ieee.es
Instituto Español de Estudios Estratégicos

Analysis
Paper

19/2022 16 marzo 2022

David Ramírez Morán

Borders for the digital raw materials

Borders for the digital raw materials

Abstract:

Data is one of the main raw materials for the 4.0 industrial revolution. Its extraction is a complex activity that is being done through different procedures, from direct capture to the instrumentation of other services to obtain additional data. It should not be forgotten that techniques exist to obtain data illegally and that this lack of legitimacy arises from different questions such as the violation of personal privacy, the lack of compliance of the protection and processing of date, the violation of the information protection measures or spying operations driven even by States.

Citizens' data, both of those belonging to a State or of foreign citizens, are required by the new business models that the 4.0 industrial revolution is generating. Having the data may result a requirement to achieve the productivity that ensures the prosperity of a State. For this reason, the management of data reaches a strategic nature that may affect national interests, beyond matters as important as the rights of citizens for privacy.

The borders to the use of data arise in a natural way as a result of the different existing regulations as well as the existing differences between the principles that drive the societies.

Keywords:

Privacy, data protection, economic interest, ethics.

Fronteras de la materia prima digital

Resumen:

Una de las principales materias primas de la revolución 4.0 son los datos. Su obtención es una actividad compleja que se está realizando por diversos procedimientos desde la captura directa a la instrumentalización de otros servicios para conseguir datos adicionales. Tampoco puede olvidarse que existen técnicas que permiten obtener datos de forma ilegítima y que esta falta de legitimidad puede ser el resultado de distintas cuestiones como son la violación de la privacidad personal, el incumplimiento de la legislación sobre protección y tratamiento de datos, la violación de las medidas de protección de la información o de operaciones de espionaje dirigidas incluso por Estados. Los datos de los ciudadanos, tanto propios de un Estado como los de ciudadanos ajenos, son requeridos por los nuevos modelos de negocio que la revolución industrial 4.0 está generando. Disponer de los datos puede resultar imprescindible para conseguir la productividad con la que asegurar la prosperidad de los Estados. Por este motivo, la gestión de los datos alcanza una naturaleza estratégica que puede afectar a los intereses nacionales, más allá de cuestiones tan importantes como los derechos de los ciudadanos a la privacidad.

Las fronteras al uso de los datos surgen de forma natural debido a las distintas regulaciones existentes así como a las diferencias existentes entre los principios que rigen las sociedades.

Palabras clave:

Privacidad, protección de datos, intereses económicos, ética.

How to cite this document:

RAMÍREZ MORÁN, David. *Frontiers* of digital raw materials. IEEE Analysis Paper 19/2021. https://www.ieee.es/Galerias/fichero/docs_analisis/2022/DIEEEA19_2022_DAVRAM_Fronteras_ENG.pdf and/or bie link[3] (consulted day/month/year)

Introduction

The raw material when developing digital products is data. Without this, there is no chance of achieving the benefits that digital transformation can bring to society. Data is continuously being generated and refined to be useful in optimising production processes and developing new relationship and business models.

The initial step for data to be used further down the chain is its capture and digitalisation. From then on, if the process has been carried out correctly in terms of accuracy and quality, it will become an element of information which, although may be simple or insignificant in itself, is essential for the complex information processing and use systems to perform their function and be useful.

Obtaining data requires, on one hand, a sensor, an element that generates that data in a format that can later be processed, and, on the other hand, a system responsible of collecting that data to make it available to those in charge of processing it. In the case of industrial or scientific data, it is the professionals who are responsible for equipping themselves with these two elements to carry out their work. However, with human activity data, where most interest currently lies for public and private organisations, the solution is not so simple. By default, data does not just appear neatly sorted and validated in databases that are easily accessible for processing.

The combination of the internet, digital services and mobile phones has provided an answer to this problem by providing a sensor, which is the phone, a way to collect the data via the Internet, and systems that receive and store the data when using the digital services available via the Internet. This approach provides an infrastructure, the material means that make it possible to collect data for further processing, but another important part is still missing.

For infrastructure to be useful, the data sources need to be used by people, and this is where the widespread use of mobile phones and the Internet and the emergence of more and more digital services that facilitate people's daily lives becomes important. Every time one of these services is used, data is generated, some more relevant and some less relevant depending on what the service provider collecting the data is interested in.

To provide a service, some data is essential for the correct technical operation of the service, such as the desired product, the information required to produce it – such as the

model of the product requested, the current location or the origin and destination of a journey – the authorisation elements necessary to be able to control access to the services, the technical data essential to carry out the communication, such as the IP address of the device and, optionally, certain information on the device used in order to provide the best possible user experience.

Of all this essential data, the service provider would only need to store some way of identifying the user and the service purchased for billing purposes for a certain period of time. For security reasons, legislation imposes additional requirements on the information to be stored for each of these transactions, requiring the metadata of the communication to be stored for possible criminal investigation purposes. It is the communications operators who bear this responsibility.

So far, this describes a viable business model that, except for digitalisation, has been in place since trade in goods and services appeared in the world: a customer required the service, paid for it and the relationship between the two parties ended.

From the end of the 18th century to the present day, there has been a revolution in this model. Competition means that customers have a choice of which supplier to use and they need to be won over. At the same time, the improvement in the quality of life leads to a demand for products that go far beyond mere subsistence, whose market can be expanded through campaigns that generate desire. Both factors make advertising appear as a new tool capable of generating value and giving rise to a new business model. During the 20th century, advertising has been a key source of media funding based on a one-way relationship. No feedback of effects was available beyond looking at the change in turnover generated. In addition, the ability to personalise it to reach the target audience was very limited.

With the advent of personal communications in the first place and then with the explosion of the Internet it is possible to address a person specifically and a total transformation takes place. Communication can be done on a one-to-one basis, which requires greater use of resources for person-to-person contact as opposed to simply placing an advertisement in a newspaper, or on a radio or TV station. For this reason, it is also necessary to have some kind of filter to optimise the available resources in order to dedicate them to those profiles that are more likely to become customers. The need to count on people's data arises and new businesses appear that provide what a service

provider needs: groups segmented by age, characteristics, interests, etc. to whom campaigns can be directed.

To segment the population you need their data and for some businesses it is relatively easy to have it as soon as there has been a previous business relationship. However, for other businesses or to get new customers, there was no solution until the emergence of these businesses. But how do they access this information? The advent of Web 2.0 was another revolution. Users began to generate content and both the requirement to identify users and the generalisation of techniques to identify individual users became widespread.

An evolution is now taking place on the Internet where data flows are turning around and it is the provider that receives more information than the user. It is this data that is fuelling the new machine learning and artificial intelligence technologies that will rule the future of humanity as advocated by leading experts in many sectors.

Profiling and associating information

One of the measures to increase the quality of data for further processing is classification. All information can be subject to a myriad of different classifications and one of them is being able to attribute the information to a particular person. From information correctly associated with a person, it is possible to generate a profile that defines that person's interests, practices and habits. However, the risks involved in this practice quickly become apparent when one considers the potential information that could be gleaned from the data collected from that user's internet usage. Issues such as health, political ideology, employment data, etc. could be extracted from data such as web pages visited. To this end, various techniques have been developed to identify the profile to which information can be associated and to trace interactions on different websites by the same user.

Cookies have for years been one of the main tools for tracking the actions of individual users across the internet. They leave a small packet of information stored on the user's computer that will later be read when accessing other websites.

To individualise the information collected from a device, there are numerous techniques that make it possible to uniquely identify the device from which a user accesses a web page even if the user has not provided information such as a user name. These are called

fingerprinting techniques[1] and involve implanting fragments of code in web pages that collect information on the type of graphics card, the fonts installed, browser extensions, the language configured, etc. One of the actors that is giving more visibility to the possibilities of identifying and tracking users is the Electronic Frontier Foundation[2], which allows tests to be carried out on the correct configuration of devices to avoid these techniques of collecting information and profiling users.

Online services require prior identification even for free versions. The provider wants to know who uses the service and when they use it. This allows the profiling of users, which means that, according to legislation, the information exchanged in the provision of this service is also in the nature of personal information. Free versions, especially mobile phone applications, are also a tool to deliver advertising to users.

More and more companies and organisations are developing their own applications for their users and customers to interact with them as opposed to the previous solution which consisted mainly of an Internet presence. These applications increasingly provide more functionalities and, at the same time, require more permissions to access the information of the terminals on which they are executed. Information such as location, obtained both from specific sensors and through indirect means of visible Wi-Fi networks, can accurately geo-locate the user's location.

Today, especially in more developed countries, a telephone number is almost equivalent to a citizen's personal identification except for devices used for business purposes only. The use or transfer of this data, in the case of a personal and non-professional device, already results in all data exchanged by the application being considered personal data.

In light of the risks of impersonation, greater security measures are being implemented, such as the second authentication factor, which, while allowing services to be provided with greater security, is also a means of information leakage. These tools require personal information, such as the user's phone number or email address, in order to receive a code to reconfirm that it is the user who is legitimately trying to access the application. In other cases, what is requested is the installation of a specific application on the device that will generate keys with which to gain access. This application may require access to the

[1]https://fingerprintjs.com/blog/browser-fingerprinting-techniques/
[2]https://www.eff.org/

internet, so it would be sending information about the user's IP address, and in many cases it is developed with permissions that may go far beyond the generation of an additional password for access. Functionalities such as geolocation can be included, which contribute to security by preventing access by the same user from very distant locations, but which represent a possible way for service providers to obtain data by simply recording these locations.

Once again, security is the argument put forward for its implementation, appealing to the age control of users so that they do not access inappropriate content, for example on video viewing pages. In return, such a platform will have a history of all the videos that user has watched.

Emails are no longer so innocent and all links, usually sent by services that facilitate mass distribution, also constitute a mechanism for identifying who has received the message, whether they have clicked on the links, etc. This is done by sending a different link for each user to whom the information has been sent, while the server can determine which user has accessed the information by simply reading the information in the URL address requested when the user clicks on the link. Similarly, e-mails are no longer just plain text but full html pages. This makes it possible to include fonts and images whose download from the server can be both logged and individualised to determine who has read an email, and when, if no action is taken to prevent it.

A modern website has multiple functionalities that can be implemented by using online services directly integrated into the website and provided by other providers. Elements such as search boxes using traditional search engines but limited to that website, tools to compile visitor statistics, use of individualised sources, content distribution services that speed up navigation, etc. make it possible to create the website using already developed solutions. On the other hand, it is necessary to upload code directly from the servers of these providers in most cases. This means providers will obtain data on the website user such as the website they are accessing, statistics on the time spent on the website, even visualisations of the user's interaction with the website, etc.

In general, every interaction currently carried out on the internet is being recorded by one or more providers who can make use of the information. Of course, the service provider we are using receives the information. However, it may not be so obvious that all other

providers from which information is downloaded for the operation of the site will also receive information about this access.

It should not be forgotten that an app installed on a mobile device is equivalent to a web application that can incorporate additional functionalities not allowed on web pages because it is installed on the device and where the user has given confidence by installing it. Therefore, the app provider will have access to much more user information if the user has allowed it, such as access to the device's content, more control over functionalities such as calls or sending and receiving messages, etc.

The next step

When a third party wants to gain unauthorised access to data protected by identification mechanisms, it is necessary to apply cracking techniques[3]. It is about using techniques that circumvent protection measures to gain access to protected information.

These tools are characterised by providing a mechanism that allows you to take full control over the devices on which you manage to install them. To do so, it exploits vulnerabilities, malfunctions in the software of the target devices and manages to install programmes that capture information from the device and send it to the storage and control infrastructures that support them, deployed on the Internet. It is not only possible to access information already on the device such as photos, e-mails or call lists, but also to install, run, delete or modify applications.

These tools can be used to obtain passwords, access the microphone or camera and forward recorded audio and video, obtain multi-factor authentication data to access or perform operations on online services, etc. In short, they can carry out all the actions that the owner of the phone can perform, and even more, without the owner's knowledge.

These are highly complex tools that face increasingly secure devices. The vulnerability that allows the tool to be installed, the fact that doing so requires minimal user intervention – sometimes just accessing a web page in a browser –, the infrastructure that supports the tool on the internet, the stealth with which it sends the collected data and the support

[3] Cracking actions are those that allow unauthorised activities to be carried out on a computer system by a third party through exploiting detected vulnerabilities or using credentials obtained by illicit means. It should be noted that this term differs from hacking, which consists of taking advantage of the characteristics of computer systems in order to carry out or facilitate the carrying out of a task by computer.

necessary for the tool to continue to perform its function even after the bugs that allowed it to be installed have been detected and fixed are the value achieved with the development of these tools. Two examples of tools used for these purposes are XKeyscore, which was mentioned in the Edward Snowden leaks and was used to access the mobile phone of former German Chancellor Angela Merkel[4] and up to 125 other high-ranking members of her government, and the Pegasus tool, developed by the Israeli company NSO.

The powerful capabilities provided by Pegasus, for example, pose a risk to anyone's privacy, which could call into question the development of such technologies. However, in the case of NSO, the company's website[5] states that "NSO creates technology that helps government agencies prevent and investigate terrorism and crime to save thousands of lives around the globe". According to this statement, it only provides its technology to government agencies and therefore assumes that the use of its technology will be subject to the rule of law.

However, Pegasus has been frequently cited in the media in the face of scandals involving its use against individuals whose profile does not correspond to these objectives, given that traces of the tool have appeared on the devices of activists, heads of government, diplomats, journalists, etc. [6]

These facts led the European Data Protection Supervisor to issue a report[7] in 2022 recommending a ban on the development and deployment of such tools in the European Union in order to protect the fundamental rights and freedoms of citizens. This report includes the exception of using them to prevent very serious imminent threats, but highlights a number of issues that need to be carefully considered even in these exceptional situations:

- Strengthening of democratic oversight of surveillance measures.

---

[4] "NSA tapped German Chancellery for decades, WikiLeaks claims". Available at https://www.theguardian.com/us-news/2015/jul/08/nsa-tapped-german-chancellery-decades-wikileaks-claims-merkel accessed 1/03/2022

[5] https://www.nsogroup.com/

[6] *Dana Priest*, *Craig Timberg* and *Souad Mekhennet*. "Private Israeli spyware used to hack cellphones of journalists, activists worldwide " https://www.washingtonpost.com/investigations/interactive/2021/nso-spyware-pegasus-cellphones/ Accessed 5/3/2022

[7] "Preliminary remarks on Modern Spyware". European Data Protection Supervisor. https://edps.europa.eu/system/files/2022-02/22-02-15_edps_preliminary_remarks_on_modern_spyware_en_0.pdf

- The strict implementation of the EU legal framework on data protection.
- Judicial review, both ex-ante and ex-post, should be real; it cannot be a mere formality.
- Strengthening of the protections offered by the criminal procedure.
- Reduce the risk that data originating from such undemocratic and abusive surveillance practices reaches the databases of the Union, such as those of Europol or Member States' law enforcement agencies.
- Stop (ab)using national security purposes for legitimising politically motivated surveillance.
- Addressing the rule of law problems.
- Empowering civil society to bring awareness and public debate forward.

Both XKeyscore and Pegasus are tools theoretically aimed at contributing to national security, although it should not be forgotten that there are numerous much less transparent actors in the cyber world that also have the capabilities to develop and deploy such tools. It is only a question of there being enough interest in the information an individual handles that someone, with the appropriate technical and financial backing, decides to use such a tool to access it.

Risks to national interests

In light of the risks posed by the availability and processing of personal data for rights considered fundamental in the EU, such as privacy and freedom of expression, data protection regulations have made it possible to establish limitations on the use that could be made of the data, as well as the rights that protect users in the event that their information has been recorded or processed.

Orbiting over this day-to-day life in which the main objective is to obtain data to feed computer systems, an architecture is also being built with which it is possible to learn more through the information that users vehemently provide. The privacy of the individual's information sometimes does not even matter to the individual, who uses it as an exchange for the free – in terms of economic cost – receipt of services. However, this information, suitably aggregated and processed, can provide data of great value to third parties with interests that may not be compatible with those of the population who have

so generously provided their data, interests, tastes, habits, opinions, state of health, state of mind and even economic capacity.

Data protection regulations focus on the protection of personal information, so only in cases where it is possible to associate data with an individual a breach of these rules occurs.

The General Data Protection Regulation[8] is the current EU standard in this respect and is a major stumbling block for activities that have traditionally been carried out for the normal operation of services provided from outside the EU.

In October 2015, the International Safe Harbor Privacy Principles, a framework under which the transfer of personal information from Europe to the United States could take place, was declared invalid. Known as the Schrems I ruling, work soon began on a second set of rules under the name EU-US Privacy Shield. The European Court of Justice also declared it invalid in October 2020 with a ruling named Schrems II after the previous one. The ECJ ruled that the transatlantic Privacy Shield agreement, which set the rules for the exchange of information between the EU and the US, was invalid because it could not protect European users' data from US legislation such as the CLOUD Act[9], which requires US companies to provide information located on their servers for national security reasons even if they are located beyond their borders. However, the court allowed some technology companies such as AWS and Google to use Standard Contractual Clauses (SCCs) as a legal mechanism for data transfers, with some adjustments.

A new framework for the exchange of personal information between the EU and the US is underway. It is referred to in some forums as Schrems III in reference to the two previous mechanisms along these lines.

Data protection regulations can also constitute an economic risk for companies that make the use and processing of data their business model. One of the world's largest market capitalisation companies, Meta, recently reported being at risk from European data protection regulation:

[8] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

[9] www.justice.gov/CLOUDAct

> "reliance on Standard Contractual Clauses in respect of European user data does not achieve compliance with the General Data Protection Regulation (GDPR) and preliminarily proposed that such transfers of user data from the European Union to the United States should therefore be suspended...".[10]

This was interpreted by some as the company threatening to shut down the service in the EU if problems with the transfer and processing of user data to the US were not resolved.

In fact, in US strategic circles, the restraining character of this regulation is highlighted:

> "*Regulations such as the European Union's GDPR have already shown to have adverse effects on free speech, consumer choice, and even scientific research. Further, experts have noted that the GDPR does little to protect privacy and instead focuses almost exclusively on its stated goal of data protection.*"[11]

Recently, the Commission Nationale de l'Informatique et des Libertés (CNIL)[12] requested a company to remove any reference from its website to the service it had installed to carry out statistical analysis on access to its website. A user complained that their data had been transferred outside the EU, specifically to the US, and that the GDPR had therefore been breached.

Even anonymised data can be a lateral route for third parties to obtain data. Given previous episodes in which military operations were threatened by the wealth of information that mobile phones distribute, sometimes even unconsciously from their users, TTPs have been adapted to avoid these risks. However, in the recent advance of Russian troops, there was again a remarkable issue as students from a US high school detected the irregularity that was occurring in traffic on the roads near the Ukrainian border at ungodly hours. At 3:15 a.m. the alert for traffic jams on these roads appeared on the navigator of their devices. It was not the mobile devices of the soldiers themselves that had led to this situation but the devices of the population who had been slowed down by the barriers that closed the roads for the passage of the convoy of vehicles approaching the border. Both the mobile phones of drivers and passengers and the

[10] KUNDALIYA dev. *Meta threatens (again) to shutter Facebook and Instagram in Europe* https://d18rn0p25nwr6d.cloudfront.net/CIK-0001326801/14039b47-2e2f-4054-9dc5-71bcc7cf01ce.pdf
[11] Alexander Kersten and Isaac A. Robinson . Data Protection or Data Utility? https://www.csis.org/analysis/data-protection-or-data-utility
[12] https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply

satnavs in the cars were sending information to servers far beyond the country's borders that automated systems were interpreting as a traffic jam. [13]

It is not the first time that the Google Maps service has taken on strategic overtones because, as a mapping information provider, it represents maps and these are blank slates that can convey many messages. In fact, the borders shown on these maps sometimes transcend mere geography and are the result, both in their layout and the type of line used to draw them, of the reflection of the company's position with respect to certain territorial conflicts.

In the Crimea conflict in 2014, Google Maps provided a different map depending on the website from where the map was loaded. When accessing the .ru domain, a border between the peninsula and the mainland appeared, this border disappeared when accessing the Ukrainian domain, .ua, while when accessing the international site, .com, a dashed border was displayed indicating the dispute over it.[14]

Conclusions

From day-to-day decisions to strategic decisions, they are undergoing a process in which the incorporation of external data is increasingly necessary. This is because this data provides benefits, because it is useful, which justifies the implementation and the costs necessary to collect them. Companies are particularly aware of this and it is now very difficult to interact anonymously on the internet.

Personal data is private property and, like tangible property, also becomes subject to protection by legislation. There are many different approaches that can be adopted to establish this protection, according to criteria that differ considerably depending on the principles governing the work of legislators. Different actors have created regulations in this regard that must accommodate security, privacy, defence of economic interests and other principles in line with their respective legislations. This creates contrasts that result in limitations to the free flow of data both in terms of exporting and retaining data.

---

[13] Rachel LermanOn Google Maps, tracking the invasion of Ukraine https://www.washingtonpost.com/technology/2022/02/25/google-maps-ukraine-invasion/

[14] Google Maps Russia claims Crimea for the federation. https://www.theguardian.com/technology/2014/apr/22/google-maps-russia-crimea-federation

States must make a decision on an ethereal commodity whose use is bubbling. It can be exported for other actors to squeeze out the growing value it is developing or take steps to ensure that this raw material, data, contributes to local growth with the productivity gains that its exploitation in Industry 4.0 promises and the large number of in many cases high-skilled jobs it leads to.

In parallel with economic interests, the issue of privacy and citizens' freedoms must also be considered. In the case of the European Union, where the GDPR pursues these objectives with restrictive measures that seek to respect these rights considered fundamental in its founding Treaty, there are many challenges that are arising in the current model in which actors that do not respond to these principles in their entirety are the main service providers, as well as those most interested in accessing citizens' information.

The dilemma between data protection and business is gaining in intensity, and in this battle, factors directly related to security are being brought to the table. Thus, laws that facilitate access to information distributed outside national jurisdictions, belonging to citizens also outside the borders of the originator of the legislation, are justified by measures such as the fight against terrorism, paedophilia and crime. Arguments such as the limitations to technological development and growth that the protective measures of other states place on general interests are also used, underestimating the importance of these legislations whose objective is the defence of the fundamental rights of citizens.

At the same time, it is ironic when one of the sectors critical of protective measures also takes a position of veiled criticism of legislative action in other states where citizens' privacy has been greatly compromised by the rapid technological growth made possible by rapid access to their citizens' personal information, and which, in turn, imposes serious limitations on the incursion into this market of foreign entities that seek to access the raw material or send it beyond their borders.

The specific problems and examples of situations where the use of data can affect national security should serve as a benchmark in determining the importance with which the proper management of information generated by a state's population should be approached. These are examples that are already occurring today and will become more numerous with the increasing emergence of services that may be affected by the nature and quality of data.

Physical borders pose barriers to the use of citizens' data in the face of incompatibilities in the regulations of different States, although it should not be forgotten that, in addition to the physical borders that arise mainly from legislation, there is another set of borders defined in this case by the citizens, dictated by the ethical and moral principles to which the processing of information must be subject.

*David Ramírez Morán**
Analyst of IEEE