



HAL
open science

Chasing data in the Intermediation Era Economy and Security at stakes

Aurélien Faravelon, Stéphane Frénot, Stéphane Grumbach

► **To cite this version:**

Aurélien Faravelon, Stéphane Frénot, Stéphane Grumbach. Chasing data in the Intermediation Era Economy and Security at stakes. 2015. hal-01107365v1

HAL Id: hal-01107365

<https://hal.science/hal-01107365v1>

Preprint submitted on 20 Jan 2015 (v1), last revised 11 Nov 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chasing data in the Intermediation Era

Economy and Security at stakes

Aurélien Faravelon¹, Stéphane Frénot¹ and Stéphane Grumbach ^{*2}

¹University of Lyon

²INRIA

Abstract

Intermediation is the action to match two types of actors (users, clients, services, etc.) in a world with incomplete information, where the matching would have been rather difficult without intermediaries. We show the increasing role of on-line intermediation platforms in the economy, and their growing responsibility for ensuring global security of people and society. Our contribution is twofold. At a theoretical level, we better define algorithmic intermediation by establishing a hierarchy of intermediation actors, based on their degree of abstraction from specific services. Then, surveying data from web analytics tools such as Alexa.com, we show that there are only few important intermediation platforms, headquartered in few countries. These results underline the strategic importance of intermediation actors in terms of security and that the more abstract they are, the more influence they reach.

1 Introduction

The growth at an unprecedented pace of the production and exchange of data in digital form has led to the extremely rapid development of new industries. Their services disrupt the old economic models and raise complex social issues, ranging from the protection of individuals and their privacy, to threats to the society as a whole and its security. New equilibria

are taking shape between corporations of the digital age that spread their activities worldwide across borders on the one hand, and states that have legal responsibilities and legitimacy over restricted and well-defined territories on the other hand.

In this paper, we show the dominant role of on-line intermediation platforms, and their increasing responsibility for ensuring global security of people and society. Intermediation players are of tremendous importance for their growing impact on the world economy, their fantastic promises of changing our lives and our organisations, as well as for the profound disruption they impose on traditional actors in both the private and the public sectors.

Intermediation, which dates from the pre-digital era, is not a new activity. The term comes from the financial sector, and designates the capacity of a financial institution - such as a bank - to match the funds of depositors with the needs of potential borrowers. The absence of complete information leads to the existence of intermediaries, able to make the match, that single actors would not achieve, while making profit out of successful matches [1]. Intermediaries are moreover in a position to create new services such as, very importantly in the banking sector, the mutualisation of risks.

The advent of new technologies for data processing has completely changed the intermediation business. The complexity of financial intermediation, for instance, has grown tremendously since the 1980s, and has led to the handling of high frequency data flows, and very abstract financial products. It has also lead

*Electronic addresses: `firstname.lastname@inria.fr`

to the development of powerful global intermediaries in sectors historically dominated by local actors.

Intermediation is made both possible and necessary by an asymmetry of information. The tremendous growth of data in the digital era has created avenues for intermediation in an unbounded number of sectors. The first new intermediation service of tremendous importance has been developed in the late 1990s. It is the search engine, which intermediates between users and the knowledge they seek, and has revolutionised our access to knowledge. Since then a powerful industry has emerged with brand news services, such as social networks, forums, reservation platforms, e-commerce, etc, which enjoy a remarkable growth.

Data fuels intermediation corporations the way crude oil fuels the traditional industry. Leaders in these sectors are now rivaling with the oil industry as the top capitalisations. On the web, the leading corporations, such as Google, Facebook or Amazon, are all in the intermediation business.

Intermediation corporations are not all alike though. Our first contribution is to establish a hierarchy of intermediation actors, based on their abstraction level. We introduce four levels that range from production actors, involved in the production of the products or services they offer, such as media corporations; then distribution actors, essentially distributing products produced by others such as Amazon or Netflix for instance; sectorial intermediation actors, such as LinkedIn or eBay; and finally intermediation platforms, which offer essential services to support intermediation as well as other services provided by third players.

We then show that intermediation actors occupy a central place in the new economy. The web has a “bowtie” structure [9]. A set of sites - at the center of the bowtie - are well-referenced and interconnected. This set is surrounded by a set of sites which mainly reference, through hypertext links, the center of the bowtie, and a set of sites which conversely are mostly referenced by other sites. The center of the bowtie is dynamic and evolves with the rise and fall of sites. The connections are also evolving with the advent of new technologies, such as dynamic web features implemented with Ajax [8] for instance, and not only

hyperlinks.

Being in the center of the bowtie is crucial for business as well as influence. Various types of analysis of the centrality of positions have been made, such as for specific top-level domains for instance. The presence of British universities on the TLD co.uk has been analysed in [7]. It has also been applied to the study of specific sectors sites such as hotel sites in order to measure the effectiveness [11] of trackers - i.e., sites which track users on other sites - in order to assess their income.

As our investigation demonstrates, intermediation actors occupy the center of the bowtie, when distribution sites linger at the extremities of the bowtie¹. Our metrics, based on the flow of visits of the top actors of the Web according to Alexa², reveal that the higher the abstraction, the larger the traffic. In addition to occupying central positions, we also show that the power of intermediation systems follows a power law, with a very rapid decrease of their influence.

Intermediation platforms are revolutionising the world of services. First, by offering new services such as the search engine of course, but more importantly by disrupting traditional services of the old economy, such as hotel reservations, transportation, the press, publishers, to name a few, which have strongly reacted against the disruption they endure. The success of the platforms is due to their capacity to offer better services than those previously offered to users. They rely on powerful algorithms to propose the best matches. They maintain a close and direct connection with their users, wherever they are on earth.

These new services are changing the rules of the game in many economic sectors [6]. As they build new services and gather numerous users and penetrate all economic sectors, intermediation platforms disrupt legacy economic models [6]. For instance, Amazon is well known for disrupting the publishing business model. As platforms acquire power, they may be in a dominant position and weigh on policies for instance [12]. Sometimes, they conflict with local legislations. Therefore, the relationship between corporations controlling these new services and states

¹See Figure 8.

²alexacom

is of tremendous importance. Geographical territories, where laws apply and which are limited by boundaries, conflict with “digital territories” which are frontierless. This conflict strongly imparts data security.

We focus on a group of 30 countries chosen for their economic importance all around the world. We show that if there is extreme concentration of power on a rather small set of platforms, the same power law applies to the countries hosting platforms, with a superpower, the USA, an emerging power, China, and then a handful of modest powers.

The dependency of most countries from foreign systems raises many issues related to trade, sovereignty, security, and values. Global issues of data flows and data storage are at the heart of today economic, political and business agendas. Edward Snowden’s revelations starting in june 2013 have made the debate on these technical issues accessible to a large audience. Consumers regularly protest against the lack of privacy protection when states pass new laws and corporations adjust their storage strategy according to data policies, left aside financial or fiscal arguments which are beyond the scope of this paper, to install their data centres.

We thus consider the domains on which platforms and states are conflicting, which include in particular labor laws, taxation, money, privacy rules as well as security. Conflicts have led to trials, new laws, as well as interruption of service for platforms over a territory, as was recently the case for Google News in Spain. We then consider the common grounds on which platforms and states are working in mutual interest, which include economic as well as security issues.

Our study relies on data collected from web analytics databases such as Alexa.com and Trafficestimate.com. Data were collected on 19 december 2014. In order to conduct our study, we have developed several javascript plugins which retrieve the necessary data and pre-process them to allow their visualization.

In the body of the paper, we study intermediation platforms. We need to precise the vocabulary we use. A *platform* is a a set of services - such as mail or social networks. A platform belongs to a *corporation*

- such as Google and Facebook - which can possess several platforms. A platform may be distributed over several *sites* - such as google.com, google.fr, etc.. Throughout our study, we consider platforms - i.e. we group the metrics of the different sites belonging to the same platform.

The paper is organised as follows. In the next section, we define more precisely the concept of intermediation and present a hierarchy of actors based on their abstraction level from specific services. In Section 3, we show the relationship between intermediation platforms and countries. Sections 4 and 5 are devoted to respectively the most powerful platforms and countries, demonstrating the extreme concentration of the intermediation power both in the industry and over the geography. Finally, in Sections 6 and 7, we consider the role of platforms as disruptive operators - especially in the field of security and then we study their relations with states.

2 The rise of intermediation

Intermediation is the action to match two types of actors (users, clients, services, etc.) in a world with incomplete information, where the matching could not be performed without intermediaries for the lack of information. In addition to achieving the matching, it gives rise to new services that only intermediaries are in capacity to deploy. This has been the case in the financial sector, where intermediation has long been used to match the funds of depositors and borrowers. What is new is the use of algorithms to carry intermediation in potentially all economic sectors.

Algorithmic intermediation is being generalised with the explosion of data and the advent of complex data analytics. Its role in the digital economy is already remarkable. It applies to a large spectrum of sectors of activity, with distinct economic models. We first propose to distinguish different types of intermediation depending upon their implication in the products or services exchanged, and then show that the level of abstraction impacts on the influence.

A service is “*abstract*” if it does not focus on a specific usage. For instance, a social network is an abstract service as it offers a wide set of functional-

ities not restricted to an economic sector, and more importantly it allows to build other functionalities on top of it using its API. In contrast, an online shop provides a limited amount of functionalities.

We distinguish four categories of intermediation actors, ranging from a low degree of abstraction to a high one. Figure 1 shows this ranking ranging from services mimicking an industrial model - they sell the goods or services they produce - from intermediation platforms which do not buy nor sell anything but connect users and services.

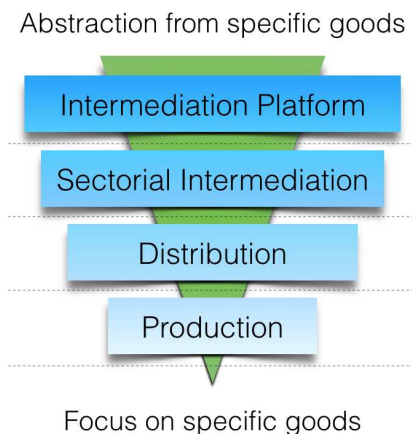


Figure 1: Degree of abstraction of the intermediation

At the bottom of the hierarchy, the PRODUCTION category encompasses industries that produce goods or services and sell them online to their customers, essentially with a very restricted form of intermediation between their services and their customers. The press constitutes a good example of such industries, with a direct relationship to their readers.

One level above, DISTRIBUTION corporations commercialise goods produced by others. Netflix is a good example of that level, giving mostly access to cultural products they do not produce but distribute to their customers. Amazon is another example for retail.

The next level, SECTORIAL INTERMEDIATION, includes corporations which provide online services which allow their users to connect with specific goods

or services. The search engine belongs to that category together with numerous commercial actors, such as Blogspot or LinkedIn for precise economic sectors.

Finally, the highest level, INTERMEDIATION PLATFORM, is constituted by corporations that offer an ecosystems on top of which others can build and distribute their services. Facebook and Google, for instance, are the most prominent actors in this category. To some extent, at this level, corporations offer a sort of global operating system disconnected from physical supports, that allows the development of unbounded types of activities.

For some corporations, their category might be arguable. Amazon, for instance is mostly in the distribution category but operate as well as sectorial intermediation while Netflow also produces cultural goods. We chose for each of them the category which corresponds to the business model of their main activity. For PRODUCTION systems, the intermediation activity might be shallow. The press for instance can be seen as intermediating between journalists and their audience. But what is of interest to us, is that although the intermediation might be shallow, there is a possibility of desintermediation of the activity, which is what is going on in the press in particular.

Corporations of all categories harvest data, which is of strategic importance for their activity. Production systems might get a good knowledge of their users - such as their purchase history or specific tastes. Yet, it is really at the level of distribution systems, that very sophisticated *recommendation* algorithms are used, such as those deployed with great success by Netflix or Amazon. For Sectorial Intermediation systems, the main challenge is to maintain direct connection with their users, while ensuring the gatekeeping to the services they propose with no middleman. As they do so, they collect a large amount of information from their users. Eventually, intermediation platforms also ensure a direct link to their users, and try to become a universal gate to all services. For that purpose, they facilitate the development of services developed by others, thanks to the API they offer. They thus collect data about their users and about the traffic generated by applications built on top of their services.

Intermediation platforms are the main players on

the web as they attract hundred of millions of users and thus collect the largest amounts of data. The Top 25 global sites according to Alexa correspond to 22 distinct platforms. Out of these 22 platforms, 21 belong to intermediation corporations: 7 are intermediation platforms and 14 belong to the sectorial intermediation category. Only one belongs to the distribution category, Amazon. Table 1 presents a classification³ of the Top 25 sites according to Alexa.

Type of service	Corporation	
INTERMEDIATION PLATFORM	US	Live, Facebook, Yahoo, Google,
	CN	QQ, Weibo
	RU	VK
SECTORIAL INTERMEDIATION	US	Youtube, Wikipedia, Twitter, LinkedIn, Bing, Blogspot, Ebay
	CN	Baidu, Taobao, Hao123, Sina, Tmall,
	RU	Sohu, Yandex,
DISTRIBUTION	US	Amazon
PRODUCTION		

Table 1: Repartition of the Top Global 25 actors

The absence of production corporations from the Top 25 may be explained by the dependence of their activities on a specific geographic territory. Customers are likely to buy goods from corporations which easily ship to their country. Moreover, intermediation corporations are over-represented because they offer so many functionalities that users are likely to have to use them.

The domination of intermediation corporations translates into a major weight in terms of activity as shown in Figure 2. We retrieved the amount of visits each corporation in the Top 25 attracts from trafficestimate.com and sum this amount for each category. Large orders of magnitude separate each level of abstraction of the intermediation. DISTRIBUTION corporations only attract 1,1 billion of visits in one

³The table contains only 22 corporations because some of them, such as Google, own several sites in the world Top 25.

month when SECTORIAL INTERMEDIATION corporations attract 9,4 billions of visits and INTERMEDIATION PLATFORMS 15,1 billions of visits⁴. Once more, as intermediation platforms offer a large range of services which can be used for pretty much anything, they are able to attract most visits. They also capture most data on the web.

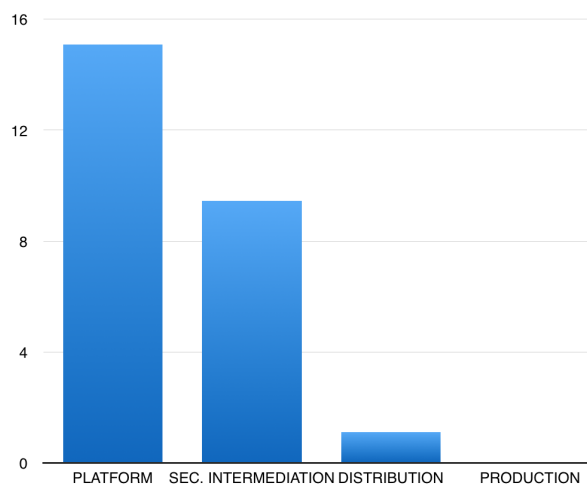


Figure 2: Top 25 global sites

3 Platforms and Countries

Large intermediation systems play a fundamental role in the economy. They offer essential services, much like electricity or water supply, which are used by people as well as corporations to support their services. Unlike essential services, such as energy, telecom, or transportation, which are strongly regulated in all countries by local authorities to ensure their fair distribution over the territory and in the whole population, the new services are provided by multinational corporations and are only starting to get the attention of the legislators. We will see in Section 7, some examples of conflicts between corporations proposing intermediation services and states.

⁴Data were retrieved on December 2014, 19

The relationship between countries and platforms is of outermost importance when considering issues such as privacy and security. Indeed, both issues are related to the legal framework implemented in each country, while platforms operate on a global basis, over a borderless territory. Hosting major digital corporations is thus a key factor in the power of a country.

The rather unbalanced distribution of platforms in the world leads, to us, to an unbalanced geopolitical situation which motivates our investigation. Figure 3 shows the difference of context in the US, China, Russia, France and the United Kingdom (UK). For each country, the proportion of national actors from the country’s Top 25 sites belonging to a specific category (such as Intermediation platform, distribution, etc.) operating on its territory is shown.

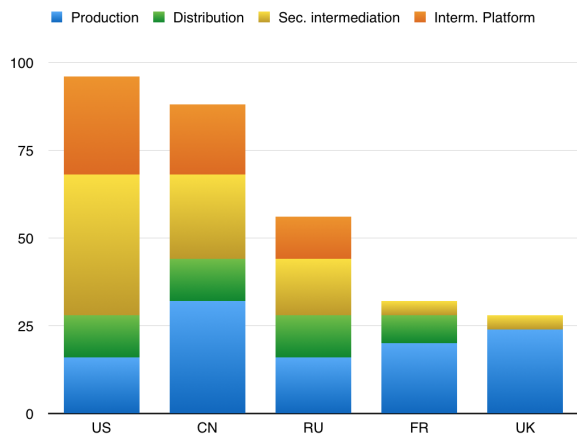


Figure 3: Proportion of national corporations in each intermeditation level of the 5 permanent members of UN Council

The Production category appears in all national Top 25 list when it was missing from the global Top 25. Users, indeed, access online newspapers and buy goods from national retailers. However, only three countries possess intermediation platforms and, in the UK and France the sectorial intermediation corporations are mostly dedicated to goods selling. Clearly enough, the US host most of the sites which

their citizens visit, the same stands for China. Both countries may thus control these sites, which are under their legal framework. Russia still owns most of the corporations which attract its citizens but in a smaller proportion. In contrast, most sites which operate in France and the UK are hosted by foreign countries. Thus, there is a conflict between digital and geographic borders.

The conflict between legal, geographic and digital borders is of great importance for data security. Knowing where sites are hosted and where datasets lie allow to assess their security level. Corporations depend on legal and political frameworks and can tune their business strategies according to them. Eventually, some countries are in a strong position - they accommodate corporations and their data and are able to perform data surveillance - when others are powerless. The regulations on data influence the relations between states, as shown by the negotiations on the Transatlantic Trade Agreement (TTIP).

Conflicting frontiers also influence the relations between countries and corporations. Countries and corporations may collaborate or share the definition of values such as privacy. Yet, sometimes the policies of corporations conflict with the ones of countries and reshape the context of cybersecurity. We have gathered some of the potential relations between countries and digital corporations in Figure 4.

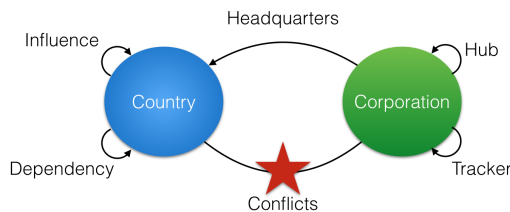


Figure 4: Relations between corporations and countries

In the rest of this paper, we strive to analyse these relations in order to capture the resulting power and economic equilibria. As we do so, we picture the security context of web data and account for current conflicts. We have considered 30 countries, which

include the Top 25 countries of the world based on their 2013 Gross Domestic Product (GDP), as published by the CIA World FactBook⁵. Additionally, we have included countries, which figure in the Top 5 GDP in their respective continents, but did not make it to the Top 25 global GDP list, to ensure a sufficient representation of each region of the World. Table 2 shows the final selection of countries.

World Area	Country
Asia and Oceania	China (CN), India (IN)
	Japan (JP), South Korea (KR)
	Taiwan (TW), Thailand (TH)
	Indonesia (ID), Australia (AU)
	United States (US)
America	Brazil (BR), Mexico (MX)
	Argentina (AR), Colombia (CO)
	Venezuela (VE), Canada (CA)
	United Kingdom (GB),
	Germany (DE), Italy (IT)
Europe	France (FR), Poland (PL)
	Spain (ES)
	Netherlands (NL), Russia (RU)
	South Africa (ZA), Egypt (EG)
	Algeria (DZ), Nigeria (NG)
Africa and Middle-East	Turkey (TR), Iran (IR)
	Saudi Arabia (SA)

Table 2: Selected countries

For each country, we have considered its Top 25 sites, according to their Alexa rank. Alexa’s Traffic ranks are based on the traffic data provided by users in Alexa’s global toolbar panel over a rolling three months period. A site’s ranking is based on a combined measure of *Unique Visitors* and *Pageviews*. Unique Visitors are determined by the number of unique Alexa users who visit a site on a given day. Pageviews are the total number of Alexa user URL requests for a site.

The restriction to the Top 25 sites might seem severe. It is in fact rather meaningful. Indeed, the Top 25 sites represent about half of the traffic of the Top 500 sites in average in most countries. The Top

⁵<https://www.cia.gov/library/>

25 sites of each country therefore amounts to a significant part of the data flows and hence give a good estimate of the overall picture. The overall number of sites, Top 25 in the Top 30 countries, amount to 419 sites, which can be seen as the most influential worldwide. Our analysis brings two main findings. First, there is only a small number of actually influent sites. Second, these sites tend to be located in a small number of countries. Influent sites thus turn out to be in almost monopolistic positions. The same stands for countries where the corporations owning these sites are headquartered. We analyse these two aspects in Section 4 and Section 5.

4 Platforms of influence

In this section, we focus on the relations between platforms and countries. We hypothesize that only a few platforms have a strong influence on the web, a small set of platforms influences most countries. Indeed, the web traffic to the platforms follows a power law. Figure 5 shows the rapid decrease of the global traffic on the Top 25 platforms in the world. When a platform is distributed over several sites (such as Google.com and Google.co.jp), we have aggregated the visits to the sites and labelled them with the platform’s name.

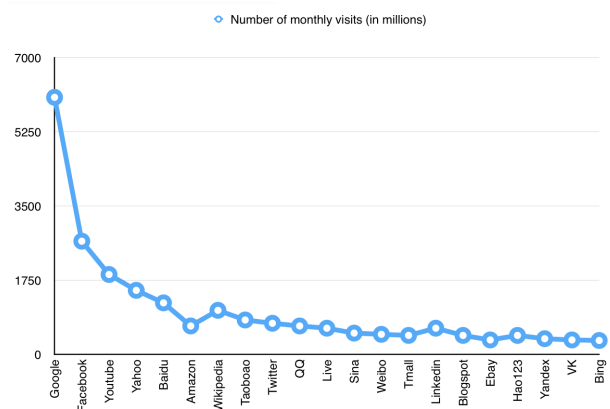


Figure 5: Global traffic of top world corporations

As visible, there are only few influent platforms: Google comes first, followed by Facebook. The Top 25 platforms attract most of the visits and, most likely, most of the data. Thus, they are major economic powers. For instance, in 2013, Amazon was larger than the next dozen Internet retailers combined⁶. As they attract users, platforms are able to provide API and turn into platforms on top of which other sites and services are built.

Yet, there is a strong imbalance between the platforms. Google receives twice as much traffic as any other platform. This observation echoes the domination of Google over the global Top 25 sites.

We have considered the international audience of platforms, defined as the number of countries for which they belong to the Top 25. In order to identify the most influent platforms, we study the international audience of the global Top 50 sites according to Alexa. For each site, Figure 6 shows the percentage of our 30 countries, for which the site belongs to the Top 25. The international influence also follows a power law. Only 9 platforms belong to 50%, or more, of the top 25 lists we study. From the 10th to the 50th sites, there are only “local” platforms. They only influence a small set of countries.

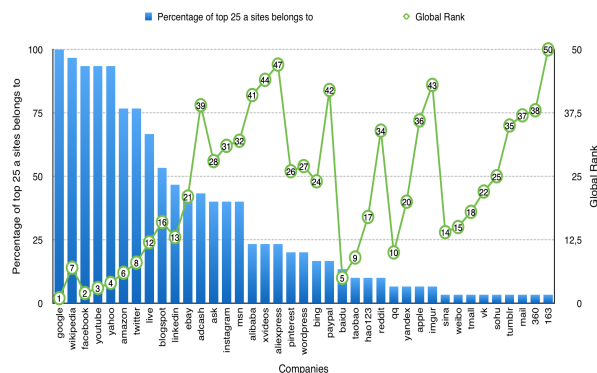


Figure 6: Top 50 sites international influence

Interestingly enough, we see that the international influence of platforms does not exactly mimic their rank. For instance, Baidu is ranked 5 worldwide on

⁶<http://www.forbes.com/>

Alexa. However, it only belongs to 4 Top 25 national lists. This situation may be explained by the rank’s mode of computation (on Alexa, a rank is a combination of average daily visitors and page views over the past three months) and the corporation’s international development strategy. Baidu mainly is available in Chinese and Japanese, it may enjoy a large number of visitors but they are likely to be located in a small set of countries. Figure 6 also shows that local policies may not hinder the development of platforms. Google is officially absent from China but Google.co.hk appears in the Chinese Top 25 sites.

At any rate, Figure 6 highlights Google’s dominant position. This place partly comes from the global use of Google services. However, not only does Google influence all the countries we study. It also attracts a large part of the traffic to the Top 25 sites of each country. Figure 7 presents the percentage of traffic Google attracts according to the traffic to the Top 25 of each country.

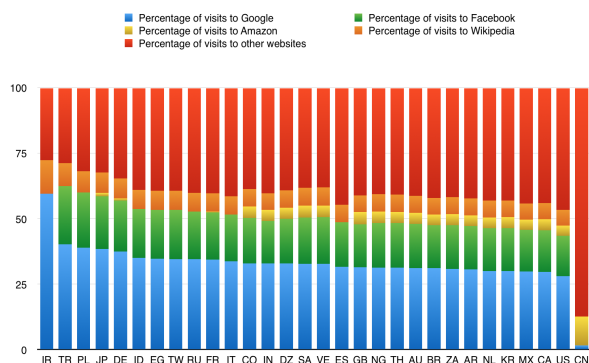


Figure 7: Influence of Google

In all countries but China - where Google does not officially work - Google gets more than a quarter of the traffic dedicated to the Top 25 sites. Interestingly enough, Google is especially strong in Europe when it is an American corporation. This situation may reflect the absence of alternative search engine in Europe for instance.

Figure 6 also shows that Google overpowers all the other platforms. Indeed, it has become customary

in Europe to discuss the Influence of the “GAFA”, namely Google, Amazon, Facebook and Apple on the digital economy. Apple may not be a major web actor as it mainly sells hardware and influence the mobile market. However, Figure 6 shows that Google oversteps the influence of other platforms with a comparable rank. Indeed, Facebook only has half of Google’s influence. Amazon does not belong to all the Top 25 lists we study and Wikipedia only has less than 20 percents of Google influence. We have underlined earlier that the Top 25 world sites represent half the traffic to the Top 500 sites. Following this observation, Google would attract roughly 10 to 12% of the traffic to the Top 500 world sites. This extrapolation emphasizes, if needed, the global power of Google.

We have studied so far the influence of platforms over the world. However, the literature on the structure of the web and the development of ecosystems on top of which other services are built, shows that some platforms may dominate the flow of visits to others and play the role of hubs. Doing so, these hubs accumulate data about users and their behaviors.

In order to test this hypothesis, we have retrieved from Alexa the “upstream” information, i.e. the set of sites users visit before visiting a given site. We have evaluated our hypothesis on the Top 25 sites in France, the result is displayed on Figure 8⁷.

We can see three main groups of sites. In the first group, sites are well connected: users surf between these sites. These sites mainly are *intermediation platforms*. The second group consists in sites from which users navigate to others. These sites thus are *hubs* to other sites. Eventually, the remaining set of sites contains endpoint sites to which users surf.

These three sets indicate complex relationships between corporations: the platforms and the hubs control the streams of visits and may collect data out of it, through cookies or social tools for instance. As they do so, they may develop new innovative services. Google, for instance, has recently announced it was able to process usage data from web users in order to determine if users were actual humans instead of forcing them to use captchas⁸.

⁷When several sites belonged to the same platform, we have grouped them.

⁸<https://www.google.com/recaptcha/intro/index.html>

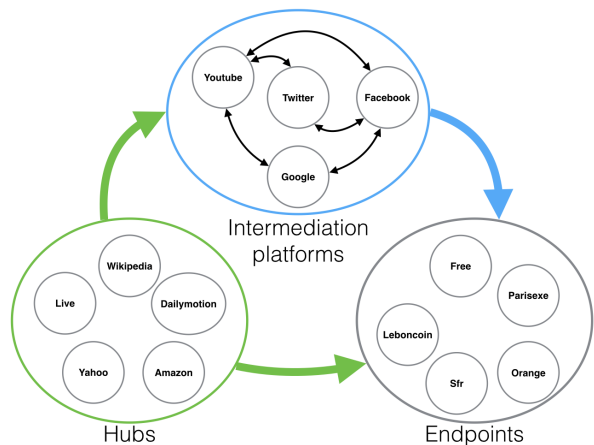


Figure 8: Platforms as major web nodes

These three sets also display another side of the relationships between countries: locally hosted sites may belong to a country’s Top 25 but the stream of visits is still centered around intermediation - and mostly foreign - platforms. Indeed, intermediation platforms and sectorial intermediation corporations are pretty much all American in the French Top 25 sites⁹.

5 Countries of influence

We have seen in the previous section the extreme concentration of activity over a few intermediation platforms, particularly for those having a global reach, leading to a power law. The same phenomenon can be observed at the level of geographical territories. Only a handful of countries host the headquarters of influential web corporations. The location of the headquarter of a corporation is of course not the only criteria. The location of the data centres is of considerable importance as well, but countries hosting headquarters most often also host large amounts of data centres.

⁹The relations between the three sets remains the same for other countries. We have checked our results for the US and China. However, the influence of foreign sites is not as important for these countries.

Figure 9 displays the countries which host influential corporations, that is those which own influential platforms. Out of the 30 countries we considered, only 11 do so. The US host most of the influential corporations, while the other countries considered, apart from China, only host one influential corporation.

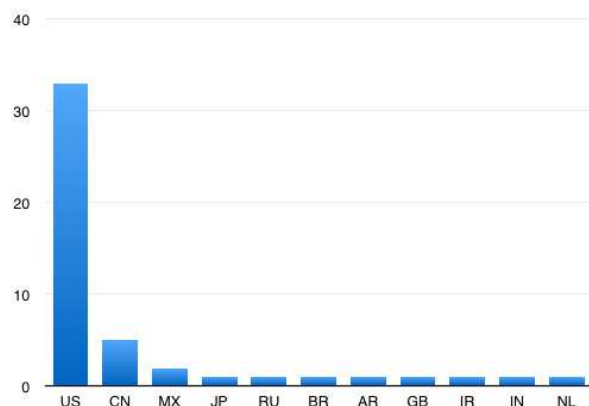


Figure 9: Number of influential platforms by countries

By hosting corporations headquarters and data centres, countries exert an influence on other countries, reached by the services of these corporations. Figure 10 displays the flow of visits and this influence for a subset of the 30 countries we study. This influence is twofold. First, by the impact of their services on local usages and practices. Second, by the capacity to handle the data of their users, which fall under their privacy and security frameworks.

We introduce basic metrics to measure this notion of influence. They can be refined of course, but the results even with rather coarse metrics are striking. We consider that a country X influences another Y , if a corporation, headquartered in X , has platforms in the Top 25 of Y . For each country Y of our investigation, we compute the number of countries influenced by a country X .

Figure 11 displays how many countries a given country influences. As most influential platforms belong to corporations headquartered in the US, they

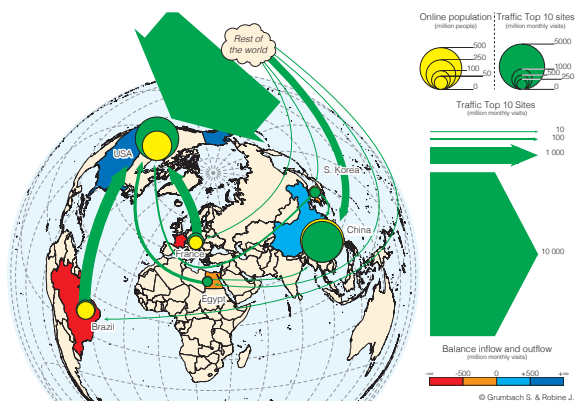


Figure 10: Traffic in top 10 sites in selected countries.

reach in fact all the the countries we studied. Chinese influential platforms reach 8 countries, while Brazilian ones do so in 5 countries, Argentinian ones in 3 countries, British and Mexican ones in 2 countries. The influential sites of the other countries only reach one country but the country where they are headquartered.

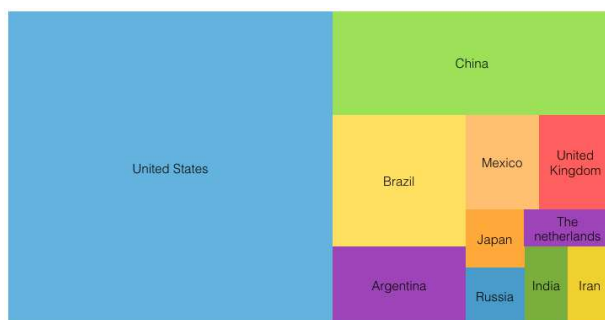


Figure 11: Size of influence zone by countries

We identify three categories of countries. “*Global powers*” influence most countries, as they host a large number of influential platforms. Only the US fits into this category. “*Regional powers*” reach a limited number of countries, generally in a regional or linguistic area. The 10 subsequent countries of Figure

11 belong to this category. Eventually, the 19 other countries we have studied do not host headquarters of any influential platform, we consider them as “*influenced countries*”.

In order to capture the difference between these categories, let us keep in mind that all Top 25 global platforms are American and consider a particular country, France, which is an “influenced country”, representative of the situation of most European countries. Its Top 25 platforms are shown on Table 3. Most of the sites are foreign platforms. If we consider the traffic, the situation is even more impressive, with only 22% of the national traffic on national platforms. Foreign platforms thus capture 78% of the traffic of the Top 25 platforms in France together with the associated harvested data.

Since, according to our measures, the Top 25 platforms represent half of the traffic to the Top 500 platforms, it means that at least 39% of the traffic to the Top 500 platforms in France goes to foreign platforms.

Type of service	Corporation
PLATFORMS	Google, Facebook, Yahoo,
SECTORIAL	Youtube, Wikipedia, Ebay, Twitter, LinkedIn, MSN, Leboncoin, Adcash,
INTERMEDIATION	commentcamarche.net, bycontext, Adservices,
DISTRIBUTION	Amazon, CDiscount, Lafnac,
PRODUCTION	Orange, Free, SFR, Lefigaro, pagesjaunes, Pole-emploi

Table 3: Repartition of the Top 25 actors in France

Actually, out of the 30 countries we considered, most follow a pattern similar to France as shown on Figure 12. Most of the traffic of all countries goes to US sites, about a third to national sites, and a tiny portion to sites of third countries.

We have selected the countries we study according to macroscopic indicators such as their GDP. Yet, influence can also be measured at a different, less political, level, such as cities. The notion of “global cities”, for instance, refers to cities which have a global influ-

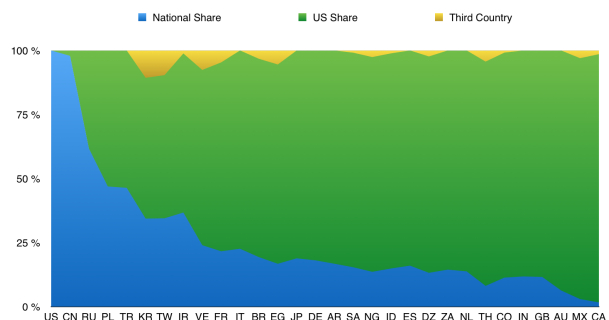


Figure 12: Ratio of sites in the Top 25 of each country headquartered in the US, nationally, or in a third country

ence¹⁰. This influence measures the business activity, the human capital, the information exchange, the cultural experience and the political engagement of the cities. In 2014, the Top 20 global cities were evenly distributed across Asia, America and Europe among 14 countries. These cities are not necessary capital cities. Over the last decade, some features, such as human capital, have become more evenly distributed.

On the contrary, our study on the concentration of intermediation platforms contrasts with the analysis of the evolution of global cities. Indeed, platforms are strongly concentrated. Some influential areas (such as Europe which hosts a quarter of the top global cities) are left aside.

6 Platforms as disruptive operators

The previous sections have focused on the concentration - i.e. the small amount - of both platforms and influential countries. We have shown that there are only a few important platforms and a handful of influential countries. As a result, top platforms have a global influence. In this section, we consider the disruptive impact of platforms.

Platforms alter existing services and create new

¹⁰<http://www.atkearney.fr/documents/>

ones by introducing new intermediaries. For instance, e-commerce platforms such as Amazon are intermediaries between customers and sellers. In the publishing sector, platforms such as Amazon challenge existing actors such as publishers by *disintermediating* them, while becoming intermediaries between service providers - such as editors - and service customers. Thus, intermediation can also be seen as a process. It consists as the ability to play the role of intermediary and, by doing so, create or alter economic sectors.

As platforms become major intermediaries, they weave themselves into the fabric of daily lives and become *basic utilities* while disrupting economic models. For instance, the so-called “sharing economy” disrupts traditional industries and collaboration models [3]. Independent workers have more and more opportunities to work for several platforms such as Amazon Mechanical Turk¹¹, Uber or make money through platforms such as Airbnb. This disruption will lead to adjustment in labor laws as the employer/employee relationship evolves and independent workers claim a new status to protect their rights¹².

Platforms have strong responsibility for cyber security, be it at an individual level or a state level. Indeed, most platforms retain users data. Their terms of use condition the power of users to handle their privacy by retrieving, modifying and erasing their data. Most platforms provide privacy management tools which allow users to alter their visibility to a certain extent and define pieces of information which are visible by default.

From a state perspective, platforms possess sensitive information and incomparable surveillance power. As a result, platforms have become partners in national security protection - they may provide useful pieces of information - but also support to opponents of governments. The interests or even values of countries may conflict with the terms of use of platforms. The relationships between countries and platforms is thus extremely complex, as demonstrated in particular by the Snowden revelations, and it will evolve dramatically in the coming decade.

As a result, beyond their economic impact, platforms are gaining an increasing importance that challenges states on their own territories and on their own prerogatives. They belong to the core of national cyber-security systems for their nowadays global strategic importance. In addition, platforms are changing the world and pushing countries to change their regulations. They intervene more and more on prerogatives of states, such as money, taxes, labor rights, to name a few. Conflicts are thus expectable phenomenon and so is cooperation between countries and platforms. We analyse these movements in the next Section.

7 Countries and platforms: conflicts and collaboration

Cyber-security is strongly dependent upon legal frameworks - which restricts which data can be collected and processed - and surveillance practices. The concentration of the headquarters of platforms extends the power of the legal frameworks of affluent countries - which applies to data belonging to foreign citizens. Platforms have also become first-class partners for surveillance programs. Table 4 recaps the main features of the legal frameworks in terms of privacy and surveillance of the eleven influential countries we have identified¹³.

Two main conclusions arise from this Table. First, platforms are crucial in surveillance strategy. The US, for instance, collaborates with corporations to retrieve data. However, headquartering a platform is not enough for a state to access its data. The US have also recently unsuccessfully tried to get data stored abroad by corporations such as Microsoft¹⁴.

Then, the concentration of influential countries and platforms and the globalization of digital activities leads to conflicts. We leave aside phenomenon such as fiscal optimisations as they constitute an object of study in their own rights (for instance, in the US, most digital corporations have received state fundings but, as they perform fiscal optimisation deprive

¹¹<https://www.mturk.com/>

¹²<http://www.theguardian.com>

¹³Data were collected on <http://uk.practicallaw.com> and <http://en.wikipedia.org>

¹⁴<http://www.wired.com>

Country	Privacy Protection	Cross Border transfer and storage laws	Surveillance policy
US	Mix of federal and state laws. First amendment protects freedom of speech.	Only a few laws to restrict the transfer of data outside the US	Several programs (PRISM, etc.) and laws (Patriot act).
CN	Privacy of people is a constitutional right. Several sectoral laws protect data.	-	Collaboration with corporations (QQ) and several programs (Great firewall)
MX	No law on personal data protection currently published.	-	Telecommunication Law provides government with censorship and surveillance prerogatives.
JP	Regulation of data processing by data controllers. Privacy may not be claimed by an individual.	No restriction, except if transfer to a searchable database of personal data.	-
RU	Mix of federal laws, international conventions and specific acts	Law applies to all actions in Russia and cross border agreements.	System of Operative Investigative Measures (SORM)
BR	Mix of constitutional principles and sectoral laws such as Internet act.	-	Conflicts with the NSA surveillance over the Internet.
AR	Specific personal data protection law.	-	Some contents are monitored and prohibited.
GB, NL	European directive on data protection. For GB: case laws.	Data must be transferred to countries which guaranty an adequate level of data protection.	Personnal data can be accessed to protect national security. Telecom corporations
IR	-	-	Some sites - such as social media - are blocked and the Internet is monitored.
IN	Constitutional and dedicated laws.	Data must be transferred to countries with the same level of data protection.	Central monitoring System (CMS) comparable to PRISM.

Table 4: Data storage and legal context

the US state from taxes). Table 4 presents conflicting definitions and regulations in terms of privacy and data protection. For instance, from a European perspective, privacy definitely is a core individual right. In contrast, Japan regulations does not consider privacy infringement as an individual cause of action and the US emphasizes freedom of speech. As a result, Europe does not consider that the US provides an adequate level of data protection and an agreement - called "Safe harbor" - had to be settled in order to allow transatlantic transfer of data [4].

Conflicting frameworks lead to various situations. The Chinese corporation, Xiaomi, for instance, has recently announced it would store its data in the United States out of privacy concerns¹⁵. However, conflicting definitions of values and data protection standards translate in various conflicts between several European countries and major platforms. Geographic territories have rigid borders within which they enforce data regulation and values. In contrast, large intermediation platforms operate worldwide. Thus, there values, embedded in terms of use may conflict with particular countries. The same stands for business strategies, for instance in terms of data storage location. Figure 13 depicts a set of conflicts between intermediation platforms and countries.

Yet, these conflicts do not prevent platforms from being mandatory partners when it comes to data protection and management. The "right to be forgotten" [10], for instance, has famously opposed Google and Europe in 2014. A European citizen claimed the right to ask for the delisting of pages which threatened their image. The EU court condemned Google to offer such an opportunity. However, as of today, there is no European actor responsible for the right to be forgotten. On the contrary, Google is in charge on its own and responsible for collecting the delisting claims and processing them according to its own criteria.

¹⁵<http://www.reuters.com>

8 Discussion

Motivated by the quick and tremendous development of intermediation, we have studied the features of the revolution that intermediation constitutes in the digital economy. As we have categorized digital activities, we have shown that intermediation overpowers all other types of activities in terms of importance: it attracts most visits on the web. Intermediation also is a concentrated phenomenon: there are only few influent platforms, headquartered in a small amount of countries. Our work emphasises the supremacy of countries such as the United States and China over other areas of the world, such as Europe.

These conclusions allow us to depict a contrasted picture of data security. Influent countries are able to enforce data surveillance and impose their views on data protection. Influent platforms have become both potential partners and opponents in defining of cyber security, be it from an individual point of view (when it comes to privacy for instance) or a state viewpoint.

Our work should be contrasted with [5] and [2]. In [5], the authors show that web trackers are mostly located in countries with permissive frameworks in terms of privacy and cross borders data transfer. The correlation of the location of data centers and data frameworks emphasizes once more their importance in business strategies. Indeed, as privacy becomes a competitive advantage, corporations turn to countries with the appropriate security context. In [2], the authors underline the importance of the US on the web: they attract most data flows. Our analysis reinforce these conclusions, intermediation platforms will set the upcoming political and economic agenda.

Our analysis must be followed up in several directions. First, we left aside several economic domains of studies, such as tax management in digital activities. We also ignored several rising types of use, such as mobile applications. Both raise stringent political and economic issues. Uber, for instance, a booming intermediation platform, mostly works on mobile.

Eventually, the figures we present are striking - they picture the domination of some platforms and countries. Yet, our metrics should be finer and completed with qualitative measures. Indeed, we believe



Figure 13: Conflicts between countries and corporations

that the higher the abstraction level of the platform, the higher the quality of the data harvested. More elaborate measures based on the abstraction level are needed to better quantify and qualify the influence we have diagnosed.

9 Acknowledgments

The authors would like to thank Nadeem Anjur and Damien Reimert for their help in developing the tools necessary to collect and analyse the data. This work has been funded by the European project Big data roadmap and cross-disciplinary community for addressing societal Externalities (BYTE).

References

[1] F. Allen and A. M. Santomero. The theory of financial intermediation. *Journal of Banking &*

Finance, 21(11):1461–1485, 1997.

- [2] C. Castelluccia, S. Grumbach, and L. Olejnik. Data Harvesting 2.0: from the Visible to the Invisible Web. In *The Twelfth Workshop on the Economics of Information Security*, Washington, DC, United States, June 2013. Allan Friedman.
- [3] M. A. Cusumano. How traditional firms must compete in the sharing economy. *Commun. ACM*, 58(1):32–34, Dec. 2014.
- [4] J. Davis. Safe harbor data privacy and security for european union employee data. *Information Security Journal: A Global Perspective*, 18(4):194–198, 2009.
- [5] M. Falahrestegar, H. Haddadi, S. Uhlig, and R. Mortier. Anatomy of the third-party web tracking ecosystem. *CoRR*, abs/1409.1066, 2014.

- [6] S. Grumbach. Intermediation platforms, an economic revolution. *ERCIM News*, 2014(99), 2014.
- [7] S. A. Hale, T. Yasseri, J. Cowls, E. T. Meyer, R. Schroeder, and H. Margetts. Mapping the UK webspace: Fifteen years of british universities on the web. *CoRR*, abs/1405.2856, 2014.
- [8] S. Ihm and V. S. Pai. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 295–312. ACM, 2011.
- [9] O. Lehmberg, R. Meusel, and C. Bizer. Graph structure in the web: Aggregated by pay-level domain. In *Proceedings of the 2014 ACM Conference on Web Science*, pages 119–128, New York, NY, USA, 2014. ACM.
- [10] A. Mantelero. The eu proposal for a general data protection regulation and the roots of the ‘right to be forgotten’. *Computer Law & Security Review*, 29(3):229 – 235, 2013.
- [11] A. Scharlr, K. W. Wöber, and C. Bauer. An integrated approach to measure web site effectiveness in the european hotel industry. *Information Technology & Tourism*, 6(4):257–271, 2003.
- [12] A. Themelis. Information and intermediation, abuse of dominance and internet ‘neutrality’: ‘updating’ competition policy under the digital single market and the google investigations(?). *European Journal of Law and Technology*, 4(3), 2013.