CIGI

CHATHAM HOUSE
The Royal Institute of International Affairs

Global Commission on Internet Governance

ourinternet.org

PAPER SERIES: NO. 36 — MAY 2016

Internet Openness and Fragmentation: Toward Measuring the Economic Effects

Sarah Box

INTERNET OPENNESS AND FRAGMENTATION: TOWARD MEASURING THE ECONOMIC EFFECTS

Sarah Box

CIGI

CHATHAM HOUSE
The Royal Institute of International Affairs

Published by the Centre for International Governance Innovation and Chatham House.

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada.

67 Erb Street West
Waterloo, Ontario N2L 6C2
Canada
tel +1 519 885 2444 fax +1 519 885 5450
www.cigionline.org

CHATHAM HOUSE
The Royal Institute of International Affairs
10 St James's Square
London, England SW1Y 4LE
United Kingdom
tel +44 (0)20 7957 5700 fax +44 (0)20 7957 5710
www.chathamhouse.org

TABLE OF CONTENTS

ABOUT THE GLOBAL COMMISSION ON INTERNET GOVERNANCE

The Global Commission on Internet Governance was established in January 2014 to articulate and advance a strategic vision for the future of Internet governance. The two-year project conducts and supports independent research on Internet-related dimensions of global public policy, culminating in an official commission report that will articulate concrete policy recommendations for the future of Internet governance. These recommendations will address concerns about the stability, interoperability, security and resilience of the Internet ecosystem.

Launched by two independent global think tanks, the Centre for International Governance Innovation (CIGI) and Chatham House, the Global Commission on Internet Governance will help educate the wider public on the most effective ways to promote Internet access, while simultaneously championing the principles of freedom of expression and the free flow of ideas over the Internet.

The Global Commission on Internet Governance will focus on four key themes:

- enhancing governance legitimacy — including regulatory approaches and standards;
- stimulating economic innovation and growth — including critical Internet resources, infrastructure and competition policy;
- ensuring human rights online — including establishing the principle of technological neutrality for human rights, privacy and free expression; and
- avoiding systemic risk — including establishing norms regarding state conduct, cybercrime cooperation and non-proliferation, confidence-building measures and disarmament issues.

The goal of the Global Commission on Internet Governance is two-fold. First, it will encourage globally inclusive public discussions on the future of Internet governance. Second, through its comprehensive policy-oriented report, and the subsequent promotion of this final report, the Global Commission on Internet Governance will communicate its findings with senior stakeholders at key Internet governance events.

www.ourinternet.org

ABOUT THE AUTHOR

Sarah Box is counsellor to the directors in the Organisation for Economic Co-operation and Development (OECD) Directorate for Science, Technology and Innovation (DSTI). At the time of writing she was working in the Digital Economy Policy Division on issues of Internet openness. Since starting in DSTI in 2007 Sarah has worked on a variety of issues including human resources for science and technology, the OECD Innovation Strategy, public research organizations and the shipbuilding industry. Prior to joining the OECD, she worked as a senior research economist at the Australian Productivity Commission, and as an economist at the New Zealand Treasury

ACRONYMS

BSA	The Software Alliance
ccTLDs	country-code top-level domains
CDNs	content distribution networks
gTLD	generic top-level domain
GVCs	global value chains
ICT	information and communication technology
IP	Internet Protocol
ISP	Internet service provider
IXP	Internet exchange point
Mbps	megabits per second
OECD	Organisation for Economic Co-operation and Development
R&D	research and development
SMEs	small and medium-sized enterprises
TLDs	top-level domains

EXECUTIVE SUMMARY

The global free flow of data underpins opportunities for economic and social growth and is the essence of the concept of Internet openness. Internet openness facilitates international trade by easing communications between suppliers and customers, improving logistics and enabling formerly excluded firms to enter global value chains (GVCs). It also spurs innovation and entrepreneurship through its role as a knowledge-sharing hub; as a platform for launching new business ideas; as a place for entrepreneurs to find financing, services and marketplaces; and as part of the information and communication technology (ICT) constellation that is innovative in its own right.

But Internet openness is not indisputably "good" and fragmentation (or restricted data flows) is not indisputably "bad." To better understand the relationship between Internet openness and economic performance, the Organisation for Economic Co-operation and Development (OECD) began collecting and using company data to provide a new perspective on global data flows across the Internet and their effects on indicators such as trade, innovation and entrepreneurship. Initial findings underscore the highly interconnected nature of today's Internet, showing users are increasingly accessing content outside their countries over infrastructure with strong global interlinkages. Future work should focus on building a global data flow data set and highlighting data hotspots to help identify economic impacts.

INTRODUCTION

Internet openness and Internet fragmentation are often portrayed as opposing forces struggling for ascendancy. If Internet openness wins, we have a world of global connections and freedoms. If Internet fragmentation wins, we have a world of silos and closed doors. This kind of scenario implies significant economic consequences and people understandably want to know what exactly, and how large, these consequences might be.

Through its work on the economic and social benefits of Internet openness, the OECD is attempting to bring new evidence to the debate. This is important because, in reality, the issue is not black and white: openness is not indisputably good and fragmentation is not indisputably bad. Governments need more nuanced information to allow policy choices that optimize the benefits of Internet openness while addressing valid concerns for digital security and privacy. Progress must be made in understanding the strength and direction of the relationship between Internet openness and governments' ultimate economic goals — such as enhanced trade, innovation and entrepreneurship — and how Internet openness itself is affected by policy and private sector actions.

Analysis of Internet openness quickly meets a practical stumbling block: how do we measure it — or, indeed, measure Internet fragmentation — when the concept of Internet openness itself is so broad, encompassing technical, economic, political and societal aspects? To make headway, the OECD chose to focus efforts on better understanding and measuring global data flows on the Internet, as an initial indicator of Internet openness. From this starting point, it has begun building a picture of global data flows and laying out a path for future analyses (OECD, forthcoming 2016). This paper presents an excerpt of that work.[1] It describes the benefits of Internet openness for international trade, innovation and entrepreneurship, and presents initial steps to better measure the global data flows enabled by Internet openness.[2]

INTERNET OPENNESS AND INTERNATIONAL TRADE

There is a growing literature on the positive effects of the Internet on trade and the potential costs of policies (notably on data localization) that introduce frictions to

1 The OECD's work on Internet openness is being undertaken in the context of the 2016 OECD Ministerial Meeting on the Digital Economy: Innovation, Growth and Social Prosperity, to be held in Mexico in June. See www.oecd.org/sti/dep-ministerial-2016.htm. The Ministerial has four central themes: Internet Openness and Innovation; Building Global Connectivity; Trust in the Digital Economy; and Jobs and Skills in the Digital Economy.

2 This paper should be read in conjunction with GCIG Paper No. 35, *A Framework for Understanding Internet Openess*, by Jeremy West.

"business as usual" data flows on the Internet. Internet openness facilitates international trade for existing businesses by making it easier for the supplier to connect with existing consumers who are located beyond the borders of the supplier's home country (or countries) and by improving logistics control. Openness can also boost trade by providing access to a wider customer base via e-commerce. And it enables new firms to enter more geographic markets and, for the most efficient ones, to enter GVCs. At the same time, Internet openness and digitization make it possible to complete transactions and deliver products, services and payments faster and more efficiently by replacing some physical trade with online trade — for example, in books and music, or with more complex products via online shipment of designs followed by local production.

GVCs are central to the trade and Internet story. Behind aggregate trade data lie a huge number of intermediate trade flows, with inputs sourced globally and stages of production shifting from location to location to complete a final product. Both goods and services may be produced in GVCs — electronics and cars are common examples where design, raw material, production and marketing inputs are spread across countries, but aircraft, clothing, film animation, law briefs and medical advice are also created in GVCs. The rise of GVCs has been made possible in part by technological advances, notably the information management systems that allow firms to coordinate their participation in GVCs. The combination of GVCs and the Internet has not only enabled firms in developing countries to more easily engage in international trade (by specializing in one stage of a chain, such as auto electronics), but also through the use of digital platforms provided by small and medium-sized enterprises (SMEs) to enable even tiny firms (micro-multinationals[3]) to connect with global suppliers and purchasers.

Seamlessly moving potentially large amounts of data across countries is an essential part of supporting intermediate and final trade flows and allowing firms to participate in GVCs. In other words, given the pervasiveness of GVCs, reductions in Internet openness could create significant impediments to trade. Small frictions may multiply into large barriers, especially if production is split into stages that entail numerous border crossings where imposed frictions multiply. The Swedish National Board of Trade (2015, 14-15) suggests that policies such as data localization requirements could lead a firm to reorganize its GVC, either moving or closing parts of its operations, with service to end-users being restricted in some cases. Stephen Ezell, Robert D. Atkinson and Michelle Wein (2013, 46-47) make a similar point, noting that localization barriers to trade, including restrictions on data, undermine firms' ability to participate in global networks because the barriers raise costs and reduce technology diffusion. The Software Alliance, more commonly known as the BSA, additionally highlights the trade-dampening effect of country-specific technology standards and other forms of "digital protectionism," such as nationally oriented information technology procurement (BSA 2014).

Internet openness is especially important for enabling smaller firms to engage in international trade. Jessica R. Nicholson and Ryan Noonan (2014, 8) comment that while localization requirements can make cross-border trade difficult for large companies, they may make it "practically impossible for small businesses that cannot afford to implement separate systems and standards in every country in which they do business." Moreover, these firm-level impacts can sum to significant negative outcomes for countries. James M. Kaplan and Kayvan Rowshankish (2015) note that as banks reduce their operations in countries with more stringent data regulations, financial services will grow more slowly, with potentially adverse consequences for development. There are also more general concerns that policies enacted to reduce Internet openness could create a "slippery slope" for additional interventions and possibly non-tariff barriers, such as local content requirements or efforts to promote "indigenous innovation" via intellectual property right restrictions. Ezell, Atkinson and Wein (2013, 38) see a risk that the contravention of the rules and spirit of the global trading system would lead to a decay where "every country is incentivized to cheat, the competition becomes cutthroat, and the global economy suffers."

INTERNET OPENNESS, INNOVATION AND ENTREPRENEURSHIP

The Internet, as a connector on a massive scale, provides the opportunity to share, access and coordinate knowledge in ways previously not possible. Knowledge sharing was the impetus behind the creation of the Internet, albeit among an initially small group of research institutions, and research-oriented knowledge-sharing networks running on the Internet remain. These help facilitate collaborative research on a global scale, with publications, patents, researchers, and academic and research institutions taking on international dimensions and drawing benefits from cross-border knowledge flows. Firms, too, leverage the Internet to share knowledge, from multinationals with diverse research and development (R&D) and production locations to small firms tapping into local universities and research institutions. And the general expansion of access to knowledge (for example, via Google searches,

3 The term micro-multinational is not well defined and should not be automatically equated with small multinational enterprises. Micro-multinationals may simply be small exporters, whereas multinational enterprises typically comprise "companies or other entities established in more than one country and so linked that they may co-ordinate their operations in various ways" (OECD 2008, 12). Ann Mettler and Anthony D. Williams (2011) discuss micro-multinationals in terms of start-ups, typically small, service-driven companies.

Wikipedia, YouTube or online education sites) to a broader range of people can also stimulate innovation. Joshua Meltzer (2015, 92) states:

> The Internet has provided an opportunity for people to connect and share ideas in a space and time essentially free of transaction costs. Significantly, it has been the open nature of the Internet — the freedom to connect, share information and exchange ideas — that has underpinned the innovation which has created new businesses such as those based on social networking and crowd funding.

The Internet also provides a platform for innovation, open to anyone who wishes to leverage it for their own venture. Several aspects of this are frequently mentioned — first, that the Internet enables "innovation at the edges"; second, that it enables "serendipitous" (or unexpected) innovation; and third, that it allows "permissionless" innovation. The term "innovation at the edges" references the Internet's end-to-end design principle, whereby the core network provides general purpose system services (sending packets of data) and is indifferent to the various applications that may be implemented in software on computers attached to the "edge" of the Internet (Blumenthal and Clark 2001). This end-to-end feature makes the Internet flexible, general and open to innovative new applications. These innovations can challenge the status quo and can bubble up from unexpected quarters (hence the idea of serendipity), including from very small firms. Finally, permissionless innovation captures the idea that market entrants need not seek approval prior to launching lawful new services, and that this lack of gatekeeping leads to a flourishing market for ideas, be it through social networks or through promoting innovation around new devices and services. Leslie Daigle (2015, 9) points to the creative destruction built into the Internet, saying, "Systemically, the Internet supports and fosters approaches that are useful; old, outdated or otherwise outmoded technologies die away."

As a source of inputs to entrepreneurs and established firms, the Internet is also becoming increasingly valuable, offering a conduit to finance, services and marketplaces. In a way, the Internet is taking outsourcing to its extreme, allowing firms to fully concentrate on their competitive advantage. This not only benefits existing firms by improving efficiency and providing headspace for new innovative activities, but also makes it easier for entrepreneurs to muster the resources to take their ideas through to commercialization. The new phenomenon of micro-multinationals, for instance, is underpinned by the availability of business services via Internet platforms (Mettler and Williams 2011), and SMEs can also reap significant rewards from boosting their digital savvy (Mettler and Williams 2012). Firms can design, develop and deliver their products and services worldwide thanks to Internet-based crowd financing, digital utilities, professional services, micro-manufacturing, innovation marketplaces and e-commerce platforms.

Lastly, the ICT sector itself is a generator of innovation, offering increased computer power and performance and new tools. This sector forms part of the economic constellation around the Internet and both nourishes and feeds off the economic and social activity enabled by the Internet. The ICT sector was relatively resilient to the 2007–2009 global economic crisis, although it has yet to regain its pre-crisis levels in some countries, and is an important venue for R&D and patenting. Advances in ICT will underpin data-driven innovation — for instance, the main enablers of the Internet of Things are big data, the cloud, machine-to-machine communication and sensors (OECD 2015a, 244).

But all this relies crucially on Internet openness — free flows of data and information, accompanied by trust in the network, are essential for the Internet to contribute to innovation and entrepreneurship. In a recent study, young entrepreneurs in Group of Twenty countries identified international mobility of data accompanied by adequate protection of personal data as a key issue, saying that this was "one of the success factors of entrepreneurs who develop international businesses, and a critical element for entrepreneurs to get access to the right data" (Accenture 2013, 36). Commentators have argued that innovation in industries such as ICT, energy, life sciences, aerospace and scientific instruments could be especially impeded by limits to data mobility, since such industries do best serving large markets in a competitive environment (Ezell, Atkinson and Wein 2013). Limiting scale economies enables weaker firms to remain in the market, thus reducing returns to more efficient firms and eroding their ability to invest in innovation. At the same time, security and privacy standards are necessary to support innovation on the Internet; for example, in Estonia, the X-Road data exchange framework enables access to publicly held data in a high-trust environment and has spawned the development of numerous new Internet businesses, including Skype (Hofheinz and Mandel 2014).

MEASURING INTERNET OPENNESS

Specific studies on Internet openness are still scarce and there is much scope for improving quantitative evidence on the links between Internet openness and economic indicators such as trade and innovation. But, as noted earlier, the concept of Internet openness is so broad that measurement is a significant challenge.

Existing studies of the Internet's macroeconomic impact have typically used various proxies of Internet presence, including adoption indicators (such as broadband penetration rates), economic indicators (such as network investment) and technical indicators (such as Internet

Protocol [IP] addresses per capita). Each of these proxies has limitations, one being lack of insight into how people, firms, industries or regions actually make use of the Internet (OECD 2012). Unfortunately, these proxies are also imperfect measures of Internet openness, as they essentially focus on access and availability.

Quantitative studies of the Internet from a digital trade angle have typically used proxies of data flows for their analysis. On the face of it, using data flow information as a measure of Internet openness has merit. If the essence of the Internet is to facilitate movement of data/information/knowledge, for whatever purpose, then measuring flows of data could shed light on current levels of openness, even if the economic value of the data flows is unknown. Changes in flows could then be related to changes in trade and other variables on the one hand, and changes in policy or other factors on the other hand (assuming we could construct robust policy indicators). In addition, as many of the risks to Internet openness are occurring at the level of data flows, measuring this aspect would be highly relevant.

However, the proxies of data flows used to date also have drawbacks:

- As Paul Hofheinz and Michael Mandel (2015) point out, using official statistics (such as trade data related to digital activity) essentially underestimates the size of cross-border data flows, because not all flows are monetized.
- While looking at the bits and bytes themselves is another option, information on the capacity of the infrastructure (such as TeleGeography statistics [McKinsey Global Institute 2014]) does not inform us of actual data flows.
- Adding capacity usage estimates or traffic estimates can bring us closer to actual data flows, but such estimates (for instance, Cisco global IP traffic forecasts [Hofheinz and Mandel 2014; 2015][4]) do not differentiate where the traffic is coming from or going to — i.e., whether start and end points are local or cross-border — or the type of flows.

In one of the few studies that have approached Internet openness more directly, Dalberg (2014) chose to use Freedom House's Freedom on the Net index to look at the economic benefits of Internet openness. This index is based on qualitative assessments and surveys, and measures the level of Internet and digital media freedom in three areas: obstacles to access (such as regulatory obstacles for Internet service providers [ISPs]); limits on content (for example, instances of filtering); and violations of user rights (such as state surveillance). However, Dalberg considered that the limited time series and country coverage did not allow statistically significant causal relationships to be established; indeed, one of its key conclusions was to urge stakeholders "to establish standard and universally measurable indicators of Internet openness" (ibid., 50).

Other efforts are emerging along the lines of the Freedom on the Net index that group together various indicators of Internet activity, including aspects that touch on Internet openness. For instance, the Boston Consulting Group's e-Friction Index agglomerates 55 indicators to indicate the ease with which people can participate in the Internet economy (Zwillenberg, Field and Dean 2014). The e-Friction Index could perhaps be interpreted as an openness index, although some of the indicators (such as company-level technology absorption or financing through local equity market) are relatively upstream from practical Internet openness; furthermore, there are significant data gaps. Another effort to draw together a variety of indicators on Internet trends comes from the Berkman Center for Internet & Society, whose Internet Monitor research project aims to shed light on Internet content controls and Internet activity worldwide.[5] As well as an index related to Internet access and infrastructure, a "dashboard" was recently launched that incorporates data on traffic, cyber attacks and website availability, among other indicators.

However, it remains the fact that there is no easy off-the-shelf solution to measuring Internet openness. As such, one goal of the OECD's work is to push the data boundaries by collecting and using data obtained from companies with global reach to provide a new perspective on global data flows across the Internet. Eventually, this work should facilitate analyzing the effects of Internet openness at a more general level than is found in case studies of individual firms or situations, and thus should help reinforce the evidence base available to policy makers.

At the time of writing, the OECD had analyzed aggregate information related to Google searches and YouTube views (see Box 1). Google and YouTube usage provide insight into the website domains that users in a country visit via Internet search, and where YouTube content is watched. While the information does not give a sense of volumes (as it was expressed in percentages), some 240 countries are covered in the tables the OECD analyzed, enabling the exploration of interlinkages. At this stage the analysis has mainly focused on OECD countries plus its key partners

4 Hofheinz and Mandel's (2015) concept of "digital density" (the amount of data used per capita in an economy) as a proxy of data usage is based on Cisco IP traffic forecasts for major countries, which are built on a series of estimates of user numbers, adoption rates, minutes of usage and bitrates to obtain a per-month traffic estimate (Cisco 2015a; 2015b). Hofheinz and Mandel (2015) acknowledge that using this as a proxy for consumption of cross-border data flows is a leap, but propose this measure gets closer to data usage than other measures of cross-border data flows.

5 See https://thenetmonitor.org/ for further details on the Berkman Center's initiative.

Box 1: Google Data Specifications

The OECD analyzed four tables of information, related to Google searches and YouTube watch time, as follows:

Source 1: Google Search — Focus on User Country

A table of 240 countries[1] (including 1 "zz" category where the country of the user could not be determined) by 101 top-level domains (TLDs — comprising 87 country domains, 13 generic domains and 1 "other" category), showing the percentage of clicks on search results by users of a particular country searching on Google (all domains) that landed on websites of each TLD.[2] This allows us to see, for instance, that in 2014, five TLDs (.com, .au, .org, .net and .uk) accounted for 96.11 percent of Australian users' Google search result clicks, with the remaining 3.89 percent of clicks going to a variety of landing page TLDs. User locations were based on IP addresses.

Time span: 2007–2014 (eight years) for most countries in the table.

Source 2: Google Search — Focus on Landing Page TLD

A table of 240 countries (including 1 "zz" category where the country of the user could not be determined) by the same 101 TLDs, showing the percentage of clicks on search results related to each landing page TLD that come from users of a particular country who are searching on Google (all domains).[3] This allows us to see, for instance, that in 2014, 25.35 percent of clicks received by .com landing page domains via Google search results came from users in the United States. User locations were based on IP addresses.

Time span: 2007–2014 (eight years) for most countries in the table.

Source 3: YouTube — Focus on Country of Uploader

A table of 240 uploading countries by 240 watching countries, allocating the percentage share of watch hours of an uploading country's YouTube videos across each watching country. There is additionally a "zz" category where the countries of uploading user and watcher could not be determined.[4] This allows us to see, for instance, that in 2014, 18.23 percent of the watch hours for videos uploaded by users from Spain were by users located in Mexico — the second-highest watch hour share after Spanish viewers (at 23.44 percent). The locations of uploading users were user-specified, and those of watching users were based on IP addresses.

Time span: 2010–2014 (five years) for most countries in the table.

Source 4: YouTube — Focus on Watching Country

A table of 241 watching countries by 250 uploading countries (each including a "zz" category where the countries of uploading user and watcher could not be determined), allocating the percentage share of a country's YouTube watch hours across different YouTube video uploading countries.[5] This allows us to see, for instance, that in 2014, Slovenian users spent 1.61 percent of their YouTube watch hours on videos uploaded by users in Italy. The locations of uploading users were user-specified, and those of watching users were based on IP addresses.

Time span: 2010–2014 (five years) for most countries in the table.

1. References to "country" should be read to include all geographic areas with two-digit country code top-level domains (ccTLDs) in the tables. These include the 193 member states of the United Nations as well as other territories.

2. As the information is in percentages, it is not possible to say how large the "zz" user category is compared to other user countries. However, the share of user clicks going to the "other" category are typically small; for instance, for all OECD key partner and accession countries, except for Luxembourg, the shares of user clicks going to the "other" category are less than one percent. In Luxembourg's case, 13–17 percent of clicks went to "other" over the 2007–2014 sample period.

3. In this table, it is not possible to say how large the "other" category is compared to the other TLDs, but we can see that the "zz" user category makes up less than one percent of clicks on TLDs in the majority (84 percent) of cases. Over the eight-year period, .co (Colombia), .id (Indonesia), .in (India), .ir (Islamic Republic of Iran), .pk (Pakistan), .sa (Saudi Arabia) and "other" saw the most frequent incidences of a high "zz" user share.

4. In this table, it is not possible to say how large the "zz" category is as an uploading country, but we can see that "zz" as a watcher accounts for less than one percent of watch hours for any country's YouTube videos in the majority (94 percent) of cases, with this share typically decreasing over the sample period. The most frequent incidences of a high "zz" watcher share were for .al (Albania), .ir (Iran), .mc (Monaco) and .mk (Former Yugoslav Republic of Macedonia).

5. In this table, it is not possible to say how large the "zz" category is as a watcher country, but we can see that "zz" as an uploader has accounted for a steadily decreasing share of each country's watch hours over the sample period. In 2010, the share of watch hours going to "zz" YouTube videos reached 15 percent in some cases (Iran and Japan), but by 2014, the share was below or close to one percent in all cases.

and accession countries (Brazil, the People's Republic of China, Colombia, Costa Rica, India, Indonesia, Latvia, Lithuania, Russia and South Africa).

Key findings and lessons from the information analyzed are highlighted below. In interpreting the results, it is important to bear in mind the following factors:

- A ccTLD for a website does not necessarily imply that the content is hosted within that country. For instance, you do not need to be based in New Zealand to register a .nz domain name, and the domain name is not required to be hosted in New Zealand.[6]
- Indeed, some ccTLDs have no substantive linkage to the country at all and instead are used much like a generic top-level domain (gTLD). Examples include Belize (.bz), the Cocos (Keeling) Islands (.cc), the Federated States of Micronesia (.fm), Lao People's Democratic Republic (.la), Montenegro (.me), Niue (.nu), Samoa (.ws), Sint Maarten (.sx), Tokelau (.tk), Tonga (.to) and Tuvalu (.tv).
- A gTLD for a website cannot be matched to a particular country, either in terms of "owner" of the site or where the content is hosted, as these domains are available for registration by Internet users worldwide (albeit with some restrictions for some domains[7]).
- The network architecture of the Internet, the extensive use of data centres ("the cloud") and the growing presence of content distribution networks (CDNs) mean that the physical route taken by data may bear little resemblance to a straightforward bilateral flow between two countries.

Insights from Information on Google Searches

The Google search information from source 1 in Box 1 shows that Internet users differ widely in the extent to which they select results in their own country's domain. For instance, in 2014, 67 percent of Google search clicks by users in Poland led to .pl domains, whereas only 13 percent of search clicks by users in Korea led to .kr domains (see Figure 1). The United States is an exceptional case; for historical reasons, gTLDs such as .com were preferred to the .us domain, which was commercially marketed at a later stage, and just 0.66 percent of US users' Google search clicks went to .us websites in 2014.

Accompanying this diversity is an almost uniform trend of users increasingly accessing content outside their countries. With the exception of Canada, Estonia, France, India, Ireland and Sweden, all countries experienced a decline in the share of Google search clicks going to their own ccTLD between 2007 and 2014. These findings might suggest a geographically wider variety of content being accessed, increased cross-country information and knowledge exchange, and potentially an increase in actual cross-border data flows, subject to the caveats mentioned earlier.

The extent to which these patterns are accompanied by changes to Google's search algorithms is an interesting question. Google's algorithms rely on over 200 "signals" to help guess what the user might be looking for in their search, including terms on websites, content freshness, the user's region and PageRank (a measure of how authoritative a webpage is).[8] An increased internationalization of the content accessed by Google users could reflect many factors and developments. It is possible that the queries issued by users over time relate to more international topics (i.e., a change in the "query mix"), thereby leading to more international results surfacing. Even for an unchanged query mix, it is possible that users over time become more interested in international sources, seeking them out in search results; this could potentially be accompanied by Google's algorithms taking account of this preference in the composition of search results. In addition, the shape of the underlying Internet is ever-changing, and to the extent that the growing number of web pages "internationalize" this base, one would expect this change to be reflected in Google's index as well. Irrespective of the precise explanation, the fact remains that many users are increasingly looking beyond their own country content.

The information on gTLDs show that a significant share of users' search clicks go to sites with a .com domain. In fact, in every country, .com domains were the most or second-most common result click, along with the country's ccTLD (with the exception of China, Korea, Luxembourg and the United States, where the .com domain was accompanied by .hk, .net, "other" and .org, respectively, in the top two). Thirteen gTLDs were included in the Google search information — .com, .org, .net, .edu, .info, .gov, .biz, .cat, .mobi, .xxx, .mil, .name, .int — with .com, .net and .org uniformly the top three gTLD clicks and cumulatively accounting for over 50 percent of search result clicks in 27 of the countries in 2014 (see Figure 2).

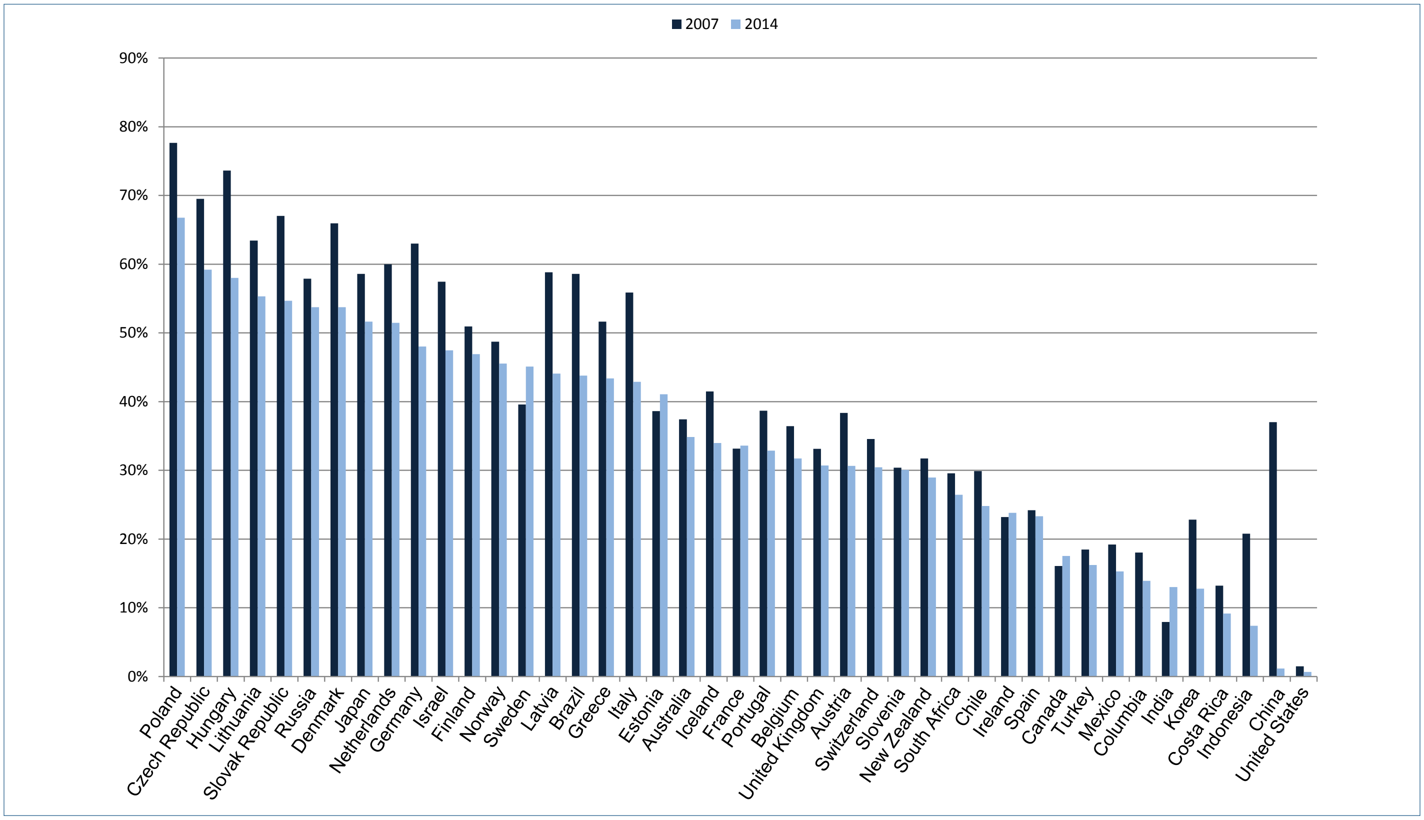
The importance of language/culture and geographic proximity can be observed in the search information. Proximate countries and those with a common language are typically among the top 10 ccTLDs in a country's search result clicks. For example, Chilean users click on

6 See more information at the .nz Domain Name Commission at https://dnc.org.nz/the-commission/faq.

7 See the Internet Corporation for Assigned Names and Numbers' list of TLDs and registrars at www.icann.org/registrar-reports/accredited-list.html.

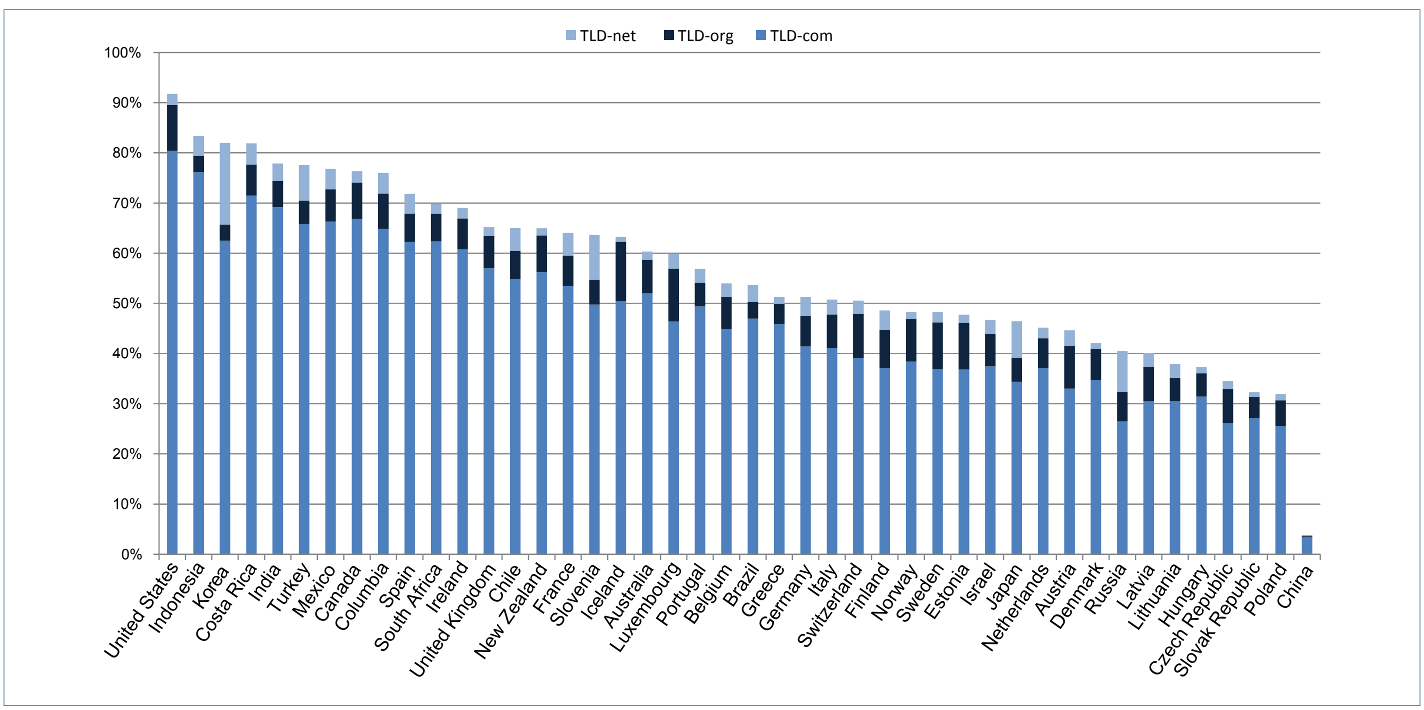
8 See www.google.com/insidesearch/howsearchworks/algorithms.html.

Figure 1: Share of Google Search Result Clicks Leading to Sites with Own ccTLD (2007 and 2014)



Source: OECD calculations, based on information from source 1 (see Box 1).
Note: Data on Luxembourg (.lu) as a search domain was available in the table.

Figure 2: Share of .com, .org and .net in Search Result Clicks by Country (2014)



Source: OECD calculations, based on information from source 1 (see Box 1).

results in the Spanish, Argentinian, Mexican, Colombian, Peruvian and American ccTLD spaces, while Swiss users click on results in the German, French, Italian, UK and Austrian ccTLD spaces. This behaviour is consistent with international trade models for goods and services that show that "gravity" — as measured by proximity, common language and so on — is an important factor driving trade links, although there may also be other effects in operation.

At the same time, the usage of the generic ccTLDs is also notable. While Tonga and Tuvalu might seem logical search result clicks for users in Australia and New Zealand — Pacific neighbours and home to immigrant communities — it is less obviously the case for Estonia and Israel, and the widespread appearance of these generic ccTLDs in top 10 search result click lists underscores the lack of a one-to-one relationship between ccTLDs and their "home countries." For instance, Tuvalu's ccTLD is often used by media companies (the .tv domain name having clear marketing value). Nevertheless, the share of total search result clicks received by such TLDs is typically small since, as clearly illustrated in Figure 2, gTLDs account for a significant share of total user clicks.

The Google search information from source 2 in Box 1 suggests that most website ccTLDs have a highly concentrated user base, accompanied by a long tail of user countries, each with tiny shares of total search result clicks. Taking the full sample of ccTLDs included in the table (excluding those that are clearly used in practice as gTLDs), 41 of 75 ccTLDs received 95 percent of search result clicks from four or fewer user countries in 2014. These were typically the country of the ccTLD plus proximate countries (either geographically or via cultural/language similarities). For instance, users from Israel and the United States accounted for over 95 percent of search result clicks to websites with Israel's ccTLD (.il), while users from South Africa, the United States and the Netherlands accounted for over 95 percent of search result clicks to websites with South Africa's ccTLD (.za). Most OECD countries received 95 percent of search result clicks from six or fewer user countries.

However, some ccTLDs have lower levels of concentration, although still with the long tail. OECD countries that stand out in this respect include Spain (12 user countries accounted for 95 percent of search result clicks in 2014), as well as Sweden, the United Kingdom, the United States and Iceland (20, 21, 27 and 50 user countries, respectively). Mexico and Colombia accounted for a significant share of Google search result clicks to websites with Spain's ccTLD (.es), followed by a number of other South American countries, plus the United States, Germany and India. The wide range of user countries behind search result clicks to websites with the United Kingdom ccTLD (.uk) is perhaps reflective of the United Kingdom's historic Commonwealth links as well as its status as a global hub.

The user base of gTLDs is unsurprisingly less concentrated than that of ccTLDs, matching their greater global availability. But one interesting observation is the variety of user countries for the gTLD .edu, which is available only to US post-secondary institutions that are accredited by an agency on the US Department of Education's list of Nationally Recognized Accrediting Agencies.[9] The Google search result clicks could be interpreted as mirroring the international attractiveness of the United States as an education destination. Users from the United States accounted for almost 71 percent of search result clicks to .edu domains in 2014; users from 27 other countries (shown in Figure 3) then accounted for a further 24 percent of the clicks.

Insights from Information on YouTube Watch Hours

YouTube is a platform for user-generated video content, from music to do-it-yourself bicycle repairs, from professional to amateur. It has been credited as a source of ideas and cross-fertilization.[10] The YouTube information in sources 3 and 4 (see Box 1) do not distinguish between types of content, but they do provide an aggregated picture of the viewing patterns of YouTube users.

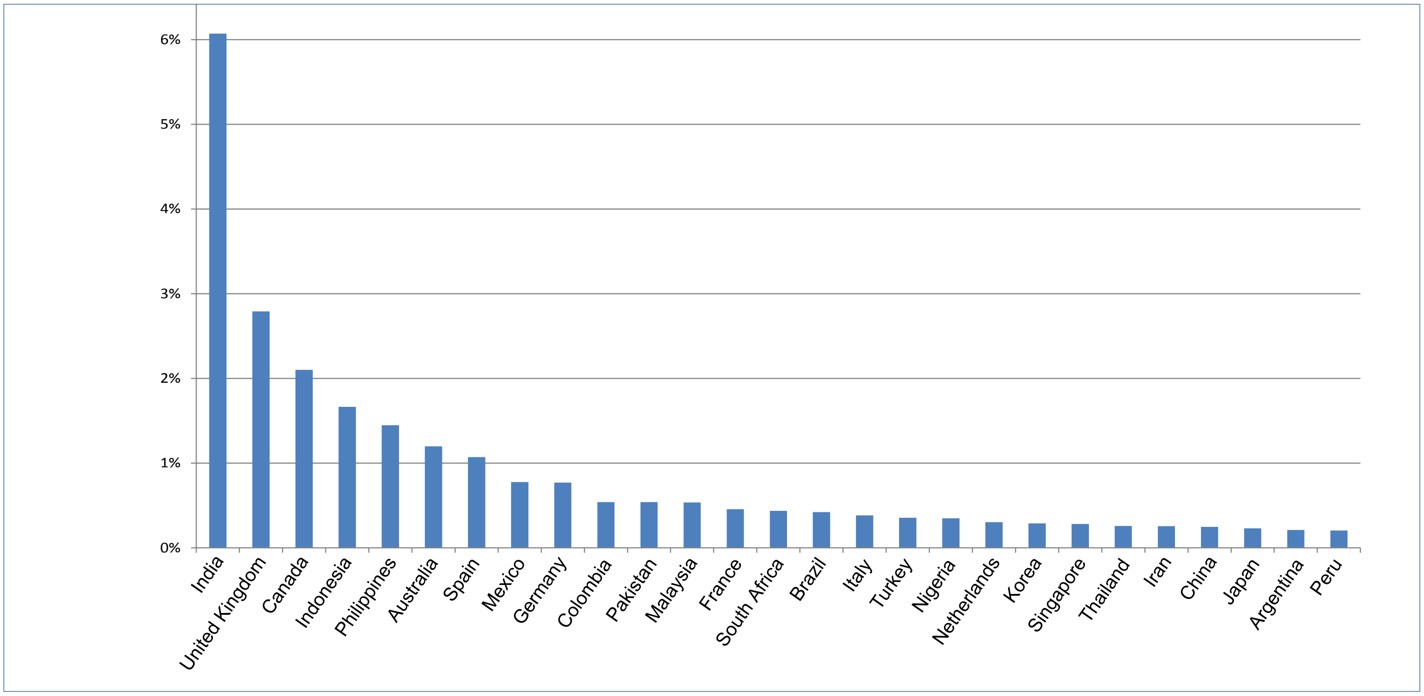
Figure 4 shows a wide variation in the extent to which content is viewed outside the country in which it is uploaded. In 2014, for instance, 85 percent of the watch hours for videos uploaded by users in Japan were from users located in Japan. Toward the other end of the scale, just eight percent of the watch hours for videos uploaded by users in Australia and Canada were from users located in those countries. For both Australia and Canada, users in the United States accounted for the largest share of watch hours for Australian- and Canadian-uploaded content (27 and 37 percent, respectively). US users were the second-largest share of viewers of Japanese YouTube content, with almost three percent of watch hours.

Figure 4 also shows how, for more than half of the examined countries, dispersion of content is becoming increasingly international. In the United States, for example, the share of watch hours for US-uploaded content accounted for by US users fell from 42 percent to 35 percent over the period 2010–2014. After US users, the top watchers of US-uploaded YouTube content in 2014 were the United Kingdom, Vietnam, Mexico, Canada, Russia, Japan, Australia, Brazil, Germany and Turkey, in that order. In contrast, Japan, Brazil, Turkey and others saw an increase in the share of local watchers in watch time for their content between

9 The .edu domain's sole registrar is Educause, an association for information technology in higher education. Eligibility for the .edu domain name is restricted. See http://net.educause.edu/edudomain/eligibility.asp.

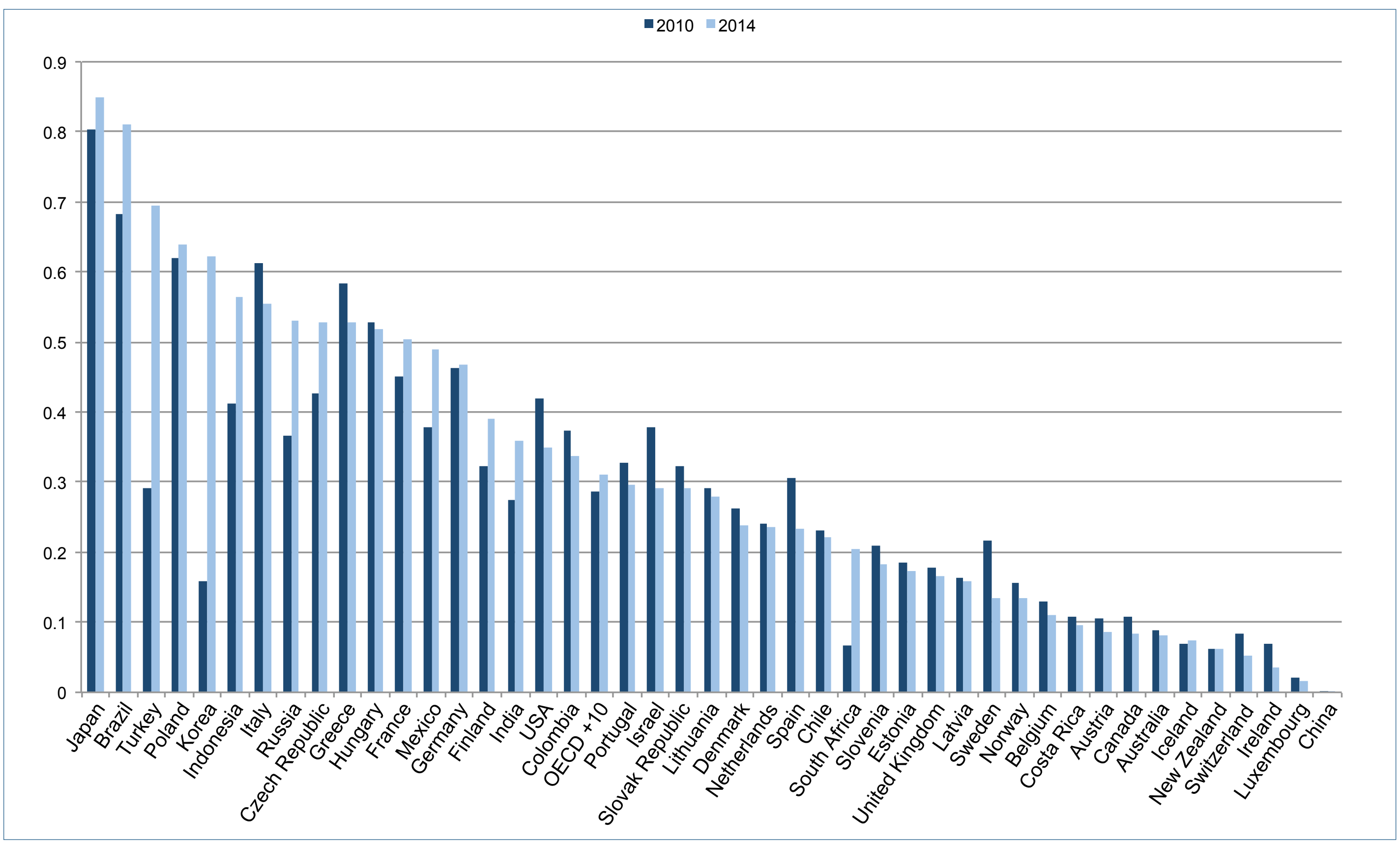
10 See, for example, McNeil (2013).

Figure 3: Top Users of .edu gTLD, Measured by Share of Search Result Clicks (2014)



Source: OECD calculations, based on information from source 2 (see Box 1).
Note: The United States is excluded from this figure. US users plus the 27 countries in the chart accounted for 95 percent of Google search result clicks to websites with a .edu TLD .

Figure 4: Views of YouTube Content Uploaded by Users in Own Country (% of Total Watch Hours for Country's Uploaded Content)



Source: OECD calculations, based on information from source 3 (see Box 1).

2010 and 2014. In some instances, this may be because the amount of local content being produced is increasing and attracting new local users; this, in turn, may be related to the penetration of smart phones, which offer another way to capture and view content.

The range of countries among watchers of a country's content sometimes points to the importance of a common language. For instance, YouTube content uploaded by Spanish users in 2014 obtained its highest share of watch hours from local viewers (23.4 percent), followed by Mexico (18.2 percent), Argentina (9.1 percent), the United States (6.1 percent), Chile (5.6 percent), Colombia (5.3 percent), Peru (3.6 percent), Venezuela (2.5 percent) and Ecuador (2.4 percent). Proximity and historical links can also be observed — in France, for instance, the highest share of watch hours of content uploaded by French users in 2014 came from France (50.5 percent), followed by the United States (5.1 percent), Belgium (4.3 percent), Canada (3.0 percent), Morocco (2.6 percent) and Algeria (2.0 percent).

Focusing on what people watch, source 4 shows that for the most part, the share of any country's watch hours spent on another country's YouTube content is numerically small (i.e., less than one percent), implying that in aggregate, people are taking a smorgasbord approach — a little bit of lots of things. However, there are three instances where this is not the case:

- All watching countries spent 10 percent or more of their watch hours on US-uploaded content, with 20 countries spending more than 50 percent of their watch hours on US-uploaded content (aside from the United States itself, these were Caribbean island nations plus Antarctica,[11] Bermuda, the Marshall Islands and several US island territories).
- Some countries' consumption of local content accounts for very high shares (over 50 percent) of total watch hours. Brazil stands out as a large consumer of its own content — 72 percent of its watch hours are on Brazil-uploaded content. Indian users also spend more than half their watch hours on local content (almost 58 percent). Other countries in this category are Japan (65 percent), Korea (62 percent), Poland (55 percent) and Thailand (66 percent).
- Certain countries' content more regularly accounts for a high share of watch hours in other countries. Spain, France and the United Kingdom stand out, with their content accounting for 10–50 percent of a relatively large number of watching countries' total watch hours (20, 38 and 45 countries, respectively). There are clear language and historical links — for instance, the countries for which French content accounts for 10–50 percent of watch hours are Algeria, Belgium, Burkina Faso, Burundi, Benin, Cameroon, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, French Guiana, French Polynesia, Gabon, Guadeloupe, Guinea, Haiti, Luxembourg, Madagascar, Mali, Martinique, Mauritania, Mauritius, Mayotte, Monaco, Morocco, New Caledonia, Niger, Réunion, Saint Pierre and Miquelon, Senegal, Switzerland, Togo, Tunisia, and Wallis and Futuna (as well as France itself).

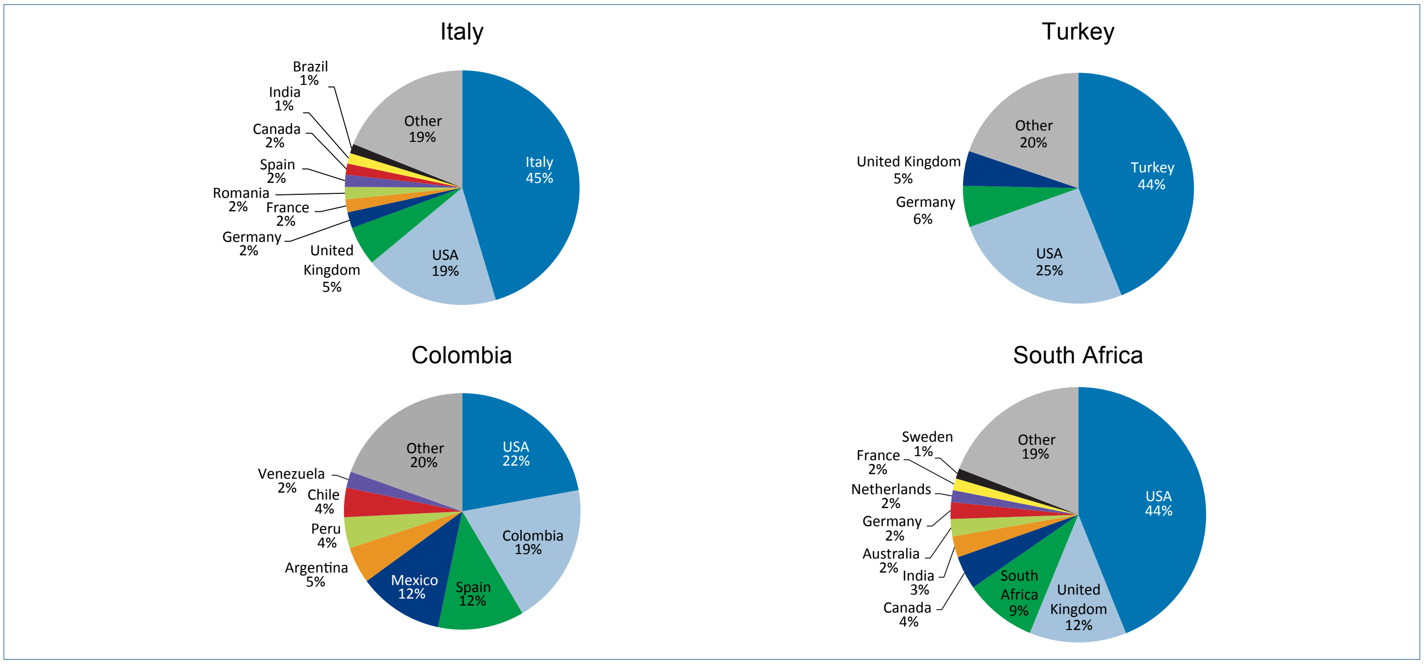
Aside from these patterns, there are also some individual cases that stand out. For instance, Indian content accounts for more than 10 percent of watch hours in several Middle Eastern countries (for example, 24 percent of watch hours for the United Arab Emirates, 15 percent for Bahrain, 12 percent for Kuwait, 15 percent for Oman and 22 percent for Qatar). Fijian users also spend a significant share of watch hours on Indian content (26 percent). This may be due to past and recent immigration patterns that have created significant Indian communities in these countries and/or to the creation of content in India that particularly appeals to Middle Eastern users.

Given the factors above, a country's watch hours typically display a long-tailed pattern, much like that of the earlier information on Google searches, where most watch hours are dedicated to content from a small group of countries and the remainder of watch hours are accounted for by small amounts of many countries' content. Four country examples are presented in Figure 5. In each case, the pie chart specifies the uploading countries (in descending order of importance) that together account for 80 percent of watch hours, with the remainder of sources aggregated as "other." It shows that for Italy, 10 countries accounted for around 80 percent of Italian YouTube watch hours in 2014, although within that a large chunk was local Italian and US content. South Africa also had 10 countries accounting for around 80 percent of its watch hours, in this case led by the United States, the United Kingdom and then local content. Eight countries accounted for 80 percent of Colombians' watch hours in 2014; this time was more evenly spread among US, Spanish, local and Mexican content. Turkey stands out, with just four countries accounting for 80 percent of watch hours, namely Turkey, the United States, Germany and the United Kingdom.

The table from source 4 also provides the possibility to observe how watch patterns have changed over the period 2010–2014 for individual countries. As an example

11 The Antarctica (.aq) TLD is administered by the Antarctica Network Information Centre Limited located in New Zealand. The .aq domain name is available to government organizations who are signatories to the Antarctic Treaty and to other registrants who have a physical presence in Antarctica. Due to the special nature of the Antarctic environment, the registrar considers a "physical presence" to include unattended installations owned or operated by the registrant and short-term visits to the ice by the registrant or its employees. Enthusiatic consumption of US-uploaded YouTube content may be partly due to the large US base in Antarctica (McMurdo Station).

Figure 5: 2014 YouTube Watch Hour Patterns: Whose Content Are They Watching?



Source: OECD calculations, based on information from source 4 (see Box 1).

of this type of analysis, Box 2 looks at seven African countries — Cameroon, Ghana, Kenya, Malawi, Nigeria, Rwanda and Tanzania. Africa was the last continent to achieve Internet connection and is still in the relatively early stages of expanding access and coverage to its population. It is interesting to see that all countries in this sample have experienced an increase in the share of watch hours attributed to locally and proximately uploaded YouTube content, although the absolute shares differ widely, doubtless reflecting their different stages of digital development.

The international sharing of YouTube content is clearly a facet of global knowledge and information flows, but its value is likely to depend greatly on the content in question, as well as on how economic and social value is measured. Subject to data availability, future work could usefully explore different categories of content, distinguishing, say, education content from other content.

From Description to Measurement?

Because of the geographic fuzziness of the information sources analyzed here, using them as a stand-alone proxy of global data flows and linking them to data on trade, innovation and other economic indicators would be misleading. In particular, the fact that gTLDs cannot be given a geographic tag makes the use of the search data to proxy data flows on a country-by-country basis unsuitable. With .com domains representing over 40 percent of search result clicks in 2014 in 20 OECD countries (over 80 percent in the United States), for example, this loss of geographic information is significant. Added to this is the lack of one-to-one relationships between ccTLDs and the location of content. While the start and end points of data flows are clearer for the YouTube information reviewed, both it and the search information have the common problem that the actual route of data flows (and thus the interdependence of global connections) is hidden behind the bilateral data points in the tables reviewed here.

However, comparing patterns in the tables with information related to infrastructure can provide additional insights into data flows and give some pointers for the direction of work. In short, Internet infrastructure has both influenced and evolved around data flows, and continues to do so in response to market and regulatory imperatives. For instance, the growth of heavy content and consumer demands for speed and quality mean that for some types of data flows there is a clear economic case for data to stay as local as possible. One example of this might be software updates, where the same content is being downloaded multiple times and where the balance of transit costs, speed/quality outcomes and storage costs makes it sensible to shift the content close to the consumer. At the same time, there remain data flows that do not lend themselves easily to localization near the customer — they may be more unique in terms of content and need to traverse regional, if not global, networks on a constant basis. One example might be financial and logistics information flows associated with international trade. Measures and interpretations of data flows may need to be nuanced to account for different contexts. The following discussion expands on this idea and proposes some next steps.

Box 2: YouTube in Africa — A Peek at Watch Patterns

Source 4 (see Box 1) allows an analysis of YouTube watch patterns across a wide range of countries — too wide for this paper to give attention to all interesting cases. However, given Africa's status as a catch-up continent on Internet connection and usage, the table below presents some simple statistics on the change in YouTube patterns in the period 2010–2014 and current watch patterns for seven countries.

There are large differences between the countries in the share of local content watched in 2014, but all showed growth in this share from 2010 to 2014. Nigeria has the strongest local following, perhaps due to its film industry and milieu generating a wealth of content for viewers. The United States and United Kingdom figure prominently in watch hours, and Nigerian content is also popular in Cameroon and Ghana (in fact, it features in the top eight of all countries in the sample). The share of watch hours spent on US content is similar to that found in OECD countries; for instance, Cameroon is comparable to Mexico and Portugal, while the others are comparable to Denmark, Estonia, Norway, Sweden and the United Kingdom, whose shares of watch hours spent on US content are in the area of 34–39 percent.

	Share of Watch Hours Spent on Local Content, 2014 (%)	Increase in Share of Watch Hours Spent on Local Content, 2010–2014 (percentage points)	Top Three Content Countries, by Watch Hour Share	Share of Watch Hours Spent on US Content, 2014 (%)	Concentration of Watch Hours — Number of Countries Accounting for 80% of Watch Hours
Cameroon	3.14	1.46	United States, France, Nigeria	25.23	14
Ghana	9.92	4.56	United States, Nigeria, United Kingdom	35.87	10
Kenya	13.59	6.88	United States, Kenya, United Kingdom	36.69	12
Malawi	2.01	1.24	United States, United Kingdom, India	38.15	14
Nigeria	25.05	19.88	United States, Nigeria, United Kingdom	32.69	7
Rwanda	9.35	5.53	United States, Rwanda, France	34.10	14
Tanzania	13.64	9.76	United States, Tanzania, United Kingdom	32.37	12

While the absolute number of watch hours is unknown, the 2010–2014 period was likely one of strong growth due to greater infrastructure provision. For instance, in 2009 there were no undersea cables connecting East Africa to the Internet, and only one cable serving the west and southern coasts. By 2013, numerous cables had been laid and some coastal countries are now served by multiple cables.[1] In-country infrastructure has also improved. There are now 37 Internet exchange points (IXPs) on the African continent (Packet Clearing House 2015) and at least two projects aim to advance regional and cross-border interconnection (AXIS and African Peering and Interconnection Forum).[2] Foreign companies are contributing — for example, in 2011 Google Global Cache[3] was made available via the Kenyan IXP. Reductions in costs and latency significantly improved the user experience for video streaming (including YouTube) and Kenyans were able to more easily consume more local content, such as Kenyan news channels and TV programs. Local provider KENET reported a 10-fold increase in Google usage after the cache was created.

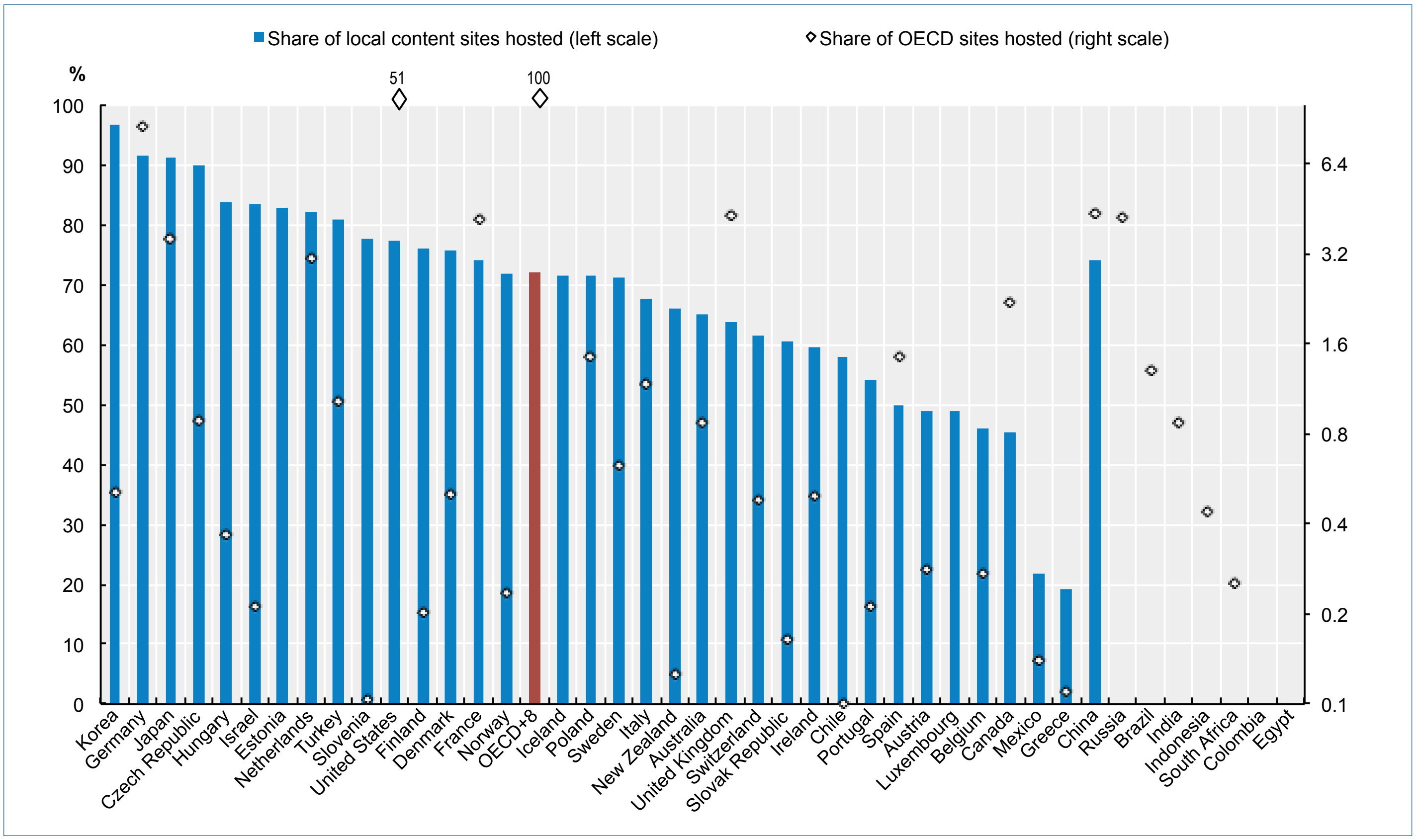
1. See maps developed by Steve Song at https://manypossibilities.net/african-undersea-cables-a-history/.

2. See http://pages.au.int/axis/about and www.internetsociety.org/events/afpif.

3. Google Global Cache is part of Google's content delivery system, whereby Google servers are placed inside the network of network providers and ISPs to serve popular Google content, including YouTube.

Source: OECD calculations, based on information from source 4; Emily Taylor.

Figure 6: Local Content Sites Hosted in Country (2013)



Source: OECD (2015b), based on Pingdom, Alexa, and datacentermap.com
Note: Based on the analysis of 429,000 ccTLDs of the top one million sites. The remaining sites, including the gTLDs, were omitted from the list, as there is no reliable public data as to where the domains are registered. Data on the share of local content sites hosted was not available for Brazil, Colombia, Egypt, India, Indonesia, Russia and South Africa.

Location, Location, Location

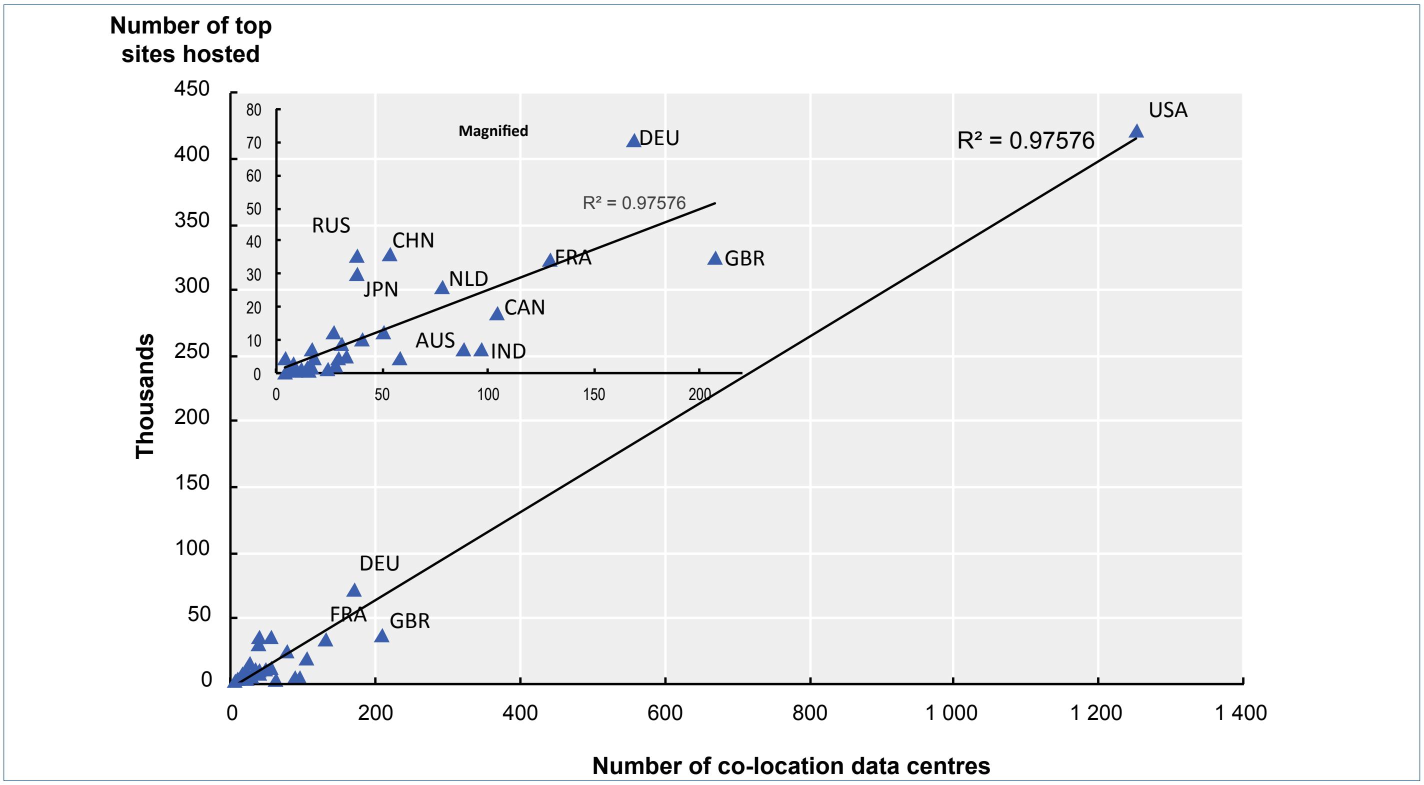
The determining factor in identifying Google search destinations (and thus data flows) is where the site is hosted, and for some ccTLDs this is predominantly offshore. Figure 6 shows to what extent countries hosted the content of their ccTLD domain in 2013. It reveals that most OECD countries host at least half the content associated with their ccTLD, but there is nevertheless a wide range of outcomes, from Korea hosting almost 97 percent of .kr sites to Greece hosting just 19 percent of .gr sites. This underscores the strong global nature of the digital economy and its associated data flows. For example, 54 percent of .pt sites were hosted in Portugal in 2013. This implies that perhaps half the time, a "local" search click to a .pt website actually entailed cross-border data flows. At the same time, Portugal also hosts foreign content (in fact, in absolute terms, Portugal hosts more foreign sites than local .pt sites), thus a share of "foreign" search clicks will stay local.[12]

The location of hosting appears to go hand-in-hand with access to efficient infrastructure. Figure 6 shows that the United States accounts for a large share of the offshore market for hosting — it hosts 51 percent of all top sites in the OECD plus Brazil, China, Colombia, Egypt, India, Indonesia, Russia and South Africa. Figure 7 reveals a clear correlation between the number of co-location data centres[13] and the number of top sites hosted in a country,

12 It is possible that the data underestimates locally hosted sites, for example, in cases where content may be presented in a national and international version — say, when a newspaper hosts a site in the country for local users and has another abroad in a location close to its international readership — or where CDNs are used to distribute data. In each case, these would have shown up as hosted outside the country in the data set (OECD 2014).

13 The OECD (2014) identifies three types of data centres: in-house data centres, located with their organization; third-party data centres or co-location facilities that offer space to clients and compete on location (sites are often around large cities, capitals and financial centres), interconnection and energy efficiency; and Internet industry data centres, say, Amazon or Facebook, for which energy and land costs are crucial.

Figure 7: Co-location Data Centres and Top Sites Hosted



Source: OECD (2015b), based on Pingdom, Alexa and datacentermap.com.
Note: Number of top sites hosted based on the analysis of 429,000 ccTLDs of the top one million sites collected in 2013. The remaining sites, including the gTLDs, were omitted from the list, as there is no reliable public data as to where the domains are registered.

suggesting that the favourable environment in the United States for setting up data centres (backhaul infrastructure, cost of energy/electricity, cost of land, regulatory environment) is an important factor in its pre-eminence. Germany is another popular location for hosting, along with France and the United Kingdom.

Logically, top hosting countries will be key conduits for data flows. For some businesses, there is a clear cost and efficiency advantage in routing data and content to data centres in these locations. Aggregating data processing, for example, can enable better control over data practices, maximize the utilization of skilled staff and improve operational efficiency. Placing this activity in the most cost-efficient location is the best business choice.

Nevertheless, for some businesses there are advantages to keeping data and content close to consumers, not all of whom are in the top hosting locations. Growth in use of CDNs and caching of content close to customers are contributing to what is, in effect, economically driven localization of some data flows. Dennis Weller and Bill Woodcock (2013) note that CDN services, such as those provided by Akamai, have supported the demand for activities such as video streaming and downloading, while some large service providers, including Google, are building their own alternatives to transit (i.e., data centres). They note that where one end of a traffic flow is a server, especially a server holding non-unique information, then the data can be replicated in many locations in order to be closer to users.

This kind of structural change in the market makes routing more direct (thus reducing costs), improves quality and increases speed of delivery. But it also makes the analysis of cross-border data flows more complex, since what may once have been multiple cross-border flows of content (for example, a music video) can become one initial cross-border flow followed by multiple local downloads from a local cache. Internet openness remains important for the content to be shared, but the magnitude of content consumption enabled by that openness is less obviously seen in cross-border data flow data.

A key piece of shared infrastructure that enables data flows to stay local when economically logical is IXPs. IXPs enable the exchange of traffic via peering between connected networks, and their global distribution plays an important role in data flow routing. Crucially, the denser their presence, the more likely it is that data can flow across shorter and faster paths between its source and the destination. An analogy is with transport networks — must travellers transit through a distant hub or can they get to their destination more directly? The shorter the distance between customers and their IXP, the lower the costs and higher the quality of data flows.

Countries with a low density of IXPs are more likely to have cross-border data flows associated with their Internet activity, partly because IXPs and data centres are often co-located,[14] and partly because even if it involves a locally hosted site, data may have no choice but to transit through an IXP in another country to gain access to the destination network. Over time, the number of IXPs has grown, particularly in emerging economies. In April 2011, Weller and Woodcock (2013, 54) counted 357 IXPs worldwide, with 25 percent in North America and 38 percent in Europe. Prior to 2011, all regions had built new IXPs, with growth especially high in Latin America, which went from 20 to 34 IXPs. This growth was welcomed by the authors, as it reduced the need to "trombone" traffic out of the country or region, allowed for more direct routing of traffic and thus improved service quality, and freed up long-haul capacity to focus on actual out-of-region traffic (ibid., 9).

As of October 2015, the global number of IXPs had grown to 452, with 60 in Latin America and 37 in Africa.[15] The impetus to build an IXP essentially comes down to cost — ISPs prefer to have an IXP in close proximity so that the cost of outbound traffic is reduced.[16] The break-even point depends on traffic volume and the ratio of local to international traffic — but at a cost of US$3.50 per unit of megabits per second (Mbps) for IP transit, an ISP could be better off joining an IXP with a traffic volume of just 2,000 Mbps.[17]

Where to from Here?

The clear takeaway is that the flow of data across the Internet is complicated — data flows come in different forms, and they do not follow political or geographic borders but, rather, economic parameters that are set by changing market conditions and the regulatory/competitive environment. How, then, can we most usefully measure Internet openness so as to link it to indicators of governments' ultimate economic policy goals?

Looking ahead, two complementary approaches could be proposed as future research paths.

Approach one: Construct a global data flow data set that more accurately tracks geographical start and end points, as well as important waypoints en route, ideally with information on the types of flows, so as to better understand the nature and volume of data flows. This approach would essentially seek to build a data flow data set that could more easily be married with economic data sets, which are typically organized by country. Possible additional data and information sources to assist with this include:

- actual traffic data, both aggregate and in certain subcategories;
- further flow data from firms;
- information on the location of .com sites;
- information on the location of key data centre sites and their throughput; and
- information on barriers to data flows, to be used in constructing proxies for modelling purposes.

This approach raises the question of whether governments should seek to establish voluntary national statistical collections of traffic data. Australia, for instance, conducts a twice-yearly survey of ISPs with more than 1,000 subscribers, collecting data on *inter alia* the number of ISPs, subscriber sectors and the volume of data downloaded.[18] It is perhaps time to explore whether such surveys should be expanded to include information on cross-border data flows. At the least, establishing a consistent cross-country methodology for collection of ISP data could enable analysis using domestic network traffic as a proxy for Internet openness, with coverage eventually expanding to cross-border data flows.

Approach two: Identify hotspots of data flow intensity (and, where possible, identify hotspots of data flow value) and overlay these with data showing the intensity and value of various economic performance variables (related to trade, innovation, entrepreneurship, productivity, and so on). In some ways, this approach would cast data flows as global data chains — similar to GVCs in the trade and production space — with intensity and value varying across different parts of the chain. Possible additional data and information sources to assist with this include:

14 The OECD (2014) notes that carrier-neutral data centres endeavour to get IXPs into their facilities, as this makes interconnection with many networks possible.

15 Data obtained from https://prefix.pch.net/applications/ixpdir/menu_download.php.

16 Weller and Woodcock (2013, Annex 4) describe how peering agreements, which comprise over 99 percent of all traffic exchange agreements, are constructed on the basis of equitable cost-revenue sharing between partners. This construction, in turn, relies on a distribution of IXPs that allows ISPs to have a similar balance of short- and long-haul paths to their traffic partners, so that neither is bearing disproportionately high costs.

17 The Internet Society (2014, 23) shows an example where traffic is destined for local termination and is either "local," "near" or "far" from the IXP. Co-location costs are estimated at US$1,000 per month, peering fees at US$2,000 per month, equipment costs at US$2,000 per month and transport into the IXP from US$2,000–$6,000 per month (depending on the distance). With an IP transit cost of US$3.50 per Mbps (estimated from information from ISPs), the break-even point to join the IXP ranges from 2,000 to around 3,140 Mbps.

18 See the Australian Bureau of Statistics Internet Activity Survey (catalogue 8153.0), www.abs.gov.au/ausstats/abs@.nsf/mf/8153.0/.

- density of data infrastructure[19]: density and composition of players at IXPs; density of interconnection agreements at IXPs; bandwidth at IXPs; IP version 6 deployment by region; and
- analysis of value added of certain Internet-related activities, similar to analysis of trade in value added.

Finally, despite the evident need for further work, there are two important conclusions regarding Internet openness that emerge from this initial analysis. First, in line with its original design, the Internet remains a highly interdependent system. Data flows frequently have international dimensions and are not necessarily predictable. Reducing openness in any part of the system could have knock-on effects across the whole system, and thus all countries have an interest in ensuring that policy decisions regarding the Internet take into account the costs and benefits of openness. Given the important role of the United States in many aspects of the digital economy, its policy decisions clearly matter, but so too do those of other countries. For instance, Figure 6 showed Germany hosted almost 8.5 percent of OECD top websites in 2013, which suggests that its policy decisions on data flows and Internet openness would likely have significant consequences across the system.

Second, Internet openness, in terms of enabling data, information and knowledge to flow across the globe, is incontrovertibly tied to open markets and competitive conditions. Firms must be able to invest in or establish access to infrastructure that allows them to efficiently and effectively provide their services, be it on a local or cross-border basis; if they cannot, customer access, choice and service quality suffer. Weller and Woodcock (2013, 45) note a frequent observation that "improvement of the Internet depends upon a circular path of improvement of each component of the Internet's infrastructure: IXPs, international connectivity, content, backbone networks, and access networks. One circumnavigates this circle endlessly, upgrading each in turn." This observation has distinct parallels with Internet openness and suggests that measures of Internet openness need to incorporate infrastructural factors.

CONCLUSION

The initial stages of the OECD's work to measure global data flows underscore the highly interconnected nature of today's Internet. Aggregate information on Google searches and YouTube watch hours suggest that users are increasingly accessing content outside their countries, highlighting the potential of the Internet for cross-country information flows and knowledge exchange. In addition, countries' Internet infrastructure and content have strong global interlinkages — one example being offshore hosting of local websites.

It is in the interests of all governments to improve the evidence base for policy making, because choices about Internet openness matter for countries' trade and innovation performance. The strong international linkages inherent in the Internet also mean that the effects of a country's Internet-related policies can spill across its borders. The OECD will continue to work with its members and partners to better understand global data flows and their effects on economies and societies.

19 Presentation of this kind of data could take inspiration from the Internet connectivity maps produced by Larry Landweber in the 1990s. See http://internethalloffame.org/news/in-their-own-words/larry-landweber-play-lab-world.

WORKS CITED

Accenture. 2013. *Entrepreneurial Innovation: How to Unleash a Key Source of Growth and Jobs in the G20 Countries: Young Entrepreneurs' Alliance Summit 2013*. Accenture. www.g20yea.com/en/wp-content/uploads/Accenture-Entrepreneurial-Innovation-G20-YEA-Report.pdf.

Blumenthal, Marjory S. and David D. Clark. 2001. "Rethinking the Design of the Internet: The End-to-End Arguments vs. the Brave New World." *ACM Transactions on Internet Technology* 1 (1): 70–109.

BSA. 2014. *Powering the Digital Economy: A Trade Agenda to Drive Growth*. Washington, DC: January. http://digitaltrade.bsa.org/.

Cisco. 2015a. "The Zettabyte Era: Trends and Analysis." White paper. May.

———. 2015b. "Cisco Visual Networking Index: Forecast and Methodology, 2014–2019." White paper. May 27.

Daigle, Leslie. 2015. *On the Nature of the Internet*. Global Commission on Internet Governance Paper Series No. 7. Waterloo, ON: CIGI and Chatham House. March 16. www.cigionline.org/publications/nature-of-internet.

Dalberg. 2014. *Open for Business? The Economic Impact of Internet Openness*. Dalberg Global Development Advisors, March. www.dalberg.com/documents/Open_for_Business_Dalberg.pdf.

Ezell, Stephen, Robert D. Atkinson and Michelle Wein. 2013. *Localization Barriers to Trade: Threat to the Global Innovation Economy*. Washington, DC: Information Technology and Innovation Foundation, September 25.

Hofheinz, Paul and Michael Mandel. 2014. "Bridging the Data Gap: How Digital Innovation Can Drive Growth and Create Jobs." Lisbon Council-Progressive Policy Institute Policy Brief 15/2014.

———. 2015. "Towards a 21st Century Agenda of Transatlantic Prosperity." Lisbon Council-Progressive Policy Institute Policy Brief 19/2015.

Internet Society. 2014. *The Internet Exchange Point Toolkit & Best Practices Guide: How to Maximize the Effectiveness of Independent Network Interconnection in Developing Regions and Emerging Markets.* Geneva, Switzerland and Virginia, United States: Internet Society February. www.internetsociety.org/sites/default/files/Global%20IXPToolkit_Collaborative%20Draft_Feb%2024.pdf.

Kaplan, James M. and Kayvan Rowshankish. 2015. *Addressing the Impact of Data Location Regulation in Financial Services.* Global Commission on Internet Governance Paper Series No. 14. Waterloo, ON: CIGI and Chatham House. May 22. www.cigionline.org/publications/addressing-impact-of-data-location-regulation-financial-services.

McKinsey Global Institute. 2014. *Global Flows in a Digital Age: How Trade, Finance, People, and Data Connect the World Economy*. McKinsey & Company, April.

McNeil, Donald G., Jr. 2013. "Car Mechanic Dreams Up a Tool to Ease Births." *The New York Times*, November 13. www.nytimes.com/2013/11/14/health/new-tool-to-ease-difficult-births-a-plastic-bag.html?_r=0.

Meltzer, Joshua. 2015. "The Internet, Cross-Border Data Flows and International Trade." *Asia and the Pacific Policy Studies* 2 (1): 90–102.

Mettler, Ann and Anthony D. Williams. 2011. "The Rise of the Micro-Multinational: How Freelancers and Technology-Savvy Start-Ups Are Driving Growth, Jobs and Innovation." Lisbon Council Policy Brief V (3). Brussels, Belgium.

———. 2012. "Wired for Growth and Innovation: How Digital Technologies are Reshaping Small- and Medium-Sized Businesses." Lisbon Council Interactive Policy Brief 12/2012. Brussels, Belgium.

Nicholson, Jessica R. and Ryan Noonan. 2014. "Digital Economy and Cross-Border Trade: The Value of Digitally-Deliverable Services." ESA Issue Brief #01-14. United States Department of Commerce: Economics and Statistics Administration, January 27.

OECD. 2008. *OECD Guidelines for Multinational Enterprises*. Paris, France: OECD Publishing.

———. 2012. "The Impact of Internet in OECD Countries." OECD Digital Economy Papers, No. 200. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/5k962hhgpb5d-en.

———. 2014. "International Cables, Gateways, Backhaul and International Exchange Points." OECD Digital Economy Papers, No. 232. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/5jz8m9jf3wkl-en.

———. 2015a. *OECD Digital Economy Outlook 2015*. Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264232440-en.

———. 2015b. *Data-Driven Innovation: Big Data for Growth and Well-Being.* Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264229358-en.

———. Forthcoming 2016. *The Economic and Social Benefits of Internet Openness.* Paris, France: OECD Publishing.

Packet Clearing House. 2015. "Packet Clearing House Report on Internet Exchange Point Locations. " www.pch.net/ixpdir/summary.

Swedish National Board of Trade. 2015. *No Transfer, No Production: A Report on Cross-border Data Transfers, Global Value Chains, and the Production of Goods*. Kommerskollegium.

Weller, Dennis and Bill Woodcock. 2013. "Internet Traffic Exchange: Market Developments and Policy Challenges." OECD Digital Economy Papers, No. 207. OECD Publishing. http://dx.doi.org/10.1787/5k918gpt130q-en.

West, Jeremy. 2016. *A Framework for Understanding Internet Openness*. Global Commission on Internet Governance Paper Series No. 35. Waterloo, ON: CIGI and Chatham House.

Zwillenberg, Paul, Dominic Field and David Dean. 2014. *The Connected World: Greasing the Wheels of the Internet Economy: A Country-by-Country e-Friction Analysis*. Boston Consulting Group. www.bcgperspectives.com/content/articles/digital_economy_telecommunications_greasing_wheels_internet_economy/.

CIGI PUBLICATIONS
ADVANCING POLICY IDEAS AND DEBATE

Global Commission on Internet Governance

The Global Commission on Internet Governance (GCIG) was established in January 2014 to articulate and advance a strategic vision for the future of Internet governance. The two-year project conducts and supports independent research on Internet-related dimensions of global public policy, culminating in an official commission report that will articulate concrete policy recommendations for the future of Internet governance. These recommendations will address concerns about the stability, interoperability, security and resilience of the Internet ecosystem. Launched by two independent global think tanks, the Centre for International Governance Innovation and Chatham House, the GCIG will help educate the wider public on the most effective ways to promote Internet access, while simultaneously championing the principles of freedom of expression and the free flow of ideas over the Internet.

GLOBAL COMMISSION ON INTERNET GOVERNANCE PAPER SERIES

The Regime Complex for Managing Global Cyber Activities
GCIG Paper Series No. 1
Joseph S. Nye, Jr.

Tipping the Scale: An Analysis of Global Swing States in the Internet Governance Debate
GCIG Paper Series No. 2
Tim Maurer and Robert Morgus

Legal Mechanisms for Governing the Transition of Key Domain Name Functions to the Global Multi-stakeholder Community
GCIG Paper Series No. 3
Aaron Shull, Paul Twomey and Christopher S. Yoo

Legal Interoperability as a Tool for Combatting Fragmentation
GCIG Paper Series No. 4
Rolf H. Weber

Innovations in Global Governance: Toward a Distributed Internet Governance Ecosystem
GCIG Paper Series No. 5
Stefaan G. Verhulst, Beth S. Noveck, Jillian Raines and Antony Declercq

The Impact of the Dark Web on Internet Governance and Cyber Security
GCIG Paper Series No. 6
Tobby Simon and Michael Chertoff

On the Nature of the Internet
GCIG Paper Series No. 7
Leslie Daigle

Understanding Digital Intelligence and the Norms That Might Govern It
GCIG Paper Series No. 8
David Omand

ICANN: Bridging the Trust Gap
GCIG Paper Series No. 9
Emily Taylor

A Primer on Globally Harmonizing Internet Jurisdiction and Regulations
GCIG Paper Series No. 10
Michael Chertoff and Paul Rosenzweig

Connected Choices: How the Internet is Challenging Sovereign Decisions
GCIG Paper Series No. 11
Melissa E. Hathaway

Solving the International Internet Policy Coordination Problem
GCIG Paper Series No. 12
Nick Ashton-Hart

Net Neutrality: Reflections on the Current Debate
GCIG Paper Series No. 13
Pablo Bello and Juan Jung

Addressing the Impact of Data Location Regulation in Financial Services
GCIG Paper Series No. 14
James M. Kaplan and Kayvaun Rowshankish

Cyber Security and Cyber Resilience in East Africa
GCIG Paper Series No. 15
Iginio Gagliardone and Nanjira Sambuli

Global Cyberspace Is Safer than You Think: Real Trends in Cybercrime
GCIG Paper Series No. 16
Eric Jardine

CIGI PUBLICATIONS
ADVANCING POLICY IDEAS AND DEBATE

The Emergence of Contention in Global Internet Governance
GCIG Paper Series No. 17
Samantha Bradshaw, Laura DeNardis, Fen Osler Hampson, Eric Jardine and Mark Raymond

Landmark EU and US Net Neutrality Decisions: How Might Pending Decisions Impact Internet Fragmentation?
GCIG Paper Series No. 18
Ben Scott, Stefan Heumann and Jan-Peter Kleinhans

The Strengths and Weaknesses of the Brazilian Internet Bill of Rights: Examining a Human Rights Framework for the Internet
GCIG Paper Series No. 19
Carolina Rossini, Francisco Brito Cruz and Danilo Doneda

The Tor Dark Net
GCIG Paper Series No. 20
Gareth Owen and Nick Savage

The Dark Web Dilemma: Tor, Anonymity and Online Policing
GCIG Paper Series No. 21
Eric Jardine

One in Three: Internet Governance and Children's Rights
GCIG Paper Series No. 22
Sonia Livingstone, John Carr and Jasmina Byrne

Combatting Cyber Threats: CSIRTs and Fostering International Cooperation on Cybersecurity
GCIG Paper Series No. 23
Samantha Bradshaw

The Privatization of Human Rights: Illusions of Consent, Automation and Neutrality
GCIG Paper Series No. 24
Emily Taylor

The Digital Trade Imbalance and Its Implications for Internet Governance
GCIG Paper Series No. 25
Susan Ariel Aaronson

A Pragmatic Approach to the Right to be Forgotten
GCIG Paper Series No. 26
Wendy Hall, Kieron O'Hara and Nigel Shadbolt

Education 3.0 and Internet Governance: A New Global Alliance for Children and Young People's Sustainable Digital Development
GCIG Paper Series No. 27
Divina Frau-Meigs and Lee Hibbard

Jurisdiction on the Internet: From Legal Arms Race to Transnational Cooperation
GCIG Paper Series No. 28
Bertrand de La Chapelle and Paul Fehlinger

Patents and Internet Standards
GCIG Paper Series No. 29
Jorge L. Contreras

Tracing the Economic Impact of Regulations on the Free Flow of Data Localization
GCIG Paper Series No. 30
Matthias Bauer, Martina F. Ferracane and Erik van der Marel

Looking Back on the First Round of New gTLD Applications: Implications for the Future of Domain Name Regulation
GCIG Paper Series No. 31
Jacqueline D. Lipton

Governance of International Trade and the Internet: Existing and Evolving Regulatory Systems
GCIG Paper Series No. 32
Harsha Vardhana Singh, Ahmed Abdel-Latif and L. Lee Tuthill

Market-driven Challenges to Open Internet Standards
GCIG Paper Series No. 33
Patrik Fälström

How to Connect the Other Half: Evidence and Policy Insights from Household Surveys in Latin America
GCIG Paper Series No. 34
Hernan Galperin

Available for free download at www.cigionline.org/publications

ABOUT CIGI

The Centre for International Governance Innovation is an independent, non-partisan think tank on international governance. Led by experienced practitioners and distinguished academics, CIGI supports research, forms networks, advances policy debate and generates ideas for multilateral governance improvements. Conducting an active agenda of research, events and publications, CIGI's interdisciplinary work includes collaboration with policy, business and academic communities around the world.

CIGI's current research programs focus on three themes: the global economy; global security & politics; and international law.

CIGI was founded in 2001 by Jim Balsillie, then co-CEO of Research In Motion (BlackBerry), and collaborates with and gratefully acknowledges support from a number of strategic partners, in particular the Government of Canada and the Government of Ontario.

Le CIGI a été fondé en 2001 par Jim Balsillie, qui était alors co-chef de la direction de Research In Motion (BlackBerry). Il collabore avec de nombreux partenaires stratégiques et exprime sa reconnaissance du soutien reçu de ceux-ci, notamment de l'appui reçu du gouvernement du Canada et de celui du gouvernement de l'Ontario.

For more information, please visit www.cigionline.org.

ABOUT CHATHAM HOUSE

Chatham House, the Royal Institute of International Affairs, is based in London. Chatham House's mission is to be a world-leading source of independent analysis, informed debate and influential ideas on how to build a prosperous and secure world for all. The institute: engages governments, the private sector, civil society and its members in open debates and confidential discussions about significant developments in international affairs; produces independent and rigorous analysis of critical global, regional and country-specific challenges and opportunities; and offers new ideas to decision-makers and -shapers on how these could best be tackled from the near- to the long-term. For more information, please visit: www.chathamhouse.org.

Communications

For media enquiries, please contact communications@cigionline.org.

CIGI

67 Erb Street West
Waterloo, Ontario N2L 6C2
tel +1 519 885 2444 fax +1 519 885 5450
www.cigionline.org

CHATHAM HOUSE
The Royal Institute of International Affairs

10 St James's Square
London, England SW1Y 4LE, United Kingdom
tel +44 (0)20 7957 5700 fax +44 (0)20 7957 5710
www.chathamhouse.org

FSC
www.fsc.org
RECYCLED
Paper made from recycled material
FSC® C023070