PPI Working Paper

Data, Trade, and Growth

Dr. Michael Mandel
Progressive Policy Institute
Mmandel@progressivepolicy.org

Presented at
"Data-Driven Productivity and Trade: Unlocking New Drivers of Growth"
Brussels, Belgium

Revised May 13, 2013
Original version February 18, 2013.

My thanks to the Sloan Foundation for funding this research. I thank Diana Carew of PPI, Alan Mauldin of TeleGeography, Steve Bauer and Bill Lehr of MIT, and Michael Kende of Analsys Mason for very useful assistance. All mistakes and errors are my own.

Cross-border data flows are the fastest growing component of international trade. According to TeleGeography, a consulting firm that keeps track of international data flows, demand for international bandwidth increased at a compound annual rate of 49 percent between 2008 and 2012.[1] By comparison, the volume of global trade in goods and services rose at an average rate of 2.4% over the same period.

The global economy could not function without cross-border data flows. Companies such as Wal-Mart are closely connected via high-bandwidth 'data pipes' with their affiliates and suppliers in other countries. Financial traders are doing transactions around the clock in every corner of the globe. Researchers in different countries collaborate online. Individuals browse foreign websites, download videos, and participate in online education originated in other countries while sitting at their home computers. Mammoth video databases such as YouTube pump out videos to watchers across the world. And large internet companies such as Google, Twitter, and Facebook have users in every country.

Because cross-border data flows are valuable, they are increasingly important as a topic for trade policy. [2] For example, the European Union is considering new data privacy regulations that could potentially act as an impediment to flows of data in and out of the EU. The original version of the regulations, announced in January 2012, specifically noted that EU rules would apply even if personal data was sent overseas.[3]

---

[1] "Global Bandwidth Research Service," TeleGeograOphy, http://www.telegeography.com/research-services/global-bandwidth-research-service/index.html.

[2] "Policy Challenges of Cross-Border Cloud Computing," Renee Berry and Matthew Reisman, *Journal of International Commerce and Economics*, May 2012.

[3] "Data protection reform: Frequently asked questions," January 24 2012 http://europa.eu/rapid/press-release_MEMO-12-41_en.htm?locale=en

On the U.S. side, cross-border data flows are one of the main subjects of the new round of trade negotiations announced by the United States Trade Representative, Ambassador Ron Kirk in January 2013.[4] In a letter to Congress, Kirk wrote:

> Internet usage has increased by 500 times since 2001, when the last multilateral trade negotiation was launched. The development of appropriate provisions to support service trade through electronic channels must therefore be an important component of any new agreement.

There's an irony here, however—in a world awash with data, we have very little data about the magnitude of cross-border data flows. The monthly trade report from the Census Bureau and the Bureau of Economic Analysis contains some information about imports and exports of telecommunications services, but as we will see below, these figures likely miss much of the increase in cross-border data traffic because of fundamental changes in the structure of global networks. Similarly, international agencies such as the ITU only collect fragmentary statistics on cross-border data flows, though they are putting more effort into estimating such figures. [5]

More important, few economists studied the economic importance of cross-border data flows, or the negative consequences of erecting barriers to trade in data. Our usual economic models for evaluating foreign trade policy are designed for goods and services, where we have a clear idea about what counts as an export or import.

By contrast, the usual categories of exports and imports don't seem to apply very well to cross-border data flows, since it's not clear that an outflow of data from a country should count as an export. Indeed, long-established conventions treat outgoing international phone calls as imports, even though both the originating network and the receiving network play an equal role in the call.

---

[4]Letter from Ambassador Ron Kirk, http://www.ustr.gov/sites/default/files/01152013%20ARK%20letter%20to%20Speaker%20Boehner_0.pdf
[5] See for example

Plan of this paper

This paper is intended as a preliminary exploration of the empirical, conceptual, and policy aspects of cross-border data flows. There are four main results.

- First, we do a rough comparison of the magnitude of cross-border flows of data with all internet/IP traffic. We find that for the United States, cross-border flows are roughly 16-25% of all U.S. data traffic. Cross-border data flows between Europe and the rest of the world equal roughly 13-16% of all European data traffic. These estimates, which should be treated as highly imprecise and tentative, suggest that the United States is more interconnected with the rest of the world than Europe.

- Second, rather than interpret these data flows as exports and imports, we develop a new framework for international trade in data based around the concept of **replication.** What we see today is companies such as Google and Amazon spending large amounts of money to build data centers around the world. In effect, these data centers are 'replication facilities', taking data generated in the rest of the world and replicating it locally, in a form that is easily accessible to local residents and businesses. As a result, the relative utility of the data rises because it can be accessed repeatedly with faster response times (latency) and at a higher quality.

- Third, we note some economic and policy implications of data trade as replication. The construction of a replication facility such as a data center in a region effectively boosts the intangible capital stock of a region by making the rest of the world more accessible, effectively boosting its economic output at a relatively low cost. However, this strong positive impact of trade in data requires that companies have an economic incentive to build replication facilities, which are expensive to build and run.

- Four, regions that tax or regulate data flows may deter companies from investing in replication facilities in those regions . In turn, these regulations may have a harmful effect on economic growth.

<u>Section I: Measurement Issues with Cross-Border Data Flows</u>

<u>Measuring Cross-Border Data Flows</u>

In an earlier paper, we discussed the idea that the production and use of data should be treated as a fundamental component of economic activity, parallel to the production and use of goods and services. [6] This approach leads naturally to an increased focus on directly measuring data generation, data flows, and data storage as a way of understanding economic activity.

One pioneer in such efforts has been Martin Hilbert, who has been developing a systematic methodology for comparing the communications capacity of various media, ranging from mobile to television.[7]Based on this work, the International Telecommunications Union (ITU) has been gradually moving towards direct measures of data flows, as opposed to indirect measures such as numbers of cellular subscriptions or broadband connections. A recent publication from the ITU notes:

> Using the unifying metric of bits per second, employed for measuring global technological capacity to communicate, it is possible to compare different communication technologies. It is also possible to analyse bits per second per capita, per technology, per country or per any other relevant socio-economic or demographic parameter. [8]

---

[6] "Beyond Goods and Services: The (Unmeasured) Rise of the Data-Driven Economy," Michael Mandel, Progressive Policy Institute, October 2012

[7] See, for example, "The World's Technological Capacity to Store, Communicate, and Compute Information," Martin Hilbert and Priscila Lopez, *Science*, April 2011.

[8]*Measuring the Information Society 2012*, International Telecommunications Union, 2012, http://www.itu.int/ITU-D/ict/publications/idi/material/2012/MIS2012_without_Annex_4.pdf

This paper follows in the same spirit of direct measurement of data flows. However, we are particularly interested in cross-border data flows, which have become increasingly important in recent years. This is a difficult and tricky area, so the results presented here should be viewed as exploratory.

For the purposes of this paper, data flow is measured in terabits per second (Tbps). Submarine cable is used to carry cross-border data flow across oceans but also often between countries on the same continent, because it's often easier and safer to maintain cables that run along the coast underwater than across difficult terrain.

Cables are typically laid with multiple strands of optic fiber, some of which are 'lit' and have the necessary equipment to be used, and some of which are 'dark' and not yet ready for use. Capacity can be increased by laying new cables, lighting dark fiber, or by improving the capacity of already lit fiber.

The telecommunications market research and consulting firm TeleGeography estimates that the U.S. had 23 Tbps of international internet capacity in 2012, with an average utilization of 29% and a peak utilization of 49%.[9] This suggests that on average, the U.S. cross-border data flow is roughly 6.7 Tbps.[10]

Is this volume of cross-border data a large number or a small number? We compare the cross-border data flow with a recent Cisco-sponsored projection of data traffic, by country and type.[11] For 2012, the Cisco study estimates that internet and IP traffic in the United States at 8 exabytes per month and 13 exabytes per month, respectively. [12] That translates into roughly 26.5 Tbps and 42.2 Tbps. [13]

---

[9] We thank Alan Mauldin of TeleGeography for providing these estimates.

[10] These figure are based on the bidirectional averages of the monthly average and peak during April of each year.

[11] See "The Zettabyte Era, " Cisco, May 2012. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf. See also http://ciscovni.com/vni_forecast/advanced.html

[12] Non-internet IP traffic in the U.S. is mainly consumer video.

Table 1 below compares the U.S. cross-border data flows with the overall U.S. internet and IP traffic. We find that cross-border data flows are roughly 25% and 16% of U.S. internet and IP traffic, respectively. To put this in perspective, U.S. exports of goods and services are 14% of U.S. gross domestic product (GDP) in 2012, while U.S. imports of goods and services are 18% of U.S. gross domestic product (GDP) in 2012.

[13] 1 exabyte=1024 petabytes; 1 petabyte=1024 terabytes; 1 terabyte = 8 terabits.

Table 1: Cross-Border Data Flows

Cross-Border Data Flows, 2012: United States

	Terabits per second (except as noted)
International internet capacity connected to the U.S.	23
Average utilization (percent)	29%
Average cross-border data flow (average international traffic)	6.7
All U.S. internet traffic	26.5
All U.S. IP traffic	42.2
Average U.S. cross-border data flow as a percent of:	
All U.S.internet traffic	25%
All U.S. IP traffic	16%

Data: International capacity and utilization estimates from TeleGeography. Traffic estimates from Cisco. IP includes both internet traffic and managed IP such as consumer video. Figures omit mobile.

Cross-Border Data Flows, 2012: Europe

	Terabits per second (except as noted)
International internet capacity (including intra-Europe)	56.5
Interregional internet capacity connected to Europe	12.6
Average utilization (percent)	39%
Average cross-border data flow (average international traffic)	4.9
All European internet traffic	30.4
All European IP traffic	37.2
Average cross-border data flow as a percent of:	
All European internet traffic	16%
All European IP traffic	13%

Data: International capacity and utilization estimates from TeleGeography. Traffic estimates from Cisco. IP includes both internet traffic and managed IP such as consumer video. Figures omit mobile.

Now, these numbers are accompanied by substantial and worrisome caveats, and the possibilities of large errors in both directions. In particular:

- **Coverage and methodology may differ.** Cisco's projections include all IP usage. TeleGeography's estimates of international capacity by country do not include private networks such as intra-corporate networks, Google and other content providers networks, and research networks. This factor would tend to underestimate the share of cross-border traffic.
- **Double-counting is inevitable**. International internet traffic is often routed through third-party countries before getting to its destination. Traffic between Moscow and New York might be routed through London and therefore show up as part of European cross-border data flows. Traffic between the Canadian cities of Vancouver and Toronto might be routed through the United States, and therefore show up as part of U.S. cross-border data flows. And since less-developed countries may have better internet connections with the U.S. and Europe than with each other, it's possible for intra-African traffic, say, to be routed through New York or London. This factor would tend to overestimate the share of cross-border traffic
- **When comparing estimates/forecasts from different sources, timing matters.** International internet capacity, as estimated by TeleGeography, has been growing at almost 50% per year. Domestic U.S. internet traffic as projected by Cisco, has been growing roughly as fast. As a result, calculating cross-border data flows as a share of internet traffic can be heavily influenced if one source is using yearly averages while the other source (TeleGeography) is using a particular point in time (April of each year). The direction of bias is uncertain.
- **Compression may distort the statistics**: Widespread and growing use of compression means that "we communicate around three times more

information through the same installed infrastructure as we did in 1986."[14] The direction of bias is uncertain.

We can calculate interregional cross-border data flows as a share of internet traffic for Europe. TeleGeography estimated that international bandwidth in Europe was 56.5 Tbps in 2012, but that 78 percent of that bandwidth was between cities in the same region. As a result, "interregional internet capacity connected to Europe" equaled 12.6 Tbps in 2012. Based on this figure, we calculate that cross-border data flows between Europe and the rest of the world equaled 16% of the region's internet traffic and 13% of the region's IP traffic.

These results, which should be viewed as highly imprecise and tentative, suggest that the United States is more interconnected with the rest of the world than Europe. The sources of error enumerated in the caveats above are potentially very significant.

<u>Contrast: Tracking the Money</u>

In contrast with the direct approach to measuring cross-border data described above, the U.S. Bureau of Economic Analysis (BEA) measures data-related exports and imports by tracking the money received from 'foreign persons' and the money paid to 'foreign persons' for data-related services. This approach is very informative. However, in this section we will show that given the changing structure of the internet, tracking the money misses some very important and growing sources of cross-border data flows.

The BEA collects much of its data on service sector exports and imports through surveys: specifically "BEA benchmark (BE-120) and quarterly (BE-125) Survey of

[14] "Mapping the dimensions and characteristics of the world's technological communication capacity during the period of digitization (1986 - 2007/2010)", ITU, December 2011, prepared by Martin Hilbert.

Transactions in Selected Services and Intangible Assets with Foreign Persons. "[15] Figure 1 shows the list of services queried in these surveys is fairly long. Many of these include cross-border data flows, including telecommunication services, database and other information services, and financial services.

[15] "Chapter 8: net exports of goods and services," Bureau of Economic Analysis, updated November 2011. http://www.bea.gov/national/pdf/ch8NetExport%20for%20posting.pdf

Figure 1. Service Transactions Tracked by BEA Survey

Transaction code	Types of transactions
	Receipts for intellectual property
1.	Rights related to industrial processes and products
2.	Rights related to books, music, etc.
3.	Rights related to trademarks
4.	Rights related to performances and events pre-recorded on motion picture film and TV tape (include digital recordings)
5.	Rights related to broadcast and recording of live events and performances
6.	Rights related to general use software
7.	Business format franchising fees
8.	Other intellectual property
	Receipts for selected services
9.	Accounting, auditing, and bookkeeping services
10.	Advertising services
11.	Auxiliary insurance services
12.	Computer and data processing services
14.	Data base and other information services
15.	Educational and training services
18.	Industrial engineering services
19.	Industrial-type maintenance, installation, alteration, and training services
20.	Legal services
21.	Management, consulting, and public relations services (including expenses allocated by a U.S. parent to its foreign affiliates)
22.	Merchanting services
24.	Operational leasing services
25.	Trade-related services, other than merchanting services
26.	Performing arts, sports, and other live performances, presentations, and events
29.	Research and development services
30.	Telecommunications services
31.	Agricultural services
33.	Disbursements to fund production costs of motion pictures
34.	Disbursements to fund news-gathering costs and production costs of program material other than news
35.	Waste treatment and depollution services
36.	Other selected services

However, the problem is that it's increasingly easy to find important cases where data crosses borders but no money changes hands. We're going to consider two such cases.

- An American economist downloads Russian money supply data from the Bank of Russia, the Russian central bank (www.cbr.ru)
- A European teenager views a video of a Stanford University class on machine learning on YouTube. (http://www.youtube.com/watch?v=UzxYlbK2c7E)

Figure 2: Screenshot of Russian Central Bank data

Центральный банк Российской Федерации

Денежная масса (национальное определение) в 2012 году
(млрд. руб.)

Дата	Наличные деньги в обращении вне банковской системы (денежный агрегат M0)	Переводные депозиты	в том числе: населения	в том числе: нефинансовых и финансовых (кроме кредитных) организаций	Денежный агрегат M1[1]	Другие депозиты, входящие в состав денежного агрегата M2	в том числе: населения	в том числе: нефинансовых и финансовых (кроме кредитных) организаций	Денежная масса в национальном определении (денежный агрегат M2)[2]	Темпы прироста денежного агрегата M2, %: к предыдущему месяцу	к 01.01.2012	к соответст-вующему месяцу предыдущего года
	1	2 = 3 + 4	3	4	5 = 1 + 2	6 = 7 + 8	7	8	9 = 5 + 6	10	11	12
01.01.2012	5 938,6	6 918,9	2 169,7	4 749,2	12 857,4	11 625,7	7 684,7	3 941,0	24 483,1	11,7	—	22,3
01.02.2012	5 670,7	6 630,5	1 913,4	4 717,0	12 301,2	11 316,4	7 776,3	3 540,1	23 617,6	-3,5	-3,5	22,3
01.03.2012	5 713,0	6 572,6	1 964,7	4 607,8	12 285,6	11 505,5	7 891,7	3 613,9	23 791,1	0,7	-2,8	21,8
01.04.2012	5 704,3	6 568,9	1 953,7	4 615,2	12 273,2	11 702,0	8 010,7	3 691,4	23 975,3	0,8	-2,1	21,2
01.05.2012	5 831,5	6 399,3	2 054,8	4 344,5	12 230,8	11 931,5	8 154,7	3 776,9	24 162,3	0,8	-1,3	20,7
01.06.2012	5 856,4	6 497,3	2 034,3	4 463,1	12 353,7	12 012,2	8 241,1	3 771,1	24 365,9	0,8	-0,5	20,9
01.07.2012	6 003,9	6 617,4	2 195,5	4 421,9	12 621,3	12 057,9	8 332,7	3 725,2	24 679,2	1,3	0,8	19,1
01.08.2012	5 976,3	6 494,6	2 148,7	4 345,9	12 470,9	12 093,5	8 380,4	3 713,1	24 564,3	-0,5	0,3	17,9
01.09.2012	5 980,0	6 313,7	2 156,6	4 157,1	12 293,8	12 279,7	8 474,4	3 805,4	24 573,5	0,0	0,4	16,7
01.10.2012	5 969,2	6 405,8	2 140,8	4 265,0	12 375,0	12 282,5	8 529,4	3 753,1	24 657,5	0,3	0,7	14,8
01.11.2012	5 931,3	6 374,0	2 063,9	4 310,0	12 305,2	12 433,9	8 657,2	3 776,8	24 739,2	0,3	1,0	15,8
01.12.2012	5 975,4	6 484,1	2 148,2	4 335,8	12 459,4	12 621,2	8 798,3	3 822,8	25 080,6	1,4	2,4	14,4

Денежный агрегат M1 включает наличные деньги в обращении вне банковской системы (денежный агрегат M0) и остатки средств в национальной валюте на расчетных, текущих и иных счетах до востребования населения, нефинансовых и финансовых (кроме кредитных) организаций, являющихся резидентами Российской Федерации.

Денежный агрегат M2 включает денежный агрегат M1 и остатки средств в национальной валюте на счетах срочных депозитов и иных привлеченных на срок средств населения, нефинансовых и финансовых (кроме кредитных) организаций, являющихся резидентами Российской Федерации.

Денежная масса (национальное определение) 2012 г.

The first example is emblematic of the broader array of cases where a U.S person downloads data from a non-U.S. source. The second example is the reverse: A non-U.S. person downloads valuable data from a U.S. source. Data crosses borders in both cases.

First, note in both cases there is no charge for content. The Russian central bank is not charging U.S. economists for downloading data, and YouTube (and its corporate parent Google) does not charge the European teenager for downloading an educational video.

So if these two cross-border data transfers are going to create a monetary footprint and show up in the BEA statistics, it will happen because the telecommunications transport across national borders involves an exchange of money.

In the first case, the economist points his or her browser to the Russian central bank's website and downloads the data (figure 2). Obviously the economist or his or her institution pays a domestic internet service provider such as Comcast or Verizon for an internet connection. But unlike an international phone call, no extra money is paid for the foreign website.

The data request is passed from network to network until it reaches the Russian central bank, which then sends the money supply figures back again. At some point, that data request passes from a U.S.-owned network to a foreign-owned network.

For the sake of clarity of the example, let's assume that the U.S.-owned network also owns the submarine cable between New York and the United Kingdom, so that the interchange between the U.S.-owned network and foreign network physically occurs in the UK. [16]

Is there an exchange of money between the U.S.–owned and the foreign-owned network? Not necessarily. Large networks are often connected by *peering*, which is an agreement to exchange traffic without exchanging money. Peering agreements are so ubiquitous that they are basically conducted on a handshake, as one authoritative OECD study shows: [17]

> A survey of 142,000 peering agreements conducted for this report shows that the terms and conditions of the Internet interconnection model are so generally agreed upon that 99.5% of interconnection agreements are concluded without a written contract.

In fact, the largest global networks—the so-called 'Tier 1' networks—almost by definition peer with every other Tier 1 network.

It might seem like peering is a barter-type agreement that should generate revenue recognition on the financial books, even if no money exchanges hands. However, peering takes place mostly in situations of balanced traffic, so the revenues and costs would net out. The accounting firm KPMG notes that: [18]

> In our experience, peering arrangements between Tier 1 telecoms do not result in the recognition of revenue even though a service is provided and value is transferred between telecoms in much the same way as under traditional interconnect arrangements.

It's worth noting that peering is one of the key reasons that you can access websites from all over the work without having additional charges added to your internet bill.

---

[16] Many large providers either own their own undersea cables, have a share of a cable, or long-term rights to use part of the bandwidth.

[17] *Internet Traffic Exchange: Market Developments and Policy Challenges,* Dennis Weller and Bill Woodcock, OECD Digital Economy Papers 207, January 2013.

[18] "Accounting under IFRS:Telecoms", KPMG, January 2010.

Now let's turn to the second example, the European teenager accessing a video of a Stanford University computer science class from YouTube. We choose this example because YouTube is owned by Google, which has three full-fledged data centers in Europe. It is quite possible that the European teenager will be downloading a copy of the video from one of Google's European data centers, especially if it has been repeatedly accessed by other Europeans.

In that case, at some point the video will have been replicated in a Google European data center, via a cross-border data transfer from one of Google's U.S. data centers. Will this cross-border data transfer between Google and one of its affiliates be booked as a money exchange? Highly unlikely, given that an arms-length interconnection between two large telecoms would typically be a peering arrangement with no exchange of money. [19]

<u>Section II: Conceptual Considerations</u>

<u>Defining telecom imports and exports</u>

So far this paper has focused on some empirical questions on the measurement of cross-border data flows. Now we shift to the broader conceptual question about whether it even makes sense to talk about telecom exports and imports.

Historically the major cross-border data flow was the conventional international phone call. In the United States, the originator of an international phone call picked up a telephone, dialed 011, the country code and phone number, and paid an international charge to his or her phone company. The provider then paid the carrier in the receiving country according to a government-mandated settlement schedule. Conversely, the recipient of an overseas call did not pay an international

---

[19] See also "YouTube's Bandwidth Bill Is Zero. Welcome to the New Net," Ryan Singel, 10/16/2009, http://www.wired.com/business/2009/10/youtube-bandwidth/

charge—instead the overseas caller paid his or her local provider in their own country, who settled up with the U.S. phone company.

Under this scheme, calls from the U.S to overseas points were classed as imports, because the foreign carrier received the payments. Calls from other countries were classed as exports, since the payments came to the U.S. carriers. So if U.S. customers made more overseas calls than they received, the telecom trade balance would be negative. Indeed, that was true for many years. According to an FCC report from 1998: [20]

> U.S. carriers owe settlement payments for the services that they bill, and are owed payments for the services that the foreign carriers bill. In addition, U.S. carriers are owed payments for switched traffic that transits U.S. points. Because U.S. customers place far more calls than they receive and because U.S. carriers terminate more collect calls which generate surcharges for the originating carrier, U.S. carriers make net settlement payments to most foreign carriers. The total net payment for all U.S. carriers grew from $0.4 billion in 1980 to $5.6 billion in 1996. [21]

Reading this explanation, however, should make it clear that this definition of telecom imports and exports is an artifact of a regulatory convention that "calling party pays" for wireline calls. Suppose instead that we had a rule that "receiving party pays" as in a collect call or an 800-number. Under that regulatory regime, an outgoing international call would be treated as an export rather than an import, since the money would be collected by the foreign carrier, which would then settle up with the domestic carrier. Similarly, an incoming call would be treated as an import. Such a shift in regulatory conventions would immediately turn a telecom trade deficit into a trade surplus, without altering the final allocation of revenues to the respective telecom carriers. In addition, outgoing and incoming international calls were physically indistinguishable, in terms of the equipment used.

---

[20] "Trends In The U.S. International Telecommunications Industry" FCC, August 1998 http://transition.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/Intl/itltrd98.pdf

[21] In fact, current BEA figures show a telecom trade deficit of roughly $5 billion for 1996.

Why It Matters Where the Data is Accessed From

This example illustrates a more general problem with identifying exports and imports of data. We are used to trade in goods and services. Imports and exports of goods can be clearly identified. Goods have a location. An automobile manufactured in Germany and exported to the United States is located in Germany, in transit, or in the United States. It can't be in more than one place simultaneously.

Data is different. The YouTube video mentioned above still is stored in Google's data centers in the U.S., even after it has been replicated in Europe. The conventional definitions of exports and imports do not obviously apply to trade.

However, we will show in this section that the concept of location is still useful when thinking about data. All things being equal, the same data has a higher relative utility if it is accessed from 'nearby' in both geographic and network terms rather than from across the world. All other things being equal, a shorter distance between data and user, in geographic and network terms, leads to:

- A shorter length of time to get a response (latency)
- A lower probability for packets to arrive out of order or late (jitter)
- A lower probability for packets to be lost along the way and have to be retransmitted.

Some of these issues are related to distance, in the geographic sense. Going through more links over longer distances means that it will simply take longer to get a response. Some of these issues are related to distance in the network sense—for some source-users pairs, the best paths don't take the straightest routes. For example, the best internet path between many African countries often led through New York or London, because of a lack of good direct overland links in Africa.

How much of a difference could distance make in terms of response time? It's simple to do latency tests from your own computer, or to use some sites that are specifically designed to measure response times from different parts of the globe. Table 2 shows the response time of queries to various central banks around the world, measured from a European site.

Table 2: Examples of Latency

	Response time from Germany
European Central Bank (www.ecb.int)	7 ms
Swiss Central Bank (www.snb.ch)	13 ms
Bank of Russia (www.cbr.ru)	57 ms
Federal Reserve (www.federalreserve.gov)	98 ms
Bank of Taiwan (www.cbc.gov.tw)	295 ms

Data: Based on data from ping.eu

So the response time, from Germany to the Federal Reserve in Washington DC, is more than ten times greater than the response time from Germany to the European Central Bank, located in Frankfurt.

A systematic measure of latency comes from Verizon Enterprise, which publishes every month a table summarizing latency across its global business-oriented network. In December 2012, latency averaged 14 ms within Europe and 42 ms within North America, compared to 80 ms on transatlantic links and 112 ms on transpacific links.[22]

How important is a 100 ms or even 200 ms difference in response times? For one, it depends how often the request is repeated. For a single request for data from a remote website, a 100 ms lag may not matter. But if a radiologist in India is reading

[22] http://www.verizonbusiness.com/about/network/latency/#overview

x-rays performed in the United States, it's far better to have the data replicated in India to shorten the response time as the radiologist goes through the slides.

Some applications of the internet are more or less immune to lags and distance. For example, email—one of the earliest internet applications--is asynchronous. That means you can send it, and you don't expect an immediate reply. On a network level, the data packets that make up the email can arrive out of order and then be reassembled at the destination without doing any damage at all to the usability of the email.

But more and more, we're moving to uses of the internet where low latency and high reliability are desirable characteristics. For example, streaming video requires that the data packets arrive in order and without much variance. That suggests that streaming video from a U.S.-based data center to a European user is less desirable than streaming the same video from a European data center to the same user.

Applications that are sensitive to the location of the user relative to the source include:

- Financial trading—high-speed trading requires very rapid response times to changes in the data.
- Search—Companies such as Google and Microsoft have reported that faster response to searches means the users do fewer searches and spend less money. [23]
- Streaming video
- Voice over IP—distance degrades the quality of the connection
- Game play—a greater distance between the player and the game server makes the game less enjoyable to play
- Social media—as social media includes more pictures and videos, distance to the server matters.

[23] http://radar.oreilly.com/2009/07/velocity-making-your-site-fast.html

- Ecommerce—Amazon has found that consumers are more likely to buy if the website responds quickly
- Real-time group collaboration—as real-time video conferencing becomes more sluggish, it becomes harder to use.
- Healthcare—All sorts of health-related operations, from reading x-rays to monitoring patients, are sensitive to lags.
- Education—any lags in educational applications are discouraging to students.
- Any application that requires interaction with a large amount of data

The appropriate concept here is the idea that shorter lags and higher quality connections increase the relative utility of data. For example, YouTube videos of cats involve in different activities have a very large following on the internet. [24] (see Figure 3). Those videos will have a higher relative utility from a nearby server rather than from one halfway across the world.

[24] "In Search of the Living, Purring, Singing Heart of the Online Cat-Industrial Complex" wired.com, Gideon Lewis-Kraus, August 31, 2012, http://www.wired.com/underwire/2012/08/ff_cats/all/

Figure 3: Screenshot of a popular cat video

Replication

Let's suppose that there is a very large database located in the United States. It could be anything—financial data, genonomics data, videos of singing cats, products for sales, global weather data.

In theory, that data can be accessed from anywhere in the world. However, those links may be of varying quality, depending on the distance and network topology. It may take a relatively long time to get a response (latency). Packets may arrive out of order, lowering the quality of real time streaming applications such as video. Congestion means that packets may drop out, especially if everyone is trying to access at once across relatively limited cross-border links.

The relative utility of data increases if it can be accessed from a source that is 'closer'—both in distance and in the internet topology. Imagine that there is a data center set up in region A with useful data that is continually being updated and changed. Region B is fairly far away from region A (think separated by an ocean). Users in both regions A and B have access to this information through the internet. However, in general, users in region A will have 'better' access than users in region B (lower latency, fewer lost packets, etc).

Given data originally created and/or stored in region A, companies have a variety of strategies for increasing the proximity of that data to users in region B. These strategies include building new data centers in region B, setting up content delivery networks (CDN) nodes, or installing local caches. We will describe each of these below, but what they all have in common is that they involve *replication* of the data.

Replication means what it says—the data is replicated in another region. (See Figure 4) [25] Weller and Woodcock, in their recent OECD paper, note that: [26]

---

[25] There are subtle differences between the terms 'replication' and 'synchronization.' However, for the purposes of this paper I use the terms 'replication' and 'synchronization' interchangeably.

[26] Weller and Woodcock, 2013

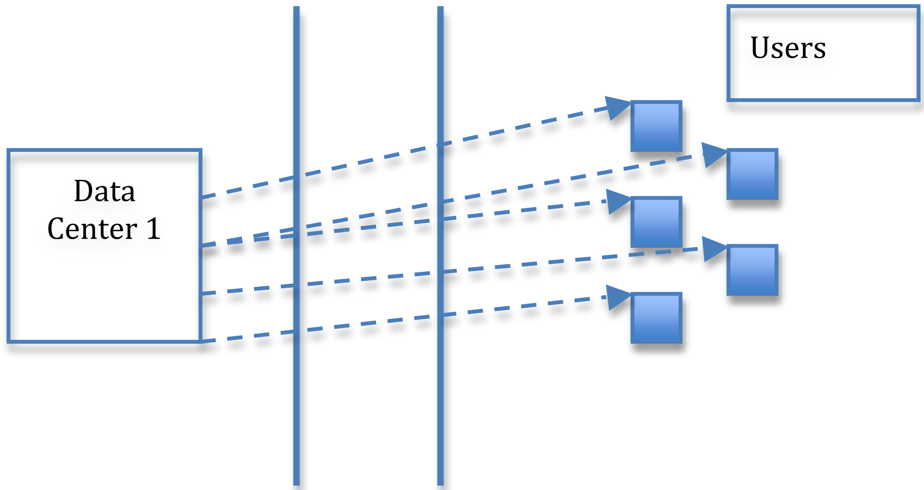
> …optimisation of price/performance of Internet services consists largely of reducing the distance between the source and destination of any packet of data, in order to allow it to travel more quickly, at a lower cost. …If the source and destination are both individuals or organizations, it is likely impractical to move them about, but if one end of the transaction is a server, particularly a server of non-unique information, as is the case of most common web servers, the data on the server may be replicated to many locations, so as to be proximal to as many users as possible."

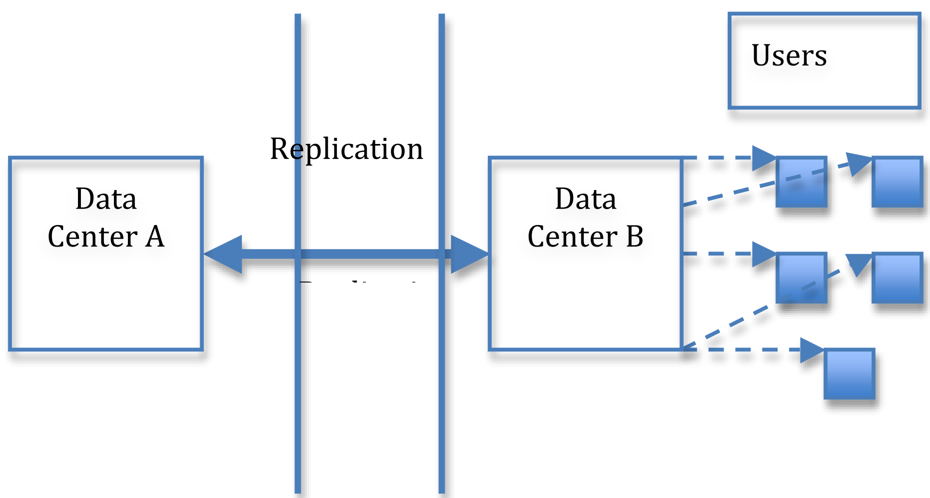
It should be noted that replication is a fundamental business activity for such large data-driven companies such as Google, Facebook, and Amazon. When such companies consider their internal cross-border movements of data, they think in terms of replication, not imports and exports. [27] Greatly simplified, when Google copies data from a U.S. data center (say, in Lenoir, North Carolina) to a European data center (say, in St Ghislain, Belgium), that move improves the relative utility of the data for Europeans. However, the data is being replicated rather than exported. No internal payments are made, and the data is still fully present in its original form in the original country.

Now, replication can either be simple or complex. Copying a file from one computer to another is a simple form of replication. Slightly more complex is the synchronization of the calendar on your phone with the calendar on your laptop. The program has to decide which entry is out-of-date and replace it with the latest changes.

[27] James C. Corbett et al"Spanner: Google's Globally-Distributed Database" https://www.usenix.org/system/files/conference/osdi12/osdi12-final-16.pdf

Figure 4: How Replication Works





On a global scale, of course, the problem of replication becomes much more difficult and requires a considerable amount of computational and physical resources. We can identify three different types of *replication facilities*:

- Duplicate Data Centers
- Content Delivery Networks
- Caches

We will discuss each of these in turn.

Duplicate data centers are exactly what they sound like. Companies will build data centers in various locations around the world that replicate part or all of the data in the other centers. The data can be generated by the company owning the data center (multinationals), be uploaded from other sources ("cloud computing"), or from search engines. The key here is to realize that data is always changing, so the data centers need to be updated at regular intervals. This generates a cross-border data flow.

Content delivery networks (CDN) are an example of one-way replication. As one study notes:[28]

> A CDN replicates content from the origin server to cache servers, scattered over the globe, in order to deliver content to end-users in a reliable and timely manner from nearby optimal surrogates.

Content delivery networks include Akamai, Limelight, Edgecast, Level 3, and BitGravity. CDNs typically have 'nodes' around the world. For example, Edgecast, a CDN founded in 2006, has locations on five continents, including Australia [29] ChinaCache, a China-based CDN operates data centers in California and New York City "to support global customers in delivering content into Asia".[30]

---

[28] "Taxonomy and Survey of Content Delivery Networks," Pathan, A. M. K., & Buyya, R. A , 2007, http://www.cloudbus.org/reports/CDN-Taxonomy.pdf
[29] http://www.edgecast.com/network/map/

Caches are another replication strategy used by large providers such as Google and Akamai. For no charge, they deploy caches within other networks that can be used to replicate content on the fly.

The Business Implications of Replication

Replication has costs and benefits, both private and public. It is not costless for a company to implement a replication strategy. That company must invest in physical equipment in the target region of replication: Computing power, storage, telecom links, buildings. To build a state-of-the art data center, like the one Facebook is building in Sweden, can cost from $700 million to well over a $1 billion. In addition, data centers are prodigious users of power, both for running the computers and for cooling, and bandwidth.

Why are companies willing to spend this kind of money? The key is that reducing latency enables them to improve the quality and accessibility of their services. That is, individual and businesses in Europe will be more willing to access Amazon or Microsoft if the response time is faster. That's straightforward.

In general, companies will require a return on replication in order to make it worthwhile. If a sporting event in the U.S. or Europe is being broadcast (replicated) around the world via a content delivery network, then the viewers of that event are likely having to pay a fee, directly or indirectly. Alternatively, revenue can come from advertising.

Growth Implications of Replication

However, if we take a step back, replication has broader public benefits that are more far-reaching than the private benefits. Investment in replication has the effect of increasing the intangible capital stock in a region. As data is accumulated in one

[30] http://en.chinacache.com/index.php/about-ccih.htmlindex.php?option=com_content&view=article&id=124&Itemid=131

region, it is available and useful in another region. It's as if we put up a building in New York, and it is replicated in Paris and Shanghai.

In that sense, replication of data from region A has a direct impact on the output and productivity of region B, because of the increase in the intangible capital stock available to region B. Users have to wait a shorter time for access to data, which makes the data more useful and boosts output and productivity. As data becomes more important to economic growth, the value of replication increases as well.

To put it another way, the building of replication facilities in region B—such as a data center which tracks and replicates data from region A—creates an economic surplus in region B. Region A pays for the creation of the data, while region B reaps all the benefits above the investment in the replication facility. Such benefits are presumably divided between the residents of the region and the enterprise.

To be more specific, let's assume that there's a very large and detailed database of construction information, drawings, and engineering information located in region A that is accessible to construction companies in region B via the Internet.

Assume that this intangible capital (construction database) is less valuable to region B if it has to be access via long-distance links. Response times are longer, and computer assisted design (CAD) programs that use the construction database run much slower and are almost unusable.

If that construction database is to be replicated within region B, it would have the effect of increasing the stock of intangible capital available to region B, even though the database was theoretically available before via the internet. To put it another away, the investment of physical capital within region B necessary for the replication would have the effect of increasing the stock of intangible capital as well. Similarly, if Google, Facebook, Microsoft, or Amazon build a data center in Europe, it has the effect of increasing the stock of intangible capital available to European businesses and consumers. North American and Asian data that were formerly accessible with a long lag are now available with a much shorter lag and higher quality.

This suggests that a replication facility in a region should be considered an extremely valuable economic asset for that region. By reducing latency and improving response time, it makes global data much more usable and valuable to the country. This is the international equivalent of the "broadband bonus."

How much more valuable? How much is a replication facility worth to a region where it is located? In theory it is possible to calculate how much a replication facility reduces average response time for a fixed geographical basket of requests. For example, given a basket of one-third data requests to Asia, one-third to North America, and one-third to Europe, it would be possible to calculate how much a replicating facility/data center in Europe reduces the average latency of responses.

However, it is much more difficult to put a monetary value on the reduction of average wait times. One possible approach is to look at a mix of different application—videos, education, health-- and analyze how each of them would be affected by a decline in latency. This is a doable but not easy exercise.

We note that international internet companies such as Google and Amazon are currently the major builders of replication facilities in Europe and Asia. What's more, in any region, they replicate far more data from the rest of the world than they collect from that country. Indeed, it seems likely that a replication facility in a region will replicate more rest-of-world data to the region than it replicates local data to the rest of the world. That's just a function of the fact that any region's data makes up a relatively small share of the global whole.

Replication vs. Exports/Imports

How does the concept of replication compare to the concept of exports and imports? The way that exports and imports are typically handled in the national income accounts imply rivalness. When a good is shipped from one country to another, it is available for consumption in the importing country and no longer available for consumption in the exporting country. Exports in general enter the calculation of gross domestic product with a positive sign, while imports enter the calculation of gross domestic product with a negative sign. [31]

Replication, by contrast, applies best to data and other nonrival assets. The replication of a database from region A to region B increases the intangible capital stock of region B without reducing the intangible capital stock of region A, or region A's ability to produce goods and services.

To use a specific example, one type of intangible capital stock is 'entertainment, literary, and artistic originals', including films. Licensing the right to show a film in a foreign country currently shows up as an export in the national income accounts. However, such a license generally does not reduce the ability of American consumers to view the film, and does not reduce the intangible capital stock of 'entertainment, literary, and artistic originals' (two ways of saying the same thing).

It might be more natural to think of licensing the right to show a film in a foreign region as an example of replication. It no longer shows up in the export account, then, or as part of domestic GDP. In fact, to the degree that the actual replication happens via a content delivery network or data center located in the foreign region with the license, the physical investment required for replication may not be in the U.S. at all.

Section 3: Policy Implications

[31] GDP = Consumption + Investment + Government output + Exports - Imports

From the previous section, we can make two observations:

- Companies make voluntary decisions about whether they will invest in a replication facility in a region. Such facilities can be expensive to build and run.
- Attracting replication facilities to a region is a big plus for economic growth. The benefit of replication facilities is the way they make global data more useful and accessible to local residents and businesses, thus augmenting existing intangible capital.

Together, these suggest that governments should be welcome the construction of additional replication facilities in their regions, just like they welcome the construction of new factories by foreign companies. At a minimum, governments that are concerned about growth should not try to make it more costly and problematic to build data centers or other types of replications facilities.

However, there have been several recent proposals for increasing the tax rate paid by international internet companies, or imposing additional regulations on them. For example, a recent paper from the French government suggested a sort of tax on data.[32] Meanwhile the proposed European data protection regulation would limit what could be done with personal data gathered in Europe, or limit the movement of data across national borders.

However, the whole point of a replication facility is to replicate data between different locations. To the degree to which the replication function is restricted or taxed, the economic rationale for investing in a replication facility is reduced. Imposes an enormous deadweight loss on growth. Weller and Woodcock note:

---

[32] "Mission d'expertise sur la fiscalitié de l'économie numérique," Pierre Collin and Nicolas Colin, January 2013. http://www.redressement-productif.gouv.fr/files/rapport-fiscalite-du-numerique_2013.pdf

> It is also the case that regulations that are not explicitly intended to apply to Internet traffic exchange may have that effect. For example, restrictions on the ability to export certain data, such as customer profiles, intended to protect security and privacy, may also limit the development of Internet topology and the growth of Internet assets in some regions. Similarly, tax policies in each country toward broadband and Internet businesses are likely to affect the choice of the locations for investment in Internet assets.

Consider, for example, a policy that makes it difficult for companies to move data across national borders. Perversely, such a policy is likely to discourage the construction of replication facilities by lowering the return on investment. That, in turn, will have the potential effect of reducing the stock of intangible capital in that region and hence lowering economic output.

The whole point of a data center is to create benefits by moving data closer to users and encourage the free flow of information. Anything that impedes that flow is likely to reduce the overall benefits.