World Trends in Freedom of Expression and Media Development:

Special Digital Focus 2015

United Nations
Educational, Scientific and
Cultural Organization

UNESCO
Publishing

World Trends in Freedom of Expression and Media Development

Special Digital Focus 2015

UNESCO

United Nations
Educational, Scientific and
Cultural Organization

UNESCO
Publishing

Published in 2015 by
the United Nations Educational, Scientific and Cultural Organization,
7, place de Fontenoy, 75352 Paris 07 SP, France

ISBN 978-92-3-100127-7

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

This publication draws from three other studies, summarising key points and highlighting the trends involved:

1. Gagliardone, I. et al. 2015. *Countering Online Hate Speech*. UNESCO Series on Internet Freedom. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002332/233231e.pdf
2. Posetti, J. (forthcoming). *Protecting Journalism Sources in the Digital Age*. Paris: UNESCO
3. MacKinnon, R. et al. 2014. *Fostering Freedom Online: The Role of Internet Intermediaries*. UNESCO Series on Internet Freedom. Paris: UNESCO / Internet Society. http://unesdoc.unesco.org/images/0023/002311/231162e.pdf

This publication was supported by Sweden.

SWEDEN

The Internet Society, the Open Society Foundations, the Center for Global Communication Studies at the University of Pennsylvania's Annenberg School for Communication, the University of Oxford, and the World Editors Forum within the World Association of Newspapers and News Publishers (WAN-IFRA) also contributed.

Editor: Rachel Pollack Ichou
Graphic design: UNESCO
Cover design: UNESCO
Illustrations: UNESCO
Typeset: UNESCO
Printed by: UNESCO

Printed in France

World Trends In Freedom of Expression and Media Development: Special Digital Focus 2015

Editor: Rachel Pollack Ichou, Project Officer, Section for Freedom of Expression, Division of Freedom of Expression and Media Development, UNESCO

Countering Online Hate Speech

Authors:

Iginio Gagliardone, Research Fellow in New Media and Human Rights, Centre for Socio-Legal Studies, and member, Programme in Comparative Media Law and Policy (PCMLP), University of Oxford

Danit Gal, MSc Candidate in Social Science of the Internet, University of Oxford

Thiago Alves Pinto, DPhil Candidate in Law, University of Oxford

Gabriela Martinez Sainz, DPhil Candidate in Education, University of Cambridge

International Advisory Committee:

Monroe Price, Adjunct Full Professor of Communication, Director, Center for Global Communication Studies, Annenberg School for Communication, University of Pennsylvania

Richard Danbury, Research Associate, Faculty of Law, University of Cambridge

Cherian George, Adjunct Senior Research Fellow, Institute of Policy Studies, Lee Kuan Yew School of Public Policy, National University of Singapore

Nazila Ghanea, University Lecturer in International Human Rights Law and Fellow of Kellogg College, University of Oxford

Robin Mansell, Professor of New Media and the Internet, Department of Media and Communication, the London School of Economics and Political Science (LSE)

Bitange Ndemo, former Permanent Secretary, Ministry of Information and Communication, Kenya

Nicole Stremlau, Head, Programme in Comparative Media Law and Policy, University of Oxford

Protecting Journalism Sources in the Digital Age

Author and Chief Researcher: Julie Posetti, former Research Fellow, World Association of Newspapers and News Publishers. (WAN-IFRA) and Editor with the World Editors Forum; Lecturer – Broadcast and Convergent Journalism, University of Wollongong.

Academic Researchers:

Ying Chan, Founding Director, Journalism and Media Studies Centre, University of Hong Kong

Marcus O'Donnell, Senior Lecturer in Journalism, Faculty of Law, Humanities and the Arts, University of Wollongong

Carlos Affonso Pereira de Souza, Vice-coordinator, Center for Technology & Society (CTS), Getulio Vargas Foundation (FGV) Law School, Rio de Janeiro

Doreen Weisenhaus, Director of Media Law Project & Associate Professor, University of Hong Kong

Graduate Research Assistants, and Undergraduate Research Contributors:

Federica Cherubini, Programme Manager, WAN-IFRA

Jake Evans, former WAN-IFRA-University of Wollongong Journalism Intern; contributor Australian Broadcasting Corporation (ABC)

Emma Goodman, Research Officer, LSE Media Policy Project, LSE

Angelique Lu, Journalist, BBC; and former WAN-IFRA-University of Wollongong Journalism Intern

Alice Matthews, Journalist, Australian Broadcasting Corporation (ABC) News; former Research Assistant, WAN-IFRA

Alexandra Sazonova-Prokouran, Student, University of Oxford; former WAN-IFRA Intern.

Jessica Sparks, Journalism/Law student, University of Wollongong; former WAN-IFRA Intern

Nick Toner, Co-founder / Publisher, VERSA News; student, Oxford of University, former WAN-IFRA Intern

Farah Wael, Coordinator, Media Development and Press Freedom, WAN-IFRA

Alexandra Waldhorn, Communications Officer, International Institute for Educational Planning, UNESCO; former Youth Advisor, WAN-IFRA

Olivia Wilkinson, Student, University of Oxford; former Intern, WAN-IFRA

Administrative support:

Ashleigh Tullis, Journalist, Fairfax Media Wollondilly Advertiser; former Intern, WAN-IFRA

International Advisory Committee

Mark Pearson, Professor of Journalism and Social Media, Griffith University

Julie Reid, Media Studies Senior Lecturer, Department of Communication Science, UNISA (University of South Africa)

Lillian Nalwoga, President, Internet Society's Uganda Chapter; Policy Officer, Collaboration on International ICT Policy in East and Southern Africa (CIPESA)

Dan Gillmor, Director of the Knight centre for digital media entrepreneurship at Arizona State University's Walter Cronkite school of journalism and mass communication

Prisca Orsonneau, Lawyer at the Paris Bar, specializing in Media Law and Human Rights; Chair, Reporters Without Borders Legal Committee

Gayathry Venkiteswaran, Executive Director, Southeast Asian Press Alliance

Mario Calabresi, Editor-in-Chief, La Stampa

Mishi Choudhary, Legal Director, Software Freedom Law Centre and SFLC.in

Fostering Freedom Online: The Role of Internet Intermediaries

Authors:

Rebecca MacKinnon, Director, Ranking Digital Rights Project, New America Foundation; Visiting Affiliate, Center for Global Communication Studies, Annenberg School for Communication, University of Pennsylvania

Elonnai Hickok, Researcher, Centre for Internet and Society

Allon Bar, Research Coordinator, Ranking Digital Rights Project

Hae-in Lim, Researcher, Ranking Digital Rights Project

Researchers:

Sara Alsherif, Researcher, Freedom of Information Program, Support for Information Technology Center

Celina Beatriz Mendes de Almeida Bottino, Instituto de Tecnologia & Sociedade do Rio de Janeiro

Richard Danbury, Research Associate, Faculty of Law, University of Cambridge

Elisabetta Ferrari, Center for Media, Data and Society, Central European University, Budapest; Doctoral student, Annenberg School for Communication, University of Pennsylvania

Grace Githaiga, Associate, Kenya ICT Action Network (KICTANet)

Kirsten Gollatz, Project Manager, Alexander von Humboldt Institute for Internet and Society

Elonnai Hickok, Researcher, Center for Internet and Society

Hu Yong, Associate Professor, Feature and Documentary Production, School of Journalism and Communication, Peking University

Tatiana Indina, Candidate of Sciences, Research Fellow, Center for the Study of New Media and Society

Victor Kapiyo, Programme Manager – Human Rights Protection, the Kenyan Section of the International Commission of Jurists (ICJ Kenya); Kenya ICT Action Network (KICTANet)

Peter Micek, Senior Policy Counsel, Access

Agustín Rossi, PhD Candidate, European University Institute; Non-resident Fellow, Global Public Policy Institute

International Advisory Committee:

Renata Avila, Global Campaign Manager, Web We Want initiative, World Wide Web Foundation; Contributor, Global Voices

Rasha Abdulla, Associate Professor and Past Chair, Journalism and Mass Communication, the American University in Cairo

Sunil Abraham, Executive Director, Centre for Internet and Society

Peng Hwa Ang, Professor, Nanyang Technological University, Wee Kim Wee School of Communication

Eduardo Bertoni: Director of the Center for Studies on Freedom of Expression and Access to Information (CELE), Palermo University School of Law

Seeta Peña Gangadharan, Program Fellow, New America's Open Technology Institute

Leslie Harris, Principal, Harris Strategy Group LLC; former President and CEO, Center for Democracy and Technology

Dunstan Allison Hope, Managing Director, Advisory Services, BSR

Rikke Frank Jørgensen, Senior Advisor, Danish Institute for Human Rights

Jeremy Malcolm, Senior Global Policy Analyst, Electronic Frontier Foundation

Pranesh Prakash, Policy Director, Centre for Internet and Society

Lucy Purdon, Program Manager: ICT, Institute for Human Rights and Business

David Sullivan, Policy and Communications Director, Global Network Initiative

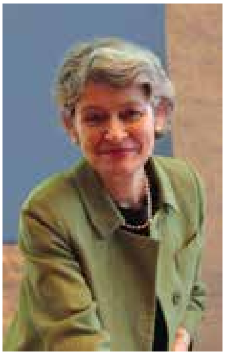
Ben Wagner, Director, Centre for Internet & Human Rights, European University Viadrina

Safety of Journalists

Author: Ming-Kuok Lim, Assistant Programme Specialist, Section for Freedom of Expression, Division of Freedom of Expression and Media Development, UNESCO

Table of Contents

Foreword by Irina Bokova, Director-General of UNESCO



As UNESCO celebrates its 70th anniversary, our founding mandate to promote 'the free flow of ideas by word and image' has never been so important, to advance the right to freedom of expression and to foster peace and sustainable development through media freedom, pluralism, independence and journalists' safety.

Throughout the World Summit on the Information Society (WSIS) and its follow-up process, UNESCO has promoted a vision of inclusive knowledge societies built on the pillars of freedom of expression, universal access to information and knowledge, respect for cultural and linguistic diversity, and quality education for all. As the United Nations reviews the past 10 years of WSIS and looks forward to supporting Member States to reach the Sustainable Development Goals (SDGs), UNESCO's work in these areas becomes all the more vital – especially at this time of revolutionary technological change and when all societies are transforming deeply.

The global communication and information environment has been transformed by the spread of digital technologies. Today, more than three billion women and men worldwide use the internet, and more than six billion have access to mobile phones. These technologies have expanded the possibilities for progress towards sustainable knowledge societies, while also raising new challenges.

In this dynamic landscape, UNESCO was called on by its 195 Member States in November 2013 to produce a comprehensive study of internet-related issues within its mandate, focusing on four areas: access to information and knowledge, freedom of expression, privacy, and ethics of the information study. The resulting study, *Keystones to foster inclusive Knowledge Societies*, explores these themes along with possible options for future action. This built on the earlier mandate provided by Member States at the UNESCO General Conference in 2011 to monitor world trends in freedom of expression and media development. It further drew on the first *World Trends on Freedom of Expression and Media Development* report that was published in 2014.

The *Keystones* report is unique in advancing the concept of 'Internet Universality', to designate an internet that is human Rights-based, Open, Accessible for all, and governed by Multi-stakeholder participation, showing how the internet that can advance a range of Sustainable Development Goals and targets – from ending poverty, achieving gender equality and ensuring sustainable consumption and production patterns to combatting climate change and promoting peaceful and inclusive societies.

This second *World Trends in Freedom of Expression and Media Development* elaborates on key aspects of the *Keystones* study. In this way, it updates the first edition of *World Trends*. While the first *World Trends* covered the breadth of issues, this second edition gives depth with a focus on four specific trends signalled in the *Keystones* study, taking forward UNESCO's role of enhancing knowledge and understanding through the high quality research relevant to building knowledge societies.

The research for this report would not have been possible without the continued support of the Government of Sweden, for which I am deeply grateful. I wish to thank also the Internet Society, the Open Society Foundations, the Center for Global Communication Studies at the University of Pennsylvania's Annenberg School for Communication, the University of Oxford, and the World Association of Newspapers and News Publishers (WAN-IFRA).

I am convinced that *World Trends in Freedom of Expression and Media Development – Special Digital Focus 2015* will become a reference for Governments, civil society, the private sector, academics as well as students, at this time when freedom of expression has never been so important.

Irina Bokova

Irina Bokova

I. INTRODUCTION

In 2011, UNESCO's 195 Member States approved a resolution during their 36th General Conference requesting the Organization to 'monitor, in close cooperation with other United Nations bodies and other relevant organizations active in this field, the status of press freedom and safety of journalists, with emphasis on cases of impunity for violence against journalists…and to report on the developments in these fields to the biannual General Conference.'

To carry out this mandate and with the support of the Government of Sweden, UNESCO embarked in 2012 on a large-scale research project with an advisory group of 27 leading international experts. Based on this research, a summary report on trends in press freedom and the safety of journalists between 2007 and mid-2013 was presented to the 37th General Conference in November 2013, in the form of an overview document highlighting relevant key findings. The final publication on *World Trends in Freedom of Expression and Media Development* was launched by UNESCO's Director-General in Stockholm, Sweden in March 2014 and then presented in all five UNESCO regions.

The first *World Trends* report filled an important gap in contemporary media and communications research. While other studies and reports offered snapshots of specific dimensions or regions, UNESCO's *World Trends* was the first to present a systematic trend analysis of the multiple aspects of media freedom, pluralism, independence, and safety, while giving special attention to gender-sensitive considerations.

Given the success of the first *World Trends* report and the need for additional research, UNESCO led a second edition in the series, focusing in depth on selected digital-era trends. *World Trends in Freedom of Expression and Media Development – Special Digital Focus 2015* provides a substantive analysis of key areas identified in the first *World Trends* as particularly relevant for further study, namely the issues of: online hate speech, protection of journalism sources, and the role of internet intermediaries in fostering freedom of expression, as well as continued focus on the safety of journalists. It also builds on issues raised in the 2015 UNESCO study titled *Keystones to foster inclusive Knowledge Societies.*

The current publication therefore contains four thematic chapters:

1. **Countering Online Hate Speech** provides a global overview of the dynamics characterizing hate speech online and some of the measures that have been adopted to counteract and mitigate it, highlighting trends in good practices that have emerged at the local and global levels. There is a comprehensive analysis of the international, regional and national normative frameworks developed to address hate speech online, and their repercussions for freedom of expression, and there is particular emphasis on social and non-regulatory mechanisms that may be considered to help to counter the production, dissemination and impact of hateful messages online.

2. **Protecting Journalism Sources in the Digital Age** draws on research covering 121 UNESCO Member States, which updates an earlier study of these countries by the NGO Privacy International in 2007. The chapter shows how legal frameworks that support protection of journalistic sources, at international, regional and country levels,

have come under significant strain in the intervening years. They are increasingly at risk of erosion, restriction and compromise. This is a trend that signifies a direct challenge to the established universal human rights of freedom of expression and privacy, and one that constitutes a particular threat to the sustainability of investigative journalism. A recommendation for consideration from this research is the proposal of an 11-point research tool for assessing the effectiveness of legal source protection frameworks in the digital age.

3. **Fostering Freedom Online: The Role of Internet Intermediaries** sheds light on internet intermediaries – the services that mediate online communication and enable various forms of online expression. It shows how they both foster and restrict freedom of expression across a range of jurisdictions, circumstances, technologies and business models. According to the UN Guiding Principles for Business and Human Rights, while states have the primary duty to protect human rights, businesses have a responsibility to respect human rights, and both should play a role in providing remedy to those whose rights have been violated. This chapter applies the 'protect, respect, and remedy' framework to the policies and practices of companies representing three intermediary types (internet service providers, search engines, and social networking platforms) across 10 countries. The three case studies highlight challenges and opportunities for different types of intermediaries within the trend of their increasing importance.

4. **Safety of Journalists** examines recent trends in the safety of journalists, presenting UNESCO statistics for 2013 and 2014, and tracking other developments up to August 2015. It follows the framework of the previous UNESCO report *World Trends* report, including physical safety, impunity, imprisonment of journalists, and a gender dimension of the issues. Additionally, the chapter examines the unprecedented trend of the strengthening of normative international standards, as well as new developments in practical mechanisms, improvement in UN inter-agency cooperation, greater collaboration with the judiciary system and security forces, and research interest in the subject.

In addition to contributing to this report and to UNESCO's comprehensive study of internet-related issues (see *UNESCO: Promoting Freedom of Expression and Media Development*), the chapters on online hate speech and the role of intermediaries have also been published as longer, stand-alone publications in the Organization's flagship Series on Internet Freedom.

Particular consideration is given throughout this new *World Trends* study to gender equality, one of UNESCO's two global priorities. As in the first *World Trends* report, gender is primarily conceptualized here as referring to women journalists' experiences and the effect of policies and practices on women.

The trends identified in this report shed light on the shifting landscape of opportunities and challenges for freedom and expression and media development, particularly those brought about by digital technologies. Through such sharing of knowledge and good practices, UNESCO works to advance human rights in the digital era: countering online hate speech, protecting journalism sources, fostering freedom online through sharing good practices for internet intermediaries, and enhancing the safety of journalists, both online and off-line.

II. UNESCO: FOSTERING FREEDOM OF EXPRESSION AND MEDIA DEVELOPMENT

UNESCO is the UN agency with a mandate to defend freedom of expression, instructed by its Constitution to promote 'the free flow of ideas by word and image'. This mission is reinforced by the Universal Declaration of Human Rights, which affirms, 'Everyone has the right to freedom of opinion and expression.' Freedom of expression, and its corollaries of freedom of information and press freedom, applies to all media, including traditional print and broadcast media, as well as newer digital media.

In 2013, UNESCO's General Conference of 195 Member States adopted Resolution 52, which recalled Human Rights Council Resolution A/HRC/RES/20/8, 'The Promotion, Protection and Enjoyment of Human Rights on the Internet', affirming that the same rights that people have off-line must also be protected online. Such rights are relevant across UNESCO's areas of competence and are critical for sustainable development, democracy and dialogue. The tracking of trends by UNESCO in regard to these rights, and especially concerning the right to freedom of expression, has been called for by Member States. Agreed at the 36th session of the General Conference, Resolution 53 required the Organization to '(m)onitor, in close cooperation with other United Nations bodies and other relevant organizations active in this field, the status of press freedom and safety of journalists, with emphasis on cases of impunity for violence against journalists, including monitoring the judicial follow-up through the Intergovernmental Council of the International Programme for the Development of Communication (IPDC) and to report on the developments in these fields to the biannual General Conference'. This underpinned the first *World Trends Report on Freedom of Expression and Media Development*,[1] which was launched in six cities worldwide, and includes six regional sub-studies. This current study continues the mandate, and uses the conceptual framework of the first *World Trends* report, which highlights issues of freedom, pluralism, independence, safety and gender. Also informing the current study is the mandate of the 37th General Conference in 2013, where Resolution 52 called for a comprehensive and consultative study of four dimensions of the internet as relevant to the remit of UNESCO. Published as *Keystones to foster inclusive Knowledge Societies,* that study examined access to information and knowledge, freedom of expression, privacy and the ethical dimensions of the information society.

Building on this background, UNESCO recognises that as digital technologies become ever more central to societies, so the issues around online freedom of expression, and its interface with the off-line world, also call for attention by the Organization. An example here is the safety of journalists and the issue of impunity, which is one of the chapters in this *World Trends* report. What happens in this realm in the practical world has a major bearing on what happens in the online dimensions – and vice versa. A lack of safety in one sphere has repercussions for safety in the other sphere. Hence, UNESCO is increasingly sensitive to the interfaces.

At the programme level, UNESCO works worldwide to promote freedom of expression on all platforms, both online and off-line and the inter-relations between the two. The focus

[1] http://www.unesco.org/new/en/world-media-trends

is on two dimensions, reflecting the output and the input sides of the communication process:

The first dimension of free expression is the right to *impart* information and opinion. This is the foundation for the right to press freedom, which refers to the freedom to publish to a wider audience. In the digital age, this right is especially relevant to anyone who uses traditional or social media. For UNESCO, effective press freedom is based on media freedom, pluralism, independence and safety. This applies to all media, including creative media and social media, and not just to the news media. Within this perspective, the matter of independence is of special relevance to those who use press freedom to do journalism. Independence depends on freedom and pluralism, and in the case of journalism, whether online or off-line, and this is founded upon the existence of professional standards for the production and circulation of verifiable information in the public interest.

In a nutshell, freedom of expression is the parent of press freedom – understood as the use of the right to impart information on a mass scale. Media freedom, pluralism, independence and safety constitute the essential enabling environment for the exercise of press freedom. It is within this context that professional journalism, as an offspring of freedom of expression, can flourish and make its contribution to building knowledge societies.

The second dimension of freedom of expression is the right to *seek and receive* information, which is the foundation of the right to information. In turn, this is one of the foundations of transparency, which is recognised as essential for development and democracy. Huge advances in transparency are enabled by digital technologies, as regards both public and private institutions, allowing for unprecedented accountability and citizen empowerment.

These two dimensions of freedom of expression are increasingly intertwined with the right to privacy, with potential synergies as well as tensions. Strong privacy can strengthen the ability of journalism to draw on confidential sources for public interest information, but it can also weaken transparency and conceal information in which there could be legitimate public interest. Weak privacy can lead to journalistic sources withholding information or practising self-censorship because of a fear of being arbitrarily monitored. Weak privacy may also enable an over-reach in transparency, amounting to an unjustified intrusion into individuals' personal lives. Trust in the benefits of digital communications can be affected by how a society addresses the right to privacy with the two dimensions of the right to freedom of expression.

Much of the work of UNESCO provides insight into how the two rights can each be respected, online as well as off-line – and where these interface, as well as how they can be balanced harmoniously in the public interest where necessary. The Organization does this through providing research, monitoring, awareness raising, advocacy, capacity building, and technical advice. UNESCO's International Programme for the Development of Communication (IPDC) also provides grant support for relevant projects for free, pluralistic, independent and secure media, whether online or off-line.

At the standard-setting level of defending online freedom of expression and privacy, UNESCO has actively been involved in and contributed to global and regional processes, including the NETmundial Internet Governance Principles and Roadmap for the future evolution of the Internet governance, the Council of Europe's Recommendation on Internet freedom, the African Declaration on Internet Rights and Freedoms, and the European Union's Seventh Framework Programme's project on 'Managing Alternatives for Privacy, Property and Internet Governance'.

In addition, the Organization advocates globally for online freedom of expression and privacy and engages with relevant stakeholders through global, regional and national fora, initiatives and meetings. Such fora include, among others, the Internet Governance Forum (IGF), the WSIS process, the NETmundial Initiative, the International Association for Media and Communications Research, the Global Media Forum, the Freedom Online Coalition and various regional IGFs.

As a result of Resolution 52, and as noted above, UNESCO produced the *Keystones to foster inclusive Knowledge Societies* study, in fulfilment of the request by Member States to focus on access to information and knowledge, freedom of expression, privacy and the ethical dimensions of the information society. The activity is reported to the 38th General Conference in the framework of the Report by the Director-General on the implementation of the World Summit on the Information Society (WSIS) outcomes.[2]

As per the mandate, UNESCO produced the study by convening an inclusive multi-stakeholder process which included governments, private sector, civil society, international organizations and the technical community. In July 2014, an online questionnaire was launched and inputs were solicited through social media and major fora, as well as directly sought from Member States and more than 300 experts and organizations, representing civil society, academia, the private sector, the technical community and intergovernmental organizations. By the end November 2014, UNESCO had received 200 solid responses to the questionnaire. Input to the study was also sought at global fora on internet-related issues, and a thematic debate on online freedom of expression and privacy was held at the 29th meeting of the Council of the IPDC in November 2014. In parallel to the multistakeholder consultations, UNESCO commissioned a series of publications on specific sub-themes to provide in-depth analysis and recommendations to its Member States and other stakeholders on internet freedom issues. These sub-studies contributed to the wider internet study, with some also being published as stand-alone volumes in the flagship Series on Internet Freedom.[3]

In addition to the sub-studies that have also contributed to three chapters in this present publication (i.e., *Countering Online Hate Speech*, *Protecting Journalism Sources in the Digital Age*, and *Fostering Freedom Online: The Role of Internet Intermediaries*), UNESCO

[2] http://unesdoc.unesco.org/images/0023/002341/234144e.pdf

[3] See UNESCO. UNESCO Series on Internet Freedom. http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/publications-by-series/unesco-series-on-internet-freedom/

has also commissioned a wide range of research within the framework of the Series on Internet Freedom, in the context of the 2013 Resolution 52 by the General Conference:

1. **Building Digital Safety for Journalism: A Survey of Selected Issues**: In light of limited global understanding of emerging safety threats linked to digital developments, UNESCO commissioned this research within the Organization's on-going efforts to implement the UN Plan of Action on the Safety of Journalists and the Issue of Impunity. In examining cases worldwide, the publication serves as a resource for a range of actors by surveying evolving threats and assessing preventive, protective and pre-emptive measures. It shows that digital security for journalism encompasses, but also goes beyond, the technical dimension. Recommendations are proposed for consideration by Member States, regarding governments, journalism contributors and sources, news organizations, trainers, corporations and international organizations.

2. **Principles for Governing the Internet: A Comparative Analysis**: This research reviews more than 50 internet-specific declarations and frameworks relevant to internet principles, prompted by a need for a specific review of the declarations and frameworks from the perspective of UNESCO's mandate. The publication shows that while each of these documents has its own value, none of them fully align with UNESCO's priorities and mandate. It there puts forward for Member States' consideration the concept of 'Internet Universality' as the Organization's clear identifier for approaching the various fields of internet issues and their intersections with UNESCO concerns. This concept has relevance to the Organization's work in many areas – including online freedom of expression and privacy; efforts to advance universality in education, social inclusion and gender equality; multilingualism in cyberspace; access to information and knowledge; and ethical dimensions of information society.

3. **Online Licensing and Free Expression**: UNESCO commissioned research on the topic of online licensing and free expression, particularly as regards journalism. Restricting access to using a means of communication is a matter directly related to press freedom; it has emerged as a complimentary approach to the longer standing practices of filtering and blocking, which impact particularly on the right to seek and receive information. From the point of view of international standards, free expression is the norm and restrictions the exception. When registration serves as licensing in the sense of being both compulsory and exclusionary, it may be seen as a form of prior restraint. Therefore, strict tests are needed to ensure that registration can be justified by the international standards of necessity, proportionality, due process and legitimate purpose. The purpose of this research is to provide contemporary and evidence-based answers to questions around the issue of by-permission publishing online that have been raised by recent policy, legal and regulatory regimes.

4. **Privacy and Media and Information Literacy**: UNESCO is conducting global research into privacy and media and information literacy (MIL). The issue of internet users having MIL competencies about the different dimensions of privacy is explored, such as public awareness of privacy rights in cyberspace, including national data protection regimes; ability to evaluate how privacy is respected in digital content and

communication that is accessed by a user; and ability to evaluate legitimate limitations of privacy online. The research seeks data on these areas, both by marshalling publicly available data in specific countries and regions, and by analysis of MIL practice in the same areas.

5. **Balancing Privacy and Transparency**: UNESCO commissioned global research on balancing privacy with transparency, the latter being assessed in terms of its relationship to freedom of expression. The research unpacks the complexity of the subject through both normative and empirical information, extending analysis to actors across individuals, civil society, private sector and government. The issue of user trust in the belief that personal data will not be illegitimately rendered transparent is also covered. Risks to personal privacy from transparency will be outlined, as are risks to transparency from privacy. There will be analysis of cases showing the issues and the lessons arising. Good practices in reconciling privacy and transparency are identified in terms of their correspondence with international standards.

6. **Privacy and Encryption**: This study discusses the availability of different means of encryption and their possible applications, providing a short overview of the state of the art in encryption technologies deployed in the internet and communications industries. It analyses the relationship between encryption and human rights, at the international level including relevant cases at national levels. The study provides an overview of legal developments with respect to government restrictions on encryption in selected jurisdictions and reviews options on encryption policy at the international level, including ideas for enhancing 'encryption literacy'.

Through the multi-stakeholder consultation and the sub-studies, UNESCO identified the four fields of research as being interdependent building blocks for the internet. Published as *Keystones to foster inclusive Knowledge Societies*, the study underlined the widespread interest in a future for the internet as an open, trusted and global resource that is equally accessible to all across the world. It analysed issues for technology and policy providing support greater and more equitable access to information and knowledge, for strengthening freedom of expression as an instrument of democratic processes and accountability, and for reinforcing the privacy of personal information.

The *Keystones* study found that freedom of expression is not an inevitable result of new technologies. Rather, freedom of expression must be supported by policy and practice, and requires trust in the internet as a safe channel for expressing one's views. Rising concerns over surveillance and internet filtering, for example, have led to the perception of freedom of expression on the internet becoming threatened, requiring major efforts to instil trust in privacy, security and the authenticity of information and knowledge accessible online, and to protect the safety and dignity of journalists, social media users and those imparting information and opinion online.

Moreover, *Keystones* established that freedom of expression online is linked to the principle of openness, particularly in regard to the international standards that advocate transparency in relation to restrictions on the right to expression. Open opportunities to share ideas and information on the internet are integral to UNESCO's work to promote freedom of expression, media pluralism and intercultural dialogue. For UNESCO, freedom of expression online is also a question of how people use their access to the internet and related ICTs to express themselves. Media and information literacy for all men and women is relevant to this question, including youth engagement and the countering of all forms of hatred, racism and discrimination, as well as violent extremism, in digital contexts, ranging from email to online video games.

To discuss the draft of the *Keystones* study, UNESCO organized a conference on 'CONNECTing the Dots: Options for Future Action', with some 400 participants representing five continents, at UNESCO headquarters in Paris in March 2015. The event provided a platform for exploring the findings of the study in preparation for its finalisation and featured presentations by a wide range of speakers from around the world. With overwhelming agreement, the multi-stakeholder gathering adopted an Outcome Document that underscored the significance of the internet for human progress and its role in fostering inclusive knowledge societies. The Outcome Document affirms the human rights principles that underpin UNESCO's approach to internet-related issues, and supports the Internet Universality principles that promote a Human **R**ights-based, **O**pen internet, which is **A**ccessible to all and characterized by **M**ultistakeholder participation (R.O.A.M). The *Keystones* analysis is that the four principles provide the guiding logic for supporting the further development of the internet in ways that will enhance access to information and knowledge, freedom of expression, privacy and ethics.

A resolution adopted by UNESCO's 196th Executive Board in April 2015 recommended that the Outcome Document of the CONNECTing the Dots conference be considered by the 38th session of the General Conference and forwarded as a non-binding input to the post-2015 Development Agenda; the UN General Assembly overall WSIS review process; and the high-level meeting of the General Assembly established by General Assembly Resolution 68/302. This Outcome Document is reflected in the options for future action, as outlined in the *Keystones* study.

In parallel, UNESCO has helped to shape the post-2015 sustainable development agenda by convening WSIS Action Line meetings and IGF events to highlight the crucial role of free, independent and pluralistic media and Internet Universality principles in the sustainable development goals. At the WSIS Forum 2015, UNESCO presented *Keystones* and organized the 10th facilitation meeting of WSIS Action Line C9 Media. Three workshops and an open forum were approved for the 10th IGF in Brazil in November 2015.

Through cutting-edge research and contributions to multi-stakeholder dialogue, UNESCO has thus engaged across the board, with the intention of strengthening the fundamental rights of freedom of expression and privacy, across online and off-line realms, within the ever-deepening digital age.

III. COUNTERING ONLINE HATE SPEECH[4]

[4] This chapter is drawn from Gagliardone, I. et al. 2015. Countering Online Hate Speech. UNESCO Series on Internet Freedom. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002332/233231e.pdf

1. INTRODUCTION

As more people participate in online discourse, so more attention is being given to the trend of hate speech on the internet. It is also evident that in the aftermath of dramatic incidents, calls have commonly been made for more restrictive or intrusive measures to contain the internet's potential to spread hate and violence, although the links between speech and action, and online content and off-line violence, are not clear cut. To understand the issue, and trends in response to it, requires analysis of several matters, starting with how we conceptualise the phenomenon. This chapter provides a conceptualisation with reference to recent debates, and further assesses evolutions in international normative standards. It recognises the trend of private internet companies as increasingly becoming actors in regard to hate speech, and the applicable international standards for them. Finally, it examines the wider significance of emerging trends in five social responses to online hate speech. These are: i) research efforts to monitor how online hate speech emerges and spreads, developing early warning systems and methods to distinguish among different typologies of speech acts; ii) coordinated actions by civil society members seeking to create national and international coalitions; iii) initiatives to encourage social networking platforms and internet service providers to play a more robust role in actively responding to online hate speech; iv) media and information literacy campaigns and initiatives aimed at preparing users to interpret and react to hateful messages; and v) news media moderation of hate expression. Finally, the chapter points to the key issues for trends ahead in regard to understandings of hate speech and jurisdictional matters, with a summation of key points.

1.1 A BROAD CONCEPTUALISATION

Hate speech lies in a complex nexus with: freedom of expression; individual, group and minority rights; and concepts of dignity, equality and safety of person. Its definition is often contested. As outlined in section 1.2 below, in international normative standards and many laws, hate speech equates to expression that advocates incitement to harm based on the target being identified with a certain social or demographic group. In the case of race-based hatred, international law includes those related expressions that could foster a climate of prejudice and intolerance. In common parlance, definitions of hate speech tend to be broader still, extending to encompass words that insult those in power or are derogatory of individuals who are particularly visible. That there exist these variations in referents shows an ongoing trend that there is no single agreed understanding, and that the term 'hate speech' continues to be shorthand whose meaning may cover a range of speech.

Conceived in this widest sense, 'hate speech' attracts concern not only because it is often seen as being an in principle affront, but also because it is often assumed that it may fuel actions that lead to the violations of rights in practice. Although the connection between expression and action is far from being automatic, especially at critical times such as during elections, sensitivity is greater. At the same time, the use of the term 'hate speech' may also be prone to manipulation: accusations of fomenting hate speech may be traded among political opponents or used by those in power to curb dissent and criticism.

For the purposes of covering the widest ground, this chapter uses the term 'hate speech' pragmatically to enable a review of the range of definitions and practices gathered under a wide rubric. Thus although a definitive conceptualisation of hate speech is elusive, this chapter uses the term to cover speech that serves degrading or dehumanizing functions. Drawing on the work of New York University School of Law professor Jeremy Waldron, the chapter recognises that an expression that can be considered hateful entails two meanings. The first is a message to the targeted group and it functions to dehumanize and diminish members assigned to this group. The second meaning is to let others with similar views know they are not alone, to reinforce a sense of an in-group. 'Hate speech' in this sense relies on tensions, which it seeks to reproduce and amplify; it unites and divides at the same time. It creates 'us' and 'them'. For this chapter, the term 'hate speech' is generally used in this wider sense of oppositional identity involving group identity, not restricting the meaning to speech where there is specific incitement of harm. It is also used without the assumption that such speech acts to stimulate harmful practical actions.

There are clearly a number of axes along which hatred can be constructed, such as race, ethnicity, language group, gender, religion, sexual preference or nationality. However, it is also clear that strong opinions about ideas should not per se be conflated with hate speech. Hate speech concerns antagonism towards people, and not abstract ideas. It does not encompass hostility to political ideologies, faiths or beliefs even if it is on this basis that human targets may be categorised; a distinction must be kept in mind to limit 'mission creep' of the term 'hate speech'.

Because hate speech as a concept has been contested as too wide-ranging and open to manipulation, a trend has emerged in recent times to promote narrower conceptions, including limiting the label only to hateful cases that could be described as 'dangerous speech' and 'fear speech'. These have been advanced to focus on the ability of speech to cause practical harm and lead to violent outcomes. While hate speech is found – in some form or guise – in many contexts, the concept of 'dangerous speech' emerged around 2010. As advanced by Susan Benesch, of the Berkman Center for Internet and Society, aims at isolating acts that have a significant probability of 'catalyzing or amplifying violence by one group against another'. The concept of 'fear speech' put forward by Antoine Buyse, director of the Netherlands Institute of Human Rights, has also been recently advanced to emphasise language that is able to progressively create a siege mentality and that may ultimately lead to legitimizing violent acts. Based on the study of mass atrocities, the idea of 'fear speech' offers a pathway for understanding whether the preconditions for violence may gradually emerge, and for possibly identifying critical points when counter measures may be most effective. Finally, there have been growing attempts to move the general debate on 'hate speech' beyond simply identifying, regulating and distinguishing countermeasures, by initiating research into who is expressing hatred and why. Such research seeks understanding of the unique characteristics and causes of a fast evolving phenomenon, as a precondition for seeking ‘solutions’. Hate Speech, both on and offline, can therefore nowadays be assessed in terms of these nuances in terms of developing responses appropriate to the particular problem at hand.

The proliferation of online hate speech, observed by Rita Izsák, the UNHRC Special Rapporteur on Minority Issues, in her report A/HRC/28/64, poses a new set of challenges. While statistics offering a global overview of the phenomenon are not available, both social networking platforms and organizations created to combat hate speech have recognized that hateful messages disseminated online are increasingly common and that there is unprecedented attention to developing adequate responses. According to Hatebase, a web-based application that collects instances of online hate speech worldwide, the majority of cases of hate speech target individuals based on ethnicity and nationality, but incitements to hatred focusing on religion and class have also been on the rise.

Online hate speech renders some legal measures elaborated for other media ineffective or inappropriate, and it calls for approaches that are able to take into consideration the specific nature of the interactions enabled by digital information and communication technologies (ICTs). There is the danger of conflating a rant tweeted without thinking of the possible consequences, with an actual threat that is part of a systematic campaign of hatred. There is the difference between a social media post that receives little or no attention, and one that goes viral. In addition, there are the complexities that governments and courts may face, for example when trying to enforce a law against a social networking platform headquartered in a different country. Therefore, while online the content of hate speech is not intrinsically different from similar expressions found off-line, there are peculiar challenges unique to it.

These challenges may be identified in terms of digital permanence, itinerancy, anonymity and cross-jurisdictional character:

- First, hate speech can stay online for a long time in different formats across multiple platforms. Andre Oboler, CEO of the Online Hate Prevention Institute, has stated: ‘The longer the content stays available, the more damage it can inflict on the victims and empower the perpetrators.’ Platforms’ architectures may allow topics to stay alive for shorter or longer periods. Twitter’s conversations organized around trending topics may facilitate the wide and quick spread of hateful messages, but they also offer the opportunity for influential speakers to shame such messages and possibly end popular threads. Facebook, on the contrary, may allow multiple threads to continue in parallel and go unnoticed outside of a narrow community, creating longer-lasting spaces when certain groups are vilified.
- Second, online hate speech can also be itinerant. When content is removed, it may find expression elsewhere, possibly on the same platform under a different name or in a different online space. If a website is shut down, it can reopen using a web-hosting service with less stringent regulations or it can relocate to a country with laws

imposing a higher threshold for hate speech. The itinerant nature also means that thoughts that would not have found wide public expression in the past may now be visible to large audiences through a range of platforms.

- Third, the endurance of hate speech materials online is unique, due to the low preservation cost and the potential for immediate revival, which ensures its continued relevance in particular spheres of discourse.

- Fourth, anonymity can also present a challenge to dealing with online hate speech. Law professors Danielle Keats Citron and Helen Norton perceive that: 'The internet facilitates anonymous and pseudonymous discourse, which can just as easily accelerate destructive behaviour as it can fuel public discourse.' Some governments and social media platforms have sought to enforce real name policies, but such measures have been deeply contested as they hit at the right to privacy and its intersection with free expression. In addition, the majority of online trolling and hate speech attacks come from pseudonymous accounts, which are not necessarily anonymous to everyone. Genuinely anonymous online communications are rare, as they require the user to employ highly technical measures to ensure that he or she cannot be easily identifiable. Yet perceived anonymity can nevertheless encourage some actors to consider that their online speech cannot be tracked to them.

- Fifth, as the internet is not governed by a single entity, concerned individuals, governments and non-governmental organizations may need to address internet intermediaries on a case-by-case basis, leaving the owners of a specific online space to decide how to deal with users' actions. Internet intermediaries risk becoming private tribunals in deciding how content should be regulated, a point that is discussed in more depth later in this report. A further complication is the internet's transnational reach, raising issues of cross-jurisdictional co-operation in regard to legal mechanisms for combatting hate speech. While there are mutual legal assistance treaties in place among many countries, these are characteristically slow. The transnational reach of many private-sector internet intermediaries may provide a more effective channel for resolving issues in some cases, although these bodies are also often impacted upon by cross-jurisdictional appeals for data. Unlike the dissemination of hate speech through conventional channels, hate speech dissemination online often involves multiple layers of actors, whether knowingly or not. When perpetrators make use of an online social platform to disseminate their hateful message, they do not only hurt their victims, but may also violate the platform's terms of service and at times even state law, depending on their location. The victims, for their part, may feel helpless in the face of online harassment, not knowing to whom they should turn for help.

Based on the analysis outlined above, this chapter covers emerging trends in online hate speech. Its focus is on developing and developed countries, but it also acknowledges that the biggest problems of online hate speech are currently in countries where there is high internet connectivity. However, this may portend similar developments elsewhere as more people become connected around the world. Some of the responses assessed here could therefore be considered for adaptation proactively and early on, rather than only after the emergence of the problem.

1.2 LEGAL AND SOCIAL RESPONSES

The most debated responses to online hate speech have focused primarily on legal definitions and legal means, but this approach involves risks and limitations.

First, law is entangled with power, and can sometimes be abused to limit legitimate speech under a claimed rationale of punishing hate speech. There can be collateral damage to speech that, even if highly objectionable to some, does not transgress international standards for freedom of expression. The key issue here is where 'hate speech' and its regulation lies in the three categories of expression identified in 2012 by Frank La Rue, the then UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression. In his report, A/66/290, he distinguished between:

- Expression that constitutes an offence under international law and can be prosecuted criminally;
- Expression that is not criminally punishable but may justify a restriction and a civil suit; and
- Expression that does not give rise to criminal or civil sanctions, but still raises concerns in terms of tolerance, civility and respect for others.

It is evident that not all categories above call for entail legal responses, and also that those which do may entail differences between criminal and civil law responses. It is also evident that social responses may have preventative and other roles to play in all instances. These points have a bearing on how different instances of hate speech are understood and addressed.

Second, the UN Special Rapporteur on minority issues observes that hate crimes rarely occur without prior stigmatization and dehumanization of targeted groups and incidents of incitement to hate. At the same time, she notes: 'Only the most egregious forms of hate speech, namely those constituting incitement to discrimination, hostility and violence, are generally considered unlawful.' This highlights that while there is a role for law, legal measures cannot be seen as an appropriate response to the full spectrum of speech that may (though does not necessarily) contribute to a climate for hate-based actions.

Third, in regard to hate speech, a purely legal lens can miss out on how societies evolve through contestation and disagreement. Although hate speech is offensive, it can also be thought of as a window into deeply rooted tensions and inequalities, which need addressing beyond pure speech issues, and beyond the online dimension.

This analysis highlights why it is important to track trends in both legal and social dimensions around hate speech. Hence, this chapter proceeds now to a broad overview of the evolutions in the most important international legal instruments that regulate hate speech, and it then follows this with particular emphasis on social responses.

2. METHODOLOGY

The research strategy for this chapter draws extensively from the more detailed UNESCO study *Countering Online Hate Speech*, which in turn combined multiple techniques of data collection and analysis, beginning with an extensive literature review, including the legal literature that examines how hate speech is addressed in different countries and continents, and ethnographic studies examining user behaviour in hate-based online spaces. Given the novelty of the phenomenon under investigation and its fast-evolving nature, the literature review also included non-academic articles published by experts on their blogs and in specialist publications and major online newspapers and magazines.

The chapter also made use of semi-structured interviews were carried out with relevant stakeholders, from representatives of social media platforms, including Facebook and Google, to members of civil society organizations, politicians and technical experts. The analysis also covered content produced by non-governmental organizations that have launched diverse media and information literacy (MIL) initiatives to counter online hate speech, and the terms of service agreements of online media platforms, including Facebook, Twitter and YouTube. The aim was to understand the actual monitoring and management of online content. In addition, the research analysed how MIL campaigns target different audiences and with what results; and the strategies adopted by anti-discrimination groups or coalitions to lobby social media organizations. While the range of views about, and responses to, online hate speech is wide, common questions were asked for each case.

3. FRAMEWORKS

3.1 FRAMEWORKS OF INTERNATIONAL LAW

Hate speech touches on contested issues of dignity, equality, security of person and freedom of expression. Hate speech is not explicitly mentioned in many international human rights documents and treaties, but it is indirectly called upon by principles related to human dignity, and equality, and freedom of expression. Certain expression may be identified as directly impugning dignity, including on a group basis. In some cases, expression may also be identified as advocating incitement to discrimination, which would violate the right to equality (although the link between speech and practice is a distinct matter). A further issue is the right to life, liberty and security of person, and whether certain expression constitutes a direct harm in this respect, as in cases of calls for attacks on the persons framed as being in a particular group.

These rights are all provided for in the 1948 Universal Declaration of Human Rights. Taking them all together, everyone has the right to freedom of expression, the right to be protected against violations of dignity and equality, and the right to life and security. In other words, everyone has the right to be protected against hate speech insofar as such speech incorporates violations of these other rights. This involves a complex balancing of rights in ways that maintain as much as possible the essence of each right, and therefore processes and criteria for achieving such balance are vital. What is important to keep in mind, however, is that proportionality, necessity and legitimacy, balancing under the rationale of countering hate speech does not over-reach in regard to freedom of expression.

The UDHR was decisive in setting a framework and agenda for human rights protection, but the Declaration is non-binding. A series of binding documents were subsequently created to offer a more robust protection for rights. Of those, the International Covenant on Civil and Political Rights is the most important and comprehensive in addressing hate speech - although it does not explicitly use the term ‘hate speech’. Other more tailored international legal instruments also contain provisions that have repercussions for the definition of hate speech and identification of responses to it, such as: the Convention on the Prevention and Punishment of the Crime of Genocide (1951), the International Convention on the Elimination of All Forms of Racial Discrimination (1969), and the Convention on the Elimination of All Forms of Discrimination against Women (1981).

The ICCPR is the legal instrument most commonly referred to in debates on hate speech and its regulation. Article 19 provides for the right to freedom of expression, setting out the right and including general strictures for legitimate limitations. However, Article 20 expressly limits freedom of expression in cases of ‘advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence’. Cognizant of tensions surrounding Article 20, the Human Rights Committee has stressed that the article is fully compatible with Article 19.

To elaborate further, in the ICCPR, the right to freedom of expression is not an absolute right. It can legitimately be limited by states under restricted circumstances that are ‘provided by law and are necessary for respect of the rights or reputation of other’ or ‘for the protection of national security or of public order (ordre public), or of public health or morals’. In General Comment 34, the Human Rights Committee explains that limitations imposed by states may include online speech under Article 19(3) of the ICCPR, noting that such restrictions ‘generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3.’

Between Article 19(3) and Article 20, there is a distinction between optional and obligatory limitations to freedom of expression. Article 19(3) states that limitations on freedom of expression ‘**may** therefore be subject to certain restrictions,’ as long as such restrictions are provided by law and necessary to certain legitimate purposes. Article 20 states that any advocacy of hatred that constitutes incitement to discrimination, hostility or violence ‘**shall be** prohibited by law’. Despite indications on the gravity of speech offenses that should be prohibited by law under Article 20, there remains complexity. In particular there is a grey area in conceptualising clear distinctions between (i) expressions of hatred, (ii) expression that advocate hatred, and (iii) hateful speech that specifically constitutes incitement to the practical harms of discrimination, hostility or violence. Thus, while states have an obligation to prohibit speech conceived as ‘advocacy to hatred that constitutes incitement to discrimination, hostility or violence’, as consistent with Article 20(2), how to interpret such provision is not clearly defined. Consequently, limitations on freedom of expression, based on the ICCPR provision, may be open to abuse. The Camden Principles, a set of standards formulated by the NGO ARTICLE 19 in consultation with human rights experts, define specific criteria to avoid misapplication of Article 20(2). Article 20 must be interpreted narrowly to avoid its misuse.

The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) of 1965 also has implications for conceptualising forms of hate speech, although it too does not explicitly use the phrase. The ICERD differs from the ICCPR in three respects. First, its conceptualisation of hate speech is specifically limited to speech that refers to race and ethnicity. Second, the ICERD imposes an obligation on state parties that is stricter than Article 20 of the ICCPR, covering the criminalisation of racist ideas that are not necessarily inciting discrimination, hostility or violence. Third, the concept of ‘advocacy of hatred’ introduced in the ICCPR is more specific than discriminatory speech described in the ICERD, since it is taken to require consideration of the intent of author and not the expression in isolation. The mere dissemination of messages of racial superiority or hatred, or even incitement to racial discrimination or violence, shall be punishable in accordance to the ICERD. In the ICCPR, the intent to incite hatred needs to be proven in order for the speech to be prohibited under Article 20(2).

The Committee on the Elimination of Racial Discrimination actively addressed hate speech in 2002 in its General Recommendation 29, A/57/18, in which the Committee recommends state parties to take measures against the dissemination of ideas 'through the mass media and the internet' of caste superiority and inferiority; justifying violence, hatred or discrimination against descent-based communities. It called for strict measures against any incitement to discrimination or violence against the communities 'including through the internet'; and for raising awareness among media professionals of the nature and incidence of descent-based discrimination. These points, which reflect the ICERD's reference to the dissemination of expression, have particular significance for the internet, in that the expression of ideas in some online contexts may immediately amount to spreading them. This is also relevant for formerly private spaces that have begun to play a public role, as in the case of many social networking platforms.

Similarly to the ICERD, the Genocide Convention aims to protect groups defined by race, nationality or ethnicity, although it also extends its provisions to religious groups. When it comes to hate speech, however, the Genocide Convention is limited only to acts that publicly incite to genocide, recognised as 'acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group'.

Specifically gender-based hate speech (as distinct from discriminatory actions) is not covered in depth in international law. The Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), which entered into force in 1981, imposes obligations on states to condemn discrimination against women and 'prevent, investigate, prosecute and punish' acts of gender-based violence. The Human Rights Committee has also expressed 'grave concern at acts of violence and discrimination, in all regions of the world, committed against individuals because of their sexual orientation and gender identity'.

The extent to which expression links to such practical actions is a subject of debate. However, the UN Human Rights Committee in General Comment 28 called for states to 'provide information about legal measures to restrict the publication or dissemination' of pornographic material which portrays women as objects of degrading treatment.

In summary, the ICCPR norms allow for possible restrictions for respect for the rights or reputations of others, or for national security or public order, or public health and morals, which provision in certain contexts may apply to expressions that could be labelled 'hate speech'. The ICCPR further requires restrictions for 'national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence'. It is evident that these two dimensions constitute norms for a conditional basis to limit certain speech that could be categorised as falling within 'hate speech', provided that such restrictions are in law and necessary. Under the ICERD, there is a normative basis to restrict dissemination of ideas of racial superiority (which would also count as protecting respect for the human right to equality).

In recent trends, and responding to this complexity and risks of abuse of international norms to restrict legitimate speech, the UN has sought to create spaces for promoting a shared understanding of what hate speech is and how it should be addressed, as well as the relevance of human rights to the online realm. Also in recent times, governing bodies of the UN General Assembly, UNHRC, and UNESCO, have definitively recognised that all human rights apply both off and on line. These combined developments set the context for grappling with hate speech on the internet.

A milestone in the process has been that of the UN Office of the High Commissioner for Human Rights (OHCHR) organizing a series of consultative meetings. These led in 2012 to the formulation of the Rabat Plan of Action on the prohibition of 'national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence'. The Rabat Plan of Action acknowledges that, despite the obligations for states that are ICCPR signatories, many legal frameworks do not contain legal prohibition of such advocacy. In addition, some laws that do so use terminology that is inconsistent with Article 20 of the ICCPR. The Plan also proposes a six-part threshold test to identify hate messages, considering context, speaker, intent, content, extent of the speech and likelihood the speech could incite actual harm. In this sense, there is not an assumption that all hateful expressions would cause or translate into actual harm. Instead what is proposed is a method to pinpoint those expressions most requiring attention.

However, in the case of *online* hate speech, the emphasis that the Rabat Plan places on national level actors and especially states may underplay the significance of private sector social networking platforms that operate transnationally. These actors can play a highly meaningful role in interpreting hate speech and allowing or constraining expression. In addition, the Rabat Plan does not give much attention to issues of hatred on grounds such as gender, sexual preference or language spoken.

In addition to allowing states to take measures to limit hate speech, international law includes some provision for individuals to bring complaints about speech: the Human Rights Committee receives individual complaints related to the ICCPR, the Committee on the Elimination of Racial Discrimination receives complaints from the ICERD, and complaints related to the CEDAW are dealt with by the Committee on the Elimination of Discrimination against Women. However, individuals may only bring a complaint against a state that has explicitly allowed for such mechanisms.

Diverse opinions on balancing freedom of expression and limitations around hate speech find pronounced manifestation in regional human rights instruments. These documents complement international treaties as they reflect regional particularities that are not specified in treaties with universal reach. Regional instruments may be particularly effective to enforce the protection of human rights as in the case of the European Court of Human Rights, which decides more cases related to hate speech than the United Nations Human Rights Committee. Nevertheless, regional human rights instruments ought not to contradict established international norms, nor impose stronger limitations on fundamental rights. Most regional instruments do not have specific articles prescribing prohibition of hate speech, but they more generally allow states to limit freedom of expression – which provisions can be applied to specific cases. The paragraphs below examine how the right to freedom of expression and its limitations are defined at the regional level and how regional documents complement other texts that allow for definition and limitation of hate speech.

The **American Convention on Human Rights** describes limitations on freedom of expression in a manner similar to the ICCPR in Article 19(3). The Convention adds a specific limitation clause prohibiting prior censorship; however, in order to offer more protection to children, it allows prior censorship for the 'moral protection of childhood and adolescence'. The Organization of American States has adopted another declaration on the principles of freedom of expression, which includes a specific clause stating that, 'prior conditioning of expressions, such as truthfulness, timeliness or impartiality is incompatible with the right to freedom of expression recognized in international instruments.' The Inter-American Court has advised that preventive measures are incompatible with freedom of expression and that states should instead employ 'subsequent imposition of sanctions on those who are guilty of the abuses'. The Court also imposes a test for states willing to enact restrictions on freedom of expression, as they need to have met previously established grounds for liability, be defined by law, sought to achieve legitimate ends, and 'necessary to ensure' such ends. Finally, the Inter-American System has a Special Rapporteur on Freedom of Expression whose comprehensive study on hate speech concluded that the Inter-American Human Rights System differs from the UN and the European approach on a key point: the Inter-American system covers only hate speech that actually leads to violence, and solely such speech can be restricted.

The **African Charter on Human and Peoples' Rights** takes a different approach in Article 9(2), allowing for restrictions on rights as long as they are 'within the law'. This concept has raised issues, and there is a vast amount of legal scholarship on the so-called 'claw-back' clauses and their interpretation. The question mainly aims at the fact that countries can manipulate their own legislation and can weaken the essence of the right to freedom of expression. However, it is important to add that the Declaration of Principles on Freedom of Expression in Africa elaborates a higher standard for limitations on freedom of expression. It declares that the right 'should not be restricted on public order or national security grounds unless there is a real risk of harm to a legitimate interest and there is a close causal link between the risk of harm and the expression'.

In 1990, the Organization of the Islamic Conference (now Organization of Islamic Cooperation – OIC) adopted the **Cairo Declaration on Human Rights in Islam**, which states that human rights should be 'in accordance with the Islamic Shari'ah'. This clause is seen by some to impact the threshold for limitations and is why OIC member states have called for criminalisation of speech that extends beyond cases of imminent violence to encompass 'acts or speech that denote manifest intolerance and hate'.

The **Arab Charter on Human Rights**, adopted by the Council of the League of Arab States in 2004, includes in Article 32 provisions relevant for online communication, guaranteeing the right to 'freedom of opinion and expression, and the right to seek, receive and impart information and ideas through any medium, regardless of geographical boundaries'. Paragraph 2 states: 'Such rights and freedoms shall be exercised in conformity with the fundamental values of society.' This position is different to that of the Human Rights Committee General Comment No. 22, which states that, 'limitations on the freedom to manifest a religion or belief for the purpose of protecting morals must be based on principles not deriving exclusively from a single tradition.'

The **ASEAN Human Rights Declaration** includes the right to freedom of expression in Article 23. Article 7 of the Declaration provides for general limitations, affirming that, 'the realisation of human rights must be considered in the regional and national context bearing in mind different political, economic, legal, social, cultural, historical and religious backgrounds.' In this regard, the Office of the High Commissioner for Human Rights has drawn attention to the Vienna Declaration's provision that, notwithstanding differences, 'it is the duty of states, regardless of their political, economic and cultural systems, to promote and protect all human rights and fundamental freedoms.'

Some regional texts are potentially more restrictive of freedom of expression than international standards. However, other regional texts contain narrower tests for assessing the legitimacy of limitations of freedom of expression. The Charter of Fundamental Rights of the European Union, which declares the right to freedom of expression in Article 11, asserts in Article 54 that the Charter must not be interpreted as implying any 'limitation to a greater extent than is provided for therein'. The European Convention on Human Rights implies a strict test of necessity and proportionality for limitations to freedom of expression. The European Court of Human Rights distinguishes between hate speech and the right of individuals to express their views freely, even if others take offence.

The Council of Europe (CoE) in 2000 issued a General Policy Recommendation on Combating the Dissemination of Racist, Xenophobic and Anti-Semitic Material via the Internet. The 2001 CoE Convention on Cybercrime regulates mutual assistance regarding investigative powers, providing signatory countries with a mechanism to deal with computer data, including transnational online hate speech. In 2003 the CoE launched an additional protocol to the convention that addresses online expression of racism and xenophobia, which imposes an obligation on member countries to criminalise racist and xenophobic online insults based on 'race, colour, descent or national or ethnic origin, as well as religion'. Nine countries outside Europe have now signed or ratified the convention.

Important to all the above, it may be noted that recent international documents such as the aforementioned Human Rights Committee General Comment 34 (2011) and the Rabat Plan of Action (2011) have emphasised the mutuality between freedom of expression and protection against hate speech. The complexity in balancing freedom of expression and limitations as regards hatred accounts for the diversity of legal conceptions of hate speech around the world, and complicates the interpretation of law in any given case. However, any legal limitations must always be considered alongside the broader right to freedom of expression. According to General Comment 34, 'the relation between right and restriction and between norm and exception must not be reversed.'

3.2 FRAMEWORK FOR PRIVATE ACTORS

The international and regional legal instruments surveyed above are evolving a framework for *states* to address hate speech within their duty to promote and protect rights, which includes balancing the right to freedom of expression with rights to dignity, equality and security of person. When dealing with *online* hate speech, however, individual states are not always the most impactful actors. Internet intermediaries such as social networking platforms, internet service providers and search engines, stipulate in their terms of service how they may intervene in allowing, restricting, or channelling the creation and access to specific content. A vast amount of online interactions occur on social networking platforms that transcend national jurisdictions and have developed their own definitions of hate speech and measures to respond to it. For a user who violates the terms of service, the content he or she has posted may be removed from the platform, or its access may be restricted to not be viewable within a specific country.

The principles that inform terms of service agreements, and the mechanisms that each company develops to ensure implementation, have significant repercussions on the ability that people have to express themselves online as well as to be protected from hate speech. Most intermediaries enter into negotiations with national governments to an extent that varies according to the type of intermediary, areas where the company is registered, and the legal regime that applies. Internet service providers are the most directly affected by national legislation because they have to be located in a specific country to operate. Search engines have increasingly tended to adapt to the intermediary liability regime of both their registered home jurisdictions and other jurisdictions in which they provide their services, either removing links to content proactively or upon request by authorities. Online social network companies show a range of variations in approach.

Whatever the diversity in the sector, it has recently become clearer that all internet intermediaries operated by private sector companies, are also expected to respect human rights. This is set out in the 2011 Guiding Principles on Business and Human Rights elaborated by the United Nations OHCHR. The document emphasises corporate responsibility in upholding human rights. To do this, internet intermediaries, in line with other companies, should assess 'actual and potential human rights impacts integrating and acting upon the findings, tracking responses, and communicating how impacts are addressed.' The UN Guiding Principles also indicate that in cases in which human rights are violated, companies should 'provide for or cooperate in their remediation through legitimate processes.' In the case of internet intermediaries and conceptions of hate speech, this means that they should ensure that measures are in place to identify hate speech and provide a commensurate response.

These principles, however, are still struggling to find concrete reference in many intermediary policy positions, as well as in concrete implementation in everyday corporate practice. One issue is the extent to which a private sector entity has the right to set terms of service that may be more restrictive of speech than what a state is required to permit in terms of international standards such as the ICCPR. This is analogous in some respects to press freedom, in which a media outlet is entitled to set its own editorial policy for information it publishes. This holds notwithstanding that though social media are distinctly based upon the expressions of users, compared to news media where expressions emanate from those employed by the platform itself. Another issue is how companies, inasmuch as they follow international human rights standards, decide on the balance of rights, such as freedom of expression in relation to privacy, equality or dignity, and what redress exists.

Finally, there are issues related to how companies make decisions when national laws are not compliant with international human rights standards, such as for legitimate limits on freedom of expression. The situation is dynamic and continues to develop.

At the same time, there is an evolving trend in terms of which internet intermediaries have been developing disparate definitions of hate speech and guidelines to regulate it. Some companies do not use the term hate speech, but have a descriptive list of terms related to it. Yahoo!'s terms of service prohibit the posting of 'content that is unlawful, harmful, threatening, abusive, harassing, tortuous, defamatory, vulgar, obscene, libellous, invasive of another's privacy, hateful, or racially, ethnically or otherwise objectionable.' Similarly, Twitter does not explicitly prohibit hate speech, but alerts its users that they 'may be exposed to Content that might be offensive, harmful, inaccurate or otherwise inappropriate, or in some cases, postings that have been mislabelled or are otherwise deceptive.' It denies all liability for content. Twitter's terms of service are complemented by Twitter's Rules, a set of conditions for users, and Twitter has responded to hate speech-related content removal requests from governments and civil society.

Other companies make explicit reference to hate speech. YouTube's terms of service, for example, seek to balance freedom of expression and limitations to some forms of content. While stating that it 'encourages free speech', YouTube declares that it does 'not permit hate speech: speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status and sexual orientation/gender identity.' This definition is thus wider than the ICCPR's call for limitation only of speech that constitutes intentional advocacy of hatred that incites discrimination, hostility or violence. It is an example of how companies can be more restrictive than international law, and even some regional or national laws on hate speech.

Facebook's terms forbid content that is harmful, threatening or that has potential to stir hatred and incite violence. In its community standards, Facebook elaborates that 'Facebook removes hate speech, which includes content that directly attacks people based on their: race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases.' Microsoft has specific rules concerning hate speech for a variety of its applications. Its policy for mobile phones prohibits applications that 'contain any content that advocates discrimination, hatred, or violence based on considerations of race, ethnicity, national origin, language, gender, age, disability, religion, sexual orientation, status as a veteran, or membership in any other social group.' The company also has rules regarding online gaming, which prohibit any communication that is indicative of 'hate speech, controversial religious topics and sensitive current or historical events'. This is another example of how companies can be more restrictive than regional or international law on hate speech: 'Controversial religious topics and sensitive current or historical events' are not necessarily prohibited in international law, nor are they automatically considered discriminatory'. Nevertheless, in order to promote what they see as a safer online community, Microsoft has chosen to impose speech restrictive regulations on certain products that it offers. On the other hand, in certain jurisdictions, these terms of service may be more liberal than local legal limits.

Typically, only a small minority of users read the terms of service, and there are different levels of 'quality' among the various types of agreement. Analysis of the trends shows that it is not only how internet intermediaries define hate speech, but also how they enforce their definitions. An issue here is the liability of these intermediaries. Many intermediaries argue that they do not generate or control content online and therefore should have only limited liability. This is interpreted to exempt them from prior screening or moderation of content, and expose them only after publication in regard to cases where their attention is drawn to content that offends law and/or their terms of service. Different legal regimes on liability exist worldwide, with different impacts, although ultimately it is likely that there will be a single jurisdictional standard that can enforce an intervention by the company to limit a particular instance of online speech, although there are complexities as to where the entity is registered, where the data is held and where it can be served.

The notion of limited liability distinguishes internet intermediaries from news media companies. There are debates, however, over the extent to which news media should have limited liability for user-generated comments on their websites. Their practices and terms of service for moderating content, as well as their self-regulatory systems such as press councils, may at any rate have significance for internet intermediaries. For internet service providers, liability in a given jurisdiction is relatively straightforward. Similarly to other internet intermediaries, they can define their own parameters when offering a service, but since they are bound by the principle of territoriality, they tend to operate in terms of the laws of the country where they offer service. This makes them more responsive than other intermediaries to external requests to remove content.

The issue becomes more complex for social networking platforms with an international reach. Given the enormous amount of data they handle, social networking platforms mainly rely on notifications from users who report content they consider inappropriate, offensive or dangerous. The platforms then decide, mainly according to their terms of service, whether or not this content should be removed or if other actions need to be taken to restrict access to it or the ability of its authors to continue using the platform's service. In the absence of multiple national jurisdictional authority over the company, and the limited capacity and reach of any single jurisdiction except that in which the operation is domiciled, many intermediaries operate according to their own overarching global terms of service.

Apart from old guidelines leaked by employees of companies to which social networking platforms outsourced some aspects of content regulation, little is known about how terms of service translate in practice into what to keep and what to filter or remove. Some have suggested that Facebook has been developing a set of objective standards to act upon speech that it considers likely to provoke violence. But Facebook officials have indicated that the company tries to avoid a textbook approach and prefers to look at context as much as possible.

It is a trend that some companies have become more attentive to users' complaints over recent times. In 2012 Facebook introduced the possibility for users who flag content that they consider inappropriate to track their reports until the issue is resolved. It has also offered tools to ‘socialize’ reporting, allowing users to privately notify the author of a specific piece of content before formally asking Facebook for it to be removed. These new opportunities represent interesting additions to other measures to respond to perceived hate speech, even if evidence is lacking about how effective they have been over time and whether or not users are satisfied with the options they are being offered. At the same time, there are big debates about commercial actors acting as tribunals over permissible speech, notwithstanding the private status of these entities and their online properties. There is further discussion of this item later in this report.

In summary, the analysis above has given consideration to the landscape of international and regional norms and laws, and the emerging trends in transnational internet intermediary companies that have emerged as the main sites and actors regarding online hate speech and its regulation. Different definitions of hate speech are evident within a complex patchwork of international policies, which are applied differently by governmental actors and companies. While all actors should seek to abide by the norms in universal treaties, the practical reality is complicated by internet intermediaries’ relative autonomy and by their major role in online communications. At the same time, state-based regulatory responses may be slow to develop, complex to implement, and be vulnerable to political interference. In this context, social responses have emerged in response to perceived online hate speech.

4. ANALYSING SOCIAL RESPONSES

The analysis here seeks to offer a nuanced picture of how concerns about hate speech and violence are being manifested in a range of social responses. The sections that follow cover issues of monitoring, mobilizing, lobbying of intermediaries, empowering users through media and information literacy initiatives, and news media content moderation.

4.1 MONITORING AND ANALYSING HATE SPEECH

The climate for hate speech is likely to become the most conducive to violence in situations where the political stakes are high, such as during elections. This section analyses the broader issues arising from practical responses developed to deal with the potential of online hate speech emerging in such situations. One response that provides a backdrop for wider observations is the UMATI research project, which began in September 2012 ahead of the March 2013 elections in Kenya; it monitored online discourse to estimate both the occurrence and virulence of hate speech. The experiences provided stakeholders with an opportunity to analyse the issues and targets of hate speech and to collectively reflect on the potential that specific speech acts have to lead to violence or not.

In 2007, Kenya held the most contested and violent elections since it returned to multipartyism in 1991, leaving more than 1,000 people dead and 600,000 displaced. This was the first election in which new ICTs became an integral part of the electoral contest. Social media, emails and SMS text messages were used to unprecedented extents to rally supporters and disseminate information, but also to spread rumours, with political and ethnic groups suggesting how their opponents were planning actions to attack, evict and kill individuals and communities. Documents were forged and disseminated online to cast doubts on presidential candidates. In the wake of the violence, Kenya set up the National Cohesion and Integration Commission, which worked with media and law enforcement officials to counter ethnic tensions.

Against this background, a group of researchers and entrepreneurs came together ahead of Kenya's 2013 electoral competition and launched UMATI (which means 'crowd' in Kiswahili), a project seeking to monitor online instances of hate speech. UMATI's overall goal was to detect signals of mounting tensions among Kenyan citizens in order to offer a picture of the different phases of the electoral contest, and to sound the alarm before violence erupted. The elections took place in March 2013, and the project lasted for nine months between September 2012 and May 2013. It tracked blogs, forums, online newspapers, and Facebook and Twitter content generated by Kenyans in English as well as in the major languages spoken in Kenya. Adopting the definition of 'dangerous speech' elaborated by Benesch as a subset of hate speech with the highest potential to catalyse violence, the UMATI team defined practical criteria to distinguish among different speech acts and weigh their potential to catalyse violence. Monitors evaluated questions based on the influence the speaker had on the online community, the statement's content, and the speech's social and historical context. As a result, speech acts could be sorted into three categories: offensive speech, moderately dangerous speech and dangerous speech. Daily monitoring, positioning speech acts along a continuum, and mapping other variables, (including the targets of hate speech and whether speech acts referred to specific events), allowed the researchers to track the evolution of hate speech over time and offer a more nuanced understanding of real and perceived risks.

The findings of the UMATI project, which mapped speech and the cases of violence, or lack thereof, during Kenya's 2013 elections, offer a wider indication of the complexities of linking online speech with actions off-line. In contrast with the previous electoral context, the 2013 elections were largely peaceful. This does not mean that hate speech was less abrasive or widespread. Despite the absence of a baseline that could allow clear comparisons, in 2013 the UMATI project still identified serious, extensive and on-going cases of hate speech and calls to violence. These speech acts, however, did not directly translate into violence on the ground. As the team suggested, other factors besides the presence of hate expressions are likely to have played a more significant part in accounting for the incidence of violent or indeed peaceful outcomes. The numerous calls to peace, coming from different corners of society, including the media, religious groups, and politicians on different sides of the political spectrum, created a climate in which acts of violence were severely condemned.

The UMATI project also offered the opportunity to test how public perceptions of hate speech compared with those used among scholars and in policy circles. As a result of a survey conducted among Kenyans, the project illustrated that the majority of those who participated in the research considered personal insults, propaganda and negative commentary about politicians as hate speech. They similarly held a broader conceptualisation of hate speech than what is stated in Kenya's 2010 Constitution, which in Article 33 prohibits 'propaganda for war, incitement to violence, hate speech or advocacy of hatred that constitutes ethnic incitement, vilification of others or incitement to cause harm'. As Nanjira Sambuli, UMATI's Project Lead explained, the awareness of how hate speech was conceptualized by Kenyans offers an opportunity for discussing not just what hate speech means, but also to place it into a broader discussion on freedom of expression.

Finally, the project offered some indications on how different social networking platforms may enable distinct ways for hate messages to spread and to be counteracted. Only 3% of the total hate speech comments collected by UMATI originated on Twitter, while 90% were found on Facebook. UMATI's final report offers some indications of why this could be the case, pointing to differences in the sites' architectures. Facebook's architecture allows for groups and pages to exist without any activity on them, and for users to engage in different behaviours in different spaces. A user may have a 'clean' timeline on his or her personal profile, while still posting hateful messages on specific groups and pages. On Twitter, on the contrary, all the user's posts are contained in a singular information domain, and can be viewed by everyone that follows the user.

When it comes to challenging hate speech, the project showed how different platforms also allowed different responses with varied effectiveness. In many instances, tweets that were considered unacceptable were shunned and their authors publicly ridiculed. In some cases the ‘offender’ was even forced to retract statements due to the crowd’s feedback, or to close his/her Twitter account altogether. As the iHub report concludes, ‘the singular conversation stream architecture found on Twitter facilitates [this type of response] since all posts are contained on a single timeline and can be viewed by all.’ Similar responses were found to be less likely to occur on Facebook, as the platform’s architecture tends to make conversations more stove-piped and less accessible to broad audiences.

Such monitoring and analysis of egregious online hate speech is potentially a trend that may come to be manifested in many other instances elsewhere.

4.2 MOBILISING CIVIL SOCIETY

Experience in Myanmar provides an example of positive responses from civil society to raise awareness and counteract voices of hatred. After passing a new constitution in 2008 and holding elections in 2010, Myanmar has embarked on a path towards greater openness and social inclusion. The government has led reforms in key sectors, including the media, where new spaces for debate have grown. In 2013, 1.2% of the population had access to the internet and 12% to a mobile phone, up from less than 1% in 2009. The two companies contracted to develop Myanmar’s ICT infrastructure have pledged more than 90% mobile coverage in five years. In this context, some have used social media to spread calls to violence. In 2014, the UN Special Rapporteur on Minority Issues expressed her concern regarding Myanmar over the spread of misinformation; hate speech and incitement to violence; and discrimination and hostility in the media and on the internet. A growing tension online has developed in parallel with cases of actual violence, leaving hundreds dead and thousands displaced, although it would be simplistic to seek direct causal links between online speech and off-line acts.

With the rapid emergence of new online spaces, albeit for a fraction of the population, deeply rooted tensions have appeared in a new form. As Facebook has rapidly become the platform of choice for those citizens making their first steps online, dealing with intolerance and online hate speech is an emerging issue. In this environment, some individuals and groups have championed a more aggressive use of the medium, especially those who feel protected by a sense of righteousness and claims of acting in defence of the national interest. Political figures have also used online media for particular causes. In social media, derogatory terms have been used in reference to minorities. In this complex situation, a variety of actors have begun to mobilize, seeking to offer responses to prevent further violence. Facebook has sought to take a more active role in monitoring the uses of the social network platform in Myanmar, developing partnerships with local organizations and making guidelines on reporting problems accessible in the Burmese language. Myanmar’s Information Minister has pledged to take further steps to fight online hate speech and has expressed interest in developing stronger ties with the USA to find effective measures to contain online hate speech. It is the creative responses from the local civil society that are analysed below.

Local civil society has constituted a strong voice in openly condemning the spread of online hate speech, while at the same time calling for alternatives to censorship. Among the most innovative responses has been *Panzagar*, which in Burmese means 'flower speech', a campaign to openly oppose hate speech. The initiative's goal is to offer a joyful example of how people can interact both online and off-line. Flowers have a powerful meaning in Myanmar, and the campaign encouraged Facebook users to update their profile with a picture of themselves holding a flower in their mouths. The campaign received significant attention both at the national and international level, but, as some activists have recognized, campaigns must take root among those living in the rural areas and among the least educated. Successful coalitions need to be created, and widely respected religious leaders need to be rallied. In addition, besides encouraging 'flower speech', violence must be denounced. Activists express awareness of the need to clarify the limits of what can and what cannot be said, and the role of the state in tackling the problem.

While initiatives such as *Panzagar* have been able to rally different groups, civil society groups do not necessarily share unanimous views on solutions to the problem of hate speech. Some are against laws that would more strictly punish hate speech; some are in favour. In light of the transition, advocates say it is important that the response to hate speech comes from civil society. Local activists have focused on local solutions, rather than trying to mobilize global civil society on these issues. This is in contrast to some other online campaigns that have been able to attract the world's attention towards relatively neglected problems. Initiatives such as those promoted by the *Save Darfur Coalition* for the civil war in Sudan, or the organization *Invisible Children* with the Kony2012 campaign that denounced the atrocities committed by the Lord Resistance Army, are popular examples. As commentaries on these campaigns have pointed out, such global responses may have negative repercussions on the ability for local solutions to be found.

The case of Myanmar is an example of how civil society organizations can proactively mobilize and create local coalitions that are able to address emergent threats. As activists have recognized, the balance between local focus, raising international attention, producing locally relevant results, and avoiding upsetting a delicate transition is fragile. However, their efforts show that a mobilization against online hatred may be an opportunity to help address off-line conflicts that are reflected online.

This experience, as with that outlined in sections 4.2 above, may be emblematic of a potential emerging trend that is broadly replicated in other countries.

4.3 LOBBYING PRIVATE-SECTOR COMPANIES

Various organizations that have combatted hate speech in other forms or have defended the rights of specific groups in the past have found themselves playing an increasingly important role online. This trend is especially evident in developed countries where internet penetration is high, and where private-sector companies are key intermediaries. This section examines campaigns and initiatives in the United States of America, Australia, and the United Kingdom of Great Britain and Northern Ireland, where issues of online hatred have emerged with regard to religion, race and gender. Organizations like the USA-based Anti-Defamation League (ADL) and Women, Action and the Media (WAM!); the Australian-based Online Hate Prevention Institute; the Canada-based The Sentinel Project; and the UK-based Tell MAMA (Measuring Anti-Muslim Attacks) have become increasingly invested in combating online hate speech by putting pressure on internet intermediaries to act more strongly against online hate speech, and by raising awareness among users.

In some cases, organizations have focused on directly lobbying companies by picking up specific, ad-hoc cases and entering into negotiations. This process may involve these organizations promoting their cases through online campaigns, organized barrages of complaints, open letters, online petitions and active calls for supporters' mobilization both online and off-line. However, it is the organizations that largely drive a specific cause. A second type of initiative promoted by some organizations is collecting complaints from users about specific types of content. This activity is particularly interesting when considered in relation to the internet intermediaries' processes of resolving cases of hate speech. While some companies have begun to publish public transparency reports listing requests that governments make for data, information, and content to be disclosed or removed, they have not released information about requests from individual users. When individuals flag content as inappropriate, they may be notified about the processing status of their complaints, but this process remains largely hidden to other users and organizations. This has the result of limiting the possibility of developing a broader understanding of what speech individuals deem to be offensive, inappropriate, insulting, or hateful. Examples of initiatives crowdsourcing requests to take action against specific types of messages include HateBase, promoted by The Sentinel Project and Mobiocracy; Tell MAMA's Islamophobic incidents reporting platform; and the Online Hate Prevention Institute's Fight Against Hate. These initiatives serve as innovative tools for keeping track of hate speech across social networks and how different companies regulate it.

HateBase focuses on mapping hate speech in publically available messages on social networking platforms in order to provide a geographical map of hateful content disseminated online. This allows for both a global overview and a more localized focus on specific language used and popular hate trends and targets. The database also consists of a complementary individual reporting function used to improve the accuracy and scope of analysis by having users verify examples of online hate speech and confirm their hateful nature in a given community. Similarly, Fight Against Hate allows for reporting online hate speech on different social networks in a single platform, which helps users keep track of how many people report the hate content, where they come from, how long it has taken private companies to respond to the reports, and whether the content was effectively moderated. Finally, Tell MAMA offers a similar function of multiple site reporting in one platform, yet focuses solely on anti-Muslim content. This reporting platform also facilitates the documentation of incidents on racial and religious backgrounds for later analysis. The reports received on the platform are processed by the organization, which then contacts victims and helps them deal with the process of reporting incidents to the appropriate law enforcement authorities. The information recorded is also used to detect trends in online and off-line hate speech against Muslims in the UK.

In discussing the importance of generating empirical data, the Online Hate Prevention Institute's CEO, Andre Oboler, stated that platforms like these offer the possibility of making requests visible to other registered users, allowing them to keep track of when the content is first reported, how many people report it, and how long it takes on average to remove it. Through these and other means, these organizations may become part of wider coalitions of actors participating in a debate on the need to balance between freedom of expression and respect for human dignity and equality. This is well illustrated in the example below, in which a Facebook page expressing hatred against Aboriginal Australians was eventually taken down by Facebook even though the page did not openly infringe upon Facebook's terms of service, but because it was found to be insulting by a broad variety of actors, including civil society and pressure groups, regulators, and individual users.

This case illustrates how a large-scale grassroots controversy surrounding online hate speech can reach concerned organizations and government authorities, which then actively engage in the online debate and pressure private companies to resolve an issue related to online hate speech. In 2012, a Facebook page mocking indigenous Australians called 'Aboriginal Memes' caused a local online outcry in the form of an organized flow of content abuse reports, vast media coverage, an online social campaign and an online petition with an open letter demanding that Facebook remove the content. Memes refer in this case to a visual form for conveying short messages through a combination of pictures with inscriptions included in the body of the picture.

The vast online support in the struggle against the 'Aboriginal Meme' Facebook page was notable across social media and news platforms, sparking further interest among foreign news channels. In response to the media commotion, Facebook released an official statement, recognizing that some content may be 'controversial, offensive or even illegal'. In response to Facebook's statement, the Australian Human Rights Commissioner asserted his disapproval of the controversial page and of the fact that Facebook was operating according the First Amendment to the U.S. Constitution on a matter that involved an Australian-based perpetrator and Australian-based victims.

The online petition was established as a further response to Facebook's refusal to remove the content by automatically answering several content abuse reports with a standard statement. The open letter in the petition explained that it viewed the content as offensive due to repeated attacks against a specific group on racist grounds and demanded that Facebook take action by removing the specific pages in question and other similar pages that are aimed against indigenous Australians. Facebook temporarily removed the pages for content review. After talks with the Race Discrimination Commissioner and the Institute, Facebook concluded that the content did not violate its terms of services and allowed the pages to continue under the requirement of including the word 'controversial' in its title to clearly indicate that the page consisted of controversial content.

A second phase came after a Facebook user began targeting online anti-hate activists with personal hate speech attacks due to the 'Aboriginal Memes' case. Facebook responded by tracing and banning the numerous fake users established by the perpetrator, yet allowed him to keep one account operational. Finally, in a third phase, Facebook prevented access to the debated page within Australia following publically expressed concerns by both the Race Discrimination Commissioner and the Australian Communications and Media Authority. However, the banned Facebook page remains operational and accessible outside of Australia, and continues to spread hateful content posted in other pages that are available in Australia. Attempts to restrict specific users from further disseminating the controversial 'Aboriginal Memes' resulted in a 24-hour ban on these actors from using Facebook.

In the following case, the organizations involved took up a long-standing online controversy and went beyond serving as intermediaries for complaints to actively and aggressively lobbying companies, demanding closer content moderation and further, permanent self-regulatory action. In 2013, the group Women, Action and the Media (WAM!) and the Everyday Sexism Project in the UK launched a shared campaign showing advertisements for prominent companies on Facebook pages that disseminated graphic content that was abusive of women. In response to the campaign, both Nissan and the insurance company Nationwide pulled their advertisements from Facebook. Upon seeing their success, the organizers, backed by online supporters and activists, began sending written complaints and photos of advertisements on hateful pages to other major companies like Dove and American Express on their social media platforms, urging them to follow suit. As a result of this campaign, 15 major companies decided to remove their adverts from Facebook.

The campaign also included an open letter written by the two aforementioned groups listing pages that promoted rape and violence against women and demanding that the pages be removed and that Facebook revise its content regulation policy. Alongside the open letter, an online petition on change.org collected over 225,000 signatures and helped raise awareness among online users. Campaign supporters took further action by initiating a large scale protest in front of Facebook's shareholder meeting, publishing the name of all prominent companies using the platform for online advertising, and calling on people to send the companies letters of complaints urging them to withdraw their advertisements from Facebook. Furthermore, activists engaged financial writers in their social media pages, asking them to analyse the potential fiscal damage Facebook could endure due to the growing number of companies pulling out. The online campaign using the hashtag #FBrape resulted in Facebook contacting the organizations concerned in a request for cooperation. The #FBrape campaign gained notable media attention only after it was successful in pressuring the company into an active fight against hateful content targeting women. It had delivered a swift blow in targeting specific companies and their advertising campaigns rather than just Facebook directly.

Facebook's response, however, was not as cooperative initially, as it maintained that the pages listed in the open letter did not violate the company's terms of service. Yet shortly after the campaign began and companies started pulling out, the offensive content was quickly removed. Facebook then released an official statement on its website, stating that it wished to clarify its terms of services and content regulation policies, and also promote cooperation with organizations working to promote freedom of speech while preventing online hate speech from targeting specific groups and individuals. Acknowledging its failure to identify and remove hate speech, the company declared that it intended to review and update its guidelines concerning hate speech moderation, provide its content moderators with better quality training, strengthen its collaboration with concerned organizations to facilitate a shared, responsive effort to better counter hateful content online and also act to hold distributers of such offensive content accountable for such actions.

In a separate, yet related instance, Twitter also took a stand against online harassment of women in collaboration with WAM! by launching a joint pilot project in the form of a reporting platform that would attempt to moderate the content reported within 24 hours. The reports filed by victims of online abuse of women are aimed to serve a dual purpose: allowing WAM! to collect data on offensive content focused on gender-based online harassment in order to explore the phenomenon in depth; and helping Twitter improve its content regulation mechanisms in relation to gender-based online discrimination and abuse. The reporting tool asks women to name the specific users harassing them or the specific tweets they find offensive, classify the type of harassment and answer general questions about how many times they have been harassed, on which platforms and whether the harassment came from one or multiple users. After the report is filed, the claims are investigated by WAM! and then passed to Twitter for further investigation and moderation. The pilot program for the reporting tool operated for three weeks, in which it claimed to have gathered 700 reports and helped over 100 people get faster responses from Twitter. WAM! plans to produce a report on the gathered data, aimed at attaining a better understanding of online hate speech against women.

It appears that the fight against perceived online hate speech is beginning to reach a number of concerned parties, from governments to technology companies and internet service providers, through to a growing number of active organizations and affected individuals. Many online communities and individuals fight against hateful content online on a daily basis alongside more formal organizations. However, this fight necessitates large-scale action in order to ensure that online hate speech can be effectively and contextually identified and remedied in the long run. It requires the empowerment of users to identify and combat hate speech without blocking legitimate speech, and in this way creating more inclusive spaces for expression.

Internet intermediaries, and social networking platforms in particular, have shown a trend to advance their responses to alleged online hate speech through careful interactions with user complaints and by making their regulation process increasingly transparent. Facebook officials have indicated they rely on multiple teams dealing with different types of content in different languages to address reported content as quickly and effectively as possible. Furthermore, Facebook has introduced a reporting dashboard that allows users to keep track of the review process in order to improve their individual communications with each user. Adopting similar mechanisms to contend with hateful speech, Twitter introduced a report button in 2013 following an online petition set up by an individual user.

In summary, there appears to be a trend where internet intermediaries work in increasingly close cooperation with campaigning organizations to provide rapid and effective responses to hate speech on their platforms. At the same time, they also state that they equally weigh complaints by individuals and treat these as seriously as they do petitions and other forms of collective action. To an extent, these companies are also beginning to issue reports to inform users of any changes in their policies and privacy settings, although few provide information about user reports in comparison to transparency reports about official governments requests. The actions of campaigning groups plays an important role, sometimes in conjunction with officials, and particularly where – for various reasons – it is impractical and/or problematic for governments themselves to address the issue.

4.4 COUNTERING ONLINE HATE SPEECH THROUGH MEDIA AND INFORMATION LITERACY (MIL)

While the previous sections have addressed mostly reactive responses to the proliferation of online hate speech, this section offers insights into attempts to provide more structural answers through education. It analyses initiatives targeting citizens, and especially youth, to make actors aware of the issues surrounding and possible responses to perceived online hate speech.

Citizenship education focuses on preparing individuals to be informed and responsible citizens through the study of rights, freedoms, and responsibilities and has been variously employed in both peaceful societies as well as societies emerging from violent conflict. One of its main objectives is raising awareness on the political, social and cultural rights of individuals and groups, including freedom of speech and the responsibilities and social implications that emerge from it. The concern of citizenship education with hate speech is twofold: it encompasses the knowledge and skills to identify hate speech, and enables individuals to counteract messages of hatred. One of its current challenges is adapting its goals and strategies to the digital world, providing not only argumentative but also technological knowledge and skills that a citizen may need to counteract online hate speech. A new concept of digital citizenship is being proposed by some organizations, which incorporates the core objectives of media and information literacy aimed at developing technical and critical skills for on line media consumers and producers and which connects them with broader ethical and civic matters.

Relevant here is global citizenship education (GCED), one of UNESCO's Education Programme's strategic work areas for 2014-2017 and one of the three priorities of the UN Secretary-General's Global Education First Initiative. GCED aims to equip learners of all ages with those values, knowledge and skills that are based on, and instil respect for, human rights, social justice, diversity, gender equality and environmental sustainability. GCED gives learners the competencies and opportunity to realize their rights and obligations to promote a better world and future for all.

Within this wider perspective, UNESCO and many others, working under the umbrella of the Global Alliance for Partnerships on Media and Information Literacy, promotes user empowerment. MIL is an umbrella concept that covers a package of literacies on- and off-line. It includes the development of the technical skills and abilities required to use digital technologies, as well as the knowledge and abilities needed to find, analyse, evaluate and interpret specific media texts, to create media messages, and to recognise their social and political influence. Multiple and complementary literacies are seen as essential for the exercise of rights and responsibilities in regard to communications.

The emergence of new technologies and social media has played an important role in this shift. Individuals have evolved from being only consumers of media messages to producers, creators and curators of information, resulting in new models of participation that interact with traditional ones. Teaching strategies are changing accordingly, from fostering critical reception of media messages to include empowering the creation of media content. There is a strong trend in MIL itself continuing to evolve as a concept, augmented by the dynamics of the internet. It is beginning to embrace issues of identity, ethics and rights in cyberspace (see Paris Declaration on MIL in the Digital Era).

Certain knowledge and skills can be particularly important when identifying and responding to online hate speech. The present section analyses initiatives aimed both at providing information and practical tools for internet users to be active digital citizens. Projects and organizations covered include:

- 'No place for hate' by the Anti-Defamation League (ADL), USA;
- 'In other words' project by the Provincia di Mantova and the European Commission;
- 'Facing online hate' by MediaSmarts, Canada;
- 'No hate speech movement' by the Youth Department of the Council of Europe;
- 'Online hate' by the Online Hate Prevention Institute, Australia.

Even though the initiatives and organizations presented have distinctive characteristics and particular aims, they all emphasise the importance of MIL and of educational strategies as effective means to counteract hate speech. They stress the ability of an educational approach to represent a structural and sustained response to hate speech, considered in comparison to the complexities involved in decisions to ban or censor online content or the time and cost that it may take for legal actions to produce tangible outcomes.

Many argue that the package of competencies within MIL can empower individuals and provide them with the competencies they need to respond to perceived hate speech rapidly as it appears. This can be particularly important given the emphasis that social networking platforms place on individual reporting of cases of abuse, incitement to hatred, or harassment.

Individuals involved in these initiatives tend to recognise the importance of normative and legal frameworks as a reference for their efforts. Most of the initiatives include education about legal instruments and procedures used to prosecute perpetrators of online hate speech, and many encourage a complementary view between legal and educational aspects.

A common denominator of the analysed initiatives is the emphasis on the development of critical thinking skills and the ethically reflective use of social media (based on human rights principles) as starting points for MIL skills to combat online hate speech. The expectation is that these MIL competencies can enhance individuals' ability to identify and question hateful content online, understand some of its assumptions, biases and prejudices, and encourage the elaboration of arguments to confront it. The initiatives discussed here also have an important role in showing that identifying online hate speech is not necessarily as straightforward as it may seem to some.

The initiatives analysed tend to be directed towards a diversity of audiences that are involved and affected by online hate speech. The participant organizations studied here particularly focus their efforts on vulnerable groups and on those prone to either being targets of hate or being recruited by hate groups. Children and youth are one of the main audiences targeted by these initiatives. Parents, teachers and the school community also tend to be considered an important audience due to their role in exposing and protecting children from hateful content. Other groups also targeted include those with the ability to shape the legal and political landscape of online hate speech, including policy makers and NGOs, and those who can have a large impact in the online community exposing hate speech, especially journalists, bloggers and activists. A summary of the different audiences targeted in the analysed initiatives can be found in Table 1.

Table 1. Audiences covered by each educational initiative

	Children	Youth	Teachers	Parents	Policy makers	Bloggers	NGOs	General audience
Anti-defamation league	X	X	X	X	X			
In Other Words					X	X	X	X
No Hate Speech Movement		X				X		X
MediaSmarts	X	X	X	X				
Online Hate Prevention		X				X	X	X

The goals of each project are closely related to the interests and needs of the initiative's intended audience. For instance, MediaSmarts has developed an online video game for children 12 to 14 years old, designed to increase their ability to recognize bias, prejudice and hate propaganda. In the video game, when the children come across varying degrees of prejudice and discrimination in the form of jokes, news or videos, they are asked to identify how such messages can promote hate and then to develop strategies to deal with these messages, either by ignoring or confronting them.

The ADL has focused much of its outreach and educational efforts on teachers and parents, providing them with essential information on how to discuss hate and violence with children, and how to encourage young people to take pertinent action. The No Hate Speech Movement organizes training sessions for bloggers and youth activists in which they can discuss their experiences with online hate speech and share best practices on how to combat it. The sessions aim to promote a grassroots understanding of hate speech and to raise awareness on the impact that bloggers and activists can have in tackling hateful content. The project 'In Other Words' has sought to influence policy makers and civil society to monitor various types of media. It advocates the use of accurate information about minorities and vulnerable groups in media representations, encouraging monitoring to avoid the dissemination of stereotypes, prejudice and other discriminatory discourse.

Despite the particularities of each initiative's content and audiences, they share three broad educational goals: to inform, to analyse, and to confront hate speech. These three aims can be seen in a continuum encompassing progressive goals with specific objectives, each one focusing on different aspects of the problem and providing specific alternatives to respond to hate online. A summary is shown in Table 2.

Table 2. Educational goals and objectives

Information	Analysis	Action
– Raising awareness about hate speech and its consequences – Conveying and disseminating information – Communicating the relevant legal framework	– Identifying and assessing hate speech – Analyzing common causes and underlying assumptions and prejudices – Recognising biased behaviours – Reporting and exposing hate speech	– Responding to hate speech – Writing against hate speech – Changing the discourse of hate speech – Media monitoring

The first educational goal focuses on conveying information on hate speech include raising awareness about online hate speech, its different forms and possible consequences. They also provide information on relevant national, regional and international legal frameworks. Examples of these initiatives can be found in multiple formats, for instance the video 'No Hate Ninja Project - A Story About Cats, Unicorns and Hate Speech' by the No Hate Speech Movement, the interactive e-tutorial 'Facing online hate' by MediaSmarts or the toolbox developed by the project 'In Other Words'.

The second educational goal is more complex and focuses on understanding, through the analysis of online hate speech. This analysis includes assessments and evaluations of the different types of online hate speech, including racism, sexism, and homophobia; and of the multiple forms in which it is presented. An important aspect of the analysis is the critical examination of hate speech in order to identify its common causes and understand its underlying assumptions and prejudices. This analytical process enables individuals to report and expose hateful content online. Examples of projects with this educational goal are the 'No Hate' discussion forum and the 'Reporting hate speech' platform. The discussion forum managed by the No Hate Speech Movement allows young people to debate what counts as hateful content and expose examples of online hate speech that they previously had encountered. The reporting platform designed by the Online Hate Prevention Institute enables individuals to report and monitor online hate speech by exposing what they perceive as hate content; tracking websites, forums and groups; and reviewing hateful materials exposed by other people.

The third educational goal identified in these initiatives focuses on fostering actions to combat and counter hate speech acts. Resources within this educational goal aim to promote concrete actions and responses to online hate speech. The actions proposed vary, depending on the focus of the project and the organization, being more or less combative and confrontational in nature; however, the main focus remains on empowering individuals to respond to and assertively combat hateful content. Examples of these initiatives are training sessions for bloggers, journalists and activists run by the No Hate Speech Movement; the teaching materials and lesson plans developed by MediaSmarts; and the media monitoring policies proposed by the project 'In Other Words'.

Whereas some organizations and initiatives focus on the content of online hate speech, others emphasise its personal aspect by drawing attention to the victims or to the general impact on the community. Regardless of their focus, most projects consider the development of digital skills as an essential aspect for preventing, exposing, and combating online hate speech. The tools and strategies analysed exhibit a variety of approaches to developing such skills, from basic 'how-to guides' to more complex and specialised training. The great array of formats discussed and analysed in the different initiatives make it possible to reach and attract very different audiences.

Exhaustive evaluations of these initiatives, however, are still lacking, and it is difficult to assess whether and to which extent they are successful in combating hate speech or affecting groups that are most likely to engage in online hate speech. For instance, even though MediaSmarts' initiatives and resources have received multiple awards and recognitions, there are no clear indications of who makes the most use of their resources and it is difficult to evaluate the results of their programmes. In the case of the project 'In Other Words', the expected results included the development of material for dissemination, but there is no information on how such material has been used since its publication or what audiences it has reached. Also in the case of the 'No Hate Speech Movement', which has developed different materials and resources (including videos, training manuals, educational tools, and the online platform to report hatred content), there are not clear and public guidelines on how to evaluate or report impact. While most of these initiatives are commendable and potentially offer powerful instruments to combat hate speech at a structural level, more information is needed in order to understand how individuals integrate newly acquired skills in their daily lives and what impact this has for their online activity. This need may be addressed as a possible emerging trend as responses to online hate speech evolve.

4.5 NEWS MEDIA CONTENT MODERATION

Outside of instances of tabloid abuses, reportage of hate speech does not generally amount to advocacy of incitement to discrimination, hostility or violence, but is rather an informative service in the public interest about realities that should be known. However, it is a trend that news media institutions frequently encounter the need to identify and respond to speech contributed by users to their online platforms. A range of systems and practices have been analysed in two studies: a review of legal and institutional dispensations in South East Europe by the Albanian Media Institute, and *Online comment moderation: emerging best practices*, produced by the World Association of Newspapers and News Publishers, which analyses the practices of 104 news organizations from 63 countries. Dealing with dynamic flows of user messages, without restricting legitimate expression, is a challenge for news media that highlights the need for policies as to how each institution defines hate speech as a foundation for what calibrated responses may be called for. This requires a monitoring system by each media house, even if only a minimal mechanism for readers to flag and report incidents for further investigation by the platform's editors. Practices of monitoring and discussion of online hate speech in the news media could be profitably shared with internet intermediary companies, notwithstanding the different standing of these two entities. The Ethical Journalism Network has promoted a five-point plan for newsrooms to identify hate speech and respond accordingly, both in news coverage and in user-comment moderation. A trend (on- and off-line) may emerge of seeking to instrumentalise journalism in the service of combatting hate speech. However, it is widely recognised that one of the antidotes to negative speech is professional reporting standards and credibly informing audiences about facts concerning the presence, status and impact of such expressions in a given society.

5. CONCLUSION AND RECOMMENDATIONS

The emergence and diffusion of online hate speech is an evolving phenomenon. The trends show a combination of measures is emerging to deal with the complexity of a phenomenon that is still poorly understood, and that societies are evolving tailored and coordinated responses from a range of actors. As this trend unfolds, effective solutions will need to be grounded in a better understanding of how different forms of expression emerge, interact, and potentially dissipate online. The emergence of each response discussed in this chapter is linked to unique circumstances, but their analysis and dissemination offers a general palette of methods that various stakeholders can adapt in terms of developing these trends to further effect. A number of broad points can be signalled concerning trends in online hate speech and responses going forward:

5.1 DEFINITION AND UNDERSTANDING

- It is likely that international institutions will continue to avoid providing stringent definitions of hate speech. This caution seems to be shared by important private players that shape online communication. Social networking platforms have avoided proposing strict rules and procedures to identify what type of content should be removed. Some have tried to 'socialise' content moderation, allowing users to resolve controversy through interactions facilitated by the platform. This provides nuance and avoids a mechanistic approach.

- Narrower definitions have been advanced, and their uptake may be adopted more widely by a range of actors, in order precisely to prioritise the most serious online hate speech in an era of massive information flows. These definitions include 'dangerous speech' and 'fear speech'. Such concepts offer tools to identify and describe particular hate speech, possibly signalling critical cases or danger zones where collective responses may be needed to avoid the spread of violence. This is important in responding to the challenge of making connections between online expressions of hatred and actual harm, such as hostility, discrimination or violence. Elements characterizing online communication, including users' perceived anonymity and the immediacy with which a given message may reach wide audiences, make the problem particularly complex. Systematic research is still lacking on the connections between online hate speech and off-line violence, and this need may elicit a research trend in coming years.

- At the same time, a narrow focus may have an underside if it is pursued exclusively. There is a risk that emphasis on the potential of a speech act to lead to violence can lead to a narrow approach that is limited to law and order responses. A concentration only on violence can point to answers that may privilege the state (as the actor that has the legitimate control of the use of violence), to the possible neglect of other actors that could advance different or complementary solutions. However, alternative interpretations of hate speech are focusing on respect for human dignity more widely, and on empowering the targets of speech acts to demand respect and be defended, thereby placing them, rather than the state or another actor, at the centre of effective responses. This approach is not devoid of problems and contradictions, as an excessive emphasis on dignity may lead to a cacophony of relativism or of support for particularistic ideas that are not compliant with human rights. But it suggests, nevertheless, that when addressing online hate speech, different perspectives should be taken into consideration and weighed against one another, both for their ability to explain this phenomenon and its complex link to actual violence, and to offer answers that reflect a more holistic approach.

- Paradoxically, the complexity of defining hate speech also offers opportunities to develop shared local interpretations of the different international standards on hate speech. Hate speech operates as a kind of 'empty signifier'. It is a term that may seem self-explanatory to most, but for which people tend to offer very disparate descriptions when asked. This may constitute a problem, for example when accusations of expressing hateful messages are used instrumentally to discredit legitimate speech or to justify cases of censorship. These are instances when criticism or ridicule of individuals, or opinions or beliefs, becomes labelled as hate speech, going far beyond the parameters spelled out by the ICCPR. The characteristic of the term as an empty signifier, however, may also offer opportunities for different actors to come together and discuss issues that may be difficult to approach otherwise. It may be that a trend to debate the issues raised by online hate speech becomes more widespread given the growing prominence of the phenomenon.

5.2 JURISDICTION

- Much of the attention towards identifying and responding to online hate speech has concentrated on governments. However, there is now a clear trend that internet intermediaries, services that mediate online communication, however, are playing an increasingly important role both in allowing and constraining expression. Many of them, especially search engines and social networking platforms, stretch across countries and regulate users' interactions based on their own definitions of hate speech with unclear relation to international human rights law. They largely rely on users' notifications of content considered inappropriate, and when a case is brought to their attention, the default response is to adjudicate it based on their own terms of service. The conditions under which internet intermediaries operate, however, in terms of how they relate to national and international rules and regulation, pressure groups and individual users, are in constant flux.

- Companies themselves and many civil society actors seem to feel particular unease when private institutions are mandated to act as tribunals and decide what should or should not be offered online. There is on-going debate about the extent to which such tribunals may differ from voluntary self-regulation, in which companies offer their own channels for individual complainants even though the latter retain the right to resolve a particular concern through national courts if unsuccessful. Such legal re-territorialisation of online spaces, however, may lead to a progressive fragmentation of the internet, with states or groups of states imposing their own rules and breaking down the potential of the internet to share expression across frontiers and bring people closer to one another. It creates a scenario in which the internet is experienced very differently in different localities, and where the norm of free flow becomes overshadowed by national or regional exceptionalism. The balance of emphasis between common standards and national differences would shift.

- Most internet intermediaries increasingly operate a use-based approach. Facebook, for example, has activated a 'social reporting' function, which offers users a way to send a message to a person posting information the user does not like but which does not violate Facebook's terms of service. Another option, though far from emerging as a trend (although it is a feature of Facebook), is effectively a 'notice and notice' facility, whereby individuals, via the intermediary, may challenge another party to remove a particular expression. Social networking platforms have sometimes changed or improved the mechanisms through which content is monitored and moderated. This approach has included degrees of cooperation with governments, but in these cases, informality could serve to reduce accountability and transparency both for states and private companies. While informality in some instances responds well to the fluid nature of online hate speech, it has the disadvantage of being ad hoc and piecemeal. In some cases, it may be the particular ability of a pressure group to 'hit the right chord' to make the difference, not the importance or validity of a specific case per se or whether the given case of perceived hate speech really exceeds legitimate expression.

- The trend of intermediary action impacting on hate speech will continue, although increasingly impacted upon by civil society groups (national and transnational) and specific governments.

5.3 COMPREHENSION

- The objectionable nature of hate messages offers apparently strong justifications for limiting them and silencing their authors, such as banning them from a platform or even from use of the internet. These justifications, notwithstanding that they may be disproportionate and fail the 'necessity' test for a limitation to be legitimate, tend to grow stronger in the aftermath of dramatic incidents. At such times, authorities may call for strong measures to contain the internet's potential to spread hate and violence, although the links between online speech and off-line violence may be tenuous. In this context, efforts to identify and understand hate speech not solely with the instrumental goal to counter or eliminate it, but also to grasp what it is the expression of, are particularly difficult. Yet, such efforts are clearly very important, despite trends to over-hasty or overly reactive responses. Research is needed to investigate who the people inhabiting extremist online spaces are, why they say what they say, and how they interpret it, as this may present findings that are often counterintuitive. Such studies are still rare, but a better understanding of the dynamics that may lead to certain types of speech could inform innovative answers not based solely on repressing. For example, are there links between economic inequalities and hate speech? How do some people successfully exploit hate speech for partisan ends, and why do many of their victims tend to come from vulnerable or disadvantaged backgrounds? Are there connections between access to education and hate speech? Answers to questions like these may point towards the need for proactive and practical policies for greater social inclusion, rather than solely to reactive responses to address hate speech understood as a symptom of deeper grievances. This remains a trend that is conspicuously underdeveloped.

- A partial trend that is emerging is that of recognising that online hate speech covers a broad set of phenomena that are conditioned partially by their different platforms. These platforms' architectures vary significantly and have important repercussions on how hate speech spreads and can be countered. A more fine-grained understanding of how each platform can enable or constrain the production and dissemination of different types of messages may thus be an important factor in developing appropriate responses.

- Large social networking platforms have primarily adopted a reactive approach to dealing with hateful messages reported by their users, and analysing whether or not they infringed upon their terms of service. Social networking platforms could, however, take a more proactive approach. They have access to a tremendous amount of data that can be analysed and combined with real life events that would allow more nuanced understanding of the dynamics characterizing online hate speech. Vast amounts of data are already collected and analysed for marketing purposes. Similar efforts could be made as part of the social responsibility mandate of the companies that own these platforms, contributing to the production of knowledge that can be shared with a broad variety of stakeholders. Pressure from external stakeholders could stimulate a trend to more transparency and data sharing.

- Diverse initiatives promoting greater media and information literacy in a range of areas have begun to emerge as a more structural response to online hate speech. Given young people's growing exposure to social media, information about how to identify and react to hate speech is increasingly important. While some schools have expressed interest in progressively incorporating media and information literacy in their curriculum, these initiatives, are still patchy and have often not reached the most vulnerable, who most need to be alerted about the risk of online (and off-line) hate speech and how to counter it. It is particularly important that anti-hate speech modules are incorporated in those countries where the actual risk of widespread violence is highest. There is also a need to include in such programmes modules that reflect on identity, so that young people can recognise attempts to manipulate their emotions in favour of hatred, and be empowered to advance their individual right to be their own masters of who they are and wish to become. Pre-emptive and preventative initiatives like these should also be accompanied by measures to evaluate their impact on students' actual behaviour online and off-line, and on students' ability to identify and respond to hate speech messages. It is of signal importance in countering online hate speech that the uptake of MIL, especially by national education authorities, becomes a prominent trend in coming years.

5.4 SUMMATION

- Defining online hate speech in detail internationally will likely continue to escape a universally observed consensus for some time. A palette of engagements with it is clearly emerging.
- The problem of online hate speech requires collective solutions. As this study has indicated, there are elements specific to the issue of online hate speech that are likely to make responses entrusting only one or a limited number of actors highly ineffective. No single actor and no single response can solve the problem of online hate speech.
- The internet stretches across borders and complex problems such as online hate speech cannot be easily addressed simply by relying upon state power. For example, identifying and prosecuting all individuals posting hateful messages would be impractical for most states.
- As proposed by the UN Special Rapporteur on Minority Issues, states could work collaboratively with organizations and projects that conduct campaigns to combat hate speech, including on the internet, including by providing financial support.
- Internet intermediaries, for their part, have an interest in maintaining a relative independence and a 'clean' image. They have sought to reach this goal by demonstrating their responsiveness to pressures from civil society groups, individuals and governments. These negotiations have so far been ad hoc, however, and they have not led to the development of collective over-arching principles aligned to international human rights law.
- As some of the individuals interviewed for this study have suggested, many users seem to have been numbed by the incidence and presence of online hate speech. More structural initiatives are needed to explain not only how certain instances can be reported, but also why this is important in creating shared spaces where dialogue can occur around hate speech. The silent or passive middle ground could be consolidated to lean away from hateful extremes, through activists' engaging with online hate speech through counter speech.

IV. PROTECTING JOURNALISM SOURCES IN THE DIGITAL AGE[5]

[5] This chapter is drawn from Posetti, J. (forthcoming). Protecting Journalism Sources in the Digital Age. Paris: UNESCO

1. INTRODUCTION

Internationally, source protection laws are increasingly at risk of erosion, restriction and compromise in the digital era. This trend presents a direct challenge to the established universal human rights to freedom of expression and privacy, and their relevance to press freedom and the role of independent journalism. In assessing this trend, it is important to start by unpacking the principles and rationale involved in journalistic source protection.

Journalists rely on ethically and legally enshrined source protection to gather and reveal information in the public interest. In these cases, sources may require confidentiality to protect them from physical, economic or professional reprisals in response to their revelations. The use of confidential sources is not at odds with professional journalistic practice that entails multi-sourcing, verification and corroboration; these are all the more relevant to credibility when such sources are used. However, without such sources, many acts of investigative story telling may never have surfaced. Even reporting that involves gathering opinions in the streets, or a background briefing often relies on trust that a journalist respects confidentiality where this is requested.

All this explains why there is a strong legal tradition of source protection internationally, in recognition of the vital function that confidential sources play in facilitating ‘watchdog’ or ‘accountability’ journalism. It also explains why there is a globally established ethical obligation upon journalists to avoid revealing the identity of their confidential sources. While journalistic professionalism excludes encouragement or condoning of law-breaking, which may take the form or unsanctioned leaking, journalists have a duty to consider the public interest significance of publishing the resulting information. In this process, maintaining confidentiality is a way not to jeopardise the flow of such information that can make an important contribution to combatting corruption and human rights violations.

However, in many cases, the legal situation does not grant recognition of such confidentiality, and journalists can still be legally compelled to identify their sources or face penalties, prosecution and imprisonment. Exceptions to legal protection might include circumstances involving grave threats to human life, when a journalist is accused of committing a crime, or if s/he witnesses a serious crime. Where the legal line is drawn, and how it is interpreted, varies around the world but the principle that sets confidentiality as the norm, and disclosure as the exception, is the generally accepted standard.

The value to society of protecting the confidentiality of sources is widely recognised as greatly offsetting occasional instances of journalists abusing confidentiality such as, for example, inventing sources or failing to verify information before publication. Such abuses invariably come to light, and they are strongly condemned by journalists’ professional organizations which stress the requirement to rely on anonymous sources only when it is necessary to do so to protect the source from exposure, in the course of public interest journalism. Accordingly, free expression standards internationally uphold the confidentiality principle. This principle shields the journalist directly by recognising their professional obligation not to disclose the identity of the source, and it shields the source indirectly through the journalist's commitment. However, this principle works in practice only if the identity of the confidential source cannot be easily discovered by other means and where there are legal limits on the use of this information if anonymity is compromised.

The need to protect the confidentiality of sources is justified in international and regional instruments (see sections 4 and 5 below) largely in terms of ensuring a free flow of information, especially in regard to information derived from whistle-blowers. Without this, a 'chilling effect' is likely, with holders of sensitive information being reluctant to come forward. As another knock-on effect, when media outlets or individuals doing journalism know or suspect that they will be put under pressure to reveal sources, they may become less likely to seek or subsequently use information supplied on condition of confidentiality, with concomitant shrinkage of public interest content as a result.

The expansion of digital means of communication and monitoring, coinciding with increased sensitivity to security issues in many countries, poses particular challenges to traditional legal protections for journalists' sources. A reporter's commitment to refuse attempts to compel them to identify their source/s in the analogue past may have provided significant protection for an anonymous source, but in the age of digital reporting, mass surveillance, mandatory data retention, and disclosure by third party intermediaries, this traditional shielding of identity can be penetrated.

Technological developments and a change in police and intelligence services' operational methods are redefining the character of privacy and of the legal classification the shielding of journalistic sources. Aided by technological advancement, law enforcement and national security agencies have shifted practice from a process of detecting crimes already committed, to one of threat prevention. In the digital age, it is not the act of committing (or suspicion of committing) a crime that may lead to a journalist or a source being subject to surveillance, but the simple act of using mobile technology, email, social networks and the internet. As a result, journalistic communications are increasingly being caught up in the nets of law enforcement and national security agencies. This is in addition to instances when the communications of particular journalists and sources may be expressly selected for targeted surveillance. A 2014 report by the Office of the High Commissioner for Human Rights noted: 'The lack of adequate national legislation and/or enforcement, weak procedural safeguards, and ineffective oversight'. Such gaps have particular relevance to the professional privacy of journalistic work, including journalists' digital communications with their sources.

Parallel to these developments, the past decade has increasingly seen restrictive anti-terrorism and national security legislation, potentially curtailing existing legal protections, including 'shield laws'. These entail moves to broaden the scope of 'classified' information and reduce exceptions that would allow coverage in the public interest, and to criminalise any disclosure (in some cases, also including publication by journalists) of any information classified as secret secrets, without there being provisions for public interest exceptions. These security trends, along with digital tracking, can impact on both journalists and their sources, and constrain, or 'chill', of public interest journalism, especially investigative journalism that relies upon confidential sources. In this complex situation, there are evolutions around the right to privacy in the digital age.

In this digital and security-conscious context, there is debate about which journalistic actors qualify for source protection in the digital era – which raises the need to inclusively define terms like ‘journalism’ and ‘journalists’ in reference to questions like, ‘Who can claim entitlement to source protection laws?’ Another issue is extending shielding laws to all acts of journalism, including digital reporting processes and journalistic communications with sources, not just after the publication of content that is based on these communications.

2. METHODOLOGY

This chapter provides quantitative data and qualitative analysis around the world linked to protection of journalists' sources in the digital age. The research was conducted by WAN-IFRA, the global news publishing association that houses the World Editors Forum (WEF), and a fuller version has been published in parallel to this report.

2.1 STRUCTURING THE RESEARCH

The researchers applied a process of 'datafication' to a 2007 report by David Banisar commissioned by Privacy International called *Silencing Sources: An International Survey of Protections and Threats to Journalists' Sources*. This process involved manually mining and keyword searching the document to a) identify every country mentioned in the report, and to b) establish which countries required additional research to strengthen the available data, thereby enabling the solid benchmarking of the 2007 research. The result was the development of a database that listed each country identified in the 2007 report, along with the different kinds of legal protections applicable globally.

There were 124 territories identified through the 'datafication' of the Privacy International report, but the limitation of the research to UNESCO Member States reduced the number of countries selected for examination to 121. It is this sub-set of countries (see Appendix 3) which constitutes the focus for the research presented here.

2.2 ENVIRONMENTAL SCAN

Once the initial data set was established, each country was assigned to a researcher or research assistant, according to language capacity, for a qualitative mapping exercise, known as an environmental scan. The process of undertaking this scan involved:

a) Preparing a literature review (focused on scholarly books, journals and major reports)

b) Online searches of legal, legislative, and relevant NGO databases in each country

c) Online searches of news websites

d) Contacting WAN-IFRA member organizations and affiliates for input

e) Contacting sources in countries

Data collection began on 1 August 2014 and ended on 20 July 2015.

2.3 ANALYSIS OF COUNTRY DATA

Once each country was studied, a subset of countries was identified where developments had been identified in the period from 2008 up to mid-2015. Ultimately, developments pertaining to legal protections for journalists' sources were recorded in 84 out of the 121 countries (69 per cent) studied.

2.4 SURVEYS

A set of online survey questions was designed to engage members of the journalistic, academic, legal, freedom of expression and online communities globally. Specifically, they were asked to: pinpoint shifts in the legal and regulatory environment pertaining to source protection since 2007; identify key experts/actors for future qualitative interviews; and suggest potential case studies. This survey was launched in October 2014 and continued until January 2015.

The relevant results of an earlier online survey, launched during the World Editors Forum in Turin, Italy in June 2014, were synthesised with the data from the survey distributed in connection with this UNESCO-commissioned study. It asked for evidence of the impact of surveillance revelations by Edward Snowden on newsrooms globally, in terms of changes in training and practice in reference to source protection, along with broader digital safety issues. Further, relevant data from the over-arching UNESCO Internet Study survey was examined in answer to the question: 'To what extent do laws protect digitally interfaced journalism and journalistic sources?'

A total of 134 people from 35 countries - representing every UNESCO region - responded to the combined surveys. The survey data was scanned for evidence of changes to legal source protection frameworks, and digital dimensions. This was used to augment the regional overviews presented below to assist in the identification of expert actors, and in the development of the thematic studies.

2.5 QUALITATIVE INTERVIEWS

Dozens of key actors with legal, journalism, and freedom of expression expertise were identified through the environmental scan and survey processes. Ultimately, 49 interviewees were selected from 22 countries on the basis of relevant expertise, and with the goal of achieving regional and gender balance.

2.6 PANEL DISCUSSIONS

Two panel discussions related to research were held during the study's final phase. The first panel was held in Washington, DC during the World Editors Forum in June 2015. The London Foreign Press Association and the Frontline Club in London jointly hosted the second panel convened in July 2015. The contributions of the panellists during both sessions were leveraged to update and strengthen the study's analysis.

2.7 THEMATIC STUDY

Many potential case studies were identified in the environmental scan and survey processes. Three thematic studies were selected for in-depth analysis to ensure representation of key issues and reflection of regional and linguistic diversity. The third, a ***model assessment tool for international legal source protection frameworks***, is presented here. This thematic study presents the development of an 11-point assessment tool for measuring the effectiveness of legal source protection frameworks in the digital era, drawing on long form qualitative interviews with international experts.

3. KEY FINDINGS AND RECOMMENDATIONS

1. 84 UNESCO Member States out of 121 studied (69 per cent) for this report demonstrated noteworthy developments, mainly with negative impact, concerning journalistic source protection between 2007 and mid-2015
2. The issue of source protection has come to intersect with the issues of mass surveillance, targeted surveillance, data retention, the spill-over effects of anti-terrorism/national security legislation, and the role of third party internet companies known as 'intermediaries'
3. Legal and regulatory protections for journalists' sources are increasingly at risk of erosion, restriction and compromise
4. Without substantial strengthening of legal protections and limitations on surveillance and data retention, investigative journalism that relies on confidential sources will be difficult to sustain in the digital era, and reporting in many other cases will encounter inhibitions on the part of potential sources
5. Transparency and accountability regarding both mass and targeted surveillance, and data retention, are critically important if confidential sources are to be able to continue to confidently make contact with journalists
6. Individual states face a need to introduce or update source protection laws
7. It is recommended to define 'acts of journalism', as distinct from the role of 'journalist', in determining who can benefit from source protection laws
8. To optimise benefits, source protection laws should be strengthened in tandem with legal protections extended to whistle-blowers, who constitute a significant set of confidential journalistic sources,
9. Source protection laws need to cover journalistic processes and communications with confidential sources – including telephone calls, social media, and emails – along with published journalism that depends on confidential sources
10. Journalists are increasingly adapting their practice in an effort to partially shield their sources from exposure, but threats to anonymity and encryption undermine these adaptations.
11. The financial cost of the digital era source protection threat is very significant (in terms of digital security tools, training, and legal advice), as is its impact on the production and scope of investigative journalism based on confidential sources
12. There is a need to educate journalists and civil society actors in digital safety
13. Journalists and others who rely on confidential sources to report in the public interest may need to train their sources in secure methods of contact and information-sharing

4. IDENTIFICATION OF KEY THEMES

The data assembled for this research confirmed the existence of four key overlapping and inter-related trends affecting the legal protection of journalists' sources in the digital age.

The key digital era themes emerging from the research undertaken for this chapter demonstrate patterns that are reflected globally: 1) source protection laws are at risk of being trumped by national security and anti-terrorism legislation that increasingly broadens definitions of 'classified information' and limits exceptions for journalistic acts; 2) the widespread use of mass and targeted surveillance of journalists and their sources undercuts legal source protection frameworks by intercepting journalistic communications pre-publication; 3) expanding requirements for third party intermediaries to mandatorily retain citizens' data for increasingly lengthy periods of time further exposes journalistic communications with confidential sources; 4) debates about digital media actors' entitlement to access source protection laws where they exist, are intensifying around the world. These themes inform the regional catalogue of developments affecting legal source protection frameworks – including legislative changes, judicial precedents, incidents and revelations – that follow.

5. INTERNATIONAL REGULATORY AND NORMATIVE ENVIRONMENTS

Protection of sources in international instruments outlined below is viewed as necessary to ensure the free flow of information, an essential element of several international human rights agreements. The presumption made is that 'exceptional circumstances' are required to justify disclosure of journalists' confidential sources. Accordingly, the need for the information about the source must be judged as essential, and only in cases where there is a 'vital interest' can disclosure be justified.

5.1 UNITED NATIONS ACTORS

5.1.1 Resolutions

2012: The Human Rights Council resolution (A/HRC/RES/21/12) on the safety of journalists passed in September 2012

2012: Resolution adopted by the UN Human Rights Council (UN Doc. A/HRC/RES/20/8) on the promotion, protection and enjoyment of human rights on the Internet that recognise the need to uphold people's rights equally regardless of environment

The first resolution noted 'the need to ensure greater protection for all media professionals and for journalistic sources'. The second affirmed that 'the same rights that people have offline must also be protected online'. This presents important support for extending legal source protection provisions for analogue journalistic processes to the digital realm.

2013: Resolution adopted by the UN General Assembly (A/RES/68/163) on the Safety of Journalists and Issue of Impunity (2013)

This resolution acknowledged that '…journalism is continuously evolving to include inputs from media institutions, private individuals and a range of organizations that seek, receive and impart information and ideas of all kinds, online as well as offline, in the exercise of freedom of opinion and expression, in accordance with Article 19 of the International Covenant on Civil and Political Rights thereby contributing to the shaping of public debate.' It further recognised shifts in definitions of 'journalism' relevant to debates about who is entitled to invoke source protection, and it referred to the value of journalism to the public interest. It noted with appreciation the UN Plan of Action on the Safety of Journalists and Issue of Impunity, which states that efforts to end crimes against journalists should cover not only formally recognised journalists, but also human rights defenders, community media workers and citizen journalists.

In November 2013, the 37th session of the UNESCO General Conference passed a Resolution on 'Internet-related issues: including access to information and knowledge, freedom of expression, privacy and ethical dimensions of the information society'.

This resolution formally recognised the value of investigative journalism to society, and the role of privacy in ensuring that function: '…privacy is essential to protect journalistic sources, which enable a society to benefit from investigative journalism, to strengthen good governance and the rule of law, and that such privacy should not be subject to arbitrary or unlawful interference.'

Responses to a survey attached to the UNESCO study of internet-issues signalled the importance of UN positions on the issue of journalistic source protection. The finalised study, which was informed by preliminary research flowing from this research, proposed (within a package of options) to UNESCO's 195 Member States that they 'recognise the need for enhanced protection of the confidentiality of sources of journalism in the digital age.' The internet study is on the agenda for UNESCO's 2015 General Conference.

In December 2013 the United Nations General Assembly (UNGA) adopted a resolution on the Right to Privacy in the Digital Age. **(A/C.3/68/167)**

This resolution was co-sponsored by 57 Member States and it called upon all States to '…respect and protect the right to privacy including in the context of digital communication. … To take measures to put an end to violations of those rights and to create the conditions to prevent such violations, including by ensuring that relevant national legislation complies with their obligations under international human rights law.' The resolution expressed 'deep concern…at the negative impact that surveillance and/or interception of communications, including extraterritorial surveillance and/or interception of communications, as well as the collection of personal data, in particular when carried out on a mass scale, may have on the exercise and enjoyment of human rights.'

It also called upon States: 'to review their procedures, practices and legislation regarding the surveillance of communications, their interception and the collection of personal data, including mass surveillance, interception and collection, with a view to upholding the right to privacy by ensuring the full and effective implementation of all their obligations under international human rights law' and 'to establish or maintain existing independent, effective domestic oversight mechanisms capable of ensuring transparency, as appropriate, and accountability for State surveillance of communications, their interception and the collection of personal data,' emphasising the need for States to ensure the full and effective implementation of their obligations under international human rights law.

The General Assembly further requested the United Nations High Commissioner for Human Rights to submit a report on 'the protection and promotion of the right to privacy in the context of domestic and extraterritorial surveillance and/or the interception of digital communications and the collection of personal data, including on a mass scale.' The Assembly, in line with the 2012 Human Rights Council resolution (UN Doc. A/HRC/20/L.13), also affirmed that 'the same rights that people have offline must also be protected online, including the right to privacy.' Through its calls to protect the right to privacy, including in the context of digital communications, this UNGA resolution is relevant to source protection. The right to privacy online applies also to journalists, and it can be related to journalists' dealings with confidential sources. Whistle-blowers – a prominent subset of journalists' confidential sources – are more likely to communicate with journalists directly online if journalists can rely on their right to privacy to help shield their professional communications.

2014: Resolution adopted by the UN Human Rights Council (A/HRC/RES/27/5) on the Safety of Journalists

The resolution acknowledged 'the particular vulnerability of journalists to becoming targets of unlawful or arbitrary surveillance and/or interception of communications, in violation of their rights to privacy and to freedom of expression.' This observation has direct application to the issues of source protection and the safety of journalists and their sources.

December 2014: UNGA Resolution on The Safety of Journalists and the Issue of Impunity (A/RES/69/185)

This UNGA resolution made two observations related to the role of journalism in shaping public debate and the 'particular vulnerability of journalists to becoming targets of unlawful or arbitrary surveillance or interception of communications in violation of their rights to privacy and to freedom of expression.'

5.1.2 Reports, recommendations, statements and comments

July 2011: Office of the International Covenant on Civil and Political Rights UN Human Rights Committee, General Comment no. 34

This Comment recognised freedom of opinion and expression as 'the foundation stone for every free and democratic society' that 'form the basis of the full enjoyment of a wide range of other human rights.' A 'free, uncensored and unhindered press and other media' is described as essential for the fulfilment of freedom of opinion and expression. The Comment called for protection of all forms of expression and the means of their dissemination, including electronic and internet-based modes of expression.

2012: Carthage Declaration - participants at the UNESCO World Press Freedom Day conference:

This declaration highlighted the challenges posed by internet communications to the maintenance of freedom of expression and privacy rights essential to the practice of investigative journalism.

June 2013: 'Report of the Special Rapporteur (Frank La Rue) on the Promotion and Protection of the Right to Freedom of Opinion and Expression' to the Human Rights Council

La Rue concluded that: 'States cannot ensure that individuals are able to freely seek and receive information or express themselves without respecting, protecting and promoting their right to privacy.' This statement underscored the relationship between the rights to freedom of expression, and access to information and privacy that underpins source protection.

In July 2013, the then UN High Commissioner for Human Rights, Navi Pillay spotlighted the right to privacy in protecting individuals who reveal human rights implicated information.

Highlighting the case of Edward Snowden, Pillay asserted that national legal systems must ensure avenues for individuals disclosing violations of human rights to express their concern, without fear of reprisals. This is relevant to confidential sources because, although the protection of journalistic confidentiality does not necessarily encompass protection of the source's act of disclosure, fear of reprisal is a factor that affects a source's confidence in a journalist's commitment to keep confidentiality. In this way, an increased fear of reprisal can increase the 'chilling effect'.

Pillay declared that the right to privacy, the right of access to information, and freedom of expression are closely linked. She also explicitly pointed to the need for people 'to be confident that their private communications are not being unduly scrutinised by the State.' The consequence of an absence of such confidence represents a 'chilling effect' on sources that could, in turn, lead to the freezing of the 'information pipe'. This perspective again has a bearing on confidentiality of journalistic sources.

2013 Report (A/HRC/23/40) of the UN Special Rapporteur on Freedom of Opinion and Expression, Frank la Rue:

The report noted that: 'Journalists must be able to rely on the privacy, security and anonymity of their communications. An environment where surveillance is widespread, and unlimited by due process or judicial oversight, cannot sustain the presumption of protection of sources.' La Rue's statement highlights how surveillance can impact on journalism, especially journalism dependent upon confidential sources.

In February 2014, the UN hosted an international expert seminar on the Right to Privacy in the Digital Age (Geneva)

At this event, Rapporteur La Rue called for a special UN mandate for protecting the right to privacy, and added: ‘Privacy and freedom of expression are not only linked, but are also facilitators of citizen participation, the right to free press, exercise of free opinion, and the possibility of gathering individuals, exercising the right to free association, and to be able to criticise public policies.’

July 2014 - Summary of the Human Rights Council panel discussion on the safety of journalists: Report of the OHCHR

The summary noted that: ‘A recurrent issue raised during the discussion was the question of whether the current legal framework was sufficient for ensuring the safety and protection of journalists and media workers. The issue was looked at in terms of both the physical protection against threats and violence and protection against undue interference, including legal or administrative.’ Further, the summary noted that the emergence of new forms of journalism (including social networks and blogs) has led to ‘greater vulnerability of the media, including illegal interference in the personal lives and activities of journalists. Such interference was to be condemned and the independence of the traditional and digital media supported.’

According to the summary, La Rue stated that privacy and anonymity of journalists were also vital elements to ensuring press freedom. Speakers also noted that: ‘bloggers, online journalists and citizen journalists played an important role in the promotion of human rights... [and] stated that the protection of journalists should cover all news providers, both professional and non-professional.’ This is relevant to the issue of the application of legal protection for journalists’ sources. Finally, the meeting heard that national security and anti-terrorism laws should not be used to silence journalists.

These points are relevant to journalists’ right to receive and report information obtained from confidential sources in the public interest, without interference

2014 UNESCO World Trends in Freedom of Expression and Media Development report

The threat of surveillance to journalism is underlined in this global report which highlighted the role of national security, anti-terrorism and anti-extremism laws as instruments ‘used in some cases to limit legitimate debate and to curtail dissenting views in the media, while also underwriting expanded surveillance, which may be seen to violate the right to privacy and to jeopardize freedom of expression.’ The report further noted that, ‘National security agencies across a range of countries have gained access to journalists’ documents, emails and phone records, as well as to massive stores of data that have the potential to enable tracking of journalists, sources and whistle-blowers.’

July 2014: 'The right to privacy in the digital age: Report of the Office of the United Nations High Commissioner for Human Rights'

The UNGA mandated this report on the protection and promotion of the right to privacy in the context of domestic and extraterritorial surveillance and/or the interception of digital communications and the collection of personal data, including on a mass scale. The report found that in the digital era, communications technologies have enhanced the capacity of 'Governments, enterprises and individuals to conduct surveillance, interception and data collection.'

The risks of 'big data' for re-identifying 'anonymous' data were also highlighted in the report. The issue of metadata collection (e.g. data that indicates patterns of behaviour - such as the number of calls between two individuals and the timing of the calls, rather than the content) is also highly relevant to source protection. The chilling effect on confidential sources, given the risk of profiling and exposure posed by the combination of data retention and the implications of big data analysis, is therefore further exacerbated.

The report also stated that 'the onus is on the Government to demonstrate that interference is both necessary and proportionate to the specific risk being addressed. Mass or 'bulk' surveillance programmes may thus be deemed to be arbitrary, even if they serve a legitimate aim and have been adopted on the basis of an accessible legal regime.' It concluded that governments increasingly rely on private sector actors to retain data (often in the context of mandatory data retention legislation that is a common feature of surveillance programs) 'just in case'. It stated that such measures are neither 'necessary', nor 'proportionate'.

Citing a European Court of Human Rights ruling, the report declares the onus should be on the State to ensure that any interference with the right to privacy, family, home or correspondence is authorised by laws that 'are sufficiently precise.' It observes the practice of States sharing their intelligence and bypassing limits on surveilling their own citizens themselves. This has evident implications for journalists, especially foreign correspondents and journalists conducting international investigations.

The role of third party intermediaries was also referenced in this report. This represented an important new dimension relevant to journalists' source protection, as there are increasing pressures on third party intermediaries which may have access to journalists' 'private' digital dealings with confidential sources (such as search engines, ISPs, telcos, and social networks) to hand data over to governments and corporations – in the context of either court proceedings or extra-judicial approaches. This process is increasingly formalised: as telecommunications service provision shifts from the public sector to the private sector, there has been, according to the report, a 'delegation of law enforcement and quasi-judicial responsibilities to internet intermediaries… the enactment of statutory requirements for companies to make their networks 'wiretap-ready' is a particular concern, not least because it creates an environment that facilitates sweeping surveillance measures.' The report noted that 'on every continent, Governments have used both formal legal mechanisms and covert methods to gain access to content, as well as to metadata.'

November 2014: UNESCO International Program for the Development of Communication (IPDC) Council decision

In 2014, the IPDC's 39 Member-State council welcomed the UNESCO Director-General's Report on the Safety of Journalists and the Danger of Impunity, which uses the term 'journalists' to designate the range of 'journalists, media workers and social media producers who generate a significant amount of public-interest journalism'. The Council also reaffirmed the importance of condemnations of 'the killings of journalists, media workers and social media producers who are engaged in journalistic activities and who are killed or targeted in their line of duty'.

May 2015: UN Office of High Commissioner for Human Rights Report on Encryption, Anonymity and the Human Rights Framework by UN Special on the Promotion and Protection of the Right to Freedom of Opinion and Expression, David Kaye

This report from the new Special Rapporteur emphasised the essential roles played by encryption and anonymity. According to Kaye, these defences – working separately or together – create a zone of privacy to protect opinion from outside scrutiny. He notes the particular importance of encryption and anonymity in hostile environments.

Kaye further highlighted the value of anonymity and encryption to journalists, researchers, lawyers and civil society seeking to protect their confidential sources and their communications with them. He observed that individuals trying to 'seek, receive and impart' information and ideas may be forced to rely on anonymity and encryption, particularly in environments where censorship is prevalent. A related issue addressed by Kaye is a trend involving States seeking to combat anonymity tools, such as Tor, proxies and VPNs, by denying access to them. It is evident that such moves may indirect undermine attempts to legally protect confidential journalistic sources in the context of digital communications.

Kaye also recognised that many States recognise the lawfulness of maintaining the anonymity of journalists' sources. However, he reported that: 'States often breach source anonymity in practice, even where it is provided for in law,' highlighting the pressures on journalists that undermine these legal provisions either directly, or progressively. Another issue the Special Rapporteur noted is the increasing prevalence and impact of compulsory SIM card registration on confidential communications, including those between journalists and their sources. He stated that over 50 countries in Africa require, or are in the process of requiring, SIM card registration involving the provision of identifiable data, and that 'such policies directly undermine anonymity…and may provide Governments with the capacity to monitor individuals and journalists well beyond any legitimate government interest.' Kaye concluded that States should support and promote strong encryption and anonymity, and he specifically recommended strengthened legal and legislative provisions to enable secure communications for human rights defenders and journalists.

6. REGIONAL INSTRUMENTS OF HUMAN RIGHTS LAWS AND NORMATIVE FRAMEWORKS

6.1 EUROPEAN INSTITUTIONS

European organizations and law-making bodies are making significant attempts at a regional level to identify and mitigate the risks posed to source protection in the digital environment.

6.1.1 Council of Europe resolutions, declarations, statements, comments, recommendations, reports and guidelines

September 2007: Guidelines of the Committee of Ministers of the Council of Europe on protecting freedom of expression and information in times of crisis adopted

These guidelines recommended that Council of Europe (COE) Member States adopt Recommendation No. R (2000)7 on the 'right of journalists not to disclose their sources of information' into law and practice.

The following principles were appended to Recommendation No. R(2000)7:

- ***Principle 1 (Right of non-disclosure of journalists)***

Domestic law and practice in member States should provide for explicit and clear protection of the right of journalists not to disclose information identifying a source…

- ***Principle 2 (Right of non-disclosure of other persons)***

Other persons who, by their professional relations with journalists, acquire knowledge of information identifying a source through the collection, editorial processing or dissemination of this information, should equally be protected under the principles established herein.

- ***Principle 3 (Limits to the right of non-disclosure)***

a) *The right of journalists not to disclose information identifying a source must not be subject to other restrictions than those mentioned in Article 10, paragraph 2 of the Convention…*

b) The disclosure of information identifying a source should not be deemed necessary unless it can be convincingly established that:

i. reasonable alternative measures to the disclosure do not exist or have been exhausted by the persons or public authorities that seek the disclosure, and

ii. the legitimate interest in the disclosure clearly outweighs the public interest in the non-disclosure, bearing in mind that:

- *an overriding requirement of the need for disclosure is proved,*
- *the circumstances are of a sufficiently vital and serious nature,*
- *the necessity of the disclosure is identified as responding to a pressing social need, and*
- *member States enjoy a certain margin of appreciation in assessing this need, but this margin goes hand in hand with the supervision by the European Court of Human Rights.*

c) The above requirements should be applied at all stages of any proceedings where the right of non-disclosure might be invoked.

- ***Principle 4 (Alternative evidence to journalists' sources)***

In legal proceedings against a journalist on grounds of an alleged infringement of the honour or reputation of a person, authorities should consider, for the purpose of establishing the truth or otherwise of the allegation, all evidence which is available to them under national procedural law and may not require for that purpose the disclosure of information identifying a source by the journalist.

- ***Principle 5 (Conditions concerning disclosures)***

a) The motion or request for initiating any action by competent authorities aimed at the disclosure of information identifying a source should only be introduced by persons or public authorities that have a direct legitimate interest in the disclosure.

b) Journalists should be informed by the competent authorities of their right not to disclose information identifying a source as well as of the limits of this right before a disclosure is requested.

c) Sanctions against journalists for not disclosing information identifying a source should only be imposed by judicial authorities during court proceedings which allow for a hearing of the journalists concerned in accordance with Article 6 of the Convention.

d) Journalists should have the right to have the imposition of a sanction for not disclosing their information identifying a source reviewed by another judicial authority.

e) Where journalists respond to a request or order to disclose information identifying a source, the competent authorities should consider applying measures to limit *the extent of a disclosure, for example by excluding the public from the disclosure with due respect to Article 6 of the Convention, where relevant, and by themselves respecting the confidentiality of such a disclosure.*

- ***Principle 6 (Interception of communication, surveillance and judicial search and seizure)***

a) The following measures should not be applied if their purpose is to circumvent the right of journalists, under the terms of these principles, not to disclose information identifying a source:

i. interception orders or actions concerning communication or correspondence of journalists or their employers,

ii. surveillance orders or actions concerning journalists, their contacts or their employers, or

iii. search or seizure orders or actions concerning the private or business premises, belongings or correspondence of journalists or their employers or personal data related to their professional work.

b) Where information identifying a source has been properly obtained by police or judicial authorities by any of the above actions, although this might not have been the purpose of these actions, measures should be taken to prevent the subsequent use of this information as evidence before courts, unless the disclosure would be justified under Principle 3.

- ***Principle 7 (Protection against self-incrimination)***

The principles established herein shall not in any way limit national laws on the protection against self-incrimination in criminal proceedings, and journalists should, as far as such laws apply, enjoy such protection with regard to the disclosure of information identifying a source.

In regards to the definition of a journalist, the Recommendation stated that the laws should protect ‘any natural or legal person who is regularly or professionally engaged in the collection and dissemination of information to the public via any means of mass communication’. The CoE’s 2007 guidelines that reference Recommendation R(2000)7 further recommended that ‘media professionals should not be required by law-enforcement agencies to hand over information or material…gathered in the context of covering crisis situations.’

2010: Report on the protection of journalists’ sources from the Council of Europe Parliamentary Assembly

The report declared that ‘the protection of journalists’ sources of information is a basic condition for both the full exercise of journalistic work and the right of the public to be informed on matters of public concern.’ Noting that source protection is often violated, it highlighted the need to limit exceptions to legal source protection provisions. It referenced the emergence of threats to journalistic source protection in the digital age. Further, it recommended that 'Member states which have not passed legislation specifying the right of journalists not to disclose their sources of information should pass such legislation' in accordance with the case-law of the European Court of Human Rights and the Committee of Ministers' recommendations.'

2011: Council of Europe Human Rights Commission issues discussion paper on Protection of Journalists from Violence

This Report by the CoE Commissioner for Human Rights directly linked journalistic source protection to journalists' safety. It also referenced a 1996 European Court of Human Rights judgement [*Goodwin v. the United Kingdom* (27 March 1996)] that '[p]rotection of journalistic sources is one of the basic conditions for press freedom.' The Court concluded in that case that, in the absence of 'an overriding requirement in the public interest,' an order to disclose sources would 'violate the guarantee of free expression enshrined in Article 10 of the European Convention on Human Rights (ECHR).' This case led the Council of Europe's Committee of Ministers to adopt Recommendation No. R (2000)7 on the right of journalists not to disclose their sources of information. The CoE reaffirmed the need for protection to ensure that the basic protections of sources were not undercut by security efforts, recalling a declaration (2005) that member states should not undermine protection of sources in the name of fighting terrorism, noting that 'the fight against terrorism does not allow the authorities to circumvent this right by going beyond what is permitted [Article 10 of the ECHR and Recommendation R (2000) 7].'

2011: Council of Europe Parliamentary Assembly adopted Recommendation 1950 on the protection of journalists´ sources.

This Recommendation reaffirmed the centrality of source protection to the democratic function of journalism. It also acknowledged the 'large number of cases' of violations of source protection in Europe and the importance of the protection of sources for investigative journalism. The recommendation required that exceptions to source protection laws be narrowly designed and meet the requirements of Article 10 of the ECHR to prevent widespread demands from authorities for source revelation. It also pointed to the importance of confidential sources within the police and judiciary, and the right of journalists not to disclose them. The problem of data retention in connection with source protection was also referenced in the Recommendation. In addition, the Recommendation made reference to the importance of applying the principles of confidential information sharing to third party intermediaries, which is relevant to the emerging threat of pressure applied to third party intermediaries to hand over data to authorities or litigants, thereby circumventing source protection laws.

It further proposed that the Committee of Ministers call on all their Member States to:

- Legislate for source protection

- Review their national laws on surveillance, anti-terrorism, data retention, and access to telecommunications records
- Co-operate with journalists' and media freedom organizations to produce guidelines for prosecutors and police officers and training materials for judges on the right of journalists not to disclose their sources
- Develop guidelines for public authorities and private service providers concerning the protection of the confidentiality of journalists' sources in the context of the interception or disclosure of computer data and traffic data of computer network

The Recommendation also indicated the need to extend source protections to non-traditional media platforms in line with changes in professional practice, publishing and distribution modes, the role of social media, and participatory audiences and sources. Nevertheless, the Recommendation also took the position that bloggers and social media actors are not journalists and therefore should not be able to claim access to source protection laws. However, the conflation of 'journalism' with 'journalists' could, in effect, exclude a significant number of bloggers who are journalistic actors – such as academic or legal bloggers, as well as activists with human rights organizations who use social media, journalism educators and their students.

The synergies between whistle-blower protections and legal frameworks designed to protect journalists from being compelled to reveal their sources are also recognised in the Recommendation.

2014 Declaration of the Committee of Ministers on the protection of journalism and safety of journalists and other media actors adopted:

This Declaration stated that arbitrary or disproportionate application of laws related to defamation, national security or terrorism 'creates a chilling effect on the exercise of the right to impart information and ideas, and leads to self-censorship.' Furthermore, it declared that 'prompt and free access to information as the general rule and strong protection of journalists' sources are essential for the proper exercise of journalism, in particular in respect of investigative journalism.' The Committee also contended that surveillance of journalists and other media actors 'can endanger the legitimate exercise of freedom of expression if carried out without the necessary safeguards, and it can even threaten the safety of the persons concerned. It can also undermine the protection of journalists' sources.' The Committee also agreed to consider further measures regarding the alignment of laws and practices concerning defamation, anti-terrorism and protection of journalists' sources with the ECHR.

January 2015: Council of Europe Committee on Legal Affairs and Human Rights, Report on Mass Surveillance/Resolution and recommendation

This Report, prepared by Rapporteur Pieter Omtzigt, on the impact of mass surveillance on human rights, addressed the implications for journalistic source protection in the context of freedom of expression and access to information. He pointed to the impact of the 'chilling effect' on journalistic communications with confidential sources and consequent limitations on the revelation of information in the public interest.

January 2015: Council of Europe Resolution and Recommendation on mass surveillance

The Council of Europe Committee on Legal Affairs and Human Rights unanimously adopted a resolution and a recommendation based on the Report discussed above on 26 January 2015. The Resolution stated that the Parliamentary Assembly is 'deeply concerned about mass surveillance practices' disclosed by Edward Snowden, which 'endanger fundamental human rights, including the rights to privacy…freedom of information and expression'. The Assembly also expressed concern over the 'collection of massive amounts of personal data by private businesses and the risk that these data may be accessed and used for unlawful purposes by state or non-state actors' as well as 'the extensive use of secret laws, secret courts and secret interpretations of such laws, which are poorly scrutinized'. The Committee invited the CoE Council of Ministers to consider 'addressing a recommendation to Member States on ensuring the protection of privacy in the digital age and internet safety in the light of the threats posed by the newly disclosed mass surveillance techniques'.

6.1.2 Council of the European Union resolutions, declarations, reports and guidelines

May 2014: Council of the European Union - 'EU Human Rights Guidelines on Freedom of Expression: Online and Offline'

These guidelines advised that: 'States should protect by law the right of journalists not to disclose their sources in order to ensure that journalists can report on matters in the public interest without their sources fearing retribution.' They added that the EU will 'support the adoption of legislation that provides adequate protection for whistle-blowers and support reforms to give legal protection to journalists' right of non-disclosure of sources'.

6.2 THE AMERICAS

In 1997, the Hemisphere Conference on Free Speech held in Mexico City adopted the Chapultepec Declaration. Principle 3 states that, 'No journalist may be forced to reveal his or her sources of information.' Building on the Chapultepec Declaration, in 2000 the Inter-American Commission on Human Rights (IACHR) approved the Declaration of Principles on Freedom of Expression as a guidance document for interpreting Article 13 of the Inter American Convention of Human Rights. Article 8 of the Declaration states that, 'Every social communicator has the right to keep his/her source of information, notes, personal and professional archives confidential.' The application of the term 'social communicator' has resonance with the 'who is a journalist?' debate in reference to shield laws.

In 2013, the IACHR report *Violence Against Journalists and Media Workers: Inter American Standards and National Practices on Prevention, Protection and Prosecution of Perpetrators* by the Office of the Special Rapporteur for Freedom of Expression defined journalists as 'those individuals who observe and describe events, document and analyse events, statements, policies, and any propositions that can affect society, with the purpose of systematizing such information and gathering facts and analyses to inform sectors of society or society as a whole'. It clarifies that this definition includes 'all media workers and support staff, as well as community media workers and so-called "citizen journalists".

6.3 AFRICA

Article 9 of the African Charter of Human Rights gives every person the right to receive information and express and disseminate opinions. The 2002 Declaration of Principles on Freedom of Expression in Africa, released by the African Commission on Human and People's Rights, provides detailed guidelines for member states of the African Union on protection of sources. It stipulates that 'media practitioners shall not be required to reveal confidential sources of information or to disclose other material held for journalistic purposes except in accordance with the following principles':

- *The identity of the source is necessary for the investigation or prosecution of a serious crime, or the defence of a person accused of a criminal offence;*
- *The information or similar information leading to the same result cannot be obtained elsewhere;*
- *The public interest in disclosure outweighs the harm to freedom of expression;*
- *And disclosure has been ordered by a court, after a full hearing.*

6.4 INTER-REGIONAL INSTITUTIONS

6.4.1 Organization for Security and Co-operation in Europe

The OSCE Representative on Freedom of the Media regularly issues statements and comments regarding breaches and threats to legal source protection frameworks. The June 2011 Vilnius Recommendation on Safety of Journalists included a recommendation to 'encourage legislators to increase safe working conditions for journalists by creating legislation that fosters media freedoms, including guarantees of free access to information, protection of confidential sources, and decriminalising journalistic activities'.

6.4.2 The Organisation for Economic Co-operation and Development

March 2014 report: 'The CleanGovBiz Toolkit for Integrity'

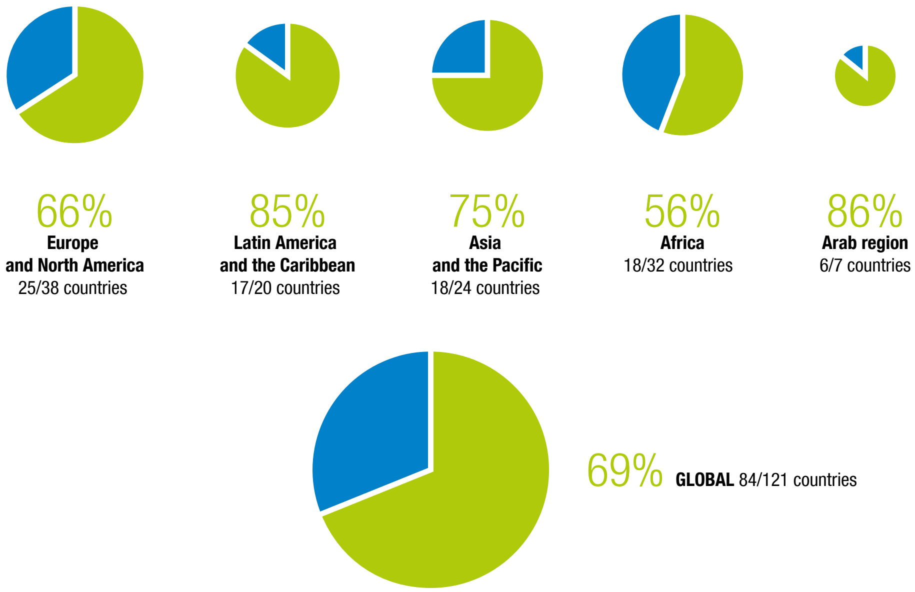
This report asked the questions: 'Are journalists guaranteed to keep their information sources private? If so, how is this ensured?' It acknowledged the importance of source anonymity since 'it can be dangerous for members of the public to provide journalists with information, especially if that information denounces serious misbehaviour or pertains to corruption.' The report stated that forcing a journalist to reveal a source in cases of corruption would be short sighted. The report, which also cited the CoE Committee of Ministers' Recommendation R(2000)7, pointed out the broader risks of unmasking journalists' confidential sources to the ability of people to impart information and the ability of the public to receive information. Further, it stipulated that such protection 'should not only include the journalists' contact persons but also their own workspace and research'. And it argued that: 'Exceptions should only be granted by a judge and only for key witnesses and serious crimes,' highlighting the importance of clearly specifying restrictions, 'so that journalists can reliably inform their potential sources about the risks involved'.

7. OVERVIEWS BY UNESCO REGION

As noted above, developments pertaining to legal and regulatory environments regarding protections for journalists' sources were recorded in 84 out of the 121 countries (69%) studied for this report during the period 2007-2015. Space does not allow for detailed analysis here, but the results, as recorded in the fuller study, illustrate primarily negative or potentially negative impact as regards source protection. These developments were identified and analysed in each of the five UNESCO regions with a particular emphasis on the key identified themes of:

1. The 'trumping effect' of national security/anti-terrorism legislation
2. The role of surveillance (mass and targeted) in undercutting protections
3. The role of third party intermediaries and data retention
4. Changes in entitlement to protection – Who is a journalist? What is journalism?
5. Other digital dimensions (e.g., anonymity)
6. Non-digital dimensions

Percentage of countries with developments in legal and regulatory environments on protections for journalism sources, 2007-2015



7.1 AFRICA

Relevant developments in source protection between 2007-mid 2015 were observed in 18 out of 32 countries (56 per cent) examined in the Africa region**.** However, in 2015, source protection laws in Africa remain limited. Legal developments affecting source confidentiality and its protection in Africa over the past eight years were largely non-digital. In several States, legal source protection frameworks have been threatened by moves to provide broad exclusions to a journalist's right to protect their sources from disclosure on 'national security' grounds and the criminalisation of breaches. Meanwhile, allegations of mass surveillance emerged as a notable theme in some countries. The developments less obviously demonstrate the digital context and the associated risks. This may be because internet penetration in this region is still low. As a result, many of the various matters related to digital newsgathering or online news publication have not yet entered the national debate in many African countries. Multiple governments currently do not see a need to regulate digital media – whether to protect or restrict journalism - in part because relatively few people have meaningful access to it. This trend may change in the future, as more users are able to regularly access online news content.

7.2 ARAB REGION

Developments occurred in six out of seven countries (86 per cent) studied in the region between 2007 and mid-2015. The most notable developments related to mass surveillance and non-digital realms. There was only one noteworthy development found in any of the countries in relation to third party intermediaries. This may correlate with the limited internet penetration, or strict internet controls in parts of the region. While internet engagement among the Arab states remains low relative to certain other regions, the increasing numbers of users means that three countries have introduced laws regulating use of the internet since 2007 with possible implications for source protection. Two of the countries studied demonstrated developments in relation to the question of who is entitled to claim source protection. Four countries of the six reflecting developments demonstrated non-digital shifts in relation to source protection.

It can be noted that the methodology applied to this study excluded a number of Arab States that have undergone dramatic transition since 2007. As a result, it is recommended that further in-depth research be undertaken in all UNESCO Arab States to ascertain the impacts of dramatically changing communications environments on source protection in the region.

7.3 ASIA AND THE PACIFIC

Of the 24 countries analysed in the Asia and Pacific region, 18 (75 per cent) have demonstrated developments in relation to the protection of journalists' sources since 2007. The impact on civil liberties following measures taken to strengthen national security, mass surveillance and data retention; the involvement of third party intermediaries; ambiguous definitions of journalists and bloggers; and a number of other digital and non-digital issues, have weakened source protection. Most notable developments are reflected in the eight countries where issues were documented in relation to national security. Seven countries implemented measures in connection to mass surveillance and data retention in the period examined, and five countries dealt with definitions of journalists and bloggers in reference to access to source protection.

7.4 EUROPE AND NORTH AMERICA

Twenty-five out of 38 (66 per cent) countries examined in Europe and North America experienced significant developments pertaining to source protection laws in the period 2007-2015. These changes reflected the key themes identified associated with emerging digital effects on legal source protection frameworks: a) national security/anti-terrorism impacts; b) surveillance; c) data retention/handover & the role of third party intermediaries; d) questions about entitlement to claim source protection; e) increased risk of source exposure due to digitally stored journalistic communications being seized during investigations.

7.5 LATIN AMERICA AND THE CARIBBEAN

Significant developments that impact on source protection coverage between 2007 and 2015 were identified in 17 of the 20 countries (85 per cent) examined in Latin America and the Caribbean – all of these countries are in Latin America. Surveillance was a clear theme in 10 of the countries studied, five of which introduced new laws that allow data retention and/or interception. Four countries have proposed variations to state secret laws or information classification laws, which in some cases allow for prison sentences for revealing such information. While many countries have laws in place to protect journalists' sources, it is increasingly evident that sources can be identified by other means such as intercepts, threats, raids, accessing stored data, and biometrics. In many of the countries under examination in Latin America, these factors, along with the classification and restriction of information in the name of national security, have rendered many protections for journalists' sources symbolic rather than substantively effective due to the impacts of corruption and organized crime.

However, three Latin American countries introduced new source protection laws.

8. THEMATIC STUDY: TOWARDS AN INTERNATIONAL FRAMEWORK FOR ASSESSING SOURCE PROTECTION DISPENSATIONS

This section maps the development of an 11-point framework for assessing the effectiveness of legal source protection systems in the digital era. It draws on long form qualitative interviews with 31 international experts across all five UNESCO regions spanning the areas of law, human rights, academia, professional journalism, and ICT experts. The interviews were conducted in person, via Skype, telephone and email between November 2014 and February 2015. Based on initial study of the issues, and in consultation with UNESCO, the researchers presented a draft eight-point standard for the experts' consideration. It was then developed and expanded into an 11-point assessment tool, based on the experts' input.

The emergent tool is designed to be applicable to all international settings for assessing the effectiveness of legal source protection frameworks within a State, in the context of established international human rights laws and principles.

Principles for assessing legal source protection frameworks internationally

A robust, comprehensive source protection framework would ideally encompass the need to:

1. Recognise the value to the public interest of source protection, with its legal foundation in the right to freedom of expression (including press freedom), and to privacy. These protections should also be embedded within a country's constitution and/or national law,

2. Recognise that source protection should extend to all acts of journalism and across all platforms, services and mediums (of data storage and publication), and that it includes digital data and meta-data,

3. Recognise that source protection does not entail registration or licensing of practitioners of journalism,

4. Recognise the potential detrimental impact on public interest journalism, and on society, of source-related information being caught up in bulk data recording, tracking, storage and collection,

5. Affirm that state and corporate actors (including third party intermediaries) that capture journalistic digital data must treat it confidentially (acknowledging also the desirability of the storage and use of such data being consistent with the general right to privacy),

6. Shield acts of journalism from targeted surveillance, data retention and handover of material connected to confidential sources,

7. Define exceptions to all the above very narrowly, so as to preserve the principle of source protection as the effective norm and standard,

8. Define exceptions as needing to conform to a provision of 'necessity' and 'proportionality' — in other words, when no alternative to disclosure is possible, when there is greater public interest in disclosure than in protection, and when the terms and extent of disclosure still preserve confidentiality as much as possible,

9. Define a transparent and independent judicial process with appeal potential for authorised exceptions, and ensure that law-enforcement agents and judicial actors are educated about the principles involved,

10. Criminalise arbitrary, unauthorised and wilful violations of confidentiality of sources by third party actors,

11. Recognise that source protection laws can be strengthened by complementary whistle-blower legislation.

Further research could develop a repository of examples of model laws and exemplar judgements that address the issues of 'exceptions' and 'necessity' provisions. A summary of such a repository could be appended to this model assessment framework.

9. GENDER DIMENSIONS

Women journalists face additional risks in the course of their work – on- and off-line. In the physical realm, these risks can include sexual harassment, physical assault and rape. In the digital sphere, acts of harassment and threats of violence are rampant. Similarly, female sources face increased risks when acting as whistle-blowers or confidential informants. These issues manifest in several ways as regards the issue of source protection in the digital era. Unpacked in greater detail below, the issues can be summarised as follows:

1. In comparison to their male counterparts, women journalists face additional risks in dealing with confidential sources
2. Women sources face greater physical risks in encounters with journalists and in revealing confidential information.
3. The physical risks confronted by both women journalists and women sources in the course of confidential communications may require their increasing reliance on digital communications, which raise particular vulnerabilities
4. Secure digital communications defences, including encryption, are arguably even more necessary for women journalists and women sources than for men.

Specific factors for consideration

1. *Female journalists and sources need to be able to communicate digitally*

Female journalists reporting conflict and organized crime are particularly vulnerable to physical attacks, including sexual assault and harassment. In some contexts, their physical mobility may be restricted due to overt threats to their safety or as a result of cultural prohibitions on women's conduct in public, including meeting privately with male sources. Therefore, women journalists often need to be able to rely on secure non-physical means of communication with their sources.

Women sources may face the same physical risks outlined above – especially if their journalistic contact is male and/or they experience cultural restrictions, or they are working in conflict zones. Additionally, female confidential sources who are domestic abuse victims may be physically unable to leave their homes and therefore be reliant on digital communications. These factors present additional challenges for women journalists and sources, in regard to maintaining confidentiality in the digital era.

2. *Digital safety and security are paramount for both female journalists and sources*

Women journalists need to be able to rely on secure digital communications to ensure that they are not at increased risk in conflict zones or when working on dangerous stories, such as those about corruption and crime. The ability to covertly intercept and analyse journalistic communications with sources increases the physical risk to both women journalists and their sources in such contexts. Encrypted communications and other defensive measures are therefore of great importance to ensure that their movements are not tracked and the identity of the source remains confidential.

The risks of exposure for confidential sources are magnified for female whistle-blowers. Therefore, they need to be able to have access to secure digital communications methods to ensure that they are at minimum risk of detection and unmasking. They also need to have confidence in the ability to make secure contact with journalists to ensure that stories affecting women are told, enabling women's participation in public interest journalism. They can also help to avoid magnifying the 'chilling' of investigative journalism dependent upon female confidential sources. Strong legal protections for confidentiality, applied in a gender-sensitive manner, are also needed, especially in regard to judicial orders compelling disclosure.

3. *Online harassment and threats*

Journalists and sources communicating through the internet, including mobile by apps, can face greater risk of gendered harassment and threats of violence. These risks need to be understood and mitigated to avoid further chilling women's involvement in journalism – as practitioners or sources.

10. CONCLUSION

There has been significant change in the realm of legal protections for journalists' sources between 2007 and mid-2015. There has been a partial trend towards preliminary recognition of challenges in terms of international actors, but there is less recognition of the issue at national state level. The developments recorded in the past eight years in 69% (84 countries from 121) of States are generally in directions that run counter to robust source protection in the digital era. The legal frameworks that support protection of journalists' sources are under significant strain in the digital era, with this protection unnecessarily subjected to collateral damage in the face of broader security trends which could result in a loss to societies of the benefits of this particular dispensation.

Impacting on source protection and freedom of expression more broadly, the right to privacy, on which journalists and whistle-blowers partly depend for confidentiality, is directly challenged. In many of the countries studied, frameworks are being undercut by national security, anti-terrorism and data retention legislation that trumps source protection laws, or they risk being weakened by surveillance and mass surveillance. Other threats arise due to pressure being applied to third party intermediaries to release data that risk exposing sources, in response to legal or state-sanctioned demands. There are also increasing challenges to technical measures that support confidentiality, such as limits on anonymity, and moves to outlaw encryption.

Furthermore, there is the question of entitlement to protection: in an era where citizens and other social communicators have the capacity to publish directly to their own audiences, and those sharing information in the public interest are recognised as legitimate journalistic actors by the United Nations, to whom should source protection laws apply? On the one hand, broadening the legal definition of 'journalist' to ensure adequate protection for citizen reporters (working on and off-line) is desirable, and case law is catching up gradually on this issue of redefinition. However, on the other hand, it opens up debates about licensing and registering those who do journalism and who wish to be recognised for protection of their sources. This is why the key tests in contemporary society for access to source protection laws are evolving towards the definition and identification of 'acts of journalism', rather than occupational or professional descriptors.

Journalists and news organizations are in the process of adapting their practices – strengthening digital security and reverting to pre-digital era methods of communication with confidential sources. But unless individual States and regional bodies revise and strengthen their legal source protection frameworks, journalists adapting reporting methods and reverting to analogue 'basics' (an option not always feasible, especially, as argued above, for women who do journalism) will not be enough to preserve source protection in the digital age. In an era of technologically advanced spy-craft, it is also necessary for States to review surveillance practises and oversight in line with UN General Assembly resolutions on privacy. In addition, States need to limit data retention and rendition laws, improve accountability and transparency measures (applied to both states and corporations in regard to journalistic data), and create exemptions for journalistic acts within over-riding national security legislation.

V. FOSTERING FREEDOM ONLINE: THE ROLE OF INTERNET INTERMEDIARIES[6]

[6] This chapter was drawn from MacKinnon, R. et al. 2014. Fostering Freedom Online: The Role of Internet Intermediaries. UNESCO Series on Internet Freedom. Paris: UNESCO / Internet Society. http://unesdoc.unesco.org/images/0023/002311/231162e.pdf

1. INTRODUCTION

As the internet has evolved, an increasingly evident trend is the role played by private sector companies. Among these, Google's search engine, Twitter's social network and Vodafone's telecommunication and internet services are examples of *internet intermediaries* because they *mediate* online communication and enable various forms of online expression. Intermediaries can also act as chokepoints, arbiters, defenders, or selective 'gatekeepers' of expression. Yet intermediaries' power can only be fully understood in the context of state power. The position of internet intermediaries in relation to states and to international human rights standards is complicated: they often operate across a variety of jurisdictions, and states expect them to comply with national laws that in turn align in varying degrees with international human rights norms. Some views have regarded these companies as a source of 'liberation technology' that will help unshackle the hands of the oppressed. Others have critiqued them for failing to do enough to protect user privacy rights and facilitating unaccountable surveillance by the private sector as well as governments. Intermediaries are showing a trend of increasing awareness that they have a powerful and positive role to play in fostering rights. However, in order to protect freedom of expression and privacy and not engage in violation of rights, they need to more closely follow international standards of transparency, necessity, proportionality, legitimate purpose, and due process.

This chapter examines recent trends in intermediaries' policies and practices towards users' freedom of expression and privacy, drawing from the 2014 UNESCO study *Fostering Freedom Online: The Role of Internet Intermediaries*. That publication helped inform UNESCO's comprehensive study on internet-related issues, mandated by Member States in Resolution 61 of UNESCO's 37th General Conference in 2013, which was published in 2015 as *Keystones to foster inclusive Knowledge Societies: Access to information and knowledge, Freedom of Expression, Privacy, and Ethics on a Global Internet*.

1.1 BUSINESS AND HUMAN RIGHTS

International human rights law has traditionally focused on state conduct, founded in declarations and agreements concluded between states. Over the past several decades, however, there has been growing recognition that businesses also have human rights responsibilities for which they should be held accountable. Because most internet intermediaries are operated by private sector companies, this chapter builds on established human rights standards for business and human rights laid out by the UN's 'protect, respect and remedy' framework. It assesses that while governments have the primary duty to protect human rights, companies also have a responsibility to respect human rights; both entities must ensure access to effective remedy.

This perspective has been elaborated in trends over the past five years. In 2011 the UN Human Rights Council (UNHRC) endorsed the UN Guiding Principles on Business and Human Rights, the result of six years of research and consultation with companies, governments and civil society by the UN Special Representative of the Secretary-General on Business and Human Rights. The Guiding Principles begin with the duty of states to protect against human rights abuses by businesses operating within their territory, and to 'set out clearly the expectation that all business enterprises domiciled in their territory and/or jurisdiction respect human rights throughout their operation'. These principles apply universally to all companies – not just internet intermediaries. They also apply universally. UN High Commissioner for Human Rights Navi Pillay wrote in her June 2014 report to the General Assembly, 'the responsibility to respect human rights applies throughout a company's global operations regardless of where its users are located, and exists independently of whether the State meets its own human rights obligations.'

This chapter identifies trends in what internet intermediaries have done and can do further to maximise freedom of expression across a range of jurisdictions, contexts, technologies and business models. To track and understand this, however, it is first necessary to elaborate on the nature of intermediaries and their relation to free expression.

1.2 INTERMEDIARIES

An intermediary, as defined by legal scholar Thomas F. Cotter, is 'any entity that enables the communication of information from one party to another'. In a 2010 report, the OECD explains that internet intermediaries 'bring together or facilitate transactions between third parties on the internet. They give access to, host, transmit and index content, products and services originated by third parties on the internet or provide internet-based services to third parties.' Most definitions of intermediaries explicitly exclude content producers, as does this chapter. More explicitly, the OECD excludes from the intermediary's function 'activities where service providers give access to, host, transmit or index content or services that they themselves originate'. In this view, publishers and other media that create and disseminate original content are *not* intermediaries. Examples of such entities include news websites that publish articles written and edited by staff or invited contributors, or digital-video subscription services that hire or invite people to produce videos and disseminate them to subscribers.

At the same time, many entities offer hybrid services and constitute intermediaries to one extent or another. To what extent social media services, for instance, are primarily intermediaries or operate a media function, is important in terms of expectations. In 2011, the Council of Europe adopted a broad definition of media using six criteria to assess when new actors count as media. These criteria include intent to act as media, exercise of editorial control, and application of professional standards. Some stakeholders, however, have raised concerns that efforts by some states to define intermediaries as 'media' have resulted in stronger restriction on freedom of expression. While there are some potential similarities between media and intermediaries in some instances, there are also significant differences in evolution. While media is generally liable for its content in legal terms, because of its editorial control, intermediaries usually have limited legal liability inasmuch as the content carried emanates from actors independent of their control (see 2.2 below).

All commercially-operated internet intermediaries studied in this chapter require users to agree to 'terms of service' before they are allowed to use the service. Sometimes such terms may restrict users' speech that is actually protected by the law in some jurisdictions. While the enforcement of such terms may sometimes resemble an editorial function, the legal basis for terms of service enforcement in the US and Europe, where internet intermediaries first emerged, is derived not from media law but from contract and commercial law.

1.2.1 Types of intermediaries

This chapter focuses on services and platforms that primarily host, give access to, index, or facilitate the transmission and sharing of content created by others. As intermediaries' importance has grown for the global knowledge economy, a number of organizations have sought to describe or categorise intermediary types by their roles and technical function. These include the OECD, the UN Special Rapporteur on freedom of opinion and expression, and civil society organizations. The table below provides a comparison of the key intermediary types that these organizations have characterized or singled out for examination.

Table 1: Categories and key examples of internet intermediaries

OECD	Special Rapporteur La Rue	ARTICLE 19	Center for Democracy and Technology	Global Partners
Internet access and service providers	Internet service providers (ISPs)	Internet service providers (ISPs)	Access providers/ ISPs Network operators and mobile telecommunications providers	Physical layer: makes communications possible
				Connectivity & code: the language or protocols of the communication
Data processing and web hosting providers		Web hosting providers	Domain registrars and registries Website hosting companies	Applications: tools to navigate content
Internet search engines and portals	Search engines	Search engines	Internet search engines and portals	
E-commerce intermediaries			E-commerce platforms and online marketplaces	
Internet payment systems				
Participative networking platforms	Blogging services Online communities Social media platforms	Social media platforms	Online service providers In general, any website that hosts user-generated content or allows user-to-user communications	

From these efforts, it is evident that different types of intermediaries perform different functions and have different technical architectures. For example, internet service providers (ISPs) connect a user's device to the internet, and then web hosting providers and domain registrars and registries make it possible for websites to be published and to be accessed online. Search engines make a portion of the World Wide Web accessible by allowing individuals to search their database and are often an essential go-between between websites and internet users. Social networks allow individual internet users to exchange text, photos and videos, as well as allowing them to post content to their network of contacts or to the public at large.

It is also apparent that different intermediary types further entail different kinds of business models. In order to provide internet access and/or telecommunications, companies must operate equipment and services within the geographical jurisdictions where customers physically reside. This type of service requires substantial investment of resources, equipment, and personnel in physical jurisdictions, requiring state permission and compliance with local law. Thus states maintain a high degree of leverage over ISPs.

The same actors do not necessarily provide telecommunications and internet access. Much internet service provision mainly rides on the technical transmission infrastructure of telecommunication, which may serve as an underlying lever to exclude or limit access to certain ISPs or to their customers' users. In turn, the ISPs may limit access at a second level independently of their relationship with the telecommunications infrastructure operators. The reliance of ISPs on telecommunications makes the network level of intermediaries particularly susceptible to regulation by states.

By contrast, other intermediary types such as web hosting providers, domain name registrars and registries, search engines, and social networks do not necessarily need to locate staff, equipment or other physical resources in the same geographical area as the users they aim to serve. The open, interoperable architecture of the internet makes it possible for a user in a given country to conduct a search on Google, set up a website with a web hosting service, or communicate with friends on Facebook without those companies having staff, offices or equipment in that country. This has the potential to distance web-based intermediaries – and their users – from control by states in which they are not headquartered or otherwise have a physical presence.

This relative independence is precisely why scholars have documented new media, particularly social media, as enhancing freedom of expression in contexts where off-line expression is subject to strong restriction by the state. In practice, however, a growing number of states assert jurisdiction over intermediaries by exercising control over the underlying tier of telecommunication providers and ISPs, which serve as chokepoints for web access. States can, and increasingly do, threaten to deny access to all users under their jurisdiction to a particular service if remotely-based intermediaries fail to comply with their laws. By targeting intermediaries at different levels, states may exercise control over users' online expression or access to information even when this is conducted outside the national jurisdiction. They may also delegate controls to intermediaries, without directly policing individuals themselves.

1.2.2 Modes of restriction

Depending on the type of intermediary and the service offered, intermediaries control how and with whom their users can communicate. They have access to information created by users as well as a range of information directly related to users. For this reason, intermediaries are key in facilitating and protecting the rights to free expression and privacy. They also serve as avenues through which governments can monitor, regulate and control individuals' online activities and access to information. The two primary ways in which freedom of expression can be restricted via ISPs, search engines, and social media can be broadly described as follows:

1. **At the network-level**, telecommunications access providers and ISPs can be used to restrict freedom of expression in three main ways:

   a) ***Filtering***: Access is blocked to entire websites, specific pages or specific keywords. Filtering is carried out either by the ISP or by the network operators that control internet flows into a jurisdiction or some combination of the two. The content still exists elsewhere on the internet, but cannot be accessed by users of the network on which the filter is deployed. Such blocking prevents users from receiving information, but can also prevent users from posting information to a specific location such as a social network.

b) ***Service shutdown***: One or more services offered by one provider or all providers can be shut down in a given jurisdiction or geographic area, preventing users in the area from accessing the internet via fixed line or mobile, sending SMS messages, etc.

c) ***Non-neutral service***: Access to certain content or applications is 'throttled' or slowed down, making it more difficult for users to access. Alternatively, users might be charged different rates for access to different kinds of content or services, or might be granted free access to specific services.

The two other intermediary types covered by this chapter, search engines and social networks, are directly affected if these restrictions are carried out at the network level. At the same time, filtering or the threat of filtering at the network level is a means by which pressure can be placed on search engines, social networks and other intermediaries to carry out restrictions at the platform level.

2. Intermediaries that operate **at the platform level** such as search engines and social networks can act to remove content completely, block it from view to particular categories of users, or deactivate user accounts. These actions are carried out by the company itself or by government authorities that have been granted direct technical access to the platform's core functions. The removal, blocking, or deactivation may take place at the request of a government, users or other third parties, or according to the intermediary's own private rules.

   The restrictions described above are an enforcement tool for different kinds of public and private governance. They are used to enforce state regulation / law or to help identify violations of state regulation, and to enforce companies' private terms of service and other rules. They are also used in some countries to enforce standards issued by private or quasi-governmental bodies.

   Freedom of expression can also be impacted by intermediaries' actions that are privacy-related (at both network and platform levels). Internet users who believe that their communications and online behaviour are being monitored or exposed in a manner that violates their privacy rights are less likely to express themselves freely while using the services of intermediaries. Privacy can be negatively affected via all tiers of intermediaries as follows:

a) **Data collection and monitoring** takes place at all layers of the internet and has the ability to restrict expression through encouraging self-censorship.

b) **Lack of security in how user data is stored or how content data is transmitted** can result in breaches of privacy, unauthorized interception, or interception by government authorities, without the active involvement of the company.

c) Different services and platforms provide **internet users with varying levels of control over their personal information** and if and how it is preserved or publicly accessible.

The following table provides a summary of the modes of restriction described above.

Table 2: Modes by which expression and privacy may be restricted via internet intermediaries either on request or on company initiative

	ISPs	Search Engines	Social Media
Network-level Restrictions	• Filtering • Service shutdown • Non-neutral service		
Platform-level restrictions		• Manipulation of search ranking • Removal or 'de-listing' of links to specific web pages or categories of web pages	• Removal of content from the platform • Blocking of content, and free expression opportunities, by restricting access of particular categories of users (including geographical location) • Account limitation or deactivation
Privacy-related chilling effects	• Collection and retention of user data for commercial or government mandated purposes • 'Real name' account registration requirements • Government requests for user data • Real-time government surveillance	• Collection and retention of user data for commercial purposes • Government requests for user data • Catalogue of individuals' personal individual via searches on their name	• Collection and retention of user data for commercial purposes • 'Real-name' identity requirements • Government requests for user data

The role that intermediaries have been playing in protecting or restricting freedom of expression is further complicated by the global nature of many companies. Multinational companies, as well as internet services with users in multiple jurisdictions, can be subject to a global patchwork of legal and regulatory regimes. Some internet companies have sought to address this dilemma by creating country specific filters and by developing company policy on handling government requests for content restriction as well as user data requests. When a company does not have any physical offices or personnel in a particular jurisdiction, it is difficult for a government to compel that company to abide by its laws or respond to its requests for content restriction. In response, some governments have resorted to filtering – or threatening to filter – content or entire services. In all this complexity, freedom of expression standards are often inadequately understood, protected, respected and remedied.

1.2.3 Commitments to freedom of expression

In light of this increasingly complex global landscape, a number of efforts have appeared as an emerging trend in recent years at the industry and governmental level to help internet intermediaries maximise respect for users' privacy and freedom of expression. For example, in 2013 the European Commission launched a 'sector guide' on how ICT companies can implement the UN Guiding Principles on Business and Human Rights, which were developed in consultation with industry, academia, civil society and governments. Some intermediaries have begun to make public commitments to respect users' rights. Since its launch in 2008, several internet companies have joined the Global Network Initiative (GNI), a multistakeholder body in which major intermediaries work together with participants from civil society, responsible investment, and academia to implement a set of core principles on freedom of expression and privacy. Of the intermediaries studied in this report, Google is a founding member of the GNI. In January 2014, Google passed an assessment process verifying that the company had satisfactorily implemented the GNI principles in handling government requests for content restriction and user data. Facebook joined the GNI in May 2013, but had not by September 2015 undergone an assessment to verify whether it has implemented the GNI principles. In 2012, a group of telecommunications companies, including Vodafone, formed the Industry Dialogue on Freedom of Expression and Privacy in an effort to develop principles and best practices.

In this context, an emerging trend can be observed in the growing number of internet and telecommunications companies that have begun publishing regular 'transparency reports', thus named for the light they shed on the volume and nature of requests to remove content – whether by government or private entities – or to disclose user data. Such transparency helps users and the public at large to understand what kinds of restrictions are being undertaken, and on whose behalf they are carried out. Among companies studied in this chapter, Facebook, Google, Twitter and Vodafone have published transparency reports. It is important to note, however, that significant variations in their scope, detail and reporting methodology make it difficult to draw meaningful conclusions about one company's respect for free expression and privacy in comparison to another. Scholars have called on companies to work with academics and advocates to establish more standardised approaches to transparency reporting. They have proposed that full transparency involves reporting more than just numbers of government requests received and complied with, but also transparency about companies' policies and practices for handling government requests as well as private enforcement mechanisms.

1.3 METHODOLOGY

This chapter reviews the case studies produced in *Fostering Freedom Online: The Role of Internet Intermediaries*, examining three intermediary types and 11 companies:

1. **ISPs & telecommunication services:** Vodafone, Vivo/Telefônica Brasil, Bharti Airtel, Safaricom
2. **Search engines:** Google, Baidu, Yandex
3. **Social networks:** Facebook, Twitter, Weibo, iWiW

The case studies include a description and analysis of the evolving legal and regulatory contexts in which internet intermediaries operate, as well as trends in company policies and practice. As UNESCO's two global priorities, Africa and gender equality each receive special sections. The chapter ends with general recommendations for all stakeholders.

Selection of the three different intermediary types was informed by the OECD's five-part classification of internet intermediaries, plus the three intermediary types singled out as examples in former UN Special Rapporteur Frank La Rue's 2011 report on the right to freedom of opinion and expression on the internet. Companies and countries of focus for each case were selected because they collectively represent a range of cultural, regional, political and legal environments from which powerful internet intermediaries have arisen.

For every country covered by the research, an in-country research team was commissioned to complete a detailed research questionnaire designed in early 2014. The questionnaires contained an average of 61 questions about the legal and political context affecting internet regulation, the policies and practices of the selected companies in the selected countries, and how the combination of particular company policies and legal contexts affect internet users, with several specific questions related to gender. Research for the questionnaires was carried out in March and April 2014, during which the researchers conducted interviews with representatives from industry, government, civil society, academia and law. To answer questions about user perspectives in each country covered by the case studies, researchers canvassed available academic research, media reports and relevant user forums. The results of these questionnaires were then analysed and distilled by the study's authors, who worked with the researchers to clarify and update the research through July 2014.

2. LAW AND REGULATION

Just as online platforms and services can be used for legitimate purposes including self-expression, education, employment and trade, they can also be used for illegitimate purposes such as theft, fraud, harassment, copyright infringement and defamatory speech. The line separating legitimate and illegitimate purpose is significantly influenced by political, religious and cultural context, resulting in multiple understandings of legitimate and illegitimate purposes throughout the world. Recognising this tension, particularly in the context of speech, the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR) and other international human rights instruments allow for certain limitations on the right to freedom of expression, while protecting the essence of the right. As former UN Special Rapporteur La Rue underscored in his 2011 report, restrictions are only compatible with international human rights standards when such restrictions are:

- Rule-based, provided by law and carried out in a transparent and predictable manner;
- Necessary and proportionate, using the least restrictive means to achieve the objective;
- Consistent with purposes cited in the ICCPR: necessary to protect the rights or reputation of others, national security or public order, public health or public morals.

In 2011, the Human Rights Committee in its General Comment No. 34 determined that limitations intended to protect 'public morals' 'must be understood in the light of universality of human rights and the principle of non-discrimination'. Restrictions applied by intermediaries should be evaluated in terms of these international standards.

Although the norm is limited liability and self-regulation, there are exceptions where intermediaries are held liable for user content that others perceive as violating privacy, defamation or other laws. A ruling in 2015 confirmed such liability in the European Court of Human Rights case of *Delfi* v. *Estonia* concerning defamatory speech, and suggested that a news portal must be aware of its content at all times, and not just remove content when the material is brought to its attention. Dissenting judges in this case said this stance was not very different from prior restraint.

Where such liability for intermediaries is the case, companies come under pressure to conduct their own monitoring and filtering to avoid possible repercussions. This contributes to a process of pre-publication controls in which some governments may come to rely on private sector companies to regulate online content without public accountability or due process. On the other hand, genuine self-regulation which references international standards on human rights may sometimes serve to protect freedom of expression and respect for the normative limitations on restriction as per the UDHR and ICCPR.

2.1 STATE COMMITMENTS AND LIMITATIONS ON EXPRESSION

While technology, business models and the scope of business carried out by internet intermediaries have evolved dramatically over the past two decades, the types of regulatory goals pursued by states remain largely unchanged, even if the methods used to pursue those goals have evolved. In many instances, there is debate about the alignment of states' regulations to ICCPR standards and the implementations of these standards. While the types of limitations should be aligned with legitimate purpose, they often fall short in safeguards of necessity, proportionality and due legal process for implementation. The predominant trend is that overreach in limitations relate to criminal defamation, national and public security, hate speech, elections, child protection, blasphemy and intellectual property.

Where limitations are legitimate, there are still numerous complexities. While there is an established right to privacy, it is not elaborately defined, particularly in the digital age. Hence limits on expression to protect privacy may give insufficient attention to the need for public interest exceptions that give priority to the public right to know. On the other hand, the 2012 UNESCO report *Global Survey of Internet Privacy* concluded that privacy laws that provide only weak protection can have a negative impact on freedom of expression.

Another example of complex balancing is seen in the case of actors seeking to curtail freedom of expression, for the reason of protecting the reputation of some citizens in turn linked to a certain notion of privacy. This is of contemporary significance in the European Union given the case of *Google Spain v AEPD,* described in shorthand as establishing a 'right to be forgotten', based upon 'a right to be delisted' on search engines, throughout the countries in this zone. As discussed later in this chapter, the decision by the European Court of Justice in May 2014 demonstrates that an individual's desire to eliminate negative information about him- or herself from the internet can conflict with the right of others to receive and impart information. Critics like Harvard professor Jonathan Zittrain have proposed a right to reply as a preferable alternative to balancing reputation and freedom of expression.

A further case of complex balancing is how the right to life, liberty and security of person can be secured while also preserving the essence of the rights to privacy and freedom of expression. This is at the heart of the debates about digital surveillance. In 2014, the UN High Commissioner for Human Rights called for reform of surveillance laws, and referred to the recommendations by global civil society for the application of the 'necessary and proportionate' principles with strong accountability, transparency and remedy. However, a survey of experts in 18 countries in 2014 showed that little surveillance reform has taken place. In many countries, new laws have continued to expand government surveillance powers. Surveillance has been documented to have a chilling impact on freedom of expression in a range of jurisdictions.

2.2 INTERMEDIARY LIABILITY

As signalled earlier, a major issue for the role of intermediaries in fostering freedom of expression is their legal liability: this relates to the question of what happens when an individual uses an intermediary service to post, share or access content that infringes laws in a given country. The kernel is the extent to which intermediaries can or should be held legally responsible – or 'liable' – for the activities of their users. Intermediary liability provisions formalize government expectations for how an intermediary must handle 'third-party' content or communications. In many jurisdictions, such legal provisions define circumstances under which intermediaries can benefit from limited liability, by setting forth criteria that intermediaries must follow in order to escape a civil or even sometimes criminal penalty for users' actions.

2.2.1 Models of intermediary liability

Many governments in regions including Europe, North America, parts of South East Asia and Latin America have laws specifically addressing intermediary liability. In other regions, particularly Africa, governments are now considering legal provisions on intermediary liability. Broadly speaking, where such regimes exist, there are three models of intermediary liability: strict liability, conditional liability and broad immunity. Exact requirements and nuances of these models vary from jurisdiction to jurisdiction, and are defined by governments and further clarified by courts. Some intermediaries explicitly comply with legal mandates relating to intermediary liability by undertaking measures such as self-regulation via enforcement of their terms of service.

- **'Blanket' or strict liability**: The intermediary is liable for third-party content even when it is not aware that the content is illegal (or even exists). The only way to avoid liability under such circumstances is to monitor, filter and remove content proactively and before publication if it is likely to be infringing. Even so, monitoring and removing content does not absolve the intermediary of liability if any infringing content is overlooked. Blanket liability regimes do not distinguish between intermediaries; all intermediaries, regardless of size or function, are liable.

- **Conditional liability:** The intermediary is potentially exempt from liability for third-party content if certain conditions are met – such as removing content upon receiving notice ('notice and takedown'), notifying the content creator of infringing material after receiving notice ('notice and notice'), or disconnecting repeat infringers upon notice. If an intermediary does not meet these stipulations, it may be liable for damages. Unlike the 'strict liability' model, this limited liability model does not compel intermediaries to proactively monitor and filter content in order to avoid liability. The 'notice-and-takedown' variety of conditional liability has been criticised as easy to abuse; furthermore, it is said to facilitate self-censorship by placing the intermediary in a quasi-judicial position responsible for evaluating the legality of content. The model is even more susceptible to abuse when it lacks elements of due process, such as the opportunity to appeal a takedown. Indeed, 'notice-and-takedown' incentivises intermediaries to remove content immediately after receiving notice, rather than investing resources to investigate the validity of the request and risk a lawsuit. Legitimate content can end up being censored as a consequence.

- **Broad immunity:** In this model the intermediary is exempt from liability for a range of third-party content without distinguishing between intermediary function and content type.

Given the key role that intermediaries and the laws that govern them play in online freedom of expression, a trend in discussions at the international level has been the pursuit of establishing common principles and best practices. For example, in December 2011, the OECD Council included 'limiting intermediary liability' as one of 14 recommended Principles for Internet Policy Making to 'promote and protect the global free flow of information online'. These principles also emphasise the importance of transparency, due process, accountability and inclusive, multistakeholder policymaking. An advisory council comprised of civil society groups endorsed the recommendation.

An emerging trend in intermediary liability policy is that it is evolving into a legal mechanism that allows governments to transpose their own interpretations of limitations to freedom of expression onto the internet, even beyond national jurisdictions. Depending on national, social and historical context, governments emphasise the restriction of different types of content, and non-compliant intermediaries may face criminal prosecution like imprisonment, civil penalties like fines, or a revocation of operating licenses. The complexities here are linked to where the content is held, where the authors are based, and where the intermediary is headquartered.

2.2.2 Special note: Intermediary liability in Africa

While internet usage is growing fastest in the developing world, there are few legal provisions related to intermediary liability in many parts of Africa. Absence of such provisions creates regulatory and procedural uncertainty. A 2014 report by an international NGO with consultative status to the UN's Economic and Social Council, the Association for Progressive Communications (APC), argued that the lack of protection for intermediaries in African countries causes intermediaries to proactively restrict content on their networks and platforms, resulting in what may be the undue restriction of users' freedom of expression.

At the same time, many African countries form part of the emerging trend of crafting intermediary liability regimes, partly in response to approaches by international bodies and major trade and aid partners to protect intellectual property rights and ensure that intermediaries take action against copyright-infringing material on their networks and platforms. For countries putting in place intermediary liability regimes, civil society groups concerned with freedom of expression such as the APC have voiced concern over potential 'cherry picking' from other countries' regimes to establish restrictive and selective regimes. Because monitoring by intermediaries for potential illegal content could compromise internet users' right to privacy and freedom of expression, strong data protection and privacy laws have been identified as an important safeguard to ensure that intermediary liability regimes are not abused for arbitrary surveillance or monitoring purposes. Indeed, while the lack of intermediary liability regimes weakens freedom of expression, the simple existence of an intermediary liability regime does not guarantee stronger protection either for intermediaries or for online freedom of expression in general. In addition, intermediaries' own terms of service may be inadequately aligned with freedom of expression standards. Competence of the courts and the presence of entities able to advocate for international standards for human rights online are key to ensuring protection of intermediaries and online freedom of expression.

2.3 SELF-REGULATION AND CO-REGULATION

Laws are not the only source of online content restriction; a company's private rules like its 'terms of service' can also circumscribe freedom of expression. In 2011, the four international rapporteurs on freedom of expression have pronounced self-regulation to be an 'effective tool in redressing harmful speech' which 'should be promoted'. In some jurisdictions, systems to set and enforce rules for online expression combine elements of public and private authority, resulting in self-regulation and co-regulatory enforcement mechanisms. The scope and power of these mechanisms are in turn heavily shaped by states' legal and regulatory contexts. Thus there is a great deal of fluidity and inter-linkage between public and private regulation. The predominant trend is of internet intermediaries engaging in some degree of self-regulation and private enforcement. The specific constitutional, legal and regulatory frameworks of a given jurisdiction, particularly its intermediary liability regime, in turn shape the extent and nature of self- regulation and co-regulation taking place. As early as 2003, self- and co-regulation were viewed favourably; a Council of Europe declaration encouraged 'self-regulation or co-regulation regarding content disseminated on the Internet' by member countries. If such systems are not to serve as censorship, they should operate in terms of criteria and processes aligned with international standards on freedom of expression. The ecosystem of options can be elaborated as follows:

- **Company self-regulation**: At the level of the individual company, this ranges from measures taken by the company to block or remove spam and viruses, to the setting and enforcement of 'terms of service', rules that users must agree to abide by in order to use the service. The terms of service of one company may be very similar to legal and regulatory requirements, whereas other companies prohibit content that is legal but deemed by the company to be undesirable or incompatible with the purpose or character of its service. Within the legal framework of at least one jurisdiction, private sector companies are normally allowed to draft their own terms for what constitutes undesirable content. However, because large intermediaries effectively serve as quasi-public spheres, some advocates have argued that these companies have a responsibility to assess the human rights implications of their private rules in order to minimise negative impact on users' rights, and that this should shape their individual policy and practice. Some governments actively encourage or even place pressure on private business to self-regulate as an alternative to formal legislation or regulation, which is inherently less flexible and usually blunter than private arrangements.

- **Collective self-regulation**: A group of private entities may jointly create industry codes of conduct or set common technical standards by which all participants agree to abide.
- **Co-regulation**: An emerging trend of self-regulation, particularly in the European Union, is being implemented as an alternative to traditional regulatory action. A regulatory regime involving private regulation that is actively encouraged or even supported by the state through legislation, funding, or other means of state support or institutional participation, has come to be known as 'co-regulation'. This model provides for greater accountability for the decisions of internet intermediaries. However, it may also co-opt these companies into a role in which their decisions follow standards that fall short of the international principles concerning limitations of free expression and privacy.

All three of the case studies in this chapter examine various models of self- and co- regulation. Proponents of industry self-regulation argue that it is preferable to government regulation because such coordination is more flexible and more effective, deters legal yet undesirable conduct in the context of a particular service's purpose, helps consumers evaluate and choose between products and services, and can lower costs. On the other hand, critics warn that self-regulation's frequent shortfalls in regard to public accountability and due process may fail to protect democratic values and neglect basic standards of justice.

2.4 INTRODUCING THE CASE STUDIES

With the issues around legal and regulatory context established above, the next three sections analyse ISPs, search engines and social networks, examining the extent to which individuals' rights are respected when their freedom of expression depends on private sector internet intermediaries. The three case studies illustrate how an internet user's freedom of expression hinges on the interplay between a company's policies and practices, government policy and jurisdictional issues. Key questions include: *To what extent do companies make concerted efforts to respect users' rights in the face of government requests and legal frameworks that are not always consistent with international human rights norms? What is the impact of private terms of service on freedom of expression? In addition to limitations on content, to what extent do company data protection practices and privacy policies, combined with government surveillance requirements, affect whether people can freely express themselves?*

Clearer understanding of these matters by all stakeholders can help foster freedom of expression online: supporting governments in formulating laws that protect online rights as well as facilitating intermediaries' respect for users' rights; helping companies improve their policies and practices to foster freedom of expression via their services; and helping civil society hold governments and companies accountable.

3. STUDY 1: ISPS – VODAFONE, VIVO/TELEFÔNICA BRASIL, BHARTI AIRTEL, AND SAFARICOM

3.1 INTRODUCTION

ISPs allow users to access and use the internet via fixed-line or wireless connections. They enable the transmission of data to and from other intermediaries over their networks. ISPs can be state-owned, partially privatised or fully privatised. Many are operated by companies whose original business focused on traditional and mobile telephone services prior to expanding into internet services. Companies that act as ISPs may also provide other services like voice calling, web hosting, cloud computing, domain name registration, email and other services. This case study focuses on the core functions of an ISP as a provider of internet access via wireless and fixed-line services.

As the Guiding Principles of the Industry Dialogue point out, telecommunications can enhance openness and transparency, and are pertinent to governments in protecting public safety and security. ISPs play a critical role in facilitating the right to freedom of expression given that internet access is a prerequisite for enabling the free flow of information globally. They act as internet 'gatekeepers' given their direct access to, and technical ability to restrict, voice or data communication on their networks. ISPs also have the ability to collect, store and access users' personal data and the content of their communications as well as metadata such as IP addresses, call record details and location. They can face legal mandates, and even extra-legal interference through informal pressures, to provide access to this information, and can also face legal requirements to facilitate real time monitoring and surveillance. For these reasons, ISPs' roles at the network level can affect users' freedom of expression on other intermediaries' services, such as search engines and social networking platforms.

The business models of ISPs generally require the investment of substantial physical infrastructure, equipment and personnel in the jurisdictions where they or the telecommunications providers operate. Thus, their policies and practices affecting freedom of expression map more closely to a jurisdiction's political and legal context than that of other intermediary types such as search engines or social networking platforms outside of their home jurisdiction. Nonetheless, ISPs do have control over a range of company business decisions, policies and practices that affect freedom of expression online.

3.1.1 The Companies

This case study examined the following ISPs:

- **Vivo Telecommunications**, also known as Telefônica Brasil, was launched in 1993. In May 2014, with 79 million cell phone subscribers, Vivo had become the largest telecommunications company in Brazil, and offers mobile, broadband, and cable services.

- **Bharti Airtel** is an Indian multinational telecommunications business founded in 1995. Airtel offers 2G, 3G and 4G wireless services, mobile commerce, fixed-line services, high speed DSL broadband, IPTV, DTH and enterprise services, including national and international long-distance services to carriers in 20 countries. Airtel is ranked as the world's fourth largest mobile operator with a subscriber base of over 200 million.

- **Vodafone** is a UK-based multinational telecommunications business founded in 1991. Vodafone is the world's second-largest telecommunication provider with a subscriber base of over 430 million customers and operating businesses in 21 countries in addition to joint ventures like Kenya's Safaricom, which Vodafone calls its 'local associate operator'.

- **Safaricom** is Kenya's largest mobile operator with 21 million subscribers. According to Bloomberg Industries, as of March 2014 Safaricom claimed 67 per cent of Kenya's mobile-phone market, as well as 79 per cent of voice traffic and 96 per cent of text messages. Safaricom is 40 per cent owned by Vodafone; the Kenyan Government owns 35 per cent and the remaining shares were publicly floated on the Nairobi Stock Exchange in June 2008.

A summary of the general findings of this case study is now presented below, highlighting the broader issues embedded in the experiences of these companies.

3.2 DIRECT RESTRICTIONS ON FREEDOM OF EXPRESSION

Restrictions carried out by ISPs are 'network-level' restrictions because they either prevent or restrict an individual's access to the internet itself or prevent or restrict access to online content, expression opportunities, and services that are offered by other types of intermediaries. Network-level restrictions made by ISPs affect the nature and extent of restrictions carried out by other intermediaries.

3.2.1 Network-level filtering

Filters are specialised software programmes that can restrict access to entire websites, types of online services, specific pages or content within websites, or webpages containing specified keywords. State-mandated filtering is usually carried out by ISPs and can be required as one of the conditions of a company's operating license in a jurisdiction. The state may also install centralised filtering mechanisms through internet exchange points that serve as gateways for internet traffic between different jurisdictions, and to and from the networks operated by different ISPs. Private or local institutions such as schools and libraries can deploy filters on their own local networks to block access to certain content. Filters can also be installed at the household level, most commonly by parents seeking to control what content their children can access. Given the availability of software filters that parents can control on their own home networks, international experts have questioned why ISPs should be legally required to filter content. For certain types of content, such as hate speech, empowering users with Media and Information Literacy capacities, and self-regulation, are sometimes viewed as more conducive to upholding freedom of expression compared to direct regulation and legal enforcement. Nonetheless there are concerns about delegating too much enforcement to private intermediaries. The paragraphs below focus primarily on state-mandated filtering by ISPs as well as other filtering that ISPs might deploy to enforce their own rules, or to participate in collective industry self- and co-regulation.

Depending on the legal context, ISPs can receive requests, recommendations, and orders for filtering from the government, private third parties, and/or regulatory organizations. Such orders can be communicated on a case-by-case basis directly to the ISP, or in the form of a general 'blacklist'. ISPs in some jurisdictions take self-regulatory or co-regulatory steps, including vetting content on their networks by company standards, as well as collaborating with hotlines, regulatory, and industry bodies to identify infringing content. ISPs can also offer individual users the option of applying filters to their home and office networks. Freedom of expression can be affected by the reasons for filtering, the practical implementation of the filtering, and the transparency by government and companies about how and why the filtering occurs. Most companies include in their terms of service prohibited types of content and activities on their networks. Specificity varies, but the predominant trend is the use of broad terms to capture many forms of disallowed content.

ISPs filter a broad range of content types in response to requests, in compliance with the law and in terms of their own terms of service*.* In the countries covered by this case study, typical types of content filtered by ISPs based on government order or legal mandate include what are deemed to be copyright-infringing materials, pornography, child-abuse images, defamation, hate speech, election-related speech and materials sensitive to national security. In general, license agreements and the law vastly limit the choices available to ISPs for challenging government filtering requests. This includes decisions about: 1) whether to comply with a request; 2) the type of public notice and explanation of the restriction provided by the service provider; and 3) whether and when to remove filters on particular content. Overbroad or inconsistently applied laws can result in the inconsistent application of filtering within a country, as well as the filtering of entire websites instead of specific infringing content within those websites, contradicting the necessary and proportionate principle. Overbroad filtering is also known as 'collateral filtering' because of the collateral damage it can inflict upon freedom of expression.

Self- and co-regulatory efforts involving ISPs vary widely depending on national context. Self-regulatory measures in the form of 'family friendly' filters made available by ISPs to users on their personal connection can risk placing the service provider in the combined role of judge, jury and police: the ISP is responsible for determining the criteria to be included in the filter, implementing the filter, and addressing complaints about mis-categorized websites. Sometimes measures that begin as self-regulatory schemes can later be turned into regulation, or formalized in the law.

BOX: Emerging trend: 'Upstream filtering'

ISPs' practice of 'upstream filtering' can hinder freedom of expression. As companies begin to filter in one jurisdiction, other jurisdictions served by the provider can be affected by these practices. This is a result of 'passing on' a filter (or other technical component) in place on the ISP's network, known as upstream filtering. The result is that content considered illegal in one jurisdiction and subsequently restricted is continued to be restricted in another jurisdiction, where it might be legal.

3.2.2 Service shutdowns and restriction

Governments sometimes order network shutdowns or restrict internet services at a regional or national level, citing reasons related to the prevention of terrorism, maintenance of public order and prevention of public unrest. The restriction can affect the entire network or a specific service. In many jurisdictions, ISPs must comply with such orders or risk legal penalty. They can also restrict or shut down the network or a service for maintenance or technical failure. The shutdown of an entire network or restriction of a service in a large area is a broad stroke that impacts all content, at risk of not meeting internationally recognised principles of proportionality and necessity. Other more narrowly targeted measures can also be taken, such as government or ISP orders to terminate or suspend an individual user's access to the internet or mobile services. Mobile telecommunications companies also receive orders from governments requiring them to send messages via their networks, which can chill freedom of expression, especially if the messages are not sent out in the government's name, because such measures 'push' certain information to users even if they do not restrict information.

ISPs generally only restrict the entire network for maintenance or reasons out of their control, but they more often suspend or terminate user accounts. The circumstances regarding when the service or the network could be affected as per company policy differs. All ISPs reserve the right to terminate, suspend or moderate service for abuse of the service or breach of the company Terms and Conditions. ISPs are faced with difficult decisions about how to comply and how to communicate with the public about their compliance.

3.2.3 Network neutrality

'Network neutrality' is the principle that ISPs should treat all data equally and not prioritize data or services for any reason, including commercial and political. Net neutrality is important for freedom of expression because it preserves an individual's choice and right to access internet content, applications, services and hardware. ISPs have access to technologies that allow them to analyse, block or slow down content and services. These practices – which result from economic reasons, bandwidth regulation and restriction of content – can threaten network neutrality. Scholars recommend greater transparency by companies as to how their broadband services work, what types of network management activities they engage in, and how such activities might affect consumers. Across jurisdictions, governments and regulators are struggling to understand how and if network neutrality should be protected by law, and what responsibility companies should have in ensuring it. Despite an emerging trend of jurisdictions proposing legislation, there still exist a number of regulatory gaps around net neutrality. Thus, practices vary from company to company. Controversy has arisen over instances where a certain service or bundle of content is 'zero-rated', in the sense of connectivity costs being waived, such as the case of Facebook's internet.org initiative. The argument in favour of such options is that providing free access to a (subsidised) part of the internet is better than nothing. Either way, there are freedom of expression and right to information issues to be considered.

3.3 PRIVACY

Service providers have access to a broad range of information about their subscribers, including metadata and communications content. According to the UN High Commissioner for Human Rights' 2014 report on the Right to Privacy in the Digital Age, internet service providers 'should adopt an explicit policy statement outlining their commitment to respect human rights throughout the company's activities' and 'should also have in place appropriate due diligence policies to identify, assess, prevent and mitigate any adverse impact'. Further, the report holds that 'even the mere possibility of communications information being captured creates an interference with privacy, with a potential chilling effect on rights, including those to free expression and association.'

Only some of the companies investigated in this case study publish privacy policies applicable to dedicated services offered locally, or clearly and comprehensively explain what user data they collect, how long they use it, and what they do with it. Despite legal mandates in many jurisdictions defining the time periods for which data must be retained, the companies studied do not specify in their terms of service or privacy policies, the exact time period for which they retain data. In most of the countries investigated, legal requirements oblige users to present government-issued identification when signing up for services. Such requirements typically apply to both post-pay and pre-paid services and are distinct from ISPs requiring personal information from the user for carrying out commercial transactions. Some jurisdictions legally obligate ISPs to verify this information before providing services to the user. This heavily reduces the space for anonymous online participation, as users' online behaviour may not only be tracked but also linked to their actual identity without the privacy protections of international standards covering legitimate limitations of rights.

3.4 TRANSPARENCY

In its 2012 report *Opening the Lines: A Call for Transparency from Governments and Telecommunications Companies*, the GNI recommends that ISPs and governments be transparent about applicable laws and operating licenses, government requests for user metadata and content, and government requests for filtering and government requests for text messages sent via the ISP's network without attribution. Transparency about laws, policies, practices, decisions, rationales and outcomes related to privacy and restrictions on freedom of expression allow users to make informed choices about their own actions and speech online. Transparency is therefore important to users' ability to exercise their rights to privacy and freedom of expression.

The practice and scope of company and government transparency about surveillance practices, filtering and service restrictions vary across jurisdictions. In none of the countries studied are ISPs legally required to be transparent about their policy or practice regarding filtering, service restrictions or surveillance measures. For ISPs and telecommunications services, the ability to be transparent with customers and users is heavily dependent on whether the government itself is transparent and also whether legal frameworks allow meaningful levels of transparency on the part of companies. In the jurisdictions studied, there is little government transparency about the nature and volume of official requests made to ISPs for filtering or service restrictions. Governments do not offer overviews or official statistics about the number and type of restriction orders they issue. Sometimes governments acknowledge restrictions or respond to allegations of restriction in the media, or to queries from other branches of the government, although such instances are not standard or consistent. The predominant trend of ISPs studied here is a lack of transparency about the extent to which they carry out filtering, policies on filtering, or explanations about the legal requirements for filtering. Some laws do not explicitly prohibit ISPs from disclosing information about surveillance and filtering, but when asked to clarify, authorities have made statements that conflict with existing practices.

3.5 REMEDY

For ISPs in jurisdictions covered by this research, potential remedy can be provided for an individual user or an entire group of users whose right to freedom of expression has been infringed. Remedy can include an investigation, a public report/explanation, the reinstatement of content or connection, or the provisioning of an alternate means through which users are able to express themselves. It is thus possible for courts or tribunals, companies, and regulatory bodies to grant remedy. The form of remedy available to users depends on the jurisdiction of the company and the user. Mechanisms for complaints and dispute resolution can potentially complement, or serve as an alternative to, systems for redress and remedy provided by government. Some governments require that companies institute private grievance and remedy mechanisms, and obtaining remedy through the courts can be time-consuming and expensive in many countries. Little international research has been done in relation to best practices for consumer protection bodies in handling cases related to telecommunications.

3.6 CONCLUSIONS

ISPs play a fundamental role in connecting users to a wealth of knowledge, opportunities, and possibilities for expression. Yet some users argue that companies need to do more to protect freedom of expression. Companies such as Vodafone have pointed out that opposing government requests can pose high risk for them in terms of business and the safety of their local employees. On the other hand, sometimes compliance may pose a risk to companies in damaging the trust of their users.

A number of general observations can be drawn from this case study's findings:

- ***Governments and companies offer little, if any, transparency about restrictions of expression made by and through ISPs.*** A predominant trend is seen in the severe lack of transparency by governments and companies across a range of jurisdictions about basic aspects of filtering practices. In the wake of the 2013 Edward Snowden revelations, public dialogues and research initiatives have focused on transparent about privacy and surveillance requests, with much less emphasis on transparency of practices that directly affect internet users' freedom of expression.
- ***On surveillance, government transparency is limited, and few companies speak up for their users.*** Two countries publish annual reports reviewing the scope of government surveillance. Vodafone publishes clear policy guidelines on how it deals with government requests for user data, while subscribers of other services are left in the dark about how their privacy is protected in the face of government or other pressures. This uncertainty is compounded by the lack of information on government user data requests. As of 2014, Vodafone was the only company to report on the number of user data requests it received from government agencies. A significant

number of users' information is requested, though no numbers about compliance are provided. Vodafone was also the only company covered in this research calling openly for greater government transparency and for legal reforms that would enable the company to report more detail about surveillance and user data requests.

- ***Company data protection and privacy practices vary widely, in tandem with the existence of data protection laws***, which are in a state of flux worldwide. A predominant trend is between weaker privacy laws and weaker privacy policies by ISPs. In countries with weaker or nascent legislation, ISPs disclose much less information about their privacy-related practices.

- ***It is difficult for individuals to hold companies and governments accountable for actions taken via ISPs that restrict users' freedom of expression in a manner incompatible with international human rights standards.*** In some jurisdictions, industry regulators can offer means by which users can report infringing content or report ISP practices that violate their rights. However, redress for violations by ISPs or government agencies of users' online freedom of expression have been limited to monetary fines, showing that recognition of and consequences for violations are limited.

- ***Public commitments by some companies to international human rights principles are an important first step, but there is a long way to go.*** As indicated earlier, in 2013, a group of telecommunications operators and vendors including ISPs launched the Telecommunications Industry Dialogue on Freedom of Expression and Privacy with a set of 'Guiding Principles' influenced by the UN Guiding Principles on Business and Human Rights. Vodafone and Telefónica were among the nine Industry Dialogue members, and their 2014 reports are billed on the ID website as a product of the companies' commitment to report annually on 'progress in implementing the principles and, as appropriate, on major events occurring in this regard'.

The Industry Dialogue has said that it is engaging in collective study of best practices in corporate transparency for their sector, as well as 'how to implement operational-level grievance mechanisms'. Members have also acted collectively to engage with governments. As their first annual report states, the Industry Dialogue's intention is to 'continue to advocate for greater government transparency on the use and scope of surveillance of communications and on actions that have the effect of restricting the content of communications, in keeping with our Principles'. The concrete impact of such company activities and commitments on internet users has yet to be studied systematically. Nonetheless, activities of the Industry Dialogue member companies thus far indicate that collective action, combined with broader stakeholder engagement, has empowered ISPs to take steps that they had previously not been willing to take on their own. The Industry Dialogue has not yet recognised that it would benefit further in terms of credibility by adding an assurance process to verify whether companies are implementing their commitments, such as the third-party assessments carried out by the GNI.

4. STUDY 2: SEARCH ENGINES – GOOGLE, BAIDU, AND YANDEX

4.1 INTRODUCTION

Search engines are a principal means by which internet users find and access information. They are important for freedom of expression because they act as an intermediary between people who seek information and people who publish information online. Most web pages are not indexed by search engines and therefore cannot be found in search engine results. Even Google, the world's largest and most popular search engine, has indexed only a small percentage of the world's web pages. There are three main reasons for this: a) the web pages have not yet been found or cannot be found by the spiders because no other websites link to them; b) they are 'invisible' to spiders because the owners of web pages and online databases have chosen to block them; c) the database structure of most websites 'hides' pages from discovery by an external spider.

Every search engine uses its own search algorithm, a complex mathematical formula that decides what results to display, and in what order, in response to a user's specific query. The algorithm's decisions about what is most relevant to the searcher are triggered in part by elements in a web page's URL, headlines and other content. Those who want their content to be viewed by large audiences can 'optimise' their websites for example, to incorporate a version for mobile devices, to maximize the probability that their page will appear near the top of a search engine's displayed results. No two search engines will produce the same results or number of results for the same query, unless their algorithms, spiders, and indexes are identical.

Freedom of expression in relation to search engines involves three potential parties: 1) individual users seeking information; 2) creators and operators of websites that are or potentially may be indexed by search engines; 3) the search engines whose algorithms have been viewed by scholars and emerging jurisprudence as a kind of editorial process, albeit not as direct as that of a media organization. This section examines how jurisdictions shape search engine policies and practices related to content restriction and content manipulation, and to varying degrees by the laws and regulations of other jurisdictions. It also analyses how three different companies headquartered in three very different national contexts have handled challenges related to online freedom of expression.

The companies

This section focuses on three search engines, run by companies that provide services beyond search:

Baidu dominates in China with between 60 and 70 per cent market share of the world's largest internet user base of over 600 million.

Yandex has more than 60 per cent market share in The Russian Federation, a country of 84.4 million internet users.

Google is the world's dominant search engine. Its market share in the United States (with about 280 million internet users) is 67.5 per cent. Google's market share is much higher in countries where there is no major local competitor, with 97 per cent in India and 90 per cent in Europe. Google's market share in The Russian Federation is about 25 per cent and less than 2 per cent in China.

The broad points emerging from this case study are now presented below.

4.2 IMPACT OF NETWORK FILTERING ON SEARCH ENGINES

The freedom of expression of search engine users can be affected when a search engine is filtered by ISPs. If the search engine's front page is filtered, then the service is wholly inaccessible to users accessing the internet via that particular ISP or national network. It is also possible for the ISP to filter only specific pages of search engine results containing specific URLs or keywords, making the service partially usable – as long as the user is not searching for content that is filtered by the ISP.

The search engine operator usually has no control over and plays no role in filtering by ISPs. However, the nature and extent of ISP filtering in a given jurisdiction affects how search engines in turn carry out their own restrictions. Thus, prior to a discussion of the policies, practices, and implementation of restrictions by the three search engines themselves, it is necessary to describe the extent and nature of ISP filtering of search engines in each country covered in this case study. In the four jurisdictions covered by this case study, four different approaches to search engine filtering were identified:

- No filtering of search engines
- Filtering of websites, but not search engines
- Limited filtering of search engines
- Extensive filtering of search engines with international character with temporary disconnections to discourage use

4.3 MEASURES TAKEN BY SEARCH ENGINES

The legal environments of the companies' home jurisdictions heavily shape their policies and practices related to content restriction. In any given jurisdiction, search engine operators may restrict or manipulate content through any or all of the following actions:

1. Remove specific pages or even entire websites from the search engine's index;
2. Program the spider not to add certain pages, websites, or sites containing certain content;
3. Program the search engine's algorithm not to deliver results for certain queries;
4. Program the algorithm to favour or 'weight' certain types of web pages over others;
5. Influence the user's understanding of certain search results by adding explanatory statements, warnings, or statements in accompanying advertising.

As with the service providers discussed in the previous case study, search engines may restrict content at the request of a government authority or other external party, or may restrict content to enforce their own terms of service and other private rules or procedures.

4.3.1 Personalisation

In 2005, Google started tailoring search results for all logged-in users to their apparent preferences and interests based on search history. In 2009 personalisation was extended to all Google searches even if the user is logged out, based on browser cookie records. Critics have expressed concern about the effect of personalisation on freedom of expression because it renders the same website more or less visible to different users depending on their prior browsing habits. The full impact of personalization on freedom of expression globally remains unclear. Some have argued that the issue is less about the degree of personalisation and more about the extent to which the user is able to understand and control the factors affecting their own searches. One recent academic study of Google searches found that personalisation varies widely depending on the query, and that personalisation was much less measurable for queries made on Google when logged out. Personalisation also occurs on Baidu and Yandex.

4.3.2 Europe and the 'right to be forgotten'

Even when operating in legal environments where freedom of expression receives strong protection, search engines are not entirely neutral arbiters of information. Adjustments are made globally to the search algorithm in order to protect users from spam and malware or identity theft, to protect children from sexual exploitation, and to comply with intellectual property law. Many more adjustments are made in response to private and government requests in specific jurisdictions around the world. The role of search engines now faces a further set of challenges in Europe – and potentially around the world – with the judicial ruling establishing the 'right to be forgotten' throughout the European Union.

A 2012 UNESCO report highlighted the inherent tensions between privacy and freedom of expression. One of many potential tensions is between the individual's desire to eliminate negative information about him or herself from the internet and the right of others to receive and impart information. On 13 May 2014, the European Court of Justice ruled in the case *Google Spain v AEPD,* brought against Google by a Spanish man who argued that an auction notice of his repossessed home appearing in Google's search results constituted a violation of his right to privacy. According to the court's ruling, internet users in Europe now have the right to demand that search engines remove links to web pages about them that are 'inadequate, irrelevant or excessive in relation to the purposes of the processing'. Furthermore, the individual's right to privacy overrides 'as a general rule' the public's interest in finding information. At the same time, the public interest may be preponderant, for example in the cases of public figures.

The ruling came under heavy criticism from free expression groups such as ARTICLE 19, the Committee to Protect Journalists and Index on Censorship, who warned that excessive enforcement of privacy rights can impinge on press freedom. This position aligns with one that recognises that press freedom represents the right to use free expression to communicate with the wider public, and that while removing links to content does not per se violate the original expression, it eliminates much of the significance of publishing in the digital age. Other digital rights advocates argued that media coverage and the free expression community overreacted, noting that Google was not deleting data, but merely blocking links from search results. Moreover, Google was given a great deal of discretion in responding to individual requests and is not compelled to remove any results prior to a court ruling.

By the end of May 2014, Google came up with a rudimentary framework for compliance with the ruling and to cover itself from subsequent cases based on the ruling. It created a public web page through which users based in Europe could request that their names be decoupled from certain search results. The removals would only take place on Google search websites specific to the European Union and would remain visible on the global search engine, Google.com. A notification that such removals had taken place would appear on the search results page.

As a member of the Global Network Initiative, Google faced the need to reconcile compliance with the ruling with its GNI commitments to be transparent about how content is restricted, as well as to interpret official requests around content restriction as narrowly as possible. On 11 July 2014 Google reported that it had received 70,000 restriction requests covering 250,000 websites since mid-May. The requests were being reviewed manually, and the company had also instituted a policy of notifying websites when the link to one of their pages was removed. *The Guardian* newspaper was one of the first news organizations to receive notifications that links to some of its stories had been removed from its search results in the EU. Wikipedia's founder Jimmy Wales condemned the process as 'censorship' after his organization received notice that several links to Wikipedia content had been removed in compliance with requests from people who were the subject of that content.

Google also set up an advisory council to investigate how it should balance privacy and freedom of expression. Senior Vice President and Chief Legal Officer David Drummond wrote that while some requests were clearly illegitimate, like politicians seeking to cover up prior misdeeds, one could sympathize with many others. In the third quarter of 2014, the company held public consultation sessions across Europe and published an online questionnaire seeking public comment. Questions included: *What is the nature and delineation of a public figure's right to privacy? How should we differentiate content in the public interest from content that is not? Does the public have a right to information about the nature, volume, and outcome of removal requests made to search engines?*

Meanwhile, following the ruling, an emerging trend of similar changes appeared around the world. For example, privacy regulators attending the Asia Pacific Privacy Authorities (APPA) forum in June 2014 in Korea discussed the possibility of 'engaging with Google and other search engines' and subsequently discussed the topic at the next APPA meeting in December 2014. The implications of implementing similar rules in other jurisdictions began to come under debate. There is also a debate sparked by a French court, which said it was not enough for Google to delist only in the national iterations of its site (e.g., google.fr, google.es), when google.com – available in Europe – retained the link. The impact here could be to block google.com from Europe, or for the company itself to block the designated information to any queries coming from IP addresses in Europe. If the alternative was for the search engine to apply the European ruling to its worldwide operations of google.com, this would constitute an over-reach of one extra-territorial jurisdiction that would not be sustainable on a global basis. As noted earlier, a solution still to be considered is a mechanism operated by search engines and allowing for a 'right to reply' in regard to links that a person finds problematic.

What this particular case, along with that of the *Delfi v. Estonia*, may indicate is an emerging trend of courts creating, rather than applying, policy and precedent, as a result of the absence of prior policy and law having been set down by representative public authorities in response to technological developments.

4.4 DATA RETENTION, COLLECTION AND SURVEILLANCE

The retention of user data by search engines combined with heightened knowledge about government surveillance practices appears to have affected the public's trust in search engines. An analysis of publicly available Google (Search) Trends data before and after June 2013 (when Edward Snowden began to release his revelations about government surveillance via internet intermediaries) sought to find 'empirical evidence of a chilling effect on users' willingness to enter [sensitive] search terms'. They examined search traffic data for 282 terms in 11 countries. Nine countries exhibited a decrease in search traffic for terms rated as 'likely to get you in trouble with the U.S. government', but an increase for terms 'not likely to get you in trouble'. In the United States, the magnitude of this drop was 2.2 per cent. Such studies indicate that at least in some societies, awareness of the lack of privacy and existence of some level of pervasive surveillance can begin to have a degree of chilling on search engine users' freedom of expression. Concerns about data collection by search engines have prompted the rise of alternatives that claim not to track or store users' digital data.

4.5 TRANSPARENCY

Members of the Global Network Initiative specifically commit to 'respect and protect the freedom of expression of their users' in the course of responding to government requests to remove content or hand over user data. They also commit to be held accountable to this commitment. There are two components of public accountability for GNI members: 'independent assessment and evaluation' of whether the companies are upholding their commitment to the GNI principles, and also 'transparency with the public'. Two years after the GNI's official launch with three company members in 2008, the practice of what has come to be called 'transparency reporting' appeared as an emerging trend.

4.6 REMEDY

There are two potential parties whose freedom of expression rights might be affected by search engines: internet users broadly, and also the creators and operators of websites, including individuals with personal blogs and websites, civil society organizations, and news organizations. Other parties can and do have grievances related to other rights – such as content creators concerned about links to file-sharing sites that violate intellectual property. This section focuses on remedy and grievance mechanisms related only to freedom of expression and not on mechanisms addressing other rights. None of the search engines studied have complaints, grievance or remedy mechanisms that can be used by internet users who believe that their freedom of expression has been violated due to the way in which a search engine governs its content.

Google offers a mechanism for website owners to challenge removal of links to their websites based on the U.S. Digital Millenium Copyright Act (DMCA). Since the company began to implement the European 'right to be forgotten' ruling, Google has reinstated links to some news articles that were originally restricted. However, the process for handling appeals for reinstatement is unclear. Prior to the roll-out of Google's new 'right to be forgotten' web form in the wake of the European court ruling, none of the search engines studied had provided mechanisms for users to seek remedy if they believed that their privacy or reputational rights have been harmed by search results. While Europeans have successfully used the courts to seek remedy for alleged privacy violations by a search engine, plaintiffs have not been successful in using courts to obtain remedy for search engines' restriction of links to their websites.

4.7 CONCLUSIONS

Search engine policies and practices related to content restriction and content manipulation are shaped by their home jurisdictions, and to varying degrees by the laws and regulations of other jurisdictions. From an analysis of three different companies headquartered in three very different national contexts, key findings can summarised as:

- ***Differences in ISP filtering regimes have a strong influence on how, and to what extent, search engines restrict their own search results.***

- ***The stricter the liability regime in a given jurisdiction, the more likely the content will be removed either proactively by the company or upon request without challenge.***

- ***While content restriction takes place on search engines at the request of authorities, it also happens for other reasons in all jurisdictions, including for reasons that the search engines deem to be in their own, the users' or the public's interest.*** This contradicts a widespread public perception that search engines are neutral arbiters of information. Some consensus among companies and freedom of expression advocates has emerged on best practice by search engines in handling government demands and take down requests from a freedom of expression standpoint, as evidenced by the Global Network Initiative's principles and implementation guidelines. However there is no clear consensus across stakeholders about how search engines should respect freedom of expression in the course of algorithmic design and other content restrictions unrelated to government requests.

- ***Transparency by companies as well as governments plays a crucial role in fostering public trust in a search engine's practices and in ensuring that freedom of expression is not restricted for illegitimate or accidental reasons.*** There are a variety of examples of why it is important that governments be transparent to their citizens about restriction demands being made on search engines as well as the network-level filtering measures that have a direct impact upon them. It is equally important that companies be transparent to users about what is being removed at government or others' request and why.

- ***Privacy concerns are growing, but only one of the three companies studied – Google – has addressed these concerns in a public and forthright way.*** Many users expect that the companies they rely upon to find information, and to have their own content found, should be more forthcoming in regard to rights-linked information. This covers as much information about data collection, storage and sharing practices as the law allows, and protecting data to the greatest extent possible within the realities of their legal and political contexts.

- ***Stakeholder engagement, commitment to principles, and remedy frameworks are important for global intermediaries in addressing tensions between freedom of expression and other rights, as well as difficult regulatory situations.*** Google's commitment to the GNI since the organization's launch in 2008, and its contribution to the development of the GNI's principles since 2006, has strengthened the company's ability to respect freedom of expression and contest government requests that it believes are not consistent with human rights norms. However on other freedom of expression and privacy-related issues not related to government demands, there has yet to emerge a global stakeholder consensus around a principled framework.

5. STUDY 3: SOCIAL NETWORKING PLATFORMS – FACEBOOK, TWITTER, WEIBO, AND IWIW.HU

5.1 INTRODUCTION

Online social networks play a vital role in social interactions and expression, providing a platform that allows for the democratisation of publishing content and information. By enabling the sharing and aggregation of user-generated content, social networks are seen by some to transform audiences into information producers, providing new tools for social cohesion and with the potential for citizens to hold governments accountable. Social networks, like most internet companies that offer free services, make their profit by targeting advertising to their customers. Third-party companies buy advertisements to appear on social networks because they expect these services to be able to identify potential buyers from within their user pool through data collection and processing. Therefore, users 'pay' for the free services they use with their personal information and privacy. The platforms evolve as they develop new ways for users to create and share data. Social networks have also increased the visibility and reach of some traditional news media, for example through 'retweeting' links or sharing other online content, thereby disseminating information vastly quicker than conventional means.

Many legal systems consider social networks as 'content hosts' because users create content on their platforms, and third parties are allowed to post and share information. Because they allow private content to be publicly shared, social networks blur the line separating the public and private spheres, raising questions about appropriate expectations for expression on such platforms. Due to the scope and impact of user-generated expression and activity on social networks, it is not easy for a company to balance a commitment to free expression, legal compliance, and user expectations, with the fiduciary duty as companies to make profits. This section examines the policies and practices of several social networking platforms in a range of national contexts. It finds that the ability of social networking platforms to respect users' freedom of expression is heavily influenced by national legal and regulatory contexts, particularly by the context of a company's home country. At the same time, companies have many options available to them in terms of how they manage and design their platforms. These choices have a critical impact on users' freedom of expression.

The Companies studied:

Facebook (www.facebook.com) is a social network with headquarters in the USA founded in 2004. As of August 2015, the company had 1.49 billion monthly active users, of which 83.1 per cent were located outside of North America. It allows registered users to maintain a personal profile through which they can share personal and contact information, photos, articles and location statuses; communicate with other users via private or public messages; search and 'friend' other users, whom they may 'tag' in photos or location updates; and join groups and interact with other members. Facebook is available on the web as well as through dedicated applications on a number of mobile operating systems.

Twitter (www.twitter.com) is a US-based micro-blogging platform founded in 2006. As of August 2015, it had 316 million monthly active users who send 500 million messages ('tweets') a day. Seventy-seven per cent of Twitter's users live outside of the United States. Twitter allows registered users to exchange messages of 140 (or fewer) characters through its website, mobile application(s) or by SMS. Users can forward such messages by 'retweeting' them and can also search for and 'follow' other users. Even unregistered people can read users' tweets, as long as they have kept their profile public (the default setting). Twitter is accessible on the web and via multiple mobile applications. Tweets can be organized via hashtags (the hash sign # followed by a word or phrase), allowing users to group related posts together. If a hashtag receives high volumes of 'retweets', it is termed to be 'trending'. Twitter does not 'require real name use, email verification, or identity authentication'.

Weibo (www.weibo.com) is a Chinese micro-blogging platform founded in 2009 that was spun off from Sina prior to its public listing of shares in the U.S. in April 2014. As of May 2015 it boasted 198 million monthly active users. Users have personal profiles, post 140-character messages (called *weibo*, which means 'microblog' in Chinese) and comment under other *weibo*, a feature that provides a 'simple way for Chinese people and organizations to publicly express themselves in real time'.

iWiW (formerly www.iwiw.hu, 'international who is who') is a now-defunct Hungarian social network that closed operations in July 2014 due to its diminishing userbase. It was founded in April 2002 as wiw.hu ('who is who'), and became iWiW in October 2005, when it unsuccessfully tried to expand and began offering its platform in multiple languages. In April 2006 it was acquired by T-Online, Magyar Telekom's business unit, and in 2008, iWiW merged with Origo.hu. Until 2011 it was invitation-only. In January 2013 it had 4.7 million registered users.

Social networks are popular around the world, but they are used differently in different cultural and political contexts. Facebook and Twitter are two of the most popular social networks with broad international user bases, and thus a study of these services can shed light on freedom of expression issues in a transnational environment. iWiW and Weibo are primarily domestic operations. Weibo is especially interesting because the Chinese social-media market is highly competitive, yet insulated from foreign competition. iWiW was chosen because it represented a domestic social networking service competing in a local linguistic and cultural context against global competitors.

The broader significance of the case study of these companies is now presented.

5.2 IMPACT OF ISP FILTERING ON SOCIAL NETWORKING PLATFORMS

Governments can require ISPs to filter social networking platforms by blocking access either to the entire website, or to specific content, groups or pages. Such filtering can also take place at national internet exchange points. Companies that operate social networking platforms have no control over actions by governments and ISPs to filter them. Some social networks, such as Facebook, Twitter and Google, have spoken out against network-level filtering in general. However, companies do exercise control over their terms of service and how they respond to government and other requests to remove content or deactivate accounts on their own platforms. Companies' decisions about such platform-level restrictions may in turn affect whether or not governments choose to filter at the network level.

Governments decide to restrict or block social networks via network-level filtering under several circumstances:

- ***Differences in jurisdictional norms:*** Unlike ISPs, social networks do not require a physical presence in a country in order to reach users in it. However, some governments use the threat of network-level filtering in an effort to compel international companies to comply with their laws.
- ***In the name of preserving national unity or national security.***
- When there is a perceived ***real-time need to control and maintain public order.***

5.3 CONTENT REMOVAL AND ACCOUNT DEACTIVATION

While social networking platforms can be the target of ISP-level filtering over which they have no direct control, social networks do have their own mechanisms to block or otherwise restrict user content. They generally require users to create an account in order to share content. Platforms operators can restrict content that users share on the platform in several ways: deleting such content; blocking it from view for users in specific jurisdictions; or shutting down – deactivating – the accounts of users who post certain content. These actions may be taken as self-regulatory measures to enforce private rules, or in compliance with government or others' requests and other legal requirements such as responding to court orders in civil cases. In some cases, users can be penalised for their legitimate online expression. The potential for legal liability of the network or the user can also lead to self-censorship.

The table below provides an overview of the different modes through which content restriction occurs, reasons and affected parties.

Table 3: Key factors affecting content restriction by social networking platforms:

Reason for restriction:	Content violates terms of service?	Modes of implementation:	Who is affected:
• government requests	• possibly	• complete removal of specific content	• all users
• law-based requests (e.g. copyright takedown notices, court orders in civil cases)	• possibly	• blocking of specific content for a specific user group or jurisdiction (content remains accessible to others)	• only users in a particular jurisdiction
• self-regulation on own initiative (terms of service and other enforcement of private rules)	• Usually	• automated (pro-active) filtering of pre-identified types of content	• only specific user groups (e.g., by age)
• user reporting (on other users' violations of terms of service)	• Usually		

5.4 PRIVACY

Social networks are veritable treasure troves of private information, revealing everything from political preferences to sexual orientation. Users implicitly entrust social networks with personal data, and governments make requests for private user information in the pursuit of civil, criminal and even national security investigations. All companies examined here have privacy policies of some form that explain how user information is used, but the policies are rarely straightforward or comprehensive. In addition, default settings have significant privacy ramifications because human beings are subject to 'default bias'. Companies studied do not offer much information about data retention.

In 2015, the UN Special Rapporteur for the promotion and protection of freedom of opinion and expression and in 2014 the UN High Commissioner for Human Rights, flagged the importance of anonymity, as linked to the right to privacy, for the exercise and protection of human rights in the internet age. Many social networking platforms, but not all, require that users sign up with their real names and enforce such policies to varying degrees and in a variety of ways.

5.5 TRANSPARENCY

5.5.1 Transparency about government and lawful requests

Both Facebook and Twitter publish datasets that have come to be known as 'transparency reports'.

Facebook's 'Government Request Report' first disclosed information about content restriction in April 2014. Facebook reports only the number of government requests that it complied with, but does not report the total number of government requests received. It also does not include court orders or copyright takedown notices in its figures. Facebook's transparency report provides detail about government requests for user data – including information about compliance rate and request types. However, it provides only very basic and incomplete information about content restriction requests.

Twitter has disclosed content removal requests since its first transparency report in 2012. In addition to what Facebook discloses, Twitter's report includes compliance rate, content withheld, and copyright takedown notices. Twitter distinguishes itself from Facebook by publishing copies of the content restriction and takedown requests it receives, on the Chilling Effects website. For its reporting on government data requests, Twitter provides details on types of requests and compliance rate, as well as data about information disclosed to authorities during emergencies.

Weibo does not publish a transparency report due to legal constraints, and it appears that the off-line and online media rarely mention state-level restrictions of content.

iWiW did not publish any type of transparency report before its operations closed down.

5.5.2 Transparency about self-regulation

While Facebook and Twitter have taken efforts to increase transparency about how they handle government and lawful requests, they share much less information with users or the public at large about how they enforce their own terms of service. None of the companies studied provide information on content they have restricted based on company policy, or any statistics about external reporting on violations of company rules. No data about the number, source or subject matter of such cases has been reported by either of these companies.

While all social networks list content that they prohibit, none of the companies studied has provided much public information about procedures for evaluating content. Industry sources have described internal rules and procedures for evaluating content in conversations with concerned stakeholders, held on condition of non-attribution, but such processes are generally not made public. It is usually through anecdotal evidence via news reports that the public learns about specific examples.

5.5.3 User notification

Companies are inconsistent in informing users when they restrict their content or hand over their user data. If content is removed due to a copyright violation, both Twitter and Facebook are legally required under the USA's DMCA to notify the user and provide information on how to file a counter-notice. Furthermore, both companies commit to inform users about requests for their data, unless the situation is an emergency or the company is legally prohibited from doing so.

For content that Facebook removes to enforce its own Community Standards and Statement of Rights and Responsibilities (SRR), the company commits to forewarn people, but Twitter does not clarify if it does the same for content that violates its terms. If it implements a foreign content restriction request, Twitter notifies the public about the restriction through a 'Tweet withheld' notice, which it also uses for copyright-related takedowns. As mentioned previously, Twitter only restricts accounts in the jurisdiction whose authorities made a valid request. Facebook displays a more generic message that 'this content is currently unavailable'; this could mean many things, and it is unknown which one applies to a particular situation. When content is restricted on Weibo, other users trying to access the post are notified that it has been deleted and are directed to a link for more information. Weibo has been reported by users to 'camouflage' messages so they remain visible only to the author, causing some authors to be unaware that their content was restricted.

5.6 REMEDY

None of the companies investigated offer a clear path to remedy for users who face image or text removal or functional restrictions, such as the user's inability to upload photos. Facebook may remove pages for alleged spam violations, but users can appeal. For suspended accounts, both Twitter and Facebook offer an appeal option. When accounts are disabled for violating Facebook's terms, users can send an appeal through a specific form. There is no information about how long it will take for a request to be processed, what the decision-making procedure is, or the severity of violations that would trigger an account suspension. Twitter's information page is also short on such information, but explains more about how to appeal. One exception is copyright, as US copyright law requires Twitter and Facebook to notify the original content uploader and inform them about counter-notices. It is unclear what sort of paths to remedy iWiW offered, if any. **Weibo** does not offer a direct appeal option or a web form; instead, users are encouraged to email the company and to indicate if they 1) disagree about administrators' operations; 2) are dissatisfied with administrators' responses after communication; 3) have questions relating to other administrative matters. On Weibo, sparse information about remedies has meant largely having to rely on anecdotal evidence. Content restriction seems to be inconsistent.

5.7 CONCLUSIONS

In the interplay between social network intermediaries' policy and practice and specific national regulatory and legal contexts, companies are better able to maximise respect for users' rights in jurisdictions where laws are relatively compatible with international human rights norms regarding freedom of expression and privacy. The legal context of the country in which a company is headquartered is particularly important for the respect of user rights. Social network companies whose home governments do not inhibit such efforts have made strides in transparency and accountability in handling government demands. Yet freedom of expression can be strongly influenced in a positive or negative direction by companies' own rules, processes and mechanisms on matters including terms of service enforcement, user privacy and identity. Companies are much less transparent and accountable with the public on these matters.

The research of Facebook, Twitter, Weibo and iWiW points to the following conclusions:

- ***Government actions against social network users may limit space for expression.*** Facing fines or even arrests, users are sometimes penalised for their online expression. A lack of clarity on what expression is allowed, along with restrictive policies, can lead to self-censorship. Companies that operate social networking platforms can help by being clear and transparent about their content restriction practices, privacy settings, and data sharing policies. They can also support individuals in cases where penalties are not in line with international human rights standards.

- Social networks do not necessarily comply with all requests for content removals; for example, Twitter has complied with only 11 per cent of such requests, showing that ***social networks do have operating space to challenge content restriction requests.*** It may be easier to resist pressures from countries other than the network's home jurisdiction, but even within the home country, some companies do not comply with all requests. Of the four social networks profiled here, only Twitter and Facebook publish their criteria, if not the actual process, for dealing with content removal requests from governments and/or third parties. Such ***published policies help users understand in what circumstances their content may be removed by external request, and can give companies a clearer framework to contest content removal demands that are not consistent with due process or international human rights.***

- ***Social networks are inconsistently transparent on government removal requests.*** Of the four social networking platforms studied, only Twitter and Facebook provide information about government requests, shedding important light on how law is enforced on their platforms. Twitter also shares the content removal request itself, where possible, with the 'Chilling Effects' website and notifies the public via messages on the platform when content is restricted based on a government request. Across jurisdictions, governments are not fully transparent about the nature and scope of content restriction and requests for user data.

- ***Some social networks do not explain how they share user data with authorities and others.*** Facebook and Twitter have published policy guidelines on how they respond to user data requests from both foreign and domestic authorised bodies. Users of the other services are not informed how their privacy will be protected in the face of requests by governments or others.

- ***None of the companies studied publish data on self-regulatory restrictions***, such as, for example, how many accounts were disabled for impersonation or how many repeat infringers were terminated. As online social networks increasingly become a central platform to individuals' online expression, users and stakeholders have a strong interest in rules and enforcement processes that are clear, predictable and to some degree independently monitored. The absence of such accountability detracts from intermediaries' legitimacy as platforms for users' freedom of expression.

- Possessing a significant amount of personal information, ***social networks carry a special responsibility to respect users' right to privacy***, a requirement for individual expression.

- ***'Real name' requirements may have a serious chilling effect on speech and require flexible implementation in order to avoid negative impacts on users' freedom of expression***. Most governments do not legally require social networks to verify their users' identities. Companies could consider the privacy and free expression ramifications of implementing a real-name policy by conducting a human rights impact assessment.

- The GNI's principles on freedom of expression and privacy and accompanying implementation guidelines have provided strong guidance for GNI companies and internet intermediaries more broadly. GNI's guidelines on transparency and process for handling government requests, grounded in international human rights norms, have had an impact on company practices of all three intermediary types studied in this chapter. Yet there is a glaring absence of similar principles, guidelines and standards for companies' self-regulatory practices, including terms of service enforcement. Given the lack of transparency and consistency in how companies enforce their terms of service and other private rules, and the impact of such enforcement of internet users' freedom of expression, ***there is a clear need for the development of guidelines and 'best practice' standards for remedy and transparency in intermediaries' self-regulation***.

6. GENDER

Among the 81 countries covered by the World Wide Web Foundation's 2013 Web Index, only half had national policies addressing gender equality online. The authors of the 2013 Web Index report point out that a 'lack of political and policy focus is compounded by failure to collect gender-disaggregated statistics.' As a result, 'the ways in which gender affects Web access and use are still poorly understood.' To see how this is related to the roles of intermediaries, it is worth presenting a brief overview of the issue of basic internet access for women in relation to men. This discussion is followed by an examination of how content restriction in some countries has affected women's access to health information and gender-related discourse. The final section discusses issues related to harassment targeting women, and how these affect women's freedom of expression online by chilling their participation in the digital information society.

6.1 ACCESS TO THE INTERNET

Internet access has empowered women, bringing about greater gender equality and economic benefits to women. Yet globally there is a significant gender divide in broadband access. Factors affecting women's access to broadband include educational and income gaps, and these are more acute for women in developing countries. The lack of access and adequate internet infrastructure affects low-income rural areas the most, and women most severely. At the same time, there is a growing trend of women increasingly accessing the internet through smart phones. Policy interventions to overcome the gender gap include expanding access to affordable platforms, developing national plans to allow for increased broadband penetration, and addressing market constraints that impact the affordability of internet platforms.

6.2 GENDER AND CONTENT RESTRICTION

In some countries, women's rights advocates have demanded broader restrictions on pornographic and 'obscene' content online, arguing that there is a connection between the online viewing of such materials and off-line violence against women. Some women assert that their rights have been violated when intermediaries fail to restrict content that has been posted on the internet with the express intention of harming them. However, women's ability to access and disseminate information and ideas about sexuality can also be stifled by restrictions, and legislation whose purpose includes protection of women can be appropriated for other purposes. Internet companies, including search engines, generally do not restrict medical information related to women, but social networks' treatment of female nudity has been a huge point of contention. Moreover, laws meant to curb pornography are also used in some countries to stamp out other content. Companies continue to struggle for the right balance in relation to broad laws that can be subject to a wide range of possible interpretations.

6.3 GENDER-BASED HARASSMENT

Because of the ease with which harassment and threats can take place via social media platforms, including stalking, hate speech, cyber mobbing, revenge porn, unwanted sexual attention and sexual coercion, an emerging trend can be observed in debates around the responsibility of intermediaries to help prevent and address online gender-based harassment. Examples in this regard are provided in the companion chapter on countering online hate speech.

6.3.1 Regulation

Divergent trends can be seen with regard to legislation related to online sexual harassment: some countries have developed specific laws, others have broad provisions that could potentially encompass online sexual harassment, while others do not have legislation addressing the topic. A specific category of online harassment that has appeared as an emerging trend in discussions by policymakers and gender rights advocates is called 'revenge porn'. Perpetrators are often bitter ex-spouses or partners, or online 'trolls' who upload what the U.S. National Conference of State Legislatures has defined as 'nude or sexually explicit photographs or videos of people online without their consent, even if the photograph itself was taken with consent'. Since 2014, at least five countries and 25 U.S. states have banned revenge porn, and several others have addressed it through tort, criminal, anti-pornography or privacy laws.

6.3.2 Policies and practices of intermediaries

Negative media coverage and pressure from civil society groups have prompted some intermediaries to proactively implement mechanisms to prevent and respond to sexual harassment. Responsiveness and enforcement, however, vary depending on the degree of company commitment and attention to the issue, public pressure and legal enforcement. In a 2014 study examining how Facebook, Twitter and YouTube handle violence against women, the APC concluded that while company approaches to violence against women differ, and the companies 'have made some effort to respond to user concerns', nonetheless 'they do not do enough'. The APC report calls on internet intermediaries to balance their commitment to freedom of expression with other human rights 'such as that to be free from discrimination and violence'. As the report pointed out, sometimes companies come up with mechanisms to report abuse only after being subject to strong public criticism. At the same time, just as social media platforms are spaces where women or men face sexual and gender-based harassment, they also enable activists to fight harassment and raise awareness. In some cases these campaigns have succeeded in bringing national attention to the issues at hand and motivated political and policy-level change.

6.4 CONCLUSION

The previous section found that global companies like Twitter and Facebook are much less transparent and accountable about how they enforce their terms of service than they are about how they handle government requests. The APC study cited in this section reinforces the need for greater dialogue and communication with all stakeholders about how social networking platforms develop and enforce their rules. Companies can work more closely with users, human rights advocates of all kinds and governments, if the problem of online gender-based violence is to be addressed in a manner that upholds and protects online freedom of expression. Indeed, the problem of online gender-based violence underscores the urgent need for a multistakeholder process to develop principles, standards and 'good practice' guidelines for how social networking platforms can communicate with and listen to users about the development and enforcement of their terms of service.

7. GENERAL CONCLUSIONS

This chapter's findings highlight key challenges for realizing the first principle of Internet Universality: human rights. It builds on the UN Guiding Principles on Business and Human Rights, according to which states have a primary duty to protect human rights, businesses have a responsibility to respect human rights, and both spheres have a role in providing remedy to those whose rights are violated. The case studies highlight the difficulties that internet intermediaries face in maximizing respect for users' right to freedom of expression when states do not uphold their own duty to protect. The cases above highlight ways in which all states have room for improvement. However, it is also clear that internet intermediaries have considerable power to influence outcomes affecting internet users' freedom of expression even when the legal and regulatory environment is not fully supportive of that aim.

7.1 STATE DUTY TO PROTECT

Part of the state's duty to protect human rights includes facilitating and supporting intermediaries' respect for freedom of expression. This chapter's findings illustrate how, to varying degrees, policies, laws, and regulations are not well aligned with that particular aspect of the state's duty to protect human rights. Issues identified in the case studies included:

1. The characteristics of intermediary liability regimes or lack thereof, as well as the regulatory objectives of the regimes, affect intermediaries' ability to respect freedom of expression. Limiting the liability of intermediaries for content published or transmitted by third parties is essential to the flourishing of internet services that facilitate expression.

2. Laws, policies and regulations requiring intermediaries to carry out content restriction, blocking and filtering in many jurisdictions are often not sufficiently compatible with international human rights standards for freedom of expression.

3. Laws, policies and practices related to government surveillance and data collection from intermediaries, when insufficiently compatible with human rights norms, serve to impede intermediaries' ability to adequately protect users' privacy.

4. Licensing agreements can affect intermediaries' ability to respect freedom of expression. This applies to ISPs in all countries and social networks and search engines in some countries.

5. Whereas due process generally requires that legal enforcement and decision-making are transparent and publicly accessible, governments are frequently opaque about requests to companies for content restriction, the handover of user data, and other

surveillance requirements. This makes it difficult for the public to hold governments and companies appropriately accountable when users' right to freedom of expression is unduly restricted which may be either directly through content interference or indirectly through the compromise of user privacy.

7.2 RESPONSIBILITY OF BUSINESS TO RESPECT

Companies' own policies and practices affect internet users' freedom of expression both positively and negatively. The case studies raise issues of terms of service enforcement, identity policies, transparency practices, the extent to which companies are willing or able to contest government requests, and policies related to privacy, data retention and data protection. Key findings include:

1. Despite the recent 'transparency reporting' trend, companies perform inconsistently in terms of what they disclose and how the information is communicated. Furthermore, companies lack transparency about how they enforce terms of service and respond to private requests.

2. Companies with clear policies and practices on handling content restriction requests are in a stronger position to contest local laws and regulations that fail to meet international standards for legitimate limitations.

3. Internal decisions among companies to restrict certain types of content and to enforce their own private rules are often welcomed by governments as a way to handle problems before they escalate into matters for the courts and law enforcement. At the same time, internal rulemaking and enforcement processes lack transparency or independent oversight mechanisms that would help to ensure that they are not subject to errors and abuses. Users in most countries studied reported incidents in which measures were taken by intermediaries against content that did not appear to violate the terms, or in which the terms were enforced in an excessively literal way, resulting in a negative impact on freedom of expression, and often without adequate means for appeal.

4. Companies in all three case studies collected similar types of data, although policies about retention and third-party sharing differed widely, as did the extent to which companies inform users about policies' existence and content. The majority of companies did not clearly explain how they handle government requests for user data, nor did they offer information on actual requests for user data or compliance with those requests. While law was a contributing factor to some of these differences, company-specific factors were also at play.

5. Whether users are allowed to use a service or create an account without having their account linked to their government-issued identity, or without having to use their real name, impacts users' freedom of expression in many jurisdictions studied.

7.3 ACCESS TO REMEDY

Remedy is the third central pillar of the UN Guiding Principles on Business and Human Rights, placing an obligation on governments and companies to provide individuals access to effective remedy. This area is where both governments and companies have much room for improvement. Across intermediary types, jurisdictions and types of restriction, individuals whose content or publishing access is restricted and individuals who wish to access such content had inconsistent, limited or no effective recourse to appeal restriction decisions, whether in response to government orders, third party requests or in accordance with company policy. While some companies have recently increased efforts to provide appeal and grievance mechanisms and communicate their existence to users, rules are enforced inconsistently and without due process.

7.4 ISSUES OF CONCERN

Company policies and practices can combine with jurisdictional contexts to produce outcomes that have a negative impact on freedom of expression. Several common categories of issues emerged:

- Over-broad law and heavy liability regimes cause intermediaries to over-comply with government requests in ways that compromise users' right to freedom of expression, or to broadly restrict content in anticipation of government demands, even if demands are never received and if the content could potentially be found legitimate in a domestic court of law.

- Intermediaries can be subject to different legal norms, and are sometimes at risk of an all-out ban by authorities disagreeing with particular content shared via their services. Internet services at times resist such pressures by closer cooperation with governments, blocking content only in the jurisdiction in question, or by wholescale deletion of said content.

- Companies decide to allow or ban certain content based on their internal policies, as well as being influenced by legal obligations following court rulings, governmental orders, civil claims, instructions by third parties, requests from monitoring groups with which the intermediary cooperates, and others. This myriad of actors involved, compounded by ambiguity of legal frameworks, often makes it unclear for individual users what content is permitted, who decides on allowed content, and how, and the potential consequences of their expression.

- The existence and nature of company policies dealing with speech related to sexual harassment, gender-based violence and exploitation or objectification of women are uneven. This is found even across the same intermediary type and jurisdiction. Companies in all three case studies had in place mechanisms allowing users to report gender abuse. These mechanisms could be used for legitimate purposes, including reporting sexual harassment, but at the same time the same mechanisms could also sometimes be used for over-reach that compromises users' legitimate freedom of expression rights.

7.5 INTERMEDIARIES AND INTERNET GOVERNANCE

In 2005, the UN Working Group on Internet Governance defined 'internet governance' as 'the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programs that shape the evolution and use of the Internet'. Thus while the term 'internet governance' is often used in the media and public debates in a narrow sense to describe the technical policymaking and coordination functions of organizations such as the International Corporation for Assigned Names and Numbers (ICANN), the concept was originally conceived to encompass a broader set of processes for determining policies and practices that shape the internet's functioning at all layers. The policy role of internet intermediaries – and policies affecting their operations – is a form of internet governance broadly defined. It is therefore useful to situate this chapter's findings in the context of global debates over core principles for internet policymaking that have a direct impact on intermediaries.

The annual Internet Governance Forum (IGF), whose creation was mandated by the 2005 Tunis Agenda for the Information Society, provides a platform for stakeholders to debate the full range of issues surrounding the internet's governance, albeit without a mandate to set policy. A number of 'dynamic coalitions' were formed to support ongoing work related to the information society, leading to the emergence in 2008 of the multistakeholder Internet Rights and Principles (IRP) Dynamic Coalition. The IRP developed a Charter of Human Rights and Principles for the Internet, with a set of ten core principles launched in 2011, including principles on freedom of expression and privacy. The September 2014 IGF in Istanbul saw the launch of a new dynamic coalition on 'Platform Responsibility' focusing on a specific category of intermediaries, i.e. 'social networks and other interactive online services', to discuss 'concrete and interoperable solutions to protect platform-users' human rights'. This new dynamic coalition has similar potential to contribute to norms for social networking services, search engines and other types of intermediaries that can be defined as 'platforms' for expression. It could serve as a focal point for developing stronger human rights-based principles and accountability mechanisms for various emerging forms of self- and co- regulation.

8. RECOMMENDATIONS

The following recommendations apply in varying degrees to governments, companies, civil society and international organizations. If online freedom of expression is to be adequately respected and protected, all of these actors must find ways to work together across borders to improve legal and regulatory frameworks, establish and implement corporate best practices, and increase awareness as well as participation by internet users and citizens. Rulemaking and enforcement related to online speech – whether carried out by governments or companies – must be compatible with and held accountable to international human rights norms. The recommendations below are offered as first steps in that direction in the hopes of fostering further discussion and building greater international consensus.

8.1 ADEQUATE LEGAL FRAMEWORKS AND POLICIES

Policy, legal and regulatory goals affecting intermediaries must be consistent with universal human rights norms if states are to protect online freedom of expression and if companies are to respect it to the maximum degree possible. Governments need to ensure that legal frameworks and policies are in place to address issues arising out of intermediary liability and absence of liability. Legal frameworks and policies affecting freedom of expression and privacy should be contextually adapted without transgressing universal standards, and they should be consistent with human rights norms including the right to freedom of expression, and contain a commitment to principles of due process and fairness. They should also be precise and grounded in a clear understanding of the technology they are meant to address, removing legal uncertainty that would otherwise provide opportunity for abuse or for intermediaries to operate in ways that restrict freedom of expression for fear of liability.

In order to better inform public and private policymaking processes, there is a need for much more qualitative and quantitative global research on the impact of company policies, practices, business models and design choices on freedom of expression. Comprehensive surveys of internet users around the world on how intermediaries affect individuals' freedom of expression in different contexts are currently lacking. More research is also needed on how legal, regulatory and policy frameworks affect intermediaries' ability to respect users' rights, as well as their impact on internet users more broadly. This chapter only begins to scratch the surface in its examination of how specific companies' policies and practices affect freedom of expression in different jurisdictions. More detailed facts about cause-and-effect between policies, practices and outcomes are needed. These facts will better equip all stakeholders to refine and adjust their policies, practices, and strategies to maximize the protection of and respect for freedom of expression rights of internet users everywhere in the world.

8.2 MULTISTAKEHOLDER POLICY DEVELOPMENT

Laws, regulations and governmental policies, as well as corporate policies, are more likely to be compatible with freedom of expression if they are developed in consultation with all affected stakeholders and take into account those interests. A genuine multistakeholder process involves all stakeholders potentially affected by the policy from the start, rather than simply seeking opinions after the basic parameters have been set and key directions already determined.

8.3 TRANSPARENCY

Transparency is important to demonstrate that governance and enforcement actions comply with pre-specified principles, rules and conditions. Greater transparency by governments about requests and requirements being placed on companies that have the potential to affect users' freedom of expression and privacy is a prerequisite for accountability in public governance of the internet. Transparency by companies is a prerequisite for accountability in how intermediaries respond to government requests, as well as their own private 'governance', which is necessary not only for the protection of users' freedom of expression, but for companies' ability to earn and maintain public trust in their services.

In this context there are two kinds of transparency: qualitative and quantitative. Qualitative transparency involves governments making publicly available the laws, legal interpretations, administrative procedures and other measures related to content restriction and surveillance. For companies, qualitative transparency involves communicating with users about processes for responding to government requests and for enforcing internal company rules and processes. Quantitative transparency refers to the publication of aggregate data about government requests and compliance rates, as well as other data that helps internet users understand what types of content are being removed, under what auspices and for what reason. The GNI and the Center for Democracy and Technology have developed transparency recommendations for governments regarding content restriction. Similar transparency measures are recommended for governments in reporting both qualitatively and quantitatively on surveillance. Companies could disclose aggregated information on the number of user data and real-time surveillance requests that they receive, and how the company responds to them, on at least an annual basis. Governments could enact legal reforms that clearly permit such transparency, and companies should be able to disclose the existence and basic details about any technical requirements for surveillance that governments impose upon them.

8.4 PRIVACY

Protecting users' right to privacy is essential for freedom to expression to flourish. Intermediaries should adopt good practices with respect to privacy and have clear and comprehensible policies in place for what user data they collect and store, how they handle it, with whom they share it, and under what circumstances authorities may access it. Such policies must be prominent and easy to access. For governments, their policies, regulations, laws and enforcement practices affecting users' privacy, including data collection and surveillance for law enforcement, should be consistent with core human rights principles. The International Principles on the Application of Human Rights to Communications Surveillance developed by a global coalition of civil society groups between late 2012 and May 2014 set forth 13 principles that governments and companies could reference in order to ensure that communications surveillance is carried out in a manner consistent with international human rights standards.

8.5 HUMAN RIGHTS IMPACT ASSESSMENT

Protection of online free expression would be strengthened if governments carry out human rights impact assessments to determine how proposed laws, regulations or policies may affect internet users' free expression and/or privacy domestically and globally, and publish the results of those assessments. Companies can also carry out human rights impact assessments to determine how their policies, practices and business operations affect internet users' freedom of expression and adapt their activities accordingly, with strategies to mitigate potential harms identified in the assessments. Such assessment processes are best informed by engagement with stakeholders whose freedom of expression rights are at greatest risk online, including the media, and civil society groups who are able to able to represent those interests.

8.6 SELF-REGULATION MUST FOLLOW PRINCIPLES OF DUE PROCESS AND ACCOUNTABILITY, AND BE CONSISTENT WITH HUMAN RIGHTS NORMS

National laws need to strengthen due process and the adherence to international human rights norms to protect the rights of internet users, but guiding principles are also essential for intermediaries' legitimacy as custodians of online content. They should be a point of reference for private terms of service enforcement processes. This aligns with international standards that require any limitations on free expression to be specified in rules and be predictable, as distinct from being arbitrary or retroactive. Self-regulation should further respect the principles of necessity, proportionality and internationally-agreed legitimate purpose. Within the context of creating a safe user experience, content restrictions deployed by intermediaries should not only be as minimal as possible, but should also avoid conflict with the key human rights principle of non-discrimination – something that links to the issue of network neutrality. In order to identify and mitigate potential adverse impacts on users' freedom of expression, intermediaries can carry out human rights impact assessments on their self-regulatory system.

The Internet Society in 2014 proposed principles and recommendations for self-regulatory processes and institutions, including specific ways in which self-regulatory mechanisms should build accountable and transparent practices. Balanced and proportionate rules, due process and judicial safeguards are essential. Periodic reviews should be built into such systems.

8.7 REMEDY

Internet users have the right to effective remedy when their rights are restricted or violated by intermediaries, states, or a combination of the two. It should be possible for people to report grievances and obtain remedy from private intermediaries as well as from government authorities, including national-level human rights institutions. In seeking remedy for restrictions or violations of the right to online freedom of expression, internet users should not necessarily be required to pursue legal action through the courts. Avenues for seeking remedies should be publicly available, known, accessible, affordable and capable of providing appropriate redress.

Depending on national context, grievance and remedy mechanisms provided by states may include redress mechanisms provided by data protection authorities, national human rights institutions, court procedures and hotlines. Grievance and remedy mechanisms provided by private intermediaries and private regulatory schemes should provide mechanisms to receive and respond to grievances from internet users, as a dimension of self-regulation. These need to be accessible, secure, and linguistically and culturally appropriate. The question of whether meaningful remedy is available to users whose freedom of expression rights have been restricted or violated should be examined as part of a company's human rights impact assessment process. Depending on the grievance and the harm identified, remedy could, but need not necessarily, involve financial compensation. Meaningful remedy measures can also include acknowledgment, apology and commitment to address the problem in the future; submitting to independent investigation or ongoing oversight; or participation in regional or sector-wide multistakeholder entities to clarify and mitigate potential restriction or violation of users' rights.

8.8 PUBLIC EDUCATION AND INFORMATION, AND MEDIA AND INFORMATION LITERACY

The composite concept of Media and Information Literacy covers the range of competencies that citizens need to fully participate in knowledge societies. In their engagement with internet intermediaries, citizens require a range of literacies concerning free expression issues. Companies and governments have a role to play in promoting these literacies formally and informally. States have an obligation to provide accessible and clear information to the public so that internet users can not only understand and effectively exercise their rights, but also recognise when their rights have been restricted, violated or otherwise interfered with. State restrictions on free expression must not only pursue a legitimate aim and comply with human rights law, but should also be clearly made known to the public. Public information should include concrete instructions on official grievance and remedy mechanisms.

Respect of internet users' rights by private intermediaries also requires informing and communicating with users about their rights, how their expression can be restricted according to the intermediary's terms of service, the reasons for those restrictions and why they are necessary, and other information needed to make an informed decision about whether or not to use the service. Educational institutions should be encouraged and incentivised to include information about the rights of internet users in curricula related to human rights, civics and government. Media should similarly be encouraged and incentivised to include content that helps to foster informed public discussion about the rights of internet users, and the obligations of states and businesses to protect and respect those rights.

8.9 GLOBAL ACCOUNTABILITY MECHANISMS

Companies and governments alike can make commitments to implement core principles of freedom of expression and privacy. In today's globally networked digital environment, these principles should be implemented in a manner that is accountable locally as well as globally. Another approach to accountability for companies is through assessment and certification by independent multistakeholder organizations. The GNI, a multistakeholder coalition, requires its members to undergo periodic assessments as part of an accountability mechanism for adherence to its principles and implementation guidelines focused on how companies handle government requests. The GNI's implementation guidelines and assessment do not currently include consumer privacy issues or terms of service enforcement, however. Other organizations and mechanisms may need to be developed to improve accountability and transparency in these areas if GNI is unable to include them in the future.

As for states, a coalition of 27 governments have joined the Freedom Online Coalition, in which member nations agree to work together to advance 'free expression, association, assembly, and privacy online worldwide'. In April 2014, the coalition's members issued the Tallinn Declaration, a set of 'Recommendations for Freedom Online.' Three multi-stakeholder working groups have been set up. The coalition holds an annual conference to which representatives from companies and civil society are invited. However it remains to be seen whether any mechanisms will emerge through which governments can be benchmarked and held accountable by global stakeholders on the extent to which they have lived up to these recommendations. Internet intermediaries will be hard pressed to fully live up to their responsibility to respect human rights unless governments fulfil their own duty to protect human rights, including freedom of expression and privacy online.

9. CONCLUSION

This chapter has been focused on the role of three types of internet intermediaries in fostering freedom of expression, with attention to the normative, legal and policy contexts in which they operate. The research is not intended to be a representative or static sample of actors, but rather to extrapolate more general insights. Various trends have been identified, with an overall increase in awareness and actions by intermediaries themselves and governments about the significance that ISPs, search engines and social networks have for freedom of expression.

The analysis above has sought to assist all stakeholders, and not least the intermediaries themselves, to identify how the gatekeeping capacity inherent in mediating internet content can be optimised for freedom of expression, as well as the right to privacy. In this way, internet intermediaries can contribute to the evolution of knowledge societies, which in turn are central to building democracy, sustainable development and peace around the world.

VI. SAFETY OF JOURNALISTS

1. OVERVIEW

This chapter examines recent trends in the safety of journalists, presenting UNESCO statistics for 2013 and 2014, and tracking other developments up to August 2015. It follows the framework of the previous UNESCO report *World Trends in Freedom of Expression and Media Development*, mandated by Member States in Resolution 53 of UNESCO's 36th General Conference, which covered the earlier period of 2007 through mid-2013, including physical safety, impunity, imprisonment of journalists, and a gender dimension of the issues.[7] Additionally, the current chapter examines the recent trends in the strengthening of normative international standards, development of practical mechanisms, improvement in inter-agency cooperation, greater collaboration with the judiciary system and security forces, and research.

The current chapter also notes that the rate of recorded killings of journalists peaked in 2012 when UNESCO registered 123 cases and there has been a small decrease in the two subsequent years. Nevertheless, the number of killed journalists remains very high. Over the period, a low proportion of Member States, where killings of journalists have taken place, have provided a response on status of judicial inquiry of the cases. From the data that has been received, it emerges that the preceding rate of impunity has remained high. At the same time, there has been much increased attention and collaborative efforts to safety of journalists and impunity at the international level, as well as in certain countries.

[7] See UNESCO. 2015. World Trends in Freedom of Expression and Media Development. Paris: UNESCO. http://unesdoc.unesco.org/images/0022/002270/227025e.pdf and Resolution 53, adopted at the 36th Session of UNESCO's General Conference in November 2011. Available at http://unesdoc.unesco.org/images/0021/002150/215084e.pdf

2. PHYSICAL SAFETY

UNESCO continues as the UN agency with a specific mandate to defend press freedom and freedom of expression, and which raises awareness about the killings of journalists, media workers and social media producers who are victimized as a result of doing journalism.[8] Ending impunity for crimes against journalists has remained an important part of this work during 2013 - 2014. On this basis, UNESCO's Director-General, through mandate of the Organization's Intergovernmental Council for the International Programme for the Development of Communication (IPDC), has continued to condemn each verified killing during the period reviewed. She has further continued to request the Member State affected to voluntarily submit information on judicial follow up. Since a decision of the IPDC in 2012,[9] States which respond may indicate whether they wish their response to be placed on UNESCO's dedicated webpage[10] which records the killings and the Director-General's statement.

Specifically, in 2013 and 2014, UNESCO's Director-General publicly condemned the killings of a total of 178 journalists, media workers and social media producers engaged in journalistic activities.

In 2013, the number amounted to 91 deaths, diminishing by a quarter compared to 2012. However, this figure still represents the second highest number of journalists killed since 2006. After several years of relative calm in Iraq, the number of journalists killed there rose to 15 in 2013, making it the most dangerous country for journalists for that year. However, by comparison, the highest and second highest recorded number of killed journalists in Iraq were 33 deaths in 2007 and 29 deaths in 2006.

In 2014, the Director-General issued public statements on 87 cases of killings of journalists. The ongoing armed conflict in Syria has continued to inflict a high toll on journalists' lives with ten journalists killed in 2014. Within the same year, in other areas,[11] eight journalists were killed in Palestine, six in Iraq, five in Libya, and five in Afghanistan. Seven journalists were killed in Ukraine.

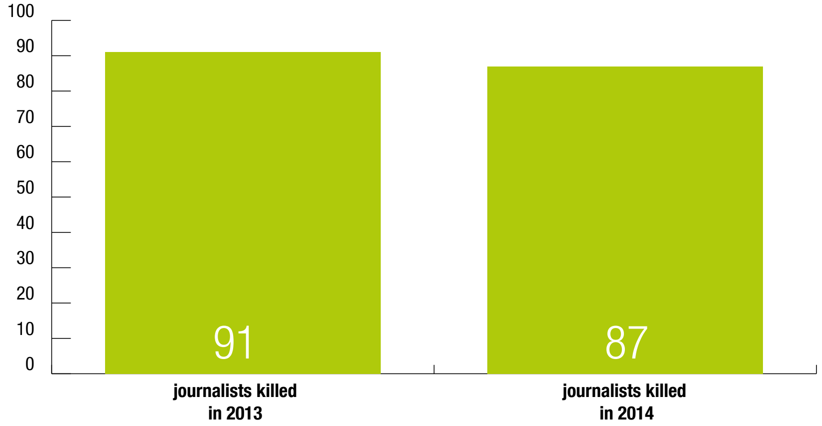
[8] See Decision 196 EX/31 on Safety of Journalists and the Issue of Impunity adopted at the 196th Session of UNESCO Executive Board. Available at http://unesdoc.unesco.org/images/0023/002323/232337e.pdf

[9] The 28th session of the IPDC Council requests the Director General 'to make available on UNESCO's website, upon request of the Member States concerned, information officially provided for the killings of journalists condemned by the Organization.'

[10] See the dedicated website UNESCO Condemns Killing of Journalists at www.unesco.org/new/en/condemnation

[11] These areas are identified in the 2013 UN Secretary-General's Report on the Protection of Civilians in Armed Conflicts, which is submitted to the UN Security Council every 18 months.

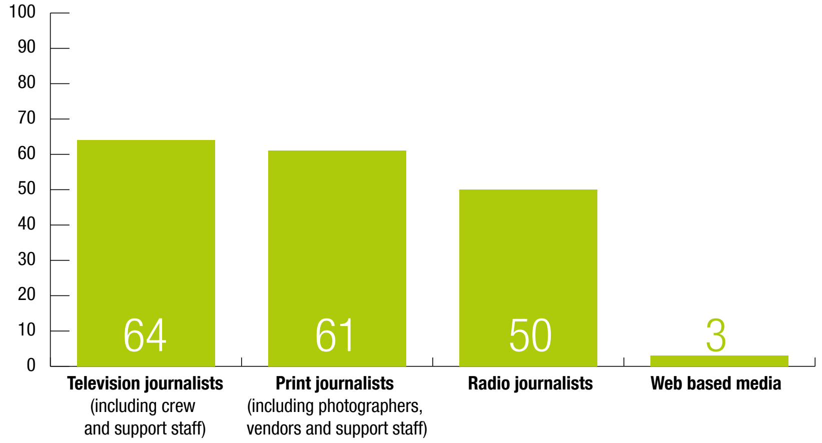
Total number of journalists killed in 2013 and 2014



As in previous years, the vast majority of killed journalists were locally based. In 2013, seven out of 91 (8 per cent) journalists killed were foreign correspondents. In 2014, the number of foreign correspondents killed increased sharply to represent nearly 20 per cent of the death toll (17 cases out of 87). Twelve of these cases took place in Syria and Ukraine.

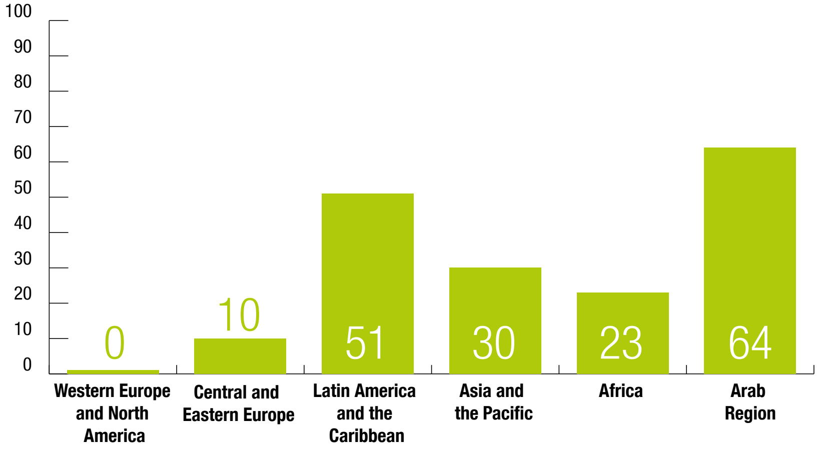
In terms of type of media, television journalists (including crew and support staff) suffered the most loss with 64 deaths in the period of 2013 and 2014. This is followed by print journalists (including photographers, vendors and support staff) with 61 deaths. Killed radio journalists numbered 50. Three journalists who worked predominantly for web based media were killed in the same period. Taken together, the 'traditional media' accounted for over 98 per cent of the loss of lives of the people who engaged in journalistic activities.

Journalists killed by type of media during the period 2013-2014



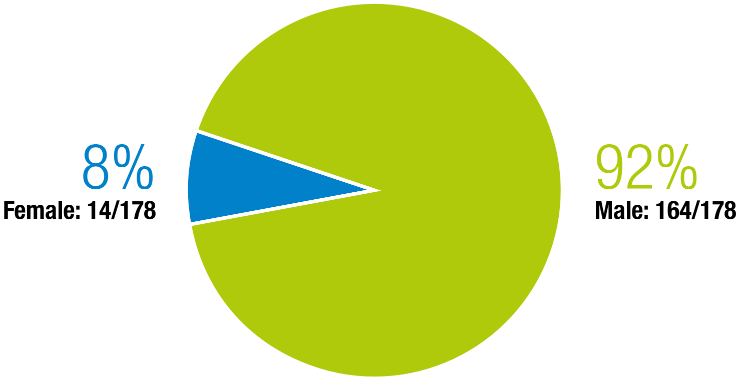
Broken down by region, a total of 64 killings of journalists (36 per cent) took place in the Arab States region, making it the most dangerous region for journalists to work in 2013 and 2014. A total of ten cases of killings of journalists took place in the Central and Eastern Europe region, 23 cases took place in the African region, 30 cases occurred in the Asia and the Pacific region, and 51 cases took place in the Latin America and the Caribbean region. No cases were recorded in the Western Europe and North America region during the two-year period under review.[12]

Total number of journalists killed per region during the period 2013-2014



In the period (2013 and 2014), male journalists continued to form the vast majority of killed journalists. A total of 164 out of 178 journalists killed were men (92 per cent).

Number of female/male journalists killed during the period 2013-2014



[12] The attacks on the French magazine Charlie Hebdo occurred just after this period.

Digital safety for journalists, which may also lead to physical danger to journalists and their sources, became more of an issue during the period. A number of media institutions experienced attacks on their websites, intrusions in their electronic communications, and the seizure of digital devices.[13]

[13] These are referenced in the UNESCO 2015 publications *Building Digital Safety for Journalism: A Survey of Selected Issues, Keystones to foster inclusive Knowledge Societies*, and in research conducted for UNESCO by the World Association of Newspapers and News Publishers (WAN-IFRA) into the protection of confidentiality of sources in the digital age.

3. IMPUNITY

A request for updates in the investigation and judicial inquiry of cumulative unresolved killings of journalists condemned by UNESCO has continued to be sent each year to Member States where the killings have taken place. From the data received, it appears that impunity has remained the predominant trend, with few perpetrators of the killings brought to justice.

Impunity refers to the effect of exemption from punishment of those who commit a crime. It thus points to a potential failure of judicial systems as well as the creation of an environment in which crimes against freedom of expression go unpunished. These features have continued to feed a vicious cycle and pose a serious threat to freedom of expression. The practice and expectation of impunity in regard to journalists' cases has implications for impunity more broadly. Journalists who work without fear help to ensure that other violations of rights cannot be hidden under a cloak of darkness. When crimes against journalists continue without punishment, this can encourage violations of numerous human rights besides freedom of expression and press freedom, as well as other forms of criminality. Elimination of the actors, along with arbitrary arrests and detention, enforced disappearance, harassment and intimidation have continued to be tactics that not only silence journalism, but also intimidate a population into self-censorship.

In June 2012, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression[14] attributed the root causes of impunity to the lack of political will to pursue investigations, exacerbated by the fear of reprisal at the hands of powerful criminal networks, inadequacies within the legal framework, judicial system, police forces and lack of resources, as well as negligence and corruption.

The most recent biennial Director-General's IPDC Report on Safety of Journalists and the Danger of Impunity, released in 2014, recorded that fewer than one in ten killings of journalists have led to a conviction. [15] The Report continued to urge Member States 'to inform the Director-General of UNESCO, on a voluntary basis, of the actions taken to prevent the impunity of the perpetrators and to notify her/him of the status of the judicial inquiries conducted on each of the killings condemned by UNESCO'.

[14] The report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/20/17) presented to the 20th Session of Human Rights Council

[15] This Report is produced in accordance with the decisions adopted by the Intergovernmental Council of UNESCO's IPDC, at its 26th, 27th, 28th, and 29th Sessions in 2008, 2010, 2012, and 2014 respectively.

The rate of response from Member States remains low, similar to earlier trends.[16] In 2013, 17 out 57 countries[17] (30 per cent) where killings of journalists had taken place and had not been resolved, responded to the formal request for information. In 2014, 13 out of 59 countries[18] (22 per cent) responded to the official request. As of 31 August 2015, 24 out of 57 countries[19] (42 per cent) responded to the latest request for information, demonstrating the possible start of an upward trend.

The responses received in 2015 cover 46% of 641 unresolved cases for the period 1 January 2006 up to 31 December 2014. This is an increase in the extent of information as compared to the previous period. For 2006-2013 inclusive, information was received in 22% of unresolved cases. However, despite the wider coverage, it is still the case that no information was received in more than half the cases.

Within the information which UNESCO did receive from Member States, the proportion of cumulative cases which are reported as being judicially resolved was 5% in 2012, rising to 8% in 2014. While there is a small increase in the percentage, and while many cases are reported as still ongoing, it is evident that impunity continues as the predominant trend. It can be extrapolated that these percentages also apply to the cases where no information was received by UNESCO, meaning that the proportion of resolved cases across the board can be gauged as continuing to be extremely low.

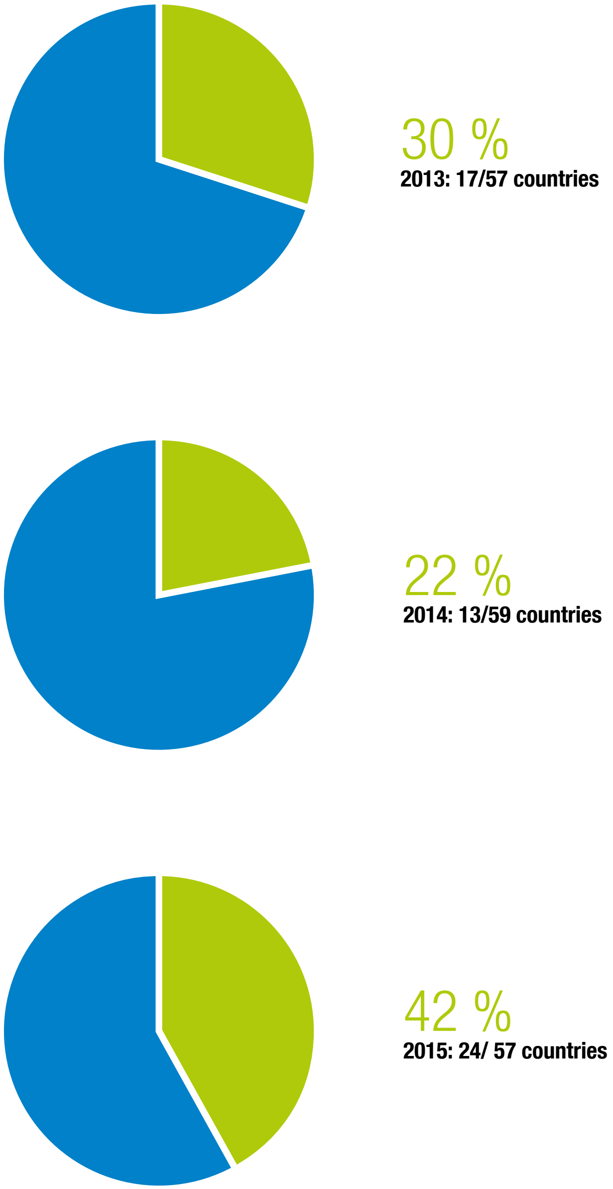
[16] In 2011, an official request for updated information was sent out to 38 countries where killings of journalists took place with 19 out of 38 countries responding over the period of 2011-2012, a figure of 50%. Considered over the longer period of killings occurred between 2007 and 2012, as reported in the 2014 UNESCO World Trends in Freedom of Expression and Media Development, 42 per cent of the Member States had provided a response by mid-2013.

[17] In 2013, 17 countries responded to the official request: Bahrain, Bolivia, Brazil, Colombia, Republic of Congo, Croatia, Democratic Republic of Congo, Honduras, Kazakhstan, Kenya, Peru, Russian Federation, Sri Lanka, Tanzania, Tunisia, Turkmenistan, and Vietnam. In the same year, 40 countries did not respond: China, Dominican Republic, El Salvador, Indonesia, Iraq, Mexico, Pakistan, Philippines, Turkey, Afghanistan, Angola, Bangladesh, Bulgaria, Cambodia, Cameroon, Ecuador, Egypt, Eritrea, Georgia, Greece, Guatemala, Republic of Guyana, Haiti, India, Iran, Kyrgyzstan, Lebanon, Libya, Myanmar, Nepal, Nigeria, Palestine, Rwanda, Somalia, Sudan, Syria, Thailand, Uganda, Venezuela, and Yemen.

[18] In 2014, 13 countries responded the official request: Colombia, Honduras, Peru, Tanzania, China, Dominican Republic, El Salvador, Indonesia, Iraq, Mexico, Pakistan, Philippines, and Turkey. In the same year, 46 countries did not respond to the request including Afghanistan, Angola, Bahrain, Bangladesh, Bolivia, Brazil, Bulgaria, Cambodia, Cameroon, Central Africa Republic, Republic of Congo, Croatia, Democratic Republic of Congo, Ecuador, Egypt, Eritrea, Georgia, Greece, Guatemala, Republic of Guyana, Haiti, India, Iran, Kenya, Kyrgyzstan, Lebanon, Libya, Mali, Myanmar, Nepal, Nigeria, Palestine, Paraguay, Russian Federation, Rwanda, Somalia, South Sudan, Sri Lanka, Sudan, Syria, Thailand, Tunisia, Turkmenistan, Uganda, Venezuela, and Yemen.

[19] By 1 September 2015, 24 countries responded to the official request: Bahrain, Brazil, Bulgaria, Colombia, Dominican Republic, Ecuador, Egypt, El Salvador, Eritrea, Greece, Guatemala, Haiti, Honduras, Indonesia, Mexico, Nigeria, Pakistan, Paraguay, Philippines, Sri Lanka, Tanzania, Turkey, Ukraine, and Venezuela. In the same year, 33 countries did not respond: Afghanistan, Angola, Bangladesh, Cambodia, Cameroon, Central Africa Republic, Republic of Congo, Democratic Republic of Congo, Georgia, Guinea, Republic of Guyana, India, Iran, Iraq, Kenya, Kyrgyzstan, Lebanon, Libya, Mali, Myanmar, Nepal, Palestine, Peru, Russian Federation, Rwanda, Somalia, South Sudan, Sudan, Syria, Thailand, Tunisia, Uganda, and Yemen.

Member State response rate to the Director-General’s requests on the status of judicial inquiry of killings of journalists (2013, 2014, up to 31 August 2015)



4. UPWARD TREND IN STRENGTHENING OF INTERNATIONAL STANDARDS ON SAFETY OF JOURNALISTS

While there has been no major shift in the safety trends described above, compared to the preceding period, there has in contrast been major progress at the normative level. International standards on the safety of journalists have been strengthened significantly in the past two years. This trend was reinforced through global reactions to the killings of journalists at the satirical magazine *Charlie Hebdo* in Paris, France in early 2015, following the brutal beheading of journalists in Syria. While that attack post-dates the period under review, it is signalled in this chapter because it occurred in a context of increasing attention to the issue internationally, and resulted in even higher international attention to the issue, including a protest march by world leaders. As with the cumulative impact of images of beheadings of journalists by extremists, and particularly violent killings perpetrated by drug-dealers against reporters over 2013-2014, the 2015 attacks in Paris have seen the world become increasingly sensitized to the seriousness of such crimes.

One indicator of the trend of increased awareness is activity at the UN level. As elaborated below, over 2012-2015, the United Nations Security Council, the General Assembly, the Human Rights Council, and UNESCO all adopted significant resolutions and decisions which unequivocally condemned all attacks and violence against journalists. Several of these statements included steps to further strengthen global monitoring and reporting mechanisms about safety, and also underlined the importance for practical measures to be taken by Member States to end impunity.

The United Nations General Assembly, the highest decision-making body of the UN system, adopted Resolutions A/RES/68/163 (in 2013) and A/RES/69/185 (in 2014) which strongly condemn all attacks on journalists and media workers, including torture, extrajudicial killings, enforced disappearances and arbitrary detention, and intimidation and harassment in both conflict and non-conflict situations. The Resolutions also strongly criticized the prevailing impunity for attacks and violence against journalists.

Furthermore, through Resolution A/RES/68/163, the UN General Assembly proclaimed 2 November to the International Day to End Impunity for Crimes against Journalists marking an important milestone in global recognition of the issues. UNESCO, which was charged to facilitate commemorations of the international day, led the inaugural round with series of events including a conference at the European Court of Human Rights in Strasbourg, France, together with the Council of Europe, the Centre for Freedom of the Media at the University of Sheffield, and the European Lawyer's Union. Elsewhere, localized events took place in New York, Tunis, Accra, and Abuja. Through these events, UNESCO sought to reach out to judicial actors, sensitizing them to the role they can play in ending impunity, and how attention to resolving the cases of attacks on journalists can contribute more broadly to strengthening the rule of law and human rights in society at large. The Strasbourg event also saw a multi-stakeholder assessment of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity. The Plan was endorsed by the UN Chief Executives Board in 2012, and was noted with appreciation by the UN General Assembly in December 2013 in Resolution A/RES/68/163.

At the Human Rights Council, the 2012 landmark Resolution A/HRC/RES/21/12 was adopted in 2012, followed in 2014 by Resolution A/RES/HRC/27/5, both covering the safety of journalists. These Resolutions called on all parties to respect their obligations under international human rights law and international humanitarian law as well as on States to promote a safe and enabling environment for journalists to perform their work independently and without undue interference.

At its 191st Session in April 2013, UNESCO's Executive Board endorsed the UNESCO Work Plan for addressing the safety of journalists and impunity of crimes committed against them. The Work Plan, with an emphasis on South-South cooperation, set out UNESCO's approach to safety, including leadership of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity. Subsequently, UNESCO's Executive Board adopted the Decision on Safety of Journalists and the Issue of Impunity at its 196th Session on 20 April 2015. This Decision reinforced the current work of UNESCO in relation to the UN Plan of Action through a multi-stakeholder approach involving all relevant actors including national authorities, UN agencies, civil society groups, academia, and the media. The Decision further confirmed that the mission of securing safety for journalism includes the safety of social media producers who generate significant amount of public interest journalism.

Additionally, the UN Security Council also adopted Resolution 2222 (on 27 May 2015) which called on parties to a conflict and all Member States to create a safe environment in law and practice for journalists to do their work. It also called for the UN Secretary-General to include consistently as a sub-item in the regular 'Report on the Protection of Civilians in Armed Conflict', the issue of the safety and security of journalists, media professionals and associated personnel.

A sign of growing recognition of the importance of the issues is reflecting by the growing number of signatories to these resolutions.

- UNGA Resolution A/RES/68/163 adopted in 2013: 54 signatories[20]

[20] The following 54 countries co-sponsored Resolution A/RES/68/163: Albania, Andorra, Argentina, Armenia, Australia, Austria, Azerbaijan, Belgium, Benin, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, Chile, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, El Salvador, Estonia, France, Germany, Ghana, Greece, Hungary, Ireland, Italy, Japan, Latvia, Luxembourg, Maldives, Mali, Malta, Mongolia, Morocco, Netherlands, Nigeria, Panama, Paraguay, Peru, Poland, Portugal, Qatar, Republic of Korea, Romania, San Marino, Serbia, Slovakia, Slovenia, Spain, Tunisia, Turkey, United States of America, and Uruguay.

- UNGA Resolution A/RES/69/185 adopted in 2014: 82 signatories[21]
- Human Rights Council Resolution A/HRC/21/12 adopted in 2012: 52 signatories[22]
- Human Rights Council Resolution A/HRC/27/5 adopted in 2014: 63 signatories[23]

The UNESCO Executive Board 196 EX/31 Decision adopted in April 2015 also received a high number of co-signatures from 47 countries.[24] Similarly, the United Nations Security Council adopted the Resolution UNSC 2222 (on 27 May 2015) with co-signatures from 49 countries.[25]

[21] The following 82 countries co-sponsored Resolution A/RES/69/185: Andorra, Argentina, Armenia, Australia, Austria, Azerbaijan, Belgium, Benin, Bosnia and Herzegovina, Brazil, Bulgaria, Burkina Faso, Cabo Verde, Central African Republic, Chile, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Egypt, El Salvador, Estonia, Finland, France, Georgia, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, Iceland, Ireland, Israel, Italy, Japan, Jordan, Latvia, Lebanon, Libya, Liechtenstein, Lithuania, Luxembourg, Maldives, Mali, Malta, Mexico, Monaco, Mongolia, Montenegro, Morocco, the Republic of Moldova, Netherlands, New Zealand, Norway, Panama, Paraguay, Peru, Poland, Portugal, Qatar, Republic of Korea, Romania, San Marino, Serbia, Slovakia, Slovenia, Somalia, Spain, Sweden, Switzerland, the Central African Republic, the former Yugoslav Republic of Macedonia, Tunisia, Turkey, Ukraine, United Kingdom of Great Britain and Northern Ireland, United States of America, and Uruguay.

[22] The following 52 countries co-sponsored Resolution A/HRC/21/12: Albania, Argentina, Australia, Austria, Belgium, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Colombia, Croatia, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Finland, Georgia, Germany, Greece, Guatemala, Honduras, Hungary, Iceland, Ireland, Kenya, Latvia, Lebanon, Libya, Liechtenstein, Lithuania, Luxembourg, Mexico, Montenegro, Morocco, Netherlands, Nigeria, Norway, Palestine, Peru, Poland, Portugal, Qatar, Republic of Moldova, Romania, Serbia, Slovenia, Sweden, Switzerland, Tunisia, Turkey, and United Kingdom of Great Britain and Northern Ireland.

[23] The following 63 countries co-sponsored Resolution A/HRC/27/5: Argentina, Australia, Austria, Belgium, Benin, Bosnia and Herzegovina, Brazil, Bulgaria, Burkina Faso, Canada, Central African Republic, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Guatemala, Honduras, Hungary, Iceland, Ireland, Italy, Latvia, Lebanon, Liechtenstein, Lithuania, Luxembourg, Maldives, Mexico, Montenegro, Morocco, Netherlands, New Zealand, Nigeria, Norway, Paraguay, Peru, Poland, Portugal, Qatar, Republic of Moldova, Romania, Saint Kitts and Nevis, Serbia, Slovakia, Slovenia, Spain, Palestine, Sweden, Switzerland, The former Yugoslav Republic of Macedonia, Tunisia, Turkey, United Kingdom of Great Britain and Northern Ireland, United States of America, and Yemen.

[24] The following 47 countries co-sponsored the UNESCO Executive Board 196 EX/31 Decision: Albania, Andorra, Argentina, Australia, Austria, Brazil, Cyprus, Czech Republic, Denmark, Dominican Republic, El Salvador, Estonia, Finland, France, Gabon, Germany, Greece, Honduras, Iceland, Ireland, Italy, Japan, Latvia, Liberia, Malawi, Morocco, Namibia, Netherlands, Nigeria, Norway, Paraguay, Peru, Portugal, Republic of Korea, Serbia, Slovakia, Slovenia, Spain, St Kitts and Nevis, Sweden, Switzerland, Trinidad and Tobago, Tunisia, Ukraine, United Kingdom of Great Britain and Northern Ireland, United States of America, and Uruguay.

[25] The following 49 countries co-sponsored the UN Security Council Resolution UNSC 2222: Albania, Angola, Australia, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Canada, Chad, Chile, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Israel, Italy, Japan, Jordan, Latvia, Lebanon, Liechtenstein, Lithuania, Luxembourg, Malaysia, Montenegro, Netherlands, New Zealand, Nigeria, Norway, Palau, Poland, Republic of Moldova, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, The former Yugoslav Republic of Macedonia, Ukraine, United Kingdom of Great Britain and Northern Ireland, and the United States of America.

5. DEVELOPMENT OF PRACTICAL MECHANISMS TO PROMOTE SAFETY AND END IMPUNITY

There has also been progress in institutional developments related to safety and impunity over 2012 and 2013. Several countries in the Latin American region have continued to develop official frameworks and institutions to deal with safety and protection, many drawing on the positive experience of Colombia. These mechanisms range from interdepartmental co-ordination systems, multi-stakeholder fora involving media and civil society representatives, and dedicated personnel and budgets. In Pakistan, a broad coalition has worked to involve many stakeholders, including government and parliamentarians, in regular discussions on safety and impunity. In Serbia, a commission of representatives of independent media, a ministry and the security services, secured the prosecution of four people for the killing of a journalist 16 years earlier.

In 2013, the Office of the High Commissioner for Human Rights (OHCHR), in collaboration with the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, published a report highlighting initiatives and good practices relating to the safety of journalists and ending impunity. The report contains an overview of the situation facing journalists, applicable law and initiatives taken by Member States, United Nations agencies and other organizations for the safety of journalists. It also identifies good practices that could assist in creating a safe and enabling environment in which journalists are able to exercise freely their profession.

On 2 April 2015, the Council of Europe launched an Internet platform aimed at protecting journalism and promoting the safety of journalists. The platform is designed to facilitate the compilation, processing and dissemination of factual information, verified by the partners, concerning serious physical threats to journalists and other media personnel, threats to the confidentiality of media sources and forms of political or judicial intimidation. The platform involves a partnership by the Council of Europe with Article 19, the Association of European Journalists, the European Federation of Journalists, the International Federation of Journalists and Reporters Without Borders.

The global trend where more and more journalism takes places through digital means is also reflected in the increased number of journalists' trainings and tools that focus on digital, especially mobile, security. This include the development of mobile phone applications that aim to empower individual journalists to better protect themselves. The International Media Women's Foundation (IWMF) has developed such an application, *Reporta,* which is designed with 'check-in', 'alerts' and 'SOS' functions. Similarly, the International Center for Journalists (ICFJ) is developing *Salama* which is a risk-assessment application.

6. IMPROVED INTER-AGENCY COLLABORATIONS

Over the 2013-14 period, there has been greater co-operation amongst UN bodies on the subject of safety. OHCHR and UNESCO contributed to the report of the UN Secretary-General's Report on the implementation Resolution A/RES/68/163 on Safety of Journalists and the Issue of Impunity. The Report, which was submitted to the UN General Assembly, provided an overview of recent trends with regard to the safety of journalists and media workers, as well as a compilation of initiatives undertaken to ensure their protection, along with recommendations.

UNODC published the *Global Study on Homicide* in 2013 which looked at gives a comprehensive overview of intentional homicide across the world. A sub-section focusing on the killings of journalists was compiled with input from UNESCO.

The United Nations Department for Public Information (UNDPI) has transmitted the information concerning the development of the UN Plan of Action on the Safety of Journalists and the Issue of Impunity to its 63 UN Information Centre (UNIC) around the globe. UN Women and UNESCO have collaborated during the period on issues pertaining to women journalists. The issue of safety of journalists has also increasingly been incorporated into the national United Nations Development Assistance Framework (UNDAF) including in Jordan, Nepal and South Sudan.

Further co-operation took place between OHCHR, ILO and UNESCO, and with the Global Forum for Media Development, in the development of draft indicators for the Sustainable Development Goals (SDGs). The SDG target 16.10 is to 'Ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements'. Discussions amongst the above cited groups reached consensus on a safety-related proposed indicator for this particular target. The proposed indicator, which may be adopted early 2016, is: 'Number of verified cases of killing, kidnapping, enforced disappearance, arbitrary detention and torture of journalists, associated media personnel, trade unionists and human rights advocates in the previous 12 months.' It is envisaged that these indicators could further mainstream the understanding that safety of journalists is a fundamental freedom in its own right as well as a target for sustainable development – and an enabler that contributes to the achievement of other SDGs.

7. TOWARDS STRONGER INVOLVEMENT OF THE JUDICIAL SECTOR IN TACKLING IMPUNITY

In the past two years, the trend towards greater engagement with the judiciary system in fighting impunity has increased, including capacity-building efforts aimed at judges and lawyers. The conference on impunity held in the European Court of Human Rights in November 2014 has already been mentioned. Also in 2014, UNESCO and Knight Center for Journalism in the Americas of the University of Texas at Austin collaborated with the former UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression and the former Special Rapporteur of the Organization of American States on delivering a massive open online course (MOOC) on freedom of expression, including safety of journalists. The course, which saw more than 800 legal actors taking part in the course within the space of one month was initially created for the Supreme Court of Mexico, and has since attracted the interests of other judicial actors elsewhere in the Latin America region. This is the result of a seed grant by UNESCO's IPDC in 2013. In 2015, the MOOC is again being offered with the support from UNESCO and the government of the state of Coahuila of Mexico.

In a number of rulings, such as that in 2013 by the African Court on Human and Peoples' Rights which ordered a reopening of the investigation into the 1998 murder of Burkinabe journalist Norbert Zongo and three others, there is increasing recognition of the importance of the effective application of the rule of law. Similar jurisprudence characterized the ruling in 2014 by the Economic Community of West African States Court of Justice, regarding the case of Gambian journalist Deyda Hydara who was killed in 2004.[26]

[26] These are recent examples. An earlier case was the 2009 judgment by the Inter-American Court of Human Rights in the case of Ríos et al. v. Venezuela which ruled that extra-judicial killings required investigations that needed to be timeous as well as conducted in a serious, fair, and effective manner.

8. STRENGTHENING COLLABORATION WITH NATIONAL SECURITY FORCES

A crucial element in the ensuring the safety of journalists is interaction with the security forces. This is especially vital during times of intensified tension and pressure, such as elections or during a street protest. UNESCO began promoting this area of capacity-building in 2013 with a series of training courses in Tunisia in collaboration with the Ministry of the Interior, and the support of the Netherlands and the Swedish International Development Cooperation Agency (Sida). The series of training sessions became the basis of UNESCO's new manual titled *Freedom of Expression and Public Order*. In 2015, similar trainings took place in Mogadishu, Somalia, in cooperation with Relief International and the United Nations Assistance Mission in Somalia (UNSOM). This emerging trend of deepening engagement between security forces and media professionals could increase both public order and freedom of expression.

9. PROMOTING A RESEARCH AGENDA ON THE SAFETY OF JOURNALISTS

Safety was further reinforced by growing insights during the period. In the review of the UN Plan of Action in Strasbourg, in November 2014, knowledge was deepened through discussion of how the contexts in which journalists are killed, including in terms of political will and capacity, call for different types of support, such as knowledge-sharing, capacity-building, sensitising stakeholders, raising awareness, safety training for journalists, and building up documentation on attacks so that justice could be exercised at a future point.

Reinforcing such nuanced insights as developed over the period under study, there has been an increase in knowledge about the causes, impacts and remedies in relation to safety and impunity. One instance of this was the piloting of UNESCO's Journalism Safety Indicators (JSI) in Pakistan, Honduras, Guatemala, and Liberia, as well as the initiation of JSI full studies in Nepal, Iraq and Kenya.

Another development has been growing interest and action in regard to academia and others who do research. The 2012 implementation strategy of the UN Plan of Action identified a significant opportunity in the area of scientific academic research on safety of journalists and the impunity. This recognition was based on the reality that a general mapping of academic research conducted in the past 20 years had yielded a relatively a small number published studies. Of these available studies, most focused on 'war reporting' or protection of journalists in armed conflict situations, although more than half of all attacks on journalists happened in non-armed conflict situations.

In an effort to develop a trend of more research in the area, UNESCO in 2014 developed a ten-point research agenda and promoted this in 2015 at special sessions in July 2015 on safety of journalists during the International Association of Media and Communication Research (IAMCR) Conference in Montreal and the Global Communication Association Conference in Berlin. More than 100 researchers participated in these events. A number of universities have expressed parallel interest in collaborating with UNESCO in research on safety.

10. IMPRISONMENT OF JOURNALISTS

As noted in the earlier UNESCO *World Trends in Freedom of Expression and Media Development*, imprisonment of journalists for their legitimate work fosters a culture of self-censorship and impinges on the broader rights of society to obtain information. It added: 'Incarceration for legitimate journalism work is unnecessary and disproportionate in terms of international standards of justifiable limitations and sanctions concerning the exercise of freedom of expression.' Given its mandate, UNESCO does not systematically collect or track data related imprisonment of journalists.[27] However, based on a wide range of sources and data, imprisonment rate of journalists worldwide is reported to have remained high in 2013 and 2014. Between 178 and 211 journalists were reported to be imprisoned in 2013 while at least 221 journalists were reportedly imprisoned in 2014.[28] This compares to at least 232 journalists reported to be imprisoned in 2012 and 179 reportedly imprisoned in 2011. Enforced or involuntary disappearance of journalists has continued to be reported as an issue. In these cases of missing journalists, who often report on criminal activities and corruption prior to their disappearance, the UN Working Group on Enforced or Involuntary Disappearances and the UN Working Group on Arbitrary Detention have been turned to by actors who seek redress or the re-opening of investigations.

[27] As indicated in the first report, many governments have maintained that that particular journalists have not been jailed for their journalism but for other reasons, and UNESCO is not mandated to assess which cases are in this category or for other reasons.

[28] Based on public data provided by major international press freedom organizations including Committee to Protect Journalists (CPJ) and Reporters without Borders (RSF).

11. GENDER PERSPECTIVE IN SAFETY OF JOURNALISTS

While women journalists account for less than eight percent of the journalists (or 14 out of 178) killed in the two-year period of 2013 and 2014, there is small increase in the absolute number of women journalists killed.[29] Furthermore, women journalists have continued to be subjected to other forms of harassments and attacks.

In the past two years, UNESCO has been supporting more research and raising more awareness in this area of safety of journalists. In March 2014, UNESCO, in collaboration with the International News Safety Institute, the International Women's Media Foundation and the Austrian Government, launched the results of a survey, titled *Violence and Harassment against Women in the News Media: a Global Picture*, in which nearly 1,000 respondents took part. This study hopes to spark more interest in research the issue of safety specific to women journalists.

Further, in 2015, UNESCO included a special gender focus within its new publication *Building Digital Safety for Journalism: A Survey of Selected Issues*. The survey, which focuses on digital threats to journalists, pointed out that women are more likely than men to face negative and indeed threatening responses online. Women journalists in particular face 'a double attack' where they are targeted both as journalists and as women.

As part of wider awareness-raising in these areas, UNESCO in the period under review regularly included a strong gender perspective in its flagship awareness raising events such as the World Press Freedom Day celebrations. In the past two years, this international event has included dedicated sessions concerning the safety of women journalists including training workshops, and various other gender related issues in the media. In 2015, responding to the 20th anniversary of the Beijing Declaration and Platform for Action, UNESCO organized three dedicated sessions on the issue of gender and media during the World Press Freedom Day celebration in Riga, Latvia.

To further highlight both the importance of safety of journalists across the world, and especially the role of women journalists, UNESCO's Director-General in 2015 appointed CNN's chief international correspondent, Christiane Amanpour, as UNESCO Goodwill Ambassador for Freedom of Expression and Journalist Safety in April 2015 in the lead up to the year's World Press Freedom Day.

[29] Six women journalists were killed in 2013 and eight women journalists were killed in 2014.

12. CONCLUSION

This chapter has reviewed trends in the safety of journalists and the issue of impunity, with statistics collected during 2013 and 2014, and with reference to some developments in 2012 and 2015. While attacks on journalists and the issue of impunity have continued to be a serious problem, there is progress evident in a number of other areas. These include the large number of UN Member States associating with UN Resolutions, and an improved response rate to UNESCO enquiries for 2014 as compared to 2013. Other improvements have been listed, covering awareness, institution and capacity-building, and knowledge generation. It is not easy to gauge whether all this heightened action has played any part in preventing the statistics from being even more serious than they are. Nor is it easy to measure whether impact will continue over a longer period of time. What is clear, however, is that there is growing global momentum towards establishing a culture in which the safety of journalists and ending impunity is guaranteed. It appears likely that this is a trend that will continue to develop as long as the problems persist, and to the extent that there is success, this will enhance the quest for peaceful knowledge societies and the achievement of the UN's Sustainable Development Goals.

VII. APPENDICES

APPENDIX 1 – INDIVIDUALS INTERVIEWED FOR COUNTERING ONLINE HATE SPEECH

Imran Awan, Deputy Director of the Centre for Applied Criminology, School of Social Sciences, Birmingham City University, United Kingdom of Great Britain and Northern Ireland

Monika Bickert, Head of Global Policy Management, Facebook, United States of America

Drew Boyd, Director of Operations, The Sentinel Project for Genocide Prevention, Canada

Ian Brown, Professor of Information Security and Privacy, Oxford Internet Institute, University of Oxford, United Kingdom of Great Britain and Northern Ireland

Laura Geraghty, No Hate Speech Movement, United Kingdom of Great Britain and Northern Ireland

Matthew Johnson, Director of Education, MediaSmarts, Canada

Myat Ko Ko, Myanmar Program Officer, Justice Base, Myanmar

Ciara Lyden, Manager Content Policy, Produce Policy, Facebook, Ireland

Andre Oboler, CEO, Online Hate Prevention Institute, Australia

Harry Myo Lin, Panzagar, Myanmar

Nanjira Sambuli, Project Lead, UMATI, Kenya

Christopher Wolf, Chair, National Committee on the Internet, Anti-Defamation League, United States of America

APPENDIX 2 – INDIVIDUALS INTERVIEWED FOR PROTECTING JOURNALISM SOURCES IN THE DIGITAL AGE

Rasha Abdulla, Associate Professor of Journalism and Mass Communication, The American University in Cairo, Egypt

Ricardo Aguilar, Investigative Journalist, *La Razón*, Bolivia

Rawda Ahmed, Lawyer, the Arabic Network for Human Rights Information, Egypt

Mahasen Al Eman, Director, Arab Women's Media Center, Jordan

Amare Aregawi, owner, Media and Communications Center and Horn of Africa Press Institute, Ethiopia

Hans-Gunnar Axberger, Professor of Constitutional Law, University of Uppsala, Sweden

Wendy Bacon, Professorial Fellow, Australian Centre for Independent Journalism, Australia

Martin Baron, Executive Editor, *The Washington Post*, United States of America

Peter Bartlett, Partner, Minter Ellison, Australia

Katarina Berglund-Siegbahn, Legal advisor, Ministry of Justice, Sweden

Catalina Botero Marino, former Special Rapporteur for Freedom of Expression, Inter-American Commission on Human Rights

Cliff Buddle, Senior Editor, *South China Morning Post*, China

Umar Cheema, Investigative Reporter, *The News*, and Founder, Center for Investigative Reporting in Pakistan

Zine Cherfaoui, Editor-in-Chief, *El Watan*, Algeria

Marites Dañguilan Vitug, Co-founder and Board Member, Philippines Center for Investigative Journalism, the Philippines

Yves Eudes, Reporter, *Le Monde*, and Co-founder, Source sûre, France

Tomaso Falchetta, Legal Officer, Privacy International, United Kingdom of Great Britain and Northern Ireland

Javier Gaza Ramos, Journalism Security & Safety Expert, Mexico

Carlos Guyot, Editor-in-Chief, *La Nación*, Argentina

Silvia Higuera, Knight Center for Journalism in the Americas, the University of Texas at Austin, United States of America

Daoud Kuttab, Journalist, Jordan

Fredrik Laurin, Director of the Investigative Unit, Swedish Public Radio (SR), Sweden

Ronaldo Lemos, Director, Institute for Technology & Society (ITS), Rio de Janeiro, and Professor, Law School, Rio de Janeiro State University, Brazil

Justine Limpitlaw, Electronic Communications Lawyer, South Africa

Henry Omusundi Maina, Director, ARTICLE 19 East & Horn of Africa, Kenya

Susan E. McGregor, Assistant Professor & Assistant Director, Tow Center for Digital Journalism, Columbia Journalism School, Columbia University, United States of America

Toby Mendel, Director, Centre for Law and Democracy, Canada

Gavin Millar QC, Lawyer, Matrix International, United Kingdom of Great Britain and Northern Ireland

Peter Noorlander, Media Lawyer, Media Legal Defence Initiative, United Kingdom of Great Britain and Northern Ireland

Gunnar Nygren, Professor, School of Social Sciences, Stockholm University, Sweden

Leanne O'Donnell, Senior Lawyer, Legal Policy, Law Institute of Victoria, Australia

Toyosi Ogunseye, Editor, *The Sunday Punch*, Nigeria

Julie Owono, Head of Africa Desk, Internet Sans Frontières, Franca

Courtney Radsch, Advocacy Director, Committee to Protect Journalists, United States of America

Marcelo Rech, Executive Director of Journalism, RBS Group, Brazil

Alan Rusbridger, Editor-in-Chief, *The Guardian*, United Kingdom of Great Britain and Northern Ireland

Gerard Ryle, Director, International Consortium of Investigative Journalists, United States of America

Rana Sabbagh, Executive Director, Arab Reporters for Investigative Journalism, Jordan

Josh Stearns, Director of Journalism and Sustainability, the Geraldine R. Dodge Foundation, United States of America

Atanas Tchobanov, Editor, Bivol.bg, and Journalist, BalkanLeaks, Bulgaria

Charles D. Tobin, Partner, Holland & Knight, United States of America

Pär Trehörning, Ombudsman, Swedish Union of Journalists, Sweden

Pedro Vaca Villarreal, Executive Director, Fundación para la Libertad de Prensa (FLIP), Colombia

Anita Vahlberg, Senior Advisor, Swedish Union of Journalists, Sweden

Dirk Voorhoof, Professor, Faculty of Political and Social Sciences and the Faculty of Law, Ghent University, Belgium

Wei Yongzheng, Lecturer, Communication University of China, Beijing, China

George Williams, Professor, Mason Professor, Scientia Professor, Foundation Director, Gilbert + Tobin Centre for Public Law, Faculty of Law, University of New South Wales, Australia

Jillian York, Director for International Freedom of Expression, Electronic Frontier Foundation, Germany

Yuan Zhen (pseudonym), Editor-in-Chief, (unnamed newspaper), China

APPENDIX 3: UNESCO MEMBER STATES STUDIED IN PROTECTING JOURNALISM SOURCES IN THE DIGITAL AGE[30]

Africa

Angola
Benin
Botswana
Burkina Faso
Burundi
Cameroon
Cape Verde
Chad
Côte D'Ivoire
Democratic Republic of the Congo
Ethiopia
Zambia
Gambia
Ghana
Kenya
Lesotho
Liberia
Malawi
Mali
Mauritius
Mozambique
Uganda
Niger
Nigeria
Rwanda
Senegal
Zimbabwe
South Africa
Swaziland
Somalia
Tanzania
Togo

Asia and the Pacific

Australia
Bangladesh
Cambodia
Bhutan
China
Timor-Leste
Fiji
India
Indonesia
Japan
Republic of Korea
Kiribati
Kyrgyzstan
Malaysia
New Zealand
Pakistan
Palau
Singapore
Sri Lanka
Philippines
Uzbekistan
Tajikistan
Turkmenistan
Vanuatu

Arab States

Algeria
Egypt
Mauritania
Morocco
Djibouti
Sudan
Syria

Europe and North America

Andorra
Armenia
Austria
Belarus
Belgium
Bosnia and Herzegovina
Bulgaria
Canada
Czech Republic
Denmark
Estonia
Finland
France
Georgia
Germany
Greece
Hungary
Iceland
Ireland
Israel
Italy
Latvia
Lithuania
Luxembourg
The former Yugoslav Republic of Macedonia
Monaco
The Netherlands
Norway
Poland
Portugal
Russian Federation
Slovakia
Spain
Sweden
Switzerland
Turkey
United Kingdom of Great Britain and Northern Ireland
United States of America

Latin America and the Caribbean

Argentina
Bolivia
Brazil
Chile
Columbia
Costa Rica
Dominican Republic
Ecuador
El Salvador
Guatemala
Guyana
Haiti
Honduras
Mexico
Paraguay
Panama
Peru
Uruguay
Venezuela
Nicaragua

[30] Member States selected based on the 2007 study by David Banisar, Silencing Sources: An International Survey of Protections and Threats to Journalists' Sources.

APPENDIX 4: PROTECTING JOURNALISM SOURCES IN THE DIGITAL AGE SURVEY QUESTIONS

1. What are the a) current and b) emerging challenges relevant to freedom of expression in a digital environment, as they apply to the practice of investigative journalism that relies on confidential sources?

2. What laws/legal instruments currently exist in your country, or region of operation, that are designed to protect journalists' sources?

3. What relevant laws/legal precedents/policies have been revoked, superseded or added since 2007 in your country or region of operation?

4. To what extent do these existing laws protect digitally interfaced journalism and journalistic sources?

5. How could/should legislation that impacts on the protection of journalists' sources be updated for the digital era?

   i) What changes to these laws are needed to better protect journalists and their sources exchanging information and publishing in the digital environment?

   ii) What changes are needed in your country/region specifically?

   iii) What changes might be better enacted through global policy instruments (e.g. UN)

6. Please identify one to three cases from your country, or region of operation, that highlight issues with journalistic source protection that you think warrant closer study. (Note: We are particularly interested in case studies which highlight the complexity of information exchange and publishing in a digital environment; the emergence of citizen journalists; the impact of national security legislation; the conflict between legislative protection of journalists' sources and other protections such as the right to privacy.)

7. Is there a need for specific protections for freedom of expression for the internet with regard to the practice of investigative journalism? Why/Why not?

8. Has legislation been drafted, or have legal cases been run or decided in your country (or region of operation), that have defined/tested the eligibility of bloggers/citizen journalists to claim protection from revealing sources under shield laws? Please provide examples.

9. Are you aware of any legislation/formal policies or legal cases where the issue and role of third party online intermediaries (e.g. Google, Facebook, Twitter) in the protection of journalists' sources has been in question (e.g. where a third party site may have access to data that, if revealed, could identify a source, and a court requires said third

party site to produce such data, thereby circumventing the journalist's legal right and/or ethical obligation to protect their source? If yes, please detail.

10. If relevant to your organisation, have you implemented policies and procedures and/or campaigns designed to make journalists and/or whistle-blowers aware of the changing digital environment as regards source protection? If yes, please elaborate.

APPENDIX 5 – PROTECTING JOURNALISM SOURCES IN THE DIGITAL AGE QUALITATIVE INTERVIEW QUESTIONS

a. Questions for lawyers, human rights activists, NGOs

1. How secure is the state of legal protection for journalists' sources in the digital era?

2) What are the key threats and challenges emerging regarding source protection in your experience?

3. How significant is the threat of mass surveillance (state and corporate) to the effectiveness of source protection laws in your region/work? (Please elicit examples).

4. What about the role of National Security/Anti-terrorism legislation (which is having a limiting effect on source protection laws) in undercutting source protection laws? How is that issue manifesting in your region? (Please elicit examples).

5. What pressures are you seeing emerging for journalistic source protection regarding third party intermediaries like Facebook, Twitter, Google, mobile companies and ISPs with regards to data retention and data handover (to courts, governments etc.)? (Please elicit examples)

6. To who should source protection laws apply in the digital era? Professional journalists (if so, how do we define them?)? All digital media actors? Or should we rather link protection to 'acts of journalism' (again, how should we define these acts)?

7. Is it actually possible any longer for journalists to promise confidentiality to sources when you consider the impacts of mass surveillance, data retention and the overriding effect of national security/anti-terrorism legislation as they undermine the legal protections traditionally deployed to allow journalists to shield their sources from unmasking?

8. How can legal source protection be strengthened in the digital era? Is it possible, for example, to consider introducing legal exclusions for journalists to protect them (or their data) from being exposed by mass surveillance?

9. What do you think of this proposed framework for assessing source protection laws in the digital era? Please comment on each point as desired. What might you exclude/add to this list to enable it to work as a tool for measuring the effectiveness of source protection laws?

   Ideally, a model source protection law might:

   1. Recognise the ethical principle and value to society of source protection

2. Recognise that protection extends to all acts of journalism, defined in inclusive terms

3. Recognise that source protection does not entail registration or licensing of practitioners of journalism

4. Affirm that confidentiality applies to the use of any collected digital personal data by any actor

5. Define exceptions to all the above very narrowly in terms of purposes allowing limitation of the principle

6. Define exceptions as needing to conform to the necessity provision, in other words, when there is no alternative

7. Define an independent judicial process, with appeal potential, for authorised exceptions

8. Criminalise arbitrary and unauthorised violations of confidentiality of sources by any 3rd party

10. What, if any, action would you like to see in your region, or internationally, with regard to strengthening the legal protection of journalists' sources?

b. Alternative question set for journalistic interviewees

1. How much confidence do you place in the legal source protections offered in your country/region in 2014?

2. How has your confidence in legal source protection been affected by the emerging digital era issues of (please elicit explanatory answers):

    a) Mass surveillance – is there any point to laws that defend journalists' right not to disclose confidential sources (e.g. shield laws) if mass surveillance risks unmasking them regardless?

    b) Data retention laws (and connected demands to hand over data being applied to third party intermediaries like Facebook, Twitter, Google, ISPs)?

    c) National security/anti-terrorism legislation (as it limits the reach of source protection laws)?

3. What impacts have these changes had on your confidential sources themselves– is there evidence of a chilling effect in play? Are they more reluctant to come forward with information than they were previously? (Please elicit examples).

4. Do you still believe it's possible to promise sources that they will remain confidential and benefit from legal protections (where they apply in your region)? And do you feel ethically comfortable doing so? If yes, why? If not, why not?

5. To who should source protection laws apply in the digital era? Professional journalists (if so, how do we define them?)? All digital media actors? Or should we rather link protection to 'acts of journalism' (again, how should we define these acts)?

6. Practically, how is insecurity around source protection changing the way in which you're undertaking investigative journalism (e.g. are you less to pursue stories dependent upon confidential sources? Are you adapting reporting practices in other ways? If so, how?).

7. How you been involved in collaborative international investigative journalism partnerships? If so, how have the above issues manifested in cross-border investigations? How are you navigating differing international legal standards and practices in such investigations? What have you gleaned about the effectiveness of source protection laws globally in this context?

8. How can legal source protection be strengthened in the digital era? For example, should States consider introducing legal exclusions for journalists to protect them (or their data) from being exposed by mass surveillance or data retention/handover? (Why/Why not?) Should there also be closer scrutiny of the limiting effect of national security/anti-terrorism laws on source protection laws? (Why/Why not?)

9. What do you think of this proposed framework for assessing source protection laws in the digital era? Please comment on each point as desired. What might you exclude/add to this list to enable it to work as a tool for measuring the effectiveness of source protection laws?

   Ideally, a model source protection law might:

   1. Recognise the ethical principle and value to society of source protection
   2. Recognise that protection extends to all acts of journalism, defined in inclusive terms
   3. Recognise that source protection does not entail registration or licensing of practitioners of journalism
   4. Affirm that confidentiality applies to the use of any collected digital personal data by any actor
   5. Define exceptions to all the above very narrowly in terms of purposes allowing limitation of the principle
   6. Define exceptions as needing to conform to the necessity provision, in other words, when there is no alternative
   7. Define an independent judicial process, with appeal potential, for authorised exceptions
   8. Criminalise arbitrary and unauthorised violations of confidentiality of sources by any 3rd party

10. What, if any, other action would you like to see in your region, or internationally, with regard to strengthening the legal protection of journalists' sources?

SELECTED BIBLIOGRAPHY

United Nations:

Committee on the Elimination of Racial Discrimination. 2002. General Recommendation 29, Discrimination Based on Descent (Sixty-first session, 2002), U.N. Doc. A/57/18 at 111 (2002), reprinted in Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies, U.N. Doc. HRI\GEN\1\Rev.6 at 223 (2003)

Office of the UN High Commissioner for Human Rights. 2012, 5 October. *Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence*. http://www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf

—. 2011. *Guiding Principles on Business and Human Rights*. New York and Geneva: United Nations.

—. 1966, 16 December. *International Covenant on Civil and Political Rights*. www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

UN General Assembly. 2015, 11 February. *Resolution adopted by the General Assembly on 18 December 2014, 69/185. The safety of journalists and the issue of impunity*. A/RES/69/185. http://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/69/185

—. 2014, 23 September. *Report of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism.* A/69/397. http://daccess-dds-ny.un.org/doc/UNDOC/GEN/N14/545/19/PDF/N1454519.pdf?OpenElement

—. 2014, 21 February. *The safety of journalists and the issue of impunity: Resolution adopted by the General Assembly on 18 December 2013*. A/RES/68/163. http://www.refworld.org/docid/53a7fab74.html

—. 2013, 20 November. *The right to privacy in the digital age. (A/C.3/68/167).*

http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/68/167

—. 1979. *Convention on the Elimination of All Forms of Discrimination Against Women*. http://www.un.org/womenwatch/daw/cedaw/

—. 1966, 16 December. *International Covenant on Civil and Political Rights*. http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

—. 1948, 10 December. *Universal Declaration of Human Rights*. http://www.un.org/en/documents/udhr/

UN General Assembly, Human Rights Committee. 2015, 5 January. *Report of the Special Rapporteur on minority issues, Rita Izsák*. A/HRC/28/64

—. 2011, 11-29 July. *General Comment 34*. http://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf

—. 2011, 14 July. *Resolution 17/19, Human rights, sexual orientation and gender identity*. A/HRC/RES/17/19

—. 2000. *General Comment 28, Equality of rights between men and women*, U.N. Doc. CCPR/C/21/Rev.1/Add.10

—. 1993. *General Comment 22, Article 18* (Forty-eighth session). Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies, U.N. Doc. HRI/GEN/1/Rev.1 at 35 (1994)

UN Human Rights Council. 2015, 22 May. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye.* A/HRC/29/32. http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/CallForSubmission.aspx

—. 2014, 2 October. *Resolution adopted by the Human Rights Council, 27/5: The safety of journalists*. A/HRC/RES/27/5. http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G14/177/81/PDF/G1417781.pdf?OpenElement

—. 2014, 23 July. *Discussion on the safety of journalists: Report of the Office of the United Nations High Commissioner for Human* Rights. A/HRC/27/35. http://www.refworld.org/docid/53eb46d34.html

—. 2014, 30 June. *The right to privacy in the digital age: Report of the Office of the United Nations High Commissioner for Human Rights.* A/HRC/27/37. www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session27/Documents/A.HRC.27.37_en.pdf

—. 2013, 1 July. *The safety of journalists: Report of the Office of the United Nations High Commissioner for Human Rights*. A/HRC/24/23. http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G13/153/19/PDF/G1315319.pdf?OpenElement

—. 2013, 17 April. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. A/HRC/23/40. http://www.refworld.org/docid/51a5ca5f4.html

—. 2012, 9 October. *Safety of journalists: resolution/adopted by the Human Rights Council*. A/HRC/RES/21/12. http://www.refworld.org/docid/50adf4812.html

—. 2012, 16 July. *The promotion, protection and enjoyment of human rights on the Internet.* A/HRC/RES/20/8. http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/20/8

—. 2012, 4 June. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue.* A/HRC/20/17. http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session20/A-HRC-20-17_en.pdf

—. 2011, 16 May. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue*. A/HRC/17/27. http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf

—. 2010, 20 April. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue*. A/HRC/14/23. http://www2.ohchr.org/english/bodies/hrcouncil/docs/14session/A.HRC.14.23.pdf

—. 2009, 16 January. *Report of the United Nations High Commissioner for Human Rights Addendum, Expert seminar on the links between articles 19 and 20 of the International Covenant on Civil and Political Rights*, A/HRC/10/31/Add.3. http://www2.ohchr.org/english/issues/opinion/articles1920_iccpr/docs/A-HRC-10-31-Add3.pdf

—. 2006, 10 September. *Incitement to racial and religious hatred and the promotion of tolerance: report of the High Commissioner for Human Rights*. A/HRC/2/6. http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G06/139/97/PDF/G0613997.pdf?OpenElement

UN Office on Drugs and Crime. 2013. *Global Study on Homicide*. http://www.unodc.org/gsh/

—. 2003. *United Nations Convention Against Corruption*. https://www.unodc.org/documents/treaties/UNCAC/Publications/Convention/08-50026_E.pdf

UN Radio. 2013, 12 July. 'Human Rights chief urges respect for right to privacy and protection of individuals revealing human rights violations'. http://www.unmultimedia.org/radio/english/2013/07/human-rights-chief-urges-respect-for-right-to-privacy-and-protection-of-individuals-revealing-human-rights-violations/

UN Security Council. 2015, 27 May. *Resolution 2222 (2015): Adopted by the Security Council at its 7450th meeting, on 27 May 2015*. S/RES/2222 (2015). http://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/s_res_2222.pdf

—. 2013, 22 November. *Report of the Secretary-General on the protection of civilians in armed conflict*. S/2013/689. http://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/s_2013_689.pdf

UNESCO

Daudin Clavaud, P. and T. Mendel. 2015. *Freedom of Expression and Public Order: Training manual*. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002313/231305e.pdf

Dutton, W. H. et al. 2011. *Freedom of Connection, Freedom of Expression: The Changing Legal and Regulatory Ecology Shaping the Internet*. UNESCO Series on Internet Freedom. Paris: UNESCO. http://unesdoc.unesco.org/images/0019/001915/191594e.pdf

Gagliardone, I. et al. 2015. *Countering Online Hate Speech*. UNESCO Series on Internet Freedom. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002332/233231e.pdf

Henrichsen, J. R., M. Betz and J. M. Lisosky 2015. *Building Digital Safety for Journalism: A Survey of Selected Issues*. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002323/232358e.pdf

MacKinnon, R. et al. 2014. *Fostering Freedom Online: The Role of Internet Intermediaries*. UNESCO Series on Internet Freedom. Paris: UNESCO / Internet Society. http://unesdoc.unesco.org/images/0023/002311/231162e.pdf

Mendel, T. et al. 2012. *Global Survey on Internet Privacy and Freedom of Expression*. UNESCO Series on Internet Freedom. Paris: UNESCO. http://unesdoc.unesco.org/images/0021/002182/218273e.pdf

UNESCO. *Global Citizenship Education*. http://www.unesco.org/new/en/global-citizenship-education

—. 2015. *Keystones to foster inclusive Knowledge Societies: Access to information and knowledge, Freedom of Expression, Privacy, and Ethics on a Global Internet*. Paris: UNESCO. http://unesdoc.unesco.org/images/0023/002325/232563E.pdf

—. 2015, 17 March. *Decision 196 EX/31 on Safety of Journalists and the Issue of Impunity*. Adopted at the 196th Session of UNESCO's Executive Board. http://unesdoc.unesco.org/images/0023/002323/232337e.pdf

—. 2014. *World Trends in Freedom of Expression and Media Development*. Paris: UNESCO. www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/world-trends-in-freedom-of-expression-and-media-development

—. 2014, July. *Internet Universality: A Means towards Building Knowledge Societies and the Post-2015 Sustainable Development Agenda*. Draft Proposed by the Secretariat. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet_universality_summary_240314_en.pdf

—. 2014, May. *Paris Declaration on Media and Information Literacy in the Digital Era*. http://www.unesco.org/new/en/communication-and-information/resources/news-and-in-focus-articles/in-focus-articles/2014/paris-declaration-on-media-and-information-literacy-adopted/

—. 2013. *Draft Medium-Term Strategy: 2014–2021 (37 C/4)*. Paris: UNESCO. http://unesdoc.unesco.org/images/0022/002200/220031e.pdf

—. 2013, November. *Resolution on Internet-related issues: including access to information and knowledge, freedom of expression, privacy and ethical dimensions of the information society*. 37th session of the General Conference. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/37gc_resolution_internet.pdf

—. 2011, 25 October – 10 November. *Records of the General Conference, 36th session*. Volume 1: Resolutions. http://unesdoc.unesco.org/images/0021/002150/215084e.pdf

UNESCO, Intergovernmental Council of the International Programme for the Development of Communication (IPDC). 2014, 20-21 November. *Decisions taken by the 29th IPDC Council Session*. Room X, UNESCO Headquarters, Paris. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/IPDC/ipdc29EN_IPDC29_FULL_DECISIONS_FINAL.pdf

—. 2012, 22-23 March. *Final Report (Twenty-eighth session)*. UNESCO: Paris. http://unesdoc.unesco.org/images/0021/002199/219910E.pdf

—. 2010, 24-26 March. *Final Report (Twenty-seventh session)*. UNESCO: Paris. http://unesdoc.unesco.org/images/0018/001896/189697m.pdf

—. 2008, 26-28 March. *Final Report (Twenty-sixth session)*. UNESCO: Paris. http://unesdoc.unesco.org/images/0016/001634/163437m.pdf

UNESCO, the International Programme for the Development of Communication (IPDC). 2014. *The Safety of Journalists and the Danger of Impunity Report by the Director-General to the Intergovernmental Council of the IPDC (Twenty-Ninth Session)*. CI-14/CONF.202/4 Rev2. Paris. http://unesdoc.unesco.org/images/0023/002301/230101E.pdf

Other inter-governmental organizations:

African Commission on Human and Peoples' Rights. 2002, 17-23 October. Declaration of Principles on Freedom of Expression in Africa, 32nd Session, Banjul

APEC Cross-Border Privacy Rules System. http://www.cbprs.org/

Association of Southeast Asian Nations (ASEAN). 2012, 19 November. *ASEAN Human Rights Declaration*. http://www.asean.org/news/asean-statement-communiques/item/asean-human-rights-declaration

Benedek, W. and M. C. Kettemann. 2013, December. *Freedom of Expression and the Internet*. Strasbourg: Council of Europe. https://book.coe.int/eur/en/human-rights-and-democracy/5810-freedom-of-expression-and-the-internet.html

Bigo et al. 2013. *National Programmes for Mass Surveillance of Personal Data in EU Member States and their compatibility with EU Law*. European Parliament. http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/493032/IPOL-LIBE_ET%282013%29493032_EN.pdf

Botero Marino, C. 2014, 22 April. *Annual Report of the Inter-American Commission on Human Rights 2013: Annual report of the office of the special rapporteur for freedom of expression,* Volume ii. Washington, D.C.: Organization of American States. http://www.oas.org/en/iachr/expression/docs/reports/2014_04_22_%20IA_2013_ENG%20_FINALweb.pdf

—. 2013, 31 December. *Violence against journalists and media workers: Inter-American standards and national practices on prevention, protection and prosecution of perpetrators*. Office of the Special Rapporteur for Freedom of Expression, Inter-American Commission on Human Rights. http://www.oas.org/en/iachr/expression/docs/reports/2014_04_22_Violence_WEB.pdf

—. 2012. *Annual Report of the Inter-American Commission on Human Rights,* volume II, Report of the Office of the Special Rapporteur for Freedom of Expression

Broadband Commission Working Group on Broadband and Gender. 2013, September. *Doubling Digital Opportunities: Enhancing the Inclusion of Women & Girls in the Information Society*. Geneva: International Telecommunication Union. www.broadbandcommission.org/Documents/working-groups/bb-doubling-digital-2013.pdf

Council of Europe. 2014, 16 April. *Recommendation of the Committee of Ministers to member States on a Guide to human rights for Internet users*. (CM/Rec(2014)6.) https://wcd.coe.int/ViewDoc.jsp?id=2184807

—. 2012, 15 April. *Mapping study on projects against hate speech online*. Strasbourg. https://www.coe.int/t/dg4/youth/Source/Training/Training_courses/2012_Mapping_projects_against_Hate_Speech.pdf

—. 2007, 26 September. *Guidelines of the Committee of Ministers of the Council of Europe on protecting freedom of expression and information in times of crisis,* 1005th meeting. https://wcd.coe.int/ViewDoc.jsp?id=1188493

—. 2003, 28 January. *Additional Protocol to the Convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*

—. 2001, 23 November. *Convention on Cybercrime*

Council of Europe, Committee of Ministers. 2003, 28 May. *Declaration on freedom of communication on the Internet*. (Decl-28.05.2003E.) https://wcd.coe.int/ViewDoc.jsp?id=37031

—. 2000. *Recommendation on 'The Right of Journalists Not to Disclose Their Sources of Information'*. http://www.coe.int/t/dghl/standardsetting/media/doc/cm/rec%282000%29007&expmem_EN.asp

Council of Europe, Commissioner for Human Rights. 2011, 4 October. 'Protection of journalists from violence: Issue discussion paper'. https://wcd.coe.int/ViewDoc.jsp?id=1899957https://wcd.coe.int/ViewDoc.jsp?id=1899957

Council of Europe, European Commission against Racism and Intolerance (ECRI). 2000, 15 December. *ECRI General Policy Recommendation N°6: Combating the dissemination of racist, xenophobic and antisemitic materiel via the internet*. http://www.coe.int/t/dghl/monitoring/ecri/activities/gpr/en/recommendation_n6/Recommendation_6_en.asp

Council of Europe, Parliamentary Assembly. 2011, 25 January. *Recommendation 1950: The protection of journalists' sources*. http://assembly.coe.int/Mainf.asp?link=/Documents/AdoptedText/ta11/EREC1950.htm

Court of Justice of the European Union. 2014, 8 April. The Court of Justice declares the Data Retention Directive to be invalid. Press Release No. 54/14. http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-04/cp140054en.pdf

Edwards, L. 2011, 22 June. *Role and Responsibility of Internet Intermediaries in the Field of Copyright and Related Rights*. Geneva: World Intellectual Property Organisation. (WIPO-ISOC/GE/11/REF/01/EDWARDS). www.wipo.int/export/sites/www/copyright/en/doc/role_and_responsibility_of_the_internet_intermediaries_final.pdf

European Commission. 2013, 16 July. *Overview on Binding Corporate Rules. Data Protection*. http://ec.europa.eu/justice/data-protection/document/international-transfers/binding-corporate-rules/index_en.htm

—. 2008, 28 November. *Framework Decision on Racism and Xenophobia*. http://ec.europa.eu/justice/fundamental-rights/racism-xenophobia/framework-decision/index_en.htm

—. 2000, 4 May. *E-Commerce Directive*. (2000/31/EC). http://ec.europa.eu/internal_market/e-commerce/directive/index_en.htm

European Court of Human Rights. 1996. *Goodwin v United Kingdom*. http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-57974

European Parliament and the Council of the European Union. 2001, 22 May. *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society*. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32001L0029

European Union. 2000. *Charter of Fundamental Rights of the European Union*. http://ec.europa.eu/justice/fundamental-rights/charter/

Horsley, W. 2012. *OSCE Safety of journalists guidebook*. Office of the OSCE Representative on Freedom of the Media. https://www.osce.org/fom/85777?download=true

Hulin, A. (Ed.). 2013. *Joint Declarations of the representatives of intergovernmental bodies to protect free media and expression*. Vienna: Organization for Security and Co-operation in Europe. www.osce.org/fom/99558?download=true

Inter-American Commission on Human Rights. 2000, 20 October. *Inter-American Declaration of Principles on Freedom of Expression*

—. 1985, 13 November. *Advisory Opinion OC-5/85*

Inter-American Court of Human Rights. 2009. 28 January. *Case of Ríos et al. v. Venezuela*. http://corteidh.or.cr/docs/casos/articulos/seriec_194_ing.pdf

ISO/IEC. 2004, March. FDIS 11179-1. 'Information technology - Metadata registries - Part 1: Framework'. http://stats.oecd.org/glossary/detail.asp?ID=4575

League of Arab States. 2004, 22 May. *Arab Charter on Human Rights*. Entered into force 15 March 2008

Ministers of the Freedom Online Coalition. *Recommendations for Freedom Online*. Adopted in Tallinn, Estonia on 28 April 2014. https://www.freedomonlinecoalition.com/wp-content/uploads/2014/04/FOC-recommendations-consensus.pdf

Organisation for Economic Co-operation and Development. 2011, 13 December. *OECD Council Recommendation on Principles for Internet Policy Making*. www.oecd.org/internet/ieconomy/49258588.pdf

—. 2011, September. *The Role of Internet Intermediaries in Advancing Public Policy Objectives*. Paris: OECD http://browse.oecdbookshop.org/oecd/pdfs/product/9311031e.pdf

Organization of American States. *American Convention on Human Rights 'Pact of San José, Costa Rica' (B-32)*. http://www.oas.org/dil/treaties_B-32_American_Convention_on_Human_Rights.htm

—. 2011. *Mandatory membership in a professional association for the practise of journalism*. http://www.oas.org/en/iachr/expression/showarticle.asp?artID=154&lID=1

Organization of the Islamic Conference. 1990, 5 August. *Cairo Declaration on Human Rights in Islam*, preambular section

Organization of Islamic Cooperation. 2013, December. *Sixth OIC Observatory Report on Islamophobia*. Presented to the 40th Council of Foreign Ministers, Conakry, Republic of Guinea

Organization for Security and Co-operation in Europe. *Decriminalization of defamation*. www.osce.org/fom/106287

—. 2011, 8 June. *Vilnius Recommendations on Safety of Journalists*. http://www.osce.org/cio/78522

Perset, K. / OECD. 2010, March. *The Economic and Social Role of Internet Intermediaries*. (DSTI/ICCP(2009)9/FINAL). Paris: OECD. www.oecd.org/internet/ieconomy/44949023.pdf

Other documents and resources

Access Now. *Telco Remedy Plan*. https://www.accessnow.org/telco-remedy-plan

ARTICLE 19. 2009, April. *The Camden Principles on Freedom of Expression and Equality*. https://www.article19.org/data/files/pdfs/standards/the-camden-principles-on-freedom-of-expression-and-equality.pdf

Broadband Stakeholder Group (UK). *Voluntary industry code of practice on traffic management transparency for broadband services*. http://www.broadbanduk.org/wp-content/uploads/2013/08/Voluntary-industry-code-of-practice-on-traffic-management-transparency-on-broadband-services-updated-version-May-2013.pdf

The Center for Internet and Society. 2014, July. *World Intermediary Liability Map (WILMap)*. Stanford, Calif.: Stanford Law School. http://cyberlaw.stanford.edu/our-work/projects/world-intermediary-liability-map-wilmap

Chilling Effects. http://www.chillingeffects.org

Committee to Protect Journalists. 2014, 1 December. *2014 prison census: 221 journalists jailed worldwide*. https://cpj.org/imprisoned/2014.php

—. 2013, 1 December. *2013 prison census: 211 journalists jailed worldwide*. https://cpj.org/imprisoned/2013.php

Electronic Frontier Foundation. *Takedown Hall of Shame*. https://www.eff.org/takedowns

—. 2014. *Who Has Your Back? Protecting Your Data From Government Requests.* https://www.eff.org/who-has-your-back-2014

European University Institute, Centre for Media Pluralism and Media Freedom. 2014. *Status of European Journalists*. http://journalism.cmpf.eui.eu/maps/journalists-status/

Facebook. *Community Standards*. https://www.facebook.com/communitystandards

Global Network Initiative. *Implementation Guidelines*. https://globalnetworkinitiative.org/implementationguidelines/index.php

—. *Principles*. http://www.globalnetworkinitiative.org/principles/

Google. *Transparency Report.* http://www.google.com/transparencyreport/

Hatebase. *Most Common Hate Speech*. http://www.hatebase.org/popular

In Other Words Project. 2013. *Toolbox.* http://www.inotherwords-project.eu/sites/default/files/Toolbox.pdf

Internet Live Stats. 2014. *Internet Users by Country*. www.internetlivestats.com/internet-users-by-country

MediaSmarts. *Facing online hate*. http://mediasmarts.ca/tutorial/facing-online-hate-tutorial

Microsoft, Windows Dev Center. '11.0 Content Policies'. *Windows apps.* https://msdn.microsoft.com/en-us/library/windows/apps/Dn764940.aspx

Microsoft, Xbox. 2014, January. *Xbox Live Code of Content*. http://www.xbox.com/en-GB/legal/codeofconduct

Necessary and Proportionate. 10 July 2013. *International Principles on the Application of Human Rights to Communications Surveillance*. https://en.necessaryandproportionate.org/text

No Hate Speech Movement. 2013. *Campaign tools and materials*. http://nohate.ext.coe.int/Campaign-Tools-and-Materials

—. 2013. *No Hate Ninja Project - A Story About Cats, Unicorns and Hate Speech*. https://www.youtube.com/watch?v=kp7ww3KvccE

Ofcom. 2014, 22 July. *Report on Internet safety measures - Internet Service Providers: Network level filtering measures*. http://stakeholders.ofcom.org.uk/internet/internet-safety-2?utm_source=updates&utm_medium=email&utm_campaign=filtering-report

Online Hate Prevention Institute. *Fight Against Hate*. http://fightagainsthate.com/

Open Rights Group. 2014, July. *Blocked! The personal cost of filters*. https://www.blocked.org.uk/personal-stories

OpenNet Initiative. About Filtering. https://opennet.net/about-filtering

Organisation for Economic Co-operation and Development. 2014, March. *The CleanGovBiz Toolkit for Integrity*. http://www.oecd.org/cleangovbiz/CGB-Toolkit-2014.pdf

Reporters Without Borders. *2014: Journalists Imprisoned*. https://en.rsf.org/press-freedom-barometer-journalists-imprisoned.html?annee=2014

—. *2013: Journalists Imprisoned*. https://en.rsf.org/press-freedom-barometer-journalists-imprisoned.html?annee=2013

Telecommunications Industry Dialogue on Freedom of Expression and Privacy. 2013, 16 March. *Guiding Principles*. http://www.vodafone.com/content/dam/sustainability/pdfs/telecom_industry_dialogue_principles.pdf

Tell MAMA (Measuring Anti-Muslim Attacks). 2014. http://tellmamauk.org

Terms of Service; Didn’t Read. https://tosdr.org/

Twitter. *The Twitter Rules*. https://support.twitter.com/articles/18311

—. 2015, 18 May. *Twitter Terms of Service.* https://twitter.com/tos?lang=en

UC Berkeley Library. *Invisible or Deep Web: What it is, How to find it, and Its inherent ambiguity*. http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html

UNESCO. *UNESCO Condemns Killing of Journalists*. http://www.unesco.org/new/en/condemnation

WAM! *WAM Twitter harassment reporting tool*. https://womenactionmedia.wufoo.com/forms/ztaetji1jrhv10/

YouTube. *Community Guidelines*. http://www.youtube.com/yt/policyandsafety/communityguidelines.html

Books, articles and reports

African Gender Institute. 2013, December. *Feminist Africa*, Vol. 18, ‘e-spaces: e-politics’. http://agi.ac.za/sites/agi.ac.za/files/fa18_web-1.pdf

Albanian Media Institute. 2014. *Hate speech in online media in South East Europe.* http://www.institutemedia.org/Documents/PDF/Hate%20speech%20in%20online%20media%20in%20SEE.pdf

Alston, P. (Ed.). 2005. *Non-State Actors and Human Rights*. Oxford: Oxford University Press.

Alves, R. 2014. ‘Trends in global collaborative journalism’, *Trends in Newsrooms 2014,* Darmstadt, Germany: WAN-IFRA, pp. 83-87

Andrejevic, M. 2014. Wikileaks, Surveillance, and Transparency. *International Journal of Communication*, 8, pp. 2619–2630

Athique, A. 2013. *Digital Media and Society: An Introduction*. Polity.

Banisar, D. 2008, November. *Speaking of terror: A survey of the effects of counter-terrorism legislation on freedom of the media in Europe*. Council of Europe, Media and Information Society Division Directorate General of Human Rights and Legal Affairs. http://www.coe.int/t/dghl/standardsetting/media/Doc/SpeakingOfTerror_en.pdf

—. 2007. *Silencing Sources: An international survey of protections and threats to journalists' sources*. Privacy International. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1706688 accessed 25/6/2014

Bankston, K., D. Sohn and A. McDiarmid. 2012, December. *Shielding the Messengers: Protecting Platforms for Expression and Innovation*. Washington DC: Center for Democracy and Technology. www.cdt.org/files/pdfs/CDT-Intermediary-Liability-2012.pdf

Barton, A. and H. Storm. 2014. *Violence and Harassment against Women in the News Media: A Global Picture*. International Women's Media Foundation and International News Safety Institute. http://www.iwmf.org/wp-content/uploads/2014/03/Violence-and-Harassment-against-Women-in-the-News-Media.pdf

Bauman, Z. et al. 2014. After Snowden: Rethinking the Impact of Surveillance. *International Political Sociology*, 8:2, 121-140

Bayley, E. 2009, 16 November. *The Clicks that Bind: Ways Users 'Agree' to Online Terms of Service.* Electronic Frontier Foundation. https://www.eff.org/wp/clicks-bind-ways-users-agree-online-terms-service

Benesch, S. 2012. Words as Weapons. *World Policy Journal*, vol. 29, no. 1, pp. 7-12

—. 2012, 12 January. 'Dangerous Speech: A Proposal to Prevent Group Violence'. New York: World Policy Institute. http://www.worldpolicy.org/sites/default/files/Dangerous%20Speech%20Guidelines%20Benesch%20January%202012.pdf

Bently, L. and B. Sherman. 2009. *Intellectual Property Law*, 3rd ed. Oxford: Oxford University Press.

Bergman, M. K. 2001, August. White Paper: The Deep Web: Surfacing Hidden Value. *Taking License*. Vol. 7, Issue 1. http://quod.lib.umich.edu/j/jep/3336451.0007.104

Beschastna, T. 2014. Freedom of Expression in Russia as it Relates to Criticism of the Government. *Emory International Law Review*, Vol. 27, No. 2. http://law.emory.edu/eilr/content/volume-27/issue-2/comments/freedom-expression-russia.html

Black, J. 1996, January. Constitutionalising Self-Regulation. *The Modern Law Review*, Vol. 59, No. 1, pp. 24-55. http://dx.doi.org/10.1111/j.1468-2230.1996.tb02064.x

BSR, with CDT. 2014, September. *Legitimate and Meaningful Stakeholder Engagement in Human Rights Due Diligence: Challenges and Solutions for ICT Companies*. http://www.bsr.org/reports/BSR_Rights_Holder_Engagement.pdf

Budish, R. 2013, 19 December. 'What Transparency Reports Don't Tell Us'. *The Atlantic*. www.theatlantic.com/technology/archive/2013/12/what-transparency-reports-dont-tell-us/282529

Business and Human Rights Resource Centre. 2010, September. *The UN 'Protect, Respect and Remedy' Framework for Business and Human Rights*. www.reports-and-materials.org/sites/default/files/reports-and-materials/Ruggie-protect-respect-remedy-framework.pdf

Buyse, A. 2014. Words of Violence: 'Fear Speech,' or How Violent Conflict Escalation Relates to the Freedom of Expression. *Human Rights Quarterly*, vol. 36, no. 4, pp. 779–97

Castells, M. 2012. *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge: Polity

Chin, Y. C. 2013, August. Regulating social media. Regulating life (and lives). *RJR 33 Online*, http://journalism.hkbu.edu.hk/doc/Regulating_social-Media.pdf

Citron, K. D. and H. Norton. 2011. Intermediaries and hate speech: Fostering digital citizenship for our information age. *Boston University Law Review,* vol. 91, pp. 1435-84

Comninos, A. 2012, October. *The Liability of Internet Intermediaries in Nigeria, Kenya, South Africa, and Uganda: An Uncertain Terrain*. South Africa: Association for Progressive Communications. www.apc.org/en/system/files/READY%20-%20Intermediary%20Liability%20in%20Africa_FINAL.pdf

——. 2012, October. *Intermediary liability in South Africa*. Intermediary Liability in Africa Research Papers, No. 3. South Africa: Association for Progressive Communications. www.apc.org/en/system/files/Intermediary_Liability_in_South_Africa-Comninos_06.12.12.pdf

Cotter, T. F. 2005. Some Observations on the Law and Economics of Intermediaries. *Michigan State Law Review*, Vol. 1, pp. 1-16. Washington & Lee Legal Studies Paper No. 2005-14. http://ssrn.com/abstract=822987

Das, S. and A. Kramer. 2013. Self-Censorship on Facebook. *Proceedings of the Seventh International Association for the Advancement of Artificial Intelligence (AAAI) Conference on Weblogs and Social Media*, pp. 120-27. www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6093/6350

Davies, S. (Ed.). 2014, June. A Crisis of Accountability: A global analysis of the impact of the Snowden revelations. *Privacy Surgeon*. www.privacysurgeon.org/blog/wp-content/uploads/2014/06/Snowden-final-report-for-publication.pdf

Defeis, E.F., 1992. Freedom of speech and international norms: A response to hate speech. *Stan. Journal of International Law*, vol. 29, pp. 57-74

Deibert, R. et al. (Eds). 2010, April. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, Mass.: MIT Press. http://mitpress.mit.edu/books/access-controlled

——. 2008. January. *Access Denied: The Practice and Policy of Global Internet Filtering.* Cambridge, Mass.: MIT Press. https://mitpress.mit.edu/books/access-denied

DeNardis, L. 2013, August. Internet Points of Control as Global Governance. Internet Governance Papers. No.2. Centre for International Governance Innovation. http://www.cigionline.org/sites/default/files/no2_3.pdf

Diamond, L. 2010, July. Liberation technology. *Journal of Democracy*, Vol. 21, No. 3, pp. 69-83. www.journalofdemocracy.org/articles/gratis/Diamond-21-3.pdf

Electronic Frontier Foundation. 2011, January. *Freedom of Expression, Privacy and Anonymity on the Internet*. https://www.eff.org/Frank-La-Rue-United-Nations-Rapporteur

Epstein, G. 3 March 2011. Sina Weibo. *Forbes Asia*. www.forbes.com/global/2011/0314/features-charles-chao-twitter-fanfou-china-sina-weibo.html

Foxman, A.H. and C. Wolf. 2013. Viral hate: Containing its spread on the Internet. Macmillan

Ghanea, N. 2013. Intersectionality and the Spectrum of Racist Hate Speech: Proposals to the UN Committee on the Elimination of Racial Discrimination. *Human Rights Quarterly*, vol. 35, no. 4, pp. 935-54. http://dx.doi.org/10.1353/hrq.2013.0053

Gillespie, T. 2010. The Politics of ‘Platforms’. *New Media & Society*, vol. 12, no. 3, pp. 347-64. http://dx.doi.org/10.1177/1461444809342738

Giroux, H. 2015. Totalitarian Paranoia in the Post-Orwellian Surveillance State. *Cultural Studies*, vol. 29, no. 2, pp. 108-140

Global Network Initiative. 2014, January. *Public Report on the Independent Assessment Process for Google, Microsoft, and Yahoo*. http://globalnetworkinitiative.org/sites/default/files/GNI%20Assessments%20Public%20Report.pdf

Goldsmith, J.L. and T. Wu. 2006. *Who controls the Internet? Illusions of a borderless world*. Oxford: Oxford University Press. http://jost.syr.edu/wp-content/uploads/who-controls-the-internet_illusions-of-a-borderless-world.pdf

Goodman, E. and F. Cherubini. 2013. *Online comment moderation: emerging best practices. A guide to promoting robust and civil online conversation*. World Association of Newspapers and News Publishers (WAN-IFRA). http://www.wan-ifra.org/reports/2013/10/04/online-comment-moderation-emerging-best-practices

Grabowicz, P. A. et al. 2012. Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. *PLoS ONE*, Vol. 7, No. 1. http://dx.doi.org/10.1371/journal.pone.0029358

Hannak, A. et al. Measuring Personalization of Web Search. *WWW '13 Proceedings of the 22nd international conference on World Wide Web*, pp. 527-538. http://www.ccs.neu.edu/home/cbw/pdf/fp039-hannak.pdf

Harvey, D., VP, Trust & Safety, Twitter. 2013, 29 July. *We hear you.* https://blog.twitter.com/en-gb/2013/we-hear-you

Herpai, G. 2013, 7 January. Unsocial network: the rise and fall of iWiW. *Budapest Business Journal*. www.bbj.hu/business/unsocial-network-the-rise-and-fall-of-iwiw_64418

Hoechsmann, M. and S. R. Poyntz. 2012. *Media Literacies. A Critical Introduction.* Oxford: Wiley-Blackwell

Hope, D. A. 2011, February. *Protecting Human Rights in the Digital Age.* BSR. https://globalnetworkinitiative.org/sites/default/files/files/BSR_ICT_Human_Rights_Report.pdf

Howard, P. N. 2010. *The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam*. Oxford: Oxford University Press

Human Rights Watch. 2014. *Liberty to Monitor All*. https://www.hrw.org/report/2014/07/28/liberty-monitor-all/how-large-scale-us-surveillance-harming-journalism-law-and

iHub Research. 2013. *Umati Final Report*. http://www.research.ihub.co.ke/uploads/2013/june/1372415606__936.pdf

Imre, A. 2009, May. National intimacy and post-socialist networking. *European Journal of Cultural Studies*, Vol. 12, No. 2, pp. 219-33

The Institute for Human Rights and Business and Shift. 2013, June. *ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights*. European Commission. www.shiftproject.org/publication/european-commission-ict-sector-guide

Intel Corporation and Dalberg Global Development Advisors. 2012. *Women and the Web: Bridging the Internet Gap and Creating New Global Opportunities in Low and Middle-Income Countries*. www.intel.com/content/dam/www/public/us/en/documents/pdf/women-and-the-web.pdf

Internet Watch Foundation. 2013. *Internet Watch Foundation Annual & Charity Report 2013*. Cambridge: IWF, www.iwf.org.uk/assets/media/annual-reports/annual_report_2013.pdf.pdf

Jellema, A. and K. Alexander. 2013, 22 November. *2013 Web Index Report*. Geneva: World Wide Web Foundation. http://thewebindex.org/wp-content/uploads/2013/11/Web-Index-Annual-Report-2013-FINAL.pdf

Johnson, E. J., S. Bellman and G. L. Lohse. 2002. Defaults, Framing, and Privacy: Why Opting In-Opting Out. *Marketing Letters*, Vol. 13, No. 1. https://www0.gsb.columbia.edu/mygsb/faculty/research/pubfiles/1173/defaults_framing_and_privacy.pdf

Kamdar, A. 2012, 6 December. EFF's Guide to CDA 230: The Most Important Law Protecting Online Speech. *EFF Deeplinks* Blog. https://www.eff.org/deeplinks/2012/12/effs-guide-cda-230-most-important-law-protecting-online-speech

Kohl, U. 2002. Eggs, Jurisdiction, and the Internet. *International and comparative law quarterly*, vol. 51, no. 3, pp. 556-582

KVG Research. 2013, December. *TV Market and Video on Demand in the Russian Federation*. Strasbourg: European Audiovisual Observatory. www.obs.coe.int/documents/205595/552774/RU+TV+and+VoD+2013+KVG+Research+EN.pdf/5fbb076c-868e-423a-bfed-dca8b66cac43

Laclau, E. and Mouffe, C. 1985. *Hegemony and Socialist Strategy. Towards a Radical Democratic Politics.* London: Verso

Learner, J. and R. Bar-Nissim. 2014. Law Enforcement Investigations Involving Journalists. *Legal Studies Research Paper Series*, no. 2014-71. School of Law, University of California, Irving

Leo, L. A., F. D. Gaer and E. K. Cassidy. 2011. Protecting Religions from Defamation: A Threat to Universal Human Rights Standards. *Harv. JL & Pub. Pol'y*, vol. 34, pp. 769-95

Limpitlaw, J. 2013. *Media Law Handbook for Southern Africa*, vol. 2. Johannesburg: Konrad-Adenauer-Stiftung Regional Media Programme. http://www.kas.de/wf/doc/kas_35248-1522-2-30.pdf?130825185204

Marquis-Boire, M. et al. 2013, March. 'You only click twice: FinFisher's Global Proliferation'. Citizen Lab. https://citizenlab.org/2013/03/you-only-click-twice-finfishers-global-proliferation-2/

Meddaugh, P. M. and Kay, J. 2009. Hate Speech or 'Reasonable Racism?' The Other in Stormfront, *Journal of Mass Media Ethics, Vol.* 24, no. 4, pp. 251-68. MediaSmarts.NDa

Lengyel, B. et al. 2013, 26 January. Distance dead or alive Online Social Networks from a geography perspective. SSRN. http://dx.doi.org/10.2139/ssrn.2207352

Levine, M., VP of Global Public Policy, Facebook. 2013, 28 May. 'Controversial, Harmful and Hateful Speech on Facebook'. https://www.facebook.com/notes/facebook-safety/controversial-harmful-and-hateful-speech-on-facebook/574430655911054

MacKinnon, R. 2012. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. New York: Basic Books.

Maireder, A. and S. Schlögl. 2014, December. 24 Hours of an #outcry: The Networked Publics of a Socio-Political Debate. *European Journal of Communication*, Vol. 29, No. 6

Marsden, C. T. 2011. *Internet Co-Regulation: European Law, Regulatory Governance, and Legitimacy in Cyberspace*. Cambridge: Cambridge University Press

Marthews, A. and C. Tucker. 2014, 24 March. Government Surveillance and Search Behavior. SSRN. http://ssrn.com/abstract=2412564

McNamee, J. 2011, January. *The Slide from Self-Regulation to Corporate Censorship*. Brussels: European Digital Rights Initiative. www.edri.org/files/EDRI_selfreg_final_20110124.pdf

Moore, M. 2007, June. Public interest, media neglect. *British Journalism Review*, vol. 18, no.2

Morsink, J. 1999. *The universal declaration of human rights: Origins, drafting, and intent*. University of Pennsylvania Press

Mossberger, K., C. J. Tolbert and R. S. McNeal. 2008. *Digital Citizenship. The Internet, Society and Participation.* London: The MIT Press

Nash, V. 2013. Analyzing Freedom of Expression Online: Theoretical, empirical, and normative contributions. In Dutton, W.H. (Ed.). *The Oxford Handbook of Internet Studies*. Oxford: Oxford University Press

Natour, F. and J. D. Pluess. 2013, March. *Conducting an Effective Human Rights Impact Assessment*. BSR. http://www.bsr.org/reports/BSR_Human_Rights_Impact_Assessments.pdf

Noorlander, P. 2014, 5 December. 'Finding Justice for Whistleblowers'. *Journalism in Europe discussion series,* Centre for Media Pluralism and Media Freedom, European University Institute

Nowak, M. 1993. *UN covenant on civil and political rights: CCPR commentary*. NP Engel Kehl

Nyst, C. 2014, July. *End violence: Women's rights and safety online project – Internet intermediaries and violence against women online. Executive summary and findings*. Association for Progressive Communications. http://www.genderit.org/sites/default/upload/flow-cnyst-summary-formatted.pdf

Omanovic, E. 2014, 20 November. *Private Interests: Monitoring Central Asia*. Privacy International*.* https://www.privacyinternational.org/?q=node/59

Osler, A. and H. Starkey. 2005. *Changing Citizenship.* Berkshire: Open University Press

Palfrey, J. G. Jr. Local Nets on a Global Network: Filtering and the Internet Governance Problem. *The Global Flow of Information.* In Balkin, J. (Ed.). Harvard Public Law Working Paper No. 10-41, p.8. http://ssrn.com/abstract=1655006

Parti, K. and L. Marin. 2013. Ensuring Freedoms and Protecting Rights in the Governance of the Internet: A Comparative Analysis on Blocking Measures and Internet Providers' Removal of Illegal Internet Content. *Journal of Contemporary European Research*, vol. 9, no. 1, pp. 138-59. www.jcer.net/index.php/jcer/article/view/455/392

Pasquale, F. A. 2010. Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries. *Northwestern University Law Review*, vol. 104, no. 1, pp. 105-74. www.law.northwestern.edu/lawreview/v104/n1/105/LR104n1Pasquale.pdf

Petrova, D. 2011, 9-10 February. 'Incitement to National, Racial or Religious Hatred: Role of Civil Society and National Human Rights Institutions'. 2011 Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred, Vienna

Pew Research Center in association with Columbia University's Tow Center for Digital Journalism. 2015, 5 February. *Investigative Journalists and Digital Security: Perceptions of Vulnerability and Changes in Behavior*. http://www.journalism.org/files/2015/02/PJ_InvestigativeJournalists_0205152.pdf

Phillips, G. 2014, 10 October. *On protection of journalistic sources*. Centre for Media Pluralism and Media Freedom, European University Institute. http://journalism.cmpf.eui.eu/discussions/on-protection-of-journalistic-sources/

Podkowik, J. 2014. 'Secret surveillance, national security and journalistic privilege – in search of the balance between conflicting values in the age of new telecommunication technologies'. University of Oslo. http://www.jus.uio.no/english/research/news-and-events/events/conferences/2014/wccl-cmdc/wccl/papers/ws8/w8-podkowik.pdf

Post, R., I. Hare and J. Weinstein. 2009. Hate speech. In *Extreme speech and democracy*. Oxford: Oxford University Press, pp. 123-38

Ramzy, A. 17 February 2011. Wired Up. *Time*. http://content.time.com/time/printout/0,8816,2048171,00.html

Rosenberg, R. S. 2011. Controlling access to the Internet: The role of filtering. *Ethics and Information Technology*, vol. 3, no. 1, pp. 35-54. www.copacommission.org/papers/rosenberg.pdf

Rotenberg, M. and D. Jacobs. 2013. Updating the Law of Information Privacy: The New Framework of the European Union. *Harvard Journal of Law & Public Policy*, vol. 36, no. 2, pp. 605-52. www.harvard-jlpp.com/wp-content/uploads/2013/04/36_2_605_Rotenberg_Jacobs.pdf

Rowbottom, J., 2012. To Rant, Vent and Converse: Protecting Low Level Digital Speech. *The Cambridge Law Journal*, vol. 71, no. 2, pp. 355-383

Russell, L. 2014. Shielding the Media: In an Age of Bloggers, Tweeters, and Leakers, Will Congress Succeed in Defining the Term 'Journalist' and in Passing a Long-Sought Federal Shield Act? *Oregon Law Review*, 93, pp. 193-227

Rustad, M. L. and D. D'Angelo. 2012. The Path of Internet Law: An Annotated Guide to Legal Landmarks. *Duke Law & Technology Review*, vol. 2011, no. 012. Suffolk University Law School Research Paper No. 11-18. http://ssrn.com/abstract=1799578

Ryngaert, C. 2008. *Jurisdiction in international law*. Oxford: Oxford University Press

Samway, M. A. 2014, October. Business, Human Rights and the Internet: A Framework for Implementation. In Lagon, M. P. and A. C. Arend (Eds.). *Human Dignity and the Future of Global Institutions.* Georgetown University Press

Savin, A. and J. Trzaskowski (eds). 2014. *Research Handbook on EU Internet Law*. Edward Elgar Publishing

Seng, D. and I. Garrote Fernandez-Diez. 2012. *Comparative Analysis of National Approaches of the Liability of the Internet Intermediaries*. Geneva: World Intellectual Property Organization. www.wipo.int/export/sites/www/copyright/en/doc/liability_of_internet_intermediaries.pdf

Sieminski, P. 2013, 21 November. Striking Back Against Censorship. WordPress *Hot Off the Press* Blog. http://en.blog.wordpress.com/2013/11/21/striking-back-against-censorship

Sparas, D. 2013, 18 June. EU regulatory framework for e-commerce. *World Trade Organization Workshop on E-Commerce*. Geneva: World Trade Organization. www.wto.org/english/tratop_e/serv_e/wkshop_june13_e/sparas_e.pdf

Stearns, J. 2013. *Acts of journalism: Defining Press Freedom in the Digital Age*. Washington, DC: Free Press. http://www.freepress.net/resource/105079/acts-journalism-defining-press-freedom-digital-age

Sunstein, C. 2013, December. Deciding by Default. *University of Pennsylvania Law Review*, Vol. 162, No. 1. http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1000&context=penn_law_review

Tuppen, C. 2012. *Opening the Lines: A Call for Transparency from Governments and Telecommunications Companies*. Global Network Initiative. https://globalnetworkinitiative.org/sites/default/files/GNI_OpeningtheLines.pdf

Van Hoboken, J. 2012. *Search Engine Freedom: On the implications of the right to freedom of expression for the legal governance of Web search engines.* PhD thesis, University of Amsterdam Faculty of Law. http://dare.uva.nl/document/357527

Viljoen, F., 2012. *International human rights law in Africa.* Oxford: Oxford University Press

Villeneuve, N. 2006, January. The Filtering Matrix: Integrated mechanisms of information control and the demarcation of borders in cyberspace. *First Monday*, vol. 11. No. 1-2

Waldron, J., 2012. *The Harm in Hate Speech*. Cambridge, MA: Harvard University Press

York, J. C. 2010, September. 'Policing Content in the Quasi-Public Sphere'. OpenNet Initiative. https://opennet.net/policing-content-quasi-public-sphere

Zingales, N. 2013, November. *Internet intermediary liability: Identifying best practices for Africa*. South Africa: Association for Progressive Communications. www.apc.org/en/system/files/APCInternetIntermediaryLiability_BestPracticesAfrica_20131125.pdf

Zittrain, J. 2006, spring. A History of Online Gatekeeping. *Harvard Journal of Law & Technology*, Vol. 19, No. 2, pp. 253-98. http://jolt.law.harvard.edu/articles/pdf/v19/19HarvJLTech253.pdf

World Trends in Freedom of Expression and Media Development – Special Digital Focus 2015 explores emerging opportunities and challenges for press freedom in the digital age. With a focus on online hate speech, protection of journalism sources, the role of internet intermediaries in fostering freedom online, and the safety of journalists, the report highlights the importance of new actors in promoting and protecting freedom of expression online and off-line. In a media environment transformed by digital technologies, this special volume in the World Trends series is a key reference for Governments, journalists, media workers, civil society, the private sector, academics and students.

UNESCO
United Nations
Educational, Scientific and
Cultural Organization

Communication and Information Sector

9 789231 001277