide positief en negatief, van hierdie toeganklikheid. Hierdie potensiaal word in baie gevalle gekoppel aan vrye toegang tot die formele kommunikasies van die wetenskap wat moontlik gemaak word deur digitalisering, die internet en ontwikkelinge in inligting- en kommunikasietegnologie. 'n Gevolg van vrye toegang tot die formele kommunikasie van die wetenskap is dat tradisionele kommunikasiekanale nie meer die poortwagters vir die publiek se begrip van die wetenskap is nie. Nuwe en verskillende kanale vir die verspreiding van die kommunikasies van wetenskap floreer in 'n samelewing wat toenemend aanlyn is en genetwerk word. Dit is dus redelik om te aanvaar dat die aandagtige nie-wetenskaplike publiek die kommunikasies van wetenskap raadpleeg. As dit die geval is, stel dit nuwe opsies bekend vir die kommunikasie van oop wetenskap wat verstaan kan word met verwysing na die vloei van inligting in kommunikasienetwerke wat kenmerkend is van die netwerkgemeenskap.

Dit is die direkte toegang tot die wetenskaplike kommunikasies deur nie-wetenskaplikes wat ondersoek word in hierdie tesis aan die hand van die volgende navorsingsvraag: 'Wat is die potensiaal van oop wetenskap in die kommunikasie van die wetenskap?'. Die vraag word beantwoord deur die teenwoordigheid van twee produkte van die wetenskap te ondersoek – toeganklike navorsingsdata en artikels in ooptoegangvaktydskrifte – in die aanlynkommunikasie van 'n spesifieke nie-wetenskaplike gemeenskap: die teeninentingbeweging. Meer spesifiek, word daar bepaal (1) of die teeninentingsbeweging melding maak van die produk soos weerspieël deur verwysings in drie aanlynruimtes (Twitter, Facebook en die web); (2) of die produk deur die teeninentingsbeweging gebruik word soos aangedui deur die vlak van hul betrokkenheid in elk van die aanlynruimtes; en (3) of daar tussengangers is in die aanlyn-kommunikasienetwerke van die teeninentingsbeweging soos bepaal deur die beweging se aanlyn-kommunikasienetwerke se gebruik van die produkte van toeganklike wetenskap te ondersoek.

Resultate toon dat die teeninentingbeweging nie toeganklike navorsingsdata gebruik nie. In die geval van artikels in ooptoegangvaktydskrifte toon die resultate dat aanlynsosialenetwerke die teeninentingbeweging toelaat om sy minderheidsposisie te versterk deur selektief te wees in terme van die entstofwetenskap wat in die beweging se aanlyn-kommunikasienetwerke gebruik word, en deur hoogs aktief te wees sonder om in diepte aandag te gee aan die wetenskaplike kennis tot hulle beskikking. Tot 'n mate is die versterking toeskryfbaar aan die teenwoordigheid van verskillende tipes, en 'n buitensporige aantal, van tussengangers.

As gevolg van die teeninentingbeweging se gebruik van artikels in ooptoegangvaktydskrifte in sy aanlyn-kommunikasie word onsekerheid oor die veiligheid van inentings geskep en versterk. Wetenskapkommunikeerders sal nuwe strategieë moet ontwikkel om die moontlike nadelige gesondheidsuitkomste van die toenames in onsekerheid en die weiering van entstowwe in die breër bevolking teen te werk. Hierdie eerste ondersoek in die moontlikhede van oop wetenskap toon dat die ontwikkeling van sulke kommunikasiestrategieë verdere navorsing sal vereis om te verstaan hoe aandag, invloed en mag werk in 'n samelewing wat toenemend deur sy globale kommunikasienetwerke gedefinieer word.

Acknowledgements

I would like to thank the following for their contributions to the completion of this thesis:

- My laissez-faire supervisors who kept me on the straight and narrow.
- Rodrigo Costas and Jonathan Dudek from the Centre for Science and Technology Studies (CWTS), Leiden University, who provided data for the chapter on intermediaries, feedback on early attempts to interpret the data, and who did their best to demystify couplings and co-citations.
- Altmetric for the confidence and foresight to permit researchers access to its data. This thesis would not have been possible without access to the Altmetric data.
- Stephane Boyera of SBC4D for effortlessly creating and fine-tuning a web crawler over several nights in Dar es Salaam.
- Johann Mouton, Nico Cloete and the proposal committee for backing not only a dark but also an ageing horse.
- Samantha and the boys for absent Sundays and for thesis-induced amplification of baseline grumpiness.
- The universe for releasing ‘I can’t quit’ by The Vaccines as the final hurdle in the 10 000m steeple chase on a wet and windy track loomed.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

Contents

List of tables

List of figures

Chapter 1. Introduction

After an air crash, when the aircraft's flight data recorder or 'black box' as it is commonly referred to, is retrieved from the sea-bed or mountain-side, there is an expectation that the device will yield an explanation for the disaster. Usually, two black boxes are recovered: the voice recorder that captures the final words of the distressed pilots, and the flight data recorder that captures instrumentation data prior to impact. It can take months before the data are released. The reason for the delay between retrieval and release is the meticulous and iterative process of reconstructing and verifying the fateful events using both the highly technical and the emotionally-charged voice data captured by the flight recorders. It is critical from an investigative point of view that the findings of the investigative team resemble as closely as possible the actual events that unfolded.

Hypothetically, what would be the likely outcome if data from the black boxes of an ill-fated aircraft were released to the public once the devices had been recovered? Would non-aviation specialists be able to make sense of the data? Would the process of reconstructing events leading to the disaster be accelerated by expanding and distributing the labour required to analyse the data? What would be the risks that may counter any gains in the speed at which events are reconstructed with the aid of the public? And who would deliver the definitive account of the events that led to the disaster?

If we think of science itself as being analogous with a black box, then it is exactly this incipient access by the public to science that we are witnessing as its workings become increasingly transparent. The increase in advocacy for transparency and accountability, operationalised as openness and access, stems in part from a degradation of trust in public institutions (Castells, 2009, 2007, 2017; Ortiz-Ospina & Roser, 2016), including those institutions tasked with conducting research and innovating for the development of society. The breakdown of trust in institutions has also seen the rise of new public management and the escalation of quality assurance models of organisational control (Power, 1997, 2000; Taubert & Weingart, 2017). The demands for accountability through greater transparency, oversight and measurement of public institutions are buttressed by claims of beneficial returns to society (Weingart, 2012). Open science is, from such a vantage, seen as being a necessary evolution towards improvement in the efficiency, quality and contribution of science to society (Jasanoff, 2006; Leonelli, Spichtinger, & Prainsack, 2015).

At the same time, advances in technology have radically transformed the interconnectedness of society resulting in new and unprecedented modes of communication and surveillance. In politics, the potential to harvest data from social media networks and to use those same social networks to influence – through targeted communication – the outcomes of democratic processes has been uncovered.[1] In the world of finance, there appear to be investment companies targeting the reputations of large, listed public companies and using the network effects of online communication media to profit from short selling.[2] If the effects of a networked society and the new potentials it creates to disrupt politics and finance are becoming more evident, then it seems reasonable to ask what the potentials are for science.

Like the black boxes of aviation, science produces highly technical 'recordings' for an expert community. These are in the form of journal articles, conference proceedings, monographs, patents and the like. Science also generates and stores vast amounts of data that underpin scientific communication. Previously, these recordings were largely inaccessible to non-scientific communities. The open science movement advocates for open access to the communications of science, including to its raw materials in the form of open research data and to science-in-progress in the form of open access journal articles.

The communication of the sciences can be divided into two systems. The first is the communication of science within the science community. The second is the communication of science with those publics outside of the scientific community. Several developments in recent years have had an impact on the communication of science, both on the 'internal', formal science communication system as well as on the communication of science with its publics (Jasanoff, 2006; Taubert & Weingart, 2018; Weingart, 2008;).

The foundational principle of scepticism means that all scientific research and communication in the formal communication system is provisional and contested, and it is common practice for majority as well as minority groups of scientists to self-organise themselves in relation to truth claims made by their peers. As in any functioning democracy, the majority tends to hold power. Choosing, temporarily at least, not to take sides, there is invariably a group of undecideds. However, when minority groups are able to leverage new communication technologies to amplify their message and

[1] See, for example, 'The scary truth that Cambridge Analytica understands' (Tharoor, 2018) and 'Cambridge Analytica, the shady data firm that might be a key Trump-Russia link, explained' (Illing, 2018).
[2] See, for example, reports in the media on the short-seller Viceroy such as 'Controversial short seller Viceroy targets Capitec AGAIN! Read explosive letter to its auditors here' (Cameron, 2018).

garner unprecedented levels of attention in relation to their size, the likelihood of swinging the undecideds increases.

Such a scenario is perhaps less likely to play out within the scientific community because of its self-imposed system of checks and balances; a system that is self-regulated because scientists value a taken-for-granted and shared objective despite any floor crossing and factionalism: the establishment of verified truths. However, external to the scientific community, the safety net of truth-seeking falls away as publics arrange themselves into majority and minority positions around contentious social issues. The undecideds are still present and often targeted with persuasive messaging by the minorities seeking to swell their numbers; and unlike in the domain of science, the common objective to establish truth is replaced by ideological objectives which are agnostic to the norms of science. What should be of concern to science, as it becomes more open to its publics, are those non-scientific, ideologically-motivated publics who are able to access knowledge-in-progress as a central component of their persuasive tactics and attention garnering communication strategies aimed at destabilising established truths. Such risk may outweigh the benefits of greater openness to scientists across the globe who are more able to access and share scientific information. As Jasanoff (2006, p. 36) writes: ‘When claims have arrived at a certain degree of robustness, then asking for renewed scrutiny of the ways in which those conclusions were reached strikes many observers not as justifiable curiosity but as ‘manufacturing uncertainty’ for political ends. When public health and safety are at stake, such needless production of uncertainty could be not entirely frivolous but downright dangerous.’

The social dynamics of the minority versus the majority are not only confined to politics or science. These dynamics play out in everyday social contexts. The social media creates new spaces for global, real-time interaction. Unlike in ‘real-world’ or place-based social contexts, the social media allows for the amplification of minority voices to attract greater attention than would be the case in face-to-face interaction and information exchange.

An example of both the amplitude and risks made possible by social media is to be found in those online communication strategies and tactics employed by the anti-vaccination movement. On 28 February 1998, British physician Andrew Wakefield published a paper in *The Lancet* in which he claimed an empirically-proven causal link between the measles, mumps and rubella (MMR) vaccine and autism in children. Wakefield’s article triggered a public backlash against vaccinations. Reasons for the reaction amid the non-scientific community and the subsequent exponential growth of what is often referred to as the ‘anti-vaxxer movement’, can be attributed to a range of additional simultaneously occurring developments. These include an apparent increase in the number of neurological diseases in children; the increasing awareness and shifts to more healthy, organic and natural lifestyles; the number of vaccines received by children (up to 50 vaccines per child in the US); the growing awareness of the ingredients in vaccines (including, for example, aluminium and other 'non-natural' and 'non-human' substances); the apparent increase in scientific studies showing the long-term, non-specific side-effects of vaccines; the lack of response from the scientific community (at least initially) and the perceived draconian response by national health regulators (for example, by imposing penalties and legal sanctions for those refusing to have their children vaccinated); and the ability for scientific and other information to be shared continuously and instantaneously by a networked community leveraging the media's quest for attention and journalism's efforts to provide balanced reporting (Nikolau, 2016; Van den Heuvel, 2013; Zimmerman, Wolfe, & Fox, 2005).

It was eventually revealed that Wakefield had manipulated his findings as part of a scheme that promised him millions of dollars. Deer (2011), in an article published in the *British Medical Journal* titled 'How the case against the MMR vaccine was fixed', describes Wakefield's scheme: 'For example, Wakefield and his associates predicted they could make more than $43 million a year from diagnostic kits alone for a condition he argued affected autistic children dubbed "autistic enterocolitis"' (Stein, 2011). Wakefield's medical license was revoked, and his paper retracted from publication; but the damage was done.

By 2005, researchers were already aware of how the 'damage' could be escalated by online content (Zimmerman, Wolfe, & Fox, 2005) or, in Kahan's (2013) terms, how the (science) communication environment had become polluted. According to DiResta and Lotan (2015), who studied the use of Twitter by the anti-vaccination community to manipulate the passing of a vaccine bill in the State of California, '[t]his anti-vax activity might seem like low-stakes, juvenile propaganda. But social networking has the potential to significantly impact public perception of events – and the power to influence opinions increasingly lies with those who can most widely and effectively disseminate a message. One small, vocal group can have a disproportionate impact on public sentiment and legislation. Welcome to "Anti-Vax Twitter". Since anti-vax activists lose on the science and are small in number, they have increasingly begun to rely on social media to inflate their presence. […] More than any other social network, Twitter helps citizens to connect and organize in the real world even if they aren't part of the same physical communities.' Zimmerman et al. (2005) studied 78 vaccine-critical websites and found that '[w]ith the burgeoning of the internet as a health information source, an undiscerning or incompletely educated public may accept these claims and refuse vaccination of their children. As this occurs, the incidence of vaccine-preventable diseases can be expected to rise'.

While global and national average vaccination rates have remained relatively stable, there are geographical enclaves where vaccination rates for vaccine-preventable diseases such as measles, mumps, whooping cough and chicken pox have decreased (Bean, 2011; CDC, 2013; Kahan, 2014; Vanderslott & Roser, 2018). By the year 2000, measles was thought to have been eradicated in the US. In 2014, the US reported a threefold increase in reported measles cases compared to the previous year (CDC, 2015). In 2017, the World Health Organisation (WHO) declared that measles had been eliminated in the UK but in 2018 between the months of January and July, 738 measles cases were reported in England (Moten et al. 2018). Several of the outbreaks in England were linked to importations from Europe where there is evidence of an unprecedented increase in measles cases, increases which are attributed to falling rates of vaccination (World Health Organisation in Boseley 2018). Increasingly clusters of vaccine refusal are not always definable in spatial terms but can constitute geographically dispersed social clusters (CDC, 2013). Scientists warn that what may seem like negligible decreases in vaccination rates can have dire health outcomes as herd immunity is compromised (Lo & Hotez, 2017).

Of equal concern is that while vaccine rates in a country such as the US have remained stable at around 90%, the perception held by the general population is that vaccination rates are in the 70-79% range (Kahan, 2014). In countries as varied as France, Russia, Japan, Italy, Greece, Iran and Vietnam, more than 20% of the population believe vaccines to be harmful (Larson, et al., 2016). These are worrying statistics given that the herd immunity threshold for most available vaccines is higher than 80%.

Changing perceptions and behaviour do not fully account for changes in vaccination rates. Constraints in the supply of vaccinations in both the developed and developing worlds also inhibit increases in vaccine coverage (Vanderslott & Roser, 2018). Nevertheless, given the evidence available, the role of communication in shaping perceptions and amplifying anti-vaccination messaging cannot be ignored; particularly if, as the US Centres for Disease Control and Prevention suggests, 'philosophical objections' rather than supply constraints accounted for 79% of measles vaccination refusals in 2012 (CDC, 2013).

Based on the evolving relationship between (open) science and the communication of science in the social media, attention is therefore given to the unknown potentials that open science may introduce for both science and for society. The specific case of the vocal anti-vaccination movement with a keen interest in science is offered as a case through which to explore such potentials.

1.1 Structure of this thesis

This dissertation is made up of 10 chapters, including this introductory chapter as its first chapter.

In Chapter 2, the evolution of the open science movement is explored in relation to its effects on the system of science communication. Specifically, the trajectories of two products of science communication, open access journal articles and open research data, are tracked. The formal system of science communication is examined to show how norms aligned with openness have been embedded in the institution of science, and the praxis of science, from the 19th century onwards. What is relatively new is the changing relationship between science and its publishers, and expectations for greater transparency and access to science on the part of external actors such as funding agencies and attentive publics. Principal among these transformative pressures are the attention paid to science on the basis of accountability and economic interest, accelerated by developments in technology (Ware & Mabe, 2015; Van Orsdel, 2008; Taubert & Weingart, 2017).

In the second part of Chapter 2, attention is paid to how technology and open science is changing science communication; not only more traditional communication flows from science to its publics but also the use of science by non-scientific publics is increasingly taking place on social media platforms where the imperative of attention appears to reign supreme. Chapter 2 concludes by identifying gaps in the present literature vis-à-vis open science and its potentials, and by presenting the research framework, objectives and questions formulated by this thesis in response to those gaps and concerns related to a poor understanding of the potentials in a transformed science communication landscape.

Chapter 3 sets out the conceptual framework of this thesis, a framework that leans on the continuity model of science communication but elaborates the model to take into account recent developments and thinking in the science of science communication. The work of Manual Castells on the network society (1996; 2004; 2009) is used as an explanatory framework for the communication of science and is set out in the second half of Chapter 3. Chapter 3 concludes with three hypotheses formulated on the basis of the preceding sections of the thesis.

Chapter 4 outlines the research design and methods used to collect and analyse the data required to test the hypotheses and answer the research questions posed by this thesis. Included in this chapter is a justification for the selection of the anti-vaccination movement as a group of non-scientists to study the potentials of open science; the methods used to create a sample of open access research products related to the issue of vaccination and autism; methods for determining whether the anti-vaccination community is accessing and using the products of open science; and methods used to establish who the intermediaries are in the online communication networks of the anti-vaccination movement. The chapter concludes with an acknowledgement of the limitations of the methods and additional considerations to be taken into account in the use of digital methods in research.

Chapters 5, 6 and 7 constitute those chapters in which the findings are presented. Chapter 5 focuses on the access and use of open research data by the anti-vaccination community, while Chapter 6 focuses on the same for open access journal articles. Chapter 7 turns its attention to the findings related to intermediaries in the communication networks of the anti-vaccination movement.

Chapters 5 through 7 restrict themselves to setting out the findings. It is in Chapter 8 that those findings are discussed in relation to the questions and objectives of this thesis. The discussion proceeds according to several themes that emerge from the findings including the production and amplification of uncertainty in online communication networks using the openly accessible products of science. Chapter 9 extends the discussion to the implications of the communication of science in an attempt to make a contribution to the science of science communication.

Chapter 10 concludes the thesis by returning to the original hypotheses and research questions. The chapter provides a concise encapsulation of the conclusions that can be drawn, both for theory and practice, based on the empirically-grounded insights gained from the research.

Additional data in support of Chapters 4 to 7 are provided at the end of the dissertation in two appendices.

1.2 A note about sources

One of the challenges of writing a thesis over a period of four years on a topic that is constantly evolving, is the shortage of published scientific literature on the topic. It is the nature of scientific publishing that books and journal articles will endure a gestation period of at least a year or two, if not longer. And this does not factor in the time to write up findings once the research has been completed. The result is a limited number of contemporaneous published books and articles providing empirical results to inform the research in this thesis with its particular interest in online communications as they relate to the communication of science.

At the same time, there is no shortage of articles in the grey literature, including in popular and professional periodicals, eager to tap into what is a media-worthy topic. Scientists, in attempts to circumvent the protracted publishing process, are also increasingly publishing articles in pre-print format to contribute to fast-moving contemporary debates and social issues.

Where relevant scientific publications in the form of journal articles, book and book chapters exist, those have been consulted; but this thesis has also had to rely on pre-prints, the grey literature and popular sources to be in tune with the most recent developments in the area of the social media and its potentials for the communication of science.

Chapter 2. Openness and science communication

2.1 Towards a more open science

Open science is an umbrella term that describes an approach to scientific research in which the inputs and outputs of the research process are shared openly and not only within the academic community but also with non-scientists such as the public, government and the business sector, among others. In open science, the processes and objects created by scientists in the pursuit of new knowledge are made universally accessible using enabling internet-based tools and platforms (Bartling & Friesike 2014; Friesike et. al. 2015). According to the OECD (2015, p. 9), '[o]pen science is the encounter between the age-old tradition of openness in science and the tools of information and communications technologies (ICTs) that have reshaped the scientific enterprise and require a critical look from policy makers seeking to promote long-term research as well as innovation.' The European Union-funded consortium Fostering the Practical Implementation of Open Science in Horizon 2020 and Beyond (FOSTER), defines open science as 'the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods' (FOSTER, 2017). Fecher & Friesike (2013, p. 1), in an attempt to define open science following a review of the literature, settle on 'Open Science can refer to [pretty] much anything: The process of knowledge creation, its result, the researching individual, or the relationship between research and the rest of society'.

Open science is as concerned with the openness of the process of science as it is with the products of science being openly accessible. The impact of an interest in the openness of the scientific process is two-fold. First, there is an expectation that the raw inputs and objects that are generated in the production of knowledge should be open. Examples of such inputs and objects include the data from laboratory experiments, trials or field research; laboratory notes; computer software; measuring equipment; presentations, papers and articles shared between scientists for validation; etc. Second, the interest extends into the constituent components of the process and products of science. Access to and categorisation of a journal article is, for example, broken down into access and labelling of text, tables, illustrations and supplementary material. Each constituent part is assigned its own digital object identifier (DOI) and platforms are created that focus solely on the publication of constituent parts (e.g. Figshare, Slideshare, Dryad and the like).

Open science is therefore an amalgam of openness across the scientific endeavour. For each stage in the process and for each category of objects created in the pursuit of scientific knowledge, a particular type of openness is demarcated: open access, open lab notes, open data, open peer review and open educational resources attest to this demarcation of openness in science.

From a historical perspective, Eamon (1985) argues that there was a progressive shift from a more secretive to a more public science from the 17th century onwards, accelerated by the disruptive technology of the printing press and a concomitant reaction against hierarchical and monopolistic knowledge systems. Following, among others, the influence of science reformers such as Bacon and Hartlib; the establishment of Théophraste Renaudot's *Bureau d'adresse* in Paris in 1633 and the crown-sanctioned establishment of the Royal Society of London in 1662; and the publication of the *Philosophical Transactions* in 1665, the institutional mechanisms that would that would govern science as a form of 'public knowledge' were in place. According to Eamon (1985, p. 346), 'the ideal of public knowledge was not taken to imply then – any more than it does today – that everyone had perfectly free access to scientific knowledge. Nevertheless, the institutionalisation of science under the auspices of the Baconian programme helped to confirm the scientist's special role in society, not as the guardian of secret knowledge, but as the purveyor of new truths bearing the authority of experimental evidence. Free communication within the scientific community became the norm'.

By the mid-20th century, sociologist Robert Merton (1973), reacting in part to the secret science of a dictatorial Nazi Germany, had proposed four norms (or ideal principles) guiding the social behaviour of scientists, one of which, the norm of communalism, dictates that the results and discoveries of science are not the property of the individual researcher but belong to the scientific community and to society at large. More recently, with the rise of the information age predicated on digitisation and information communication technologies, the discourse around 'openness' has predominantly been in opposition to the extractive and restrictive positioning of knowledge as a private good (Boyle, 2003; Chan & Costa, 2005). The opposition is based on the simple premise that the sharing and reuse of the products of science is less dependent on the services offered by intermediaries such as publishers. Proponents of open science have emerged in opposition to the 'enclosure' of research data, publications, software and processes, or at least to their control by third parties, and advocate instead for their reuse without the impediments of cost and permissions (Evans, 2005).

However, while the open science movement mobilised with transformative intentions, it is not immune to commercial interests (Lawson, Gray, & Mauri, 2016; Taubert & Weingart, 2017). As a result, there is a counter-movement towards utilitarian and instrumentalist 'openness', with less of a focus on the potential of openness for the advancement of science, and an increased emphasis on business models designed to mine openness and extract material value (Taubert & Weingart, 2017). But the economisation of open science is not necessarily evenly distributed or applied across all of its products. Some products are more lucrative than others, and the ability to extract value will most likely depend on the relative positions of power held by social actors invested or interested in any particular 'product' of science. As Fecher and Friesike (2013, p. 8) point out: 'the argument for open research data is one that is researcher-centric while argument for open access is less so and leans more towards access as a human right and its potential contribution to development'. Their suggested five schools of thought in open science also clearly shows that different configurations of actors (scientists, politicians and citizens) have vested interests in each of the schools. That they do not include actors such as corporations (e.g. Google, Microsoft and the like), research funding agencies or government as equally important stakeholders in open science does not detract from the fact that multiple and varying arrangements of stakeholders are active, depending on the specific stream or sub-domain of open science.

In the following two sections (2.3 and 2.4), two products of science are explored in greater depth: open access scholarly publications and open research data. This follows section 2.2 in which consideration is given to the norms and values in science as it is these norms and values that are important in our understanding of the potentials of open science.

Both sections 2.3 and 2.4 trace the historical trajectories of each product to surface the actors and their motivations for a more open science. Section 2.3 suggests a reactivation of institutional norms particular to science; a reactivation triggered by a breakdown in a critical component of the science communication process – publishing – attributable to a clash in the guiding logics of science and the market which, in turn, can be traced to digitisation and information communication technologies. Section 2.4 suggests that pressures for the publication of open research data originate mainly from the perceived economic benefits of open access to research data but that the outcome is the same as for open access scientific publications: open science accompanied by an interplay of conflicting norms and institutions and their communication networks overlap. These differences and convergences are explored briefly in section 2.5 before section 2.6 concludes the discussion on open science.

Section 2.7 explores the relationship between developments in information communication technologies and the communication of science. In particular, the section gives consideration to how access to the formal system of science communication is made possible by new technologies. Section 2.8 then goes on to explore some of the potentials of open science, specifically linked to universal access to open access scholarly publications and open research data. Following the review of the literature conducted in sections 2.2 to 2.7, gaps in the literature are articulated, the framing of the research is presented, and the research questions posed are set out in sections 2.9, 2.10 and 2.11 respectively.

2.2 A reassertion of communism?

The external pressures on science from the state and society in the form of demands for greater efficiency, relevance and transparency may be the direct consequence of the times we live in but being buffeted by external pressures is not a novel state of affairs for science as an institution. Institutions respond strategically, adapt and change at a guarded pace (Scott, 2014). What makes the pressures for open science particularly remarkable is that the demands for openness are not limited to pressures that emanate from outside of science; there is increasing pressure from within the institution for a more open science. This is particularly so in the case of open access to the scientific publications which has been catalytic in the broader open science movement. Whether the interest in open research data has an equal level of support is less clear as is shown in the section on open research data that follows.

2.2.1 Norms in science

In the theory of institutions, there is a group of theorists who place emphasis on the normative rules that prescribe behaviour in social systems. Within such social systems '[n]orms specify how things should be done; they define legitimate means to pursue valued ends' (Scott, 2014, p. 64). Robert Merton, widely regarded as the father of the sociology of science, proposed a normative structure of science to account for the behaviour of scientists (Merton, 1973). Merton's norms of science emerged from a school of thought referred to as the 'institutional sociology of science' pointing to the inherently organisational approach adopted by Merton, in particular his understanding of the institution as a socially constructed, self-regulated organisational form within which actors follow norms as taken-for-granted scripts that shape the behaviours of those institutionally constrained actors: 'The key was to analyse science *also* as a social institution. As an institution, science must have norms – just as political, economic and religious institutions cannot exist without norms' (Merton in Bucchi, 2004, p. 22). In a social system that is largely autonomous (science has carved out a monopoly over the validity of truth claims), self-regulated and dependent on voluntary participation in many of its core activities (e.g. participation in the assessment of the discoveries of peers), norms are particularly salient and necessary in binding together the community of scientists (Anderson, Martinson, & De Vries, 2007).

Merton proposed four institutional imperatives or norms: (1) universalism; (2) communism; (3) disinterestedness; and (4) organised scepticism (see Table 1) (Merton, 1973). The focus in this part of the thesis is on the norm of communism or 'communalism'[3] for its obvious relevance to open access publishing and open research data as mechanisms for making knowledge accessible to all. However, the norms of universalism, disinterestedness and of organised scepticism equally contribute to a science communication system designed to meet the common objective in science of producing truths about the world we inhabit (Jamieson, 2017). The importance of these norms becomes more apparent in later sections of this thesis where it is suggested that the impotency of the norms of science in other communication environments affects directly the potentials of open science.

Communalism is a normative position which holds that results and discoveries are not the property of any individual scientist but belong to the scientific community *and* to society at large. This imperative is based on the supposition that the process of knowledge creation is cumulative. In other words, a truth claim, or a new discovery, results not only from individual effort but from the cumulative efforts of a community of scientists. Linked to this imperative is the fact that the individual scientist accumulates no recognition for his/her truth claim or discovery unless the scientist publicises his/her discoveries and, in so doing, makes it publicly accessible. As Merton (1968, p. 611) states: 'The institutional conception of science as part of the public domain is linked with the imperative for communication of findings. Secrecy is the antithesis of this norm; full and open communication is its enactment.'

And yet secrecy is not uncommon in science. As Mitroff (1974) points out, Merton's norms as motivation for scientific behaviour make sense when considered within an institutional context, but less so when considered in the context of personal or individualistic motivations of scientists.

[3] For the purposes of this discussion, the term 'communalism' will be used in place of 'communism' due to the indelible pejorative association that the latter term has with a particular system of social organisation.

Table 1: Norms and counter-norms in science

Norm	Counter-norm
Universalism Scientific claims and finding are judged independently of the personal or social attributes of their proponents.	***Particularism*** A scientist's personal and social characteristics are factors which influence how his or her scientific work will be judged.
Communalism Results and discoveries are not the property of the individual researcher but belong to the scientific community and to society at large.	***Individualism*** Property rights are extended to include protective control over results and discoveries.
Disinterestedness Scientists pursue their primary aim of advancing knowledge and indirectly receive individual rewards for doing so.	***Interestedness*** The individual scientist seeks to serve his or her own interests and those of the restricted group of scientists to which he or she belongs.
Organised scepticism Every researcher is obliged to scrutinise every truth claim carefully, including his or her own, suspending final judgement until the necessary confirmations become available.	***Organised dogmatism*** The scientist must believe in his or her own findings with utter conviction while doubting the findings of others.

Source: Adapted from Mitroff (1974) in Bucchi (2004, p. 18)

2.2.2 Normative dissonance within science

Merton's explanation has been criticised for presenting an ideal state which is seen as prescriptive rather than as descriptive of how science functions in practice (Mitroff, 1974). Ideal-type norms provide explanatory value when considering the impersonal character of science but fall in value when considering the personal character of science (Mitroff, 1974). The concern with the tension between an ideal, impersonal science and the everyday behaviour of individual scientists, led Mitroff to suggest a set of counter-norms to Merton's original normative structure. In opposition to the norm of communalism in science, Mitroff posits the counter-norm of 'individualism'. Individualism as a normative position in science holds that property rights (ensuring secrecy and/or restricted dissemination) are deployed by scientists to exert protective control over results and discoveries (Mitroff, 1974).

In an attempt to reconcile norms and counter-norms as the normative basis for explaining scientific behaviour, it was suggested that both sets of imperatives be regarded as flexible 'ideological-rhetorical' repertoires from which scientists may draw in order to make sense of their actions in particular contexts. For example, while secrecy may be condemned in one situation, it may well be valued in another. Mitroff (1974, p. 579) goes as far as to suggest that the tension between scientific norms and counter-norms may in fact be 'necessary for existence and ultimate rationality of science'.

Merton himself proposed the concept of 'sociological ambivalence' (1976) to describe the often diverse and conflicting values and norms that scientists must navigate and draw on to give legitimacy to their behavioural choices as they 'do' science. Anderson et al. (2007) show from their study of over 3 000 scientists that substantial normative dissonance exists between norms regarded as ideal and scientists' perceptions of the behaviour of other scientists, and that the behaviour of other scientists was perceived to be more counter-normative than normative. Mitroff suggests that, based on the reality of sociological ambivalence in science, it may be informative to distinguish between 'well-defined' and 'ill-defined' problems when evaluating the dominance of norms or counter-norms. In the case of 'well-defined' problems in science, Mertonian norms of science are likely to be dominant whereas in the case of 'ill-defined problems', counter-norms are likely to be dominant because ill-defined problems put emphasis on the personal traits and behaviours of their formulators (Mitroff, 1974).

2.2.3 Science logics and values

Merton's norms and Mitroff's counter-norms activate within the domain of science, whether it be at the collective or the individual level. And as Jasanoff (2012, p. 102) states: 'Much of the authority of science […] rests as well on its success in persuading decision-makers and the public that the Mertonian norms present an accurate picture of the way science really works'. Bimber and Guston (1995) refer to such claims resting on the specialness of science as the 'sociological exceptionalism' of science. And Merton himself opens his essay 'The Normative Structure of Science' by pointing out that '[s]cience, like any other activity involving social collaboration, is subject to shifting fortunes. Difficult as the notion may appear to those reared in a culture that grants science a prominent if not commanding place in the scheme of things, it is evident that science is not immune from attack, restraint, and repression' (Merton, 1973, p. 267). Science exists and functions within a complex arrangement of inter-institutional relations, incorporating numerous value spheres each with its own 'institutional logic' (Friedland & Alford, 1991; Thornton & Ocasio, 2008): 'an important source of the institutional tension and change experienced by both organisations and individuals in everyday life involves jurisdictional disputes among the various institutional logics' (Scott, 2014, p. 91). Olsen introduces the term 'logic of appropriateness' to capture the normative referencing of socially constructed behaviour (Olsen, 2008).

For Friedland and Alford (1991), the core institutions of society are the capitalist market, the bureaucratic state, families, democracy and religion. Based on a review of empirical studies, Thornton and Ocasio propose a typology of six institutional sectors as a reworking of those proposed by Friedland and Alford: markets, corporations, professions, states, families and religion (Thornton & Ocasio, 2008). While Friedland and Alford identified Christianity as a central logic of the West, Thornton and Ocasio (2008) generalize the logic type to 'religion', thereby at least implicitly suggesting the universality of institutional logics. However, religions are not replaceable systems; they may differ vastly, fulfil different functions and have different relationships to other institutional logics. For the purposes of this thesis, the important question arises of why there is no logic of science as a central institutional logic of modern society. A logic of science could possibly be captured by the logic of professionalism or perhaps that of 'community' – an institutional logic added by Thornton et al. to the six originally proposed (Thornton, Ocasio, & Lounsbury, 2012). However, the point remains that the central logics appear fuzzy, arbitrary and incomplete (Wenzlaff, 2014). For the purposes of this enquiry, it is therefore presumed that the pursuit of knowledge for the purpose of establishing truth is the central logic of science and, in this sense, science is brought closer to the core institutional logic of the profession which for Thornton et al. (2012) alludes to the primacy of the value of 'personal expertise' and Friedland the value of 'knowledge':[4]

> Institutional logics, I would argue, return us to that element from which the new institutionalism fled: value. […] [V]alue is central to an institutional logic: a presumed product of its prescribed practices, the foundation stone of its ontology, the source of legitimacy of its rules, a basis of individual identification, a ground for agency, and the foundation upon which its powers are constituted (Friedland, 2012).

Merton himself was attuned to the importance of values in his seminal essay. He refers to the 'ethos' of science: 'that affectively toned complex of *values* and norms which is held to be binding on the man of science' (Merton, 1973, pp. 268-269). The norms of science are legitimised *in terms of* institutional values, and the core institutional value of science is the production of truth. This value is similar to Friedland's logic of knowledge in the core institution of the professions, with the important caveat that in science, knowledge must be verified by peers from within science for it to hold any value.

What emerges from the above synopsis of norms and values in science as an institution is an apparent differentiation between norms in science and the values of science. And it is possible that there are correlations between problems *in* science and the norm/counter-norm and

[4] It is beyond the scope of this thesis to enter into a theoretical debate as to the distinctiveness of science as a central institutional logic.

impersonal/personal dichotomies, and also between problems *of* science and the core values (logics) of science. Open science, from this perspective, may therefore not only be about the norms in science (the norm of communalism in particular) but about activating the core values (or logics) of science.

2.2.4 Encounters between science, the state and the market

In the recent history of science, there have been several occasions in which a meeting of different and potentially dissonant values – institutional logics – between science and other institutional domains have taken place.

Pinker (2002) describes the politicisation of science when it turns its attention to an examination of human nature. Dreger (2015) chronicles the clash between science and culture in her own work on intersex treatment and rights. She also writes about similar clashes between science and culture in other controversies, including those sparked by the publication of *A Natural History of Rape* in which authors Palmer and Thornhill argue that rape should not be de-sexualised; the accusations of human rights atrocities allegedly committed by anthropologist Napoleon Chagnon against the Yanomamo of Brazil;[5] and Rind, Tromovitch and Bauserman's published research (commonly referred to as the 'Rind Paper') showing that children are not, on average, as universally devastated by sexual abuse as is held by popular belief. The Rind Paper is the only scientific product to be condemned by an act of Congress in the United States – a stark example of how politicians align with cultural values in an encounter with science.

In the period following World War II, politicians came to realise the value of science for both development and political power (triggered in the West by the 'Sputnik Effect'). Committees were formed to provide expert advice to governments on science policy. As Elzinga and Jamison (1995, p. 582) point out: 'A new breed of experts came into being, bridging the values and norm systems of the state and the academy'. Thus, with the emergence of science policy advisors, we witness an encounter between the values of science and those of the state in the 'politics of science' (Jamieson, 2017; Scheufele, 2014). The increased interest in science for its political and development value transformed science. Investment in science in the post-war period increased, driving its growth – more scientists, more universities, more students, more journals, more patent applications and the creation of science councils and other funding structures (Bucchi, 2004; Elzinga & Jamison, 1995).

[5] This accusation against Chagron has been refuted. See, for example, Pinker's (2002) exposition of the ideological motivations behind the accusations levelled at Chagron and their unfounded empirical basis.

This growth would ramp up the value of the formal science communication system and its products to capital markets resulting in increased levels of interest and activity from large commercial corporations. The intersection of science, the state and the market would ultimately change the relationship between science and a critical cog in the system of knowledge production: scientific publishing.

2.3 Open access

Science as an enterprise in the production and consumption of knowledge is unique in several respects. It is unique in that the producers of knowledge are the same as the consumers of knowledge. This is not dissimilar from subsistence agriculture but, in the case of knowledge production, there is a commercial component in a more complex value chain. Knowledge producers outsource the dissemination of newly produced knowledge to publishers. And, in another unique twist, the knowledge producers retain control over the quality assurance process to ensure that the value of the knowledge produced is maintained.

2.3.1 A pact between science and publishing

What seems apparent from this distinctive arrangement in the production and dissemination of knowledge is that a pact or social contract was established between science and the publishing industry. How this arrangement came about is most likely the result of the emergence of the scientific journal in the 17th and 18th centuries in England and Europe (Eamon, 1985; Fyfe, McDougall-Waters, & Moxham, 2015).

In the period prior to the establishment of the scientific academies in the mid-17th century, much of the non-oral communication between scientists was via the exchange of letters (McKie, 1979). The invention of the printing press and increasing levels of education, led to the wider dissemination of scientific knowledge as books became the primary mode of scholarly exchange (Eamon, 1985). The establishment of the Royal Society of London in 1662, the *Académie des Sciences* in Paris in 1666, and other scientific societies was centred on the witnessing of experimental performances among a community of natural philosophers. These academies printed accounts of their discussions in journals such as the Royal Society's *Philosophical Transactions* (Kronick, 1976). In contrast to scientific books, these scientific societies and their periodical publications bound the spread of knowledge to its accreditation by a community of scholars (the prototype for the current system of peer review). Thus, the societies became the gatekeepers for codified scientific knowledge, and scientific journals became the primary means of scholarly communication.

In the period between 1750 and 1798, 69 new scientific periodicals were published (McKie, 1979). According to Levere (1981) , there were approximately 750 scientific periodicals in 1800, and several thousand by 1850. A significant development in the 19th century was not only the continued growth in the number of scientific journals but the fact that most of the journals were no longer coupled to the scientific societies but rather to commercial publishers (Meadows, 1980). Brock (1980) provides empirical support for this claim: of 535 British scientific periodicals published between 1824 and 1900, 64% were commercial.

Meadows (1979) points out that historians have opted by and large to focus on the establishment and early history of the scientific period, resulting in a lacuna for the 150-year period between the 1800s and the end of World War II, thus making it difficult to trace the entry of commercial publishing into scientific endeavour. There are, however, some indicators of how the dissemination of science shifted from the societies to the commercial publishers. McKie describes Lorenz Crell, founder of the first specialised scientific journal (in Chemistry) and many other journals, as a 'tireless correspondent and editor' (McKie, 1979, p. 15); and Henry Oldenburg, one of two secretaries of the Royal Society and private financier of the Society's *Philosophical Transactions*, as having 'carried on an extensive foreign correspondence' (McKie, 1979, p. 10). Oldenburg has been described as serving the function of clearing house for European science (Levere, 1981). Therefore, it could be that as more scientific knowledge was published in books and journals, and as the dissemination of knowledge increased between the scientific capitals of Europe, editors and secretaries could no longer adequately meet demand. Commercial publishers with more efficient means of distribution and more competitive printing costs resultant from economies of scale, were well-placed to meet the growing demand for science publications in Europe. As Meadows (1979, p. 5) writes, during 'the latter part of the nineteenth century […] [t]he problem was that the amount of literature, particularly in chemistry, was growing beyond the point where leisurely handling methods of the period could function efficiently'. And Levere (1981, p. 157), in his review of *The Development of Science Publishing in Europe*, writes: 'The commercial journals offered faster publication than the sponsored ones; they gave intelligence of foreign scientific journals; and generally kept the scientific societies on their toes.'

However, at this phase in the history of scientific publishing, commercial publishers were still probably 'motivated more by prestige than by profits' (Strasser & Edwards, 2015). According to

Meadows (1979, p. 21), even commercial journals may have been published for extended periods of time despite incurring losses for the publisher. The scientific journal *Nature* did not return a profit in the thirty years after the first issue was published in 1869. Such a prolonged period of losses would not be tolerated by contemporary commercial journal publishers. After World War II, however, scientific publishing became a more lucrative pursuit. The renewed political interest in science following the War was accompanied by increases in science funding and in the growth of the profession. And in the 1950s and 1960s, as higher education systems in much of the developed world massified, commercial publishers, attuned to the profit potential of an expanding and stable higher education market, entered into academic publishing (Strasser & Edwards, 2015).

This move lead to what Thornton and Ocasio describe as a clash within scientific publishing between competing editorial and market 'logics' (Thornton & Ocasio, 1999). The motivation on the part of publishers to participate in science for the sake of professional prestige is overtaken by a motivation to exploit the science market to make profit. By the 1990s, economies of scale and increased profits enabled commercial publishers to offer more efficient editorial management systems and broader distribution of scientific publications. This development resulted in many of the scholarly societies (who were directly managing the editorial and distribution components in the production of their journals) to outsource to the large, commercial enterprises. According to a study conducted in 2012, 64% of all academic articles were published by commercial publishers, 30% by non-profit societies and 4% by university presses (Ware & Mabe, 2015).

The rapid digitisation of the publishing industry and the rise of the internet was to have a profound impact on the cost of production and distribution of scientific publications. At first, scientists themselves exploited the benefits of digitisation and the world-wide web to explore new models of disseminating knowledge. The late 1980s saw the emergence of the first electronic journals, and e-print servers were set up by physicists in the early 1990s. But commercial publishers were also exploiting the opportunities made possible by a digital world and by 2005, studies showed that between 62% and 90% of journals were available online (depending on the methods used by the study) (Mark Ware Consulting, 2006).

At the same time, there was a drop in the subscription of print journals, and publishers responded by pushing up prices. As profit margins were squeezed, consolidation began to take place, and by 2013, the top five most prolific publishers accounted for more than 50% of all papers published (Larivière, Haustein, & Mongeon, 2015). According to *The Economist*, Elsevier, the largest of the big five was 'in rude financial health. In 2010, it made a £724m ($1.16 billion) profit on revenues of £2 billion, a margin of 36%' (Economist, 2016). Larivière et al. (2015) show how Reed-Elsevier's Scientific, Technical and Medical division's operating profits have risen consistently in the period 1991 to 2013, reaching $2 billion in 2013, while profit margins have remained consistent within a 30-40% band for the same period. In comparative terms, such profit margins put Elsevier 'on a comparable level with Pfizer (42%), the Industrial & Commercial Bank of China (29%) and far above Hyundai Motors (10%), which comprise the most profitable drug, bank and auto companies among the top 10 biggest companies respectively' (Larivière, Haustein, & Mongeon, 2015, p. 10). On average, the profit margins of commercial scientific publishers are an estimated 35%, compared to 20% for society publishers and 25% for academic presses in the late 2000s (Swan & Brown, 2008).

2.3.2 The rise of open access

In December 2001, the Budapest Open Access Initiative (BOAI) launched a global campaign for open access science publishing. According to the Initiative: 'It didn't invent the idea of OA [open access]. On the contrary, it deliberately drew together existing projects to explore how they might "work together to achieve broader, deeper, and faster success." But the BOAI was the first initiative to use the term "open access" for this purpose, the first to articulate a public definition, the first to propose complementary strategies for realising OA, the first to generalise the call for OA to all disciplines and countries, and the first to be accompanied by significant funding' (Budapest Open Access Initiative, 2012). It defines open access as follows:

> By 'open access' to [peer-reviewed research literature], we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited (Budapest Open Access Initiative, 2012).

The Budapest Initiative was bolstered with the founding of the Public Library of Science (PLOS) in 2000 by an advocacy group of scientists led by the biochemist Patrick O. Brown and the biologist Michael Eisen.

In 2003, the Max Planck Society and the European Cultural Heritage Online project convened a gathering of international experts with the purpose of capitalising on the borderless, non-material dissemination opportunities made possible by the internet and digital content by developing a new web-based research environment. It promoted an 'open access paradigm as a mechanism for having scientific knowledge and cultural heritage accessible worldwide' (Berlin Declaration on Open Access, 2003). The outcome of the meeting was the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Berlin Declaration on Open Access, 2003). In the same year, the Public Library of Science group launched *PLOS Biology*, its first open access, non-profit journal.

The Registry of Open Access Repository Mandates and Policies (ROARMAP) is an international registry charting the growth of open access mandates and policies adopted that require or request their researchers to provide open access to their research outputs. By type of policy-maker, ROARMAP's registry indicates the following number of open access policies (Registry of Open Access Repository Mandates and Policies, 2017):

- Funder 79
- Research organisation (e.g. university) 557
- Funder and research organisation 54
- Multiple research organisations 9
- Sub-unit of research organisation 70
- TOTAL 769

The number of open access journals has increased by more than 900% from 2000 to 2011, with the average number of articles doubling (Laakso & Bjork, 2012). As of April 2014, more than 50% of the scientific papers published annually from 2007 to 2012 could be downloaded for free from the internet (Archambault, et al., 2014). The finding from a 2017 study that 45% of all scholarly literature in 2015 is open access, is in a similar range (Piwowar, et al., 2017).

About 800 (7%) of the approximately 11,000 journals included in Journal Citation Reports, are open access; and of the 21,500 journals tracked by Scopus, approximately 3,500 (16%) are open access titles (Kramer, 2017). In 2017, the Directory for Open Access Journals (DOAJ), established in 2002 with 34 journals, listed 2,538,235 articles from 9,645 journals across 123 countries in its directory (Directory of Open Access Journals, 2017).

2.3.3 Ruptures between science and its publishers

In response to the American Anthropological Association's (AAA) call for proposals to publish the Association's 22 publications when its contract with Wiley-Blackwell would come to an end in 2017 (Waterston, n.d.; Heller, n.d.), academics from within the Association submitted a proposal centred on an open access, co-operative publishing model (Raikhel, 2015). The argument put forward was that there are sufficient resources with the academic system that can be mobilised to produce and disseminate the AAA's publications without having to rely on a commercial publisher, without compromising on quality and without having to limit universal access to the Association's publications (Raikhel, 2015). The authors of the proposal submitted to the AAA state: 'This model is also deeply attuned to our scholarly values because it is collaborative, communal, and cooperative' (Jiménez, Willinsky, Boyer, & Da Col, 2015).

The call to rethink the role of commercial publishers in the production of science does not appear to be a solitary call from a single academic field. From the other side of the disciplinary spectrum, there is a similar plea:

> There are signs that researchers might be tiring of the grip that the elite journals have on the biomedical sciences. Over 12,000 people have so far signed the San Francisco Declaration on Research Assessment (DORA), which began in 2012 as a commitment for research to be assessed 'on its own merits rather than on the basis of the journal in which the research is published' (Economist, 2016).

While the underlying frustration here is what is believed to be a misguided and detrimental practice of assessing the quality of research by relying on the journal impact factor as a metric rather than a frustration with academic publishing per se, it is nevertheless the commercial journal publishers who control (that is, restrict) access to academic journals. By doing so, the scope for citations is corralled by the commercial publishers. Of the 82 original signatories of DORA (2012), none were commercial publishers, and, at the time of writing, none could be found in the list of the 600+ subsequent signatories.[6]

[6] Signatories of DORA are typically academic institutes, learned societies, research or funding councils, universities and/or their libraries, and open access publishers. *The Journal of Cell Science* is among the first signatories, but it is a not-for-profit publishing organisation managed by practicing scientists who endeavour to profit science, not shareholders. FEBS Journal also signed the original declaration, but as a journal; the journal's publisher, Wiley, has not signed the Declaration.

Other examples of increasing resistance to the existing publishing model in science and of an increasing rift between scientists and their publishers abound.

On 21 January 2012, Timothy Gowers, a Fields medallist in 1998, published a blog post in which he makes public his boycott of research journals published by Elsevier (Gower, 2012). Elsevier owns more than 2 000 journals, including *Cell* and the *Lancet*. Dr Gowers has emboldened others to do the same by signing a pledge on a website called 'The Cost of Knowledge' (The Cost of Knowledge, n.d.). At the time of writing, 15 977 researchers had signed the online pledge. Among the objections in the pledge are the high prices charged by commercial publishers for subscriptions to their journals and that commercial publishers such as Elsevier restrict the free exchange of information: 'The key to all these issues is the right of authors to achieve easily-accessible distribution of their work' (The Cost of Knowledge, n.d.).

In 2006, the entire editorial board of *Topology*, a mathematics journal published by Elsevier, resigned, citing concerns about high prices restricting access to knowledge (Economist, 2016). The Editorial board of *K-theory*, a mathematics journal owned at that time by Springer, a German publishing firm, quit in 2007 (Economist, 2016). More recently, in 2015, all 6 editors and all 31 editorial board members of the journal *Lingua*, a top journal in linguistics, resigned in protest of Elsevier's policies on pricing and its refusal to convert the journal to an open-access publication (Jaschik, 2015). Again, cost and limits placed on the exchange of information are commonly cited as the motivation for these acts of protest.

Some open access initiatives not only pit academics versus publisher but attempt to bring together multiple actors in the knowledge creation process. CERN (European Organization for Nuclear Research), for example, has embarked on an ambitious plan to make all its research in particle physics open access, with a global collaboration inspired by the way scientists came together to make discoveries at the Large Hadron Collider. The initiative is called SCOAP or, to give it its full name, the Sponsoring Consortium for Open Access in Particle Physics Publishing. SCOAP was initiated in 2014 and consists of a worldwide collaboration of more than 3,000 libraries, funding agencies, research centres in 44 countries and three intergovernmental organisations. In its first three years, SCOAP made more than 13,000 open access articles available (Barbour, 2017).

When the commercial publishers of science are perceived to be 'robber capitalists' (Taubert & Weingart, 2017) it invites others to reinterpret the illegal sharing of the objects of science as morally justified. This has led to the emergence of a more militant Robin Hood-like form of resistance to the knowledge monopoly of the commercial publishers. Open access guerrillas raid the stores of the commercial scholarly publishers and make their loot publicly available online. Examples include Library Genesis (Library Genesis, n.d.) and Sci-Hub (Sci-Hub, n.d.). Indignant students for their part, with limited access to monographs and journals, are only too happy download pirated academic content (Czerniewicz, 2016), and it is likely that a younger generation of more tech-savvy academics are following suit. Aaron Swartz, a leader of the guerrilla open access movement advocated as follows:

> We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access. With enough of us, around the world, we'll not just send a strong message opposing the privatization of knowledge — we'll make it a thing of the past. (Swartz, 2008)

After downloading thousands of MIT’s pay-walled journals and making them freely available, Swartz faced 35 years in prison following the actions of federal and university lawyers. Aaron Swartz committed suicide at the age of 26.

From this abridged history of scientific and open access academic publishing, three observations can be made. The first is the shift in the primary drivers behind the call for open access: from a civil society organisation (the Open Society Foundation) in the case of the Budapest Initiative in 2001, to an organisation of multi-disciplinary autonomous scholarly institutes (Max Planck Institutes) in the case of the Berlin Declaration, to discipline-specific and university-member societies (the American Anthropological Association and the American Society for Cell Biology) to high-profile and self-organising groups of academics. This could be seen as a shift in the support for open access that lies closer to the academic heartland.

The second is an acknowledgement that while the academy is increasingly involved in the debate on the benefits and challenges of open access scientific publishing, change will be slow and measured. Perhaps frustratingly so for the open access advocates located outside of the academy. As Bohannon writes:

> We would need a real revolution. By revolution I mean a cultural revolution among academics. They would have to totally change the way they do business and, despite

> having the reputation of being revolutionary, academics are pretty conservative. As a culture, academia moves pretty slow [sic] (Bohannon, 2016).

The third observation is the progression from promoting open access as a model that unlocks the benefits afforded to science by the internet and digitisation to open access as a model that aligns with the core values of science. According to Fitzpatrick (2015):

> I have been wondering of late about the extent to which the problem is the degree to which our thinking about the goals of OA has gotten derailed by our focus on the business model of OA — and even worse, by a more-or-less exclusive focus on one particular business model that can simply be taken up without the need to reconsider the purposes or values of scholarly communication. […] My all-too-nascent idea, after all, was based on my experiences with MediaCommons, which led me to hope that groups of scholars could take control of the systems through which they publish by creating collective, cooperative, scholar-organized and -governed publications on open networks.

The recurring invocation of scholarly values and the references to the increasingly exclusionary practices of commercial publishers, reaffirm universal access to knowledge as a core value in science, and imply that the incumbent publishing model is in some way inherently anti-communal. This invocation of values is a critical stage in the deinstitutionalisation of publishing practice in science as it provides possible evidence of the application of social pressure in the deinstitutionalisation process (Scott, 2014). Pinfield (2015) conducted a textual analysis of open access literature from 2010 onwards and finds that the focus in the current open access discourse is on how open access can be made to work in practice, having moved on from a discussion of whether it should happen at all. This suggests an in-principle acceptance of open access within science, although methods and models remain contested. Therefore, while a guerrilla-led revolution is unlikely, a significant evolution is highly probable.

Evolution is a slow and unpredictable trajectory to plot. There is clearly discontent within some quarters of the science community when it comes to the relationship between science and publishing. Technology may have disrupted the system of knowledge dissemination, but profit-oriented commercial publishing, loathe to relinquish its foothold in a lucrative market, is seeking to exploit new opportunities presented by the disruption.

While many are therefore hopeful of a democratised knowledge commons with the advent of the internet and digitisation, evidence is beginning to emerge of a less utopian future for scientific publishing. Larivière et al. (2015) find that the top commercial publishers have in fact benefited from the digital era, as it has led to a dramatic increase in their share of the scientific literature published, and to a greater dependence by the scientific community on commercial publishers. De Filippi and Vieira (2014) highlight the risks of commodification: that the exploitation of the commons by one agent is likely to preclude others from doing the same and the information commons are increasingly controlled by large corporations. There is evidence of such commodification in the area of cloud computing, where the content holder has become as powerful, if not more powerful, than the copyright owners (De Filippi & Vieira, 2014). And a leading advocate of the open access movement, in a candid reflection on developments over the past 15 years, acknowledges that:

> Unfortunately, over the years, it seems like a lot of the core intentions and principles behind the Open Access conversation are being overshadowed by discussions around cost models of open access publishing. The irony is that while Open Access was supposed to improve the two-way models of knowledge access and dissemination, what we are seeing is that some models are creating new divides. […] What we are seeing is that publishers who have traditionally benefited from closed publications are now reaping even greater rewards from Open Access models! (Hillyer, 2017)

Commercial interest in science, propelled by a capitalist market logic of surplus accumulation therefore remains and will seek to defend its position despite any breakdown in the pact between science and commercial publishing. This breakdown can be seen as misalignment in the communal values of science necessary for it to pursue its logic of truth-seeking and profit accumulation as the central logic of markets.

2.3.4 Conclusion

In the 18th and 19th centuries, as the communication of science became more voluminous, more widely dispersed, and tightly coupled with matters of quality through the review of publications by a network of peers, science turned to commercial publishers for assistance. Science accepted the commercial imperatives of the publishers because the services provided were commensurate with the value added, and because there was a fair balance between the scientific imperative of access and the commercial imperative of profit. What is clear from the evidence is that there has been a disruption in the social compact between science and publishing. Increased editorial efficiencies and new forms of distribution are made possible by the internet and digitisation, and such developments make possible the enactment of 'full and open communication'. But the publishing industry has, by and large, both resisted change that could allow for greater access and started to experiment with ways of exploiting greater openness, in both instances in the interest of protecting high profit margins.

When the social compact between science and publishing – one that held in balance the value of access to knowledge in science and the value of profit in capitalist markets – is ruptured, there is a breakdown in relations, and a more vehement push from *within* the academy to restore the communal role of scientific publishing. Hence, while systemic change may be slow, open access scientific publishing cannot be dismissed; the alignment between the core values of science and an approach to the dissemination of knowledge that is more communal, catalyses the disruptive and restorative effects of open access scientific publishing in a global network of digitally connected scientists.

2.4 Open research data

The section above on open access has shown that at first it was civil society and then the academic community, or at least emerging factions within that community, that advocated for the removal of control over the dissemination of academic publications. Commercially-motivated actors are responding by exploring opportunities to commodify the commons. Nevertheless, the academic community remains a powerful force in the protection of open access to the knowledge products within science. In the specific case of open research data, the motivations for universal access and the concentration of power may not mirror the state of play in open access. This section explores the rise of open research data, and considers the apparent differences in the trajectory of open research data viz. open access.

2.4.1 Open (government) data

Open data is defined as data that can be 'freely used, modified, and shared by anyone for any purpose' (Open Knowledge, 2015). Open research data – as distinct from open government data or open corporate data – is that data which are open and emanate from the scientific endeavours of university academics and other credentialed researchers within the domain of science.

Initially, in the early 2000s, the focus of the open movement, of relevance to the scientific community, was on publications (particularly journal articles) in the form of open access. But open data was not far behind. Early initiatives, however, targeted government-held data rather than scientific data. And some important changes that predate a populist open movement were taking place as early as the 1980s within government (Gray, 2014).

In the 1980s, information policy in the US government focused on cost-cutting, innovation and growth. According to McDermott, information policies of the time 'considered information an economic resource, rather than a public good' (McDermott, 2008). In 2002, an influential paper by Peter Weiss of the US National Weather Service argued that 'the US and the EU should commit to move forward together to take the practical steps necessary to establish internationally harmonized open and unrestricted data policies for all public sector information [to realise its] wealth creating possibilities' (Weiss, 2002, p. 3). Building on the work of Weiss, an OECD working group on the information economy, developed recommendations for member countries, urging them to review their public sector information policies to seize the opportunity to develop 'new information products and services particularly through market-based competition' (OECD, 2008, p. 4).

In 2007, particularly in the UK and US, there emerges a more broad-based push for access to specific government datasets in the drive towards ICT-driven government services (often referred to as 'eGov'). In December of the same year, thirty open government advocates gathered in Sebastopol, California, and drafted a set of eight principles for open government data. These have become known as the 'Sebastopol Principles' and are still referred to as one of the gold standards for determining the openness of a dataset. At the same time, in Europe, an argument for creating social and economic value from government data (collected using public funds) began to find traction. And in the UK, political pressure mounted for increased accountability through open data in the wake of public scandals around, for example, parliamentary spending irregularities.

In 2009, shortly after being elected president of the US, Barack Obama pledged his support for open government data. What followed was the formation of the global Open Government Partnership and the emergence of .gov portals in the US, UK and other developed countries. More recently, a number of charters such as the International Open Data Charter have been published by open data advocacy groups and are aimed at eliciting formal commitments from national governments specific to the publication of open data. Country-level open data rankings, such as the Global Open Data Index and the Open Data Barometer, have also been developed to spur competitive behaviour among national governments.

The emergence of calls for the publication of open research data is, to some extent, a logical extension of the open government data movement. This is perhaps to be expected given that there are intersections between public funding and research. Public universities and research institutes as sites of research activity are examples of such intersections.

2.4.2 The emergence of open research data

In science, data began to appear in the public domain following the formalisation of science communication in the 18th century (Gross, Harmon, & Reidy, 2002). The 18th century is characterised by the appearance of the formal components of journal articles such as the introduction and conclusion as sections of articles, and of tables, figures and citations. The 19th century is notable for the emergence of passages of text that explain empirical data, the more frequent use of visuals, higher citation density and the appearance of methods sections in journal articles. In the 20th century, structured quantitative data is presented in journal articles with increased use of statistics and a wider range of data types (Gross, Harmon, & Reidy, 2002). As the number of journal articles published doubled every 15 years since 1950 (Mabe, 2003), the amount of scientific data published has also increased (although not at the same rate as for the growth in journals). The number of research data repositories listed on re3data.org (as at February 2016) was in excess of 1,500.

In 1997, the US National Research Council argued that 'full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research' (US National Research Council, 1997). And by 2007, the OECD published a set of Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007). A 2009 report by the US National Academies of Science recommends that: 'all researchers should make research data, methods, and other information integral to their publicly reported results publicly accessible, in a timely manner, to allow verification of published findings and to enable other researchers to build on published results' (National Academy of Science, 2009). In 2010, the European Commission published 'Riding the Wave: How Europe can Gain from the Rising Tide of Scientific Data. Final report of the High Level Expert Group on Scientific Data' (European Commission, 2010).

Many nations signed up to the principles of open data through membership of the International Council of Science (ICSU) (Royal Society Policy Centre, 2012). And there has been an increase in the relationship between funding for scientific research and the contractual obligation to publish open research data. For example, the European Union's Horizon2020 research funding programme requires all funded research to publish open research data after 2017. The European Commission has elevated open science to one of the Commission's top policy priorities and it has launched initiatives such as the European Open Science Cloud, the drafting of the Open Science Policy Platform, the assembly of an expert group in 2016 to inform the implementation of open science policies, and the launch of various funding programmes in support of open science. The US National Institute of Health (NIH) in the US, the Wellcome Trust, the EU and the Bill & Melinda Gates Foundation have all introduced conditions requiring funded research projects to publish open data (Eisfeld-Reschke, Herb, & Wenzlaff, 2014; McKiernan, et al., 2016). According to the JISC and RLUK-funded Sherpa Juliet project, globally there are at least 34 funders who require open data archiving and 16 who encourage it (Hahnel, 2015).

Within science, these developments are accompanied by the emergence of academic inquiry on the relevance and impact of open research data. Several publications begin to appear in the mid-2000s, including *Public Sector Information in the Digital Age* (Aichholzer & Burket, 2004), 'Data publication in the Open Access initiative' (Klump, et al., 2006), 'Let data speak to data' (Nature, 2005), 'Sharing detailed research data is associated with increased citation rate' (Piwowar, Day, & Fridsma, 2007), and 'The poor availability of psychological research data for reanalysis' (Wicherts, Borsboom, Kats, & Molenaar, 2006), to name a few.

More recently, empirical studies on the benefits of open research data continue to be published. The focus is on the possibility of open research data leading to greater efficiency in research, improvements in the quality of research, accelerating scientific discovery, increased research uptake, and improving accountability to the scientific community as well as to the public. The latter is regarded as important given that research funding is often sourced from taxpayers; the public has therefore 'earned' the right to access scientific research.

The relationship between the replicability and quality of scientific research, and greater transparency in terms of the findings (open research data), is a particular area of interest. To illustrate the extent of the 'reproducibility problem', Begley and Ellis (2012) report that in the case of clinical trials, from a sample of 53 'landmark studies', 89% of the scientific findings could not be confirmed, and, in another such study, 75% of studies could not be validated. Ioannidis et al. (2009) find that the main reason for failure to reproduce findings was that data was unavailable. The claim is made that a 'lack of reproducibility for whatever reason is primarily a waste of public resources but it also affects follow-on benefits such as knowledge transfer and the ability to spin new business off academic innovation' (Sharif, Ritter, Davidson, & Edmunds, 2017).

Other studies, attuned to the value of linking the publication of open research data to the reputational economy of science, focus on the relationship between the publication of the open research data and citations (Piwowar & Vision, 2013; Tenopir, et al., 2011).

2.4.3 The value of open research data

Ubiquitous access to research data combined with the potential of linking and analysing big datasets, has introduced a new awareness of the value of data, including data generated by the scientific community (Stodden, 2010). The claim is also made that in a so-called 'post-fact world', the value of data as the key input for predicting everything from market sentiment to election polls to particle behaviour to weather forecasting, takes on more rather than less value (Davies, 2016).

As in the case of open access, there is an invocation of communalism, that is, that data is not the property of the individual researcher but belongs to the scientific community. Making research data open to other scientists contributes positively to the advancement of knowledge.

While open science is claimed to generate social benefits, the enthusiasm of policy-makers to promote open science is also premised on an economic (rather than scientific) approach that commodifies knowledge in the digital economy; this is nowhere more apparent than in the mantra of data being the new oil (Haupt, 2016). For example, Vice-President of the European Commission, Neelie Kroes, addressing the 2013 Annual Innovation Forum in Brussels in 2013: '[Data is] a fuel for innovation, powering and energising our economy' and '[u]nlike oil, of course, this well won't run dry: we've only just started tapping it' (Kroes, 2013). In another address, Kroes stated that 'data is the new oil for the digital age. How many other ways could stimulate a market worth 70 billion Euros a year, without spending big budgets? Not many, I'd say' (Kroes, 2012). The data as fuel/oil metaphor suggests that data is an abundant resource to be extracted, refined and traded as a new commodity. The invocation of the metaphor is reflective of the way in which science and its products have come to be framed in terms of its economic potential, productivity and competitive advantage (Çalışkan & Callon, 2009). This paradigm is in sharp contrast to a more human-centred and critical-social discourse of knowledge as a public good (Kuhlen, 2010), which sees the rights to produce and share knowledge as essential for social equity and human well-being (Hall & Tandon, 2017).

2.4.4 Conclusion

As in the case of open access, open research data has attracted the interest of the market and of politicians as a driver of economic growth. However, unlike open access, evidence of a concerted push from within science for the publication of open research data is scarcer (Fecher, Friesike, & Hebing, 2015). Scientists cite lack of time to prepare and curate data for publication; reluctance to let others benefit from their labour or personal investment; concern that their findings may be scooped by other scientists; and fear of errors in their published open research data (Bishop, 2016). Many of these concerns are more closely associated with the counter-norms of science (particularism, individualism and interestedness) than the norm of communalism argued to be driving a shift to open access scientific publishing.

These differences and convergences, and their relevance to open science is explored below.

2.5 Comparing open access and open research data

From the above investigation on the emergence of and developments in open access and open research data, several differences can be noted.

First, open research data is a relatively new addition to the products of science. Access to empirical research data within science has only emerged recently in conjunction with advances in technology and concomitant with the emergence of open government data. Previously, access was limited to the information products in the form of tables of aggregated data or statistics, or graphs created from underlying raw datasets and published as supporting material in journal articles for the science community. Access to raw research data was neither required nor possible. Scientific articles, in contrast, have been available in full within the scientific community from the latter half of the 18th century. And while access to journal articles was not universal, the arrangement between science and publishers extended the distribution of journal articles to a large proportion of the scientific community. Therefore, while open access introduces universal access to an existing scientific product that was already widely accessible within science, open research data introduces universal access to a new product that was previously only partially presented. Consequently, from the scientist's perspective, open research data is not only a matter of unprecedented access but of greater scrutiny from peers, policy-makers and the public.

Second, in the case of journal articles and open access, there is in place an existing and long-standing arrangement between science and publishers on how to disseminate content. Control over the dissemination of knowledge is shared between science and the publishers. Open access is, in part, a challenge to how that control is shared. In the case of open research data, by contrast, there is no arrangement in place in terms of dissemination. This applies equally to an absence of data standards as it does to lack of consensus on the modes of distribution, who should incur the cost to convert raw data into well-described and usable research data, and the incentives for scientists to do so. The impetus to make research data openly available may be coming from funding agencies and from the open science movement, but the processes and venues for making data available, are not uniform. Scientists can select from a variety of online platforms to publish their data, while publishers are moving to create new online platforms for the dissemination of datasets. The data journal is an example of this emergence.

Third, at least in the formative stages, are the motivations driving greater openness. In the case of open access, as shown above, the dissemination of knowledge in the form of an arrangement between the scholarly publishing industry and science was already institutionalised, and open access, at least to some degree, is a reaction against the commercialisation of knowledge as new technologies bring into question the value of the arrangement. In the case of open government data, the value of the data is not established, and government officials and supranational agencies need to develop value propositions for open government data. Gray (2014), in a more detailed examination of the genealogy of open government data, argues that 'the rise of open data has coincided with a focus on technological innovation, public sector efficiency and economic growth in official transparency discourse rather than on social justice and meeting the needs of citizens' (Gray, 2014, p. 1).

Given the historical trajectory of open research data, there are also differences in the actors and their interests when compared to open access. In the case of open access, the economisation of openness by large scholarly publishing corporations emerged much later because those publishers initially regarded open access as transient and therefore as posing no threat to the extant arrangement between commercial publishing and science. In the case of open data, publishers have also been slow to respond but they are not in the same position of power in relation to its dissemination as they are in the publication of scholarly books and journals. The publication of open research data has been driven primarily by policy-makers and funders, both influenced by the open government data movement that has been advocating for openness on the basis of greater efficiency and innovation, and, more recently only, on the basis of transparency and accountability. Pressures on scientists to publish open data continue on the basis of its economic potential.

There are additional pressures from within science for research data to be more open in order to improve the replicability and reliability of research findings. But at this stage it is not clear whether these demands from within science are attributable to a reactivation of the Mertonian norm of communalism in a way suggested to be the case for open access scientific publishing.

2.6 Open science as the new, contested reality

While the balance of external and internal forces on science for greater openness may differ, in both the cases of open access scientific publishing and open research data, the communal benefits for science and the economic and political benefits for a broader collective of non-science stakeholders are held in tension. The outcome is more open science. Trends in science funding policy and the data on the number of open science outputs such as journal articles and research data are indicative of permanence (OECD, 2015).

Open science can clearly no longer be relegated to the class of technologies that rush onto the scene only to be surpassed by the next 'game changer' in a matter of months. While platforms and formats may change along with advances in technology, the principle of openness in science appears to have reached critical mass.

At the same time, the motivations for openness within science and from stakeholders external to science are neither in alignment nor can it be assumed that they are similar for the different products of open science. Science, the market, politics and publics have different, often competing, interests in advocating for open science, and the intersection of these different interest groups as they make their respective cases exposes science and its products to approaches that are agnostic to the values and norms of science that are in place to protect the production of knowledge for the establishment of truth.

The question that remains is whether the gamut of potentials made possible by open science are fully understood, including the manifestation of possibly undesirable and detrimental potentials. One approach to anticipating possible outcomes and, hopefully, foreseeing unintended consequences, is to develop a better understanding of the nature of access to the products of open science. The following section examines access in relation to openness with a particular interest in how changing information communication technologies and, by implication, a changing science communication environment, may be creating new potentials.

2.7 Open science and a changed science communication landscape

With regard to the communication of science, an increase in the accountability of publicly funded research institutions has been accompanied by a concomitant increase in pressure on scientists to engage with the public (Scheufele, 2013; Scheufele, 2014; Weingart & Guenther, 2016; Yeo & Brossard, 2017). The expectation is that the public, as a return on its investment in science, is entitled to know about scientific discoveries, and can reap certain benefits by way of this access to shared knowledge (e.g. health and educational benefits through informed decision-making and behaviour).

Scientists engaging with the public is certainly not a new phenomenon. In the past, however, engagement with the public was voluntary or self-motivated; in contrast, the scientists of today are expected to engage with the public to the extent that public engagement is becoming part of their performance assessments (Jensen, Rouquier, Kreimer, & Croissant, 2008). In support of increased public engagement by scientists, research has shown that scientists who engage with the public are not sell-outs, fame-seekers or quacks, but tend to have a higher impact on science as measured by citation counts (Jensen, Rouquier, Kreimer, & Croissant, 2008).

The rise in the accountability of science is not dependent on but is certainly accelerated by advances in information and communications technologies (ICTs), prime among which is the internet. These new technologies make possible increased transparency and accessibility in real time and on a global scale. In terms of the domain of science in particular, the emergence and entrenchment of ICTs in society have had a series of impacts on the communication of science. These include:

1. Infinite amounts of space for storing articles, data and other scientific products.
2. Unlimited and theoretically ubiquitous access to publications and data.
3. Changes in the modes of dissemination (for example, email and distributed print-on-demand).
4. Accelerated review and publishing processes.
5. Automation in the collection, aggregation and analysis of data.

The digitisation of the traditional print media and the advent of online social networks have further disrupted the communication of science (Brossard, 2013; Scheufele, 2013; Southwell, 2017) and are likely to continue to play a part as social media in the general population becomes normalised (Ware & Mabe, 2015).

If open science provides the opportunity for the public to access science and digital media provides the space for interaction between scientists and its publics, then it is conceivable that changes would occur in how the communication of science is mediated (Landrum, 2017; Scheufele, 2014). As Dickel and Franzen (2016, p. 3) point out, open science 'enables new modes of knowledge reception via non-scientists'. According to Peters (2013, p. 14102) '[m]ost scientists assume a two-arena model with a gap between the arenas of internal scientific and public communication. They want to meet the public in the public arena, not in the arena of internal scientific communication. […] Changes in the science–media interface may be expected from the ongoing structural transformation of the public communication system.' At stake is the trust placed in science by society (Schäfer, 2016; Weingart & Guenther, 2016).

There has been an unprecedented increase in the number of scientific sources available, leading to what Davies (2016) describes in his analysis of post-truth politics as an 'oversupply of facts' mired by the many methods and the origins of funding fueling that oversupply.

The media have traditionally been the primary interface between science and the public: 'The public of mass-democracies is almost exclusively represented and continually updated by the media. Everything we know about the world, and thus about science, we know through the media' (Weingart, 2011, p. 338). And it is the science journalist who has traditionally kept the public informed on the latest developments from the world of science (Schäfer, 2017). We are, however, witnessing a simultaneous decline in science journalism (Scheufele, 2013; Schäfer, 2017), an increase in the scramble for attention among a variety of stakeholders (Weingart & Guenther, 2016), and the emergence of internet-enabled social media as new, informal, interpersonal channels of communication between scientists and the public (Southwell, 2017).

The supply side is certainly more diverse as communication technologies allow for a decentralisation of the channels of communication. The mass media is no longer at the centre as individuals and minority groups are able to broadcast their own content, bypassing traditional media, and both attract and surpass the levels of attention garnered by the mass media because of the reach and density of new communication networks such as Twitter, Facebook, YouTube and Instagram (Schäfer, 2017; Southwell, 2017; Wu, 2016). This shift from vertical, hierarchical modes of communication to horizontal, networked communication has led Castells (2009) to describe this new supply-side-driven form of communication, 'mass self-communication'.

Bucchi (2018) describes a 'crisis of mediators' in which new scientific research is increasingly fed in real time into the public domain without being filtered by communication professionals. This connects unfiltered (or open) science communication directly with popularism and social trends. The non-scientific public must therefore be highly adept at discerning which communication sources of scientific information to trust (Kahan, Scheufele, & Jamieson, 2017; Scheufele, 2013), assuming they are receptive to such communications at all (Schäfer, 2017; Schäfer, Füchslin, Metag, Kristiansen, & Rauchfleisch, 2018).

On the demand side, the number and type of participants in the communication of science has broadened, and the narration of science is seen to be shifting towards conversation and contestation among a much bigger, more diverse community (McCallie, et al., 2009; Weingart, 2017). This may reshape how scientists communicate. There are those who suggest, for example, that scientists should make the most of the direct access afforded by digital technologies by producing secondary outputs; academic outputs that summarise scientific discoveries in layman's terms for consumption by the public (Bornmann, 2014; Bornmann & Marx, 2014). According to Hornig Priest (2009), the effect on the communication of science is a more active, diverse, multi-directional and personal type of communication with audiences who apply their own ethics, values and beliefs to its understanding.

Much of the science communication literature and, to some extent, the above description of the impact of information communication technologies on science communication, focus on how science can either use new technologies to communicate with publics or how it should be responding to new forms of communication between publics. An additional dimension to consider concerns access to the products of science (journal articles, research data, conference presentations, etc.) by publics and how these products are used in communications outside of the formal system of science communication for which they are intended. An increase in the ability of publics to access the products of science in their digital forms via the internet, the presence of the products of science on social media platforms such as Twitter (Haustein, Costas, & Lariviere, 2015; Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), and the relatively low levels of use of social media by scientists (Bar-Ilan, et al., 2012; Van Noorden, 2015), suggest that the products of science are being injected into the communications of publics via those online media platforms (Haustein & Costas, 2015). As Grand et al. (2014, pp. 1-2) state: non-professionals can now access 'previously uncharted territories' where previously 'only the refined, polished outputs of research appeared in public'.

There is an emerging acknowledgement of how the web and social media are impacting processes related to science communication (Southwell, 2017). Holliman (2011, p. 2) writes:

> This enmeshing of the 'social' and the 'technological' through digital media, tools, technologies and assembled network communities provides interesting opportunities for science communication researchers to explore how the sciences continue to be communicated, but are also sometimes discussed and shared with opportunities for collaboration, contributions and also confrontation.

Brossard (2013) and, more recently, Davies and Hara (2017) provide surveys of the current research in this area. However, the studies to-date have tended to be more interested in the use of social media for communicating the findings of science than in how information extracted from science is used in the social media. In other words, there is a tendency to think about science communication in the digital age as being about the potential of new media to communicate information about science, and less about how the digital communication environment makes possible the exchange and use of unmediated scientific information by non-scientific publics. The dearth of research concerned with the content of communications as they pertain to science in online social networks attests to an underappreciation of how science is used in the social media (Southwell, 2017; Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013).

The use of information extracted from the formal communications of science is an important distinction to make for at least two reasons. First, the use of information established by the scientific method in online communication is not the same as the spread of mis- or disinformation in the social and other media.[7] Deliberate falsehoods or so-called 'fake news' are fabricated truths whereas the results and findings extracted from the formal communications of science are scientifically proven accounts of the natural and social world. While it is certainly possible for scientists to fabricate truths and to publish those truths in the formal communications of science, such fabricated truth claims are subject to scrutiny by peers in a way that fake news is not.

Second, the information extracted from open science may take place at a point in the formal communication process where a truth claim has not yet been fully accepted (or rejected) by the scientific community. In other words, access to and the use of 'prototype' truths extracted from the formal communications of open science is possible. These truths may or may not pass the scrutiny of

[7] See Born and Edgington (2017) for the differences between misinformation and disinformation.

the scientific community in the fullness of time, and the potentials of those unsettled truths circulating in communication networks that are not guided by the norms of science are unknown.

There are a few empirical studies that have centred their enquiry on the products of formal science communication and their use in online communications. Nelhans and Lorentzen (2016) explored the prominence of scientific articles in conversations on Twitter and found that they were mainly mentioned for self-promotion, to initiate online conversations or as supportive evidence in discussions. While their methods did not preclude non-scientific publics, their primary interest was in the use of scientific articles on Twitter for conversations between scientists. Vainio and Holmberg (2017) also explored tweets mentioning scientific articles and found that scientific articles were tweeted to promote ideological views especially in instances where the article represented a topic that divides general opinion. The focus of the study was narrow – it included only articles written by Finnish scientists.

In terms of topics that divide opinion, Hopke and Simis (2017) explored discourse over hydraulic fracturing and the shale industry on Twitter during a period of heightened public contention regarding the application of the technology. Results showed that the discourse was dominated by activists. In the case of the contested issue that is the focus of this thesis – vaccination – several studies have been undertaken to explore the use of the web and social media by the anti-vaccination movement. These studies focus predominantly on the online strategies and tactics of the anti-vaccination community on the web and in the social media (Bean, 2011; Bennato, 2017; Kata, 2012; Mitra, Counts, & Pennebaker, 2016; Moran, Lucas, Everhart, & Morgan, 2016; Sanawi, Samani & Taibi, 2017). At the time of writing, no studies could be found that focus on how a *non-expert public* such as the anti-vaccination community is using the products of *open* science; nor are there in-depth analyses of their level of engagement with those products.

One of the consequences of the products of science being fed into interpersonal communication in the social media is the dislocation of those products from a communication space governed by their institutional values and norms to a space where such values and norms no longer inform behaviour. Instead, the communication products of science are fed into highly politicised communication spaces (Jamieson, 2017; Scheufele, 2014).

Technology and openness have created new modes of communication between science and its publics. The outcome is the potential for scientific products such as journal articles and open research data to be dragged by attentive publics into their conversations on controversial political, economic and social issues, conversations that are increasingly taking place in the social media. For open science to be both judiciously deployed and purposeful, a deeper understanding of use of the products of open science by non-scientific publics is therefore needed.

2.8 The new potentials of open science

New modes and trajectories of communication between science and its publics introduce new potentials. These potentials may take the form of benefits to science and to society, but they may also introduce risks, most likely as the result of unforeseen consequences. An unquestioning faith in the potential of technology to advance society mutes the concerns expressed by socially-attuned observers, and the unanticipated consequences and risks may only emerge once the damage has been done. Referring to the founders of Google and Facebook as examples, Naughton (2017) reports that ‘it never seems to have occurred to them that their advertising engines could also be used to deliver precisely targeted ideological and political messages’. The founder and ex-CEO of Twitter, Even Williams, has lamented the use of the platform for unintended, confrontational and nefarious purposes by some of its users: ‘I thought once everybody could speak freely and exchange information and ideas, the world is automatically going to be a better place. I was wrong about that’ (Streitfeld, 2017).

Evidence of the potentials of open science is still limited (Leonelli, Spichtinger, & Prainsack, 2015; Levin & Leonelli, 2017). Absent in much of the science communication literature are the potential risks of open science, although there are signs that a consideration of the potential risks is emerging (Bishop, 2016; Dickel & Franzen, 2016; Jasanoff, 2006; Lewandowsky & Bishop, 2016). Where the effects or impacts of open science are considered, the emphasis is often on science itself, and on the beneficial impacts (Bishop, 2016).

In the case of open access publishing in science, studies have been done on the citation benefits of open access publishing (Dorta-González & Santana-Jiménez, 2017; Eysenbach, 2006; Piwowar, et al., 2017; Wang, Liu, Mao, & Fang, 2015), on the cost benefits of open access (Swan & Brown, 2008), and in countering information asymmetries across the globe (Chan, Kirsop, & Arunachalam, 2011). Tennant et al. (2016) summarise the benefits of open access publishing as follows:

1. Academics benefits: Although evidence is variable, there is a generalisable citation advantage for researchers who publish open access and there are additional benefits in

terms of the visibility and dissemination of research to both academic and non-academic audiences.

2. Economic benefits: There is limited evidence that open access to the scientific knowledge in the research literature is of value to innovative enterprises as well as for the delivery of governmental and non-governmental services. Open access also has potential financial benefits by introducing savings for publishers, university libraries and funders of research.
3. Societal benefits: There is strong evidence for open access advancing citizen science and in making scientific knowledge more accessible to researchers in developing countries.

There is an emerging body of research on open research data, specifically on the potential benefits of open research data in terms of the replicability and verification of research findings, and the citation benefits of sharing research data (Lewandowsky & Bishop, 2016; Piwowar, Day, & Fridsma, 2007; Piwowar & Vision, 2013; Tennant, et al., 2016). However, the focus, as with open government data, tends to be on the supply of data, and the mechanisms and incentives to stimulate supply, such as research on the attitudes of and the extent to which academics are sharing research data (Ferguson, 2014), and changes to policy relating to the publication of open data by funder agencies and academic journals (Hahnel, 2015). Bishop (2016) draws attention to the risks of open research data to science in the form of *p*-hacking and publication bias. Some empirical studies have been conducted on the demand side, that is, on the use of open research data (Peters, Kraker, Lex, Gumpenberger, & Gorraiz, 2015), but these are few and far between and do not differentiate between scientific and non-scientific use.

The benefits of making open research data publicly available are summarised by FOSTER (drawing on the OECD 2015 and the Royalty Society 2012) as follows:

1. Efficiency: Increased access to scientific products can improve the effectiveness and productivity of the research system by reducing duplication as well as the costs of creating, publishing and reusing data, and by multiplying opportunities for participation in the research process.
2. Quality and integrity: Access to open research data creates opportunities for wider evaluation and scrutiny by the scientific community, thus allowing for more accurate replication and validation of research results.
3. Economic benefits: Increased access to research data can foster spill-overs to systems of innovation. Open research data can reduce delays in the re-use of data by firms and entrepreneurs, and promote a more direct path from research to innovation to produce new

products and services. It can also contribute to an increase in awareness and informed decision-making among citizens and consumers.

4. Public disclosure and engagement: Open research data provides transparency in the outcomes of public funded research and could restore trust and support for public policies and investments. It also promotes citizens' engagement and active participation in the scientific process, including the collection of data.
5. Global benefits: Open research data can promote collaborative research and more rapid knowledge transfer across multiple contexts across the globe to improve understanding of challenges that require coordinated international action.

Tennant et al.'s assessment of open access (and to a lesser degree open research data) undertook to 'provide information on the various benefits and drawbacks of open access to scholarly research' (Tennant, et al., 2016, p. 14). However, amid the list of benefits enumerated, only a single risk is identified: that of predatory publishers exploiting the pay-to-publish models used by certain open access publishers.

While openness in science may be valued and even inevitable, according to Jasanoff (2006, p. 42), it should be 'purposefully cultivated and judiciously deployed in order to serve its intended functions well'. Leonelli and Prainsack (2015) state that all 'too often, open science is cast as a tool to enhance the transparency and accountability of research, with little critical reflection on the potential confusion and obscurity created by releasing large quantities of information'. Similar concerns have been expressed elsewhere (Dickel & Franzen, 2016; Franzen, 2016; Leonelli, Spichtinger, & Prainsack, 2015). A Royal Society report (Royal Society Policy Centre, 2012), acknowledges that there are risks inherent in making research data open, and that these risks warrant restriction on openness. The warranted restrictions pertain to commercial interests, personal information, safety and national security.

While not commenting on open science per se but rather on the possible impacts of the increasing interconnectedness or 'coupling' between science and the mass media, Weingart (2012) cautions that knowledge production itself may be disrupted. And such disruption may well be amplified in a context where science is more open, and the mass media is more present. From a journalist's perspective, Julia Belluz (2015) writes:

> We now live in an age of unprecedented scientific exploration. Through the internet, we have this world of knowledge at our fingertips. But more information means

> more bad information, and the need for skepticism has never been greater. I often wonder whether there is any value in reporting very early research. Journals now publish their findings, and the public seizes on them, but this wasn't always the case: journals were meant for peer-to-peer discussion, not mass consumption. Working in the current system, we reporters feed on press releases from journals and it's difficult to resist the siren call of flashy findings. We are incentivized to find novel things to write about, just as scientists and research institutions need to attract attention to their work. […] But this cycle is hurting us, and it's obscuring the truths research has to offer.

There are further concerns that science hype for the purpose of attention-seeking (Weingart, 2017) and a communication environment in which scientific information is tainted by political or other non-scientific interests (Weingart & Guenther, 2016) may damage the legitimacy of science as an institution.

According to Davies (2016), we are witnessing a fundamental shift from facts to data, from knowing the past to knowing the future, and we are caught in an interregnum state of uncertainty or possibly even ignorance (Davies, 2015) about how to deploy data to represent reality. Ignorance extends to the ultimate ends to which a 'sublime grid of quantification' is being put to use (Davies, 2015). If this is the case, science is being opened to an army of data alchemists who in a realm of increasing uncertainty peddle claims about reality – claims that may harbour risk and only serve to propagate greater uncertainty.

2.9 Gaps in the literature

From the above review of the literature in the fast-evolving area of interest that lies at the intersection of science communication, information and communication technology, social (media) networks and sociology, it is apparent that there are few, if any, empirical studies that place the products of open science at the centre of their investigations *and* that focus on the actual use of those products by non-scientific publics in their ideologically-motivated online communications. Combining both considerations in a single study will provide useful early insights for science communication practitioners and theorists alike.

From a broader perspective, in his review of 20 years of public engagement with science, Irwin (2014) points to a failure of the social sciences to contribute to an understanding of science–public relations that neither discards the deficit model nor defaults to an instrumentalist role for public engagement with science research. His reflection exposes a need for new conceptions that account for the messiness and complexities in which the public's engagement with science is mired. This is a view echoed by Schäfer (2016). In his response to Weingart and Guenther (2016) on trust in science communication, he highlights the urgent need for more empirical research to enlighten conceptual work on mediated trust in science. Acknowledging the complexity of communicating science when the boundaries between politics and science blur, Scheufele (2014) is prompted to progress beyond the dominant models of science communication – that is, the knowledge deficit and public engagement with science models – and proposes a third model of science communication as political communication.

Irwin, Schäfer and Scheufele each expose a need for research that explores new models for the communication of science. These new models must be cognizant of and capture the effects of open science in communication environments that are fundamentally different. With regard to the latter, Weingart and Guenther (2016, p. 5) state: 'It is astonishing to observe how internet-based communication technology has escaped critical assessment almost entirely'. While there is emerging research in terms of the opportunities and threats of open science for some actors such as researchers, academics, policy-makers and publishers, there is a lacuna in the existing body of knowledge pertaining to the potential use of open science by non-scientists, particularly in relation to the communication of science.

2.10 Framing the research

Based on the historical overview that foregrounded differences in the motivations and in the actors advocating for greater openness in science, the apparent lack of research on the connections to the products of science, and the calls for new conceptualisations of science communication, this research sets out the following argument:

1. Stakeholders external to the domain of science (such as funders and the public), who are in a more expectant and frequent interaction with public research institutions, are increasing their demands for open science, including the publication of open access scholarly publications and open research data.
2. There has been an unprecedented expansion in the number of publications and datasets available in open formats.

3. Open science, enabled by rapid developments in ICTs, has resulted in the traditional media no longer being the only gatekeeper to the public's understanding of science; nor are academic publishers the gatekeepers of formal communication within science.
4. Scientists are relieved of control over their research as they are increasingly required to make their research openly available.
5. Consequently, the non-scientific public has, theoretically at least, unencumbered access to open research.
6. It is reasonable to expect an increase in the use of open research by non-scientific publics, either directly or via intermediaries, for both personal and/or political reasons.

However:

1. Little is known about the use of open research by the public; there are gaps in the existing literature about the potential effects of an open science on society and on the communication of science.
2. The 'ambiguous embrace' of the social sciences (Irwin 2014) has done little to advance theories on the communication of science that are contextually sensitive and politically attuned (Scheufele 2014).

Therefore:

Research on the use of open science by non-scientists will provide new insights into the potential of open science while simultaneously contributing to the advancement of empirically-grounded theory on the communication of science.

2.11 Research aim and research questions

The aim of this study is to establish whether those located outside of the formal science communication system, now more universally accessible, are using open access publications and open research data. The reason for doing so is to gain insight into the overarching concern as to what the potentials of openness are both for society and for the future of science communication. Persistent states of controversy over established scientific truths can damage the scientific communication environment; producing empirical evidence that advances our understanding of the science communication environment is therefore essential to the protection of that environment (Kahan, Scheufele, & Jamieson, 2017).

The primary question posed by this research is:

What are the potentials of open science in the communication of science?

The research will endeavour to answer a series of sub-questions, each of which is aimed at revealing insights related to the new communication potentials made possible by an open science. The sub-questions posed are as follows:

- Are non-scientists accessing the products of open science?
- Are non-scientists engaging with the products of open science?
- Who are the non-scientists mediating connections to the products of open science?
- How can science communication respond to the new potentials of open science?

Chapter 3. Conceptual framework

This chapter presents a conceptual framework for research that, initially at least, draws mainly on the science communication literature. The limitation to that body of literature is imposed because it allows for a more focused framework; a focus that would be difficult to maintain if concepts from other relevant fields such as sociology, communication studies and computer science were to be introduced into the framing. However, once a conceptual framework has been established, it is enriched in the second half of the chapter by incorporating concepts introduced by sociologist Manuel Castells on the flow of information in communication networks.

Fecher and Friesike (2013) identify five schools of thought within open science: the infrastructure school, the public school, the measurement school, the democratic school and the pragmatic school. The democratic school within the open science movement is concerned with *access to* knowledge, rather than democratising the *creation* of knowledge and making it comprehensible to non-experts (public school), bolstering the *efficiency* of the research and dissemination processes by opening up the science value chain (pragmatic school), deploying *technology* to make possible new research processes at unprecedented speed and scale (infrastructure school), or exploring alternative approaches to *measuring* the impact of science (measurement school).

The communication of science as a field of study overlaps most notably with the public school of thought in open science. Much of the science of science communication literature deals with how science is communicated to non-scientific audiences. However, science communication also concerns itself with the communications of science and within science – between scientists – and this interest overlaps with the democratic school of thought's emphasis on access; access to the scientific communication process by non-scientists, made possible by the publication of open access research publications and by open research data,[8] which introduces a latent form of science

[8] It would be reasonable to question the description of open research data as 'communication'. There is an obvious difference between for example, talking or writing a letter, and presenting someone with a list of vowels, consonants and diphthongs. This could be seen as analogous to the difference between the writing up of scientific findings for publication and simply making available on a website a large dataset from which the findings were extrapolated. However, the raw material of science is no less socially constructed than the building blocks of language (Johnson, 2018). Neither exist in the wild. The scientist must take many decisions – informed by principles of both science and society (e.g. ethical considerations) – in designing the processes for collecting data and in determining how data will be categorised, stored and shared. These decisions shape and frame the data. In this sense, and for the purposes of this research, open research data is described as a latent form of science communication. In addition, even if one were to dismiss open research data as a

communication. Even when scientists are not actively communicating with non-experts, non-experts have access to the communication between scientists, and are at liberty to interpret those communications for their own benefit. It is for this reason that this research on science communication is situated within the democratic school of thought.

The research is also concerned only with access to knowledge that takes place within the formal science communication system. In other words, informal communication that takes place between scientists to formulate and articulate hypotheses, develop research designs, plan the research process and the like, and which are usually private in nature, are not included. What is included are the formal communication processes that are typically centred around the publication of truth claims by scientists in such a manner that they are recognised, accessible and verifiable by their peers (Taubert & Weingart, 2017). Within the formal communication of science, this thesis focuses only on the behaviour of non-scientists and how they are accessing the products of the formal system of science communication made possible by open science.

There is a body of literature that concerns itself with the democratisation of knowledge production in order to produce 'socially robust knowledge' (Weingart, 2008). Open science can be situated within the 'socially robust knowledge' discourse. The focus is on a contemporary re-versioning of science in which there is a shift from a scientifically ordered downstream knowledge production—implementation—uptake trajectory to one in which society participates upstream in all the stages of a more open and inclusive trajectory. Examples include citizen science and open peer review. This research project does not concern itself with the upstream integration of society in science. Instead, the focus is on access afforded by greater openness and, indirectly, the impact on society itself, via the re-use of open scientific products. Any societal impact may or may not pose indirect epistemological risks, not least of which may include challenges to the credibility of science (Weingart, Engels, & Pansegrau, 2000).

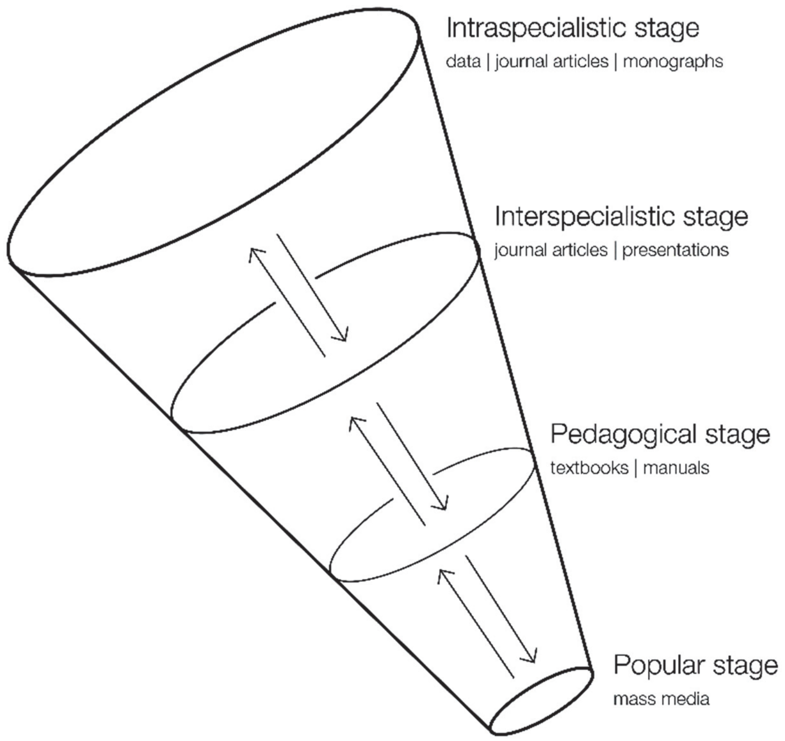
---

form of communication, its uptake by intermediaries and the fact that these intermediaries repackage data as information for consumption by others, situates open research data as an important component within the science communication system. A description of the open research data repository, Dryad, reads as follows: 'The vision of Dryad is a *scholarly communication system* in which learned societies, publishers, institutions of research and education, funding bodies and other stakeholders collaboratively sustain and promote the preservation and reuse of data underlying the scholarly literature.' The distinction between knowledge production and knowledge reception, and the fact that they decouple in an open, digital world (Dickel & Franzen, 2016), supports the view that open research data demands attention as a form of science communication.

3.1 A continuity model of formal science communication

The formalised process of science communication has been described as a closed and progressive one. Cloître and Shinn's (1985 in Bucchi 2004) continuity model of science communication as a funnel-shaped continuum is illustrative of such an understanding of the communication of science (see Figure 1). They suggest four stages in the formal science communication process: intraspecialist, interspecialist, pedagogical and popular.

Figure 1: Continuity model of science communication (Bucchi 2004)



At the intraspecialist stage, communication is between specialists in a specific field, discipline or sub-discipline. Communication is described as vague, 'esoteric' and highly abstract. Typical communication products at this stage include journal articles published in specialised scientific journals, monographs, empirical datasets, and graphs and illustrations. The interspecialist stage is characterised by communication via texts in bridging journals such as *Nature* or journals with a more inter-disciplinary ambit. The presentation of academic papers at conferences attended by scientists working in a similar field but in different areas of speciality, is also a form of interspecialist communication. Typical communication products include journal articles, conference papers and presentations. The first two specialist stages describe communication within science. In the subsequent two stages, communication begins to shift to audiences outside of the science, with concomitant influence on how science is communicated. In the pedagogical stage, the theoretical corpus has been established through consolidation in the previous two stages. This is the stage of 'textbook science' and scientific fact. Knowledge is communicated as established paradigm and acknowledges the cumulative nature of science in establishing the facts of the observable world. Communication types include textbooks and manuals. The popular stage is characterised by the communication of science in the mass media, and is notable for reduction of distinctions, nuances and truth disclaimers into elementary statements and formulae, and the use of metaphor and of imagery to communicate the facts of science (Bucchi, 2004).

The stages in the continuity model describe the internal trajectory and palimpsest-like continuity of communication *within* the sciences (and their associated processes for knowledge codification) via peer-to-peer communications to the external communication of science (post-codification of knowledge) by scientists to the public via a variety of media channels. The funnel shape is representative of changes in the communication of science as it depicts the pared progression and 'solidification' of knowledge creation from vast, contested and abstract, to a stage at which knowledge has become acquiesced, unanimous and specific. As Bucchi (2004, p. 116) illustrates: 'specialist exposition – the "science of the journals" – is provisional and tentative'; only once validated by peers through an iterative and protracted process, are the discoveries of science made public, or popularised, if at all.

3.2 Deviant trajectories in the communication of science

In a context of increasing access to science by the public, a new possibility in the communication of science is introduced – that of the public accessing scientific discoveries before validation by peers has been completed: 'Open Science is about removing all barriers to basic research, whatever its formats, so that it can be freely used, re-used and re-hashed, thus fuelling discourse and accelerating generation of innovative ideas' (Grigorov & Tuddenham, 2015). And the motivations of the public for doing so may well extend beyond stimulating debate and innovation as the communication of science becomes politicised (Scheufele, 2014).

Thus, the possibility of what Bucchi terms 'alternative' or deviant trajectories of science communication manifest alongside the 'routine trajectories' described by the continuity model of science communication (Bucchi, 2004, p. 120; Bucchi, 1996).

Bucchi's deviant trajectories describe those instances in which scientists elect to bypass the routine trajectory to communicate directly with the public at an early stage in the formal communication of science.[9] These deviant trajectories outside of the routine process of science communication are excluded from this study. This study focuses on emerging forms of deviant trajectories made possible by open science; trajectories that are less an outcome of purposive attempts by scientists to deviate from the routine trajectory of communication, and more an outcome of both scientists and non-scientists accessing specialist scientific communications made possible by open science. As such, this research focuses its investigation on the open segment of the science communication system, and within this open system, it will focus specifically on open research data and on open access journal articles. In other words, open data and the open access journal articles are the communication objects under study to the exclusion of other possible communication objects such as open access monographs, chapters in books, figures and illustrations, or open educational resources such as textbooks and other study materials.

For the purposes of this research, texts and data generated and shared by researchers outside of the science communication system are excluded. This may include research data from non-governmental organisations, private research facilities, consultancies, think-tanks, etc. This is not to suggest that research data or texts from non-university sources do not enter into the science communication system. Researchers from NGOs and consultants are known to hold honorary academic posts such as research fellows or extraordinary professors, and these credentials confer more ready access to knowledge validation and codification systems of science. But it is only when they submit data or written products (e.g. journal articles or monographs) to the scientific process of validation and codification that their products fall within the ambit of this investigation.

3.3 Non-scientific publics

The contemporary science communication literature most often uses the plural 'publics' to indicate a multiplicity of sub-groups or communities, each with varying levels of skills, knowledge and interest, that constitute the generic public (Grand, Wilkinson, Bultitude, & Winfield, 2014). At an individual level, the descriptions for the members of sub-groups include 'non-scientist', 'lay person' and 'non-expert'.

[9] see Bucchi (2004, pp. 117-122) for examples of deviant trajectories of science communication.

Bucchi (2004, pp. 111-3) has drawn attention to the problems of using the distinction 'expert' and 'non-expert' to cleave into two distinct groups those who possess expert knowledge and those with lay knowledge. Scheufele (2013), acknowledging the diversity of the publics of science, suggests a typology of non-scientists. The typology extends along two continua: (1) an attentive–non-attentive continuum to capture the fact that not all non-scientists are attentive to science; and (2) an expert–non-expert continuum to capture the fact that not all non-scientists are equally tooled to engage with science.

Schäfer et al. (2018) identify four segments within publics: (1) the 'sciencephiles', with strong interest in science, extensive knowledge and a belief in its potential, and who use a variety of sources intensively; (2) the 'critically interested', with strong interest in and support for science but with less trust in it, and who use similar sources but are more cautious toward them; (3) the 'passive supporters' with moderate levels of interest, trust and knowledge, and tempered perceptions of science, who use fewer sources; and (4) the 'disengaged', who are not interested in science, do not know much about it, harbour critical views toward it, and encounter it mostly through television.

In this research, the term 'non-scientist' is used to delineate those members of the public who are not part of the scientific community, where the scientific community comprises those whose primary purpose is the creation of new knowledge and, as such, operate within the formal science communication system to codify knowledge. To use the continuity model of science communication (see above) to draw the distinction more clearly, those individuals who produce knowledge products in the inter- and intraspecialist stages of the communication trajectory are scientists; everyone else is considered a non-scientist.

3.4 Intermediation and trust in science communication

Intermediaries fulfil an important role in bridging science and society by facilitating in one way or another the transfer of specialist scientific knowledge to non-expert audiences (Kahan, Scheufele, & Jamieson, 2017; Landrum, 2017). Their roles as communicators are seen as being linked to several goals: (1) increasing and directing research inquiry; (2) disseminating the findings of science to both the science community and to multiple publics; (3) connecting science and policy; and (4) engaging new stakeholders in their bridging attempts (Landrum, 2017). Intermediaries in the communication of science may include professional societies, private foundations, scholarly publishers, science journalists, government agencies and university communication offices (Landrum, 2017).

Irwin (2014) in his reflection on the evolution in the thinking of science communicators away from the deficit model and towards the public engagement model, questions the hasty migration away from deficits:

> Deficits are fundamental to many forms of communication and as such can never be discarded. […] [T]he point was (and is) not to deplore the existence of 'deficits' but to note the regularity and persistence of certain deficit forms. […] I have never thought that talk exists in a social, political and institutional vacuum. (Irwin, 2014, p. 73)

Scheufele (2014) also draws attention to the limits of the public engagement of science model. He argues for an expanded conception of the public engagement model of science communication to account for the socio-political dimensions of science communication. He suggests the inclusion of 'mediated realities' to account for the fact that most encounters between science and its publics do not occur in any direct form of engagement, but rather via indirect channels such as the media and other intermediaries (including online search engines, high-profile bloggers, content aggregator applications and influential individuals in the social media).

While Scheufele's expansion adds an important level of complexity to the communication of science, it does not seem to acknowledge the possibility, in a world of open science, of another form of indirect communication: when science is open, and scientists communicate in the formal science system of communication, actors in the public domain have access to their communications. Unlike Scheufele's (2014) expanded model, the boundaries between science and society are more porous allowing non-scientists to access science-to-science communications directly, without any effort on the part of the scientist. This is a distinct form of science communication that is both deviant and latent in nature. And it seems reasonable to expect that within these deviant communications, one would also find intermediaries who mediate between the formal communication of science and publics.

As the knowledge products of science, including research data and journal articles, become more accessible, attentive non-scientist publics are presented with a saturated and pluralistic universe of intermediaries, and they must choose whom to trust as authoritative sources of 'scientific' information (Blöbaum, 2016). The challenge for intermediaries is to stand out from the crowd of intermediaries and to establish and maintain credibility with their target audiences (Landrum, 2017). This has prompted Kahan et al. (2017) to refer to 'elite intermediaries' to differentiate between those trusted intermediaries (professional science journalists, government agencies and publishers) and all other intermediaries.

How trust is established between non-scientist publics and intermediaries in the communication of science is, however, poorly understood (Scheufele, 2014; Weingart & Guenther, 2016). Schäfer (2016) argues for a greater acknowledgement within the field of science communication of the role that trust plays in the intermediation process of communicating science. In the eyes of some audiences, government agencies, the media and even publishers are not to be trusted. Weingart and Guenther (2016) argue that trust is in part a factor of intent in relation to the public good. Those whose intentions are in the public interest (for example, firemen) are trusted more than those perceived to harbour intentions that are self-promoting (for example, politicians). But the markers or social cues for establishing trust aren't always available or may be replaced with new cues when the communication of science is interpersonal and takes place in online communication networks such as the social media. In such forms of communication, it remains an open question why publics are receptive to the communications of selected non-scientific intermediaries (Southwell, 2017).

3.5 An elaborated model of science communication

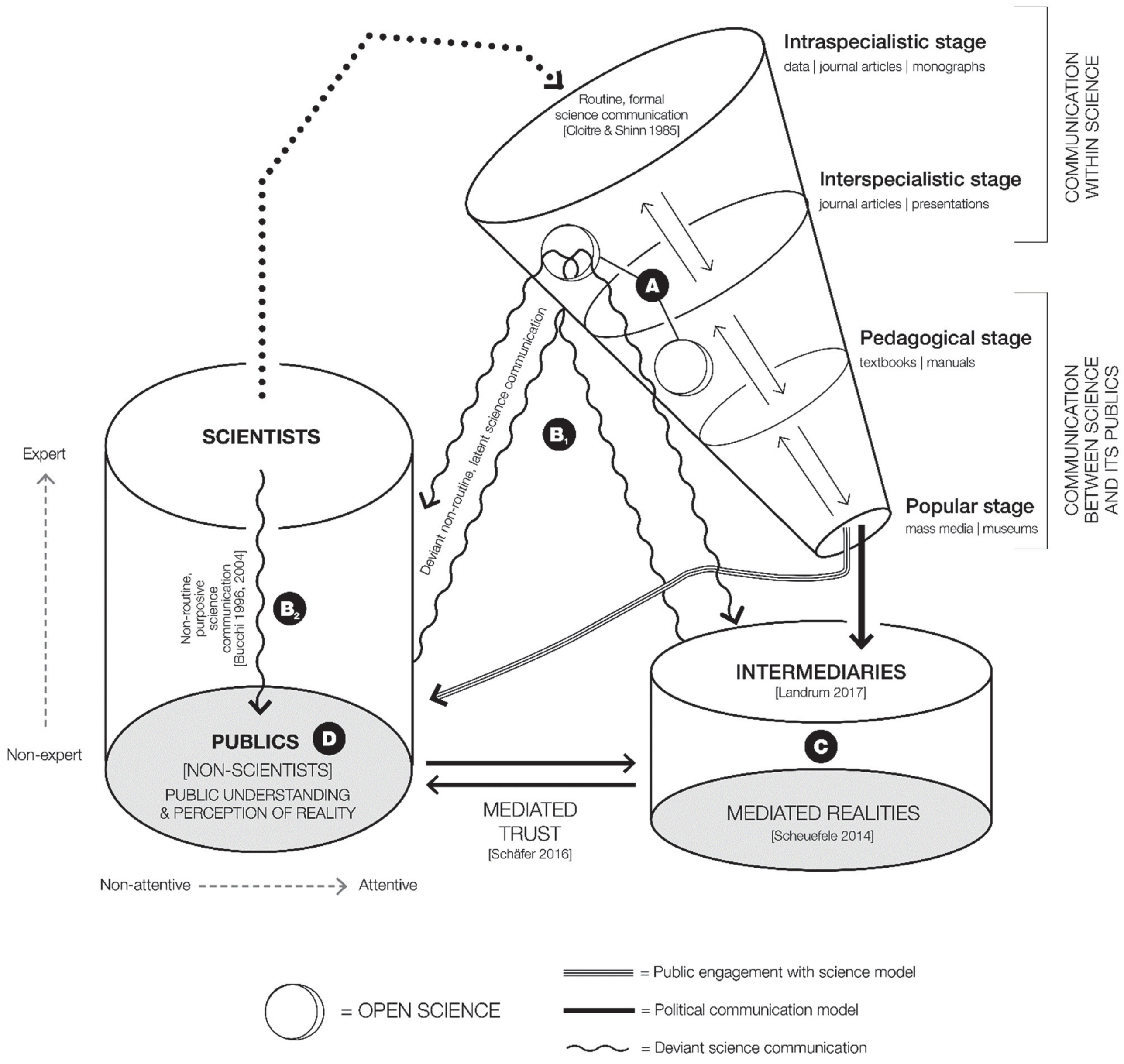
Given the above description of the process of formal science communication in the digital age, the continuity model first proposed by Cloître and Shinn (1985 in Bucchi 2004) (see Figure 1) is elaborated as shown in Figure 2.[10] In particular, the continuity model is described in greater detail to reflect a changed communication environment by adding four components represented by A, B, C and D in the figure.

The first of the elaborations to the model are the two openings in the funnel (A). These are added to indicate that the formal communication of science is more porous than was previously the case. Science is more open, and this makes possible access to the formal communications of science in the inter- and intraspecialist stages. Access to these two early stages by non-scientists is represented by the wavy lines (B_1) which connect those outside of the formal science communication process to the products of that process (e.g. journal articles, research data, monographs, conference proceedings and the like). This form of access to science communication is different from the purposive, non-routine, deviant communication (B_2) described by Bucchi (1996, 2004). Access requires no active engagement by scientists and is therefore described as being a latent form of communication. The

[10] It is worth noting than in a presentation titled 'The Crisis of Mediation: Science Communication in the Digital Age', Bucchi includes an image of the continuity model with the headline "The good old days…?" (Bucchi, 2014).

fact that the lines are looped with arrows at their end-points pointing away from the science communication funnel indicates the access and retrieval of the products of science by publics. Common to both purposive (B_2) and latent (B_1) deviant communications is the fact that knowledge not yet codified by consensus within science (that is knowledge that has progressed to the pedagogical and popular stages) is being made accessible to non-scientific publics.

Figure 2: Conceptual framework for the formal communication of science and new modes of access to the communication products of science



Access to the formal science communication process is either direct or it may take place via intermediaries (C) such as science journalists reporting in the media, documentaries on television, popular books, museums, science exhibitions and the like. Unlike the public engagement with science model in which publics interact directly with the 'end-products' emerging from the formal communication process, indirect or mediated science communication shapes how science and its truth claims are received and perceived by publics (Scheufele, 2014). Such mediation is inevitably political and may be located either between the formal outputs at the popular stage or may take place as intermediaries retrieve knowledge-in-progress from the intra- and interspecialist stages of the formal science communication process.

The publics (D) who receive information from and about science from various sources are either attentive to or disinterested in those sources of information, and they will vary in their level of expertise in relation to the scientific information communicated. It is presumed that those who retrieve information from science are inherently attentive to science or at least to fields within science that overlap with their interests.

3.6 Communication networks

Communication cannot be separated from the transmission of thought: 'there is no thinking except as aftermath or preparation of communication' (Collins, 2000, p. 2). If the continuity model describes the process of science communication (that is, the transmission of scientific thought), and the elaborated model accommodates recent developments in the communication of science by introducing new pathways for the communication of science, then it is the network that describes the interaction of actors and objects that propel the process of communication in the modern age (Castells, 1996; 2009). As a socially constructed space, the relationships between social actors (and objects) in networks of communication in the age of digital technology is key to understanding the delivery, reception, use, re-use and impact of science communication.

Digital media and infrastructure create an integrated environment based on flows of information. Increasingly, this environment provides the primary setting for human agency. Castells's (1996, 2009) network society provides a particularly useful grand narrative for any analysis of communication in general and of science in particular. Mainly this is because Castells's description of the network society accommodates a contemporary paradox, that is, the persistent authority of and trust in science (Weingart & Guenther, 2016), and a simultaneous loss in the trust of contemporary institutions and their authority in society (Bucchi, 2018; Ortiz-Ospina & Roser, 2016). Post-modernism, a relatively recent sociological macro-theory, is anti-authority and, as some have claimed, explains the erosion of scientific authority as everyone becomes an expert of their own reality (Kata, 2012). In such a world, everything is contextually bound and with no transcendental truth, authority is in perpetual flux. This has serious implications for science if the central project of science is the production of universal truths and, as such, assuming a position of authority in society over truth claims (Collins, 2000).

A network understanding offers a way out of this dilemma; an understanding that can explain both the erosion of trust and attempts to displace the authority of science, as well as its continued authority. A shift to an understanding of communication that relies on networks, and one that depends on the non-hierarchical relationship between nodes across the globe and in the spatially unbounded flow of information between nodes in real time, makes possible both the authority of the truth in certain networks and the production of uncertainty in others. The effects of global communication networks are experienced without exception by an increasingly connected species; but it is an experience that is shaped socially by interconnectivity at a global scale rather than by individual relativism.

Rather than provide an extensive account of Castells's theory of the network society in the information age, this section will limit itself to a brief extrapolation of a few key concepts that, when linked together, cast a particular light on the functioning of science communication in a world in which there have been fundamental shifts in the modes of production and of communication.

3.7 Information flows in communication networks

According to Castells, the basic elements of the network society are not material, but the intangible flows of information produced by and processed through media: Information to communicate among people, to control processes across time and distances, to check and re-evaluate existing information, and to produce more and new information.

3.7.1 The space of flows

The key concept that Castells introduces is that of the space of flows, defined as 'the material support of simultaneous social practices communicated at a distance. This involves the production, transmission and processing of flows of information' (Castells, 2010, p. xxxii).

The space of flows, according to Stadler (2006), is that stage of human action whose dimensions are created by dynamic movement, rather than by static location. The operative words in Stalder's definition are 'movement' and 'human action'. Without movement, the space of flows would cease to exist, and the space of places would resume centrality. Movement takes place through human action and creates specific social conditions in the day-to-day realities of humans that are characteristic of the network society.

It is not that networks are new but, as Castells explains, digital information networks introduce new realities of communication and therefore, by implication, of social relations. The space of flows brings distant elements (things and people) into an interrelationship that is characterised by being continuous and in real time (Castells, 1996). From a historical perspective, this conflation of spatial and temporal separation is new.

The space of flows as networked communication consists of three elements: (1) the medium through which information flows, (2) the information that flows and (3) the nodes through which the flows circulate. There is always a close relationship between the medium of the flows and their contents (Stalder, 2006). Flows travels from one node to another; nodes focus movement into flows, they connect various flows to one another, and they also connect flows with places.

In a world in which information circulates through digital media, nodes and flows are becoming increasingly differentiated. As volume and pace of the flows increases, nodes and flows are becoming more and more different logically, while functionally they are being integrated ever more tightly (Stalder, 2006).

Flows and nodes are mutually constitutive: flows without elements of structure would be noise and nodes without flows would be isolated and irrelevant. The interconnections between the nodes constitute the patterns in the flow of information and provide stabilisation within fluid environments making possible navigation and purposeful, systematic action.

While the nodes stabilise the flows of information and endow a certain consistency and continuity to the environment, they are themselves subject to the dynamics of the environment. These dynamics, which are produced by the interactions of the nodes and shaped by the media that channel the flows, are themselves not reducible to any single node but are the result of the combination of all flows, of the interaction of all nodes at the given time reflecting their different capacities to influence. The dynamics, however, are not random; they have discernible patterns in which they develop.

The social reality of the space of flows is therefore neither immaterial nor self-contained; rather it deeply affects the material world from which it is inseparable (Castells, 2010; Stalder, 2006).

3.7.2 Global communication networks

According to Castells (2009), there are multiple global communication networks. However, the contours between global information networks are not always sharply defined. Networks overlap and are influenced by one another, and networks compete and defend themselves from one another. One cannot therefore understand one network without reference to other networks, although Castells argues that it is the global financial network that dominates in the current global capitalist economic dispensation (Castells, 2009). Other global networks include: transnational production, management, and the distribution of goods and services; highly skilled labour; science and technology, including higher education; the mass media; the internet networks of interactive, multi-purpose communication; culture; art; entertainment; sports; international institutions managing the global economy and intergovernmental relations; religion; the criminal economy; and the transnational NGOs and social movements that assert the rights and values of a new, global civil society (Castells, 2009).

The different communication networks function according to different programs, programs that are determined for each network by those who wield network-making power (Castells, 2009). Castells (2004, pp. 32-33) describes the creation of network programs and the centrality of communication in this process as follows:

> [A]ll these networks do have something in common: ideas, visions, projects generate the programs. These are cultural materials. In the network society, culture is by and large embedded in the processes of communication, in the electronic hypertext, with the media and the internet at its core. So, ideas may be generated from a variety of origins, and linked to specific interests and subcultures (for example, neoclassical economics, religious fundamentalism of various kinds, the cult of individual freedom, and the like). Yet, they are processed in society through their treatment in the realm of communication. And, ultimately, they reach the constituencies of each network on the basis of the exposure of these constituencies to the processes of communication. Thus, control of, or influence on, the apparatuses of communication, the ability to create an effective process of communication and persuasion along lines that favour the projects of the would-be programmers, is the key asset in the ability to program each network. In other words, the process of communication in society, and the organizations of this process of communication (often, but not only, the media), are the key fields in

> which programming projects are formed, and where constituencies are built for these projects. They are the fields of power in the network society.

A network is therefore defined by the program that assigns the network its goals and its rules of performance; in other words, the core logic of the network. A network's program consists of codes for the evaluation of performance and criteria for success or failure in the network. To transform the outcomes of any specific network, a new program emanating from outside the network must displace the existing program of the network, and control over communication is a key determinant in the outcome of attempted displacement.

The program also determines the speed at which information is communicated and accepted (Castells, 2009; 2010). In the global financial network, information is exchanged in real time and on the basis of implicit trust which is ensured by nodes created within the network with the express purpose of conferring unquestioning trust in the information that flows across the network (Stalder, 2006).

This is markedly different from how information flows in the global network of science. The network of science is programmatically sceptical. Collins (2000) argues for the distinctiveness of networks of scientific thought. Pels (2003: 2-3) argues for 'the preservation of a place of quiet, stillness, and unhastened reflection, which must be incessantly negotiated for and demarcated against the speedy cultures of the "outside" world'. Science must constantly struggle against the interests of other networks to protect and maintain its autonomy. Davies (2016), homing in on the rise of data in what he terms 'post-truth' politics, makes the claim that it is possible to live in a world of data but no facts, a world where data is a proxy for sentiment rather than the factual basis on which statements about reality rest. While Davies is concerned about politics rather than science, the displacement of fact by data is no less relevant.

Weingart et al. (2000) proposed the concept of communication spheres of science, politics and the mass media. Each sphere has its own rules that allow the sphere to attain its unique objective. In the case of the mass media, the primary medium for communicating science at the time, the main objective is to capture attention. Captured attention can be monetised, and this brings the spheres of communication and financial markets into close proximity (Wu, 2016). Actors are connected within and between spheres, and information laden with cues (credibility, legitimacy, authority, relevance, novelty, etc.) flows between them. While conceptually similar, networks complement the notion of communication spheres by inserting the interconnectedness of nodes (communicators) and the relationship between those nodes.

The online social media as global communication networks combine the elements of media attention and of socially structured communication. Social media networks are therefore programmatically a network of attention consisting of nodes clustered primarily according to social norms, values and beliefs. This does not mean that attention in social media networks can't be monetised in the way that the mass media depends on monetisation, but it is not the primary logic within social media networks. The distinction is premised on the fact that attention-seeking in social media networks operates at the individual level while the monetisation of attention is a product of mass attention (Wu, 2016). In other words, individuals seek attention in social media networks, and only if they reach a critical mass, does monetisation of the attention attracted become possible. The most popular social networks such as Twitter, Facebook and Instagram may have different affordances (Valenzuela, Correa, & Gil de Zuniga, 2017) but they have in common simple metrics such as number of followers, friends and likes which provide clear cues of the potential for monetisation (Wu, 2016).

3.8 Publics as networked communities

Communication networks are the patterns of contact and exchange that are created by flows of signal and information between communicators through time and space (Monge & Contractor, 2003, p. 39). In other words, as users interact in social media spaces, they form connections that emerge into complex social network structures. These connections are indicators of content sharing, and the network structures reflect patterns of information flow. Patterns of connections among users within social media take the form of social networks and the structure of a network is an indicator of its unique patterns of information flow.

But from these complex network structures also emerge communities – clusters of social actors (or nodes in the language of networks) – in close proximity with one another. The reasons for the emergence of clusters in communication networks may be because of shared aspiration, purpose or identity. Regardless of the underlying cause, the communications between social actors reveal their proximity and closeness.

A group that attracts much of Castells's attention is the social movement, defined as a collective of social actors acting self-consciously to effect change of any kind (Stalder, 2006). Such movements are central to this thesis because it is the use of more readily accessible science in communication networks by a specific social movement to effect change that is of interest. In this sense, use at the individual level holds little interest; it is only when agency becomes collective, when individuals act with reference to other actors who are acting similarly, that they form a social movement. Such collection action must be purposive in the sense that there is a self-defined long-term objective and there is acknowledgement that action advances toward the objective.

Social movements are, according to Castells, self-conscious. Social movements are able to present an analysis centred around fundamental questions of identity and appropriate action (as defined by the movement irrespective of whether the formulation is meaningful to outsiders), define who their members are, and put forward a strategy for collective action (within their own frame of reference) to improve the conditions of their members.

Social movements are also no longer bound by place to the same extent as they were in a pre-network society era. Because they are no longer seen as purely reacting collectively to social ills but rather as constructing identity, social movements are more readily dislocated from place and are able to function in real time on a global scale (Stalder, 2006).

The globalisation of social movements made possible by developments in information and communication technologies has increased the potential effectiveness of social movements; change that may be for better or for worse. Castells therefore remains ambivalent about the social potential of the network society (Stalder, 2006).

To revisit the elaborated science communication model presented above, publics can be reconceived in the network society as social movements dislocated from place but bound by common, self-defined identities and able to define a strategy for collective action to improve the conditions of the movement's members as reflected in their patterns of communication. In this sense, the anti-vaccination movement fits comfortably within the conceptual framework developed.

3.9 Revisiting intermediation and trust in networks

The ability to exercise power in networks (where power is the structural capacity to impose one's will over another's will) varies according to the program of the network (Castells, 2009). Programmers hold the power to program or reprogram networks in relation to the principal objective of the network; switchers hold the power to control the connecting points between networks to unlock common objectives (Burt, 2001; Castells, 2009). Crucially, within this formulation, people, collectives and organizations are understood as social actors who through joint action have the ability to disrupt the dominant power switches, and/or to push for change through what Castells calls counterpower (Castells 2009).

The roles of intermediaries can be reformulated in network terms as those centrally located actors who are in a position to exercise power in communication networks. Their influence is a particular type of power made more potent by the effects of the network. Theirs is the power to influence because of their special and relative position as switchers and/or as the reprogrammers of communication networks which, in turn, allows them to control the flow of information – in Burt's terms (Burt, 2001), they exploit structural holes in networks. These intermediaries are not individuals but rather nodes or configurations of nodes in networks that, by virtue of their position in or between networks, are able to make new connections (switch) or to challenge the dominant programme of a specific network (reprogrammer). Their position in networks affords these actors a high degree of attention from those in their own networks but also, potentially, from those in other networks. In the network society, their position of influence is therefore always relational – it is dependent on their ability to connect to others in the network and to mobilise collective rather than individual action (Stalder, 2006).

In networks, new information cues emerge alongside more traditional social cues that allow social actors to evaluate who holds influence (Lin, 2008). These cues are particularly strong in social media networks where they present as followers and likes – simple metrics that indicate to others in the network the centrality of a potential influencer.

According Blöbaum (2016), the digitization of the media brings trustors and trustees closer together, and trustors have access to more sources allowing them to check and verify information more effectively. This understanding resonates with a network approach in which all social actors are connected to all other actors. It does, however, assume that recipients of information in the digital domain will make use of the opportunities presented by digitisation and the internet to check information. In online communication networks, it is possible that this does not happen, and that recipients take the information presented to them at face value. The reason for this is a structural condition of networks – the logic or programme of the network may determine that information must flow rapidly, negating the possibility for fact-checking and/or deferred decision-making. Instead of an increase of trust between actors in communication networks because of a multitude of readily accessible information sources, trust may in fact be implicit in certain networks because the network demands it.

The suspension of judgement is possible only if trust is implicit. In the case of some communication networks, trust mechanisms may be created purposefully to allow information injected into the network to be taken at face value (Stalder, 2006). In other cases, it is possible that trust may be 'borrowed'. Either way, if a communication network thrives only if information flowing through it is taken at face value, then it is also true that the level of engagement with that information is both rapid and shallow to prevent flows from becoming viscous and nodes from becoming inactive.

If science remains as trusted as surveys consistently indicate (Schäfer, 2016), then the information taken from the products of science may be sufficiently trusted in some networks for that information also to be taken at face value. The irony, of course, is that while scientific information may travel in some communication networks with a high degree of trust attached to it, in the scientific network itself, all judgement on the trustworthiness of a piece of information is suspended until it passes through a series of verification processes.

To guarantee that the information that circulates within communication networks can be taken at face value, the network must be separated structurally from other potentially competing networks. The constant flow of information must be protected to avoid interruptions from external events which would destroy the face value of the information. Trusted nodes that intermediate information ensure the functioning and the survival of the communication network.

3.10 Hypotheses

Based on the above elaboration of the continuity model and the application of a network approach to understanding some of the dynamics in the communication of science in the information age, the following hypotheses are formulated:

1. Non-scientists are accessing scientific products made accessible by open science.
2. While certain publics may be highly attentive to science and are accessing its products, the level of their engagement with the products of science in social media networks is low.
3. Intermediaries in the communication of science are not only located at the interface between popular science and the media, but in the new, deviant flows of scientific information made possible by open science.

H_1 Non-scientists are accessing scientific products made accessible by open science

The digitisation of the communication products of science made available on the world wide web introduces at the very least the potential for both scientists and non-scientists alike to access those products in the form of open access journal articles, open research data and the like in ways previously unimaginable. Open science both broadens and deepens the potentiality of that access by affording access to non-scientists. It is hypothesised that non-scientists are in fact accessing certain products of science, particularly those produced in the early stages of formal science communication, at a point when neither consensus within the scientific community nor the establishment of scientific fact has taken place.

H_2 The level of engagement by non-scientists with the products of science is low

Global communication networks depend on decontextualized information being circulated in real time (and therefore at high speed) across the network. Rather than suspending judgement in order to evaluate information before arriving at a decision, decision-making is rapid. This requires information in the network to be read at face value and precludes the time-consuming verification of information circulating in the network. Social media networks create environments that mimic, at least in part, the global financial and other networks. When information is harvested from scientific products by non-scientists and fed into social media networks, they too are taken at face value within non-scientific online communities, especially by those communities that can be described as social movements. Consequently, the level of engagement with the content of scientific products is low.

H_3 Intermediaries are located in the new flows of scientific information

To guarantee that the information that circulates within communication networks can be taken at face value, the network must be separated structurally from other potentially competing networks. In the global financial network, it is the clearing house that makes this possible by institutionalising a system of trust designed to protect the programme of the network against events unfolding in a chaotic environment. Without this buffer, the exchange of information would slow down considerably because the value of the information would have to be verified outside the network itself. The clearing house in the global financial network therefore protects the constant flow of information from being interrupted by external events which would destroy the face value of the information. Networks other than the global financial network require similar central, trusted nodes that intermediate information to ensure the functioning and the survival of the communication network.

Table 2 Shows the relationship between the research questions and the above hypotheses as well as additional fine-tuning of the hypothesis in order to operationalise the research.

Table 2: Research questions and hypotheses

	Question	**Hypothesis**
1.1	Are non-scientists accessing open research data?	Non-scientists are accessing (connecting to) open research data.
1.2	Are non-scientists accessing open access journal articles?	Non-scientists are accessing (connecting to) open access journal articles.
2.1	What are non-scientists' levels of engagement with open research data?	Non-scientists' level of engagement with open research data is low.
2.2	What are non-scientists' levels of engagement with open access journal articles?	Non-scientists' level of engagement with open access journal articles is low.
3.1	Who are the non-scientists mediating connections to the products of open science?	Intermediaries are located in the new potential flows of scientific information.

Chapter 4. Method

4.1 Introduction

The methodology for collecting and analysing data was designed to answer the research questions posed by this study related to the use of two products produced in the formal communication of science: open access journal articles and open research data, and, more specifically, whether these products are being used by non-scientists in their communications in the online social media and the world wide web.

The anti-vaccination movement was selected as a particular case of an active online social movement attentive to science. Methods were selected that would generate informative results on whether the anti-vaccination movement is accessing the products of open science, whether they are using those products in support of their objectives, and whether there are members of the movement that intermediate in the access and use of open access journal articles and open research data by the broader anti-vaccination movement.

The results generated from this approach are likely to answer the overarching interest of this dissertation, that is the potentials of open science in the communication of science.

This chapter starts by providing a description of the research design along with justifications for the choices made in selecting the approaches described. This is followed by an overview of the data sources used in the study, before the methodology used to collect and analyse data is presented. The methodology comprises several approaches, each of which was designed to answer the three research questions posed by this study. The methodology section starts by describing how a sample of relevant, traceable open access journal articles and open research datasets was created. This was an important step in the process as it formed the foundation for each of the subsequent steps in the process. The chapter concludes by outlining the limitations of the research design and other considerations to be taken into account in relation to the design of the study.

4.2 Research design

4.2.1 A digital methods approach

Kahan et al. (2017) identify four domains in empirical research on science communication, one of which is the communication *of* science which concerns the transfer of information from science to non-scientific audiences. It is within this domain in the study of science communication that this research endeavour is situated. Given the interest in furthering the understanding of science communication in this particular domain with its interest in non-scientific audiences, it was deemed necessary for the research to be designed in a manner that would maximise the likelihood of the results being grounded in the social dynamics that characterise the communications of a non-scientific public in online spheres.

Furthermore, given the digital nature of the 'subject' being studied, that is, data and publications as digital objects; its presence in online communications; and the increase in empirical research that relies on data from the social media (Zimmer & Proferes, 2013), the web and other publicly accessible digital sources to reconstruct social dynamics (Kozinets, 2010; Rogers, 2015), this study relied predominantly on digital methods. Digital methods are defined as 'techniques for the study of societal change and cultural condition with online data' (Rogers, 2015, p. 1); in other words, approaches that rely on the web and other online spheres as constituting a dataset.

The primary source of data for this study was therefore the social media (specifically Twitter and Facebook) and the world wide web (specifically the websites of anti-vaccination individuals, groups and organisations). Zimmer and Proferes (2013) have shown that in the period from 2006 to 2012, 380 studies across 17 disciplines relied on Twitter as a data source. Beer (2013) argues for a better appreciation of what he terms 'the politics of circulation', changed exponentially by new (social) media infrastructures, that determines how scientific knowledge is produced, disseminated and interpreted by multiple publics. However, Brossard (2013) points out that within the field of science communication studies, there is a dearth of empirical studies that exploit the potential value of Twitter in particular as a communication platform in a rapidly changing science communication landscape.

A digital methods approach aligns with this study's particular interest in the globally networked communication of science and makes possible the study of such communication in a research endeavour with relatively limited resources. Further benefits of a digital methods approach include the removal of dependency on subjects to respond to requests for data (by completing questionnaires or agreeing to be interviewed) and cost efficient data methods of data collection (data can be accessed from any location, there are no transcription costs, and many of the data collection and analysis instruments are available at no cost) (Kozinets, 2010). There is also a growing acceptance that 'if handled appropriately, data about internet-based communication and interactivity could revolutionise our understanding of collective human behaviour' (Watts, 2007 in Rogers, 2015, p. 2).

There are three approaches to using the web as a source of data for the study of social phenomena, each of which is determined by how the data are accessed (Rieder, 2013):

1. Direct database access to a data owner's servers where such access is reserved for in-house researchers or on the basis of cooperation between a data owner and external researchers. Certain data owners do also make data 'donations', for example Twitter transferred its complete archive to the US Library of Congress. The data made accessible in these ways are generally very large and well structured, but often anonymised or aggregated.
2. Access to data via a sanctioned application programming interface (API) where such access is made possible by machine interfaces provided by many Web 2.0 services to third-party developers with the objective of stimulating application development and integration with other services in order to provide additional functionality and utility to users. APIs provide well-structured data, but are generally limited in terms of which data, how much data, and how often data can be retrieved. Conditions can also vary considerably between services. Twitter is, for example, fairly open while Facebook is relatively restrictive in terms of what data can be accessed. Data owners also retain the right to modify or close their data interfaces, which can lead to substantial problems for researchers.
3. User interface crawling which is typically done manually and/or with the assistance of bots or crawlers. Such an approach can circumvent the limitations of APIs, but often at the price of much manual data collection and analysis, as well as the technical and legal uncertainties if a data owner's permission is not explicitly granted.

This research project followed the third approach because it provided (1) relatively unfettered access to the data and (2) access to most granular data available.

4.2.2 The case of the anti-vaccination movement

Open science enables collaboration, co-operation and contributions to science by non-scientists (for example, citizen science) (Grand, Wilkinson, Bultitude, & Winfield, 2014). However, open science also enables increased contestation as non-scientists are able to access both codified or 'settled' knowledge as well as knowledge that is contested and 'in-progress'. The focus of this research is on instances of contestation as this is likely to generate increased levels of activity in the social media, and such activity is critical to measure the digital traces of access to and use of open research data and open access journal articles by non-scientists.

A case study approach was chosen because it takes empirical research beyond experimental, lab-based designs that are limited in terms of scalability, reliability and relevance to the more complex social conditions outside of the lab (Babbie & Mouton, 2001; Stake, 1995). Kahan et al. (2017) suggests that there is a need for such research in the study of science communication.

The focus of this research is on the contested issue of vaccinations and their side-effects; in particular, the claim that there is a causal link between vaccination and autism. Several factors support the inclusion of the case of the anti-vaccination movement as the focus for this study. First, the anti-vaccination movement is attentive to science in a way that other social movements may not be (for example, the Occupy Wall Street Movement or the Arab Spring movements). The movement refers to science in its online anti-vaccination messaging (Bean, 2011; Bennato, 2017; Kata, 2012; Leask, 2015), and assumes a stance that is contrary to scientific position on vaccination. Second, according to Kahan et al. (2017), the science communication environment that surrounds the issue of vaccination was neglected by science communicators and this has allowed controversy to flourish. There is therefore a need to better understand the consequences of this neglect centred on the particular issue of vaccination. Third, like many social movements, the outcomes of their lobbying and advocacy can lead to real-world societal change, but unlike most contested issues that target change at a local level, the anti-vaccination movement's attempts to influence the decision-making of individuals can have catastrophic health outcomes for society at large.

As social actors interact via social media platforms such as Twitter and Facebook, they make connections that form complex social network structures. These connections are indicators of content sharing, and the resulting network structures reflect patterns of information flow (Himelboim, Smith, Raine, Shneiderman, & Espina, 2017). Southwell (2017) argues that the structure of networks is key to understanding the communication of science.

Social network research has consistently found that given the opportunity to interact freely, individuals and other social actors tend to form subgroups of connected individuals who are more interconnected with one another than with others in their social network. A key characteristic of self-organised networks in which social actors are free to create and delete connections is that they share a common small world structure (Himelboim, Smith, Raine, Shneiderman, & Espina, 2017). This makes the highly vocal, globally mobilised anti-vaccination movement ideal to study the network properties of interest to this research: that information is taken at face value in certain online communication networks and that intermediaries are important network actors in enabling action based on information exchanged in the network.

This specific focus on the link between vaccination and autism is supported by the fact that Moran et al. (2016) found that 65.8% of 480 anti-vaccination websites studied focused specifically on autism as a disease associated with vaccines, followed by brain damage (42.5%) and other neurological disorders (40.4%). The selection of the vaccination-autism controversy is also supported by the large number of dedicated autism-related resources available online (Al-jawahiri & Milne, 2017) as well as the number of journal articles on the topic.

4.2.3 Use of altmetrics

This study is unique in that it places the communication products of science at the centre of its inquiry. All communication under study is centred on either an open access journal article or an open research dataset. These objects may not always be ‘consumed’ in digital format (it is possible to print and read a paper version of a downloaded article), but they are always accessed in digital formats because the products of an open science are, by definition, digital.

The Thompson Reuters Data Citation Index (DCI) indexed 344 data repositories (as at 14 February 2017) and allows for the attribution of citations to datasets in academic journals. However, lack of attribution by *non-scientists* poses a challenge in establishing reliable links between data users outside of the science community and the sources of open research data. A similar problem is faced with the attribution of journal articles.

Rather than looking for attribution in academic journals, the world wide web and the social media provide an alternative source of reference as they are comprised of a more heterogenous population and allow for mentions of scientific research to be tracked back to source (by means of affordances such as hash tags, bookmarks and hyperlinks). The reliance on alternative, web-based rather than published scientific sources situates this study in the realm of alternative metrics or 'altmetrics'.

The use of altmetrics was augured by Cronin et al. (1998, p. 1320): 'Scholars may be cited formally, or merely mentioned en passant in listservs and others electronic discussion fora, or they may find that they have been included in reading lists or electronic syllabi. Polymorphous mentioning is likely to become a defining feature of Web-based scholarly communication'. More recently, Priem et al. (2011) have argued that 'metrics based on "traces" of use and production of scholarly output on social media platforms could help to improve scholarly communication and research evaluation'. Haustein et al. (2016) highlight that the attraction of altmetrics is linked to the possibility of measuring the use and impact of scientific products on the broader non-scientific community. However, the extent to which altmetrics is a proxy for societal impact has been questioned and led to suggestions of the use of social media data in more interactive approaches or at least approaches that do not attempt to replicate the principles of scholarly attribution in non-scientific contexts (Robinson-Garcia, Van Leewen, & Rafols, 2018).

The design of the research, mindful of the limitations of relying on simple metrics as indicators of use, includes more interactive approaches (including content and network analysis) to attain more reliable and valid results. The following section provides an overview of how the different elements of the research design were operationalised.

4.3 Overview of methodology

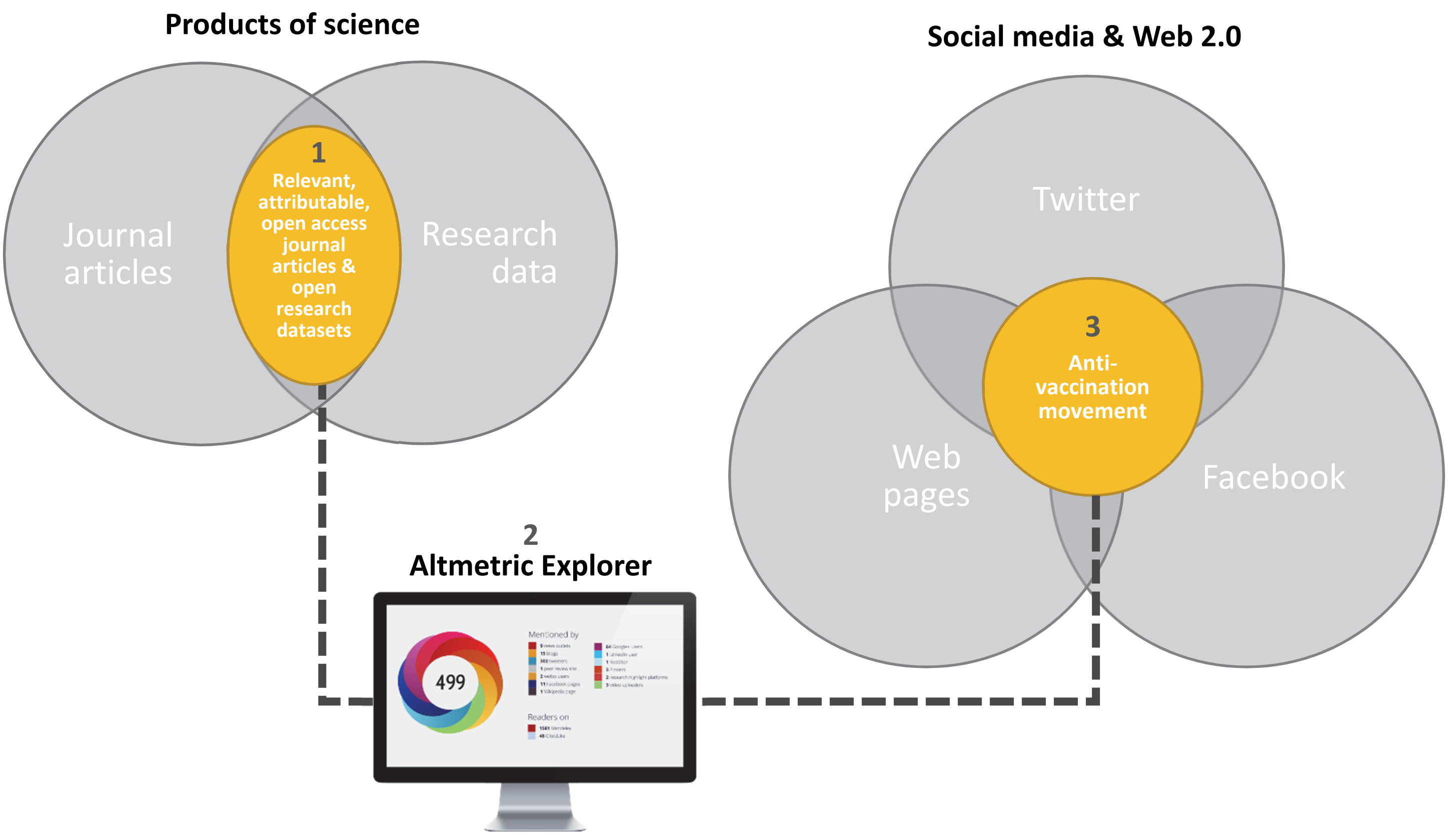
The process of linking or attributing the communications of the anti-vaccination movement in the social media and on the web with relevant communication products of open science is illustrated in Figure 3 below.

The figure illustrates the need to (1) create a sample of journal articles and datasets that are both open and relevant to the anti-vaccination movement from the population of published scientific journal articles and datasets; (2) identify mentions to these products in the social media and on the web; (3) attribute those mentions to the anti-vaccination movement.

Once these three steps were completed, it was possible to conduct further analysis to answer questions related to connections and use, as well as the presence of intermediaries in the online communication of the anti-vaccination movement as it relates to the products of open science.

These initial steps and subsequent multiple methods are described in detail in the sections that follow. Additional detail on the exact steps followed in each case is provided in Appendix 2.

Figure 3: Overview of methodology



4.4 Method for creating a sample of open, relevant and attributable products of science

Because of the reliance on DOIs to identify and track the connections to and the use of scientific articles and datasets, a sample of journal articles and datasets with DOIs was required. And because there is little existing literature of empirical research on the subject of access and use by non-scientists to guide the sampling, two different types of data source were mined to increase the chances of including in the sample products that are indeed accessed and used by non-scientists. In addition, data sources were limited to those that could provide both DOIs for an article *and* its related datasets. By creating a sample of only articles and their related datasets, it could be ensured that both the article and its dataset are related to the same topic. Because there are many published articles without any related published data, but almost all research datasets are related to a published article, the data sources were, in the first instance, limited to those that contained research datasets. Once relevant datasets with DOIs were identified, it was possible to source the DOIs for the related article.

Two approaches were used to create a sample of open and relevant open access journal articles and open research data. The first was a top-down approach. This approach entailed the identification of open access scientific products, published or indexed within the formal science publication system, and that would be of interest to the anti-vaccination movement. This approach is content-specific, but it is agnostic in terms of the relevance of the content to the anti-vaccination movement. In other words, while articles and datasets were relevant to the subject of the relationship between vaccination and autism, their relevance to the anti-vaccination movement was unknown. This introduced the possibility of a sample of subject-relevant but politically-irrelevant articles and datasets, and, consequently, non-existent or low levels of use by the anti-vaccination movement. To counter this possibility, a second bottom-up approach was adopted. In this approach, the online content produced and shared by the anti-vaccination was tracked in order to identify references to scientific content related to the vaccines–autism debate.

The first source of data for identifying research datasets related to the autism-vaccine debate was a collection of online research data repositories. Three research data sources were referred to in order to harvest DOIs that could be used to link references in the social media to specific autism-related open research data: open research data repositories; data journals; and bibliometric databases. From these three sources, a single list of DOIs was produced.

The second source was a collection of web pages produced by members of the non-scientific community under study. Purposive sampling was used to create a sample of dataset DOIs from online collections and repositories, while snowball sampling was used to create a sample of dataset DOIs from the anti-vaccination web pages.

The sampling processes are described in detail below.

4.4.1 Top-down (purposive) sampling

Data sources were identified by reviewing the literature, by drawing on personal expertise and by consulting the Thomson Reuters' Data Citation Index. Three non-field specific, open research data repositories were identified: Dryad, the Harvard Dataverse and Figshare.

Dryad[11] is a research data repository developed by a group of leading journals and scientific societies in evolutionary biology and ecology to provide online access to datasets associated with a published journal article. As of February 2017, 111 journals integrated their data submissions with the Dryad repository. Dryad describes the datasets in its repository as 'scientific and medical data'. Authors are charged a fee of USD120.00 (as at February 2017) to submit data to Dryad. Access to the data is open under a creative commons public license. The Dryad curators review submitted data files and perform some levels of quality control, including editing the metadata descriptions of the data files submitted. All datasets are assigned a DOI.

The Harvard Dataverse[12] is an open source web application developed at Harvard's Institute for Quantitative Social Science, along with others worldwide, for the sharing, preservation, citation, exploration and analysis of research data. As of February 2017, the Dataverse contained a total of 47 981 datasets of which 47.1% are in the social sciences followed by medicine, health and life sciences at 14.6%. There is no cost to upload or to access data on the Dataverse. Harvard Dataverse's default data usage license agreement for all uploaded materials is a Creative Commons Zero ('CC0') Public Domain Dedication Waiver. All submissions are assigned DOIs to allow for data citation.

Figshare[13] is a repository where researchers can publish research outputs online in a citable, shareable and discoverable manner. Unlike Dryad and the Harvard Dataverse, Figshare was developed to accommodate a variety of non-traditional research outputs such as illustrations, graphs, tables, presentations, working papers and datasets. Figshare is a private company. It 'receives support' from Digital Science and is listed as a product on the Digital Science website; Digital Science, in turn, is 'operated by' global media company, the Holtzbrinck Publishing Group (owners of, amongst others, Nature, Springer, BioMed Central and Palgrave Macmillan). There is no cost to submit or access Figshare research data. Figshare does not curate data submitted. Research outputs are published on Figshare under the Creative Commons 4.0 licenses, including CC-BY and CC-0. Figshare recommends to users assigning a CC-0 licenses for data and databases. All research outputs made publicly available on Figshare are allocated a DOI at point of publication. Researchers are not required to make the research outputs public at the time of submission.

[11] http://datadryad.org
[12] http://dataverse.org
[13] https://figshare.com

Thirty-three (33) autism-specific data repositories were identified from research done by Al-jawahiri and Milne (2017). The authors do not, however, provide a satisfactory definition of open data, claiming that payment of a 'reasonable fee' for a dataset does not preclude it from being open. Nor do they make reference to the other fundamental characteristics of open data such as accessibility, timeliness, the assignment of open licenses and the machine-readability of the datasets. The autism-related data sources identified by Al-jawahiri and Milne (2017) were evaluated for openness using four key characteristics of open data: available online, accessible, machine-readable and published under an open license. These four key characteristics were derived from the FAIR principles (Force11, 2017) proposed by a group of experts to promote optimal use of research data and methods in what they call the 'eScience ecosystem': Data should be Findable; data should be Accessible; data should be Interoperable; and data should be Re-usable. Of the 33 data sources and excluding those sources already known to the researcher (Dryad and Figshare), only 5 were deemed to be open research data repositories: NeuroVault, SFARI Gene, Universal Protein Resource (UniProt), Biological General Repository for Interaction Datasets (BioGRID) and PeptideAtlas. The main reason for many of the autism-related data repositories being categorised as 'closed' is the registration and vetting process that is required before being able to access and use the research data.

Each of the seven open research data repositories – the six open research data repositories in the Al-jawahiri and Milne (2017) sample plus Figshare – were analysed to determine whether they provided any statistics on the use of the datasets stored in the online repository. Where such statistics were made available to the public, searches were done for datasets related to autism and vaccination, and the most granular data made available from these searches was analysed for evidence of who downloaded the data and for what purpose.

Harvesting DOIs for relevant products from multiple sources

To move beyond aggregated download statistics and, consequently, to determine who is using open research data, it was necessary to find relevant datasets that have been assigned DOIs.

Open research data repositories. Of the 5 open research data repositories identified in the Al-jawahiri and Milne (2017) sample, none provided unique identifiers that would allow the researcher to track mentions or citations to the data by third parties in the social media. At most, open repositories provided DOIs to the published article associated with the dataset.

The research data repositories that were both open and that assigned DOIs to datasets in all instances[14] were searched for autism-relevant datasets using keyword searches ('autism'; 'vaccination'; 'epidemiology'; 'ASD'). The selection of keywords was based on those used in previous studies related to the topic of vaccination and that relied on keywords to identify relevant content for analysis (Mitra, Counts, & Pennebaker, 2016; Moran, Lucas, Everhart, & Morgan, 2016). Keywords were also selected based on the results from tests of keyword searches across the different repositories. The searches returned the following results: Dryad 58 and Harvard Dataverse 105 results. From the searches conducted across Dryad and the Harvard Dataverse, duplicate results and low-relevance results were removed to produce a list of unique and relevant open datasets with DOIs: Dryad 56 and Harvard Dataverse 44.

Figshare was treated differently because of (1) the number of results returned for 'autism' (n=316) and (2) the fact that some datasets have Figshare DOIs while others have journal article DOIs even though the result is for a dataset. To reduce the number of results, a search for 'autism and vaccination' was done. However, this did not reduce the number of search results (n=320). The Figshare results for the 'autism and vaccination' search query was consequently sorted by 'popularity' (based on a calculation done by Figshare using view and download data) in order to prioritise those datasets more likely to have been accessed by non-scientists. The results were found to be poor in terms of relevance. Results were therefore screened manually from the most popular to the least popular results as presented in the Figshare list of results, until 30 relevant results were found in decreasing order of popularity. One complicating factor in the case of Figshare, was the fact that while some results contained Figshare DOIs, many results contained DOIs for PLOS ONE and PubMed journals. Only results with a Figshare DOI were retained on the basis that the journal article DOIs would be included for analysis in the data journals dataset.

DOIs from the three data repositories were cleaned to remove duplicate DOIs, and to improve the relevance of the DOIs. DOIs for datasets that were not directly related to vaccination were removed. DOIs for datasets related to animal vaccination studies were retained on the basis that non-scientists may extrapolate the impacts of vaccination on animals to humans (regardless of the scientific basis for doing so). DOIs for datasets related to vaccination studies that were not directly linked to autism were retained. The reasons for retaining DOIs for datasets of non-autism specific vaccination studies was two-fold: (1) to increase the sample size; (2) while autism is claimed by the anti-vaccination

[14] Dryad and the Harvard Dataverse assign DOIs to all their datasets; Figshare assigns DOIs to datasets made publicly available (the determination is made by researchers at the time of uploading their datasets to the Figshare website).

movement as a specific outcome of vaccinations, the effects of other vaccinations are also used by the community to campaign against vaccinations more generally.

A total of 42 DOIs were harvested from the three data repositories. See Table 23 in the Appendix for a complete list of DOIs.

Data journals. Data journals are academic publications published with the primary purpose of making research datasets available. Data journals allow the author(s) to focus on the data itself rather than producing an extensive analysis of the data which is typically published in traditional academic journals. According to the Australian National Data Service, data journals seek to promote scientific accreditation and re-use, improve transparency of scientific method and results, support good data management practices, and provide an accessible, permanent and resolvable route to the dataset. There are two types of data journals: those journals that only publish data papers ('pure') and those that publish any type of paper including data papers ('mixed') (Candela, Castelli, Manghi, & Tani, 2015). Akers (2014) found 73 journals that either exclusively publish data papers or that publish data papers as one article type. Candela et al. (2015) produced a list of 116 data journals published by 15 different publishers. From this list, six journals were selected based on the extent to which they publish exclusively data papers and their journal's subject relevance (see Table 3).

Table 3: List of data journals included for analysis

Journal name	URL	Short description of data journal taken from the journal's website
Dataset Papers in Science	https://www.hindawi.com/journals/dpis/	Dataset Papers in Science is a peer reviewed, open access journal that publishes dataset papers in a wide range of subjects in science and medicine. The journal is run by a relatively large, international Editorial Board of experts in all subject areas of the journal.
GigaScience	https://academic.oup.com/gigascience	*GigaScience* aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, we publish 'big-data' studies from the entire spectrum of life and biomedical sciences. To achieve our goals, the journal has a novel publication format: one that links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.
F1000Research	https://f1000research.com/	The first Open Science journal for life scientists
Journal of Open Health Data	http://openhealthdata.metajnl.com/	*Open Health Data* is a peer-reviewed journal publishing papers that describe openly-accessible datasets with high reuse potential. We work with trusted data repositories to ensure that associated datasets are professionally archived, preserved, and openly available.
Scientific Data	http://www.nature.com/sdata/	*Scientific Data* is a new open-access, online-only publication for descriptions of scientifically valuable datasets. *Scientific Data* exists to help you publish, discover and reuse research data.
PLOS ONE	http://www.plosone.org	*PLOS ONE* is an international, peer-reviewed, open-access, online publication. *PLOS ONE* welcomes reports on primary research from any scientific discipline.

Searches on the respective journals' website were conducted to identify relevant data articles. In the case of Dataset Papers in Science, Gigascience, Open Health Data and Scientific Data, a search for 'vaccine and autism' was conducted. These keywords were selected on the same basis as for searches conducted of the data repositories.

No filtering of results was required as all of these journals are 'pure' data journals. Dataset Papers in Science and Open Health Data returned no results; Gigascience returned one result and Scientific Data returned two results. F1000Research and PLOS ONE are 'mixed' journals and this necessitated a more refined search approach. In the case of F1000Research, the 'vaccine and autism' search was limited to 'data note' in order to exclude research articles from the search results. F1000Research returned no results. PLOS ONE does not allow users to differentiate between data articles and research articles. The search for 'vaccine and autism' was therefore limited to 'research articles' in the journal 'PLOS ONE'. This returned 109 results.

The results were parsed for relevance – article titles indicating that a study examined links between vaccines and autism were retained while article titles indicating that a study focused on attitudes towards autism, on the characteristics or treatment of autism, or on infectious disease more broadly, were excluded. Parsing was terminated after the first 60 results as it became evident that results were no longer relevant. This process produced a list of 16 research PLOS ONE article DOIs. Articles were then analysed for dataset-specific DOIs – in other words, DOIs assigned to tables and supplementary material[15] – to create a list of 88 DOIs. Six relevant articles published in PLOS ONE identified during DOI harvesting of open research data repositories were added to the PLOS ONE list. This increased the number of articles to 22, and the number of DOIs to 117.

In total, 120 unique dataset DOIs were harvested from the six data journals.

Bibliometric databases. The Thompson Reuters Web of Science database was used to identify open research data related to the issue of vaccination and autism. A search of the Web of Science Core Collection was done using the terms 'autism' or 'ASD' in the 'Topic' field. These keywords were selected on the same basis as for searches conducted of the data repositories.

The search returned 17 376 results. A search within results using the term 'vaccin*' reduced the number of results to 855. Results were filtered to include only 'open access' publications of the 'article' document type. This reduced the number of results to 42. Of the 42 results, 7 did not have DOIs and 6 were duplications of articles in the data journals list. The remaining 29 articles were

[15] PLOS ONE assigns DOIs to three types of 'data': tables, figures and supplementary material (which consist of tables, figures, text). In keeping with the FAIR requirement of open research data, figures were excluded as they are neither machine-readable (interoperable) nor do they provide data in sufficient detail to make the data re-usable.

checked for any dataset DOIs by referring to the journal web page of each of the articles. A total of 5 articles were found to have 24 dataset DOIs.

Table 4: Number of vaccines–autism datasets DOIs harvested by source

Source	No. of DOIs
Open research data repositories	42
Data journals	120
Bibliometric database	24
TOTAL	**186**

Using the sample of autism-vaccination-related dataset DOIs created from open research data repositories, data journals and from the Web of Science, a matching list of DOIs for the journal articles associated with those datasets was created:

1. From the list of 42 data DOIs sourced from the data repositories, 26 associated article DOIs remained after datasets for which there were no DOIs (11) and articles that were not available open access (5) were removed from the list.
2. The list of 120 dataset DOIs from data journals produced a list of 25 associated journal article DOIs.
3. From the 24 dataset DOIs harvested from the Web of Science, 5 journal article DOIs were identified.

A total of 56 open access journal article DOIs were therefore identified (see Table 23).

The number of journal article DOIs compared to dataset DOIs was significantly lower for both the data journals and Web of Science data DOIs because for every article published in a (data) journal, there are several dataset DOIs. One additional article DOI was added for which a dataset was available but that was not indexed by the three sources. This process resulted in a list of 57 article DOIs. The 57 article DOIs were then queried on 28 June 2017 against the Altmetric database to determine whether those articles were referenced by non-scientists. The determination was done by recording the Altmetric Attention Score[16] for each DOI as well as the number of mentions of the article on Twitter by what Altmetric defines as 'members of the public'.[17] Mentions of open access articles on Twitter by members of the public were interpreted as an indicator of access.

[16] The Altmetric Attention Score is an automatically calculated, weighted count of all of the attention a research output has received across 15 different online media. For a detailed breakdown of the weightings and how the score is calculated, see https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-

[17] Altmetric only provides detailed demographic data for Twitter: 'Altmetric categorizes users from Twitter based on their posting history and profile information. Where Twitter data are available for an article, counts for each user category and

A limitation of the process described above was that it offered no certainty that the potential users of the products of open science would include individuals or organisations from the anti-vaccination movement. It was therefore decided to include an additional round of DOI harvesting that would ensure the inclusion of the anti-vaccination movement in the group of users. In other words, while the first approach progressed from the certainty of a set of open research datasets on the relationship between vaccination and autism to an unknown number and type of data users, the second approach progressed from the certainty of a group of potential open research data users from the anti-vaccination movement to an unknown collection of open research datasets. This second approach is described below.

4.4.2 Bottom-up (snowball) sampling

To complement the purposive, top-down approach of harvesting DOIs for research datasets related to the link between vaccines and autism, a bottom-up approach was adopted that followed a snowball sampling approach.

The researcher joined an anti-vaccination group on Facebook. From within that group, an active individual with a Twitter account was identified. The posts of the Facebook group and the tweets of the anti-vaccination individual were checked intermittently for a period of one year (January to December 2016). Any post or tweet that appeared to link to a web page claiming a link between vaccination and autism, *and* that appeared to reference scientific research, was tagged and archived using the application Pocket[18] for later analysis.

Starting with the archived web pages, snowball sampling was then used to identify and archive additional relevant web pages that referenced scientific research. Snowball sampling was used not because the population of anti-vaccination web pages[19] was difficult to reach for sampling purposes, but because an efficient and cost-effective technique was needed to sample from a large online

geolocation data are included in the Demographics tab of article details page. […] To compile a table of Twitter demographics, we look at keywords in profile descriptions, the types of journals that users link to, and follower lists to assign each profile a category: Member of the public - somebody who doesn't link to scholarly literature and doesn't otherwise fit any of the categories below; Researcher - somebody who is familiar with the literature; Practitioner - a clinician, or researcher who is working in clinical science; Science communicator - somebody who links frequently to scientific articles from a variety of different journals / publishers.' See https://help.altmetric.com/support/solutions/articles/6000060978-how-are-twitter-demographics-determined-

[18] https://getpocket.com/

[19] A web page is defined as 'a computer file on the web, displayed on a monitor or mobile device, which could provide text, pictures, or other forms of data'. Web pages therefore include social media pages, pages containing video or audio files, PDF pages, as well as 'traditional' web pages that feature primarily text and images. Each such web page is characterised by the fact that it has a unique URL.

population of individual web pages[20] spread across many platforms and spheres on the internet. Also, sampling had to be selective because only pages with references to scientific research were of interest. Sampling was done between May and June 2017. Sampling was terminated once 224 references to scientific research were identified. The point of termination was determined by the fact there was an obvious and rapid increase in the occurrence of the number of references to the same scientific source from different web pages.

During the snowball sampling of anti-vaccination web pages referring to scientific sources, 167 references from an anti-vaccination web page to a scientific source of one type or another were identified. These references were analysed to determine whether they connect to open access journal articles. The open access journal article DOIs (or PubMed IDs) and the web pages from the snowball sample therefore provided a sample from which to determine whether non-scientists in general and the anti-vaccination movement in particular are connecting to a specific scientific product. Connections by members of the public were interpreted as an indicator of access.

The advantage of the snowball sampling of anti-vaccination web pages was that it ensured that the stance of the content vis-à-vis vaccination was known – the pages were authored by members of the anti-vaccination movement. However, while the stance was known, the type of scientific product being referenced by the authors of those pages was not known, and the sample had to consist of open access articles with identifiers to allow for tracing using the Altmetric database. It was therefore necessary to determine the type of scientific product to create a sample of open access journal articles with article identifiers.

Of the 167 web pages identified during the snowball sampling of anti-vaccination web pages for references to a scientific product of one type or another, 151 web pages provided either DOIs or PubMed IDs to journal articles or article abstracts. 76 articles with DOIs for which only an abstract was publicly available were removed from the list to limit the sample to full-text open access journal articles. In the 12 cases where only an abstract for an article is publicly available but an anti-vaccination website has published a full-text version of the article online, the article was included. Retracted articles were also included only if a full-text version of the article remained online post-retraction. The result was a final list of 75 journal article DOIs from the 168 web pages.

[20] Moran et al. (2016) identified 27,600 web pages using search terms such as 'baby vaccinations,' 'childhood vaccinations autism,' and 'MMR side effects'.

4.5 Methods for determining access by non-specific users

For both products of science, the investigation of use started by collecting and analysing data on the use of products related to the topic of vaccination-autism but that was not specific in terms of use by the anti-vaccination movement. This was done to establish that there is in fact interest in these products of science in the broader population before conducting analyses of use by the anti-vaccination movement.

In the case of open research data, download statistics were used as a weak proxy for use by non-specific publics. The three open research data repositories – Dryad, Dataverse and Figshare – were queried on 14, 15 and 17 February 2017 respectively for download statistics using the datasets identified during the sampling process and the search functionality provided by each repository. Downloads were relied on as an indicator of non-specific use because of the non-existent or low levels of attention for datasets as revealed by the Altmetric Attention scores for the vaccines–autism datasets in the sample.

In the case of open access journal articles, however, the Altmetric Attention scores were high enough to use the Altmetric Explorer to conduct a high-level analysis of use. 132 article DOIs (57 from the purposive sample and 75 from the snowball sample) were queried using the Altmetric Explorer to determine the Attention Score of each article as well as a high-level disaggregation of publics who have made mention of each article on Twitter as calculated by Altmetric Explorer according to four categories: public; scientists; practitioners; and science communicators.

4.6 Methods for determining access to and use of open research data and open access journal articles by the anti-vaccination movement

Access describes the possibility of gaining entry; in the context of information, that entry is comprised of the possibility of connecting to and retrieving information. When access is open, it is presumed that the possibility of connecting to and retrieving information is universal.

The act of connecting to a scientific product as an information object was used as an indicator of a product having been accessed. Such connections on the world wide web take the form of hyperlinks that link directly to a scientific product elsewhere on the web. In the case of the authoring and publication of content in the social media and on the web, hyperlinks are either embedded in a web page, in a tweet or in a post. Each is indicative of a connection having been made. A connection may also assume the form of a 'mention' in which a direct reference to another product appears (e.g. the title of the product, its host, etc.) without the inclusion of a hyperlink. In the case of reading content on the web, access can be determined via views of a web page, bookmarking of web page, and retrieval of content can be determined in the form of download statistics.

Given the interest in the connection between non-scientific publics and the products of open science, and the reliance on digital methods to observe those connections, the data made available by Altmetric was used. Specifically, the 'Altmetric Explorer'[21] provides data on the sharing and purported use and impact of scientific products with a unique digital object identifier (DOI) in the social media and other Web 2.0 platforms.

None of the indicators available from Altmetric provide reliable evidence of the actual use or impact of scientific products. Links, bookmarks, downloads and mentions do not provide evidence that the product linked to, bookmarked, downloaded or mentioned was in fact read, or used. Additional methods – content analysis to determine level of engagement as a proxy for use – were developed to measure the use of scientific products by the anti-vaccination movement in the social media and on the web.

In the two subsections that follow, the methods used to determine access to and use of open research data by the anti-vaccination movement is presented, followed by the methods used to determine access to and use of open access journal articles by the anti-vaccination movement.

4.6.1 Analysis of the open research dataset DOIs

The 186 DOIs for the open research datasets containing data related to the vaccines–autism debate from the purposive, top-down sample were cross-referenced against the structured, relational database of Altmetric cleaned and maintained by CWTS at the University of Leiden for the period up to June 2016. The CWTS Altmetric database was queried for references to Twitter. For the 186 dataset DOIs queried, 15 results in the form of Altmetric IDs were returned. These 15 Altmetric IDs were then analysed manually using the Altmetric Explorer to determine the tweeter, the type of Twitter account (organisation, individual or bot), number of tweeters per Altmetric ID, whether the tweeter was from the science community or not, the number of followers of the tweeter, the level of engagement (low or high),[22] whether hashtags were used, whether any Twitter handles were included in the tweet, links in the tweet, and a description of the tweet.

[21] https://www.altmetric.com/explorer/
[22] Haustein et al. (2015) suggest 'incorporating level of engagement as a criterion in the identification of automated tweets; namely, the degree to which the content of the tweet demonstrates manipulation beyond copying

In the case of the bottom-up snowball sample, each of the 224 references in the sample was analysed to collect the following data: name of the parent website; URL of parent website; web page URL; web page title; link to scientific source; type of scientific source (abstract, full-text article, review or letter published in an academic journal; scientific monograph; dataset from scientific research study; presentation to scientific audience [e.g. conference]; advisory from a government agency; other [e.g. patents]); and digital object identifier (DOI) for the source or any other identifier if a DOI was not available (e.g. PubMed identification number).

Following coding, the 224 references were double-checked manually for relevance and for the presence of an active link to a scientific source by following the link and reading the abstract or other available description of the scientific source. Following this process, 23 pages were found to be non-relevant (i.e. not specific enough in terms of presenting a link between autism and vaccination) or did not provide a working link to the scientific source. References to the same scientific source were removed by applying the remove duplicates function in Excel to the column in the data table that contained 'Other reference' (i.e. DOIs, PubMed IDs or any other reference). 33 duplicates were found and removed. The final sample of references or links found on web pages that presented a connection between autism and vaccination, amounted to 167. These results were then sorted by 'source type' in Microsoft Excel to determine whether any of the references included a reference to a research dataset.

4.6.2 Analysis of the open access journal article DOIs

Using a combination of the Altmetric Attention Score and the number of mentions by members of the public on Twitter, the top 10 open access journal articles were selected for closer analysis to discover how those articles are being used by non-scientists. In other words, articles that returned the top 10 scores from a combination of their Altmetric Attention Scores and tweets by members of the public were selected for further analysis. The 10 articles were drawn in equal proportion from the two batches, that is 5 articles were selected from the sample of articles identified using the bottom-up approach (research data repositories, data journals and the Web of Science); and 5 articles were selected from the sample created using the bottom-up approach (the web pages of anti-vaccination websites). See Table 11 for a list of selected articles as well as Table 24 in the Appendix for the full list articles.

bibliographic data'.

Table 5 shows the number of articles in each batch from which the top 5 articles were selected for further analysis. Associating the total of 131 open access journal articles with a corresponding open dataset ensured that both the article and its dataset relate to the same topic or area of interest. In other words, by controlling for content and attention, it can be assumed that both the article and its dataset hold the same face-value in terms of interest to the anti-vaccination movement.

Table 5: Number of article DOIs associated with datasets on vaccines and autism

Source	No. of DOIs (or PubMed ID)
Batch 1: Open research data repositories, data journals and bibliometric database	56
Batch 2: Snowball sample	75
TOTAL	**131**

Each online sphere – Twitter, Facebook and the web pages of anti-vaccination websites – was analysed independently and with some variation in the analysis was done owing to different affordances and levels of access to the data of each sphere.

In the case of Twitter, accounts were categorised according to their stance (that is, whether they are anti-vaccination accounts), and the level of activity and engagement of anti-vaccination accounts in relation to the sample of 10 open access journal articles was determined. In the case of Facebook, fewer articles were analysed because of restricted access to Facebook data, forcing the researcher to rely solely on manual identification of anti-vaccination Facebook accounts and subsequent analysis in terms of the activity and level of engagement of those accounts. In the case of the web pages of anti-vaccination websites, the stance of the authors of those web pages was already known and level of activity could not be quantified in a manner possible for the social media. Web pages were therefore only analysed for level of engagement.

In the subsections that follow, the methods of analysis for Twitter, Facebook and web pages are described in detail.

4.6.2.1 Stance, activity and level of engagement on Twitter

The following Twitter data were collected from the Altmetric Explorer results for the top 10 open access journal articles for preliminary analysis:

1. Total number of tweets that mention the article
2. Total number of Twitter users who tweeted a mention of the article
3. Number of Twitter mentions by the public

The above data were analysed to determine connections (mentions) to open access journal articles in the social media as well as connections to the articles by different publics.

This preliminary analysis was followed by more in-depth analysis of the Altmetric Twitter data to determine (1) the number of anti-vaccination accounts mentioning one of the 10 open access journal articles, (2) the number of tweets by the anti-vaccination movement, and (3) the level of engagement by the anti-vaccination movement with the articles.

In the case of the top-down sample, each tweet for the four articles with the fewest number of mentions was analysed manually by the researcher. This was possible due to the relatively low number of tweets for the bottom four articles in this sample. However, this approach was not possible for the top articles by Attention Score in the top-down sample and for all the articles in the bottom-up sample because of the sheer number of tweets (between 382 and 3539).

A programmer from the company SBC4D was engaged to develop an application that could crawl a batch of URLs and query each URL for a predetermined set of terms.[23] The application returned the number of times each term could be found for each of the URLs. The application made is possible to run a query of known anti-vaccination Twitter hashtags for the full sample of Twitter account URLs for each article.

The SBC4D application held some advantages over a similar known application, the Lippmannian Device.[24] The Lippmannian Device is a proven tool previously used in research that relied on digital methodologies (Rogers, 2015). The Lippmannian Device queries the Google search engine with specific keywords for specified URLs and makes the results available for further analysis in the form of a tag cloud and an html table. The most common use of the tool is researching the presence as well as the ranking of particular sources within Google engine results. By using keywords that are indicative of the stance of a particular website, the Device is used to determine the partisanship of websites. Because each Twitter account has a unique URL, it is therefore possible to determine the partisanship of an account. However, a major drawback of the Device is that is can only process batches of 100 URLs at a time, and queries are regularly interrupted because of Google's use of Captcha technology to block suspicious crawls. This requires the user to constantly monitor crawls and to complete Captcha authentications to restart the crawl. For a batch of 100 URLs and two query terms, the number of Captcha authentications can approach close to 15. The SBC4D

[23] See https://dev.sbc4d.com/cdv/fsv/geturl.php
[24] https://wiki.digitalmethods.net/Dmi/ToolGoogleScraper

application requires no authentication and there is no limit on the number of URLs that can be queried. The advantage of the Lippmannian Device is the 'depth' of its crawls. Test showed that the Device does an extensive crawl of a URL whereas the SBC4D application limits its crawl to all the content on the first visible page of a Twitter account.

To test the accuracy of the SBC4D application against that of the Lippmannian Device, the Twitter account URLs in batch 1.5 were used. The URLs of the Twitter accounts of all the tweeters in the Altmetric results for each of the open access journal articles were entered into the Lippmannian Device and their partisanship determined by entering hashtags commonly used by the anti-vaccination movement: '#antivax' and '#vaxxed'.[25] The selection of these terms was determined by applying the following criteria: (1) their identification in previous studies investigating the use of Twitter in the anti-vaccination debate (Dredze, Wood-Doughty, Quin, & Broniatowski, 2017; Mitra, Counts, & Pennebaker, 2016; Radzikowski, et al., 2016); tests of other terms using the Lippmannian Device; (3) additional terms noted by the researcher while creating the snowball sample of anti-vaccination web pages referring to scientific research;[26] and (4) the frequency with which the terms are used on Twitter based on tweet activity analysis done using the online analytics tool Symplur Signals.[27]

The number of results was set to 20 to speed up the processing of the URLs entered. The language was set to English as only tweets in English would be analysed. Twitter account URLs were processed in batches of 100. The result was a word cloud for each batch of 100 results indicating the frequency with which the commonly used words appeared in the Twitter accounts entered into the device. The tweets in the Altmetric results for those accounts found to use any of the search terms were then analysed manually as described above. This was necessary because the hash tags '#antivax' and '#vax' are not used exclusively by the anti-vaccination movement; pro-vaccination advocates also occasionally use the terms as hashtags to draw the attention of the anti-vaccination movement to relevant scientific publications.

The SBC4D application was then used to query the same Twitter account URLs in batch 1.5. A broader set of keywords was used because the SBCD application is able to run queries much faster than the Lippmannian Device and to compensate for the SBC4D application's relatively shallow

[25] The actual search terms used were 'vaxxed' and 'antivax' – without the hashtags. The Lippmannian Device would only return result if the hashtag symbol was omitted from the search terms.
[26] Milani (2016) also finds that despite the many tools available for identifying and analysing Twitter hashtags, some are still only discovered by chance in the research process. Mitra et al. (2016) point out that due to the transient nature of social media, it is not possible to rely solely on terms found to be in common use in the past.
[27] See https://signals.symplur.com

crawl: antivax, vaxxed, vaccineinjur, vaxfax, vaccinesafety, informedconsent, vactruth. Results were found to be satisfactorily comparable, especially if it is taken into consideration that the anti-vaccination stance of all Twitter accounts returning a score of more than 1 would be verified by manually inspecting the account. In other words, the difference between scores is less important than both methods returning a score of at least 1 each URL. Of the 311 URLs in the test batch, the Lippmannian Device identified 46 Twitter accounts as potentially anti-vaccination while the SBC4D application identified 37. However, after manual checking of the accounts, 12 accounts with a score of 1 or more were found to be false positive results; in other words, pro-vaccination accounts using anti-vaccination hashtags in their Tweets to attract the attention of the anti-vaccination Twitter community. If these false positive results are excluded, then the Lippmannian Device returns 36 results and the SBC4D application 34 positive results. Based on the outcome of this test, the SBC4D application was used to crawl all Twitter account URLs in each of the 5 batches in the bottom-up sample.

The SBC4D application only detects those Twitter accounts most likely to be anti-vaccination based on the appearance of keywords that appear on the first page of a given account. Accounts that do not use the prescribed keywords may nevertheless be anti-vaccination. Similarly, an account that may have used the keywords in the past, but those keywords no longer appear on the first pages of the account, may also not return a positive result. These limitations mean that the SBCD application's results are conservative estimates of the number of likely anti-vaccination Twitter accounts.

As indicated above, the possibility exists that accounts for which the SBC4D crawler returns positive results may in fact be false positives because some pro-vaccination Twitter accounts inject keywords (hashtags) commonly used by the anti-vaccination movement into information streams to lure them into an exchange (Conover, et al., 2011). To account for false positive results, all positive results returned by the SBC4D crawler were checked manually, and all accounts found to be pro-vaccination were recorded as such and removed from the sample.

The results of the SBC4D application crawls were saved as Microsoft Excel files for further analysis to determine (1) the number of unique anti-vaccination Twitter accounts in each batch, (2) the number of tweets by the anti-vaccination movement, and (3) the level of engagement with the scientific article by each of the anti-vaccination accounts.

To determine whether a Twitter account was anti-vaccination, each account was coded as either ‘anti’, ‘pro’, ‘neutral’ or ‘unknown’. An account was deemed to be anti-vaccination if any consistent anti-vaccination sentiment was expressed in the Twitter account description, in the banner image of the account or in the most recent tweets on the first page of the account, or, failing the availability of a description or of an informative description, based on the sentiment expressed in a linked website, blog post or online document. An account was coded as neutral only if an explicit statement was found indicating impartiality and there was evidence of posts representative of both sides of the vaccination debate. Accounts were coded as unknown when it was not possible to make a determination regarding stance. Appendix 2.1 provides in detail the steps followed to determine the unique number of anti-vaccination Twitter accounts in each batch.

To determine the proportion of total tweets as recorded in the Altmetric data attributable to anti-vaccination Twitter accounts, the anti-vaccination Twitter accounts were compared to the list of all accounts and tweets to identify all tweets belonging to the anti-vaccination accounts. Appendix 2.2 sets out in detail the steps followed to calculate the number of tweets by the anti-vaccination Twitter accounts.

For articles 1.1 to 1.5, Twitter mentions in general were found to be low, and for articles 2.1 and 2.2, Twitter mentions by the anti-vaccination movement specifically were relatively low. Levels of engagement analysis was therefore limited to those articles frequently mentioned by anti-vaccination accounts on Twitter, that is, articles 2.3 (812 anti-vaccination tweets); article 2.4 (672 anti-vaccination tweets) and article 2.5 (545 anti-vaccination tweets). For practical reasons, not all Tweets could be analysed for level of engagement. A simple random sample of 100 tweets was generated for each of the three articles. Each random sample was generated by creating a random number for each anti-vaccination tweet using Microsoft Excel. The list of random numbers and their associated tweets were then sorted from largest to smallest in order to provide a random list of tweets. The first 100 tweets were selected from the list for analysis. See Appendix 2.3 for a detailed account of the process followed to select accounts and determine level of engagement.

The determination of the level of engagement on Twitter by the anti-vaccination movement was done by reading each tweet in the Altmetric Explorer datasets for the three open access journal articles. Each Tweet was analysed using a 6-point scale of engagement. The scale was developed based on the suggestion by Haustein et al. (2016) that those actions on the web that result in online visibility and traceability be categorised along a continuum of access, appraisal and application. In an earlier study on the identification of Twitter audiences Haustein and Costas (2015) set out to measure the degree to which audiences engage with tweeted journal articles. They exclude retweets and use the dissimilarity between the content of the tweet and the title of the journal article as an indicator for engagement. They provide as reason for this approach the fact that only original content constitutes engagement and also that automated bots are frequent retweeters. The scale for level of engagement developed here departs from such an interpretation of retweets because although retweets are regarded as the lowest form of engagement, they nevertheless are assumed to play an important role in communication strategies that rely on retweets to (re)amplify the messages contained in the titles of journal articles.

Progression from access to application indicates increased levels of engagement by actors with digital objects such as web, pages, images, journal articles, datasets and the like. An article may generate many tweets and retweets that mention an article, but such activity may not be the result of the content of the article. For example, a retracted article may generate many mentions to the article in relation to its retraction, but such activity is not necessarily indicative of engagement with the content of the article. The scale for level of engagement attempts to measure increasing levels of engagement in relation to the content of each article rather than in relation to the degree of activity on Twitter.

The original scale was tested and refined using tweets for article 1.1 in order to produce the scale in Table 6.

Table 6: Scale for level of engagement with journal articles in anti-vaccination tweets

ENGAGEMENT →					
ACCESS		**APPRAISAL**		**APPLICATION**	
1 LOW	**2 LOW**	**3 MEDIUM**	**4 MEDIUM**	**5 HIGH**	**6 HIGH**
Retweet OR Tweet that is copied from an earlier tweet OR broadcasting existing tweet to other accounts	Tweet article title OR tweet link and hashtags OR reply to existing tweet with link and hashtags	Tweet direct quotation from article abstract or summary	Tweet a description of the article findings in own words OR a direct quotation from the body of the article	Tweet consists of an interpretative statement or graphic pertaining to the article content	Tweet consists of an interpretative statement followed by a discussion thread consisting of at least a reply from another user and a response from the author of the tweet in which content from the article is used to substantiate the author's position

Tweets that no longer existed or to which access was restricted were included in the sample of 100 tweets but could not be analysed for obvious reasons.

4.6.2.2 Stance, activity and level of engagement on Facebook

Research that relies on data from Facebook confronts several challenges, prime among them the restrictions on access to the data (Rieder, 2013). For this reason, it was not possible to use the SBCD data crawler to determine the stance of Facebook accounts vis-à-vis vaccination. Analysis of Altmetric Explorer Twitter data revealed that there were three articles in the sample that received the most attention by the anti-vaccination movement: articles 2.3, 2.4 and 2.5.

The decision was therefore made to undertake a manual assessment of Facebook accounts for a single article to determine the stance of Facebook accounts and, subsequently, the level of engagement on the part of the anti-vaccination movement with that article.

Article 2.3 was selected for analysis because it was one of three articles that received the highest number of mentions by the anti-vaccination movement on Twitter, and because of the three articles, it had the highest Altmetric Attention Score.

Manual determination of stance involved coding each Facebook account according to one of the following: anti-vaccination, pro-vaccination, neutral, unknown. An account was deemed to be anti-vaccination if any consistent anti-vaccination sentiment was expressed in the 'About' section of the account and in the most recent posts on the first page of the account, or, failing the availability of any About text or of an informative description in the About text, based on the sentiment expressed in a linked website, blog post or online document. An account was coded as neutral only if an explicit statement was found indicating impartiality and there was evidence of posts representative of both sides of the vaccination debate. Accounts were coded as unknown when it was not possible to make a determination regarding stance. There were several reasons for not being able to determine the vaccination stance of a Facebook account, including a URL that no longer resolves to the original post, an insufficient number of posts to make a determination with any degree of confidence and posts in a language other than English with no or limited machine translations available. See Appendix 2.4 for a more detailed account of the process followed to determine stance.

The total number of mentions to article 2.3 in the Altmetric data was 444. After 198 duplicate accounts were removed from the Altmetric dataset, 246 unique Facebook accounts remained. The account name description ('Outlet or Author' field) and the title of the post ('Mention Title' field) of these 246 accounts were checked for any strong indications of stance. Once the fields in Altmetric dataset had been exhausted, each Facebook mention was checked manually online to verify and/or determine stance.

There were some cases where the Altmetric dataset included a Facebook account as the source of a post mentioning article 2.3 when in fact the account in question was tagged by another account. For example, Altmetric lists the account 'Things anti-vaxers say' as the source of a post mentioning article 2.3 when in fact it is the account 'Edgar Diaz' that tags 'Things anti-vaxers say' to draw that account's attention to the post. In such cases, rather than making a determination of the stance of the tagged account listed in the Altmetric dataset, the account doing the tagging was recorded and it was analysed to determine its stance.

To determine the number of Facebook posts as recorded in the Altmetric data attributable to anti-vaccination Facebook, the anti-vaccination Facebook accounts were compared to the list of all Facebook accounts and posts to identify all posts belonging to the anti-vaccination accounts. Appendix 2.5 sets out in detail the steps followed to calculate the number of posts by the anti-vaccination Facebook accounts.

To determine the anti-vaccination movement's level of engagement with article 2.3 on Facebook, a 6-point scale of engagement as shown in Table 7 was developed.

Table 7: Scale for level of engagement with journal articles in anti-vaccination Facebook posts

ENGAGEMENT →					
ACCESS		APPRAISAL		APPLICATION	
1 LOW	2 LOW	3 MEDIUM	4 MEDIUM	5 HIGH	6 HIGH
Repost OR shared post OR a post that is copied from a previous post	Post article title OR post article link	Post is a direct quotation from article abstract or summary	Post is a description of the article findings in own words OR a direct quotation from the body of the article	Post consists of an interpretative statement or graphic pertaining to the article content	Post consists of an interpretative statement followed by a discussion thread consisting of at least a reply from another user and a response from the author of the post in which content from the article is used to substantiate the author's position

The scale is a modification of the scale developed for Twitter with adjustments to take into consideration the differences between the two social media platforms in terms of how contents is published and shared. The purpose of using the Twitter scale as a foundation was to ensure a certain level of similarity and comparability between the two scales.

Only posts in English were analysed – of the 182 unique anti-vaccination accounts identified, 164 accounts posted in English. See Appendix 2.6 for further detail on the process followed to determine the level of engagement with article 2.3 by the anti-vaccination movement on Facebook.

4.6.2.3 Level of engagement on the web

During the snowball sampling of anti-vaccination web pages referring to scientific sources, 167 web pages were identified that made reference to a scientific source of one type or another. Of these, 70 pages included DOIs or PubMed IDs but only 34 web pages provided either DOIs or PubMed IDs to *full-text* open access journal articles. In the 12 cases where only an abstract for an article is publicly available but an anti-vaccination website has published a full-text version of the article online, the article was excluded. Retracted articles were included but only if a full-text version remained online post-retraction. Web pages that provided lists or indexes of articles, and provided no supporting text to contextualise or describe the articles were also excluded. For example, the 'Research and Resources'[28] page of the Children's Medical Safety Research Institute (CSMRI) provides functionality that allows visitors to the page to explore CSMRI-funded and other research using a keyword search of its articles database that returns links to the articles on their parent websites, but provides no information about each of the articles. The CSMRI pages was excluded. The page '30 Solid Scientific Studies that Prove that Vaccines Cause Autism'[29] also provides a list of articles but prefaces the list with several paragraphs of copy to provide context for the articles that follow. This page was included in the sample.

The determination of level of engagement by members of the anti-vaccination movement with open access journal articles via web pages was done by developing a 6-point scale of engagement that corresponds as closely as possible to the scales used to analyse engagement on Twitter and Facebook, while taking into account the differences in how content is constructed and shared on social media and web pages. As with engagement on Twitter, the scale was developed based on Haustein et al.'s (2016) suggestion that engagement on the web be categorised along a continuum of access, appraisal and application, and that progression from access to application indicates increased levels of engagement by actors with digital objects such as web, pages, images, journal articles, datasets and the like. The scale was tested and refined using three randomly selected web pages. The final 6-point scale used is presented in Table 8.

[28] https://www.cmsri.org/research-resources/
[29] http://yournewswire.com/30-solid-scientific-studies-that-prove-vaccines-cause-autism/

Table 8: Scale for level of engagement with journal articles on anti-vaccinations web pages

ENGAGEMENT →					
ACCESS		**APPRAISAL**		**APPLICATION**	
1 LOW	**2 LOW**	**3 MEDIUM**	**4 MEDIUM**	**5 HIGH**	**6 HIGH**
Republication (repost) of a previously published article or blog	Includes only the title or a direct quotation from the article abstract as a reference to the article	Includes only a direct quotation from the article abstract, plus comment(s) by the author	Includes a description of the article in own words AND/OR a direct extract from the body of the article	Includes an interpretative statement/ narrative, table or graphic pertaining to the article	Includes an interpretative statement/narrative, table or graphic pertaining to the article followed (1) by a discussion thread consisting of at least a reply from another user and a response from the author of the web page in which content from the article is used to substantiate the author's position OR (2) references to and reasoned counter-arguments to pro-vaccination articles

It is important to note the scale for level of engagement on the web does not in any way attempt to measure or assess the validity of arguments presented by the anti-vaccination movement with reference to open access scientific journal articles; the scale only seeks to measure the level of engagement with the content of those scientific articles in the construction of arguments.

Only web pages written in English were analysed for level of engagement.

4.7 Method for identifying intermediaries in the online communications of the anti-vaccination movement

To determine who the intermediaries are in the anti-vaccination movement's online social communication networks, basic network analysis was done. Networks were limited to those on the social media platform Twitter because its data were easily accessible as input data for the creation of network graphs.

The software package NodeXL Pro[30] version 1.0.1.396 (Hansen, Shneiderman, & Smith, 2010), was used for all network-metric calculations and for generating network graphs. NodeXL Pro input data was either extracted from the Altmetric database and imported into NodeXL Pro, or NodeXL Pro was used to import data directly from Twitter via its API.

[30] https://www.smrfoundation.org/nodexl/

In the first instance, two ego networks were generated using NodeXL Pro. These networks relied on Twitter mentions to journal articles from the snowball sample (article 2.2 and article 2.3) as recorded in the Altmetric database.

To explore the relationships between Twitter accounts, their possible clustering into distinctive sub-groups, and the presence of intermediary Twitter accounts, a known active tweeter from each of the anti-vaccination and pro-science communities (@LotusOak and @doritmi respectively) was used to generate a sociogram using NodeXL Pro. In this case, data were imported directly via the Twitter API.

Neither of these two methods produce network graphs that represent the relationships between users of social media (agents) whose tweets (event) link to an open access journal article on the topic of vaccination and autism (affordance) as illustrated in Figure 4. One solution to this problem would be to create a relational network graph by using NodeXL Pro to search Twitter for all mentions to a journal article, and to create a network based on the connections between those Twitter accounts that mention the article. However, NodeXL Pro is only able to access and query the most recent 7 days of Twitter data. This places limits on the usefulness of the software to create relational networks using the DOI or URL of a journal article as a unique identifier because articles are mentioned over a period much longer than 7 days, and mentions can, at times, be sporadic. Twitter accounts and, to a lesser degree, hashtags, are more 'stable' identifiers. But the problem with hash tags is that they are an unreliable indicator of a mention to a scientific product on Twitter.

Using the CWTS Altmetric database it was possible to create a relational network based on both Twitter accounts and mentions made by those accounts to open access journal articles in the form of hyperlinks. To create such a network, multiple products of science (research objects) are required to create instances where two Twitter accounts (agents) tweet (event) a link (affordance) to several (more than two) open access journal articles (research object) (see Figure 4). Each pair of Twitter accounts linking to a journal article in this manner is a couple, hence the description 'tweeter coupling network'.

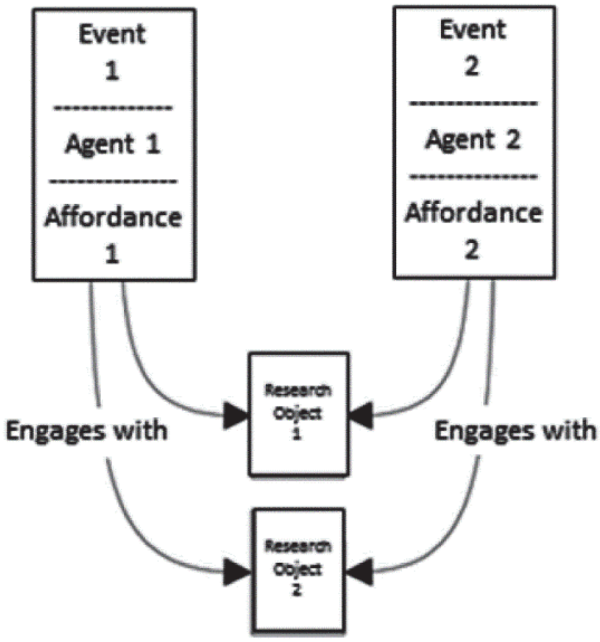
Using 113 of the 131 DOIs identified for open access journal articles on the topic of vaccination and autism (see Table 5), all Twitter accounts that mentioned any of the 113 articles were identified using the Altmetric database.[31] This produced a total of 12 207 unique Twitter accounts and 21 490 tweets. Following Costas et al. (2017), a matrix of coupled accounts (or Twitter handles) in which

[31] The reason for the use of only 113 articles as opposed to 131 articles was the exclusion of articles that had an Altmetric Attention score of zero, and the inclusion of only those articles for which a DOI was available (i.e. in some instances, only a PubMed ID was available, and those articles were excluded).

each pair mentions the same article more than twice was created. This data consisting of account (or handle) pairs was then imported into NodeXL Pro as network edges making possible the plotting of a network graph showing the connections between those Twitter accounts that mentioned more than two open access journal articles from the sample of 113 articles.

Using a combination of (1) the visual grouping of the network using the Harel-Koren Fast Multiscale algorithm, which is one of NodeXL Pro's two lay-out algorithms; (2) graph metrics calculated by NodeXL Pro (for example, centrality and betweenness); and (3) the manually-annotated data generated previously when determining the stance of Twitter accounts mentioning open access journal articles related to vaccination and autism, further analysis of a variety of network graphs was done to identify centrally located Twitter accounts in the network.

Figure 4: Social media coupling



Source: Costas et al. (2017)

4.8 Limitations and additional considerations in relation to research design

There is a fair degree of skepticism regarding the value of the web as an adequate site for the study of cultural and social issues. This researcher found the web as a source of data to be useful in developing an understanding of communication in the online social media as it relates to the communication of science. The web as the source of data was also found to circumvent many of the challenges associated with research that relies on interviews or surveys as sources of empirical data.

These include the time required to collect data from human subjects, low response rates, inability to conduct post-collection verification and the cost associated with more traditional methods of data collection. Nevertheless, using the web as a site of study does pose some challenges.

It is not uncommon for the owners of data sources on the web (e.g. Facebook and Twitter) to change the rules on how their data can be accessed and re-used. This may affect not only the researchers' ability to access data directly but may also compromise the functionality and usefulness of applications used to collect and analyse online data. When the web is the data source, there are also challenges with the stability of the data – websites disappear from the web for a variety of reasons, users of social media close their accounts or change their account profiles (affecting either the content of those accounts or the accessibility of the account's content).

All of these fluid conditions on the web place limits on the replicability of this study and on the ability to verify some of the data referenced in the research. They also confirmed the challenges of relying on the web as a data source as the conditions of data access changed, tools became redundant and, in some cases, data previously used in analysis disappeared during the data collection and analysis processes.

A variety of tools have been developed to compile metrics using the social media as a source of data, including Altmetric which was used in this study. Reviews of the strengths and weaknesses of these tools (Sugimoto, Work, Larivière, & Haustein, 2017) suggest that one of their weaknesses is the reliance of Altmetric on documents that have a unique object identifier. This excludes certain document types and results in a bias towards journal articles (Sugimoto, Work, Larivière, & Haustein, 2017). That documents from the grey literature are excluded in this study is unproblematic because of its focus on two specific products of science in the formal communication of science, one of which is the journal article. However, it is acknowledged that the dependency on digital object identifiers may have excluded some relevant open datasets and possibly some open access journal articles from the results.

A second issue raised by Sugimoto et al. (2017) in their review is the observed discrepancy between social media counts when comparing results obtained from different aggregators. Based on these observed discrepancies, it is acknowledged that a limitation of this study is its dependency on Altmetric data and the possibility that different results may be obtained when using any of the other available altmetric analysis tools.

The emphasis on online communities and cultures in the research should not be read as an indication of the neat separation of the online from offline world of experience. It is acknowledged that the offline and online worlds are interconnected. And not only in the sense that values, meaning and identity are imported from offline realities into virtual realities, but that virtual realities are increasingly shaping offline values, meaning and identity. This acknowledgement was a contributor to the selection of Castell's 'theory' of the network society as framework for understanding the use of scientific productions by non-scientific social movements. Nevertheless, it is also acknowledged that this thesis did not incorporate into its design an examination of the offline worlds of the anti-vaccination movement as it may relate to their use of open access journal articles and open research data. The research therefore presents only a partial if important and understudied component of communication in social networks.

The researcher is not an expert in medical research and was not in a position to judge the merits of the methods, findings and arguments put forward by scientists and non-scientists to prove or refute the association between vaccination and autism. This did not, however, place any limits on the analysis of the connections to and the use of the products of open science by non-scientists as collecting evidence for such connections and use did not rely on (and could possibly even have been hampered by) the researcher's understanding or stance on the vaccines–autism issue.

Ethical guidelines related to the use of online data for the purpose of research are inconsistent and evolving (Kozinets, 2010; Sugiura, Wiles, & Pope, 2017). This has led some scholars to argue for approaches that rely on 'goodwill' (Kelley, 2016) or 'contextual integrity' (Nissenbaum, 2011) rather than on strict guidelines. According to Kozinets (2010, p. 142), 'analysing online community or culture communications or their archives is not human subjects research if the researcher does not record the identity of the communicators and if the researcher can legally and easily gain access to these communications or archives.' Access to and use of data on social media platforms such as Twitter and Facebook is typically legalised by the user (Rogers, 2015), although concerns remain as to the extent which users will ever fully appreciate the contractual terms of use (Nissenbaum, 2011). The research presented in this thesis, conducted using online data under conditions consistent with data in the public domain made available under terms consented to by the individuals or legal entities, is exempt from restrictions pertaining to private data, including obtaining the consent of 'participants'. Moreover, the data used in this study consisted only of content made available by the users of the social media platforms and by the authors of web pages, and did not require any interaction between the researcher and members of the community under study. The names and details of selected social media accounts and web pages were included in this study because data and other information from those sources were not shared in confidence and could easily and legally be accessed by anyone with the inclination to do so. The inclusion of the names of selected social media accounts and web pages in this thesis makes it possible for future research to verify and/or expand on the findings presented without placing those named at risk.

Chapter 5. Connections to and use of open research data by the anti-vaccination movement

To determine whether non-scientists are accessing and using open research data, the anti-vaccination movement was selected as a case study. The online communications of the anti-vaccination movement were analysed to find evidence of access and use.

This chapter presents findings of this analysis in two parts. The first part presents findings from analysis on whether there is any evidence that open research datasets related to the topic of vaccines and autism are being accessed. While the results presented in this first part are non-specific in terms of *who* is accessing open research data, the results do provide some evidence of access among a broader population of potential users of open research data.

The second part of the chapter presents more focused findings related to whether a specific group of non-scientists – the anti-vaccination movement – is accessing and using open research data on the topic of vaccination and autism as evidenced by mentions to relevant open research data in the movement's online communications. The second part of the chapter also includes some observations on access to and use of open research data by the scientific community.

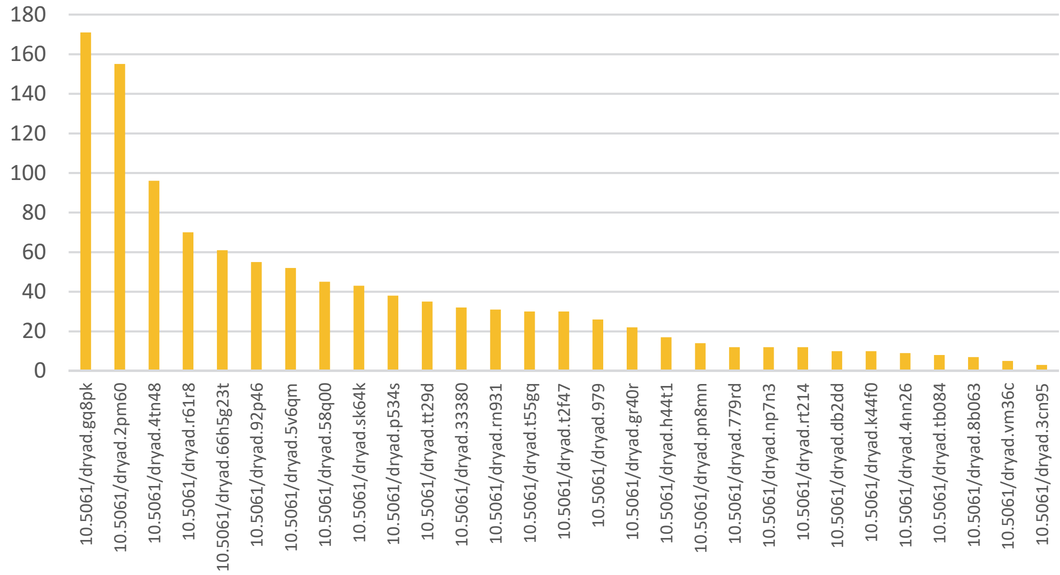
5.1 Connections to open research data on the topic of vaccination and autism by the broader population

Three online open research data repositories – Dryad, the Harvard Dataverse and Figshare – were referenced to determine whether open research data on the topic of vaccination and autism are being accessed by the broader population. Data on number of downloads of datasets identified in the sampling process was used as an indicator of access.

Dryad

Dryad does report on downloads for each data item but 'does not track, and cannot report, the identity of individuals who view or download files' (Dryad, 2018). Dryad encourages authors to use third-party tools such as Altmetric and the Data Citation Index for usage metrics, made possible by the assignment of DOIs to datasets by Dryad. Figure 5 shows the number of times 29 vaccine-related datasets were downloaded from Dryad. The data shows that the number of downloads ranges from as low as 3 to a peak of 171 downloads. On average, each dataset was downloaded 38 times.

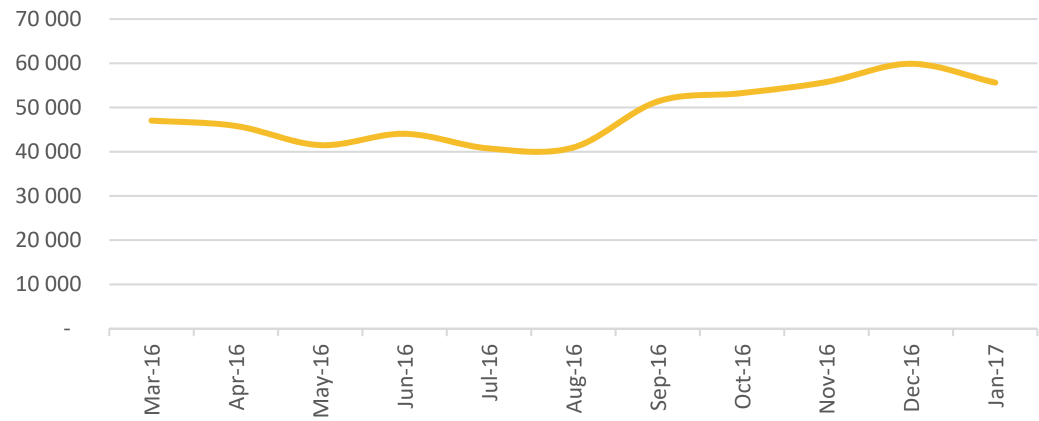
Figure 5: Number of times vaccine-related datasets downloaded from Dryad, as at 14 February 2017 (n=29)



Harvard Dataverse

Dataverse provides, as a weak proxy for use, the number of data files downloaded from the distributed dataverses. In February 2017, it reported 2 186 210 downloads. Figure 6 shows the number of downloads across all datasets for an 11-month period.

Figure 6: Number of datasets downloaded from the Dataverse (March 2016 to January 2017), as at 15 February 2017



Source: http://dataverse.org/metrics

Only 7 vaccine–autism-related datasets were found on the Dataverse. Data for number of times datasets were downloaded and corresponding dates of publication are shown in Table 9. As was the case for datasets downloaded from Dryad, there is a large discrepancy between the most (n=5,041) and the least (n=4) downloaded dataset.

Table 9: Number of times vaccine-related datasets were downloaded from the Dataverse, as at 14 February 2017 (n=7)

Dataset DOI	Number of downloads	Publication date
hdl:1902.1/15088	5041	23.09.2010
doi:10.7910/DVN/FF4DI7	573	15.03.2016
doi:10.7910/DVN/P04NKK	20	16.05.2016
doi:10.7910/DVN/XNUYPB	10	09.01.2017
doi:10.7910/DVN/C85WNZ	6	21.05.2015
hdl:1902.1/CP2009P2009-03	6	29.11.2011
doi:10.7910/DVN/26370	4	01.07.2014

Figshare

Only 6 vaccine–autism-related datasets with unique Figshare DOIs were found on Figshare. Data for number of times datasets were downloaded and corresponding dates of publication are shown in Table 10. Again, the data reveals evidence of the availability of research open datasets on the topic of vaccination and autism and of open research datasets being downloaded, although not as often as from the other two open research data repositories.

Table 10: Number of times vaccine-related datasets downloaded from Figshare, as at 17 February 2017 (n=6)

Dataset DOI	Number of downloads	Publication date
doi.org/10.6084/m9.figshare.1271925.v2	50	16-Dec-14
doi.org/10.6084/m9.figshare.3085885.v1	18	21-Mar-16
doi.org/10.6084/m9.figshare.1418309.v1	16	19-May-15
doi.org/10.6084/m9.figshare.1418313.v1	11	19-May-15
doi.org/10.6084/m9.figshare.1092545.v2	9	29-Aug-14
doi.org/10.6084/m9.figshare.2062575.v1	7	12-Jan-16

Autism-specific open data repositories

NeuroVault, SFARI Gene, Universal Protein Resource (UniProt), Biological General Repository for Interaction Datasets (BioGRID) and PeptideAtlas are autism-specific research data repositories. None of the repositories were found to provide publicly-accessible download statistics.

Conclusion

The number of file downloads in the range 40,000 to 60,000 downloads per month from the Dataverse for the period March 2016 to January 2017 (Figure 6) shows that open research datasets are being downloaded. The findings for the downloads of specific vaccine–autism-related datasets from Dryad, the Dataverse and Figshare show that datasets of relevance to the vaccine–autism debate are indeed available from open research data repositories, and that these datasets are being downloaded by the broader population. The number of downloads varies considerably but this is to be expected given (1) that there is an observable relationship between when the dataset was published and the number of downloads, and (2) that not all datasets will be of equal relevance or interest to users.

The download data available from the three open research data repositories do not reveal anything about who is downloading the data nor how the data are being used, if at all.

5.2 Connections to open research data by the anti-vaccination movement

To gain more fine-grained insights into the data downloads, research datasets that have DOIs assigned to them were analysed using the Altmetric Explorer. The use of the Altmetric Explorer enables the identification of content (tweets, posts, pages) and their associated accounts in the social media and on the web that make mention of scientific content (articles, data and the like) where such content has been assigned a DOI.

For the 186 relevant datasets with DOIs harvested from open research data repositories, data journals and other sources, only 15 (8%) were mentioned on Twitter. Most of the tweeters are organisations and almost all are scientists or scientific organisations. Most of the datasets (12, 80%) received only a single mention on Twitter in the form of what appears to be an automated post by the repository to which the dataset has been uploaded. Mostly, the connections to datasets are infrequent, and only from and for the science community. There is no evidence of the use of open research data by non-scientists, in this case, those from the anti-vaccination movement.

From the second approach of following the anti-vaccination movement on social media and snowball sampling from those social media accounts and web pages making reference to scientific sources in support of a claimed connection between autism and vaccination, use of open research data was also found to be non-existent. Of the web pages in the sample, none made reference to open research data.

Two instances were, however, found of anti-vaccination web pages making reference to statistics, in both cases published by US government agencies: the US Department of Health and Human Services' National Vaccine Injury Compensation Program Monthly Statistics Report[32] and The Centers for Disease Control and Prevention's (CDC) Autism Spectrum Disorder (ASD) Data and Statistics.[33]

The US Department of Health and Human Services' National Vaccine Injury Compensation Program Monthly Statistics Report consists of government rather than scientific data on the number of petitions filed for compensation, and the number and value of compensations paid. The title of the web page making reference to the statistics confirms the source of the data as being from government: 'Vaccines DO HARM – New Government Data Shows Real Extend [sic] of Vaccine Injury'.[34] The page reads as follows:

> The government's own document […] reveals statistics of over 18,000 petitions filed for compensation by families of children who were seriously injured, maimed or killed by vaccines in the United States alone. […] The document reveals that a shocking 633 petitions for serious vaccine injury or death have been filed in fiscal year 2017. This follows 1,120 petitions that were filed in 2016.[35]

As a side-note, the government document referred to was removed from the web by the US Department of Health and Human Services. However, a copy of the document was posted on an anti-vaccination website,[36] pre-emptively as it turned out, 'in case the government attempts to "memory hole" the report'.[37] The act of keeping a statistical government document online and in circulation after it has been retracted by government could be interpreted as an attempt by the anti-vaccination movement (1) to draw attention to the claimed control and manipulation of data by government agencies and (2) to keep as a reference a source of data because of the weight it lends to the position taken by the anti-vaccination movement.

Unlike the vaccine injury compensation data, the Centers for Disease Control and Prevention's statistics can be described as scientific in that they are compiled from data collected by the 'Autism and Developmental Disabilities Monitoring (ADDM) Network […] surveillance system that provides

---

32 http://www.naturalnews.com/files/monthlywebsitestats04_01_17.pdf
33 https://www.cdc.gov/ncbddd/autism/data.html
34 http://www.truthlibrary.info/articles/health/vaccines-do-harm-new-government-data-shows-real-extend-of-vaccine-injury/
35 Ibid.
36 http://www.naturalnews.com/files/monthlywebsitestats04_01_17.pdf
37 http://www.truthlibrary.info/articles/health/vaccines-do-harm-new-government-data-shows-real-extend-of-vaccine-injury/

estimates of the prevalence and characteristics of ASD among children aged 8 years whose parents or guardians reside in 11 ADDM Network sites in the United States' (Christensen, Baio, Braun, & et al., 2016). The methods, findings and selected data are published in a peer reviewed report 'Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012'. However, while the anti-vaccination web page cites the summary statistics taken from the Centers for Disease Control and Prevention's website (see Figure 7), it does not cite or use the underlying data available from the report; it limits its use to the high-level statistics to support its argument:

> According to statistics from the Centers for Disease Control and Prevention (CDC), one out of 68 children suffered from autism in the year 2012. In the year 2000, one in 150 children suffered from autism.[38]

Figure 7: Table of statistics on the prevalence of autism cited by an anti-vaccination website

Identified Prevalence of Autism Spectrum Disorder
ADDM Network 2000 – 2012
Combing Data from All Sites

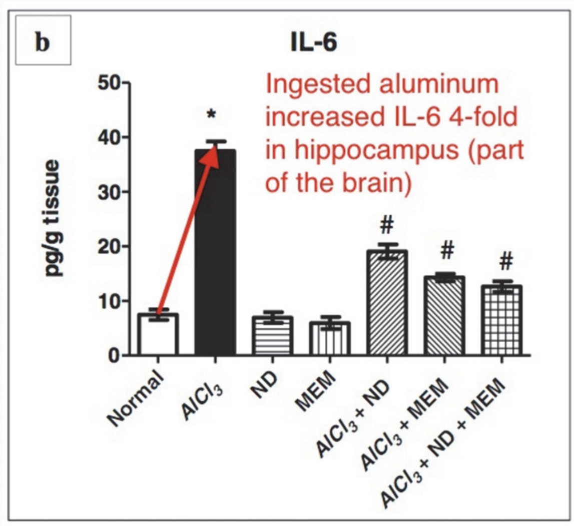
Surveillance Year	Birth Year	Number of ADDM Sites Reporting	Prevalence per 1,000 Children (Range)	This is about 1 in X children...
2000	1992	6	6.7 (4.5 – 9.9)	1 in 150
2002	1994	14	6.6 (3.3 – 10.6)	1 in 150
2004	1996	8	8.0 (4.6 – 9.8)	1 in 125
2006	1998	11	9.0 (4.2 – 12.1)	1 in 110
2008	2000	14	11.3 (4.8 – 21.2)	1 in 88
2010	2002	11	14.7 (5.7 – 21.9)	1 in 68
2012	2004	11	14.6 (8.2 – 24.6)	1 in 68

The second example is the author of an anti-vaccination web page's[39] extensive use of graphs and a table from several scientific publications to support his argument that the aluminium adjuvant used in vaccines causes brain injury and autism. The author of the web page uses no fewer than nine figures and one data table from nine different scientific sources to support his argument. Each item is annotated and explained with reference to the connection between aluminium adjuvant and autism (see Figure 8). There is, however, no evidence that the author engaged with the underlying

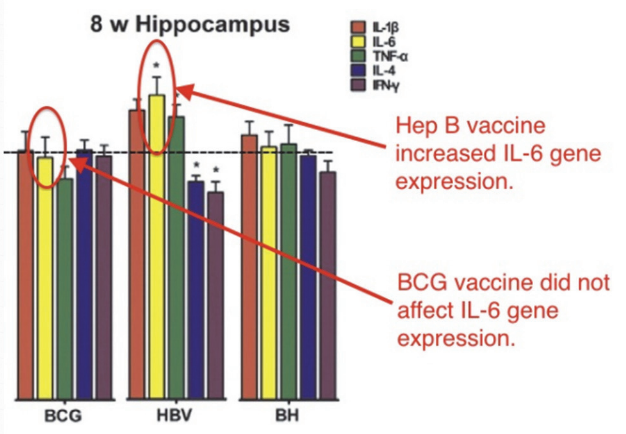
[38] https://vactruth.com/2017/02/23/autism-epidemic/
[39] http://vaccinepapers.org/aluminum-inflammation-interleukin-6/

data from which the figures and tables were produced. For eight of the nine sources, no supplementary data was published by the authors of the journal articles. However, for one of the articles, supplementary data was published, albeit not as open data.[40] Given that the author had obtained access to the pay-walled journal articles from which the graphs were extracted, it is reasonable to assume that he could also have accessed the supplementary data.

Figure 8: Examples of graphs used on an anti-vaccination website (annotations in red and captions to graphs added by the author of the anti-vaccination web page)



Above: Ingestion of 3.4 mg/kg/day soluble aluminum (from 17mg/kg/day AlCl3) caused a 4-fold increase in IL-6 in the hippocampus, a part of the brain. This is a vaccine-relevant dosage. The aluminum elevated other signs of brain inflammation, and caused adverse effects on learning, memory, and behavior. "ND" and "MEM" refer to nanodiamond and memantine, which reduced some of the adverse effects of aluminum. From Alawdi et al 2016.



Above: Hepatitis B vaccine caused an increase in IL-6 gene expression in rat brain (in the hippocampus) at age 8 weeks. BCG vaccine had no effect on IL-6 gene expression. Hep B vaccine contains Al adjuvant, and BCG vaccine does not. Elevated IL-6 causes autism. From L et al. 2015.

5.2.1 Data sharing within the scientific community

Al-jawahiri and Milne (2017) identify and describe 33 data resources available to the scientific community for autism research. According to the authors, these resources comprise either open or big datasets. Following an analysis of the data sources using the FAIR principles (Force11, 2017), it was found that of the 33 datasets, only 7 (21%) meet the FAIR criteria for open data. Almost half of the data repositories (16, 48%) failed to meet the criteria for openness because they require prospective data users to register and be vetted before being able to access the autism research data. This finding is consistent with that of Capocasa et al.'s (2016) research on data accessibility that found that only 3 in 46 research biobanks followed a fully open data policy.

[40] See http://vaccinepapers.org/wp-content/uploads/BCGhepB-vaccines.pdf and http://www.jni-journal.com/article/S0165-5728(15)30029-1/addons

5.3 Conclusion

The findings show that a politically active non-scientific community, the global anti-vaccination movement, is not accessing open research data on the purported link between vaccination and autism. At most, there is evidence of the use of non-textual scientific information in the form of numerical tables and graphs, neither of which are indicative of the use of raw scientific data by the anti-vaccination movement.

Evidence in the form of the number of online data repositories containing research data relevant to study of vaccinations and autism – 33 according to Al-jawahiri and Milne (2017) and 46 according to Capocasa et al. (2016) – and the low proportion of these data accessible by the public – 21% according to an analysis of the Al-jawahiri and Milne (2017) data and 7% according to Capocasa et al. (2016) – suggests that the scientific community is sharing research data within the self-regulated boundaries of the scientific community and that scientists appear to be cautious about making the data publicly available.

Possible explanations for restrictions imposed by the scientific community on access to open research data by a non-scientific community are presented in Chapter 8; as are possible explanations for the anti-vaccination movement's use of information as opposed to data.

Chapter 6. Connections to and use of open access journal articles by the anti-vaccination movement

The findings in Chapter 5 show that there is no evidence of the anti-vaccination movement using open research data related to the claimed link between vaccination and autism. Without evidence of connection to open research data, it was not possible to extend the inquiry to *who* the users of open research data within the anti-vaccination movement are, nor *how* they may be using open research data. Consequently, the inquiry was broadened to include a second type of scientific product: open access journal articles. By doing so, the likelihood of finding connections and use was increased, and a new line of inquiry was opened: an examination of differences in the connections to the use of two products by a specific non-scientific community.

To determine connections to and the use of open access journal articles by non-scientists, open access journal articles on the topic of the claimed relationship between autism and vaccinations and mentions to those articles by the anti-vaccination movement, was selected as the area of focus. Data from the Altmetric Explorer database that contains mentions to specific journal articles in the social media and other online sources was used to determine connections and use.

The first part of this chapter presents findings related to the level of attention and connections by the non-scientific public in general to open access journal articles. The second part of the findings relates to mentions as connections made specifically by the anti-vaccination movement to 10 selected open access journal articles. This section includes findings on the relative size and activity of the anti-vaccination movement on Twitter and Facebook. The third and final section of this chapter presents findings on the use of open access journal articles on Twitter, Facebook and web pages, by applying level of engagement as a proxy for use.

6.1 Connections to open access journal articles by non-scientific publics

It is instructive to consider the Attention Scores and the composition of publics mentioning the articles on Twitter for all articles in the top-down purposive sample and the bottom-up snowball sample. Of the 57 open access journal articles linked to open research datasets related to the claimed link between autism and vaccination, 49 (86%) had an Altmetric Attention Score of 1 or more. The highest Altmetric Attention Score was 511. To put this score into perspective, the highest Altmetric Attention Score on 27 June 2017 for the 11,351,607 research products indexed in the Altmetric database was 10,127. The median Altmetric Attention Score for the 57 articles was 4 and the average Attention score was 32.[41] The number of articles tweeted by at least one member of the public was 47 (82%). The highest number of mentions on Twitter by members of the public was 215, the median score was 2, and the average number of mentions was 11.[42]

Of the 75 open access research publications referenced by the anti-vaccination web pages (i.e. the snowball sample), 69 (92%) had an Altmetric Attention Score of 1 or more (as at 29 June 2017). The highest Altmetric Attention Score was 3,641. The median Altmetric Attention Score for the 75 articles was 32 and the average score was 217. The number of articles tweeted by at least one member of the public, was 64 (85%).[43] Table 11 shows those articles from the two samples that were found to have the highest Altmetric Attention Score, which is taken as an indicator of the level of attention garnered by an open access journal article from a broad spectrum of publics in the social media and on other Web 2.0 platforms.

Table 11: Levels of attention for the highest scoring open access journal articles (October 2017)

Article ref.	DOI	Title	Date of measurement	Altmetric Attention Score
1.1	10.1371/journal.pbio.1002198	Imperfect vaccination can enhance the transmission of highly virulent pathogens	18 Oct	511
1.2	10.15585/mmwr.ss6503a1	Prevalence and characteristics of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012	18 Oct	311
1.3	10.1371/journal.pone.0003140	Lack of association between measles virus vaccine and autism with enteropathy: A case-control study	18 Oct	306
1.4	10.1186/2047-217x-3-18	GWATCH: A web platform for automated gene association discovery analysis	18 Oct	113
1.5	10.1371/journal.pbio.1001368	The evolutionary consequences of blood-stage vaccination on the rodent malaria *plasmodium chabaudi*	18 Oct	96
2.1	10.1001/jama.2015.3077	Autism occurrence by MMR vaccine status among US children with older siblings with and without autism	17 Oct	3,674
2.2	10.1016/j.vaccine.2014.04.085	Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies	24 Oct	2,989
2.3	10.1080/15287394.2011.573736	A positive association found between autism prevalence and childhood vaccination uptake across the US population	24 Oct	1,336
2.4	10.1186/2047-9158-3-16	Measles-mumps-rubella vaccination timing and autism among young African-American boys: A reanalysis of CDC data	24 Oct	1,048
2.5	10.1186/2047-9158-2-25	A two-phase study evaluating the relationship between thimerosal-containing vaccine administration and the risk for an autism spectrum disorder diagnosis in the United States	24 Oct	1,018

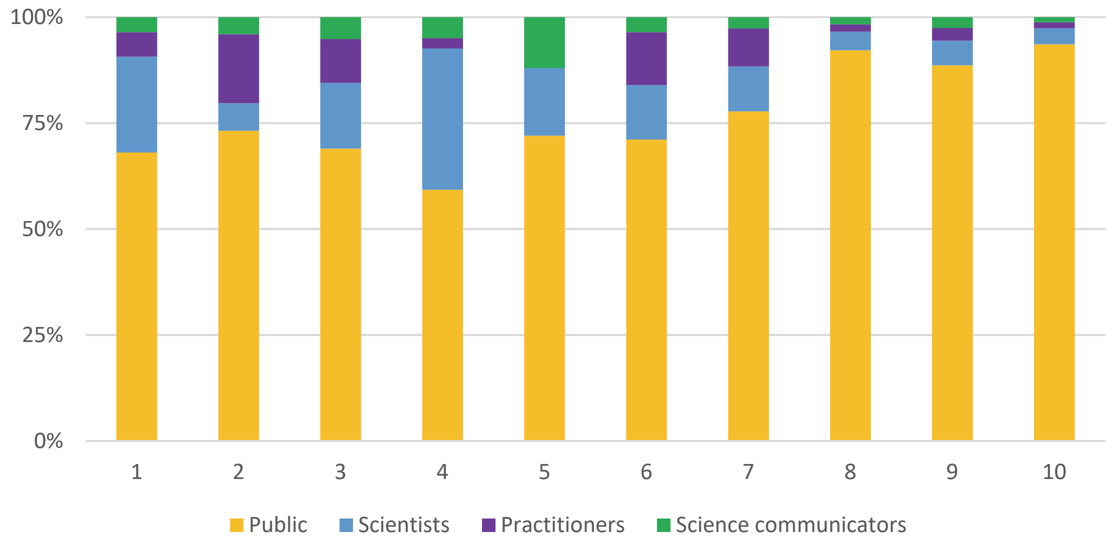
[41] The median and average Altemtric.com attention scores could not be calculated for the entire database of 11,351,607 research outputs because exports of data from the Altmetric Explorer are limited to 1,000,000 research outputs.
[42] No comparable data available from Altmetric for the entire database of 11,351,607 research outputs.
[43] As for data sourced from open data repositories and journals, no comparable data available from Altmetric.

Altmetric does provide disaggregated data by type of user, but only for data from Twitter. Altmetric provides disaggregated data for the following broad groups: the public; scientists; professionals; and science media/journalists. Based on an analysis of the five articles with the highest Altmetric Attention Score in the purposive and snowball samples respectively, Figure 9 and Table 12 show that on Twitter, the public accounts for the majority of the attention attracted by the 10 journal articles, regardless of whether the article is one that is relevant to the vaccines–autism debate on a web-based repository or whether it is referenced on an anti-vaccination web page.

Table 12: Distribution of connections on Twitter by publics for a sample of open access journal articles on vaccines–autism

Article ref.	No. of Twitter users	Public	Scientists	Practitioners	Science communicators
1.1	310	211	70	18	11
1.2	123	90	8	20	5
1.3	58	40	9	6	3
1.4	81	48	27	2	4
1.5	25	18	4	0	3
2.1	3,186	2,265	410	397	113
2.2	2,767	2,150	293	248	74
2.3	1,588	1,452	69	27	27
2.4	941	826	54	28	24
2.5	1,403	1,307	53	20	17

Figure 9: Breakdown of attention by attentive publics for 10 open access journal articles (%)



Comparatively speaking, the findings show a marked difference between the levels of attention when comparing the top-down sample with the bottom-up sample. This is to be expected to the extent that the top-down sample of articles (1) was created by using keyword searches, and (2) is agnostic of any potential engagement by non-scientific publics. Furthermore, while Altmetric does index blogs, it does not index the entire web, and is therefore unlikely to index all the anti-vaccination web pages that contain references to open access journal articles.[44] Conversely, the bottom-up sample was created from instances where an article had already captured the attention of at least one member of the anti-vaccination community (the author of the web page).

Nevertheless, the degree of comparative difference is unexpected, and reveals that even though a journal article – the content of which is seemingly relevant given the political agenda of the anti-vaccination movement – attracts high levels of attention on the web among the scientific community, this is not necessarily an indicator that the article will attract the attention of the anti-vaccination movement.

These findings suggest that non-scientists are connecting to the products of open science (in this case, open access journal articles) if we use level of attention and mentions on the social media platform Twitter as a proxy for connections. The findings do not, however, provide any insight into the composition of Altmetric's 'public' category; nor their stance towards vaccination. The findings also do not account for the possibility that practitioners and even scientists may be aligned ideologically with some of the communities that constitute the public category.

In order to be more specific about who is accessing open access journal articles and to differentiate publics by stance, disaggregation of the data was done to isolate the anti-vaccination movement.

6.2 Connections to open access journal articles by the anti-vaccination movement

6.2.1 Connections to 10 open access journal articles from Twitter

Further disaggregation of attentive publics on Twitter was done by determining the number of anti-vaccination Twitter accounts in each sample of Twitter accounts that mention one of the 10 journal articles.

The findings in Table 13 show that the proportion of anti-vaccination accounts to all accounts mentioning one of the 10 articles did not exceed 18%. In other words, no more than 1 in 5 mentions to an open access journal article related to the autism-vaccination debate were from Twitter users

[44] The fact that Altmetric does not count all web pages accounts for the fact that six pages returned an Altmetric score of zero when it is known that at least one web page links to an open access journal article.

whose stance was anti-vaccination. The findings do nevertheless confirm that unlike in the case of open research data, the anti-vaccination movement is accessing open access journal articles.

The table also shows that articles accessed fall into two broad groups: one group of 5 articles (1.1, 1.5, 2.3, 2.4 and 2.5) in which the mentions by anti-vaccination accounts was between 11% and 18% relative to all unique accounts mentioning the article on Twitter, and a second group of 5 articles (1.2, 1.3, 1.4, 2.1 and 2.2) in which fewer than 2% of mentions originated from anti-vaccination accounts.

Table 13: Number and proportion of anti-vaccination accounts on Twitter

Sample	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5
No. of unique accounts	310	123	58	81	25	3,187	2,775	1,567	931	1,397
No. of verified anti-vaccination accounts	36	1	1	0	4	35	40	218	166	157
% anti-vaccination accounts	11.6%	0.8%	1.7%	0%	16.0%	1.1%	1.4%	13.9%	17.8%	11.2%

6.2.2 Activity of the anti-vaccination movement on Twitter

Further analysis of the data is possible to determine the proportion of tweets (as opposed to accounts) posted by the anti-vaccination movement and which make mention of one of the 10 open access journal articles.

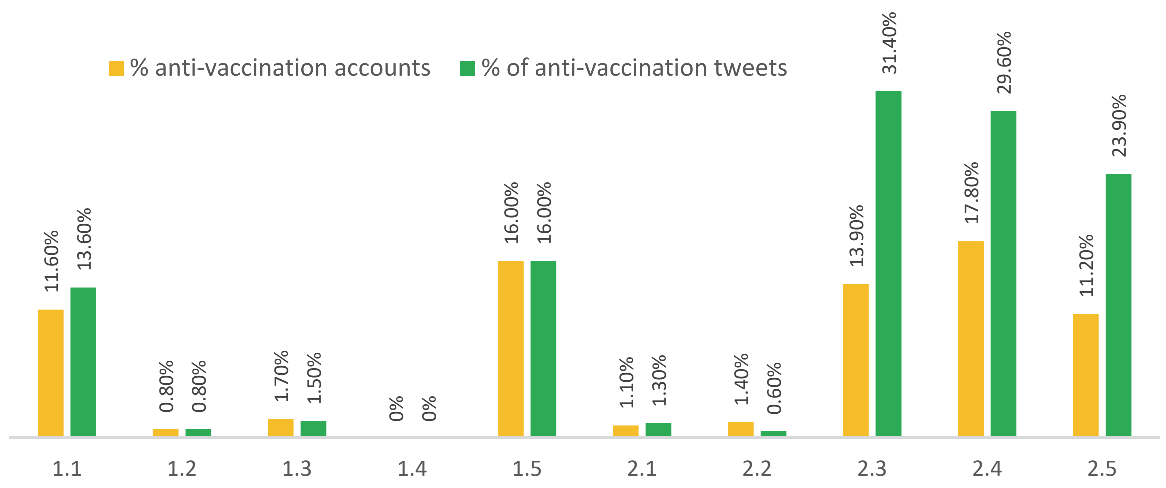
Table 14 shows the proportion of tweets by the anti-vaccination movement compared to all tweets that make mention of one of the 10 open access journal articles. The proportion of tweets varies by article and again present in two distinct groups that correspond with the two anti-vaccination Twitter account groups.

Table 14: Number and proportion of anti-vaccination tweets

Sample	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5
Total no. of tweets	382	131	69	93	25	3,551	3,509	2,589	2,268	2,282
No. of anti-vaccination tweets	52	1	1	0	4	45	21	812	672	545
% of anti-vaccination tweets	13.6%	0.8%	1.5%	0%	16.0%	1.3%	0.6%	31.4%	29.6%	23.9%

Figure 10 compares the proportion of anti-vaccination Twitter accounts with the proportion of anti-vaccination tweets for each of the 10 articles. The graph shows that for those articles that appear to be of interest to the anti-vaccination movement (that is, articles 1.1, 1.5, 2.3, 2.4 and 2.5), the proportion of tweets by the anti-vaccination movement is equal to or higher than the proportion of anti-vaccination Twitter accounts for the same open access journal article. The difference is most pronounced in the cases of articles 2.3, 2.4 and 2.5.

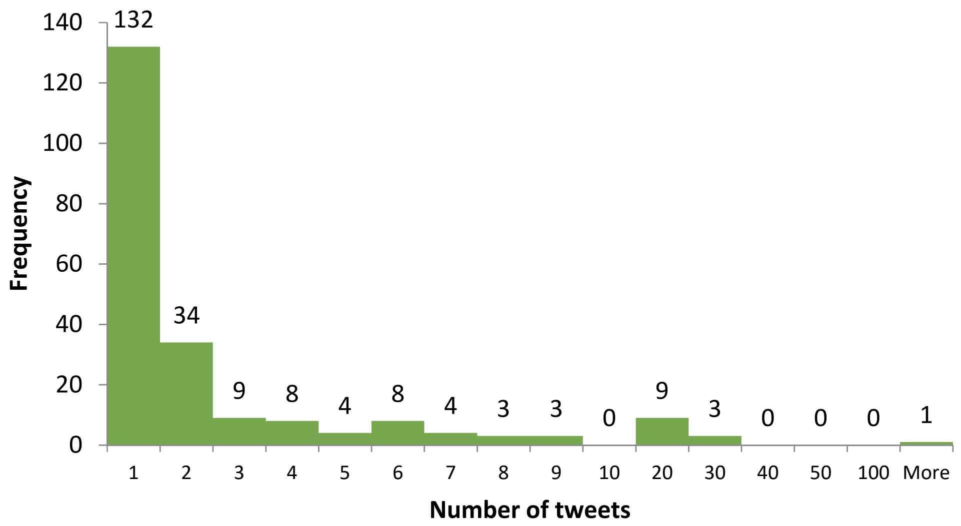
Figure 10: % of anti-vaccination accounts compared to % of anti-vaccination tweets by article



It is possible to determine who the most active Twitter accounts are in the samples of unique anti-vaccination accounts mentioning the three most frequently mentioned open access journal articles (that is, articles 2.3, 2.4 and 2.5).

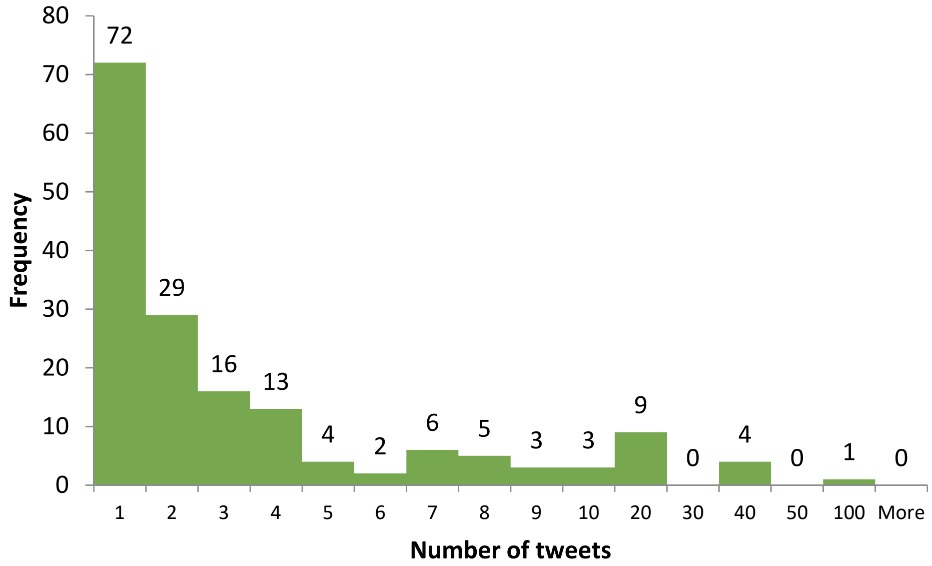
There were 218 unique Twitter accounts mentioning article 2.3 in 812 tweets. Figure 11 shows the frequency with which each unique account tweeted a mention to the article.

Figure 11: Article 2.3: Frequency of mentions by unique accounts on Twitter (n=812)



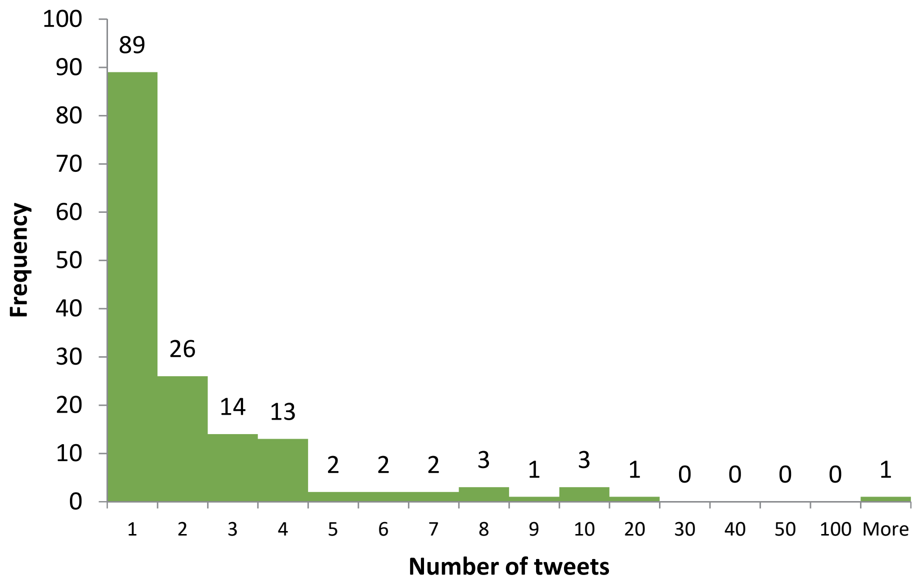
There were 167 unique Twitter accounts mentioning article 2.4 in 672 tweets. Figure 12 shows the frequency with which each unique account tweeted a mention to the article.

Figure 12: Article 2.4: Frequency of mentions by unique accounts on Twitter (n=672)



There were 157 unique Twitter accounts mentioning article 2.5 in 545 tweets. Figure 13 shows the frequency with which each unique account tweeted a mention to the article.

Figure 13: Article 2.5: Frequency of mentions by unique accounts on Twitter (n=545)



The data in Figure 11, Figure 12 and Figure 13, show a skewed distribution in which the majority of Twitter accounts mention an article only once. The data also show that for all three articles, there are a few accounts that mention the article more than 10 times, and in all three cases there is one account that mentions the article 100 or more times.

Without relevant comparative data, it is impossible to assess whether these are high levels of activity. Nevertheless, mentioning an article more than once does seem to indicate more than a incidental level of activity, and mentioning it more than 100 times certainly does seem to indicate activity at the upper end of the activity scale.

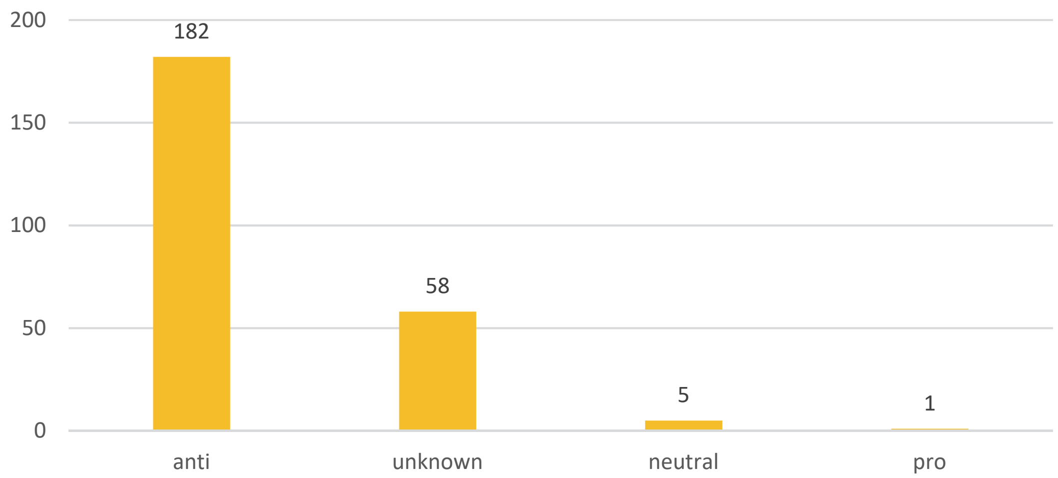
Further analysis of these accounts with a focus on the content of their tweets and what this may reveal about the anti-vaccination movement's level of engagement with open access journal articles is done in section 6.3.

6.2.3 Connections to article 2.3 by the anti-vaccination movement on Facebook

There are limitations of a technical nature that preclude the analysis of Facebook data to the same extent as is possible with, for example, Twitter data. Such limitations meant that it was not possible to use automated tools to analyse a large number of Facebook accounts to determine their stance vis-à-vis vaccination. It was therefore decided to focus only on article 2.3, and to conduct manual analysis of all 246 Facebook accounts that mention article 2.3.

Figure 14 shows that the majority (182, 74%) of Facebook accounts that mentioned the open access journal article 2.3 'A Positive Association found between Autism Prevalence and Childhood Vaccination uptake across the US Population' were anti-vaccination. The remainder of the accounts were either neutral (5, 2%) or it was not possible to determine the stance of the account (58, 24%). Only 1 (0.4%) Facebook account that mentioned the open access journal article in question was found to be pro-vaccination.

Figure 14: Stance of Facebook accounts making mention of a single open access journal article vis-à-vis vaccination (n=246)



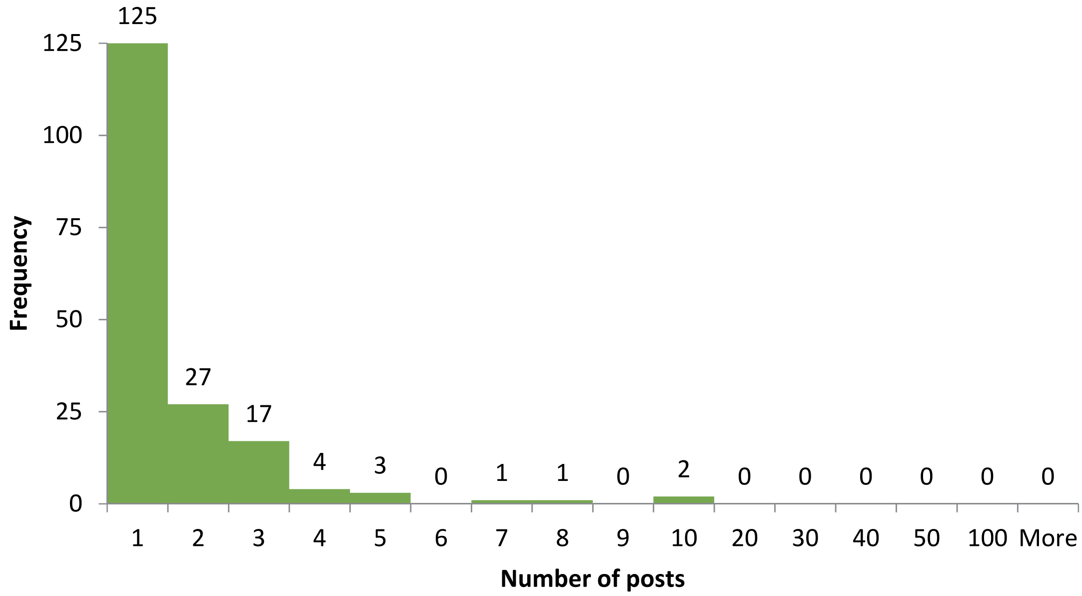
If the unknown accounts are excluded, the proportion of anti-vaccination accounts increases to 97% of all known Facebooks accounts mentioning article 2.3. This finding that almost all accounts on Facebook mentioning article 2.3 are anti-vaccination is in stark contrast to the findings for Twitter where only 13.9% of accounts found to mention the same article were anti-vaccination accounts.

6.2.4 Activity of the anti-vaccination movement on Facebook

In terms of number of Facebook posts mentioning article 2.3 in relation to the number of anti-vaccination Facebook accounts, it was found that the 182 anti-vaccination accounts contributed 294 posts. This equates to 74% (182 out of 246) accounts contributing 78% (294 out of 375) of the posts mentioning article 2.3. In other words, each anti-vaccination account posts, on average, one mention of the article. In the case of Twitter, 13.9% of the accounts contributed 31.4% of the tweets; more than an average of 2 tweets per account mentioning article 2.3.

Figure 15 shows the frequency with which each of 182 unique anti-vaccination Facebook accounts posted a mention to article 2.3 from a total of 294 mentions. Figure 15 shows a lower frequency of posts per account than was the case for Twitter. Most Facebook accounts (125, 69%) post only once; only 2 (1%) accounts posted 10 times. No accounts posted more than 10 times on Facebook whereas in the case of Twitter 13 (6%) accounts tweeted more than 10 times.

Figure 15: Article 2.3: Frequency of mentions by unique accounts on Facebook (n=294)



6.2.5 Connections by the anti-vaccination movement to open access journal articles from web pages

During the DOI snowballing sampling process, 75 connections from anti-vaccination web pages to open access journal articles were found. This provides evidence (1) that the anti-vaccination movement is making reference to open access journal articles from its web pages, and (2) of the *potential* use of open access journal articles to support its ideology and political agenda. It is to the use of open access journal articles by the anti-vaccination movement that the next section turns its attention.

6.3 Use of open access journal articles by the anti-vaccination movement

Online use of open access journal articles was measured by assessing the level of engagement with open access journal articles by the anti-vaccination movement in its online communications. The approach adopted is an attempt to go beyond views, downloads or mentions as proxies for the use of online content. The level of engagement was measured for selected open access journal articles in the social media (Twitter and Facebook) and on the web pages of anti-vaccination websites.

Level of engagement was measured on a 6-point scale in three categories: access (a score of 1 or 2); appraisal (a score of 3 or 4); and application (a score of 5 or 6). A score of 1 represents the lowest level of engagement while a score of 6 represents the highest level of engagement.

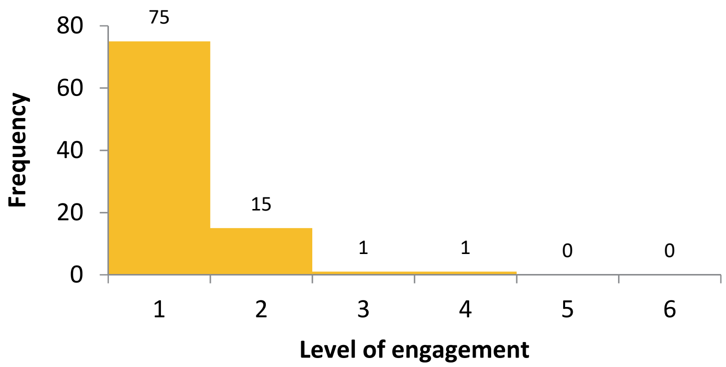
6.3.1 Level of engagement on Twitter

Based on the findings of the anti-vaccination movement's presence and activity on Twitter, only three articles were selected to assess the community's level of engagement with those articles: 2.3, 2.4 and 2.5. The selection of these three articles was determined by the fact that they are the articles that garnered the most attention from the anti-vaccination movement on Twitter.

Article 2.3

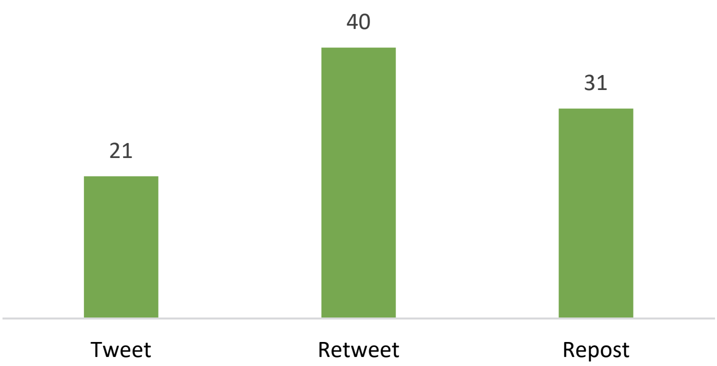
The findings for the levels of engagement with Article 2.3 on Twitter are shown in Figure 16 and Figure 17. Figure 16 shows that the distribution of scores for level of engagement on Twitter fell predominantly in the access category: 90 (98%) tweets scored either 1 or 2 on the scale. Only 2 tweets fell in the appraisal category.

Figure 16: Article 2.3 Level of engagement on Twitter (n=92)



The low levels of engagement with article 2.3 is explained in part by the finding shown in Figure 17 that many of the tweets were either retweets (40, 43%) or reposts[45] (31, 34%). In other words, of the 92 tweets, only 21 (23%) consisted of original content likely to return higher scores for level of engagement.

Figure 17: Article 2.3: Tweets, retweets and reposts as an indicator of engagement (n=92)



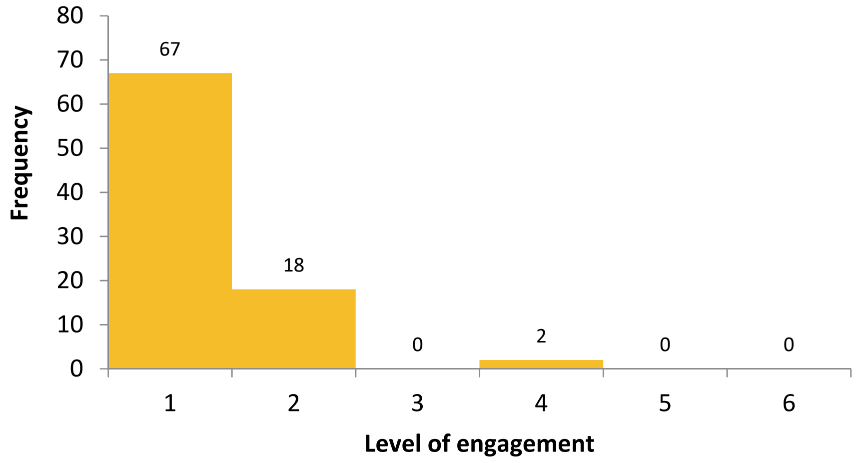
Article 2.4

The findings for the levels of engagement with Article 2.4 on Twitter are shown in Figure 18 and Figure 19. Figure 18 shows that the distribution of scores for level of engagement on Twitter fell predominantly in the access category: 85 (98%) tweets scored either 1 or 2 on the scale. Only 2 tweets fell in the appraisal category.

---

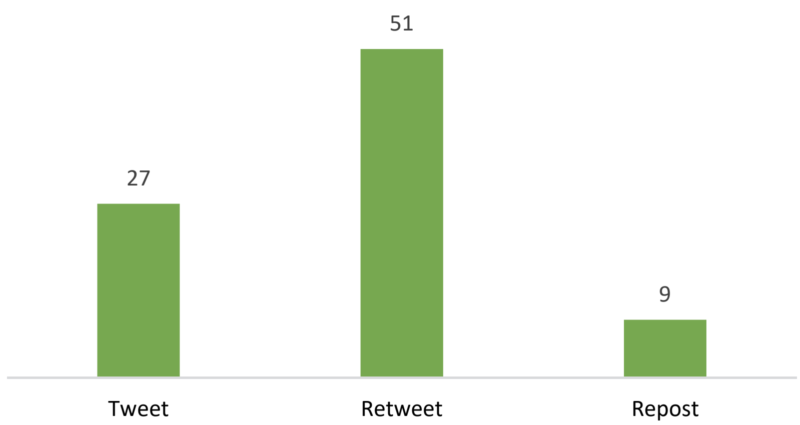
[45] A repost is defined as occurring when an account creates a new tweet or a comment that uses the exact same content as a previous tweet by the same account.

Figure 18: Article 2.4 Level of engagement on Twitter (n=87)



As in the case of Article 2.3, the low levels of engagement with article 2.4 is explained by the finding shown in Figure 19 that many of the tweets were either retweets (51, 59%) and reposts (9, 10%). Of the 87 tweets by the anti-vaccination movement, only 27 (31%) consisted of original content.

Figure 19: Article 2.4: Tweets, retweets and reposts as an indicator of engagement (n=87)

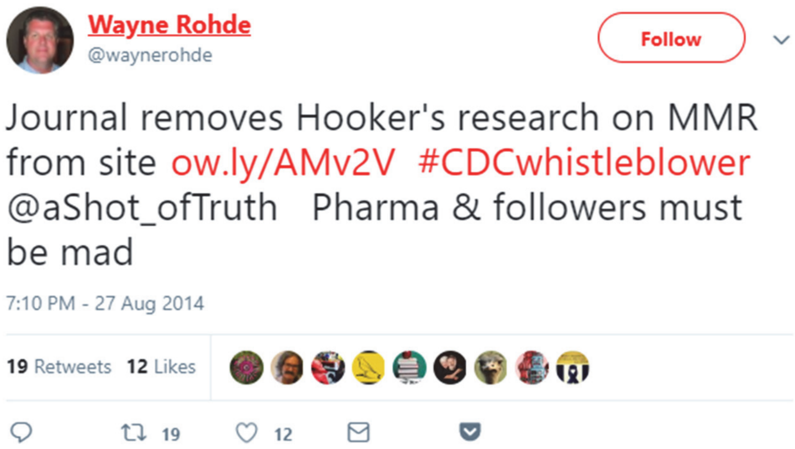


Three observations can be made in relation to level of engagement by the anti-vaccination movement with Article 2.4 on Twitter. The first is that many of the tweets for this article do not relate to its content per se but to the fact that the article was retracted: 'An expression of concern has been published for this article. This article has been retracted. See Transl Neurodegener. 2014; 3: 22'.[46] See Figure 20 for an example of a tweet on the retraction of the article. The motivation

[46] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4128611/

behind these tweets is to 'prove' collusion between CDC, the pharmaceutical industry and science, and there is consequently little engagement with the actual content of the article.

Figure 20: Example of a tweet regarding the retraction of article 2.4



The second observation is the number of tweets providing an alternative link to the article post-retraction (see Figure 21). The intent of these tweets is to inform the community that the article remains accessible and, as such, available to them to support their campaign regardless of the fact that the scientific community has retracted the article. Again, the posting of a link does not indicate a high level of engagement with the content of the article; particularly so if the community does not engage with the scientific motivations for its retraction and elects instead to interpret the article's removal as being politically motivated.

Figure 21: Example of a tweet providing an alternative link to retracted article 2.4

The third observation is the use of 'broadcast' tweets in the case of mentions to this article. To illustrate: An anti-vaccination Twitter account will tag another anti-vaccination account by prefixing a (re)tweet with the Twitter handle of another member of the movement. The tweet may also include a call to action, a link or a hashtag (e.g. '#CDCwhistleblower'). Often the tagged anti-vaccination accounts will have a much larger number of Twitter followers and/or be more active on Twitter than the tweeter. In the example below (Figure 22), the tweeter had 2,888 followers while the tagged account @TannersDad had almost ten times as many followers (21,400). And while the tagged accounts @ceestave and @NOWinAutism had numbers of followers closer in number to that of the tweeter, both accounts are highly active. The tweeter @MarcellaPiperTe had tweeted 13,200 times (since joining in June 2014), while the tagged accounts @ceestave had tweeted 58,500 times (since May 2009) and @NOWinAustism 59 700 times (since August 2014). @TannersDad is also a highly active account with 222,000 tweets (since joining in October 2008).[47]

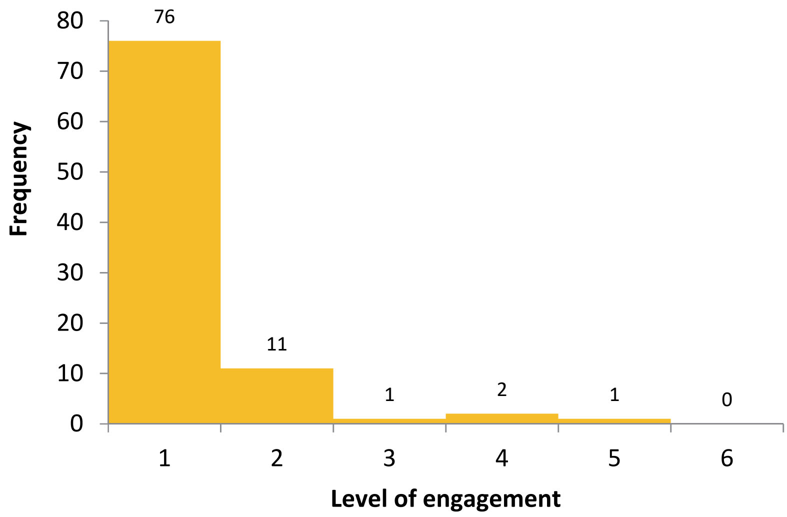
Figure 22: Example of a tweet broadcasting article 2.4 to other Twitter accounts

Article 2.5

The findings for the levels of engagement with Article 2.5 on Twitter are shown in Figure 23 and Figure 24. Figure 23 shows that the distribution of scores for level of engagement on Twitter fell predominantly in the access category: 87 (96%) tweets scored either 1 or 2 on the level of engagement scale. Only three tweets fell in the appraisal category, and Article 2.5 was the only article in which a tweet was scored as being in the application category.

[47] All followers and number of tweets as on 26 March 2018.

Figure 23: Article 2.5: Level of engagement on Twitter (n=91)

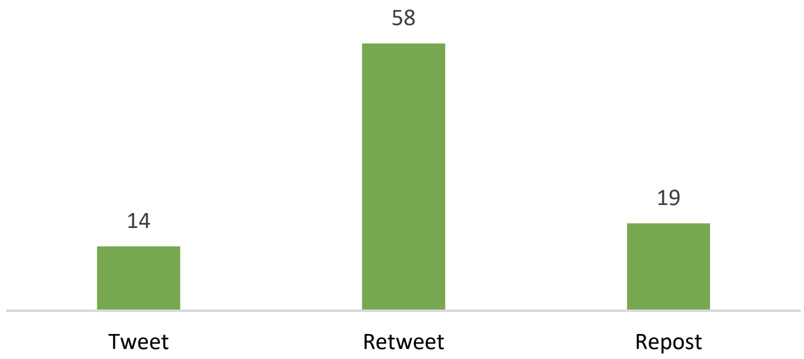


The tweet that scored 5 on the level of engagement scale was a retweet of a previous tweet by the same account.[48] However, additional information was added to the tweet thread in the form of data published by the US Federal Drug Administration (FDA) and information from a journal article (including underlined text from the methods section). The first four comments comprise selected extracts the FDA on the presence of mercury in vaccines, for example, '#Flu #vaccine FLUZONE, p.18: Each 0.5 mL dose contains 25 mcg #mercury, 0.25 mL (infant) dose - 12.5 mcg mercury'. The fifth comment mentions a journal that 'Found 7.6-FOLD Increased Risk of #AUTISM from Exposure to #Thimerosal' and includes an image of the results section of that article with the finding highlighted in red.[49] The tweet was deemed to indicate a level of interpretation consistent with being categorised on the 'application' end of the scale.

The overall low level of engagement with article 2.5 is again explained by the finding shown in Figure 24 that many of the tweets were either retweets (58, 64%) and reposts[50] (19, 21%). In other words, of the 91 tweets, only 14 (15%) consisted of original content thereby limiting the possibility of higher scores for level of engagement.

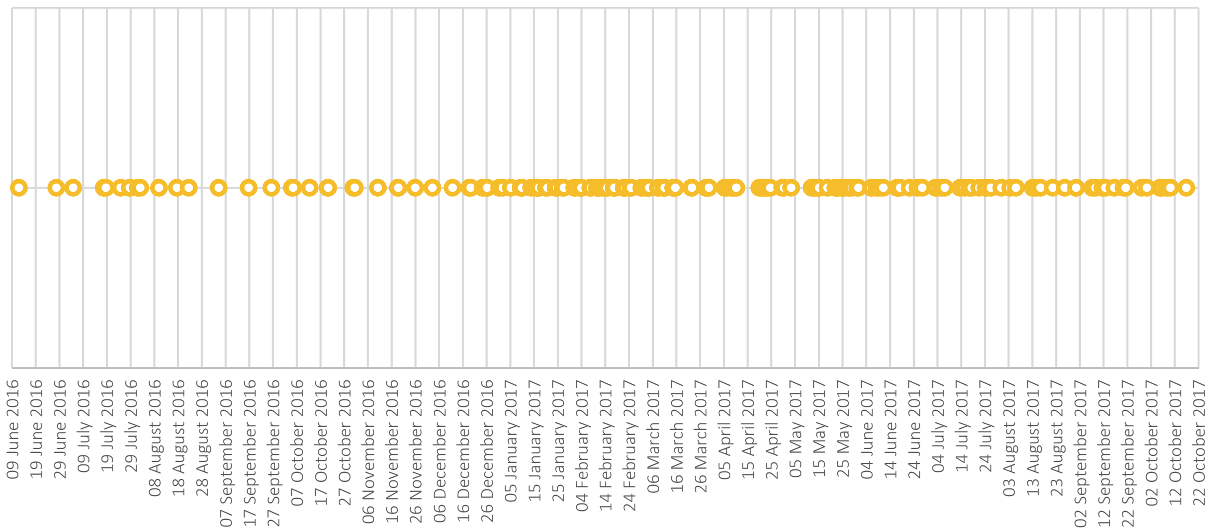
[48] See the tweet in question at http://twitter.com/LotusOak/statuses/916317625407430657
[49] https://twitter.com/LotusOak/status/909814655044001794
[50] A repost is defined as occurring when an account creates a new tweet or a comment that uses the exact same content as a previous tweet by the same account.

Figure 24: Article 2.5: Tweets, retweets and reposts as an indicator of engagement (n=91)



Furthermore, of the tweets that mention article 2.5 in the random sample of anti-vaccination accounts, 55% (50 of 91) were retweets by the account @LotusOak. Figure 25 shows that the Twitter account @LotusOak tweets consistently from June 2016 to October 2017. What is notable is the fact that these are not unique tweets; the majority of the tweets are @LotusOak retweeting or reposting the same tweet that reads as follows: '#STUDY: #Thimerosal-containing #Vaccines & #Autism Risk http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3878266/ … It's still in multi-dose vaccines'.[51] That @LotusOak retweets the same content verbatim is further evidence of a consistently low level of engagement on Twitter as far as this scientific product is concerned.

Figure 25: Tweet frequency of @LotusOak to article 2.5 (n=197)



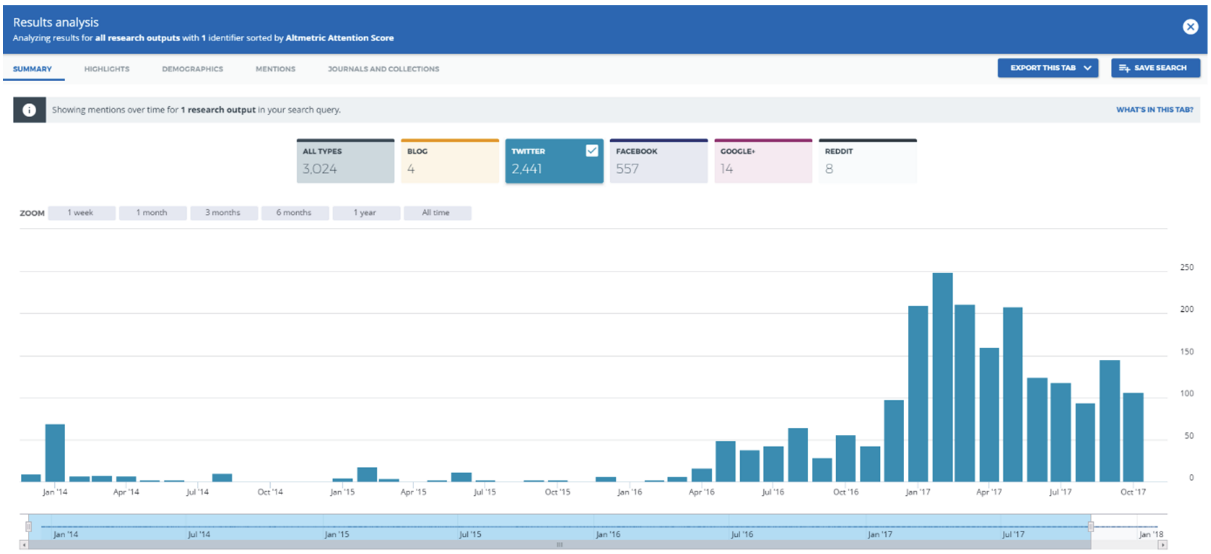
There is also evidence of the tweet being modified in this example, possibly to reignite attention in the anti-vaccination Twitter community. Figure 26 shows that activity on Twitter increased

---

[51] https://twitter.com/LotusOak/statuses/832598920568111104

remarkably from April 2016, even though the article was published and first mentioned on Twitter in December 2013. On 16 September 2017, the Twitter account @LotusOak reposts the same tweet but on 18 September 2017, content is added to the tweet in the form of five separate comments to support the association between mercury in vaccines and autism.[52] While the exact reason for the increase in interest several years after publication remains unknown, a single influential Twitter account modifies the content of an original tweet to propel and support the information flows linked to findings from a specific open access journal article in the Twitter network.

Figure 26: Time-line for all mentions to article 2.5 on Twitter[53]

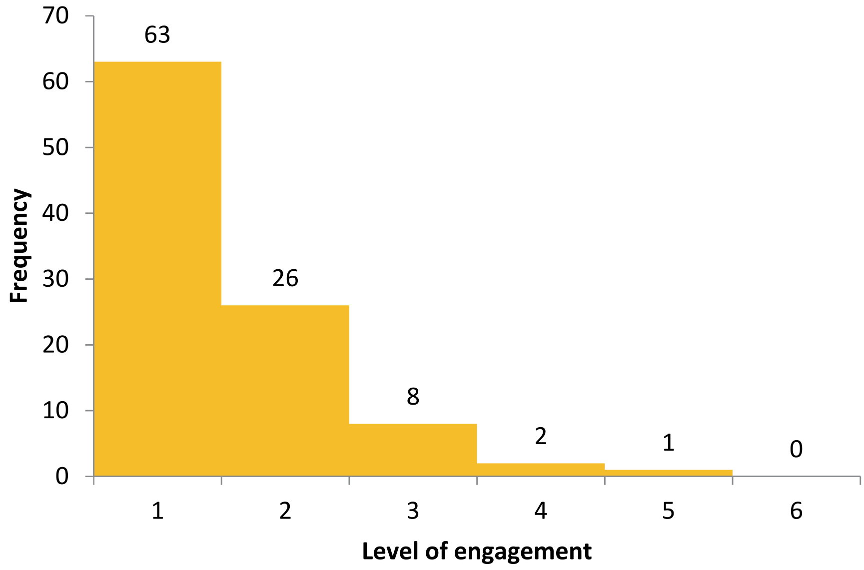


6.3.2 Level of engagement on Facebook

The finding for the anti-vaccination movement's level of engagement with article 2.3 on Facebook is shown in Figure 27. The findings presented in the figure show that the distribution of scores for level of engagement on Facebook fell predominantly in the access category with 89 (89%) posts scoring either 1 or 2 on the scale; 10 (10%) posts scored 3 or 4, falling in the appraisal category; and only one (1%) post on Facebook fell in the application category with a score of 5.

[52] See https://twitter.com/LotusOak/status/909060995640971264 and https://twitter.com/LotusOak/status/909814655044001794
[53] See https://www.altmetric.com/explorer/outputs/attention?identifier_list_id=ed9ebaba-ed85-45d2-8eae-b61fb7207b20

Figure 27: Article 2.3: Level of engagement on Facebook (n=100)



As was the case for engagement on Twitter, the low level of engagement can be explained by the frequent reposting of content on Facebook.

Several tropes stand out from the sample, and these are shown in Table 15 along with an indication of the frequency with which these tropes appeared in the Facebook posts of the anti-vaccination movement.

Table 15: Common themes or tropes appearing in Facebook posts mentioning article 2.3

	Trope	No. of posts	% of total
A	'There is no science that shows vaccines cause Autism, except in these published studies which show vaccines cause Autism'	**14**	**14%**
B	'50 Peer reviewed studies from the US National Library of Medicine National Institutes of Health that link Autism to Vaccines'	**5**	**5%**
C	'25 Studies from the US National Library of Medicine National Institutes of Health that link Autism to Vaccine'	**15**	**15%**
D	'Did your doctor mention this? A positive association found between autism prevalence and childhood vaccination uptake across the U.S. population.'	**2**	**2%**
A + B	'There is no science that shows vaccines cause Autism, except in these published studies which show vaccines cause Autism. OVER 50 PEER REVIEWED SCIENTIFIC Studies from the US National Library of Medicine National Institutes of Health that link Autism to Vaccines'	**5**	**5%**
	TOTAL	**41**	**41%**

In the sample of 100 posts there was also evidence of tropes evolving and being blended. For example, 'There is no #science that shows the correlation between #vaccines and #Autism; except these published studies that show a correlation between vaccines and Autism' becomes 'There is no science that shows vaccines cause Autism, except in these published studies which show vaccines cause Autism' and then 'There is no science that shows vaccines cause autism, except in these peer reviewed published studies from the medical literature which show that vaccines cause autism' in subsequent reposts. The first evolution sees the omission of 'correlation', possibly because of the frequent backlash received by the anti-vaccination movement when claiming correlation proves causation. In the next evolution, 'peer reviewed' and 'from the medical literature' is added, possibly to amplify the legitimacy of the findings. Trope A is also blended with Trope B over time to create a hybrid post 'There is no science that shows vaccines cause Autism, except in these published studies which show vaccines cause Autism. OVER 50 PEER REVIEWED SCIENTIFIC Studies from the US National Library of Medicine National Institutes of Health that link Autism to Vaccines' which appears 5 times in the sample.

6.3.3 Level of engagement on the web

The findings for level of engagement by the anti-vaccination movement with open access journal articles on the topic of vaccination and its potentially harmful side-effects via web pages is shown in Figure 28. The figure shows a higher level of engagement compared to the levels of engagement on Twitter and Facebook: 13 (38%) web pages fall into the access category, 13 (38%) fall into the appraisal category, and 8 (24%) fall into the application category. The relatively high number of web pages in both the appraisal and the application clusters differentiates the findings on the level of engagement on the web with those on Twitter and Facebook.

Figure 28: Level of engagement with open access journal articles on the web (n=34)

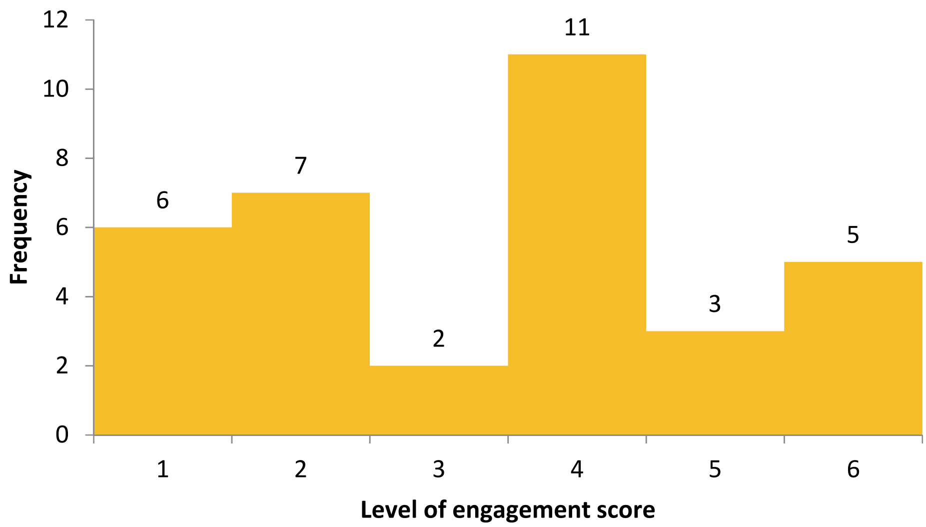
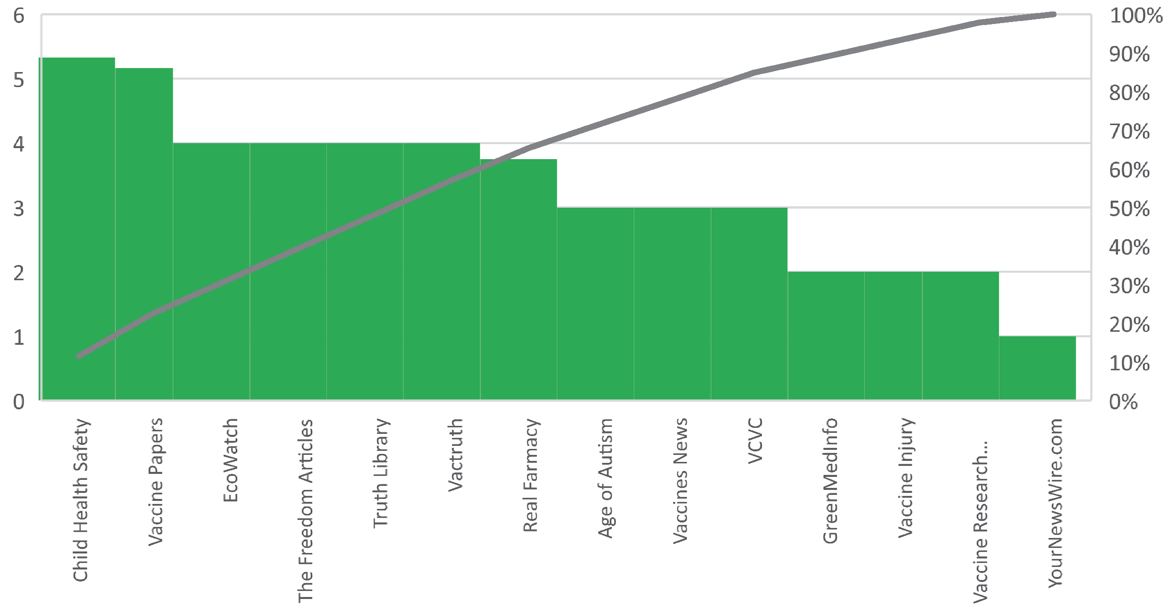


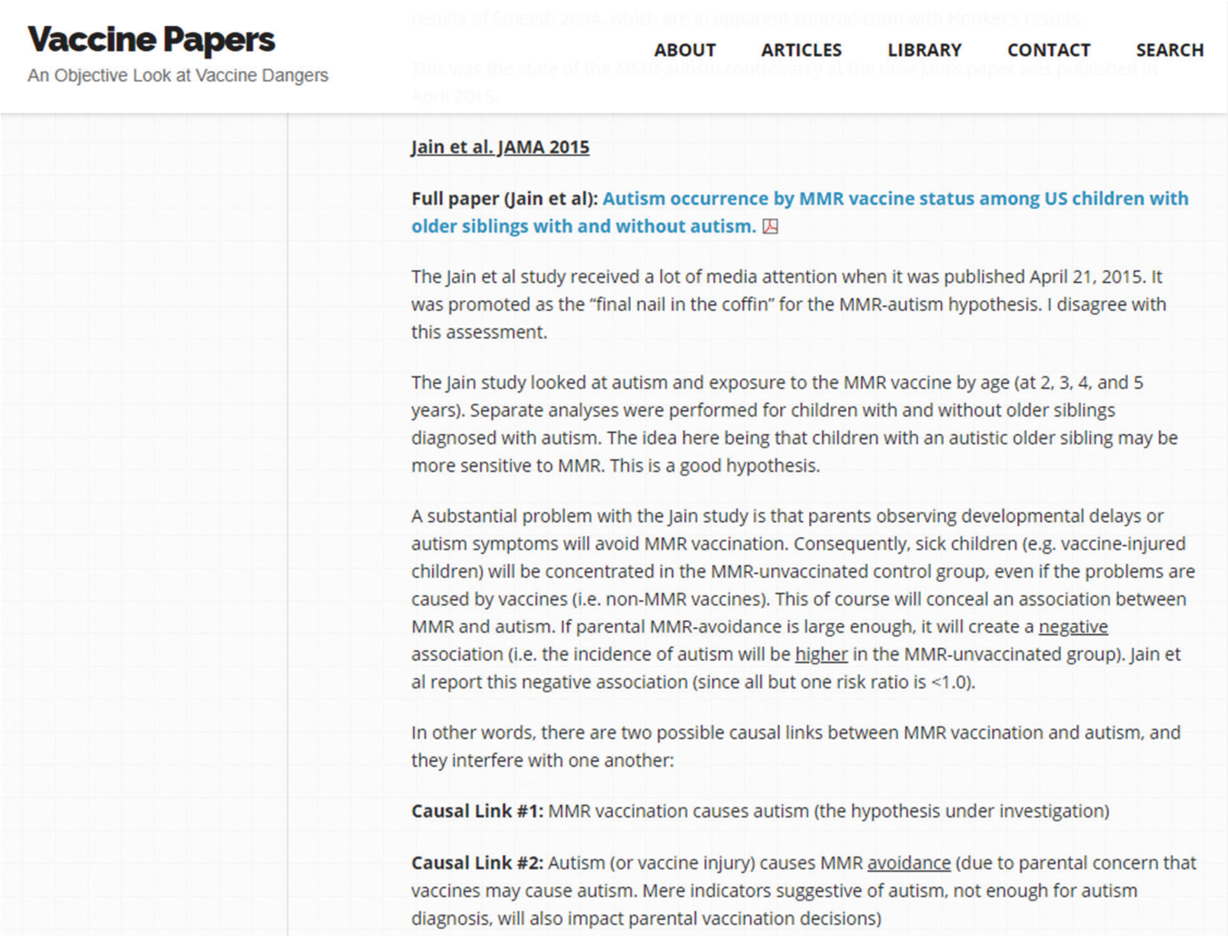
Figure 29 shows that of the 14 unique websites that published the 34 web pages, 2 stand out as publishing web pages in which the authors of those pages engage closely with the content of an open access journal article: (1) Child Health Safety and (2) Vaccine Papers.

Figure 29: Average level of engagement score for web pages published by unique anti-vaccination websites (n=14)



Authors of articles published on these two anti-vaccination websites present findings in their own words, they reinterpret findings by engaging critically with the methods and/or analyses presented in the original study, often focusing on a scientific paper that shows no association between vaccination and autism, and they refer to other scientific research to support their reanalysis. An additional mark of their close engagement with the journal article is that the authors of these articles reply to questions and challenges posed in the comments section of the web page, often posted by pro-vaccination individuals, in order to provide additional clarity in support of their reanalysis.

Figure 30: Extract from a highly engaged anti-vaccination web page


Vaccine Papers
An Objective Look at Vaccine Dangers

ABOUT ARTICLES LIBRARY CONTACT SEARCH

Jain et al. JAMA 2015

Full paper (Jain et al): Autism occurrence by MMR vaccine status among US children with older siblings with and without autism.

The Jain et al study received a lot of media attention when it was published April 21, 2015. It was promoted as the "final nail in the coffin" for the MMR-autism hypothesis. I disagree with this assessment.

The Jain study looked at autism and exposure to the MMR vaccine by age (at 2, 3, 4, and 5 years). Separate analyses were performed for children with and without older siblings diagnosed with autism. The idea here being that children with an autistic older sibling may be more sensitive to MMR. This is a good hypothesis.

A substantial problem with the Jain study is that parents observing developmental delays or autism symptoms will avoid MMR vaccination. Consequently, sick children (e.g. vaccine-injured children) will be concentrated in the MMR-unvaccinated control group, even if the problems are caused by vaccines (i.e. non-MMR vaccines). This of course will conceal an association between MMR and autism. If parental MMR-avoidance is large enough, it will create a negative association (i.e. the incidence of autism will be higher in the MMR-unvaccinated group). Jain et al report this negative association (since all but one risk ratio is <1.0).

In other words, there are two possible causal links between MMR vaccination and autism, and they interfere with one another:

Causal Link #1: MMR vaccination causes autism (the hypothesis under investigation)

Causal Link #2: Autism (or vaccine injury) causes MMR avoidance (due to parental concern that vaccines may cause autism. Mere indicators suggestive of autism, not enough for autism diagnosis, will also impact parental vaccination decisions)

With the available data, it is not possible to determine whether websites that publish pages containing highly engaged content are also active on social media and are equally engaged on those platforms. However, a crude comparison between the level of engagement of selected websites (that is, for those that contained more than 3 web pages in the sample) and the activity of their Twitter and Facebook accounts, does show that those websites that are the most engaged are also those with the lowest levels of activity on Twitter and Facebook (see Figure 31 and Figure 32). This may suggest that members of the anti-vaccination movement select different online media depending on how closely they engage with scientific articles. Put differently, social media platforms such as Twitter and Facebook may be selected by those in the anti-vaccination movement who do not wish to engage closely with the scientific content but who nevertheless seek to leverage science to further their cause.

Figure 31: Level of engagement of selected pages from 5 anti-vaccination websites versus activity on Twitter

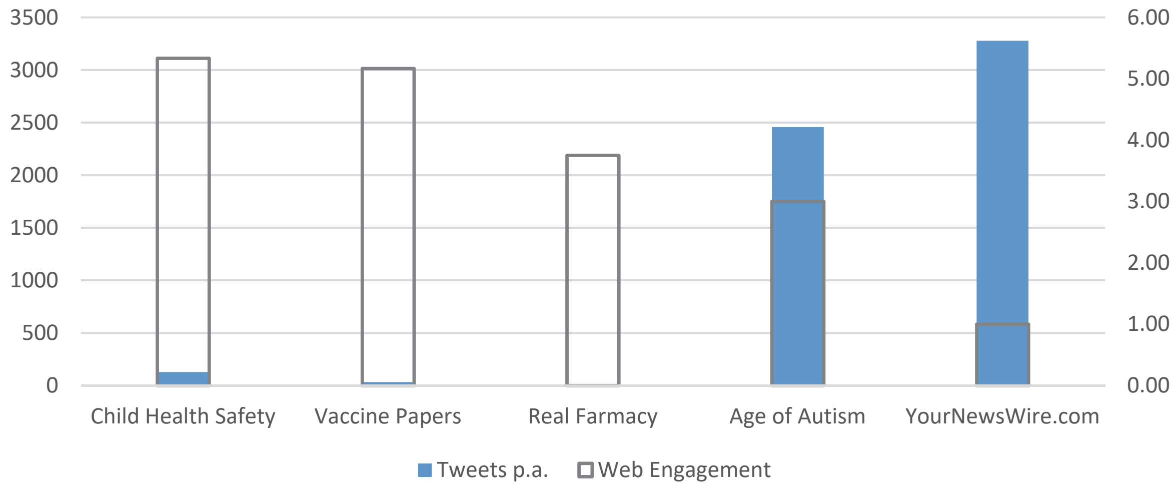
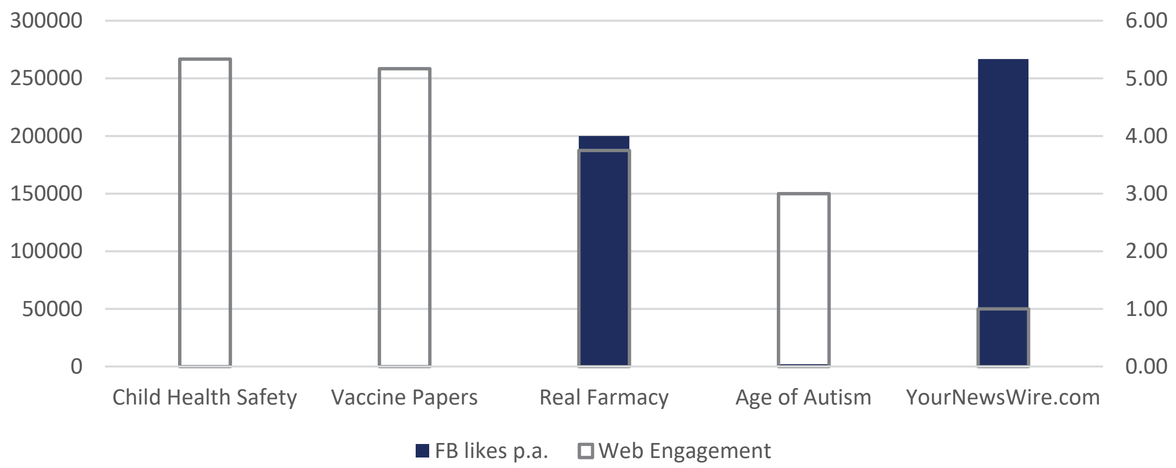


Figure 32: Level of engagement of selected pages from 5 anti-vaccination websites versus activity on Facebook



6.4 Full versus partial access to journal articles

Of the 167 references to open access journal articles in the sample, over half (52.7%) were to the abstracts of pay-walled journal articles (see Table 16). In other words, a certain degree of access is made possible by digitisation and the internet alone, and even without open access, the products of science are available for the creation of new divergent trajectories in the communication of science.

At the same time, however, the anti-vaccination movement was found to have made available online the full-text versions of 12 pay-walled articles. And in the analysis of the use of open data, there was one striking example where a member of the anti-vaccination movement extracted graphs from several journal articles to support his hypothesis of a link between the aluminium adjuvant in vaccines and autism. These examples could point to the fact that full-text open access scientific articles hold additional value when compared to the abstracts of those articles.

Table 16: Open access publication references by the anti-vaccination movement, by publication type as at 22 June 2017 (n=167)

Publication type	Number	%
Abstract	88	52.7
Full-text article	68	40.7
Other	11	6.6
TOTAL	**167**	**100**

6.5 Conclusion

The findings presented in the first part of this chapter show that non-scientists in the form of the anti-vaccination movement are connecting to the products of open science (in this case, open access journal articles) via social media and other Web 2.0 platforms. It was found that open access journal articles on the link between vaccination and autism are mentioned online even though the open datasets for those same articles are not.

At the aggregated level, clear differences were found between two samples when comparing level of attention and the communities of attention. Those articles identified from references made by the anti-vaccination movement were found to have higher level of attention scores and are mentioned by a larger proportion of the public than is the case for articles sourced from research repositories, data journals and bibliometric indexes.

At the disaggregated level, it was found that the anti-vaccination movement constitutes a small and variable proportion of the total number of social media accounts mentioning scientific articles. For one group of open access journal articles, 11% to 18% of Twitter accounts were found to be anti-vaccination; in the case of a second group of open access journal articles, the proportion of anti-vaccination Twitter accounts dropped to 1% to 3%. This indicates an interest in selected articles only. For the journal article with the highest proportion of anti-vaccination Twitter mentions, almost all known Facebook accounts mentioning the article were found to be anti-vaccination.

For those articles with a higher proportion of mentions from anti-vaccination accounts on Twitter, it was found that the activity (proportion of mentions) for those accounts, far exceeded their representation (proportion of unique anti-vaccination Twitter accounts). In other words, anti-vaccination Twitter accounts tweet more than their representation suggests. This indicates that the anti-vaccination movement ‘punches above its weight’ when a scientific article that supports its ideological position is accessible and fed into the flow of information in its social media networks.

High levels of activity are confirmed by the findings of a closer examination of the frequency with which each anti-vaccination Twitter account tweets a mention to articles 2.3, 2.4 or 2.5. Table 17 shows that a large proportion of anti-vaccination accounts (between 33% and 40%) tweet or retweet a mention to an article more than once, and using 10 mentions as being indicative of a high level of activity, several accounts (between 5 and 12) do so 10 or more times.

Table 17: Frequency of mentions by anti-vaccination Twitter accounts

Article	More than 1 mention		More than or equal to 10 mentions	
	Number	%	Number	%
2.3	73	33	12	6
2.4	81	49	17	10
2.5	63	40	5	3

Despite high levels of activity in the social media, the level of engagement by the anti-vaccination movement with open access journal articles on both social media platforms is low. The frequent reposting of content and the relatively small proportion of original content is the main contributor to low levels of engagement on Twitter and on Facebook. On both social media platforms, in addition to frequent reposting, evidence was found of content being modified and blended to reignite levels of attention, if not engagement.

It was found that the low levels of engagement by the anti-vaccination movement with open access journal articles in the social media cannot be generalised to similarly low levels of engagement on the web. Overall, engagement by the anti-vaccination movement tends towards appraisal and application in the case of web pages. Two exceptional cases were found in which the authors of anti-vaccination web pages engage closely with open access journal articles. In these cases, web pages included an interpretative statement, table or graphic pertaining to the open access journal article referenced and, in some cases, included references to and reasoned counter-arguments to pro-vaccination articles, or were followed by a discussion thread consisting of at least a reply from another user and a response from the author of the web page in which content from the article is used to substantiate the author’s position. Findings also suggest that those highly engaged with scientific content on the web are relatively inactive in the social media.

In addition to findings on connections to articles by the anti-vaccination movement as well as the community's levels of activity and engagement in the social media in relation to open access journal articles, it was noted that full access to journal articles is not a precondition for mentions to or the use of those articles in support of the ideologically motivated political agenda of the anti-vaccination movement. In other words, open science per se does not introduce any risks associated with the use of the products of science for political purposes; digitisation and access via the internet are sufficient conditions for doing so.

Regardless of the underlying enablers of access to the products of science, the findings in this chapter, when read collectively, show how the social media as self-referential networks operating outside of the norms of science, are used by a group of non-scientists to amplify its minority position without the need to engage closely with the scientific knowledge at its disposal.

Chapter 7. Intermediation in the anti-vaccination movement's Twitter communication network

The findings in Chapter 6 centered around access to the formal communication of science by a non-scientific public made possible by new and unprecedented levels of access to the products of science in the formative stages of knowledge production. These are new trajectories in the communication of science that are less an outcome of purposive attempts by scientists to deviate from the routine and formal trajectory of communication, and more an outcome of a non-scientific public accessing specialist scientific communications in the form of open access journal articles.

This deviant form of science communication is neither linear nor does it proceed on a one-to-one basis between product and non-scientific reader. Instead, communication is networked, with the products of science being connected to relatively new online communication networks (the social media) such as Facebook, Twitter and the like. In these networks, as Chapter 6 has shown, communication can be amplified by a relatively small but highly active and well-organised social movement.

It would seem reasonable to expect that, based on our understanding of communication networks, there would be a handful of key actors in these movements who play a strategic role in the amplification of messaging extracted from the openly accessible products of science. These actors may mediate communication between science and the movement, between members of the movement, or both. It is to the existence of such intermediaries in a new deviant trajectory in the communication of the science and the role they play in such communication that this chapter turns its attention.

7.1 Centrality in article ego networks

When conducting the data analysis for level of engagement on Twitter (see Chapter 6), the account @LotusOak stood out as a possible nodal actor in the anti-vaccination movement's engagement on Twitter with the scientific articles in the sample. For example, of the tweets that mention article 2.5 in the random sample of anti-vaccination accounts, 55% (50 of 91) were retweets by @LotusOak. The fact that the Twitter account @LotusOak has 139,000 Twitter followers (as at 3 Feb 2018) further contributes to the account's nodal position and its ideal location within the anti-vaccination network to broadcast anti-vaccination content to a large number of Twitter accounts.

In the case of article 2.3, a similar pattern emerges if *all* Twitter accounts that mention the article are analysed. The graph in Figure 33 represents an ego network for article 2.3. The graph was generated using Altmetric data on the links between article 2.3 (the ego) and unique Twitter accounts (the alters). The Twitter accounts (or alters) are arranged according to edge weight, that is, those Twitter accounts closer to the centre have a greater edge weight than those located at the periphery. Edge weight simply equals the number of times the Twitter account has linked to Article 2.3. The 10 Twitter accounts with the highest edge weights are depicted in red. All 10 accounts located at the centre of the ego network are anti-vaccination. The Twitter account @LotusOak has the highest edge weight of 204; this is 6.8 times higher than the next nearest account with an edge weight of 30 (see Table 18).

Figure 33: Connections between Article 2.3 and Twitter accounts that mention the article (n=1,569)

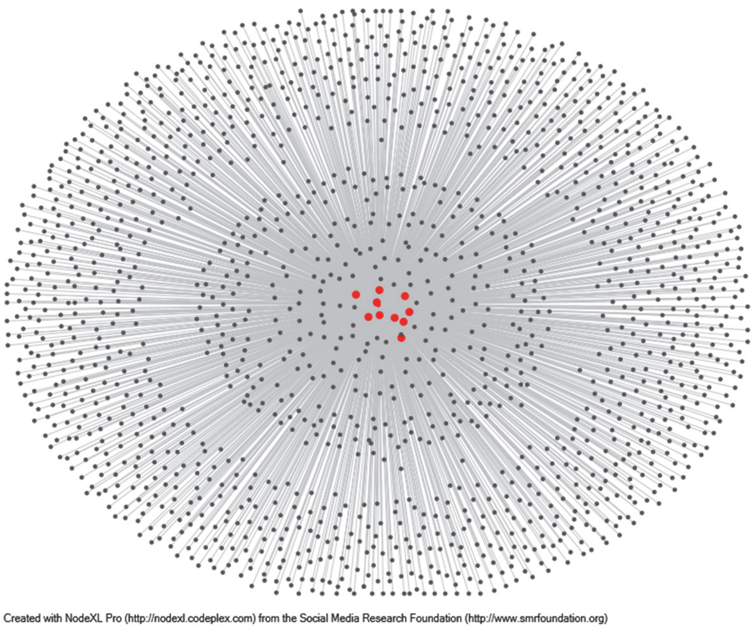


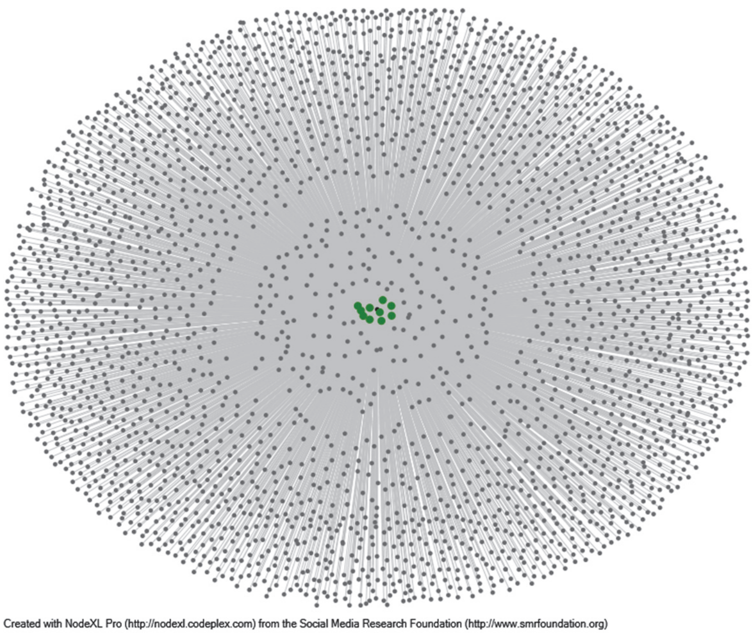
Table 18 shows the comparative edge weights for Article 2.3 and Article 2.2. Article 2.2 was selected for comparison because it is an open access journal article that disproves a causal link between vaccination and autism and, as such, garnered a low degree of attention from the anti-vaccination movement (see Table 13). In addition to the comparatively high number of connections from the Twitter account @LotusOak to Article 2.3, the data in Table 18 reveals that the 10 most connected Twitter accounts for Article 2.2 are all pro-science (indicated in green). The data shows that, comparatively, the pro-scientists are grouped more closely together around Article 2.2 than is the case for the anti-vaccination Twitter accounts' clustering around article 2.3.

Figure 34 illustrates in the form of an ego network the central position of the top ten pro-science accounts (alters) in relation to Article 2.2 (ego).

Table 18: Comparison edge weights for Twitter accounts mentioning two journal articles

	Article 2.2		Article 2.3	
	Account	**Edge weight**	**Account**	**Edge weight**
1	@doritmi	55	@LotusOak	204
2	@VaccineUK	46	@VeritasDolor	30
3	@DocBastard	42	@mission2heal	27
4	@dlmetcalf	36	@Vbalance03	25
5	@GeoffSchuler	35	@andrewmorrisuk	24
6	@TonyBaduy	26	@Sheeple101	19
7	@SecuLawyer	24	@ParentAlerte	14
8	@Takethatdoctors	22	@SNCCLA	13
9	@kath2cats	21	@debnantz	13
10	@TakeThatChem	20	@xileenie	13

Figure 34: Connections between Article 2.2 and Twitter accounts that mention the article (n=2,777)



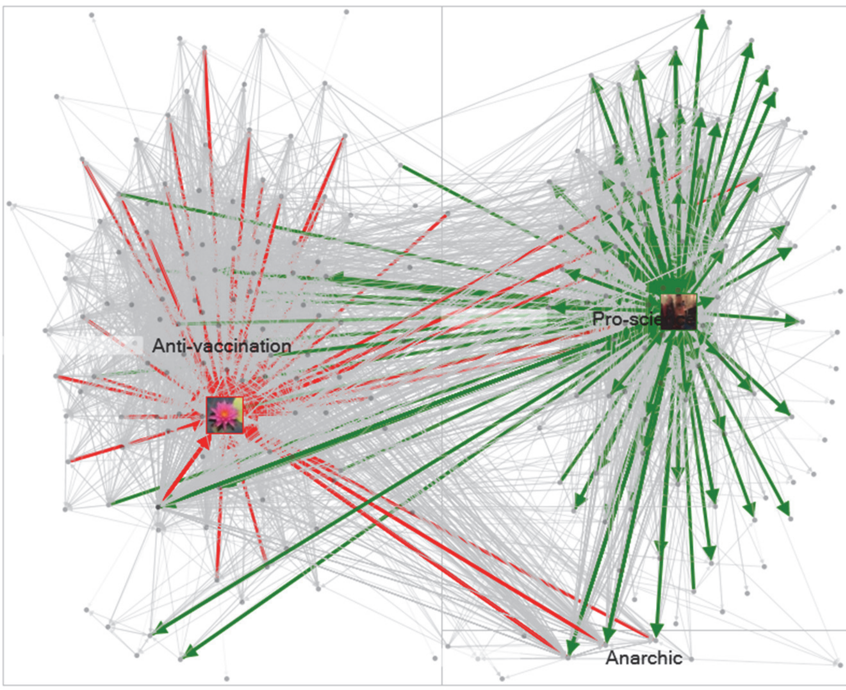
7.2 An anti-vaccination pro-science Twitter account network

Neither of the above ego networks reveals anything about the relationship between Twitter accounts connecting to open access journal articles on the topic of vaccination and autism.

Figure 35 uses two accounts to create such a relational network graph, or sociogram. The accounts used as input data were those identified as being central in the anti-vaccination and pro-science communities resepectively (see Table 18). Figure 35 shows the presence of three distinct groups based on the follower networks of @LotusOak (left-hand side of the graph and red vertices) and @doritmi (right-hand side of the graph and green vertices). The first is a group of tightly connected anti-vaccination Twitter accounts with @LotusOak as one of the accounts occupying a central position in that group. The second group consists of pro-science Twitter accounts with @doritmi as one of the central accounts in that network. A third, much smaller group, can be described as consisting of anarchic Twitter accounts that hold no position regarding vaccination, and seek only to fuel tensions between the bi-partisan groups. The presence of such a group is consistent with research that found that there are Twitter troll accounts whose sole purpose is to create discord in online communication networks (Broniatowski et al. 2018).

The number of inward pointing red vertices reveals that @LotusOak is an account that is followed by many accounts in the network but does not follow many other Twitter accounts; in the case of @doritmi, there is a balance between inward- and outward-pointing vertices indicating that this account has roughly the same number of Twitter followers as followees. This may suggest more of an information exchange role for @doritmi and more of an information broadcasting role for the anti-vaccination Twitter account @LotusOak.

Figure 35: Sociogram for connections between two opposing Twitter accounts and their followers



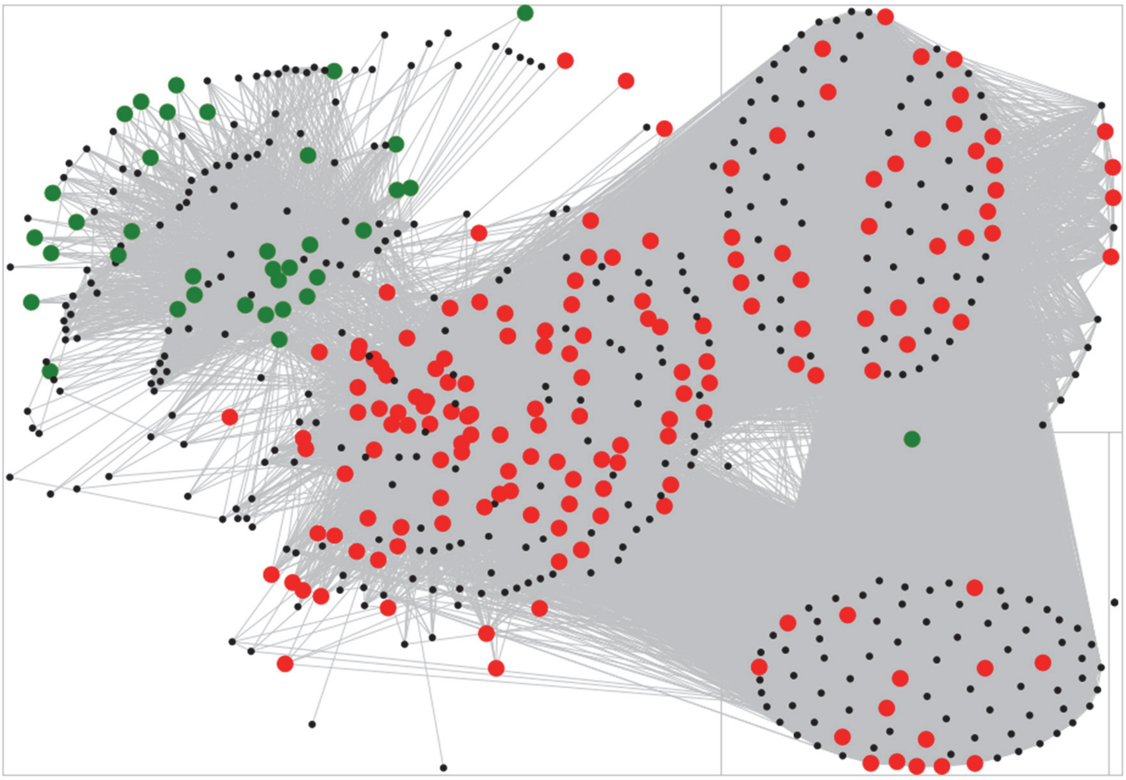
Created with NodeXL Pro (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

7.3 A tweeter coupling network for open access journal articles

The sociogram (Figure 35) is a graphic representation of the relationships or connections between the followers of two Twitter accounts, one anti-vaccination and one pro-science, both of which occupy central positions in their respective online communities. However, the graph does not capture how information extracted from the products of the formal communications of science flows or connects actors in the communication network. In other words, it is agnostic in terms of the content shared between Twitter accounts and, in terms of the interest of this thesis, whether the content exchanged across the connections in the network is sourced from the products of science. To bring the products of science into the communication network as the central structuring determinant of the connections between actors, a different approach is needed.

Figure 36 shows a tweeter coupling network in which each pair of Twitter accounts shares a mention to more than 2 of 113 open access journal articles related to the topic of vaccination and autism. Red discs show known anti-vaccination accounts while green discs show known pro-science Twitter accounts.

Figure 36: Tweeter coupling network (n=12,207)



Created with NodeXL Pro (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

The graph shows several noteworthy characteristics. The first is the clustering of Twitter accounts into three distinct groups. The network graph shows that the clusters exhibit both high degrees of interconnectedness (between clusters) as well as high degrees of intraconnectedness (within clusters).

The two groups on the right-hand side of the graph are composed of anti-vaccination accounts. It is possible that the groups are clustered around one of several core principles that form the basis of the group's resistance to vaccination such as homeopathy or pro-choice or anti-GM foods. A third group (on the left of the graph) consists of both pro-science and anti-vaccination accounts, although there is also a discernible division within this group according to stance: pro-science accounts are clustered in the top left while anti-vaccination accounts are clustered in the right-hand side, closer to the two exclusively anti-vaccination groups.

The two exclusively anti-vaccination groups and the anti-vaccination sub-cluster in the mixed group form a clear triangle. This confirms previous research on networks which suggests that paths of strong ties tend to close into triangles, creating clusters of highly connected groups (Himelboim, Smith, Raine, Shneiderman, & Espina, 2017).

Other notable characteristics of this graph include the fact that the vertices between the three anti-vaccination clusters are demonstrably denser than the links between the anti-vaccination sub-group and the pro-science Twitter accounts in the same group.

While the existence of both mixed and stance-polarised groups is of interest, this network graph created using data on shared connections to the products of open science, also provides important evidence on who is mediating between science and the anti-vaccination movement.

An indicator of intermediation is the centrality of Twitter accounts in the network. Centrality can be measured using three metrics: degree centrality, betweenness centrality and eigenvector centrality. The algorithms for calculating these centrality metrics extend the analysis of network influencers beyond number of tweets or retweets to take into account their position in a network relative to others in the same network, and provides some insights into how information flows in the network.

Degree centrality is a simple count of the number of connections from one node to other nodes in a network. In other words, it indicates how many links a specific node has to other nodes. In-degree centrality measures for each node the number of incoming connections (or followers in the case of Twitter), and out-degree centrality measures the number of outgoing connections (or followees in the case of Twitter). In the tweeter coupling network, neither in- nor outdegree centrality measures are available because the vertices are undirected. In other words, the tweeter coupling data does not capture directionality in the relationship between Twitter accounts; only that they connect to the same journal articles via hyperlinks in their tweets.

It should be noted that the number of connections in the tweeter coupling network of 552 Twitter accounts (nodes) may be artificially inflated as confirmed by the relatively high density score (graph density = 0.163 or 16.3%) for an online social network of this size. This is because the edges in the network were created using data in which every couple was known to connect to more than two journal articles. This high degree of association may have increased the likelihood of connections being formed between couples and, conversely, reduced the number of isolated couples in the network. While an artificially high network density precludes making any deductions in relation to high levels of information flow in the network based on the density of the network, it does nevertheless allow for the comparison of differences in the number of connections between tweeters or groups of tweeters *within* the network.

Table 19 shows those network nodes with the highest degree centrality scores. Noticeable is the fact that all the Twitter accounts with the highest degree centrality scores are anti-vaccination. The account with the highest degree centrality is @itsmepanda1 – 404 other nodes in the networks connect to this anti-vaccination account. The pro-science Twitter account with the highest degree centrality score is @dkegel with a score of 140 followed by @doritmi with a score of 135.

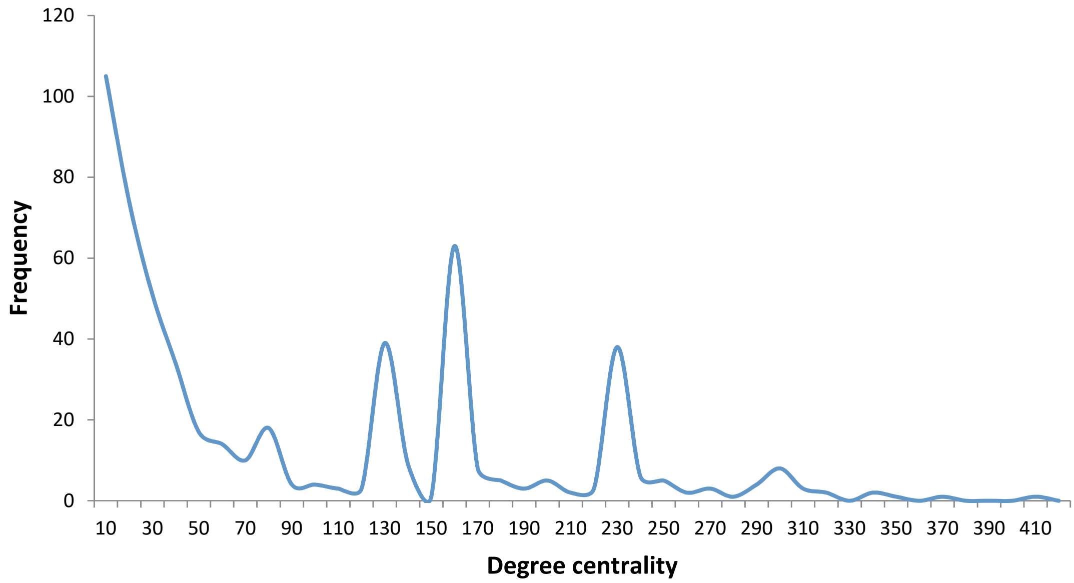
Table 19: Top 20 accounts by degree in the tweeter coupling network

Account	Stance	Degree
@itsmepanda1	Anti-vaccination	404
@LaLaRueFrench75	Anti-vaccination	365
@debnantz	Anti-vaccination	344
@eTweeetz	Anti-vaccination	340
@Biegenzahn	Anti-vaccination	335
@libertylives277	Anti-vaccination	319
@EMcCra2	Unknown	318
@aspiritcan	Anti-vaccination	309
@SNCCLA	Anti-vaccination	304
@VaxChoiceVT	Anti-vaccination	303
@luckykelsey	Anti-vaccination	300
@doctorsensation	Anti-vaccination	299
@Just4TheCause	Anti-vaccination	299
@Stutzy6	Anti-vaccination	297
@mission2heal	Anti-vaccination	296
@xileenie	Anti-vaccination	296
@gramps97	Anti-vaccination	294
@pHarmaNemesis	Anti-vaccination	293
@qtbeauty	Anti-vaccination	289
@nocompulsoryvac	Anti-vaccination	287

If degree centrality is an indicator of an intermediary position in the network, then a distribution of degree centrality scores would show whether there are a typical number of intermediaries in the tweeter coupling network given that the distribution of degree centrality in social networks typically follows a power law distribution (Kahle, 2011). Figure 37 shows that the distribution is typical of a social network – a high number of accounts with low degree centrality scores followed by a rapid drop-off in the number of low degree centrality accounts, and a long tail of accounts with high degree centrality. What is atypical are the spikes in mid-range degree centrality at scores in the following bands: 120-130, 150-160 and 220-230.

A closer examination of the bands in which these spikes occur reveal that all three bands consist of Twitter accounts located in the anti-vaccination clusters, primarily the two exclusively anti-vaccination clusters. The accounts in these bands link to an unusually high number of other accounts in the network. These anomalies in the mid-range degree centrality scores could be interpreted as indicating a network with more intermediary accounts than is normally the case.

Figure 37: Centrality degree distribution for a tweeter coupling network



Eigenvector centrality accounts not only for the node's own degree, but also the degrees of the nodes to which it connects. A node, using this measure, is important because it connects to other important nodes in the network. Eigenvector centrality is therefore an extended form of degree centrality, in which not only the number of connections a node has but also how central those connections are, is calculated. Table 20 shows the top 20 Twitter accounts in the network by eigenvector centrality. As was the case for degree centrality, most of the accounts are anti-vaccination and there is a high degree of similarity between the degree centrality and eigenvector centrality scores of the top 20 Twitter accounts in the tweeter coupling network. In other words, anti-vaccination Twitter accounts are not only central in the network, they also connect most often to other Twitter accounts with a high number of connections.

Table 20: Top 20 accounts by Eigenvector centrality in the tweeter coupling network

Account	Stance	Eigenvector centrality
@itsmepanda1	Anti-vaccination	0.005465
@LaLaRueFrench75	Anti-vaccination	0.005442
@debnantz	Anti-vaccination	0.005383
@Biegenzahn	Anti-vaccination	0.005357
@libertylives277	Anti-vaccination	0.005338
@EMcCra2	Unknown	0.005333
@eTweeetz	Anti-vaccination	0.005325
@aspiritcan	Anti-vaccination	0.005300
@SNCCLA	Anti-vaccination	0.005293
@VaxChoiceVT	Anti-vaccination	0.005289
@Stutzy6	Anti-vaccination	0.005283
@luckykelsey	Anti-vaccination	0.005280
@xileenie	Anti-vaccination	0.005276
@gramps97	Anti-vaccination	0.005275
@doctorsensation	Anti-vaccination	0.005274
@Just4TheCause	Anti-vaccination	0.005269
@nocompulsoryvac	Anti-vaccination	0.005257
@pHarmaNemesis	Anti-vaccination	0.005256
@qtbeauty	Anti-vaccination	0.005248
@ParentAlerte	Anti-vaccination	0.005235

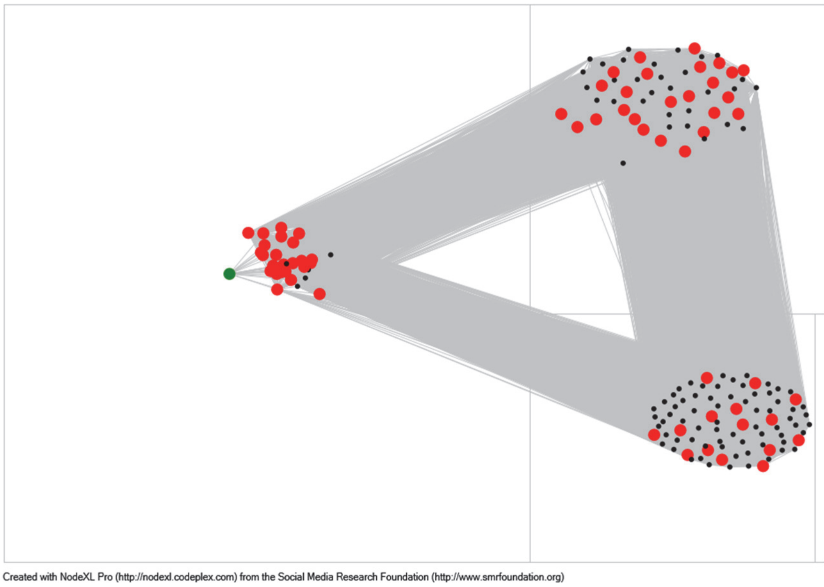
Betweenness centrality indicates how important each node is in providing a bridge between different parts of the network. This bridge represents the extent to which a node is part of transactions among other nodes and is therefore an important indicator of which nodes facilitate the flow of information between other nodes in the network.

Table 21 shows the 20 Twitter accounts with the highest betweenness centrality. Given the structure of the network, it is not surprising to find pro-science accounts in the top 20 as one would expect there to be pro-science accounts that bridge between their own sub-group and the anti-vaccination sub-clusters in the mixed group. The bridging position of the pro-science Twitter account @dkegel is illustrated in Figure 38.

Table 21: Top 20 accounts by betweenness centrality in the tweeter coupling network

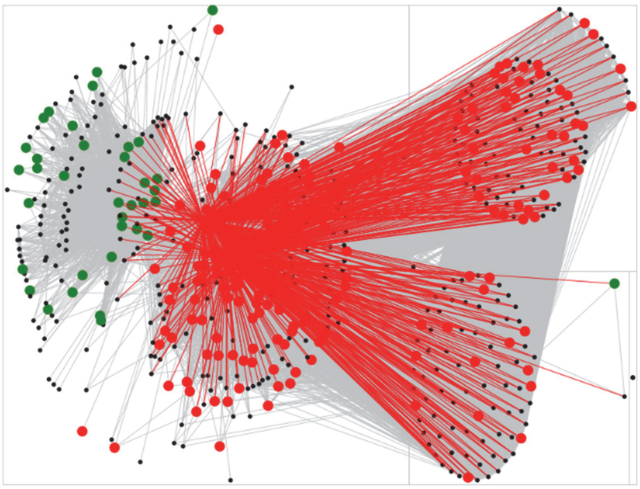
Account	Stance	Betweenness centrality
@itsmepanda1	Anti-vaccination	18087.753
@eTweeetz	Anti-vaccination	10467.082
@dkegel	Pro-science	8694.468
@doritmi	Pro-science	7540.706
@PaulWhiteleyPhD	Unknown	7082.546
@Anwar_Hashem	Pro-science	6530.001
@Biegenzahn	Anti-vaccination	6398.226
@LaLaRueFrench75	Anti-vaccination	5834.115
@mission2heal	Anti-vaccination	4889.009
@debnantz	Anti-vaccination	4873.072
@VaxChoiceVT	Anti-vaccination	3271.135
@libertylives277	Anti-vaccination	3106.091
@doctorsensation	Anti-vaccination	2942.356
@miles_wilma	Unknown	2791.991
@BlaKiss4	Pro-science	2717.097
@justthevax	Unknown	2694.971
@EMcCra2	Unknown	2209.234
@Just4TheCause	Anti-vaccination	2028.555
@DocBastard	Pro-science	1995.022
@kath2cats	Pro-science	1968.830

Figure 38: Illustration of the bridging role played by a pro-science Twitter account



The bridging role played by the top two anti-vaccination Twitter accounts – @itsmepanda1 and @eTweetz – is shown in Figure 39.

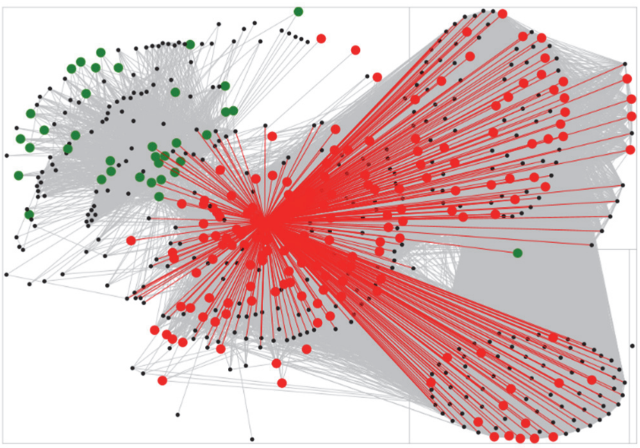
Figure 39: Illustration of the bridging role played by two anti-vaccination Twitter accounts



Created with NodeXL Pro (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

The figures below show two different patterns of intermediation in the anti-vaccination movement on Twitter. Figure 40 shows that the anti-vaccination Twitter account @itsmepanda1 connects the three anti-vaccination clusters in the network and also links the anti-vaccination Twitter accounts in the network with the pro-science cluster. This suggests that this account mediates *between* sub-networks within an aligned group, as well as between non-aligned sub-networks.

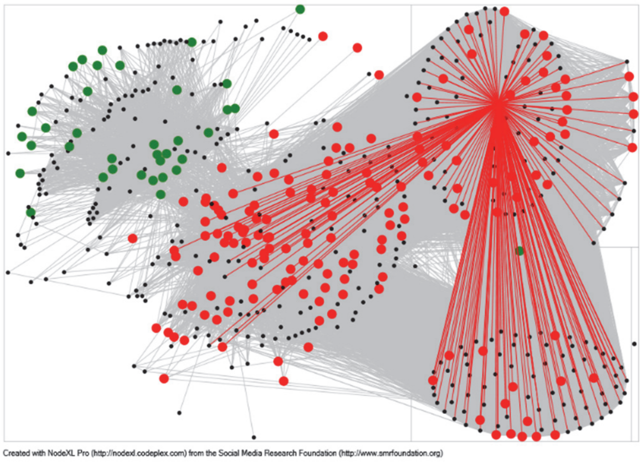
Figure 40: Example of an intermediary Twitter account in a tweeter coupling network: @itsmepanda1



Created with NodeXL Pro (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

In Figure 41, however, the Twitter account @LotusOak only connects the three anti-vaccination clusters; and there is also no evidence of shared mentions on Twitter between @LotusOak and the pro-science accounts in the network. This suggests that intermediation by this anti-vaccination Twitter account is limited to intermediation *within* the anti-vaccination movement.

Figure 41: Example of an intermediary Twitter account in a tweeter coupling network: @LotusOak



It is perhaps worth remembering the uniqueness of the tweeter coupling network. Unlike many other such social media networks generated for analysis, this network has been generated based on shared mentions of a specific *product* of science: the journal article. In addition, it is a scientific product that is more readily accessible in full by non-scientists because it has been published open access. The network relies on mentions made by account pairs to more than two open access journal articles on the topic of vaccination and autism. In some instances, a pair of Twitter accounts will link to more than three articles. The more articles each pair links to is an indicator of their shared interest in the same scientific information, a conceptual connection based on the fact that they chose to mention specific articles from the available 113 journal articles, and of their level of activity in injecting scientific information into networked information flows.

The minimum number of shared articles is three and the maximum number of shared mentions was found to be 13. Figure 42 shows that the distribution of coupled mentions follows a power law distribution, one that is common in social media networks (Kahle, 2011). Most coupled Twitter accounts mention only three of the 113 open access journal articles while very few couples mention five or more articles. This suggests that mentions of open access journal articles on the topic of vaccination and autism is highly selective.

Figure 42: Distribution of number of open access journal articles mentioned on Twitter by coupled accounts (n=24,817)

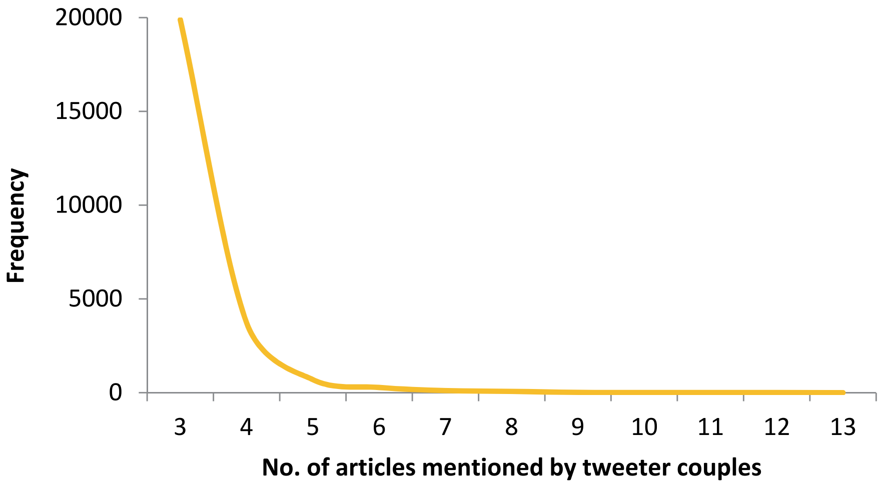
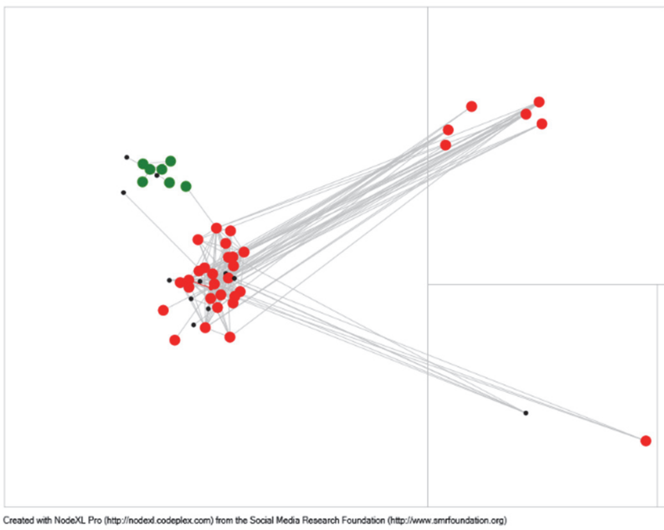


Figure 43 shows only those Twitter account couples that mentioned seven or more open access journal articles (seven being roughly the mid-way point between the minimum and maximum number of mentions). The graph shows that more active tweeter couples are known to be anti-vaccination; that there are tweeter couples that consist of accounts in different anti-vaccination sub-groups; and that the smaller group of pro-science couples are disconnected from the anti-vaccination clusters with the exception of one couple made up of one pro-science account and one anti-vaccination account. This latter finding suggests the anti-vaccination and pro-science clusters are mentioning (linking to) different scientific articles; in other words, the highly active couples in the clusters with opposing views are selecting and mentioning different journal articles.

Figure 43: Tweeter coupling network for tweeter couples mentioning more than 7 open access journal articles



7.4 Conclusion

The two anti-vaccination only groups are arranged into sub-networks that reveal a typical hub-and-spoke arrangement characteristic of broadcast networks (Himelboim, Smith, Raine, Shneiderman, & Espina, 2017). These groups are centred around one or two Twitter accounts, and the other accounts repeat much of what the hub account tweets. Intermediation in these networks is therefore evident but only at intra-(sub)network level.

However, unlike typical hub and spoke networks, the density of connections between the three anti-vaccination clusters is relatively high. This indicates much higher levels of information exchange between clusters ostensibly because of the high degree of cohesion attributable to a shared interest in anti-vaccination advocacy.

The mixed group consisting of both anti-vaccination and pro-science accounts, shows a different type of network structure; one that can be described as being polarised (Himelboim, Smith, Raine, Shneiderman, & Espina, 2017). In this part of the network, connections between the pro-science and the nearest anti-vaccination cluster are less dense while in-cluster density remains high. Here anti-vaccination intermediaries play two roles. They bridge between the anti-vaccination and pro-science clusters, and they interconnect the pro-science cluster with the other two anti-vaccination clusters. Given the divisiveness of the anti-vaccination debate, it is unlikely that the bridging intermediaries play a conciliatory role; instead these are bridges where contestation and further polarisation with the support of homogenous networks on either side of the bridge foment.

In sum, the findings of this chapter indicate the presence of intermediaries in the flow of scientific information in the network. There are different types of intermediaries with different functions in relation to the distribution of information in the network. While there are sub-clusters within the anti-vaccination movement, they remain highly connected to one another and disconnected from the pro-science cluster in relative terms. The distribution of scientific information within the network is selective, further reinforcing the existence of echo chambers despite the high degree of inter-cluster interconnectivity and relatively high degree of centrality in the anti-vaccination clusters.

Chapter 8. Discussion

Chapters 5, 6 and 7 presented findings on whether open science is creating new potentials for non-scientific publics to access and use the products of science. The chapters did not offer much by way of explanation or venture into the realm of speculation; nor was effort expended on integrating findings into a coherent narrative. This chapter seeks to remedy these self-imposed limitations by discussing the findings from all three chapters in relation to the key question posed by this thesis: What are the potentials of open science in the communication of science?

The discussion proceeds according to emergent themes pertinent to the communication of open science with non-scientific publics. This approach allows for a more integrated discussion of the findings while remaining mindful of the sub-questions that this thesis set out to answer, each of which is aimed at revealing more fine-grained insights on the communication potentials made possible by open science:

- Are non-scientists accessing the products of open science?
- Are non-scientists engaging with the products of open science?
- Who are the non-scientists mediating connections to the products of open science?

The final sub-question posed – how science communication can respond to the potentials of open science – is dealt with in Chapter 9.

8.1 Expertise as a barrier to use

This thesis considered open access to two types of scientific product: research data and journal articles. It did not, however, differentiate between the products in terms of the usability of those products by non-scientific publics.

The findings presented in Chapter 5 show that a politically active non-scientific community, the global anti-vaccination movement, is not accessing open research data on the purported link between vaccination and autism. At most, there is evidence of the use of non-textual scientific information in the form of numerical tables and graphs, neither of which are indicative of the use of raw research data. At the same time, there is evidence that the same non-scientific community is both accessing and using open access journal articles (many of which are related to the same scientific study as the open research dataset included in the analysis).

This finding suggests that open research data is somehow different to open access research articles in a way that makes open research data less attractive to non-scientific publics. This difference could be accounted for by the differences in expert knowledge required to read data and text. Because of the level of expertise (knowledge) and skill (training) required to read data, open research data is most likely 'immune' to use by non-scientific publics because of its inherently technical and complex nature. Non-scientific publics are therefore not able to interpret the raw open research data to make it usable for political or other purposes.[54]

That some evidence was found of the use of tables and graphs – that is, data that has been interpreted by an expert and presented in aggregated and/or graphic formats – attests to the likelihood that a lack of expertise to interpret open research data presents a barrier to its reuse by non-scientific publics. It also points to the dependency of non-expert publics (and scientists) on intermediaries to make research data interpretable and, by implication, usable.

Whether the expertise barrier is a constant limiting factor in the use of open research data by non-scientific publics across all scientific fields would need to be verified. It could well be the case that the raw, empirical data on genetics and microbiology that often form the basis of studies on vaccination and its side-effects are particularly immune to use by non-scientific publics whereas research data from studies in other contested areas such as hydraulic fracturing or mass immigration may require lower levels of expertise.

A context-determined ascription of usability and usefulness is an additional factor which may influence the use of open research data: 'data sharing obstacles may be understood in terms of a policy–practice divide, whereby data sharing policies fail to recognise data sharing as a relational practice [...R]aw data have limited scientific value when taken out of their contexts of production, signalling that meaning is not inherent to the data but to the data-in-context' (Mauthner & Parry, 2013, pp. 47, 58).

[54] It is equally true that the expertise required to interpret data will also preclude other scientists from being able to use open research data because of field- or discipline-specific knowledge required to make sense of the data (Mauthner & Parry, 2013). The absence of established, shared standards for descriptive metadata and for data formatting, place additional limitations on the use of open research data in interdisciplinary science or when working with large, complex datasets (Tenopir, et al., 2015).

Whether the expertise barrier becomes more porous over time also needs to be considered. As the anti-vaccination movement gains in notoriety and is able to hold the attention of a broad spectrum of a global audience through the use of social and other media, it may begin to attract the interest and support of those who do possess the expertise and skills to interpret open research data.[55] If this 'floor-crossing' scenario does indeed play out, further questions are worth consideration. First, what would be the incentives for anti-vaccination data experts to engage with open research data if the strength and efficacy of the movement depends less on the veracity of the science disproving the link between vaccinations and autism, and more on the headline findings and their value in amplifying the anti-vaccination message in global communication networks? In other words, if, as the findings of this study suggest, engagement is likely to remain shallow and if such engagement is sufficient to attract attention in social media networks (and possibly influence decision-making outside of those networks), then open research data may be of little interest to minority groups such as the anti-vaccination movement.

Second, would those scientists who support the minority anti-vaccination position from within science abandon, as Mitroff (1974) suggests, the norms of science in favour of particularism, individualism, interestedness and organised dogmatism? And third, if, as Mitroff (1974) further suggests, 'ill-defined problems' are more likely to activate the counter-norms of science, then to what extent are non-scientific publics able to destabilise science, that is, contribute to the manufacturing of ill-defined problems as science becomes more open, more democratised, and as global communication networks increasingly make possible such disruption?

8.2 Openness and access within and between

'It is foolish to feed an ass lettuces when thistles suffice him,' writes Roger Bacon in his *Opus Majus*, citing Aristotle to invoke a moral obligation on the part of the scientist to be circumspect when dealing with the secrets of nature (Eamon, 1985, p. 325). From ancient history to the Middle Ages, the communication of science was reserved for selected experts but kept from epistemological, innovative and even political competitors as well as from non-experts variously described as the 'vulgar', the 'rank and file', and the 'multitude'. By the time of the 17th century, however, with the emergence of new communication technologies such as the printing press, legal mechanisms for protecting the interests of discoverers, and the emergence of institutions for the promotion of

[55] See section 8.5 below for a more comprehensive examination on the attraction held by highly visible minority movements for experts, including medical practitioners and scientists.

scientific activity, there was widespread realisation that the stagnation of knowledge was attributable to the non-public nature of science (Eamon, 1985). Along with this realisation came a new willingness to share discoveries and new knowledge *within* the growing community of European scientists (Eamon, 1985; Fyfe, McDougall-Waters, & Moxham, 2015), and to enshrine in science its public and communal character (Merton, 1973). More recently, as illustrated in Chapter 2 of this thesis, social pressures stemming from, amongst others, a deterioration in the trust of public institutions, has led to calls for greater transparency and accountability of public institutions, including producers of scientific knowledge such as universities and publicly-funded research institutes. The outcome has been a gradual opening of science and its products to the public. While such openness holds benefits for sharing knowledge *within* science, it opens up new channels of exchange *between* science and the public.

This study posited that open science in general and access to open research data and open access journal articles by ideologically-motived publics in particular, could catalyse new potentials that may either benefit or pose risks to society. In the case of the anti-vaccination movement, risk presents as a more likely potential: access to and use of scientific data and publications could be used as powerful evidence to convince those who are undecided about vaccination not to vaccinate; to mobilise sufficient support to petition successfully for amendments to health policy; or, at the very least, to increase the number of vaccine refusers in the broader population thereby compromising herd immunity to highly infectious diseases.

However, from the findings presented in the earlier chapters, it cannot be said unequivocally that it is open science alone that introduces such risk. An examination of all the references to scientific articles – that is, both open access and pay-walled articles – showed that in 53% of cases only the abstract of the article was accessible and referenced by the anti-vaccination movement, while in 41% of cases the reference was to a full-text, open access article. The associated risk of mis-use is therefore not a factor of openness per se but of access, even if that access is limited to the component parts of scientific products (such as the title or abstract of a journal article). In other words, without open access, ideologically-motivated groups are still able to access sufficient content from the products of science, including journal articles, that play a contributing role in the scientific research process that proceeds from hypothesis to fact, from truth claim to settled truth.

There was evidence, however, of actors in the anti-vaccination movement accessing pay-walled articles and pirate archiving those articles to afford access to the entire anti-vaccination movement. There was also evidence of retracted journal articles and scientific reports being archived online by the anti-vaccination movement. In these cases, the anti-vaccination movement is effectively interfering with the scientific process of self-regulated, organised scepticism by withdrawing contested and rejected truth-claims from the scientific process, and preserving those claims outside of science so that they may continue to play a role as weaponised 'facts' in an ideologically motivated battle against the consensus position in science.

It should be added that the pirate archiving of journal articles was undertaken in the main by a specific group within the anti-vaccination movement, one that counter to the trend, was found to be highly engaged with the formal publications of science. Thus, while their attempts at preservation could at one level serve the advocacy efforts of the broader movement, it is also possible that archived articles are being kept in circulation outside of the scientific community to allow the highly engaged sub-group to draw on the articles in their ongoing interrogation of vaccine science.

While the focus of this study was not on the effects of open science on the scientific community, certain insights emerged that are relevant to a discussion on openness in science. For example, the findings in Chapter 5 show that most of the research data repositories relevant to the study of vaccination and its side-effects are described as 'open', but that they fail to meet the criteria for openness as per the FAIR principles (Force11, 2017). In fact, most of the repositories control access to the data.

It is possible that access to the research data is restricted by virtue of the fact that the data contain sensitive, personal information. Alternatively, access may be restricted on the basis of excluding interested non-scientific publics. The findings show that, in some cases, access control is, in fact, in place due to the presence of confidential patient data. However, in several other cases where the data has been anonymised, access controls remain in place (examples include IAN, ASD-UK, ABIDE and NIMH). This precludes non-scientists from accessing and using the data, and suggests that openness is being interpreted by scientists in this field of study as allowing primarily for broader access to research data *within* the scientific community.

In addition to ubiquitous access to research data, of which there are some examples, there is therefore a parallel semi-open channel in the formal science communication continuum. What was previously a honey-comb arrangement within and between academic disciplines has evolved, in the case of autism research, into a more porous structure that enables higher degrees of access and sharing across and between scientists. This finding resonates with emerging developments in the broader open data movement. There is a growing acknowledgement that publishing open data, and particularly open government and open corporate data, is unlikely to become institutionalised practice for all types of data, and that advocating instead for data sharing in these instances may be a more pragmatic approach.[56]

To conclude, one of the central assumptions of this thesis was that open science necessarily introduces new potentials in the communication of science, potentials which may carry unknown or unacknowledged risk to science and society. Ideologically-motivated groups such as the anti-vaccination movement, with increased access to the products of science, could deploy the authority of science to bolster vaccine refusal in the broader public.

The reality, it seems, is different. First, open science creates opportunities for increased sharing within science and, while not the focus of this research, some evidence was found of this taking place among scientists working on autism research. Second, the risks introduced by unintentionally providing access to ideologically-motivated groups are not only the consequence of open science but of prior enabling conditions including the digitisation of scientific products, and the rise of the internet and of the world wide web creating a world in which everyone and everything is increasingly connected to everyone and everything else. Third, the risks are exacerbated because most members of the anti-vaccination movement do not need to engage deeply with the products of science to support the spread of its anti-vaccination message. The following section explores in greater detail how information from the products of science is used by the anti-vaccination movement in its online communications in such a manner so as to leverage the effects of networked communication.

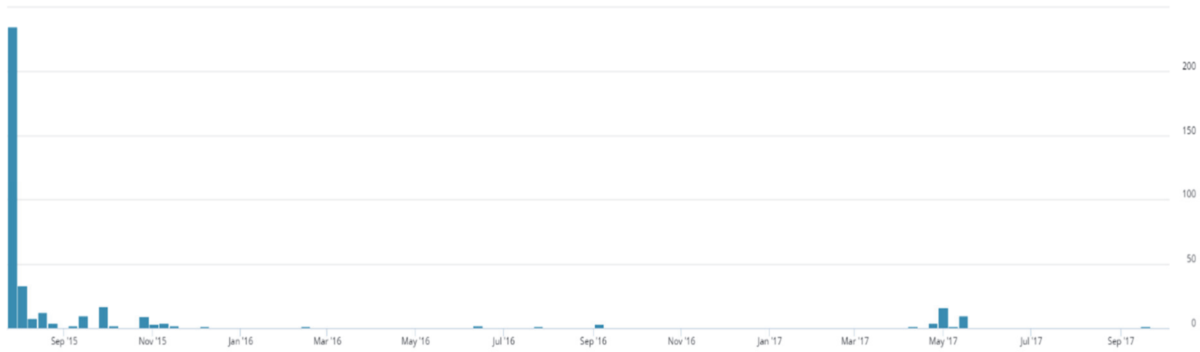
8.3 The production of uncertainty

Referring to the journal article 1.1 ('Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens'), one tweet reads: 'Expect this study to become the newest ammunition for anti-vaccination movement'. This prediction is confirmed to the extent that 36 of 310 (11.6%) tweets mentioning article 1.1 were from the anti-vaccination movement, and this proportion is comparable to the proportion of mentions to other journal articles that indicate support for vaccine refusal. However, what is surprising is that attention from the anti-vaccination movement spiked in early May 2017 when @VaxCalc tweeted 'Australian Pertussis epidemic: bacteria is evolving, becoming more

[56] Compare, for example the Open Data Institute's post 'Data sharing is not open data' (Tennison, 2014) in March 2014 with a post by the same organisation two years later titled 'Widening the debate on data sharing' (Wells, 2016). Also see the GovLab's Data Collaborates Project (GovLab, n.d.) and its explicit choice of 'data' instead of 'open data' in the project title and the project tag line: 'creating public value by exchanging data'.

lethal, evading vaccine (technical paper) http://b.autovist.com/ebfc0d35' and @4health replied to the tweet as follows: '@VaxCalc Imperfect vaccines an [sic] lead to more virulent strains https://t.co/d5IMewqmYi' (see Figure 44). This is just less than two years after the first tweets to the article were posted on 27 July 2015. However, since 2017, there are no mentions of article 1.1 by the anti-vaccination movement on Twitter.

Figure 44: Twitter activity for the journal article 'Imperfect vaccination can enhance the transmission of highly virulent pathogens' (article 1.1)



It is not clear why activity declined post-May 2017. One possible explanation is the limited influence of the Twitter accounts @VaxCalc and @4health in the anti-vaccination Twitter network, and their limited ability to stoke further interest in the community of anti-vaccination Twitter users. In other words, continuous circulation of anti-vaccination information is dependent on the broadcasting of information by centrally located, highly connected community members (see Section 8.7). Kumar et al. (2018) have shown that highly active members in online communities are more likely to initiate interaction and conflict with other communities. However, 'while these interactions are initiated by the highly active users of the source community, the attackers and defenders who actually get mobilized to participate in the negative mobilization are much less active than them' (Kumar, Hamilton, Leskovec, & Jurafsky, 2018, p. 5). This suggests that those highly active anti-vaccination Twitter accounts are not only more likely to instigate interaction with the online pro-science community, but they play an important role in mobilising less active members who, by taking up the cause, further boost the information flows of the anti-vaccination movement in the broader social media network.

A second explanation for the withering of interest in article 1.1 post-2017 is that activity levels in social networks related to the mentioning of scientific products is influenced by content being broadcast in other communication networks. For example, a high-profile court case on compulsory vaccination, or the appearance of a pro- or anti-vaccination advocate in the traditional mass media, may trigger a resurgent interest and an increase in activity in social media networks (Tucker, et al., 2018; Zannettou, et al., 2017).

The findings on the number of mentions to specific articles and those from the tweeter coupling network, suggest that network actors are selective in terms of the scientific content injected into their information flows. A high degree of variance was found in terms of the proportion of anti-vaccination accounts mentioning scientific articles on Twitter. Articles accessed fell into two broad groups: one group of 5 articles (1.1, 1.5, 2.3, 2.4 and 2.5) for which mentions by anti-vaccination Twitter accounts were between 11% and 18% and a second group of 5 articles (1.2, 1.3, 1.4, 2.1 and 2.2) for which fewer than 2% of all mentions originated from anti-vaccination accounts.

In selecting the 10 articles for analysis, no attention was paid to how the stance of the article may be interpreted by readers based on its title or its findings (as described in the article abstract); given that all articles were in one way or another relevant to the vaccines–autism topic, the sole criterion for selection was the article's level of attention in online communication networks.

Table 22 shows that there is a relationship between the proportion of anti-vaccination accounts mentioning an article and the indicative stance of the article vis-à-vis vaccination. Unsurprisingly, those articles whose titles and findings are clearly supportive of an anti-vaccination stance are more likely to be mentioned by the anti-vaccination movement than those articles that provide no support or contradict an anti-vaccination stance. This relationship between the title, the findings and likelihood of being mentioned by the anti-vaccination movement on Twitter occurs despite any cautionary statements by the authors of the journal article that their findings require further research to be confirmed.

Table 22: Vaccination stance of 10 open access journal articles and proportion of Twitter anti-vaccination accounts connecting to those article

Article ref.	Indicative stance: Title	Indicative stance: Findings	% of anti-vaccination Twitter accounts that mention the article
1.1	**ANTI-VAC**	**ANTI-VAC**	**11.6**
1.2	**NEUTRAL**	**PRO-VAC**	0.8
1.3	**PRO-VAC**	**PRO-VAC**	1.7
1.4	**NEUTRAL**	**NEUTRAL**	0.0
1.5	**ANTI-VAC**	**ANTI-VAC**	**16.0**
2.1	**NEUTRAL**	**PRO-VAC**	1.1
2.2	**PRO-VAC**	**PRO-VAC**	1.4
2.3	**ANTI-VAC**	**ANTI-VAC**	**13.9**
2.4	**ANTI-VAC**	**ANTI-VAC**	**17.8**
2.5	**ANTI-VAC**	**ANTI-VAC**	**11.2**

See Table 28 in the Appendix for the textual analysis done to determine stance.

The point to note here is (1) that the anti-vaccination movement is selective in terms of which scientific products it refers to in its online communications, and (2) that the products accessed are those produced early in the science communication process (the intra- and interspecialist stages in the formal science communication continuum) at a stage when truth claims are still contested and in flux. This finding confirms the observation that publics with a limited understanding of how science works – in particular, that science is iterative and self-correcting – may select and exaggerate the findings of individual studies (Kahan, Scheufele, & Jamieson, 2017) and supports the large body of work in the science communication literature on ‘phenomena of selection’ (Akin & Landrum, 2017, p. 455). The anti-vaccination movement, rather than being made to wait for settled truth claims to emerge at the end of the hierarchically structured science communication continuum, access via increasingly networked communication ‘unsettled’, single-study truth claims and interpret them as universal truth. These ‘scientific truths’ hold value for the anti-vaccination movement because they confer legitimacy on its cause in the eyes of other non-scientific communication networks.

According to Castells (1996), value is what the network determines it to be. In the case of the global anti-vaccination network, which could be situated within the larger global anti-establishment communication network, only that information that supports the ideology/beliefs of the network holds currency and is therefore worth exchanging. Besi et al. (2014) find that Facebook users prominently interacting with alternative information sources – i.e. are more exposed to unsubstantiated claims – are more prone to interact with intentional and parodistic false claims. In the case of scientific information, the findings of this research show that those scientific articles that articulate a causal relationship between vaccinations and adverse health or that express doubt about the efficacy and safety of vaccines, gain currency in the network. Conversely, scientific articles that disprove the dangers of vaccination hold no value and therefore do not circulate within the communication network of the anti-vaccination movement.

While the exchange of selected scientific information on the effects of vaccination serve to reinforce the belief systems of the anti-vaccination movement, they have the opposite effect on other networked communities who are both present in social media networks and therefore inevitably connected to the online anti-vaccination movement. The selective harvesting of information from open science and the communication of that information in social media networks, produces uncertainty in other online communities, even when there is consensus within the scientific community, as is the case for vaccine safety. The production of uncertainty can therefore be understood as an attack on the information flows of other networks and is aimed at destabilising trust in the information that circulates in those networks.

The production of uncertainty in non-scientific networks is particularly worrisome in relation to health choices because non-scientific publics follow a different network logic than those of scientists. In science, following the programmatic logic of the global science network to establish truths about the world around us, the unquestioned normative position in the face of uncertainty is to suspend all truth-claims until such time as sufficient evidence is available and all the available evidence has been considered and vetted by peers. In social communication networks, such norms do not apply. The program of social networks is attention-seeking because human identity is increasingly defined in the space of information flows (Castells, 1996, 2004, 2009; Stalder, 2006) and attention-dependent online advertising is the source of income for the owners of social media platforms (Wu, 2016). As such, communication in online social networks seeks to attract a critical mass of attention and neither ideologically-motivated communities nor the owners of social media platforms, both of whom are locked into attention-seeking behaviour, derive value from adhering to the norms designed to establish truth.

The situation is further complicated by the appeal of uncertainty for other, more traditional media such as the print and broadcast media. Scientific uncertainty, when elevated to levels of public panic or crisis, is certainly newsworthy and are sure to be reported by the traditional media. But uncertainty also allows the traditional media to pivot or change direction to prevent loss of attention as new scientific solutions are presented in attempts to reduce uncertainty; solutions which are themselves often not yet accepted by the scientific community (Weingart, Engels, & Pansegrau, 2000). Trending topics on Twitter (that is, those dominant topics that hold the attention of the largest number of Twitter users) are largely picked up from traditional news sources and are amplified by repeated retweets on Twitter (Asur, Huberman, Szabo, & Wang, 2011). Uncertainty, then, manufactured in social media networks, and picked up by traditional media, creates a vicious cycle of uncertainty in networks of communication. As the Cultural Cognition Project reports:

> Many of those who have undertaken to inform the public are themselves deeply confused. The primary source of information on vaccine risks consists in ad hoc risk communication by journalists, advocacy groups, and even some individual public health professionals who consistently mischaracterise the nature, extent, and sources of public concerns over vaccines. Rather than assure the public, this empirically uninformed style of risk communication itself poses a risk of exciting

group animosities and other dynamics that could magnify unfounded fears of vaccines. (Cultural Cognition Project, no date)

8.4 The amplification of uncertainty

In the theory on communication, amplification is the process of intensifying or attenuating signals during the transmission of information (Kasperson, et al., 1988). How the amplification of information emanating from scientific sources takes place in the online communications of the anti-vaccination movement is explored in the sections that follow.

8.4.1 High levels of activity

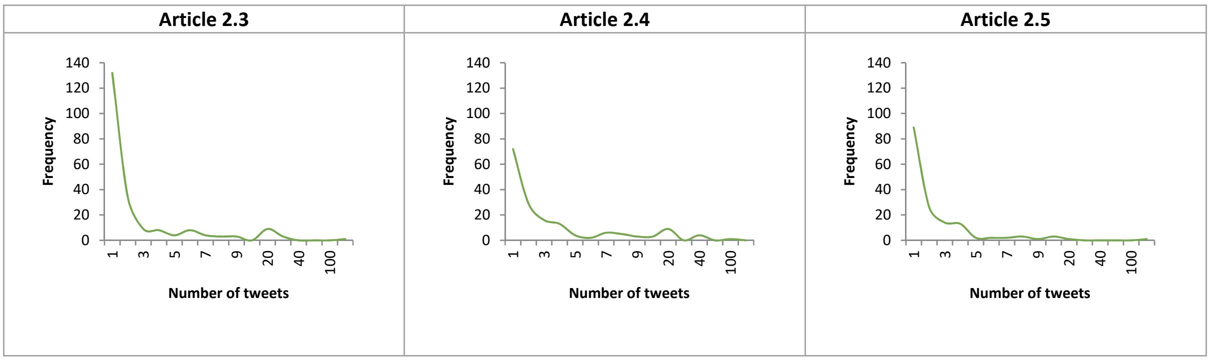
In general, the most influential tweeters are more active than the less followed tweeters although it is not clear whether these individuals are widely followed due to their high posting volume, or whether they are prolific because their audience is sufficiently large (or appreciative) (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013).

High levels of activity in the social media also increases the probability of content consistent with a particular stance appearing at the top of the content feeds of those who follow highly active accounts. Being listed at the top of content feeds, in turn, increases the chances of the content being shared with others in the social media network (Lerman & Hogg, 2014).

The findings of this research showed that the anti-vaccination movement ‘punches above its weight’ when a scientific article that supports its ideological position is accessible and inserted into its networked information flows on Twitter (see Figure 10). In other words, the proportion of tweets by the anti-vaccination movement for selected open access journal articles is higher than the proportion of anti-vaccination Twitter accounts that mention the same open access journal article. It is the structure and affordances of networked communication that makes possible such ‘network effects’.

The distribution of activity for the three articles show a skewed distribution (Figure 45) in which most Twitter accounts mention an article only once. However, the data also show that for all three of these open access journal articles, there are a few Twitter accounts that mention the article more than 10 times, and in all three cases there is one Twitter account that mentions the article 100 or more times.

Figure 45: No. of tweets by the anti-vaccination movement for three open access journal articles



The most active Twitter account in the anti-vaccination movement was found to share the same content repeatedly across the network for an extended period of time. If we take as a starting point that the anti-vaccination movement in terms of its network properties is highly homogeneous, then research by Piedrahita et al. (2018) on how contagion dynamics emerge when networked actors repeatedly contribute to activity around a collective cause may be significant. They conclude that 'to the extent that digital technologies are inserting networks in every aspect of social life, our results suggest that we should expect to see more instances of large-scale coordination cascading from the bottom-up' (Piedrahita, Borge-Holthoefer, Moreno, & González-Bailónd, 2018, p. 334). And according to Asur et al. (2011), rather than a large number of followers, the most effective strategy to propagate information (at least in terms of creating trending topics) on Twitter is to retweet; the number of retweets for a topic correlates strongly with the length of time the attention of the network is held.

8.4.2 Low levels of engagement

It is not only the suspension of judgement or decision-making in the face of manufactured uncertainty in social media networks that introduces risk. An additional concern is that in addition to uncertainty, or perhaps because of it, the engagement in the social media with scientific information on the part of the anti-vaccination movement was found to be low.

As shown in Table 22, the titles and abstracts of journal articles alone are strong indicators of the likely interest in a scientific article by the anti-vaccination movement. Combined with low levels of engagement, this suggests that scientific content is treated at face value, and that scientific information flows through social media networks with little need for actors in these communication networks to engage any deeper with face-value statements. This finding that engagement with scientific content on Twitter and on Facebook is shallow, confirms findings by Thelwall et al. (2013) that the content of tweets linking to journal articles are unlikely to contain insightful responses to the content of those articles.

The explanation for face-value engagement, according to Stalder (2006), describing the work of Castells, is that certain networks rely and depend on information being taken at face value. Trusted intermediaries (clearing houses) are established as central nodes in the global financial network; they facilitate the rapid exchange and transfer of information that can be trusted at face value across the network. This is critical because the network depends on real-time decision-making by financial traders.

The global science network operates differently, as do social media networks. The different programs of networks determine the speed at which information is accepted as accurate before it is acted upon. The global science network, for example, is programmatically sceptical – scientists are more likely to interrogate information received from others in the network before acting. Social media networks are less so because they are programmatically a network of attention (Wu, 2016), and to retain attention, information must flow constantly regardless of the accuracy of the information.

There were some observable differences between social media platforms with regard to attention, engagement and the use of scientific information for ideological purposes, a finding which confirms that of Valenzuela et al. (2017) that Facebook and Twitter are similar in their participatory effects but the paths through which their influence occurs are distinct. Facebook excels at promoting political engagement by leveraging social pressure made possible by the affordances of the platform. Twitter performs better at injecting and disseminating information because of its heterogeneous social network structure.

From the findings of this research, Facebook appears to be a network where members are tightly connected and less likely to encounter contrary opinions. Partially, this is because Facebook allows its users to exercise greater control over friends, access to content, and, therefore, with whom they interact in the network. In network terms, all connections in the Facebook network are bi-directional. This finding confirms that of a previous study (Quattrociocchi, Scala, & Sunstein, 2016) that found that Facebook users tend to promote their favourite narratives, form polarised groups and resist information that does not conform to their beliefs, creating what the researchers called 'echo chambers'. When deliberately false information was introduced into these echo chambers, it was absorbed and viewed as credible only if it conformed with the primary narrative. When more truthful information was introduced to correct falsehoods, it was ignored, or it reinforced the users' false beliefs.

Twitter, a unidirectional network, shows much higher levels of interaction between different communities (see Figure 35 and Figure 36). Levels of engagement were, however, lower on Twitter than on Facebook.

8.4.3 Conclusion

A social movement that holds a view contrary to that of science (in this case, the anti-vaccination movement) is both highly active and selective in terms of the information accessed from open science and fed into its communications in the social media. The movement produces uncertainty in online communication networks; uncertainty that cascades across online communication networks programmed for attention and devoid of the normative guidance of science designed for settling truth claims.

The anti-vaccination movement also displays low levels of engagement with the products of science on those same social media platforms. Engagement is low because there is little incentive to engage more deeply – information injected into social media communication networks flows more rapidly and widely when it is taken at face value, and this is critical to holding the attention of the network and for the amplification of uncertainty in socially networked communications.

8.5 Trust and intermediation in social media networks

The findings from the network analysis in Chapter 7 indicate that rather than being a single community, there are distinct sub-groups that constitute the anti-vaccination movement. At the same time, there are different types of centrally located actors in the communication network of the anti-vaccination movement, each playing important roles in the dissemination of information within the community and to the broader socially networked public.

In this part of the discussion, the focus is on exploring some of the complexity related to the composition of the anti-vaccination movement that is apparent from the findings of this research as well as the effects of the positions occupied by intermediaries in the observed communication arrangements of the anti-vaccination movement on trust and influence in social media networks vis-à-vis the communication of science.

8.5.1 Segmentation

The findings from this research supports the disaggregation of publics into segments according to their perceptions of science. However, it is not only within nationally bounded publics that segmentation is possible; it is equally possible to segment *within* publics bound by their shared ideology. Findings show that within the anti-vaccination movement there is a sub-group that corresponds with Schäfer et al.'s (2018) categorisation of 'critically interested' publics. Authors of content on anti-vaccination websites such as the Vaccine Papers, show a strong interest in science, some distrust in the findings of science and a dispassionate approach to providing alternative interpretations of scientific results published in accessible journal articles.

At the same time, the category 'disengaged', that is, those who are disinterested in science, do not know much about it and harbour critical views toward it, remains too general based on the findings of this research. The findings show that within the anti-vaccination movement, there is a sub-group who are interested in science and harbour critical views towards it but are not highly engaged with vaccination science. This sub-group is better described as 'interested and disengaged'. The findings of this research also show that while it cannot account for other media through which the anti-vaccination movement encounters science, there is evidence that social media platforms such as Twitter and Facebook constitute an additional medium for its engagement with science.

The findings therefore show that while the ideological objective of the anti-vaccination movement may be shared, their methods for achieving their objectives are not. From the analysis of level of engagement, there emerged a difference in approach taken by a sub-group of anti-vaccination advocates; one which attempts to engage with science on its own terms and which distances itself from the aggressive advocacy tactics used by others in the anti-vaccination movement. In effect, this sub-group appears to subscribe to the Mertonian norms of science; in particular, the norms of universalism and disinterestedness in scientific enquiry.

The most prominent example of this 'objective' approach is that of the Vaccine Papers website. Following an analysis of the scientific evidence of the possible link between aluminium adjuvant and autism, 'flamingolady' posts the following comment:

> It is always all about the money, follow it and therein lies the truth. I know, contradiction in terms, but the lies that the pharmaceutical companies tell have become truth.

To which Vaccine Papers responds:

> Please do not make this a partisan issue. Its [sic] not.

Statements on the Vaccine Papers website further confirm a commitment to objective inquiry:

> Vaccinepapers.org provides detailed, science-based and objective information about the dangers of vaccines. […] We denounce the personal attacks, insults and harassment pursued by some in the vaccine debate (on all sides). We do not do this. […] We assume that both sides in this debate share the same goal: a medical system that maximizes health. […] The vaccine debate has been in the gutter for far too long. There are important issues raised by recent scientific findings that deserve consideration. Society will benefit the most from a rational, polite debate about them.

In addition to the 'scientific' sub-group, there are other sub-groups within the larger anti-vaccination movement who will deny their position as being anti-vaccination. They claim a more nuanced position. According to Hobson-West (in Leask, 2015), these are the 'reformers' as distinct from the 'radicals' in the anti-vaccination movement. Within the reformers, the following sub-groups were evident based on their specific objections to vaccination:[57]

1. *The moderation* sub-group who are against the use of multi-dose vaccines (such as the MMR vaccine) or against specific vaccines (such as the HPV vaccine) or against the number of vaccinations administered in early childhood.
2. *The homeopathic and natural healing* sub-group are opposed to the use of specific non-natural and/or toxic ingredients or adjuvants such as aluminium or foetus cells in vaccines.
3. *The pro-choice* sub-group are opposed to the limitations placed on the rights of parents to choose whether or when to vaccinate their children.

There sub-groups became evident during the content analysis of tweets and Facebook posts. They may account for the anti-vaccination clusters in the tweeter coupling network (see Figure 36).

[57] There are motivations that cut across these three sub-groups: parents who are concerned for the health of their children, individuals who question the integrity and influence of the pharmaceutical industry, and those who have lost trust in the independence and goodwill of public health institutions.

8.5.2 Polarisation

While segmentation within the anti-vaccination movement is evident, social communication networks may polarise political positions (Tucker, et al., 2018; Yeo, Xenos, Brossard, & Scheufele, 2015).

Barberá (2015) challenges the widely held view that social media intensifies polarisation. He contends that social media usage reduces political polarisation because social media platforms like Facebook and Twitter increase incidental exposure to political messages and the social media facilitate exposure to messages from those who are more likely to provide novel information. Consequently, the use of social media leads to an increase in exposure to a wider range of political opinions than those normally encountered offline. The findings of this research do not support such exposure to diverse points of view. Network analysis showed that there is clearly some interaction between the pro-science and anti-vaccination clusters, but connections are few compared to dense connections within and between the anti-vaccination clusters suggesting the existence of echo chambers in which the owners of Twitter accounts in the network are primarily exposed to selected content and like-minded views.

A consequence of polarisation is that it leaves little room for more nuanced, moderate positions. In other words, the communication in the social media may not accommodate such grey areas, or it may create the perception that there are no moderates between the extremes. The risk of the social media creating such polarised positions is that even if some of the more moderate positions hold merit and are worth considering at the policy level, they may be muted by the amplified noise generated by minorities occupying more extreme, polar positions. See, for example, Sunstein (2017) for a more detailed account of the detrimental impact of the social media and polarisation.

A further consequence of polarisation in social media communication may be the fuelling of vitriolic bi-partisan skirmishes rather than measured and rational debate. If this is the case, it may alienate the science community and preclude constructive engagement of the kind advocated by science communicators. It certainly seems more likely that scientists would engage in rational debate and would choose to turn their backs on vitriol fuelled by extreme bi-partisan groups. Some evidence was found of such a response from science. One of the highly engaged anti-vaccination websites, Child Health Safety, on a web page that scrutinises a paper by two Japanese scientists proving that MMR vaccine does not cause autism,[58] laments the fact the despite several attempts to solicit

[58] See https://childhealthsafety.wordpress.com/2009/06/03/japvaxautism/

feedback from the authors of the article, the journal editor and other experts, no response was received.

8.5.3 Missing voices

The global networks whose programs have a vested interest in vaccination are numerous. However, in the tweeter coupling network graph (Figure 36), only three clusters are evident, and these clusters are representative of two global networks (Castells, 2009): (1) the anti-vaccination and the anarchic clusters as representative of the global anti-establishment network, and (2) the pro-science cluster as representative of the global science network. In other words, clusters representing other global networks and that have a vested interest in vaccination, are not represented. These include, for example, a pharmaceutical company cluster (representing the global financial network) with a vested interested in the manufacture and sale of vaccines; a media cluster (representing the global media network) with a vested interest in either profiting from the attention garnered by the anti-vaccination movement or in reporting objective information to the public; and a cluster of public health and other institutions with a vested interest in mitigating the health risks of vaccine refusal.

Sanawi et al. (2017) generated a similar social network graph with clustering using the Twitter hashtag ‘vaccination’. Clusters representative of science and anti-establishment networks were present in their network; but they also found media and government clusters. Sanawi et al. (2017) were more interested in the influence of individual network actors and their network was broader in the sense that a hashtag used to generate the network did not depend on links to scientific publications where those publications were limited in their scope to the topic of vaccination and autism. Why it is that neither the media nor government feature in the tweeter coupling network mapped by this study needs further investigation. It is possible that they are present in the mixed pro-science/anti-vaccination cluster. Certainly, one would expect public health institutions to be pro-science; less so the media that is unlikely to hold a consensus view on vaccination. But this seems unlikely to provide the full answer.

Absent in both this study and that of Sanawi et al. (2017), is a pharmaceutical cluster. The explanation for their absence offered by the anti-vaccination movement would be that there is collusion between the media and the vaccine manufacturers as they have common parent companies. The pharmaceutical companies therefore rely on their sister media companies to engage in communication on the vaccine issue.[59] Whether they are co-conspirators or whether there are other reasons for a noticeable absence of the pharmaceutical industry in a communication network that represents the use of scientific publications in the social media on the topic of vaccination and autism, also requires further investigation.

8.5.4 Influence

There is also evidence that in the social media in general and on Twitter specifically, influence (as measured by the number of retweets) follows a power law distribution, with only a handful of tweeters sufficiently connected for their tweets to reach a large audience (Kahle, 2011; Kwak, Lee, Park, & Moon, 2010).

In a typical population distribution, the extreme positions are in the minority, and the majority is arranged around the centre. In the case of politics, the centre is targeted by both of the extreme positions (often referred to as the 'left' and the 'right') to bolster their support. In online social networks, the distribution of influence follows a different pattern, one referred to as a power law distribution. In such an arrangement in relation to influence, most of the population making up the network hold very little influence, and there is long tail representing the few who hold much of it. This suggests that in social networks, groups who may occupy extreme, minority positions in terms of their political views, may nevertheless occupy highly influential positions in social media networks, and may capitalise on their position to influence the majority holding the centre.

This combination of minority position and high influence is shown to be the case for the anti-vaccination movement. The movement's position is contrary to that of the concensus position in science which is that vaccines are both necessary and safe. However, in the case of the social media network Twitter, the anti-vaccination movement is feeding scientific evidence taken from open access journal articles into an online communication network in which it occupies a position of relatively high influence.

The tweeter coupling network is a network which represents all Twitter mentions of 113 open access journal articles on the topic of vaccination and autism. It shows a typical power law distribution but with minor variations. There are spikes in levels of influence in the mid-range of the tail. Using followers and followings on Twitter, Kwak et al. (2010) report similar deviation from a power law

[59] See, for example, this anti-vaccination web page which links pharmaceutical and media corporations: http://realfarmacy.com/22-medical-studies-that-show-vaccines-can-cause-autism/

distribution due to the presence of many celebrities with large numbers of followers on Twitter. Influence as a factor of followers is therefore unusually high on Twitter. However, in the case of influence in a network as a product of the number of connections to 113 scientific articles shared between Twitter users, the anomalies cannot be explained by the presence of celebrities. Influence in this network is exerted by Twitter accounts that are known to be anti-vaccination. This shows that in addition to occupying positions of relatively high influence in the distribution, the anti-vaccination movement is also able to deviate from the norm in the tail of the distribution to increase its influence.

From the tweeter coupling network, the evidence also suggests that there is no single, obviously identifiable intermediary located between science and the anti-vaccination movement in the Twitter communication network. There is a mix of both anti-vaccination and pro-science Twitter users interacting with one another in one group, and two separate anti-vaccination groups whose information sharing is predominantly within each of those two groups.

The fact that the Tweeter coupling network graph shows a group made up of both members from the anti-vaccination movement and those who are pro-science, was unexpected. Coupling occurs when two Twitter accounts mention the same journal article on more than two occasions. The assumption made is that these two accounts are conceptually connected. At some level, there is a conceptual connection between the anti-vaccination and the pro-science community on Twitter – they share an interest in the effects of vaccination. However, they remain quite distinctive in terms of their stance towards vaccinations and this may explain the appearance of two distinct sub-groups within the mixed-stance group.

The evidence also suggests that there are least two types of intermediaries. The first type of intermediary consists of those Twitter users who intermediate *within* the anti-vaccination sub-groups in the network. The second type of intermediary consists of those Twitter users who intermediate *between* sub-groups in the network.

Between network intermediation can, in turn, assume two forms. Intermediation can be between different communities *within* online social media networks (e.g. between community x and community y on the Twitter social media network) who share a common interest in a particular topic. As shown in the sociogram for the Twitter social network in Figure 35, intermediation in the vaccination–autism topic network is between those who are anti-vaccination and those who are pro-science.

Between network intermediation can also be between different communication networks, for example, between the social media network and the science communication network. The ego networks (Figure 33 and

Figure 34) show connections between a single product of science and a large number of alters who are from different networks as confirmed by Altmetric data (see Figure 9). Communication takes place within the science communication network, non-scientists connect to the open formal communications of science (e.g. journal articles), and communication then takes place in a new network, in this case an online social media network (as shown in Figure 35). Rather than belonging to multiple groups (Grabowicz, Ramasco, Moro, Pujol, & Eguiluz, 2012), these intermediaries play an important role in spreading information by acquiring it from one communication network and injecting it into another communication network.

These different forms of intermediation in the tweeter coupling network suggest that the selection of information from the early stages of the science communication process is done by specific actors in the social movement's network, and that the distribution of the information in the social communication networks of the social movement is undertaken by other actors. Findings also showed what appear to be an unusually high number of centrally positioned actors in the skewed distribution of centrality in the tweeter coupling network. Their location in the network suggests a high proportion of within-group intermediaries who amplify selected scientific information by redistributing that information to others within social media networks. These findings warrant further research based on the knowledge that highly active provocateurs may initiate conflict in online networks, but it is those who are less active who maintain the engagement between communities in conflict (Kumar, Hamilton, Leskovec, & Jurafsky, 2018).

There appear to be specialist functions for certain nodes in the online communication networks of social movements – the switchers connect between science and social movements who are organised such that the amplifiers spread selected scientific information across their own as well as other social media networks (Zannettou, et al., 2017).

8.5.5 Risk

Risk can be amplified by social factors when risk signals, often triggered by the media, are received, interpreted and passed on by a variety of social actors (Kasperson, et al., 1988). Previous research has interpreted these social amplification effects as being place-bound (Petts & Niemeyer, 2004, p. 19): 'it was evident that concerns, understanding and beliefs had been reinforced by the communities in which people live. Maternal responsibilities and knowledge are maintained, supported and developed through a close social network, which in urban areas seems to be amplified by spatial polarisation'. However, if social amplification is dislocated from place and takes place in the space of flows exemplified by online social (media) networks, then the amplification of risk may be increased.

The amplification of risk in the space of flows has a bearing on how change is effected. Miller (2017) argues that social media communication is not transformative but phatic and, as a consequence, the social media do not mobilise political action. The assumption is that change proceeds from the real-time, global interpersonal connections and communications in the space of flows made possible by the social media to action in the space of places against concentrated, hierarchically structured power such as an oppressive regime or Wall Street. In other words, change requires a switch from the space of flows to the space of places. Based on a review of the evidence, Miller (2017) argues that the social media activism is not transformative or politically goal-orientated.

The evidence presented in this thesis suggests that the social media can effect change along different lines. Change is made possible by the production of uncertainty and, in certain spheres of social life such as health and well-being, is equally capable of driving change. Change proceeds from the real-time, global interpersonal connections and communications not to place-bound action but to an increase in influence over a diffused and interconnected mass public, more of whom seek meaning in the space of flows. In this sense, every connected individual is a target in attacks on what is held to be certain, and uncertainty has the potential to alter the decisions taken by individuals. In this change process, a shift from the space of flows to space of places is not required because change is effected by disrupting the flows of information in other communication networks by amplifying uncertainty. This includes disruption in the flows of scientific certainties. It is unlikely that all change can be effected in this manner but it is short-sighted to suggest that the social media cannot be used by social movements to effect change, particularly when decision-making at the individual level poses risk to the masses.

While the amplification of risk in the social media may increase the probability of detrimental health outcomes, the opportunity for science to counter risk through effective communication may also be undermined by the attraction of the social media as a communication venue. This thesis limited its investigation to the indirect or latent communication of science made possible by ubiquitous access to the products of science and it did not therefore investigate the direct communication of science by scientists or via any of its trusted intermediaries (such as universities, scientific bodies or even science journalists) to non-scientific audiences. However, there are indications of direct communication between scientists and the anti-vaccination movement as evident in the findings.

During the data collection phase of this thesis, it was not uncommon to find pro-science accounts and anti-vaccination accounts using hashtags and mentions to attract one other's attention. Often this was done to draw the attention of the opposing party to a journal article supporting their position. In the case of the tweeter coupling network, it was found that that pro-scientists, many of whom can be presumed to be scientists, and members of the anti-vaccination movement are clustered in the same group by virtue of mentions to open access journal articles on vaccination.

These findings suggest that pro-scientists could in fact be the unwitting transmitters of information from science to the anti-vaccination movement in the social media. In other words, scientists may be the ones drawing attention to the products of science using the affordances of the social media platform. These findings echo to some extent the argument put forward by Weingart (2017) that scientists, from the point of view of scientific norms, are guilty of 'inappropriate' attention-seeking communication practices, and these practices may put at risk the reputation of science. Scientists, their funders or their host institutions stir up hype around a new discovery (often published in a journal article), rather than letting the discovery follow the established path through the iterative, protracted and self-correcting system of truth verification. When they do so, they run the risk of making unsubstantiated, decontextualized and even fraudulent claims that, when exposed, cause damage to the reputation of the trusted public institution of science. Similarly, communication between scientists and ideologically-motivated groups on social media platforms may, instead of recalibrating the opinions of minority-group members, undermine the authority of science and the effectiveness of its communications.

8.5.6 Trust and attention

Unlike the global financial network that has created centrally located and trusted nodes in the networked flows of information to ensure that information in the network can be taken at face value, as well as self-regulating structures within the network and the assurance of external, government intervention in the case of the threat of collapse, social media networks have no such mechanisms to ensure trust in the information exchanged.[60] This may account for why ideologically-

[60] It is not that the structures put in place by the global financial network are infallible. The point is that there have been attempts to self-correct, that is, to protect the network's program of surplus accumulation. There has been no concerted attempt to self-correct across social media networks.

motivated social movements active in global social media networks appear to borrow or import trust from the open communication networks of a social institution that is trusted by the public: science.[61] As the Vaccine Research Library proclaims: 'We have more than 7,000 links to abstracts and full text from mainstream, scientific literature […]. If their own literature isn't a 'reliable source', then what is?' (Vaccine Research Library, 2015).

In the real world (the space of places), social cues confer authority and trust; in networks (the space of flows), these cues are not necessarily linked to class, cultural status or other traditional social cues (Lin, 2008). Network social capital or 'network capital' may present itself as a new type of capital that accumulates in virtual networks as socially networked actors attract and consolidate the attention of others in the network. Attention-seeking as a strategy to gain an influential position in networks accounts for the migration of unexpected actors to parts of the communication network where they are most likely to attract attention. Medical professionals and scientists are not immune to such attention-seeking behaviour in their quest to extend their influence over others, and it is for this reason that there are doctors and scientists to be counted in the online communication networks of the anti-vaccination movement. The anti-vaccination rock band, The Refusers, is a further example of an attempt to seek influence across networks. In some instances, existing capitals (cultural or symbolic) from the space of places may be leveraged to attract attention in the space of flows (networks) so as to accumulate network capital. For example, doctors converting their cultural capital (expertise) to network capital by switching allegiance from the professional network of medicine to that of the anti-vaccination movement's communication network.

There is also the potential for the conversion of network capital to economic capital. Fake photo sites on Twitter post doctored historical photos (e.g. Stalin playing baseball). The photographs posted are known to be fake but nevertheless hold popular appeal. Once these Twitter accounts have amassed a large number of followers, they leverage their network capital to attract economic capital as advertisers are prepared to provide financial rewards for the attention that these accounts can bring to their brands via their follower networks.[62] In other words, social media does not conform to expected rules and social hierarchies that confer authority or trust – a fake account can attract more attention and, by implication, yield more influence in a social media network than the account of a trusted, authoritative source, including that of a scientist.

---

[61] See Schäfer (2017) on science as a trusted institution and Lin (2008) on open networks and the accumulation of social capital.

[62] See Reply All podcast #48 by Gimlet: https://gimletmedia.com/episode/48-i-love-you-i-loathe-you/

The motivations behind scientists and medical professionals' participation in the social media or their motivations for expressing their allegiance to the anti-vaccination movement, may well be attention-seeking. Nevertheless, their presence and their inferred authority in these communication networks, destabilises traditional social cues of authority by creating the perception of divided positions on which there is, in reality, scientific consensus. In such a scenario, who to trust becomes unclear. It may be clear within relatively closed networks with shared norms and values who to trust (Burt, 2001), but to outsiders, where those norms are no longer shared, it becomes increasingly difficult to identify trusted sources. This has implications for uncertain parents and policy-makers alike who find themselves participating in online communication networks where trust has been destabilised, as this research has shown, by active minority groups exploiting the attention imperatives of those networks. As Southwell (2017, p. 223) states, communication of scientific information in the social media 'can undermine scientific authority, complicate decision-making and fuel the propagation of rumours and misinformation'.

How centrality is established in networks and how 'network capital' is accumulated to establish a position of trust and influence in an online social network, a question that was not explored in this thesis, remains opaque. If, as some have suggested (Muller, 2017), influence is a new form of power in the network society, then it becomes increasingly important to understand better not only who the trusted influencers are, but how they establish and protect their positions of influence.

Castells's (1996, 2004) concept of network programs offers a potentially fruitful line of enquiry in terms of how influence may be established. In pre-network society networks, the program of each network determines its own measure of success for nodes in the network. In the financial network programmed for profit, success is wealth measured by the quantum of surplus accumulated; in the network of scientists programmed to pursue truth, success is reputation measured by extent to which other scientists cite a particular scientist as the source of truth. More successful nodes are promoted up the hierarchical structures of these networks and such promotion, in turn, confers authority and a degree of control over lower reaches of the network. In the global communication networks that define the network society, these measures of success remain relevant, but they are insufficient. They are insufficient because placed-based strategies for drawing attention to success – clubs, schools, committees, social events, and the like – are no longer as effective in a world defined by the space of flows. More significantly, the hierarchical structure of traditional networks is replaced by the horizontal arrangement of network nodes precluding promotion as a strategy for attracting attention. What is therefore required in the network society are additional strategies for attracting attention.

It is not that attention replaces criteria for success as determined by the programs of global networks, but that attention becomes an indispensable component in a mutualistic relationship.

In social media networks, the problem faced by nodes in the network is that there is no central, unifying network program and attention becomes the only programmatic criterion for determining success in the network. In crude terms, success is number of followers, number of likes or number of retweets – any one of a number of metrics indicative of the quantum of attention attracted by a node in the network relative to the other nodes in the same network.

This explanation of the role of attention in the establishment of influence in networks remains speculative. This research has shown how the anti-vaccination movement is able to attract disproportionate levels of attention in online communication networks to exert influence of what is certain or true. Further research and conceptual development are needed to move towards a more comprehensive theory of attention, influence and power in the network society. Developing such an understanding will be critical for the science of science communication as it seeks to inform effective strategies for the communication of science to networked publics.

Chapter 9. Implications for science communication

In certain circumstances, being confronted by uncertainty has negligible consequences – decision-making can simply be deferred or suspended. But for the uncertain parent of a new-born child who faces a time-bound decision on whether to vaccinate, being confronted by amplified anti-vaccination messaging seemingly supported by science, presents the very real possibility of the parent electing not to vaccinate. And this decision would be taken despite the fact that the consensus position within science, based on available scientific evidence, is supportive of vaccination. These findings are echoed in a recent article:

> The broader public health implications of propagating these memes and articles make anti-vax activities more than a bizarre online curiosity. Most of the material that the […] accounts tweet are designed to erode confidence in vaccination. The goal is to make new parents question everything … Public health officials are concerned. 'It is important to remember that today, the vast majority of people follow the recommended vaccine schedule – they take the advice of their doctors, supported by professional medical bodies and the WHO,' says Gary Finnegan, editor of Vaccines Today. 'However, it is essential that when people go online for information they are left with the clear impression that vaccines are safe and effective.' If that's going to change, the people fighting misinformation need to understand how it gets spread in the first place. (DiResta & Lotan, 2015)

This thesis has attempted to contribute to an understanding of how information from open science is used by the anti-vaccination movement in its online communication. The question for science communication is: How should science respond?

A consideration of the implications of this case study research for the communication of science as a whole is an ambitious undertaking. Therefore, two points of departure are followed to structure a discussion on implications as informed by the findings of the research. The first point of departure relates to implications for the communication *in* science; in other words, the formal communication of science. The second point of departure relates to those implications for the communication *of* science; in other words, communication outside of science typically communicated by science journalists, museums, foundations, universities and the like through popular media and exhibitions to non-scientific publics.

The chapter ends by broadening the implications for science by exploring its status as institution and global network, and its relationship with other institutions such as the market and the state as they relate to the communication of science. The chapter concludes with areas for further research that emerge either because of the limitations of this study or because of their importance in furthering the science of science communication.

9.1 Implications for communication in science

Online repositories such as the US Government's National Center for Biotechnology Information's (NCBI) PubMed.gov are vast, openly accessible digital collections of scientific information. The NCBI suggests that it 'advances science and health by providing access to biomedical and genomic information' (NCBI, n.d.).

Along with the scientific community who may benefit from such a central collection of scientific information, politically or ideologically motivated minority groups are also able to access the repository of scientific information; scientific information that has not yet been codified into factual knowledge through a prescribed process to determine consensus. Armed with selected and unvetted truth claims, minority groups deploy the network effects of the social media to amplify uncertainty in non-scientific communication networks where the norms of science do not apply. Such amplification of uncertainty increases the possibility or likelihood of undecided groups being confronted with information that challenges or destabilises their position of certainty.

The findings have shown that open science is not the only change that has allowed scientific information to flow through social media networks in general, and within non-scientific, ideologically-motivated social media communities in particular. Non-scientists are just as likely to rely on abstracts or article titles in their online communications when the full-text version of a relevant journal article is inaccessible to them. And the titles and abstracts of journal articles are all available online, published either on journal websites or on the personal online pages of the authors (for example, on Research Gate or Academia.edu).

Moreover, the finding that the level of engagement with the content of journal articles is low, suggests that certain non-scientific communities have little incentive to access the full-text versions of journal articles.

The response from science to such uncontrolled access to its internal, formal communications may be for it to become more insular. Data sharing which imposes controls over access may be preferred to open research data. This research found evidence of groups or teams of scientists choosing to share research data with each other and with their vetted peers, but not with those unvetted by the group. Controlled access to journal articles could also be introduced. The online science networking platform ResearchGate applied such control by only allowing scientists with approved institutional email addresses to register for its service. While ResearchGate was not controlling access to content, it did create a mechanism that attempted to restrict membership to a specific globally networked community.

Imposing access controls will undoubtedly also remove some of the reported benefits of open science. At this point in time it seems unlikely that a return to controlled access is likely or even possible. It should therefore be taken for granted that the formal communications of science will remain accessible to non-scientific publics. And while the benefits of openness should still be weighed against its potential risks in attempts at developing responsible approaches to how science is made open to non-scientists (Jasanoff, 2006), science will need to remain vigilant to the effects of unfettered access to its formal communications.

9.2 Implications for the communication of science in the social media

While the barriers to participation in online communications are relatively low, the requirements of internet access and a suitable device may nevertheless preclude participation in global communication networks. Publics on social media are therefore only a segment of the global population. More so in the case of specific social media networks if one considers that in countries such as China, Facebook and Twitter are not available at all. At the same time, in some developing countries, data charges are waived to access social media platforms like Facebook. Nevertheless, those on social media networks (in the place of flows) who numbered 2.23 billion active users on Facebook[63] (Statsita, 2018) and 335 million on Twitter[64] (Statista, 2018), can wield significant influence over those in the real world (the space of places) and science communicators can therefore ill-afford to ignore the social media.

At the same time, trying to target those who are set in their non-census views on vaccination using social media may prove futile (Chan, Jones, & Albarracin, 2017). Adversarial approaches are also

[63] As at the second quarter of 2018.
[64] Ibid.

unlikely to be effective (Leask, 2015) and this research has shown that the value placed on attention in social media networks is likely to render any adversarial communication strategies targeted at the anti-vaccination movement unproductive. In fact, such attacks are likely be to the benefit of the anti-vaccination position as they bestow additional attention on the vaccination controversy. It may be more productive to target those who are undecided about their position on vaccination, or, based on the finding that the anti-vaccination movement is segmented, target more moderate groups within the anti-vaccination movement. It may also be more effective to identify those occupying central positions in online communication networks – a task that is relatively easy to accomplish using the data made available by social media platforms and network analysis software – and to target communication at those individuals in an attempt to cripple key nodes in the communication networks and/or to create new pathways (switches) conducive to disseminating information that is consistent with the consensus position on vaccination.

According to DiResta and Lotan (2015), 'people don't organise in groups around everyday life-saving measures; there is no pro-seatbelt activist community on Twitter. The recent emergence, lack of central coordination, and weak connections seen among pro-vaccine Twitter users, who often use the hashtag #vaccineswork, means that the pro-vax message is not amplified to the same degree' as the anti-vaccination message. The findings of this thesis confirm the use of social media to amplify anti-vaccination messaging, and science communicators may need to be more strategic about how they maintain a comparative level of presence in the social media. This is likely to have a minimal effect in disrupting the echo chambers of the anti-vaccination movement but is important in extending the reach of pro-science messaging to those holding more moderate views as well as to parents and policy-makers undecided about the benefits and risks of vaccination.

Post-modernists cast doubt over scientists as the sole arbiters of truth (even if they retain exclusivity over its certification) (Kata, 2012). The contemporary communication landscape clearly places new and greater responsibility on scientists and their institutions, who are increasingly active in communicating with the 'end user' and who are not always well-prepared to deal with the dynamics and potential risks of such engagement. During the heated debate that ensued about vaccination in Italy in 2016, an immunologist who had committed to engage in discussion through his own Facebook page eventually decided to abruptly cancel all comments by claiming, 'Here only those who have studied can comment, not the common citizen. Science is not democratic' (Bucchi, 2018).

And yet scientists are drawn to and are encouraged to participate in the social media. To some extent, the benefits are obvious. The social media provides a relatively pervasive and user-friendly network with low barriers of entry. Participating in such a network makes possible connections across the scientific community, possibly useful in the dissemination and discovery of new truth claims, and for crossing paths with potential collaborators on future scientific endeavours. In effect, the social media provides a new mechanism for community-building. This sits well with the normative position of communalism in science which holds that results and discoveries are not the property of any individual scientist but belong to the scientific community *and* to society at large. Merton's statement cited in the introductory chapter bears repeating:

> The institutional conception of science as part of the public domain is linked with the imperative for communication of findings. Secrecy is the antithesis of this norm; full and open communication is its enactment (Merton, 1968, p. 611).

The social media is an effective means of communication. At the same time, however, as this thesis has attempted to show, the social media as a global online communication network does not subscribe to the norms of science. Most patently, it is a space of information flows where organised dogmatism, rather than organised scepticism, flourishes because of the social media's attention-imperative.

If science wishes to communicate effectively in social networks, it is therefore not the scientist who should be communicating on Twitter or Facebook, unless it is for reasons other than the pursuit of truth. Instead, scientists or their host institutions may be better off associating with those who wield influence in social networks to deliver their messages.[65] In other words, this calls for new arrangements (interactions) between scientists and *trusted* intermediaries who are more adept at social media communication. This may include collaborating more actively with pro-vaccination groups. While there are not many of them, there are examples of pro-vaccination groups attempting to restore the 'passion asymmetry' of the 'silent majority'; a case in point being a group of Californian mothers who relied on social media and other tactics to lobby for the passing of mandatory vaccinations in that US state (DiResta, 2017b).

[65] It is possible that scientists themselves may wield sizeable influence in non-scientific online communication networks and are thus are able to leverage their position to challenge consensus positions. Professor Tim Noakes' contested use of social media to dispense unconventional advice serves as an example (Gous & Child, 2018).

A different approach is the role assumed by the growing number of fact-checking organisations. These organisations assess the truthfulness of claims made by politicians, by the news media and others. An example of this type of approach is the Sense about Science[66] and Skeptoid[67] initiatives.

The problem with fact checking is that it is unlikely to shift the beliefs and practices of the anti-vaccination movement or other social movements that are beholden to minority belief systems. Research has shown that increasing factually accurate messaging to those with strong contrary beliefs, may in fact further entrench their beliefs (Kahan, 2014; Mitra, Counts, & Pennebaker, 2016). This is referred to as the 'backfire effect' (Nyhan & Reifler, 2010). As Mitra et al. (2016, p. 270) conclude from their study: 'These findings suggest that health officials attempting to simply correct conspiracy fuelled false claims might be counterproductive. Thus, newer methods to counter the harmful consequences of anti-vaccination beliefs are needed'.

Moreover, fact checkers typically only activate after a claim has sparked a controversy and spread across online communication networks. This implies that they too are part of the attention apparatus and that they have limited success in countering the views of those who have settled into a polarised position. Nevertheless, if trusted fact checkers are able to occupy central, intermediary positions in communication networks, then this may at least result in an increase in the likelihood of scientifically sound communication reaching those whose beliefs are not polarised.

One of the network effects of feeding selected scientific information from the early stages of the science communication process into social media networks is the production and amplification of uncertainty. Such uncertainty may pose threats to the credibility of science as an institution in the eyes of the public (Kahan, Scheufele, & Jamieson, 2017). Institutions such as science react to threats to their credibility (the extent to which they are trusted) by making taken-for-granted norms more explicit (Weingart, 2017). In network terms, the network's programmers must defend the logic of the network from attacks made by other networks or emanating from the network environment (Stalder, 2006). By making the institutional norms of science more explicit or by enforcing the terms of participation in the global science communication network, scientists should, in theory, refrain from non-normative attention-seeking in other communication networks or face sanction from their own.

[66] http://senseaboutscience.org/
[67] https://skeptoid.com/

Nevertheless, science communicators are unlikely to let go of the potentials for the effective communication of science via the social media, particularly given that such communications may have positive outcomes for both science and its attentive audience. Nor will all scientists retreat from the attraction of attention bestowed on them in non-scientific communication networks. The question then becomes how to communicate science in the social media so that the positive benefits can be reaped, and the negative outcomes avoided. Southwell (2017) suggests three strategies: (1) creating interaction spaces; (2) increasing communication confidence; and (3) heeding content factors. All three strategies are aimed at increasing levels of interpersonal communication for the benefit of science.

The suggestion most relevant to the findings of this thesis is that of leveraging network effects to create community conversation forums and reinforce the consensus position of science before controversies ignite. Putting aside the question of how one would anticipate controversies emanating from a wide range of scientific discoveries, this research has shown that there is little incentive for minority groups who cling to strong ideologically-motivated positions to engage deeply with factual bases underpinning the consensus position. In cases where sub-groups may wish to engage in such interaction (e.g. Vaccine Papers), their influence over others in the anti-vaccination movement is questionable based on their non-central positions in social media communications networks such as Twitter. Moreover, consensus positions are not necessarily universal and there are likely to be contrarian positions within the scientific community. With access to the publications of those dissenters within science, minority social movements will feed such information selectively into the information flows in the social media. It is therefore not clear how safe spaces or forums will counteract the amplification of uncertainty in pervasive social media networks.

A modified approach is to increase *direct* contact between science and social movements who hold non-consensus positions (Pettigrew & Tropp, 2006; Petts & Niemeyer, 2004). Whether the social media is suitable for such engagement which is predicated on direct contact is debatable, although research on online conflicts has shown that direct engagement between attackers (anti-vaccination movement) and defenders (pro-science community) when initiated by defenders can reduce the likelihood of negative outcomes (Kumar, Hamilton, Leskovec, & Jurafsky, 2018).

According to the Kahan (2014; 2013), greater effort should be made to protect the science communication environment. He argues that the uneven levels of public resistance to different vaccines is linked directly to how the risk associated with vaccination was communicated in the first place. Deviation from communication that is empirically grounded and adheres to scientific norms increases the probability of controversy and vaccine refusal.

The problem with the protection of the communication environment approach is that it rests on the assumption that there exists a public prepared to either support or resist a specific vaccine depending on how the communication of science is managed. This thesis has found that there is a segment of the anti-vaccination movement that is vehemently opposed to *all* vaccines. Kahan acknowledges that there are 'enclaves' of strident opposition to vaccination. It is unlikely that they will be swayed by communications emanating from a protected science communication environment. Moreover, as the evidence has shown, not only will the group remain unaffected, but it is able to amplify its anti-vaccination message to the levels that may extend its influence into the broader public despite its relatively diminutive size.

To return to a point made earlier on in this chapter, there are also questions to be asked about the relationship between a protected science communication environment and open science: Does making science more accessible to all publics protect or jeopardise the integrity of the science communication environment? The findings of this research suggest that minority groups unwavering in their opposition to vaccination are accessing empirically-based scientific information to shore up their online anti-vaccination communication strategies.

The discussion on the communication of science and the social media suggests two options. The first is to accept the pervasiveness of the social media and the unlikelihood of being able to protect the science communication environment from the distortions of the social media. Such an acceptance leads to further choices such as (1) using the social media more effectively for the communication of science by relying on insights emerging from empirical research on how best to do so, or (2) deploying alternative forms of communication such as direct, face-to-face engagement with central individuals or groups in the anti-vaccination movement in attempts moderate their extreme stance on vaccination.

The second option is to intervene in the communication environment to protect society from the adverse consequences of the social media. Such an interventionist approach would benefit from understanding the network society, those social institutions that depend on global communication networks, and the relationship between them.

9.3 Science and the social media: networks and institutions

Southwell (2017) identifies several intermediaries in the communication of science. The list includes the usual suspects, but it also includes social networking sites. The inclusion of the latter is surprising and hints at category confusion. The usual suspects intermediate from within institutionalised frameworks – and their communication is shaped by the norms of those institutions. This is not so in the case of the social media. The social media provide a set of online affordances (profiles; posting content; hashtags; mentions) through which actors communicate. There are no institutionalised norms in social media networks, other than the multitude of different norms that the communicators bring to the networks.

To be sure, networks (in the Castellian sense) and institutions (in the neo-institutional sense) share some characteristics. Both are social arrangements consisting of social actors. The behaviour of actors in both institutions and networks are prescribed by programs or logics. Both have self-defined rules for inclusion and, by implication, exclusion. Neither is dependent on individual actors for their survival. But there are differences also. Networks are more adaptable and flexible – they can rapidly delete and add nodes to adapt to external shocks. Networks are horizontal; institutions are vertical. Institutions are cumbersome, their default position is to resist change emanating from external pressures.

Science is both an institution and a global communication network. Scientists communicate in real time at a global scale according to a program of truth-seeking and their actions in the network are shaped by the institutionalised norms of science. The social media is clearly not an institution. Is the social media a communication network in the Castellian sense?

Social media are networks of communication making possible the instantaneous flow of information across the globe. There are programmers of social media networks but in the technical rather than the social sense. The key difference between the global networks of finance, science, etc. and social media networks is that social media networks are controlled by actors external to the network.[68] In other words, while certain social actors constitute the communication network, there are different actors who determine the program of the network. Those actors who are in control of the

[68] The locus of control is the key differentiator here. As noted in the introduction to this dissertation, challenges to the programs of networks can certainly be mobilised using the social media but most often these attacks emanate from *within* the network by those seeking to exploit vulnerabilities in the program – either for personal gain in the short-term or to meet a broader objective which requires the reprogramming of the network.

technology that make the social network possible[69] are firmly rooted in the global financial network, even though they may be exploiting a sense of belonging in the anomic society of today. The program of social media networks is therefore to create a surplus of capital. And the mechanism for creating a surplus of capital is attention. By creating technology that holds the attention of global audiences, the controllers are able to sell advertising as a means of converting attention into profit.[70]

There are several social media networks each controlled by different corporate actors (predominantly Google, Facebook, Twitter, 10Cent). These corporate actors are in competition for financial capital and while different social media networks may be connected by virtue of the central affordance of the world wide web – the hyperlink – they do not co-operate.

What does this mean for science and its communication? On the one hand, it is unlikely that science can separate itself from the global networks of communication. There are risks in participating in these communication networks, but equally, there are benefits. It seems premature to speculate on whether the benefits outweigh the risks. We may be underestimating the risk introduced by the production of uncertainty, especially if it is likely to have material impacts on the reputation and credibility of science as the arbiter of truth. At the same time, there are some who augur the demise of the social media as society self-corrects for the untenable distortions in reality fuelled by the attention imperative of the social media (DiResta, 2017a).

The public could theoretically collapse social media networks by starving them of attention. The reason for doing so would be growing disenchantment with the inevitable network effects of social media, predominantly in the form of the distortion of information (as this thesis has shown to be the case for the anti-vaccination movement). This outcome is by no means certain, and it is impossible to predict what other equally disruptive media may emerge in the future.

Within this uncertain communication environment, the resilience of science as an institution and its value to society transcendental to the pressures for it to communicate to publics in the media of the moment, places it in a strong position to navigate lightly scathed through current vulnerabilities that are a consequence of its increasing openness.

[69] They are also to varying degrees owners of the content created by the network although this form of ownership is under intense scrutiny and has led to the promulgation of legal instruments such as the General Data Protection Regulation (GDPR).
[70] Wu (2016) points out how Google initially sought to avoid the commodification of attention but eventually chose that path anyway when it opened its products to advertisers.

9.4 New encounters between science, the state and the market?

Science may be resilient, but the likelihood also exists that the state will need to intervene to correct distortions created by the social media. In Uganda and Tanzania, governments are attempting to curtail the effect of social media and other forms of socially networked online communication through taxes and other mechanisms (Aglionby & Pilling, 2018; Kuckler, 2018). Federal and national governments have already intervened in other global networks such as the global financial market when these networks were distorted by changes in technology (DiResta, 2017a). Recent scandals around the sale and use of personal data by Facebook and Cambridge Analytica has stirred renewed interest in the role of governments in regulating near-monopolistic owners of social media networks (Shah, 2018). There is also evidence of self-regulation as Facebook, Twitter and others invest in technologies to moderate content on their platforms although not necessarily along the same lines (Daily Maverick, 2018). Recently, the messaging application WhatsApp placed limits on the number of times a message can be forwarded in its network (Hern, 2018). This was in response to a spate of rumour-fuelled killings in India attributed to false information spreading unchecked across the WhatsApp network (Griffin, 2018).

Science will need to tread carefully as the state and the market jostle to 'correct' the distortions of the social media, including the communications of science. Science will need to find the sweet spots (safe spaces; trusted and influential intermediaries) and avoid the pitfalls of hype, high levels of attention, low levels of engagement with (scientific) content in an increasingly complex communication environment. To do so, it will need to deploy its own foot soldiers in the social sciences and other disciplines to better understand the communication environment and the opportunities for effective transmission of scientific truths in that environment.

9.5 Further research on open science communication

The research presented in this thesis was a first attempt at creating a better, empirically-based understanding of the potentials of open science; specifically, the potentials arising from increased access by non-scientists to those products produced in the formal communication of science. That the research focused on a non-scientific social movement opposed to vaccinations meant that the potentials identified were in the form of risks. The anti-vaccination movement campaigns for health practices that go against the accepted scientific position on vaccination and as such their anti-vaccination campaigning is seen as introducing new health risks to society. The study of other social movements' use of the products of open science may reveal more positive potentials. Similarly, research on other social movements may confirm the findings of the single case presented in this thesis. Both endeavours are needed to be able to assert more generalisable conclusions needed to advance the science of science communication.

Other, more specific, areas for further enquiry emerged from this research. Although not an exhaustive list, possible areas for further research are explored below.

There is further research to be done on the network effects and its relevance to the communication of science. This research made only a very modest attempt to explore such network effects and there remains the need for more in-depth network analysis of the potentials of open science. For example, the interaction between open networks and network closure as it relates to science as an open communication network and the strategies used by ideologically-motivated minorities in the social media warrants further study. Open networks allow for an increase in weak ties between networks and the extraction of value previously not possible, while tightly connected closed networks (equivalent to the echo chambers of the social media) allow for the realisation of value (Burt, 2001). The weak connections between networks are described as 'structural holes' in the social structure of networks. According to Burt structural holes are an 'opportunity to broker the flow of information between people, and control the projects that bring together people from opposite ends of the hole' (Burt, 2001, p. 35). This is not that dissimilar from Castells's (1996; 2009) notion of network switchers. Such 'bringing together' may be productive in commodity markets, but in a marketplace of ideas structured around exchange in social media networks it may be unavoidably conflictual (DiResta, 2017a).

Some anti-vaccination websites were found to be highly engaged with the products of science (see section 6.3.3). This was in stark contrast to the low levels of engagement in the social media (see section 6.3.1 and 6.3.2). Evidence was also found of self-archiving paywalled journal articles to allow for permanent access to these selected articles by members of the anti-vaccination movement (see section 6.4). The websites where these highly engaged pages are published also have social media accounts; however, they are relatively inactive on the social media (see section 6.3.3). The question that emerges from these findings is: How is bridging taking place between the highly engaged, inactive part of an online community and the highly active, low engagement part of the community, if at all? If the highly engaged are the intermediaries between science and minority social movements, then this raises further questions related to emergence and influence of intermediaries (see section 8.5.4) in social media networks. For example: What is the point of entry to the closed communication networks of online communities, either directly from another communication network (e.g. science) or from an intermediary in the network? It is not clear from this research, for example, the paths followed by the products of science between and within networks: from first point of contact to the spin cycle of social media echo chambers.

Burt's (2001) proposition for the emergence and influence of intermediaries as being a function of their ability to exploit structural holes provides one account of how attention and influence manifests in communication networks. Other lines of enquiry related to the existence, creation and protection of network social capital are also needed (Lin, 2008) to explain the control and influence exerted by central actors in online communication networks, as are accounts of non-human intermediaries such as algorithms in exerting influence over the flow of information in online communication networks.

This research limited its scope to the interactions between non-scientists in the social media and science, and anchored those interactions to the products of science. Further research, similarly anchored, is needed to broaden the scope of information flows in communication networks by, for example, undertaking research on the interaction between multiple communication networks including the traditional media, the social media and other networks in the production of uncertainty. Further research is also needed to explore why stakeholders with a vested interest in vaccination, notably the pharmaceutical manufacturers and public health institutions, were absent in the online communication networks in this study.

Another line of further enquiry relates to developing a more nuanced understanding of the composition of social movements (see sections 7.4 and 8.5.1), including the different strengths and bases for beliefs held within movements, and roles played by different group members. Such an understanding may assist in the formulation and delivery of more targeted and effective science communication strategies on controversial topics. Such communication would seek to avoid those holding unshakable positions and target those whose positions are uncertain (Leask, 2015).

In sum, additional empirical research on pathways, bridges, the production and amplification of uncertainty, and the multiplicity of roles and characteristics of members within social movements, will contribute to a more fine-grained understanding of the potentials of open science and how science can respond to the detrimental effects of greater openness in the network society.

Chapter 10. Conclusion

The genesis of this thesis was an intuition that open science may pose risks to society. Minority groups such as the anti-vaccination movement that are highly organised, globally networked and ideologically motivated can access the products of science and could exploit global communication networks such as social media to influence uncertain individuals with potentially dire consequences.

The initial interest was in that movement's use of open research data; in particular, the possibility that they may be accessing and reinterpreting research data to bolster their claims that vaccines are harmful. Rather than relying on articles published in scientific journals, written by scientists who, in their eyes, are in the pockets of government, the media and pharmaceutical companies, the raw data would provide 'uncontaminated' evidence of the links between vaccination and autism.

However, contrary to expectation, no evidence could be found of the use of open research data by a vociferous minority social movement. Instead, the evidence points to the use of selected scientific information, extracted with little engagement from open access journal articles, by a highly active minority group to produce and amplify uncertainty in the broader population using social media networks. The social media environment, devoid of scientific norms to steer action toward the establishment of truth, provides an ideal communication substrate, as does the networked nature of online communications. Online communications networks in the form of the social media enable relatively small social movements to exploit the affordances of those networks to amplify their messaging. They also create opportunities for different types of intermediaries to insert themselves into information flows to exert relatively high levels influence for the benefit of the minority social movements to which they belong.

In 1939, British Prime Minister Neville Chamberlain travelled in a Lockheed Electra from London to Munich to negotiate with Hitler, Mussolini and Daladier for the peaceful settlement of the Sudeten crisis. An event preceding his departure for Germany captures simultaneously the seepage of politics into communication, and the risks of openness. While driving to Parliament to address Members, Chamberlain's secretary adds the following text to his speech:

> Signor Mussolini has informed Herr Hitler that while Italy will fulfil its obligations to Germany, it nevertheless requests that mobilisation be postponed to 24 hours. Herr Hitler has agreed. (Harris, 2017, p. 118)

Chamberlain finds his secretary's account insufficient, and dictates the following addition:

> Whatever views Honourable Members may have had about Signor Mussolini in the past, I believe that everyone will welcome his gesture of being willing to work with us for peace in Europe. (Harris, 2017, p. 119)

Chamberlain is attuned to the fact that while he will deliver his speech to the Members of the House, Mussolini will undoubtedly receive word of its content. The added text is therefore not aimed at the House but at Mussolini (and his receptive ego).

This historical account captures the ever-political nature of communication. But is also shows an acute awareness of the potential of access to the communications of others. It is the potential risk of open science that has been illustrated in this thesis, that is, the ability of outsiders to access and use the products of the formal communications of science for political gain; use which exploits the affordances of online communication networks devoid of the normative compass that guides the vetting of truth-claims within the formal communication of science among peers.

A more contemporaneous attempt at political detente illustrates the point that communication in the social media is rudderless, that is, that it becomes unhitched from the norms that program other communication networks. Reporting ahead of the unprecedented meeting between President Trump and President Kim Jong-un in 2018, *The New York Times* reports that 'the meeting will [in addition to the unpredictable personalities of the two leaders] also exemplify a world that is less constrained by the usual guardrails of international norms or policy processes – a world in which virtually anything can happen' (Fisher, 2018, p. 3).

No wonder then, that President Trump chooses to communicate using the logic-less social media. And that minority groups alert to both its amplification effects and its potential to disrupt certainty in other communication networks, choose to communicate using social media networks where attention is the common currency. Science itself will need to remain attentive to the effects of the social media as its own communications become more open, and take seriously the science of science communication in this shifting communication environment if it is to protect its position as the arbiter of truth.

Social media glossary

ACCOUNT: A person, group or organisation that has created a profile on a social media platform for the purpose of creating content for, sharing content with and/or receiving content from others on the same social media platform.

HANDLE: The name of a social media account, prefixed with the @ symbol on Twitter.

HASHTAG: A word or phrase preceded by the '#' symbol. Hashtags are a simple way to mark the topic (or topics) of social media messages and to make them discoverable to people with shared interests. On most social networks, clicking a hashtag will reveal all the public and recently published messages that also contain the hashtag.

FEED: The social media content format that provides users with a steady stream of information, including posts, comments and content from advertisers.

FOLLOWER: On Twitter, someone who chooses to link their account to another Twitter account and by virtue of that link, sees all the tweets and retweets of the linked account in their feed.

FRIEND: On Facebook, a two-way connection between two accounts requiring both accounts to endorse the relationship.

MENTION: A Twitter message or post containing a '@username' to alert that user to the message. A mention can also refer to the insertion of a hyperlink to a web page or electronic document in a message or in a post to allow the reader of the message to access the page or document by clicking on the hyperlink.

POST: Content created in the form of text, image, audio, video or a combination of any of these. For example, a meme is the combination of an image and text. A post is created by an account that may be controlled by another human or by a machine (a bot).

RETWEET: A tweet originated by a different Twitter account and re-shared to the followers of the Twitter account.

TAG: Tagging is a social media functionality, most often used on Facebook and Instagram, that lets users link back to the profile of the person shown in a photograph or targeted by the update.

TWEET: A message of limited character length created and shared on the social media platform Twitter. A tweet may be accompanied by or consist solely of an image or video. A tweet is created by an account that may be controlled by another human or by a machine (a bot).

TWEETER: The Twitter account that creates a message on Twitter.

Appendix 1: Supporting datasets

Table 23: DOIs for datasets on 'vaccination and autism' harvested from three research data repositories, 14-17 February 2017

Source	Date	DOI	Long reference
Dataverse	14.02.2017	doi:10.7910/DVN/C85WNZ	Simonyan, Vahan, 2015, 'Deep sequencing for evaluation of genetic stability of influenza A/California/07/2009 (H1N1) vaccine viruses', doi:10.7910/DVN/C85WNZ, Harvard Dataverse, V1; California-2009-pandemic-vaccine-viruses.z02 [fileName]
Dataverse	14.02.2017	doi:10.7910/DVN/26370	Glassman, Amanda; Zoloa, Juan Ignacio; Duran, Denizhan, 2014, 'A Commitment to Vaccination Index: Measuring Government Progress toward Global Immunization', doi:10.7910/DVN/26370, Harvard Dataverse, V1
Dataverse	14.02.2017	hdl:1902.1/15088	Abhijit Banerjee, Esther Duflo, Rachel Glennerster, Dhruva Kothari, 2010, 'Improving Immunization Coverage in Rural India', hdl:1902.1/15088, Harvard Dataverse, V4
Dataverse	14.02.2017	doi:10.7910/DVN/XNUYPB	Raj, Anita, 2017, 'Ananya', doi:10.7910/DVN/XNUYPB, Harvard Dataverse, V2, UNF:6:sJMyVahXyHdSgcXCOKiP0A==
Dataverse	14.02.2017	hdl:1902.1/CP2009P2009-03	Field Research Corporation., 2011, 'California Poll: 2009-03 --September 18 - October 5, 2009', hdl:1902.1/CP2009P2009-03, Harvard Dataverse, V3
Dataverse	14.02.2017	doi:10.7910/DVN/P04NKK	Dizon-Ross, Rebecca; Dupas, Pascaline; Robinson, Jonathan, 2016, 'Governance and Effectiveness of Public Health Subsidies', doi:10.7910/DVN/P04NKK, Harvard Dataverse, V1, UNF:6:FqKFNFxZYDV4feiL59dmjw==
Dataverse	14.02.2017	doi:10.7910/DVN/FF4DI7	Richard Gershon, 2016, 'NIH Toolbox Norming Study', doi:10.7910/DVN/FF4DI7, Harvard Dataverse, V3, UNF:6:bTWnBDugEcmwBgMfC5TIKg==
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.t55gq	Croucher NJ, Finkelstein JA, Pelton SI, Parkhill J, Bentley SD, Hanage WP, Lipsitch M (2015) Data from: Population genomic datasets describing the post-vaccine evolutionary epidemiology of Streptococcus pneumoniae. Scientific Data http://dx.doi.org/10.5061/dryad.t55gq
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.db2dd	Kinnear CL, Strugnell RA (2015) Data from: Vaccination method affects immune response and bacterial growth but not protection in the Salmonella Typhimurium animal model of typhoid.PLOS ONE http://dx.doi.org/10.5061/dryad.db2dd
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.58q00	Magpantay FMG, Domenech de Cellès M, Rohani P, King AA (2015) Data from: Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity. Parasitology http://dx.doi.org/10.5061/dryad.58q00
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.tb084	Hattingh HL, Sim TF, Parsons R, Czarniak P, Vickery A, Ayadurai S (2016) Data from: Evaluation of the first pharmacist-administered vaccinations in Western Australia: a mixed-methods study. BMJ Open http://dx.doi.org/10.5061/dryad.tb084
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.rn931	Atkins K, van Hoek AJ, Watson C, Baguelin M, Choga L, Patel A, Raj T, Jit M, Griffiths U (2016) Data from: Seasonal influenza vaccination delivery through community pharmacists in England: evaluation of the London pilot. BMJ Open http://dx.doi.org/10.5061/dryad.rn931
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.2pm60	Steckelberg A, Albrecht M, Kezle A, Kasper J, Mühlhauser I (2013) Data from: Impact of numerical information on 'risk knowledge' regarding human papillomavirus (HPV) vaccination among schoolgirls: a randomised controlled trial. GMS German Medical Science http://dx.doi.org/10.5061/dryad.2pm60
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.779rd	Gachohi JM, Njenga MK, Kitala PM, Bett BK (2016) Data from: Modelling vaccination strategies against Rift Valley fever in livestock in Kenya: model code. PLOS Neglected Tropical Diseases http://dx.doi.org/10.5061/dryad.779rd
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.t2f47	Vermeer-de Bondt PE, Schoffelen T, Vanrolleghem AM, Isken LD, van Deuren M, Sturkenboom MCJM, Timen A (2015) Data from: Coverage of the 2011 Q fever vaccination campaign in the Netherlands, using retrospective population-based prevalence estimation of cardiovascular risk-conditions for chronic Q fever. PLOS ONE http://dx.doi.org/10.5061/dryad.t2f47
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.8b063	Woodroffe R, Donnelly CA, Ham C, Jackson SYB, Moyes K, Chapman K, Stratton NG, Cartwright SJ (2016) Data from: Ranging behaviour of badgers Meles meles vaccinated with Bacillus Calmette Guerin. Journal of Applied Ecology http://dx.doi.org/10.5061/dryad.8b063
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.4nn26	Dunne EM, Lehmann D (2016) Data from: Aboriginal and non-Aboriginal children in Western Australia carry different serotypes of pneumococci with different antimicrobial susceptibility profiles. Pneumonia http://dx.doi.org/10.5061/dryad.4nn26
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.33380	Dey A, Molodecky NA, Verma H, Sharma P, Yang JS, Saletti G, Ahmad M, Bahl SK, Wierzba TF, Nandy RK, Deshpande JM, Sutter RW, Czerkinsky C (2016) Data from: Human circulating antibody-producing B cell as a predictive measure of mucosal immunity to poliovirus. PLOS ONE http://dx.doi.org/10.5061/dryad.33380
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.np7n3	Wieten RW, Jonker EFF, van Leeuwen EMM, Remmerswaal EBM, ten Berge IJM, de Visser AW, van Genderen PJJ, Goorhuis A, Visser LG, Grobusch MP, de Bree GJ (2016) Data from: A single 17D Yellow Fever vaccination provides lifelong

			immunity; characterization of Yellow-Fever-specific neutralizing antibody and T-cell responses after vaccination. PLOS ONE http://dx.doi.org/10.5061/dryad.np7n3
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.h44t1	Barclay VC, Sim D, Chan BHK, Nell LA, Rabaa MA, Bell AS, Anders RF, Read AF (2012) Data from: The evolutionary consequences of blood-stage vaccination on the rodent malaria Plasmodium chabaudi. PLOS Biology http://dx.doi.org/10.5061/dryad.h44t1
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.k44f0	Hassan A, Naz A, Obaid A, Paracha RZ, Naz K, Awan FM, Muhammad SA, Janjua HA, Ahmad J, Ali A (2016) Data from: Pangenome and immuno-proteomics analysis of Acinetobacter baumannii strains revealed the core peptide vaccine targets. BMC Genomics http://dx.doi.org/10.5061/dryad.k44f0
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.sk64k	Parlane NA, Shu D, Subharat S, Wedlock DN, Rehm BHA, De Lisle GW, Buddle BM (2014) Data from: Revaccination of cattle with Bacille Calmette-Guérin two years after first vaccination when immunity has waned, boosted protection against challenge with Mycobacterium bovis. PLoS ONE http://dx.doi.org/10.5061/dryad.sk64k
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.r61r8	Reither K, Katsoulis L, Beattie T, Gardiner N, Lenz N, Said K, Mfinanga E, Pohl C, Fielding KL, Jeffery H, Kagina BM, Hughes EJ, Scriba TJ, Hanekom WA, Hoff ST, Bang P, Kromann I, Daubenberger C, Andersen P, Churchyard GJ (2014) Data from: Safety and immunogenicity of H1/IC31®, an adjuvanted TB subunit vaccine, in HIV-infected adults with CD4+ Lymphocyte counts greater than 350 cells/mm3: a phase II, multi-centre, double-blind, randomized, placebo-controlled trial. PLOS ONE http://dx.doi.org/10.5061/dryad.r61r8
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.92p46	Gunning CE, Erhardt E, Wearing HJ (2014) Data from: Conserved patterns of incomplete reporting in pre-vaccine era childhood diseases. Proceedings of the Royal Society B http://dx.doi.org/10.5061/dryad.92p46
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.tt29d	Bunyasi EW, Tameris M, Geldenhuys H, Schmidt B, Luabeya AKK, Mulenga H, Scriba TJ, Hanekom WA, Mahomed H, McShane H, Hatherill M (2015) Data from: Evaluation of Xpert® MTB/RIF assay in induced sputum and gastric lavage samples from young children with suspected tuberculosis from the MVA85A TB vaccine trial. PLOS ONE http://dx.doi.org/10.5061/dryad.tt29d
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.5v6qm	Kandasamy R, Gurung M, Thapa A, Ndimah S, Adhikari N, Murdoch DR, Kelly DF, Waldron DE, Gould KA, Thorson S, Shrestha S, Hinds J, Pollard AJ (2015) Data from: Multi-serotype pneumococcal nasopharyngeal carriage prevalence in vaccine naïve Nepalese children, assessed using molecular serotyping. PLOS ONE http://dx.doi.org/10.5061/dryad.5v6qm
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.3cn95	Bilinski AM, Fitzpatrick MC, Rupprecht CE, Paltiel AD, Galvani AP (2016) Data from: Optimal frequency of rabies vaccination campaigns in Sub-Saharan Africa. Proceedings of the Royal Society B http://dx.doi.org/10.5061/dryad.3cn95
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.gr40r	Reynolds JJH, Hirsch BT, Gehrt SD, Craft ME (2015) Data from: Raccoon contact networks predict seasonal susceptibility to rabies outbreaks and limitations of vaccination. Journal of Animal Ecology http://dx.doi.org/10.5061/dryad.gr40r
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.4tn48	Read AF, Baigent SJ, Powers C, Kgosana LB, Blackwell L, Smith LP, Kennedy DA, Walkden-Brown SW, Nair VK (2015) Data from: Imperfect vaccination can enhance the transmission of highly virulent pathogens. PLOS Biology http://dx.doi.org/10.5061/dryad.4tn48
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.rt214	Ramos R, Garnier R, González-Solís J, Boulinier T (2014) Data from: Long antibody persistence and transgenerational transfer of immunity in a long-lived vertebrate. The American Naturalist http://dx.doi.org/10.5061/dryad.rt214
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.p534s	Velando A, Beamonte-Barrientos R, Torres R (2014) Data from: Enhanced male coloration after immune challenge increases reproductive potential. Journal of Evolutionary Biology http://dx.doi.org/10.5061/dryad.p534s
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.979	Moser JM, Carbone I, Arasu P, Gibson G (2007) Data from: Impact of population structure on genetic diversity of a potential vaccine target in the canine hookworm (Ancylostoma caninum). The Journal of Parasitology http://dx.doi.org/10.5061/dryad.979
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.pn8mn	França CT, Hostetler JB, Sharma S, White MT, Lin E, Kiniboro B, Waltmann A, Darcy AW, Li Wai Suen CS, Siba P, King CL, Rayner JC, Fairhurst RM, Mueller I (2016) Data from: An antibody screen of a Plasmodium vivax antigen library identifies novel merozoite proteins associated with clinical protection. PLOS Neglected Tropical Diseases http://dx.doi.org/10.5061/dryad.pn8mn
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.vm36c	Greene JM, Dash P, Roy S, McMurtrey C, Awad W, Reed JS, Hammond KB, Abdulhaqq S, Wu HL, Burwitz BJ, Roth BF, Morrow DW, Ford JC, Xu G, Bae JY, Crank H, Legasse AW, Dang TH, Greenaway HY, Kurniawan M, Gold MC, Harriff MJ, Lewinsohn DA, Park BS, Axthelm MK, Stanton JJ, Hansen SG, Picker LJ, Venturi V, Hildebrand W, Thomas PG, Lewinsohn DM, Adams EJ, Sacha JB (2016) Data from: MR1-restricted mucosal-associated invariant T (MAIT) cells respond to mycobacterial vaccination and infection in nonhuman primates. Mucosal Immunology http://dx.doi.org/10.5061/dryad.vm36c
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.66h5g23t	Duenas-Decamp MJ, O'Connell OJ, Corti D, Zolla-Pazner S, Clapham PR (2012) Data from: The W100 pocket on HIV-1 gp120 penetrated by b12 is not a target for other CD4bs monoclonal antibodies. Retrovirology http://dx.doi.org/10.5061/dryad.66h5g23t
Dryad	14.02.2017	http://dx.doi.org/10.5061/dryad.gq8pk	Kilongosi MW, Budambula V, Lihana R, Musumba FO, Nyamache AK, Budambula NLM, Ahmed AA, Ouma C, Were T (2015) Data from: Hepatitis B virus sero-profiles and genotypes in HIV-1 infected and uninfected injection and non-injection drug users from coastal Kenya. BMC Infectious Diseases http://dx.doi.org/10.5061/dryad.gq8pk

Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.1092545.v2	Phillips, Andrew (2014): Potential future impact of a partially effective HIV vaccine in a southern African setting. Data set of simulated outcomes. figshare. https://doi.org/10.6084/m9.figshare.1092545.v2 Retrieved: 09 16, Feb 17, 2017 (GMT)
Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.3085885.v1	Bélisle Pipon, Jean-Christophe; Ringuette, Louise (2016): Review of the Clinical Studies Used for Marketing Approval of the Vaccine Bexsero®. figshare. https://doi.org/10.6084/m9.figshare.3085885.v1 Retrieved: 12 37, Feb 17, 2017 (GMT)
Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.1418309.v1	Stahl, Randal (2015): MS TIC Peak Area Data for 17 Compounds in Vaccinated and Unvaccinated Male WTD. figshare. https://doi.org/10.6084/m9.figshare.1418309.v1 Retrieved: 12 50, Feb 17, 2017 (GMT)
Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.1271925.v2	Mbongue, Jacques; Nicholas, Dequina (2014): First Round of DC+ and vaccine Sequest 3752 peptides. figshare. https://doi.org/10.6084/m9.figshare.1271925.v2 Retrieved: 14 06, Feb 17, 2017 (GMT)
Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.1418313.v1	Stahl, Randal (2015): XCMS results for male WTD, vaccinated, prior to challenge and 5 months post BTb challenge.. figshare. https://doi.org/10.6084/m9.figshare.1418313.v1 Retrieved: 14 08, Feb 17, 2017 (GMT)
Figshare	17.02.2017	https://doi.org/10.6084/m9.figshare.2062575.v1	Cohen, Noah (2016): Rhodococcus equi eBeam Vaccine ELISA Data. figshare. https://doi.org/10.6084/m9.figshare.2062575.v1 Retrieved: 14 08, Feb 17, 2017 (GMT)

Figure 46: Process for harvesting DOIs for research datasets on autism and vaccination

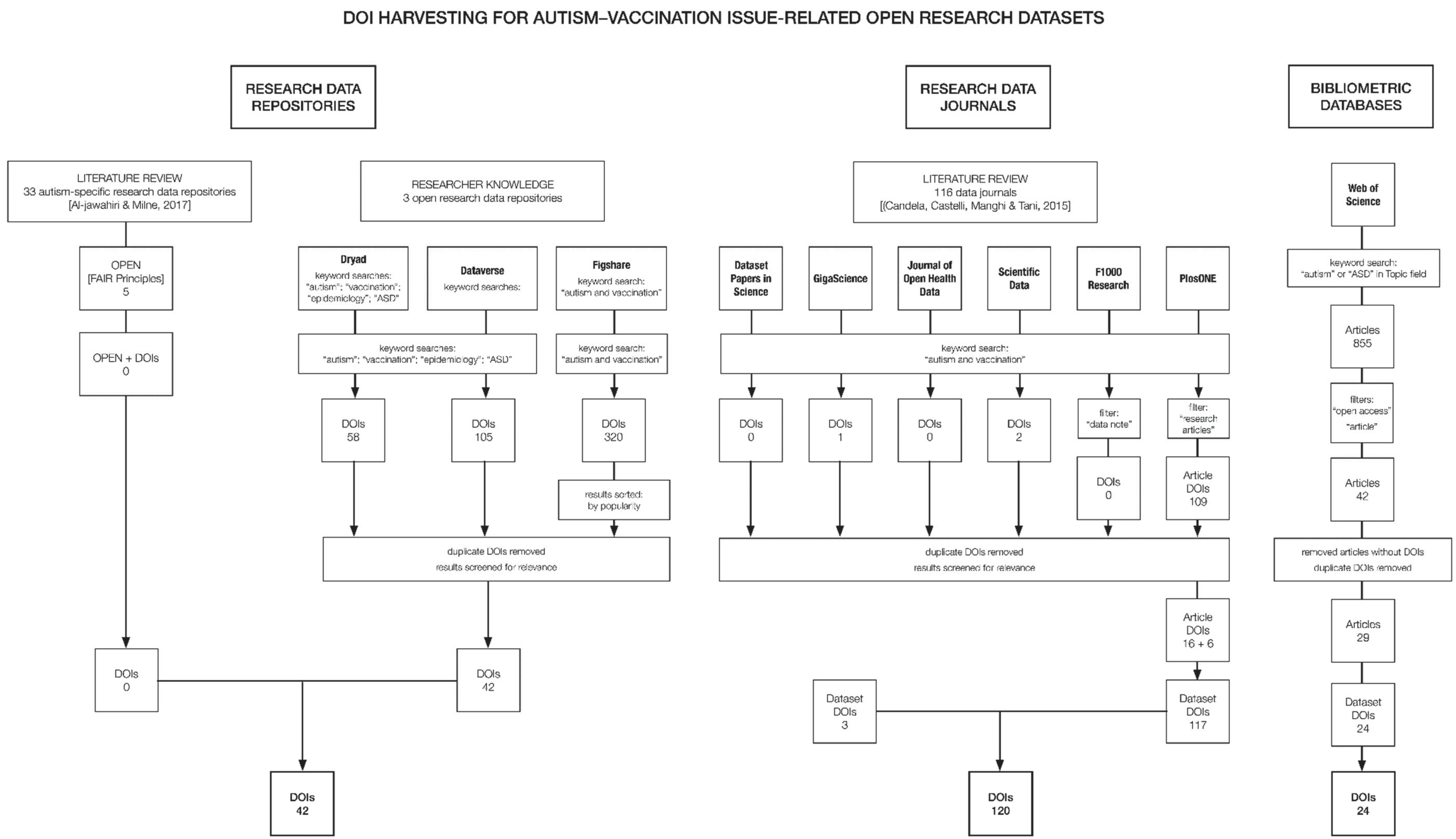


Table 24: Open access journal articles referenced by publics: Altmetric Attention Scores and members of the public tweeting (n=75, 29 June 2017)

Article DOI	Altmetric score	Twitter: Members of the public
10.1001/jama.2015.3077	3,641	2,235
10.1016/j.vaccine.2014.04.085	2,827	2,037
10.1080/15287394.2011.573736	1,156	1,248
10.1186/2047-9158-3-16	1,044	830
10.1186/2047-9158-2-25	965	1,092
10.1001/jamapediatrics.2016.3609	814	263
10.1073/pnas.1500968112	667	458
10.1056/NEJMoa021134	646	258
10.1016/j.jinorgbio.2011.08.008	644	542
10.1371/journal.pone.0146797	376	141
10.1111/j.1469-7610.2005.01425.x	283	168
10.1007/s00401-016-1629-y	279	302
10.1001/jamapediatrics.2015.0418	259	248
10.1542/peds.2009-2489	240	51
10.3390/e14112227	202	131
10.1016/j.vaccine.2015.12.067	177	142
PMID: 15585776	159	31
10.1155/2013/801517	149	49
10.3238/arztebl.2011.0099	145	93
10.1371/journal.pone.0027897	112	79
10.1016/j.vaccine.2011.09.124	112	76
10.1371/journal.pone.0058058	111	74
10.1542/peds.2010-2989	90	42
10.1177/0300060517693423	85	76
10.1371/journal.pone.0068444	79	9
10.1054/mehy.2000.1281	77	25
10.1192/bjp.bp.108.059345	61	10
PMID: 20628439	60	13
10.1186/1742-2094-7-20	56	25
PMID: 23118818	53	50
10.2310/7010.2006.00032	50	8
10.1093/brain/124.9.1821	47	29
10.1136/bmj.322.7284.460	46	13
10.1542/peds.113.5.e472	45	6
10.5897/JPHE2014.0649	42	36
10.1289/ehp.9120	41	3
10.2105/AJPH.94.6.985	40	0
10.1186/1742-2094-10-46	32	18
10.1039/c3em00374d	32	4
10.1038/sj.mp.4001529	29	10
10.1016/j.healthplace.2008.02.001	28	6
PMID: 14976450	25	15
10.1176/appi.ajp.2010.10020223	25	1
10.1038/sj.mp.4001177	22	5
10.1289/ehp.0901713	22	5
10.1093/toxsci/kfg126	19	6
PMID: 23400264	19	1
10.1016/S0165-5728(98)00021-6	17	0
10.1177/0961203316629558	14	16
10.1155/2012/190930	13	5

PMID: 22791642	10	4
10.1371/journal.pone.0056927	10	2
10.1055/s-2006-924577	9	4
10.1289/ehp.8881	7	4
10.1093/oxfordjournals.aje.a116479	7	1
10.1093/hmg/ddq307	6	0
10.1371/journal.pone.0049767	4	5
10.1111/j.1651-2227.2010.01883.x	4	3
10.1016/j.chemosphere.2016.02.092	3	3
10.1016/S0264-410X(02)00165-2	3	3
10.1542/peds.2010-3481	3	3
10.15537/smj.2016.4.13611	3	2
10.4172/2167-7689.S12-001	3	0
10.1016/j.jneuroim.2015.08.019	2	2
10.3844/ajbbsp.2008.167.176	2	2
10.3844/ajbbsp.2008.208.217	2	2
10.4172/1745-7580.1000069	2	0
10.1016/j.jtemb.2012.11.001	1	0
10.1093/infdis/jiw058	1	0
10.1016/S0161-813X(01)00067-5	0	0
10.1177/039463201002300406	0	0
10.1289/ehp.7712	0	0
10.15761/JTS.1000186	0	0
10.15761/JTS.1000187	0	0
10.1588/medver.2008.05.00182	0	0
Median	**32**	**8**
Average	**217**	**147**

Table 25: Comparison of Lippmannian Device and SBC4D application using Twitter account URLs in batch 1.1 (for all results that return a score of more than 1)

URL of Twitter account	Lipmanian Device score	SBC4D application score
http://twitter.com/aspiritcan	26	10
http://twitter.com/eTweeetz	26	8
http://twitter.com/LaLaRueFrench75	24	18
http://twitter.com/debnantz	17	11
http://twitter.com/itsmepanda1	14	4
http://twitter.com/PharmaNemesis	14	2
http://twitter.com/nocompulsoryvac	12	0
http://twitter.com/peakdavid	11	17
http://twitter.com/joegooding	10	0
http://twitter.com/CplBart	8	16
http://twitter.com/Vbalance03	8	12
http://twitter.com/wrinklyrebel	6	4
http://twitter.com/DocMeehan	6	2
http://twitter.com/VaxCalc	5	16
http://twitter.com/MaryJo__Perry	5	6
http://twitter.com/itsbaxter	5	1
http://twitter.com/AKruckel	3	36
http://twitter.com/jackiesebell	3	10
http://twitter.com/dmilat68	3	2
http://twitter.com/VaxChoiceVT	3	2
http://twitter.com/OregonVacTruth	2	220
http://twitter.com/QandAMamma	2	33
http://twitter.com/_bhickman	2	12
http://twitter.com/RonPaul4Constit	2	6
http://twitter.com/DaWeiLeigh	2	2
http://twitter.com/gmawarrior	2	2
http://twitter.com/xileenie	2	2
http://twitter.com/4thAnon	2	0
http://twitter.com/Angelapaints	2	0
http://twitter.com/DianeTriple7	2	0
http://twitter.com/libertylives277	2	0
http://twitter.com/theEcoNEST	2	0
http://twitter.com/VACTRUTH	1	335
http://twitter.com/DavidsAutismDad	1	6
http://twitter.com/iloveecopsych	1	5
http://twitter.com/gramps97	1	0
http://twitter.com/Just4TheCause	0	8
http://twitter.com/DerBHat	0	4
http://twitter.com/Anwar_Hashem	0	2

Table 26: Number and proportion of anti-vaccination users/accounts per journal article on social media on Twitter

Stance determined by manual checking of each Twitter account*

	TOP-DOWN, PURPOSIVE SAMPLE				
	1.1	1.2	1.3	1.4	1.5
Unique accounts	310	123	58	81	25
Number of verified anti-vaccination accounts	36	1	1	0	4
% anti-vaccination accounts	11.6%	0.8%	1.7%	0%	16.0%

* Sample 1.1 calculations based on use of the Lippmannian Device.

Stance determined using SBC4D crawler: https://dev.sbc4d.com/cdv/fsv/geturl.php

	BOTTOM-UP, SNOWBALL SAMPLE				
	2.1	2.2	2.3	2.4	2.5
Tweets	3551	3509	2589*	2268	2282
Duplicate accounts in sample	364	733**	1021	1337	885
Unique accounts	3187	2775	1567	931	1397
Known anti-vaccination accounts from previous batch	24	12	118	25	20
Known pro-vaccination accounts from previous batch	27	14	0	2	0
Unique accounts crawled	3136	2749	1449	904	1377
Number of verified anti-vaccination accounts	35	40	218	166	157
% anti-vaccination accounts	1.1%	1.4%	13.9%	17.8%	11.2%
Number of false positive accounts	62	71	2	19	1

* of which 2 Twitter accounts that account for 3 tweets no longer exist
** of which 1 Twitter account no longer exists

Table 27: Number and proportion of anti-vaccination tweets

	TOP-DOWN, PURPOSIVE SAMPLE				
	1.1	1.2	1.3	1.4	1.5
Total no. of tweets	382	131	69	93	25
No. of anti-vaccination tweets	52	1	1	0	4
% of anti-vaccination tweets	13.6%	0.8%	1.5%	0%	16.0%
	BOTTOM-UP, SNOWBALL SAMPLE				
	2.1	2.2	2.3	2.4	2.5
Total no. of tweets	3551	3509	2589	2268	2282
No. of anti-vaccination tweets	45	21	812	672	545
% of anti-vaccination tweets	1.3%	0.6%	31.4%	29.6%	23.9%

Table 28: Indication of stance vis-à-vis vaccination by article title and abstract

Article ref.	Title	Indicative stance: Title	Findings (from abstract)	Indicative stance: Findings
1.1	Imperfect Vaccination Can Enhance the Transmission of Highly Virulent Pathogens	ANTI-VAC	Our data show that anti-disease vaccines that do not prevent transmission can create conditions that promote the emergence of pathogen strains that cause more severe disease in unvaccinated hosts.	ANTI-VAC
1.2	Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years: Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012	NEUTRAL	For 2012, the combined estimated prevalence of ASD among the 11 ADDM Network sites was 14.6 per 1,000 (one in 68) children aged 8 years. combined ASD prevalence estimates were similar for 2010 and 2012, including in subgroups defined by sex and race/ethnicity.	PRO-VAC
1.3	Lack of association between measles virus vaccine and autism with enteropathy: a case-control study	PRO-VAC	We found no differences between case and control groups in the presence of MV RNA in ileum and cecum. Results were consistent across the three laboratory sites. GI symptom and autism onset were unrelated to MMR timing.	PRO-VAC
1.4	GWATCH: a web platform for automated gene association discovery analysis	NEUTRAL	Here we present a dynamic web-based platform – GWATCH – that automates and facilitates four steps in genetic epidemiological discovery.	NEUTRAL
1.5	The Evolutionary Consequences of Blood-Stage Vaccination on the Rodent Malaria Plasmodium chabaudi	ANTI-VAC	virulence evolved; AMA-1-selected parasites induced greater anemia in naïve mice than both control and ancestral parasites. Our data suggest that recombinant blood stage malaria vaccines can drive the evolution of more virulent malaria parasites	ANTI-VAC
2.1	Autism occurrence by MMR vaccine status among US children with older siblings with and without autism	NEUTRAL	receipt of the MMR vaccine was not associated with increased risk of ASD, regardless of whether older siblings had ASD. These findings indicate no harmful association between MMR vaccine receipt and ASD even among children already at higher risk for ASD	PRO-VAC
2.2	Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies	PRO-VAC	Findings of this meta-analysis suggest that vaccinations are not associated with the development of autism or autism spectrum disorder	PRO-VAC
2.3	A Positive Association found between Autism Prevalence and Childhood Vaccination uptake across the U.S. Population	ANTI-VAC	The results suggest that although mercury has been removed from many vaccines, other culprits may link vaccines to autism.	ANTI-VAC
2.4	Measles-mumps-rubella vaccination timing and autism among young African American boys: a reanalysis of CDC data	ANTI-VAC	The present study provides new epidemiologic evidence showing that African American males receiving the MMR vaccine prior to 24 months of age or 36 months of age are more likely to receive an autism diagnosis.	ANTI-VAC
2.5	A two-phase study evaluating the relationship between Thimerosal-containing vaccine administration and the risk for an autism spectrum disorder diagnosis in the United States	ANTI-VAC	the present study provides new epidemiological evidence supporting an association between increasing organic-Hg exposure from Thimerosal-containing childhood vaccines and the subsequent risk of an ASD diagnosis.	ANTI-VAC

Key: red = anti-vaccination; orange = neutral; green = pro-vaccination stance.

Appendix 2: Data collection and analysis processes for Twitter and Facebook

2.1 Process for determining anti-vaccination Twitter users

A. Download the Twitter data for the relevant article from Altmetric's Data Explorer.

B. Open the file in MS Excel.
Delete all columns except for the columns: Mention type, Mention date, Outlet or author, Mention URL.
Sort all data by (1) Mention type (A-Z) and (2) Date (newest to oldest).
Delete all rows containing non-Twitter data.
Check for empty cells in the 'Outlook or author' column. If there are any empty cells, use the URL in the adjacent 'Mention URL' column to check whether the account is still active. If it is no longer active, indicate 'Account no longer exists' in the 'Outlook or author' column. If the account exists or has been renamed, capture the missing/new account name (Twitter handle) in the 'Outlook or author' column.
Add new column 'Twitter account URL'.
Copy and paste all Twitter handles from the 'Outlet or author' column into the new column.
Search and replace the new column only: Replace '@' with 'https://twitter.com/'.
Record the number of Tweets.
Save file as '[batch number]_B ______ TWITTER PREP' in Microsoft Excel format.

C. Open file [batch number]_B and save as '[batch number]_CD ________ TWITTER UNIQUE ACTORS'.
Create duplicate worksheet and rename it 'ACTORS DUPLICATES'.
Delete column 'Mention type'.
Delete all 'Account no longer exists' rows.
Select all columns and apply the Conditional Formatting function: Highlight cell rules > Duplicate values.

D. Create a new worksheet called 'ACTORS UNIQUE'.
Apply the Remove Duplicates function to the data in the worksheet 'ACTORS UNIQUE' by selecting the duplicates to be removed in the column 'Outlets or author'.
Record the number of duplicates.
Calculate and record the number of unique users.
Insert the batch number in each row of the dataset.
Create a new worksheet 'KNOWN STANCE'.
Copy Twitter account URLs of known vaccination stance Twitter accounts from previous batches into the worksheet 'KNOWN STANCE'.
Copy the known anti-vaccination accounts at the bottom of the column 'Twitter account URL' in the worksheet 'ACTORS UNIQUE'. Apply Conditional Formatting > Highlight cell rules > Duplicate values.
Insert the text 'anti-vax' in Column E next to the duplicate accounts. Delete the known accounts that were added to the bottom of the column 'Twitter account URL'.
Repeat the process for known pro-vaccination accounts and insert the text 'pro-vax' for those duplicate accounts.
Record the number of known 'anti-vax' and 'pro-vax' accounts.

E. Create a new Excel document.
Copy and paste the unknown values from 'Outlet and authors' column in the document '[batch number]_CD _______ TWITTER UNIQUE ACTORS' into the new document.
Record the number of records/accounts.
Save as a csv file with file name '[batch number]_E __________ TWITTER CRAWLER INPUT'.

F. Open the URL https://dev.sbc4d.com/cdv/fsv/geturl.php in a web browser.
Select 'Choose file' and select the csv file.
Enter the keywords as follows:
antivax::::vaxxed::::vaccineinjur::::vaxfax::::vaccinesafety::::informedconsent::::vactruth
Click 'Submit' and wait for the crawl to complete.
Save the crawl results as '[batch]_F ________ CRAWLER EXPORT RESULT TWITTER'.
Record keywords and date of crawl in the file.

G. Open Excel documents 'CD' and 'F'.
In file F: Calculate the total score for each user by adding up the frequency score for each of the keywords queried by the crawler.
In file CD: Insert two new columns after the 'Twitter account URL' column.
Copy both the Twitter account URLs and total scores from file F to file G.
Insert new Column E and enter the function 'Column C=Column D' to verify that the data inserted from F is in the same order as the data in G. All results in new column E should = TRUE.
Sort crawler data only by 'Twitter crawler score' (largest to smallest), and 'Mention date' (newest to oldest).
Save file as '[batch]_G __________ TWITTER ACCOUNT ANALYSIS'.

H. Open file G. Delete column 'Mention URL'. Add new columns 'Language', 'Website' and 'Notes'.
Manually confirm the anti-vaccination status of each Twitter account with a crawler score of 1 or more by: (1) opening the account in a web browser, (2) reading the account banner image, bio text and (if necessary) the latest 20 tweets (using Google Translate for non-English accounts), (3) recording the stance as 'anti-vax', 'pro-vax', 'neutral', 'unknown', (4) recording the main language of the account, (5) recording the web address provided in the bio (if any), giving preference to Facebook web addresses, and (6) providing notes to substantiate the recorded stance.
On completion, sort data by 'STANCE'.
Record number of anti-vax, pro-vax, neutral and unknown accounts in Table 15.
Save file as '[batch]_H __________ ACCOUNT ANALYSIS COMPLETE'.

2.2 Process for determining number of tweets by anti-vaccination accounts

1. Open MS Excel files B and H.
2. Create new sheet 'ANTI-VAX' in B. Select all anti-vaccination accounts in H. Copy and paste from file H into file B.
3. To identify all the values in the 'Twitter account URL' column containing the account URLs for ALL tweets that match the known anti-vaccination accounts listed in the ANTI-VAX worksheet: In file B, insert a new column E named 'Match anti-vax'. Insert and apply the following function to all cells in the column: =IF(ISNUMBER(MATCH(D2,'Anti-vaxx'!A:A,0)),'Yes','No')
4. Sort by 'Match anti-vax' (z-a) and 'Mention date' (newest to oldest).
5. Record number of tweets by anti-vaccination accounts.
6. Save as '[batch]_Q __________ TWEETS'.

2.3 Process for measuring level of engagement using tweets

1. Open MS Excel file Q.
2. Add rows for 'Article title', 'Article abstract', 'Article URL', 'Article DOI', 'Article Altmetric URL', and add content for article.
3. Add new columns: 'Random number', 'Level of engagement', 'Tweet copy', 'Notes', 'Retweeted by'.
4. In the Random number column, add and apply to all cells =RAND(). Copy random numbers and paste back into the same column as values. Sort all data by the Random number column from largest to smallest.
5. Number all mentions from 1 onwards. Select and highlight the first 100 mentions.

6. Open the mention URL in Twitter. Copy and paste the content of the Tweet into the Tweet copy column. Capture the level of engagement score in the relevant column and provide a justification for the score in the Notes column. If a tweet was a retweet of a previous tweet, capture the Twitter handle of the original tweeter in the Retweeted by column.

2.4 Process for determining anti-vaccination Facebook accounts

1. Download the Twitter data for the relevant article from Altmetric's Data Explorer.
2. Open the file in MS Excel.
   Delete all columns except for the columns: Mention type, Mention date, Outlet or author, Mention Title, Mention URL.
   Sort all data by (1) Mention type (A-Z) and (2) Date (newest to oldest)
   Delete all rows containing non-Facebook data.
   Record the number of Facebook posts.
   Save file as '2.3 Altmetric – Mentions - FB PREP' in Microsoft Excel format.
3. Open file 2.3 Altmetric – Mentions – FB PREP and save as '2.3 B Altmetric – Mentions – STANCE'.
   Delete all rows for which the 'Outlet or Author' field is blank.
   Select all columns and apply the Conditional Formatting function: Highlight cell rules > Duplicate values.
4. Create duplicate worksheet and rename it 'ACTORS UNIQUE'.
   Apply the Remove Duplicates function to the data in the worksheet 'ACTORS UNIQUE' by selecting the duplicates to be removed in the column 'Outlet or Author'.
   Record the number of duplicates.
   Calculate and record the number of unique users.
5. Duplicate worksheet ACTORS UNIQUE and name the new worksheet 'STANCE'.
   Create a new column 'Stance'. Capture in this column the codes 'anti', 'pro', 'neutral', 'unknown' as the analysis of the Facebook accounts progresses. When complete, sort data by Stance (a-z).
   Record number of 'anti', 'pro', 'neutral', 'unknown' accounts.
6. Open file B and save it as '2.3 C Altmetric – Mentions – FB LEVEL OF ENGAGEMENT'. Check for non-English accounts and record them as 'anti NON ENG' in the 'Stance column'. On completion, sort data by 'STANCE'.
   Record number of English anti-vax accounts.

2.5 Process for determining number of Facebook posts by anti-vaccination accounts

1. Open MS Excel file B. Rename the worksheet with the original data 'anti POSTS'. Add a new column H 'Anti Outlet or Author'. Copy the anti-vaccination accounts into this new column.
2. To identify all the values in the 'Outlet or Author' column containing the account URLs for ALL Facebook posts that match the known anti-vaccination Facebook accounts:
   a. Delete all the rows for which no 'Outlet or Author' data is available.
   b. Insert a new column F named 'Match anti-vax'.
   c. Insert and apply the following function to all cells in the column: =IF(ISNUMBER(MATCH(C2,H:H,0)),'Yes','No')
3. Sort by 'Match anti-vax' (z-a).
4. **Record number of posts by anti-vaccination accounts.**

2.6 Process for measuring level of engagement on Facebook

1. Open MS Excel file C.
2. Create a copy of worksheet 'anti POSTS' and name the new sheet 'FB Engagement'. Delete all 'no' entries. Add new columns: 'Random number', 'Level of engagement', 'Post copy', 'Notes'.

3. In the Random number column, add and apply to all cells =RAND(). Copy random numbers and paste back into the same column as values. Sort all data by the Random number column from largest to smallest.
4. Number all mentions from 1 onwards. Select and highlight the first 100 mentions.
5. Open the mention URL in Facebook. Copy and paste the content of the Post into the 'Post copy' column. Capture the level of engagement score in the relevant column and provide a justification for the score in the Notes column.

References

Aglionby, J., & Pilling, D. (2018, April 22). Tanzania's bloggers face hefty fees for right to post online. *Financial Times*. Retrieved from https://www.ft.com/content/36098722-4623-11e8-8ae9-4b5ddcca99b3

Aichholzer, G., & Burket, H. (2004). *Public Sector Information in the Digital Age: Between markets, public management and citizens' rights.* Cheltenham: Edward Elgar.

Akers, K. (2014, May 9). A growing list of data journals*. Data@MLibrary*. Retrieved from https://mlibrarydata.wordpress.com/2014/05/09/data-journals/

Akin, H., & Landrum, A. (2017). A recap: Heuristics, biases, values, and other challenges to communicating science. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 455-460). New York: Oxford University Press.

Al-jawahiri, R., & Milne, E. (2017). Resources available for autism research in the big data era: A systematic review. *PeerJ*, 1-21. doi:10.7717/peerj.2880.

Anderson, M., Martinson, B., & De Vries, R. (2007). Normative dissonance in science: Results from a national survey of US scientists. *Journal of Empirical Research on Human Ethics, 2*(4), 3-14.

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). *Proportion of Open Access Papers Published in Peer-Review Journals at the European and World Levels 1996-2013.* Brussels: European Commission. Retrieved from http://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf

Asur, S., Huberman, B., Szabo, G., & Wang, C. (2011). Trends in social media: Persistence and decay. *arXiv*, 1-12. Retrieved from https://arxiv.org/abs/1102.1402

Babbie, E., & Mouton, J. (2001). *The Practice of Social Research (South African edition).* Cape Town: Oxford University Press.

Barberá, P. (2015). How social media reduces mass political polarization: Evidence from Germany, Spain, and US. Unpublished paper. Retrieved from http://rubenson.org/wp-content/uploads/2015/10/barbera-tpbw.pdf

Barbour, V. (2017, February 16). How the insights of the Large Hadron Collider are being made open to everyone. *Cosmos*. Retrieved from https://cosmosmagazine.com/technology/how-the-insights-of-the-large-hadron-collider-are-being-made-open-to-everyone

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social web. *17th International Conference on Science and Technology Indicators* (pp. 98-109). Montreal: Science-Metrix.

Bartling, S., & Friesike, S. (2014). *Opening Science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Heidelberg: Springer Open.

Bean, S. (2011). Emerging and continuing trends in vaccine opposition website content. *Vaccine, 29*, 1874-1880.

Beer, D. (2013). *Popular Culture and New Media.* Basingstoke: Palgrave Macmillan.

Begley, C., & Ellis, L. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*, 531-533. doi:10.1038/483531a.

Belluz, B. (2015, August 5). This is why you shouldn't believe that exciting new medical study*. Vox*. Retrieved from http://www.vox.com/2015/3/23/8264355/research-study-hype

Bennato, D. (2017). The shift from public science communication to public relations: The Vaxxed case. *Journal of Science Communication, 16*(2), 1-11.

Berlin Declaration on Open Access. (2003, October 22). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.* Retrieved from http://openaccess.mpg.de/Berlin-Declaration

Bessi, A., Scala, A., Rossi, L., Zhang, Q., & Quattrociocchi, W. (2014). The economy of attention in the age of (mis)information. *Journal of Trust Management, 1*(1), 1. doi:10.1186/s40493-014-0012-y.

Bimber, B., & Guston, D. (1995). Politics by the same means: Government and science in the United States. In S. Jassanoff, G. Markle, J. Peterson, & T. Pinch, *Handbook of Science and Technology Studies* (pp. 554-571). Thousand Oaks: Sage.

Bishop, D. (2016). Open research practices: Unintended consequences and suggestions for averting them (Commentary on the Peer Reviewers' Openness Initiative). *Royal Society Open Science*. doi:10.1098/rsos.160109.

Blöbaum, B. (2016). Key factors in the process of trust: On the analysis of trust under digital conditions. In B. Blöbaum, *Trust and Communication in a Digitized World: Models and concepts of trust research* (pp. 3-26). Heidelberg: Springer.

Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science, 352*(6285), 508-512. doi:10.1126/science.352.6285.508.

Born, K., & Edgington, N. (2017). *Analysis of Philanthropic Opportunities to Mitigate the Disinformation/Propaganda Problem.* Washington DC: William and Flora Hewlett Foundation.

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics, 8*, 935-950. doi:10.1016/j.joi.2014.09.007.

Bornmann, L., & Marx, W. (2014). How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons. *Scientometrics, 98*(1), 211–219.

Boseley, S. (2018, 21 August). Resurgence of deadly measles blamed on low MMR vaccination rates. *The Guardian*. Retrieved from https://www.theguardian.com/society/2018/aug/20/low-mmr-uptake-blamed-for-surge-in-measles-cases-across-europe

Boyle, J. (2003). The Second Enclosure Movement and the construction of the public domain. *Law and Contemporary Problems, 66*(1/2), 33-74.

Brock, W. (1980). The development of commercial science journals in Victorian Britain. In A. Meadows, *The Development of Science Publishing in Europe* (pp. 95-122). Amsterdam: Elsevier.

Broniatowski, D. A., Jamison, A. M., Qi, S., Al Kulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health,* published online before print. doi:10.2105/AJPH.2018.304567.

Brossard, D. (2013). New media landscapes and the science information consumer. *PNAS, 10*(3), 14096-14101. doi:10.1073/pnas.1212744110.

Bucchi, M. (1996). When scientists turn to the public: Alternative routes in science communication. *Public Understanding of Science, 5*(4), 375-394. doi:10.1088/0963-6625/5/4/005.

Bucchi, M. (2004). Communicating science. In M. Bucchi, *Science in Society: An introduction to social studies of science* (pp. 105-123). London: Routledge.

Bucchi, M. (2014, October 16). The crisis of mediators? Science communication in the digital age. Retrieved from Slideshare: https://www.slideshare.net/mcsunipd/bucchi-40348975

Bucchi, M. (2018). Credibility, expertise and challenges of science communication 2.0. *Public Understanding of Science, 26*(8), 890-893. doi:10.1177/0963662517733368

Budapest Open Access Initiative. (2012, September 12). Ten years on from the Budapest Open Access Initiative: Setting the default to open. Retrieved from http://www.budapestopenaccessinitiative.org/boai-10-recommendations

Burt, R. (2001). Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. Burt, *Social Capital: Theory and research* (pp. 31-56). New York: Aldine De Gruyter.

Çalışkan, K., & Callon, M. (2009). Economization, part 1: Shifting attention from the economy towards processes of economization. *Economy and Society, 38*(3), 369-398.

Cameron, J. (2018, May 22). Controversial short seller Viceroy targets Capitec AGAIN! Read explosive letter to its auditors here. *BizNews*. Retrieved from https://www.biznews.com/sa-investing/2018/05/22/short-seller-viceroy-nails-capitec-again

Candela, C., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology, 66*(9), 1747-1762.

Capocasa, M., Anagnostou, P., D'Abramo, F., Matteucci, G., Dominici, V., Destro Bisol, G., & Rufo, F. (2016). Samples and data accessibility in research biobanks: An explorative survey. *PeerJ, 4:e1613*. doi:10.7717/peerj.1613.

Castells, M. (1996). *The Rise of the Network Society.* Oxford: Blackwell.

Castells, M. (2004). *The Network Society: A cross-cultural perspective.* Cheltenham: Edward Elgar.

Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication, 1*, 238-266. Retrieved from http://ijoc.org/index.php/ijoc/article/viewFile/46/35

Castells, M. (2009). *Communication Power.* Oxford: Blackwell.

Castells, M. (2010). Preface to the 2010 Edition of The Rise of the Network Society. In M. Castells, *The Rise of the Network Society* (Revised ed., pp. i-xliv). Oxford: Blackwell.

Castells, M. (2017). Rethinking development in the global information age. In J. Muller, N. Cloete, & F. van Schalkwyk, *Castells in Africa* (pp. 67-94). Cape Town: African Minds.

CDC. (2013, September 12). CDC telebriefing on the National Immunization Survey, Vaccine for Children Program, and recent measles outbreaks in the US. Retrieved from https://www.cdc.gov/media/releases/2013/t0912_measles-outbreaks-data.html

CDC. (2015, January 29). Transcript for CDC telebriefing: Measles in the United States, 2015. Retrieved from https://www.cdc.gov/media/releases/2015/t0129-measles-in-us.html

Chan, L., & Costa, S. (2005). Participation in the global knowledge commons: Challenges and opportunities for research dissemination in developing countries. *New Library World, 106*, 141-163. doi:10.1108/03074800510587354.

Chan, L., Kirsop, B., & Arunachalam, S. (2011). Towards open and equitable access to research and knowledge for development. *PLOS Medicine, 8*(3), e1001016. doi:10.1371/journal.pmed.1001016.

Chan, M., Jones, C., & Albarracin, D. (2017). Countering false beliefs: An analysis of the evidence and recommendations of best practices for the retracting and correction of scientific information. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 341-349). New York: Oxford University Press.

Christensen, D., Baio, J., Braun, K., Bilder, D., Charles, J., Constantino, J. N., … Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 Years – Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR Surveillance Summaries, 16*(3), pp. 1-23. Retrieved from https://www.cdc.gov/mmwr/volumes/65/ss/pdfs/ss6503.pdf

Collins, R. (2000). *The Sociology of Philosophies: A global theory of intellectual change.* Cambridge, MA: Harvard University Press.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, G., Flammini, A., & Menczer, F. (2011). Political polarization on Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 89-96). Association for the Advancement of Artificial Intelligence.

Costas, R., De Rijcke, S., & Marres, N. (2017). Beyond the dependencies of altmetrics: Conceptualizing 'heterogeneous couplings' between social media and science. Paper presented at *Altmetrics17: The dependencies of altmetrics*. Toronto. Retrieved from Altmetrics: http://altmetrics.org/wp-content/uploads/2017/09/altmetrics17_paper_4-1.pdf

Cronin, B., Snyder, H., Rosenbaum, H., Martinson, A., & Callahan, E. (1998). Invoked on the web. *Journal of the American Society for Information Science, 49*(14), 1391-1328.

Cultural Cognition Project. (no date). Protecting the vaccine science communication environment. Retrieved from http://www.culturalcognition.net/the-vaccine-science-communicat/

Czerniewicz, L. (2016, April 7). How do students access the resources they need? Survey finds only one in five obtain all resources legally. *LSE Impact Blog*. Retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2016/04/07/how-do-students-access-the-resources-they-need/

Daily Maverick. (2018, August 8). Twitter chief defends not booting Infowars. Retrieved from https://www.dailymaverick.co.za/article/2018-08-08-twitter-chief-defends-not-booting-infowars/

Davies, W. (2015, January 12). The data sublime. *The New Inquiry*. Retrieved from http://thenewinquiry.com/essays/the-data-sublime/

Davies, W. (2016, August 24). The age of post-truth politics. *New York Times*. Retrieved from http://mobile.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.html?

Davies, S.R., & Hara, N. (2017). Public science in a wired world: How online media are shaping science communication. *Science Communication, 39*(5), 563-568. doi:10.1177/1075547017736892.

Deer, B. (2011). How the case against the MMR vaccine was fixed. *BMJ, 342*(c5347).

De Filippi, P., & Vieira, M. (2014). The commodification of information commons: The case of cloud computing. *The Columbia Science & Technology Law Review, XVI*, 102-112.

Dickel, S., & Franzen, M. (2016). The 'problem of extension' revisited: New modes of digital participation in science. *Journal of Science Communication, 15*(1), 1-15. doi:10.22323/2.15010206.

Directory of Open Access Journals. (2017, August 7). Directory of open access journals. Retrieved from https://doaj.org/

DiResta, R. (2017a, May 23). There are bots. Look around. *Ribbon Farm*. Retrieved from https://www.ribbonfarm.com/2017/05/23/there-are-bots-look-around/

DiResta, R. (2017b, April 14). This dramatic graph shows how the pro-vaccine movement can win. *Slate*. Retrieved from http://www.slate.com/articles/health_and_science/medical_examiner/2017/04/this_dramatic_graph_shows_how_the_pro_vaccine_movement_can_win.html

DiResta, R., & Lotan, G. (2015, June 8). Anti-vaxxers are using Twitter to manipulate a vaccine bill. *Wired*. Retrieved from https://www.wired.com/2015/06/antivaxxers-influencing-legislation/

DORA (2012, December 6). *San Francisco Declaration on Research Assessment*. Retrieved from https://sfdora.org/read/

Dorta-González, P., & Santana-Jiménez, Y. (2017). Prevalence and citation advantage of gold open access in the subject areas of the Scopus database. *arXiv*, 1-32. doi:1708.06242.

Dredze, M., Wood-Doughty, Z., Quin, S., & Broniatowski, D. (2017). Vaccine opponents' use of Twitter during the 2016 US presidential election: Implications for practice and policy. *Vaccine, 35*, 4670-4672. doi:10.1016/j.vaccine.2017.06.066.

Dreger, A. (2015). *Galileo's Middle Finger: Heretics, activists, and the search for justice in science.* New York: Penguin Press.

Dryad. (2018, 1 24). Frequently asked questions. Retrieved from https://datadryad.org/pages/faq

Eamon, E. (1985). From the secrets of nature to public knowledge: The origins of the concept of openness in science. *Minerva, 23*(3), 321-347.

Economist. (2016, March 19). Taking the online medicine*.* Retrieved from http://www.economist.com/news/science-and-technology/21694990-old-fashioned-ways-reporting-new-discoveries-are-holding-back-medical-research

Eisfeld-Reschke, J., Herb, U., & Wenzlaff, K. (2014). Research funding in open science. In S. Bartling, & S. Friesike, *Opening Science: The evolving guide on how the internet is changing research* (pp. 237-254). Heidelberg: Springer Open.

Elzinga, A., & Jamison, A. (1995). Changing policy agendas in science and technology. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch, *Handbook of Science and Technology Studies* (pp. 572-597). Thousand Oaks: Sage.

European Commission. (2010). *Riding the Wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data.* Brussels: European Commission.

Evans, P. (2005). The New Commons vs. the Second Enclosure Movement: Comments on an emerging agenda for development research. *Studies in Comparative International Development, 40*(2), 85-94.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLOS Biology, 4*(5), e157. doi:10.1371/journal.pbio.0040157.

Fecher, B., & Friesike, S. (2013, May). Open science: Five schools of thought. *RatSWD Working Paper Series*, pp. 1-11.

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLONE ONE, 10*(2), e0118053. doi:10.1371/journal.pone.0118053.

Ferguson, L. (2014, November 3). How and why researchers share data (and why they don't). *Wiley Exchanges*. Retrieved from http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/

Fisher, M. (2018, April 25). High stakes for Trump and Kim. *The New York Times International Edition*, p. 3.

Fitzpatrick, K. (2015, October 21). Opening up open access: Moving beyond business models and towards cooperative, scholar-organized, open networks*. LSE Impact Blog*. Retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/21/opening-up-open-access-kathleen-fitzpatrick/

Force11. (2017). Fair Data Principles. Retrieved from https://www.force11.org/group/fairgroup/fairprinciples

FOSTER. (2017). Open science definition. Retrieved from https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

Franzen, M. (2016). *Open science als wissenschaftspolitische problemlösungsformel*? In D. Simon, A. Knie, S. Hombostel, & K. Zimmerman, *Handbuch Wissenschaftspolitik* (pp. 279-296). Wiesbadden: Springer Fachmedien. doi:10.1007/978-3-658-05455-7_23.

Friedland, R. (2012). Review of the institutional logics perspective: A new approach to culture, structure, and process. *Management, 15*, 583-595. doi:10.3917/mana.155.0583.

Friedland, R., & Alford, R. (1991). Bringing society back in: Symbols practices, and institutional contradictions. In W. Powell, & P. DiMaggio, *The New Institutionalism in Organizational Analysis* (pp. 232-263). Chicago: Chicago University Press.

Friesike, S., Widenmayer, B., Gassmann, B., & Schildhauer, T. (2015). Opening science: Towards an agenda of open science in academia and industry. *Journal of Technology Transfer 40*, 581-601. doi:10.1007/s10961-014-9375-6.

Fyfe, A., McDougall-Waters, J., & Moxham, N. (2015). 350 years of scientific periodicals. *Notes and Records, 69*, 227-239. doi:10.1098/rsnr.2015.0036.

Gous, N., & Child, K. (2018, July 9). Tim Noakes's battle with council over tweet comes to an end. *BusinessDay*. Retrieved from https://www.businesslive.co.za/bd/national/2018-07-09-tim-noakess-battle-with-council-over-tweet-comes-to-an-end/

GovLab. (n.d.). Data collaboratives. Retrieved from http://datacollaboratives.org/

Gower, T. (2012, January 21). Elsevier – My part in its downfall*. Gowers's Weblog*. Retrieved from https://gowers.wordpress.com/2012/01/21/elsevier-my-part-in-its-downfall/

Grabowicz, P., Ramasco, J., Moro, E., Pujol, J., & Eguiluz, V. (2012). Social features of online networks: The strength of intermediary ties in online social media. *PLOS ONE, 7*(1), e29358. doi:10.1371/journal.pone.0029358.

Grand, A., Wilkinson, C., Bultitude, K., & Winfield, A. (2014). Mapping the hinterland: Data issues in open science. *Public Understanding of Science*, 1-17. doi:10.1177/0963662514530374.

Gray, J. (2014). Towards a genealogy of open data. *SSRN*, 1-38. Retrieved from https://www.ssrn.com/abstract=2605828

Griffin, A. (2018, July 3). WhatsApp: After killings in India, how the messaging app is being used to spread deadly fake news. *The Independent*. Retrieved from https://www.independent.co.uk/life-style/gadgets-and-tech/news/whatsapp-india-killings-latest-update-explained-app-fake-hoax-rumours-a8428746.html

Grigorov, G., & Tuddenham, P. (2015). Open science: A first step towards science communication. *Geophysical Research Abstracts, 17*.

Gross, A., Harmon, J., & Reidy, M. (2002). *Communication Science: The scientific article from the 17th century to the present.* Oxford: Oxford University Press.

Hahnel, M. (2015). 2015: The year of open data mandates. *Figshare*. Retrieved from http://figshare.com/blog/2015_The_year_of_open_data_mandates/143

Hall, B., & Tandon, R. (2017). Decolonization of knowledge, epistemicide, participatory research and higher education. *Research for All, 1*(1), 6-19. doi:10.18546/RFA.01.1.02.

Hansen, D., Shneiderman, B., & Smith, M. (2010). *Analyzing Social Media Networks with NodeXL: Insights from a connected world.* Burlington MA: Morgan Kaufman.

Harris, R. (2017). *Munich.* London: Hutchinson.

Haupt, M. (2016, May 2). 'Data is the new oil' – A ludicrous proposition. *Medium*. Retrieved from https://medium.com/twenty-one-hundred/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294

Haustein, H., Bowman, T., Holmberg, K., Tsou, A., Sugimoto, C., & Lariviere, V. (2015). Tweets as impact indicators: Examining the implications of automated 'bot' accounts on Twitter. *Journal of the Association for Information Science and Technology, 67*(1), 232-238. doi:10.1002/asi.23456.

Haustein, S., & Costas, R. (2015). Identifying Twitter audiences: Who is tweeting about scientific papers? *ASIS&T SIG/MET Metrics 2015 Workshop*, (pp. 1-3).

Haustein, S., Bowman, T., & Costas, R. (2016). Interpreting 'altmetrics': Viewing acts on social media through the lens of citation and social theories. In C. Sugimoto, *Theories of Infometrics and Scholarly Communication* (pp. 372-406). Berlin: De Gruyter. doi:10.1515/9783110308464-022.

Haustein, S., Costas, R., & Lariviere, V. (2015). Characterising social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLOS ONE, 10*(5), e0127830. doi:10.1371/journal.pone.0120495.

Heller, M. (n.d.). 2015 AAA President's report. *Anthropology News*. Retrieved from http://www.anthropology-news.org/index.php/2016/01/05/2015-aaa-presidents-report/

Hern, A. (2018, July 20). WhatsApp to restrict message forwarding after India mob lynchings. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/jul/20/whatsapp-to-limit-message-forwarding-after-india-mob-lynchings

Hillyer, B. (2017). Confessions of an open access advocate: Lesley Chan. *OCSDnet*. Retrieved from https://ocsdnet.org/confessions-of-an-open-access-advocate-leslie-chan/

Himelboim, I., Smith, M., Raine, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. *Social Media + Society*, 1-13. doi:10.1177/20563051176915.

Holliman, R. (2011). Telling science stories in an evolving digital media ecosystem: From communication to conversation and confrontation. *Journal of Science Communication, 10*(4), 1-4.

Hopke, J., & Simis, M. (2017). Discourse over a contested technology on Twitter: A case study of hydraulic fracturing. *Public Understanding of Science, 26*(1), 105-120.

Hornig Priest, S. (2009). Reinterpreting the audiences for media messages about science. In R. Holliman, E. Whitelegg, E. Scanlon, S. Smidt, & J. Thomas, *Investigating Science Communication in the Information Age: Implications for public engagement and popular media* (p. 223). Oxford: Oxford University Press.

Illing, S. (2018, April 4). Cambridge Analytica, the shady data firm that might be a key Trump-Russia link, explained. *Vox*. Retrieved from https://www.vox.com/policy-and-politics/2017/10/16/15657512/cambridge-analytica-facebook-alexander-nix-christopher-wylie

Ioannidis, J., Allison, D., Ball, C., Coulibaly, I., Cui, X., Culhane, A., . . . Van Noort, V. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics, 41*, 149-155.

Irwin, A. (2014). From deficit to democracy. *Public Understanding of Science, 23*(1), 71-76. doi:10.1177/0963662513510646.

Jamieson, K. (2017). The need for a science of science communication: Communicating science's values and norms. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 15-24). New York: Oxford University Press.

Jasanoff, S. (2006). Transparency in public science: Purposes, reasons, limits. *Law and Contemporary Problems, 69*(21), 21-45.

Jasanoff, S. (2012). *Science and Public Reason.* London: Routledge.

Jaschik, S. (2015, November 2). Language of protest. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2015/11/02/editors-and-editorial-board-quit-top-linguistics-journal-protest-subscription-fees

Jensen, P., Rouquier, J.-B., Kreimer, P., & Croissant, Y. (2008). Scientists who engage with society perform better academically. *Science and Public Policy, 35*(7), 527-541.

Jiménez, A. C., Willinsky, J., Boyer, D., & Da Col, G. &. (2015, October 20). Why an open-access publishing cooperative can work: A proposal for the AAA's journal portfolio. *Cultural Anthropology*. Retrieved from http://www.culanth.org/fieldsights/743-why-an-open-access-publishing-cooperative-can-work-a-proposal-for-the-aaa-s-journal-portfolio

Johnson, J. (2018). *Toward Information Justice: Technology, politics, and policy for data in higher education administration.* Dordrecht: Springer.

Kahan, D. (2013). A risky science communication environment for vaccines. *Science, 342*, 53-54.

Kahan, D. (2014). *Vaccine Risk Perceptions and Ad Hoc Risk Communication: An empirical assessment.* New Haven: Cultural Cognition Project, Yale Law School.

Kahan, D., Scheufele, D., & Jamieson, K. (2017). Introduction: Why science communication? In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of The Science of Science Communication* (pp. 1-11). New York: Oxford University Press.

Kahle, S. (2011). Preferential attachment. In A. Barnett, *Encyclopedia of Social Networks* (pp. 718-720). Thousand Oaks: Sage.

Kasperson, R., Renn, O., Slovic, P. B., Emel, J., Goble, R., Kasperson, J., & Ratick, S. (1988). Social amplification of risk: A conceptual framework. *Risk Analysis, 8*(2), 177-187. doi:10.1111/j.1539-6924.1988.tb01168.x.

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm: An overview of the tactics and tropes used online by the anti-vaccination movement. *Vaccine, 30*, 3778-3789.

Kelley, B. (2016). Toward a goodwill ethics of online research methods. *Transformative Works & Cultures, 22*.

Klump, J., Beterlman, R., Brase, J., Diepenbroek, M., Grobe, H., Hock, H., . . . Wachter, J. (2006). Data publication in the open access initiative. *Data Science Journal, 5*, 79-83.

Kozinets, R. (2010). Conducting ethical netnography. In R. Kozinets, *Netnography: Doing ethnographic research online* (pp. 136-156). London: Sage.

Kramer, D. (2017, May 1). Steady, strong growth is expected for open-access journals. *Physics Today*, 24-28. doi:10.1063/PT.3.3550

Kroes, N. (2012, March 5). European Commission press release*. European Commission*. Retrieved from http://europa.eu/rapid/press-release_SPEECH-12-149_en.htm

Kroes, N. (2013, March 26). European Commission press release*. European Commission*. Retrieved from http://europa.eu/rapid/press-release_SPEECH-13-261_en.htm

Kronick, D. (1976). *A History of Scientific and Technical Periodicals: The origins and development of the scientific and technical press, 1665-1790.* New York: Scarecrow.

Kuckler, H. (2018, July 25). A social media tax, freedom of speech and Uganda. *Financial Times*. Retrieved from https://www.ft.com/content/1dc7f8f8-8ece-11e8-bb8f-a6a2f7bca546

Kuhlen, R. (2010). Ethical foundations of knowledge as a commons. *Proceedings of the International Conference Commemorating the 40th Anniversary of the Korean Society for Library and*

Information Science (pp. 1-15). Seoul: Korean Society for Library and Information Science. Retrieved from http://www.kuhlen.name/MATERIALIEN/Publikationen2010/RK-ethical%20foundation%20commons-final-to-print-030910.pdf

Kumar, S., Hamilton, W., Leskovec, J., & Jurafsky, D. (2018). Community interaction and conflict on the web. *WWW 2018: The 2018 Web Conference, April 23-27, 2018, Lyon, France* (pp. 1-11). New York: ACM.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or mass media? *WWW 2010, April 26-30, 2010, Raleigh, North Carolina, USA* (pp. 591-600). New York: ACM.

Laakso, M., & Bjork, B.-C. (2012). Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine* 10, 1-9. doi:10.1186%2F1741-7015-10-124.

Landrum, A. (2017). The role of intermediaries in communicating science: A synthesis. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 253-260). New York: Oxford University Press.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLOS ONE, 10*(6). doi:10.1371/journal.pone.0127502.

Larson, H., Figueiredo, A., Xiahong, Z., Schulz, W., Verger, P., Johnston, D., . . . Jones, N. (2016). The state of vaccine confidence 2016: Global insights through a 67-country survey. *EBioMedicine, 12*, 295-301. doi:10.1016/j.ebiom.2016.08.042.

Lawson, S., Gray, J., & Mauri, M. (2016). Opening the black box of scholarly communication funding: A public data infrastructure for financial flows in academic publishing. *Open Library of Humanities, 2*(1), e10. doi:10.16995/olh.72.

Leask, J. (2015). Should we do battle with anti-vaccination activists? *Public Health Research & Practice, 25*(2), e2521515. doi:10.17061/phrp2521515.

Leonelli, S., & Prainsack, B. (2015, August 2). To what are we opening science? Reform of the publishing system is only a step in a much broader re-evaluation. *LSE Impact Blog*. Retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2015/04/21/to-what-are-we-opening-science/

Leonelli, S., Spichtinger, D., & Prainsack, B. (2015). Sticks and carrots: Encouraging open science at its source. *Geography and Environment, 2*, 12-16. doi:10.1002/geo2.2.

Lerman, K., & Hogg, T. (2014). Leveraging position bias to improve peer recommendation. *PLOS ONE, 9*(6), e98914. doi:10.1371/journal.pone.0098914.

Levere, T. (1981). Book review: A. J. Meadows, ed., Development of Science Publishing in Europe; pp. x, 269. Amsterdam, New York, Oxford: Elsevier Science Publishers, 1980. Dfl. 100.00. *Victorian Periodicals Review, 14*(4), 156-158.

Levin, N., & Leonelli, S. (2017). How does one 'open' science? Questions of value in biological research. *Science, Technology & Human Values, 42*(2), 280-305. doi:10.1177/0162243916672071.

Lewandowsky, S., & Bishop, D. (2016). Research integrity: Don't let transparency damage science. *Nature, 529*, 459-461. doi:10.1038/529459a.

Library Genesis. (n.d.). Library Genesis[2M]. Retrieved from http://gen.lib.rus.ec/

Lin, N. (2008). A network theory of social capital. In D. Castiglione, J. Van Deth, & G. Wolleb, *The Handbook of Social Capital* (pp. 50-69). Oxford: Oxford University Press.

Lo, N., & Hotez, P. (2017). Public health and economic consequences of vaccine hesitancy for measles in the United States. *JAMA Pediatrics, 171*(9), 887-892. doi:10.1001/jamapediatrics.2017.1695.

Mabe, M. (2003). The growth and number of journals. *Serials, 16*(2), 191-197.

Mark Ware Consulting. (2006). *Scientific Publishing in Transition: An overview of current developments.* Bristol: Mark Ware Consulting.

Mauthner, N., & Parry, O. (2013). Open access digital data sharing: Principles, policies and practices. *Social Epistemology, 27*(1), 47-67.

McCallie, E., Bell, L., Lohwater, T., Falk, J., Lehr, J., Lewenstein, B., . . . Wiehe, B. (2009). *Many Experts, Many Audiences: Public engagement with science and informal science education.* Washington DC: Center for Advancement of Informal Science Education (CAISE).

McDermott, P. (2008). *Who Needs to Know? The state of public access to federal government information.* Lanham: Bernan.

McKie, D. (1979). The scientific periodical from 1665 to 1798. In A. Meadows, *The Scientific Journal* (pp. 7-18). London: Aslib.

McKiernan, E., Bourne, P., Brown, C., Buck, S., Kenall, A., Lin, J., . . . Yarkoni, T. (2016). How science helps researchers succeed. *eLIFE* 5, 1-19. doi:10.7554/eLife.16800.

Meadows, A. (1979). *The Scientific Journal.* London: Aslib.

Meadows, A. (1980). *The Development of Science Publishing in Europe.* Amsterdam: Elsevier.

Merton, R. K. (1968). *Social Theory and Social Structure.* New York: Free Press.

Merton, R. K. (1973). The normative structure of science. In R. K. Merton, *The Sociology of Science* (pp. 267-278). Chicago: Chicago University Press.

Merton, R. K. (1976). *Sociological Ambivalence and Other Essays.* New York: Free Press.

Milani, E. (2016, October 4). SciHashtag: Twitter hashtags for scientists. *Public Understanding of Science Blog*. Retrieved from https://sagepus.blogspot.co.za/2016/10/scihashtag-twitter-hashtag-science.html

Miller, V. (2017). Phatic culture and the status quo: Reconsidering the purpose of social media activism. *Convergence, 23*(3), 251-269. doi:10.1177/1354856515592512.

Mitra, T., Counts, S., & Pennebaker, J. (2016). Understanding anti-vaccination attitudes in social media. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)* (pp. 269-278). Palo Alto, California: The AAAI Press.

Mitroff, I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review, 39*(August), 579-595.

Monge, P., & Contractor, N. (2003). *Theories of Communication Networks.* Oxford: Oxford University Press.

Moran, B., Lucas, M., Everhart, K., & Morgan, A. (2016). What makes anti-vaccine websites persuasive? A content analysis of techniques used by antivaccine vaccine sentiment. *Journal of Communication in Healthcare, 9*(3), 151-163. doi:10.1080/17538068.2016.1235531.

Moten, M., Phillips, M., Saliba, V., Harding, N., & Sibal, B. (2018). Measles: Is it still a threat? *British Journal of General Practice*, published online before print. doi:10.3399/bjgp18X697961

Muller, J. (2017). Universities and the 'new society'. In J. Muller, N. Cloete, & F. Van Schalkwyk, *Castells in Africa: Universities and Development* (pp. 17-31). Cape Town: African Minds.

National Academy of Science. (2009). *Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age.* Washington DC: National Academy of Science.

Nature. (2005). Let data speak to data. *Nature, 438*(7068), 531.

Naughton, J. (2017, November 19). How a half-educated tech elite delivered us into chaos. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2017/nov/19/how-tech-leaders-delivered-us-into-evil-john-naughton

NCBI. (n.d.). Welcome to NCBI. National Health Center for Biotechnology Information (NCBI). Retrieved from https://www.ncbi.nlm.nih.gov/

Nelhans, G., & Lorentzen, D. G. (2016). Twitter conversation patterns related to research papers. *Information Research, 21*(2). Retrieved from http://www.informationr.net/ir/21-2/SM2.html

Nikolau, L. (2016, September 15). Vaccines: Education and access aren't enough in a skeptical Europe, survey finds. *Humanosphere*. Retrieved from http://www.humanosphere.org/global-health/2016/09/vaccines-education-and-access-arent-enough-in-a-skeptical-europe-survey-finds/

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus, 140*(4), 32-48.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Bahvior, 32*(2), 303-330.

OECD. (2007). *Principles and Guidelines for Access to Research Data from Public Funding.* Paris: OECD Publications.

OECD. (2008). *OECD Recommendation of the Council for Enhanced Accessed and More Effective Use of Public Sector Information.* Paris: OECD.

OECD. (2015). Making open science a reality. *OECD Science, Technology and Industry Policy Papers* 25, 1-108. doi:10.1787/5jrs2f963zs1-en.

Olsen, J. (2008). Explorations in institutions and logics of appropriateness: An introductory essay. In J. March, *Explorations in Organizations* (pp. 191-202). Stanford, CA: Stanford University Press.

Open Knowledge. (2015, August 2). Open definition. Retrieved from http://opendefinition.org/

Ortiz-Ospina, E., & Roser, M. (2016). Trust. *Our World in Data*. Retrieved from: https://ourworldindata.org/trust

Pels, D. (2003). *Unhastening Science: Autonomy and reflexivity in the social theory of knowledge*. Liverpool: Liverpool University Press.

Peters, H. (2013). Gap between science and media revisited: Scientists as public communicators. *PNAS, 110*(sup. 3), 14102-14109. doi:10.1073/pnas.1212745110.

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2015). Research data explored: Citations versus almetrics. *15th International Conference on Scientometrics and Informetrics (ISSI 2015).* Retrieved from http://arxiv.org/abs/1501.03342

Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751-783. doi:10.1037/0022-3514.90.5.751.

Petts, J., & Niemeyer, S. (2004). Health risk communication and amplification: Learning from the MMR vaccination controversy. *Health, Risk & Society, 6*(1), 7-23.

Piedrahita, P., Borge-Holthoefer, J., Moreno, Y., & González-Bailónd, S. (2018). The contagion effects of repeated activation in social networks. *Social Networks, 54*, 326-335. doi:10.1016/j.socnet.2017.11.001.

Pinfield, S. (2015). Making open access work: The 'state-of-the-art' in providing open access to scholarly literature. *Online Information Review, 39*(5), 604-636. doi:10.1108/OIR-05-2015-0167.

Pinker, S. (2002). Political scientists. In S. Pinker, *The Blank Slate: The modernisation of human nature* (pp. 105-120). London: Allen Lane.

Piwowar, H., Day, R., & Fridsma, D. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE, 2*(3), 308.

Piwowar, H., Priem, J., Larivière, V., Alperin, J., Matthias, L., Norlander, B., . . . Haustein, S. (2017). The state of OA: A large-scale analysis of the prevalence and impact of open access articles. *PeerJ Preprints* 5, e3119v1. doi:10.7287/peerj.preprints.3119v1.

Piwowar, H., & Vision, T. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1-25. doi:10.7717/peerj.175.

Power, M. (1997). *The Audit Society.* Oxford: Oxford University Press.

Power, M. (2000). The audit society – second thoughts. *International Journal of Accounting* 4, 111-119.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011, September 28). Altmetrics: A manifesto. *Altmetrics*. Retrieved from http://altmetrics.org/manifesto/

Quattrociocchi, W., Scala, A., & Sunstein, C. (2016, September). Echo chambers on Facebook. *Harvard John M. Olin Discussion Paper Series*, pp. 1-15. Retrieved from http://www.law.harvard.edu/programs/olin_center/papers/pdf/Sunstein_877.pdf

Radzikowski, J., Stefanidis, A., Jacobsen, K., Croitoru, A., Crooks, A., & Delamater, P. (2016). The measles vaccination narrative in Twitter: A quantitative analysis. *JMR Public Health and Surveillance, 2*(1), 1-15.

Raikhel, E. (2015, November 10). *An open access publishing cooperative for anthropology? Somatosphere*. Retrieved from http://somatosphere.net/2015/11/an-open-access-publishing-cooperative-for-anthropology.html

Registry of Open Access Repository Mandates and Policies. (2017, August 7). How many open access policies are there worldwide? *Registry of Open Access Repository Mandates and Policies*. Retrieved from http://roarmap.eprints.org/dataviz2.html

Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. *WebSci, 2-4 May* (pp. 1-10). Paris: ACM.

Robinson-Garcia, N., Van Leewen, T., & Rafols, I. (2018). Using altmetrics for contextualised mapping of societal impact: From hits to networks. *Science and Public Policy*, scy024. doi:10.1093/scipol/scy024.

Rogers, R. (2015). Digital methods for web research. In R. Scott, & S. Kosslyn, *Emerging Trends in the Social and Behavioural Sciences* (pp. 1-22). New York: John Wiley.

Royal Society Policy Centre. (2012). *Science as an Open Enterprise.* London: The Royal Society.

Sanawi, J., Samani, M., & Taibi, M. (2017). #Vaccination: Identifying influencers in the vaccination discussion on Twitter through social network visualisation. *International Journal of Business and Society, 18*(S4), 718-726.

Schäfer, M. (2016). Mediated trust in science: Concept, measurement and perspectives for the 'science of science communication'. *JCOM, 15*(5), 1-7.

Schäfer, M. (2017). How changing media structures are affecting science news coverage. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 51-59). New York: Oxford University Press.

Schäfer, M., Füchslin, T., Metag, J., Kristiansen, S., & Rauchfleisch, A. (2018). The different audiences of science communication: A segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public Understanding of Science*, 1-21. doi:10.1177/0963662517752886.

Scheufele, D. (2013). Communicating science in social settings. *PNAS, 10*(3), 14040-16047.

Scheufele, D. (2014). Science communication as political communication. *PNAS, 111*(4), 13585-13592.

Sci-Hub. (n.d.). Retrieved from https://sci-hub.io/

Scott, W. (2014). *Institutions and Organizations: Ideas, interests, and identities* (4th edn). Thousand Oaks: Sage.

Shah, H. (2018). Use our personal data for the common good. *Nature, 556*, 7. doi:10.1038/d41586-018-03912-z.

Sharif, N., Ritter, W., Davidson, R., & Edmunds, S. (2017). An open science 'state of the art' for Hong Kong: Making open research data available to support Hong Kong's innovation policy. *SocArXiv*, 1-43.

Southwell, B. (2017). Promoting popular understanding of science and health through social networks. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 223-231). New York: Oxford University Press.

Stake, R. (1995). *The Art of the Case Study.* Thousand Oaks: Sage.

Stalder, F. (2006). *Manuel Castells.* Cambridge: Polity.

Statista. (2018). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2018 (in millions). Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Statista. (2018). Number of monthly active Facebook users worldwide as of 2nd quarter 2018 (in millions). Retrieved from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Stein, R. (2011, January 11). Wakefield tried to capitalize on autism-vaccine link, report says. *The Washington Post*. Retrieved from http://voices.washingtonpost.com/checkup/2011/01/wakefield_tried_to_capitalize.html

Stodden, V. (2010). Open science: Policy implications for the evolving. *Journal of Science Communication, 9*(1), 1-8.

Strasser, B., & Edwards, P. (2015). *Open Access: Publishing, commerce and the scientific ethos.* Bern: Swiss Science and Innovation Council.

Streitfeld, D. (2017, May 20). 'The internet is broken': @ev is trying to solve it. *The New York Times*. Retrieved from https://www.nytimes.com/2017/05/20/technology/evan-williams-medium-twitter-internet.html

Sugimoto, C., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology, 68*(9), 2037-2062. doi:10.1002/asi.23833.

Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical challenges in online research: Public/private perceptions. *Research Ethics, 13*(3-4), 184-199. doi:10.1177/1747016116650720.

Sunstein, C. (2017). *#Republic: Divided democracy in the age of social media.* Princeton: Princeton University Press.

Swan, A., & Brown, S. (2008). To share or not to share: Publication and quality assurance of research data outputs*.* A report commissioned by the Research Information Network. Retrieved from https://eprints.soton.ac.uk/266742/

Swartz, A. (2008, July). *Full text of 'Guerilla Open Access Manifesto'*. Retrieved from https://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.txt

Taubert, N., & Weingart, P. (2017). Changes in scientific publishing. In P. Weingart, & N. Taubert, *The Future of Scientific Publishing* (pp. 1-31). Cape Town: African Minds.

Tennant, J., Waldner, F., Jacques, D., Masuzzo, P., Collister, L., & Hertgerink, J. (2016). The academic, economic and societal impacts of open access: An evidence-based review. *F1000Research, 5*, 632. doi:10.12688/f1000research.8460.1.

Tennison, J. (2014, March 17). Data sharing is not open data*. Open Data Institute*. Retrieved from https://theodi.org/blog/data-sharing-is-not-open-data

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE, 6*(6), e21101. doi:10.1371/journal.pone.0021101.

Tenopir, C., Dalton, E., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE, 10*(8), e0134823. doi:10.1371/journal.pone.0134826.

Tharoor, I. (2018, March 22). The scary truth that Cambridge Analytica understands*. The Washington Post*. Retrieved from https://www.washingtonpost.com/news/worldviews/wp/2018/03/22/the-scary-truth-that-cambridge-analytica-understands/

The Cost of Knowledge. (n.d.). The cost of knowledge. Retrieved from http://thecostofknowledge.com

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics,* 17, 1-8.

Thornton, P. H., & Ocasio, W. (1999). Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958-1990. *American Journal of Sociology, 105*(3), 801-842.

Thornton, P., & Ocasio, W. (2008). Institutional logics. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby, *The Sage Handbook of Organisational Institutionalism* (pp. 99-129). Thousand Oaks: Sage.

Thornton, P., Ocasio, W., & Lounsbury, M. (2012). *The Institutional Logics Perspective: A new approach to culture, structure, and process.* New York: Oxford University Press.

Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., . . . Nyham, B. (2018). *Social Media, Political Polarization, and Political Disinformation: A review of the scientific literature.* Washington DC: William and Flora Hewlett Foundation.

US National Research Council. (1997). *Bits of Power.* Washington DC: US National Research Council.

Vaccine Research Library. (2015, December 16). Scream 146: Call for autobots to suppress vaccine information. *Vaccine Research Library*. Retrieved from http://vaccineresearchlibrary.com/scream-146-call-for-autobots-to-suppress-vaccine-information

Vainio, J, & Holmberg, K. (2017). Highly tweeted science articles: Who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics*, 112, 345-366. doi: 10.1007/s11192-017-2368-0.

Valenzuela, S., Correa, T., & Gil de Zuniga, H. (2017). Ties, likes, and tweets: Using strong and weak ties to explain differences in protest participation across Facebook and Twitter use. *Political Communication*, 1-18. doi:10.1080/10584609.2017.1334726.

Van den Heuvel, K. (2013, July 22). Jenny McCarthy's vaccination fear-mongering and the cult of false equivalence. *The Nation*. Retrieved from https://www.thenation.com/article/jenny-mccarthys-vaccination-fear-mongering-and-cult-false-equivalence/

Vanderslott, S., & Roser, M. (2018). Vaccination*. Our World in Data*. Retrieved from https://ourworldindata.org/vaccination

Van Noorden, R. (2015). Online collaboration: Scientists and the social network. *Nature, 512*(7513), 126-129. doi:10.1038/512126a.

Van Orsdel, L. (2008). The state of scholarly communications. *The Serials Librarian, 52*(1-2), 191-209. doi:10.1300/J123v52n01_16.

Vision, T. (2010). Open data and the social contract of scientific publishing. *BiScience, 60*(5), 330-331.

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics, 103*(2), 555-564.

Ware, M., & Mabe, M. (2015, March). An overview of scholarly journal publishing. *The STM Report*, pp. 1-180.

Waterston, A. (n.d.). Anticipations, from the President, Alisse Waterston. *Anthropology News*. Retrieved from http://www.anthropology-news.org/index.php/2016/03/01/anticipations/

Weingart, P. (2008). How robust is 'socially robust knowledge'? In N. Stehr, *Knowledge and Democracy: A 21st century perspective* (pp. 143-158). New Brunswick: Transaction.

Weingart, P. (2011). Science, the public and the media – Views from everywhere. In M. Carrier, & A. Nordmann, *Science in the Context of Application* (pp. 337-348). Dordrecht: Springer.

Weingart, P. (2012). The lure of the mass media and its repercussions on science. In S. Rodder, M. Franzen, & P. Weingart, *The Sciences' Media Connection: Public communication and its repercussions* (pp. 17-32). Dordrecht: Springer.

Weingart, P. (2017). Is there a hype problem in science? If so, how is it addressed? In K. Jamieson, D. Kahan, & Scheufele, D., *The Oxford Handbook of the Science of Science Communication* (pp. 111-118). New York: Oxford University Press.

Weingart, P., & Guenther, L. (2016). Science communication and the issue of trust. *JCOM, 15*(5), 1-7.

Weingart, P., Engels, A., & Pansegrau, P. (2000). Risks of communication: Discourses on climate change in science, politics, and the mass media. *Public Understanding of Science, 9*, 261-283.

Weiss, P. (2002). *Borders in Cyberspace: Conflicting public sector information policies and their economic impacts.* National Weather Service, National Oceanic and Atmospheric Administration.

Wells, P. (2016, March 2). Widening the debate on data sharing. *Open Data Institute*. Retrieved from https://theodi.org/blog/widening-the-debate-on-data-sharing

Wenzlaff, F. (2014, March 27). Reflections of the 10th New Institutionalism Workshop, 20-21 March, Rome – Part I. *Science Studies*. Retrieved from: http://science-studies.com/reflections-of-the-10th-new-institutionalism-workshop-20-21-march-rome-part-i/#comments

Wicherts, J., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726-728.

Wu, T. (2016). *The Attention Merchants: The epic struggle to get inside our heads.* London: Atlantic.

Yeo, S., & Brossard, D. (2017). The (changing) nature of scientist-media interactions: A cross-national analysis. In K. Jamieson, D. Kahan, & D. Scheufele, *The Oxford Handbook of the Science of Science Communication* (pp. 261-272). New York: Oxford University Press.

Yeo, S., Xenos, M., Brossard, D., & Scheufele, D. (2015). Selecting our own science: How communication contexts and individual traits shape information seeking. *The ANNALS of the American Academy of Political and Social Science, 658*(1), 172-191. doi:10.1177/0002716214557782.

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Sirivianos, M., . . . Blackburn, J. (2017). The web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources. *IMC '17, November 1–3, 2017, London, United Kingdom.* New York: ACM. doi:10.1145/3131365.3131390.

Zimmer, M., & Proferes, N. (2013). A topology of Twitter research: Disciplines, methods, and ethics. *Aslib Journal of Information Management, 66*(3), 250-261.

Zimmerman, R., Wolfe, R., & Fox, D. (2005). Vaccine criticism on the world wide web. *Journal of Medical Internet Research, 7*(2).