E.L. Rev. 2020, 45(6), 857-869

European Law Review
2020
Scraping personal data from internet pages - a comparative analysis of the Polish Bisnode decision and the US hiQ Labs v LinkedIn Corp judgment

Zuzanna Gulczynska [1]

Subject: Information technology
Other Related Subject: Human rights. Legal systems.

Keywords: Comparative law; Data collection; Data protection; Data subjects; Personal data; Poland; Privacy; United States;

Cases:

Bisnode, Re unreported 15 March 2019 (Poland)

hiQ Labs Inc v LinkedIn Corp unreported 14 August 2017 (D (US))

*857 Abstract

This article comments on the recent case law developments in the European Union (EU) and the US relating to the scraping of personal data. The Polish Data Protection Authority Decision in Bisnode *and the US* hiQ v LinkedIn *judgment show that the practice of scraping not only raises different legal problems in the two jurisdictions but also brings about different (and possibly conflicting) approaches to the regulation of scraping. The discrepancy lies in the difference between the American "right to privacy" and the European "protection of personal data". Moreover, the EU and the US give different weight to commercial interests of data scrapers, as compared to the rights of individuals. Finally, from a procedural point of view, the two systems grant different positions to individual data subjects as opposed to the data controllers and processors. The different approaches to scraping reflect a broader disagreement on the approach to data protection, which slowly reaches its limits and creates a division in the otherwise boundaryless digital world.*

Introduction

Properly analysed data—so called smart data[1]—has become a valuable product. Data analytics allows to discover information, draw more accurate conclusions and support better decision-making. However, the success of data analytics depends on the amount and quality of data fed to the algorithms. Data acquisition is therefore the crucial starting point of any data analytics endeavour and one of the methods deployed to that end is data scraping.

Data scraping, or web scraping, can be defined as "the act of extracting large amounts of information from a website using automated software programs called bots".[2] It allows the collection of public data, ***858** i.e. data "posted on the internet without password protection reserving it for specific users",[3] from across the Internet. Recent estimations indicate that bots' activity amounts to nearly 40 per cent of the entire internet traffic,[4] which renders scraping one of the most important methods

of collecting data. It is used by a wide array of actors, from Internet giants such as Google, who use bots to catalogue public websites, to analytics start-ups, such as hiQ, who offer data analytics services to various industries.[5]

Data scraping poses several legal challenges, relating mainly to copyright and contract law[6] but also to other, more specific areas of law, such as the computer fraud and abuse or state trespass laws in the US[7] and data protection legislation in the European Union (EU).[8] Even though technical possibilities in the digital world are not limited by borders, the reality of data scraping is conditioned by legal regulations. In the EU and the US, the rules governing data scraping are far from being homogeneous. The bone of contention is the protection of personal data and the position of individual data subjects in the process. Whereas in the US the processing of personal data is regulated via a patchwork of (largely sectoral) legislation, rendering the protection of data subjects very fragmented (and in many areas non-existent), the EU has adopted a strong horizontal regime centred around an individual's right to have his/her personal data protected. This divergence has important consequences for the practice of scraping. Indeed, the EU and US set different limits and conditions to processing publicly accessible personal data. This, in turn, can hamper the free flow of data between the two regions which asks for a more aligned approach towards data scraping.

The differences in the legal approaches to data scraping between the EU and the US have been made apparent in two recent decisions: the Polish Data Protection Authority's decision in the Bisnode case[9] (partially overruled on appeal)[10] and the preliminary injunction issued by the District Court for the District of Northern California in hiQ v LinkedIn[11] (confirmed on appeal).[12] These cases show that the rules that are (potentially) applicable to data scraping in the EU and the US accentuate different legal problems, and in some circumstances can even lead to a deadlock dividing the digital world. However, the debate on scraping is not over just yet, since appeals against both decisions are pending before the supreme jurisdictions (the Polish Supreme Administrative Court and the US Supreme Court) at the time of writing. Given the obligation of last instance courts in the EU to request a preliminary ruling from the Court of Justice of the European Union (CJEU) on matters of EU law,[13] the Polish case is likely to set a precedent for the entire EU. It is therefore an excellent moment to evaluate the approach adopted so far in the legal battle over scraping in the EU and the US. ***859**

Factual background

The facts of the two cases are similar. In the Polish case, the company Bisnode obtained personal data of persons conducting business activity in Poland, including their names, names of their enterprises and types of business activities, addresses, email addresses and telephone numbers. These data were scraped from publicly available sources, in particular from the public registers accessible online including the Register of Entrepreneurs of the National Court Register, Central Registry and Information on Economic Activity and Database of the Polish Central Statistical Office. Bisnode, whose main business activity consists of providing data analytics services, collected the said data to create commercial reports and issue addresses or telephone lists for its clients.

In the US, the start-up hiQ collected in a similar manner personal data from publicly accessible LinkedIn accounts. These data would be used by hiQ to provide its clients with insights about their employees, such as which employees are flight-risk or what skills they possess.

Even though both cases concerned the scraping of large amounts of personal data from publicly available sources for commercial purposes, the legal issues in each of the two legal regimes—the Polish legal system (governed in this case by EU-wide rules) and the US one—appear to be fundamentally different.

The European General Data Protection Regulation (GDPR) imposes the obligation on data controllers to inform the persons whose data they process, when the data is not collected directly from them.[14] In order to comply with this obligation, Bisnode sent information emails to roughly 700,000 data subjects. However, it was not in possession of email addresses of the remaining

6.5 million persons, whose data they had processed at the time. For these data subjects, Bisnode decided to rely on the exemption to the obligation to inform.[15] Bisnode justified its choice by the disproportionate effort (both financial and organisational) that the provision of such information would involve, should it be carried out via regular mail. Instead, the company published on its website a data protection policy containing the information required by art.14(5)(b) GDPR. This came to the attention of the Polish Data Protection Authority who decided to investigate Bisnode's compliance with the obligation to inform.

In the US, hiQ also faced obstacles to its data scraping activities—it was served with a cease-and-desist letter from LinkedIn requesting it to stop accessing the LinkedIn website and using the information available thereon. LinkedIn claimed, among others, a violation of its copyright but also of the LinkedIn Users Agreement and the privacy of LinkedIn users. It is in this context that hiQ started legal proceedings against LinkedIn oriented towards declaring its scraping activities as compatible with the legislation in place. Ancillary to its claim for a declaratory judgment, hiQ filed a motion for preliminary injunction to preserve the status quo of its scraping practice until the final judgment is issued.

Decisions

Bisnode case

According to the Polish and EU legislation, the President of the Personal Data Protection Office (further referred to as the Polish data protection authority, Polish DPA or DPA) is the authority competent in matters of personal data protection and the supervisory authority within the meaning of the GDPR.[16] By virtue of the GDPR, the EU data protection authorities are vested with corrective powers allowing to order ***860** data controllers and processors to bring their processing operations into compliance with the GDPR,[17] as well as to impose administrative fines for data protection breaches.[18] The Polish DPA has put these powers to use in the case of Bisnode.

In its Decision of 15 March 2019,[19] the DPA found that Bisnode was in breach of the information obligation stemming from art.14(1)–(3) GDPR. According to this provision, where personal data have not been obtained directly from the data subject, the controller is obligated to provide the data subject with certain information, in particular relating to the categories of data concerned, legal basis and purposes of the processing, rights relating to personal data and the controller's contact details.

The Polish DPA considered that the publication on a website of a data protection policy relating to personal data acquired from publicly available registers, in the context where Bisnode was in possession of data subjects' phone numbers and/or addresses, does not amount to the fulfilment of the controller's information obligation. Instead, in such a context, such information should be provided to each data subject individually.[20]

The decision also assessed the applicability of the exemption provided for by art.14(5)(b) GDPR that had been invoked by Bisnode. This provision releases data controllers and processors from the information obligation, where providing such information to the data subject proves to be impossible or would involve a disproportionate effort. In this regard, a distinction has been made between the persons whose contact details were not included in the public registers (and were therefore unknown to Bisnode), on the one hand, and those whose contact details were available in the public registers scraped by Bisnode, on the other. As to the first category, the Polish DPA considered that there was little factual possibility and no legal basis for acquiring contact details of data subjects concerned. The provision of information was therefore indeed impossible or would involve a disproportionate effort. The exemption of art.14(5)(b) GDPR was applicable. However, with regard to those data subjects, the contact details of whom were in the possession of Bisnode, a regular provision of information would not require such disproportionate effort.[21] In particular, the DPA rejected the company's argument that sending registered letters to each data subject would constitute a cost exceeding the company's yearly revenues. The DPA drew attention to the fact that data

controllers are free to choose methods by which they inform individuals of the processing of their data, as long as they are in the position to prove it.[22]

The Polish DPA ordered Bisnode to comply with the information obligation within three months from the decision's date and imposed a fine of 943,470 PLN (approximately €210,000). One of the aggravating factors in determining the amount of the fine was the fact that, by claiming that its financial interests constituted an exonerating circumstance, Bisnode lowered the value of the rights of data subjects and considered them inferior to the financial interest of the company.[23]

Bisnode challenged the decision before the Voivodeship Administrative Court, which in its judgment of 11 December 2019,[24] largely confirmed the DPA's assessment. It recognised, however, that even though the public register of entrepreneurs contains contact details, not all of these details are likely to be up to date. Therefore, the Court upheld only in part the DPA's decision: in the scope relating to natural persons who are currently conducting business activity and natural persons who have suspended such business activity. Indeed, the Court considered that the contact details of these individuals are most likely accurate ***861** and should be used to provide the information required by art.14 GDPR. Differently from the DPA however, the Court agreed with Bisnode that it is probable that the contact details of the persons who no longer conduct business activity available in the register are not up to date. Moreover, there is no legal basis for Bisnode to acquire their current contact details from other sources. It would therefore be impossible to fulfill the information obligation with regard to the persons who no longer conduct business activity. Hence, the Court overruled this part of the decision and referred it back to the DPA for a reassessment. Since the number of the data subjects affected by the breach was therefore lower (by approximately 3 million people), the DPA will have to adjust the amount of the fine accordingly.

The DPA's decision was thus partially overturned by the Voivodeship Administrative Court because the DPA failed to check whether Bisnode was actually capable of receiving the up-to-date contact details of those data subjects who no longer conduct business activity. Despite this partial overruling, the core rationale of the judgment follows the assessment of the DPA. It is therefore safe to say that the right to information stemming from art.14 GDPR has been given a broad interpretation, putting the interests of data subjects before any possible commercial interests of data controllers. Since Bisnode has appealed the judgment of the Voivodeship Administrative Court, the final say on the information obligation in the context of scraping is left to the Supreme Administrative Court in Poland (and the CJEU, if the former refers a preliminary question).[25] However, considering the fundamental rights and data protection-related case law of the CJEU,[26] it seems unlikely that the assessment will differ much from the one presented by the DPA and the Court at first instance.

hiQ v LinkedIn

In the US, the factually similar case raised completely different legal questions. The pending hiQ v LinkedIn litigation has also put the legality of data scraping under US law under the magnifying glass. Differently from the European case, however, it is not the privacy of LinkedIn users as such that is at the centre of legal dispute, but rather the question of the access to and the use of publicly available information (in this case personal data) for commercial purposes. The main legal question of the dispute concerns intellectual property law, and more specifically the applicability and possible violation of the Digital Millennium Copyright Act.[27] It also touches upon the Computer Fraud and Abuse Act[28] and California's Penal Code[29] (which both criminalise unauthorised access to computers), as well as the common law tort of trespass. The applicant also raises competition law questions with regard to LinkedIn's behaviour claiming that it excludes hiQ from business.[30]

In its application, hiQ first sought a declaratory judgment establishing that it is not in violation of any legislation in place. Secondly, hiQ filed a motion for a preliminary injunction in order to ensure that it ***862** can continue scraping data from LinkedIn up until the final judgment is issued.[31] Under US federal procedural law, and differently from the main claim, a plaintiff seeking a preliminary injunction needs to "merely" establish that,

"he is likely to succeed on the merits, that he is likely to suffer irreparable harm in the absence of preliminary relief, that the balance of equities tips in his favor, and that an injunction is in the public interest."[32]

Amongst these four conditions, the two last elements, i.e. the balance of hardships and the public interest requirement, are particularly interesting from the privacy law perspective. Therefore, even though hiQ v LinkedIn has not been decided on the merits yet, it is interesting to analyse the preliminary injunction granted to hiQ by the US District Court for the District of Northern California (District Court),[33] affirmed by the 9th Circuit Court of Appeals (Court of Appeals)[34] and currently under review of the US Supreme Court.

The District Court in its assessment of the hiQ motion for a preliminary injunction followed the above-mentioned four-conditions test. After establishing that hiQ raised a serious question on the merits, and thus fulfilled the first condition for issuing a preliminary injunction, the court moved to the assessment of the existence of an irreparable damage and the balance of equities. hiQ claimed that without a preliminary injunction allowing it to keep scraping data from LinkedIn, it would go out of business.[35] LinkedIn, on the other hand, deployed users' privacy rights to prove damage. More specifically, LinkedIn asserted that hiQ's continuous data collection not only constitutes a "tacit invitation" to gain unauthorised access to LinkedIn's computers,[36] but also threatens LinkedIn users' privacy and therefore puts at risk the trust bestowed by them in LinkedIn.[37]

In its assessment of harms, the District Court did not follow the defendant's argument. Instead, it considered that LinkedIn users had little *actual* privacy expectations.[38] Moreover, LinkedIn's concerns about users' privacy seemed to be at odds with its practice of allowing other third parties to access user data (in particular via the paid service "Recruiter" which allows head-hunters to receive full profiles of any LinkedIn member, without consent nor knowledge of the latter).[39] Additionally, since hiQ claimed that it would go out of business, the District Court concluded that the balance of hardships tips sharply in hiQ's favour.[40] The Court of Appeals not only fully agreed with the assessment of the District Court, but also reminded that "LinkedIn has no protected property interest in the data contributed by its users, as the users retain ownership over their profiles." ***863** [41]

Even when a plaintiff raises serious questions on the merits and proves that he is likely to suffer an irreparable harm, with the balance of hardships tipping in his favour, due consideration still needs to be paid to the public interest before any preliminary injunction is issued. As reminded by the Court of Appeals,

"[w]hereas the balance of equities focuses on the parties, '[t]he public interest inquiry primarily addresses impact on non-parties rather than parties,' and takes into consideration 'the public consequences in employing the extraordinary remedy of injunction'."[42]

In the case at hand, both parties to the dispute agreed that the public interest lies where the free flow of information is maximised. hiQ argued that the access to publicly available data cannot be limited by a private party such as LinkedIn. LinkedIn, in its turn, claimed that it is precisely third parties deploying user data that would create a chilling effect on the users, who would no longer want to make their personal data available online for anybody to use.[43] The Court followed the position of hiQ, considering on the one hand that,

"the actual privacy interests of LinkedIn users in their public data are at best uncertain, [as] those who opt for the public view setting expect their public profile will be subject to searches, data mining, aggregation, and analysis."[44]

Additionally, as put by the Court of Appeals,

"giving companies like LinkedIn free rein to decide, on any basis, who can collect and use data — data that the companies do not own, that they otherwise make publicly available to viewers, and that the companies themselves collect and use — risks the possible creation of information monopolies that would disserve the public interest."[45]

With this analysis of public interest, the District Court considered, and the Court of Appeals further confirmed, that hiQ should be granted a preliminary injunction and ordered LinkedIn to,

"withdraw the cease and desist letters to hiQ dated May 23, 2017 and June 24, 2017 [and to] refrain from issuing any further cease and desist letters on the grounds therein stated during the pendency of [the] injunction."[46]

Appraisal

Both the Polish DPA Decision in the Bisnode case and the preliminary injunction issued in the US hiQ v LinkedIn dispute concern the widely used practice of web scraping. Both Bisnode and hiQ made (re)using of personal data accessible on publicly available pages the core of their businesses. Finally, both companies allege that preventing them from free data scraping will close their doors. This is how far the similarities go. Even though from the factual point of view, the two cases are nearly identical, from the legal point of view, not only do they raise different legal problems, but they also highlight the different approaches to the regulation of scraping. The main discrepancy between the regimes is the position of individual data subjects as opposed to the data controllers and processors. These different takes on scraping reflect a broader disagreement on the approach to data protection, which slowly reaches its limits. ***864**

Before analysing the differences between the EU and the US approaches to data scraping, it is important to point out the terminological issue, i.e. the different understanding of the concept of privacy on the two sides of the Atlantic.

Terminological differences

The roots of the right to data protection lie in the right to privacy. In the EU, however, the right to the protection of personal data grew apart as a self-standing right separate from the right to privacy. This is reflected by the Charter of Fundamental Rights of the EU, which distinguishes the right to privacy (art.7) and the right to protection of personal data (art.8).[47] Consequently, personal data is protected regardless of whether it is private or public and the lawfulness of its processing depends only on the existence of either the consent of the person concerned, or another legitimate basis laid down by law.[48] In the same spirit, the goal of the EU secondary legislation, in particular the GDPR, is not to protect the confidentiality of personal data as such but rather to enable data subjects to retain control over it and ensure a strong and effective protection of data.[49]

In the US, there is no right to protection of personal data independent from the right to privacy. Personal data is protected to the extent necessary to safeguard the right to privacy.[50] In addition, differently from the EU, in the US there are no overarching privacy rules. Instead, privacy is protected through a patchwork of legislation covering certain sectors[51] or addressing particular problems,[52] which leaves several areas and matters unregulated.

Legal approach to the practice of scraping personal data: fundamental rights v commercial interests

This different understanding of the concept of privacy in the EU and the US influences the way the two legal systems approach the practice of scraping personal data. The Bisnode and hiQ v LinkedIn disputes exemplify this discrepancy. In the Polish case,

Bisnode was investigated and fined by the President of the Personal Data Protection Office—an administrative authority vested with the mission of enforcing data protection rules. The hiQ v LinkedIn case, on the other hand, is a private litigation focused on copyright, contract and competition law. Whereas the goal of the DPA was to protect fundamental rights law, the hiQ v LinkedIn case aims at ensuring the company's business continuity (which *in casu* amounted to securing its practice of data scraping). Consequently, whereas the data protection considerations are at the heart of the Bisnode case, in hiQ v LinkedIn, the practice of scraping has not been questioned because of the privacy concerns it raises, but because of the commercial interests it threatens.

Even though, hiQ v LinkedIn treats privacy questions only incidentally (i.e. as a part of the balance of hardships and public interest assessment within the appraisal of the hiQ's motion for an injunctive relief), both parties to the dispute, as well as the courts, discussed extensively the impact of scraping of personal ***865** data on the data subjects' privacy. What is striking, however, is that even though the arguments that have been put forward by the parties and the subsequent discussions of the courts are very similar to those triggered in the Polish case, the conclusions of the two jurisdictions are utterly different.

In the EU system, the rights of individuals with respect to the protection of their personal data are given effect via, among others, the GDPR. Considering the constitutional roots of the personal data protection,[53] it is not surprising that the obligation to provide information enshrined in art.14 GDPR has been taken very seriously by the Polish DPA. The provision was interpreted strictly and the DPA refused to apply the exonerating concept of disproportionate effort. In this regard, it is interesting to note that the argument of the high costs that would have to be endured by the company to comply with the obligation to inform was rejected, in spite of Bisnode's claim that it would "critically disturb the functioning of the Company to the extent which [sic] could imply the need to terminate conducting activity in Poland".[54] The DPO considered that,

"[a f]ailure to fulfil the above-mentioned obligation, due to financial expenses claimed by the Company, indicates lowering of the value of the rights of the data subjects, whose personal data are being processed by the Company, in relation to the value of Company's finances — which cannot be considered as a valid argument in the light of the requirements of the [GDPR]."[55]

The Voivodeship Administrative Court concurred with this statement and found that art.14(5)(b) GDPR,

"is not about financial expenditure on the mere implementation of a fully achievable obligation, but about the real difficulty in being able to implement it at all. The administrator's financial interest is not and cannot be — on the basis of [the GDPR] — a value that prevails over the [data subject's] right to obtain … information about … the legitimate interests pursued by the administrator, as well as the right to request the administrator to access personal data, rectify it, delete it or limit its processing, and the right to object to the processing."[56]

In the hiQ v LinkedIn case, hiQ raised the bankruptcy argument as well, however, differently from the Polish case, it appeared to be the one to tip the balance of equities in its favour at the cost of the users' privacy rights. Similarly to the Polish DPO, the District Court recognised the privacy interests of data subjects and found that setting one's profile to public "does not imply that [the user] wants any third parties to collect and use that data for all purposes".[57] After considering all the arguments, however, it assessed that such LinkedIn users in fact likely do not have any *real* privacy expectations and accordingly, "even if some LinkedIn users retain some privacy interests …, those interests did not outweigh hiQ's interest in continuing its business".[58]

To conclude, in the case of GDPR, fundamental rights cannot be overridden by financial interests, even when this results in the decrease of competitiveness on the market, the loss of financial liquidity or even the need to terminate business activity. In the US, the right to conduct business prevails. ***866**

Position of data subjects

It is true that LinkedIn did not raise the argument of users' privacy for the sake of protecting them from scrapers like hiQ, but rather to protect their own commercial business, and more precisely "the goodwill [it] has developed with its members".[59] Similarly, since the case is a private law dispute, the Court did not balance the equities between on the one side hiQ and on the other side data subjects but between two commercial entities—hiQ and LinkedIn. In fact, the Court of Appeals reiterated that "LinkedIn has no protected property interest in the data contributed by its users, as the users retain ownership over their profiles."[60] It is therefore the only logical conclusion that the interests of individuals were not taken into account… or is it not? LinkedIn might have no protected property interests in the users' data, but neither does hiQ. It is the data subjects themselves who have (protected?) interests in their data. Their absence in the proceedings highlights a bigger problem relating to the position of individuals with respect to personal data-related disputes.

In the European legal order, under the GDPR, DPAs observe compliance and ensure enforcement of data protection rules, amongst others through issuing binding decisions and imposing administrative fines.[61] However, these powers exist in parallel with the right of individual data subjects to institute private proceedings against anyone unlawfully processing their data.[62] On top of that, data subjects can mandate a not-for-profit body, organisation or association to exercise their right to lodge a complaint with a DPA or institute legal proceedings on their behalf.[63] The underlying reasoning of these procedural safeguards is to maximise the effectiveness of the fundamental right to data protection,[64] which translates, amongst others, into ensuring that the data subjects retain as much control as possible over their data.[65]

In the US, however, data subjects do not have any legal possibility to challenge the practice of scraping of their (publicly available) data on the basis of the privacy concerns.[66] The hiQ v LinkedIn case, where data subjects were not parties to the proceedings, highlights this issue. Yet, the problem was considered as potentially affecting the public interest. As a consequence, hiQ and LinkedIn, as well as the courts, had to engage in a discussion over what the interests of individuals are.

hiQ asserted that once individuals share their data with others, there is no expectation that those people would not share it with someone else. Consequently, making a LinkedIn profile public means irrevocably including it in the public—i.e. freely accessible—space. According to hiQ, once this choice is made, there is no turning back,[67] and any assessment of individuals interests at this stage is therefore overdue.

LinkedIn, in its turn, considered that any scraping by third parties would unlawfully compromise the privacy of the LinkedIn users, both those who chose to have their accounts "private" as those who did not opt-in for that option but who had it guaranteed via the LinkedIn Privacy Policy. The main issue pointed ***867** out by LinkedIn is that scraping allows for an anonymous surveillance of users' behaviour which can be further "reported" to their employers.[68]

In this regard, the District Court wondered whether there is,

"any kind of a setting that is available to LinkedIn users who might want to be able to be out there for all purposes, including being subject to collection by bots [as there may be] some advantages to it".[69]

One of such advantages—offered by companies like hiQ—is that,

"people might be seen as valuable, as 'keepers'. … And this is their subtle way of letting their employers know that there's free agency out there, and they might want to keep them."[70]

By this, (despite granting hiQ's request), the Court implicitly indicated that both hiQ and LinkedIn's arguments might be valid, depending on the individual situation of each and every LinkedIn member.

Differently from the District Court, the Court of Appeals did not express any doubts as to the intention of data subjects and concluded that,

"as to the publicly available profiles, the users quite evidently intend them to be accessed by others, including for commercial purposes — for example, by employers seeking to hire individuals with certain credentials."[71]

According to the Court of Appeals, it is therefore in the interest of individuals (or at the very least, not against it) to allow for scraping.

Finally, in its amicus curiae brief, the Electronic Privacy Information Center (EPIC)—a non-governmental organisation advocating for the privacy rights—claimed that LinkedIn users "provide personal data for professional networking purposes and do not expect that [it] will be acquired and monetized by unknown third-parties",[72] and that LinkedIn is obliged to protect this data. More specifically, EPIC indicated that, even when a profile is public,

"[it] serves an important purpose for some users: to allow colleagues, clients, and/or potential employers to locate their digital resume and make a professional connection even if they are not subscribed to LinkedIn. The purpose of the public profile is not to enable data analytics companies to scrape, aggregate, and monetize user data."[73]

This discussion underlines the difficulty in assessing what the privacy expectations and interests of the individuals are. This, together with the lack of legal standing of individuals with respect to the processing of their personal data highlights a broader disagreement on the approach to data protection on the two sides of the Atlantic. ***868**

As this paper has shown, the disagreement on the approach to data protection between the EU and the US lies in the fundamental differences in the two legal systems' image of the individual as bearer of legal interests.[74] As described above, this has its roots in the different assessment of the value of the recognition of data protection as a right independent from the right to privacy.

Digital world divided

The disagreement about the place of the privacy and data protection in the EU and the US grows in importance with the Bisnode and hiQ v LinkedIn disputes. This is because the decisions and judgments issued in the two cases conflict with one another. On the one hand, the Bisnode decision imposes the obligation to notify the individuals, who can then require the company to stop processing their data. On the other hand, the injunction issued in hiQ v LinkedIn not only allows, but even requires companies to give priority to commercial interests over the protection of personal data. It is true that the US case is a preliminary order and remains without prejudice to the outcome of the main proceedings. It might well be that the novelty of the issue coupled with the constitutionally protected freedom of expression and information influenced the outcome of the preliminary proceedings, deliberately leaving the final say on the practice of scraping to the main proceedings. Nevertheless, despite the temporary character of the injunction issued against LinkedIn, in the context of the Bisnode decision, one could imagine that these two conflicting requirements will soon create an impasse. This is because in hiQ v LinkedIn, it is not just "any company" that is at stake—it is LinkedIn which is owned by Microsoft.

Microsoft, as a transnational company operating in both the EU and the US, needs to meet the legal requirements in force in both regions. Instead of complying with multiple privacy regulations, Microsoft decided to level up its data protection standards and to extend the GDPR rights to all consumers around the world.[75] The hiQ v LinkedIn judgment places it in a difficult position. The promised standards cannot be implemented in the case of data scraping, as this would directly contravene the injunction imposed by the US courts. Even if Microsoft decides to comply with the injunction only on the territory of the US, it is not excluded that data scraped by hiQ includes also data of EU citizens. In such scenario, Microsoft is *obligated* by virtue of EU law to respect the GDPR.[76] The necessity to comply with both legislations entails a number of consequences.

One such consequence is a possible infringement of the founding principles of the GDPR: the protection of personal data by design through technical measures.[77] Indeed, the US court required LinkedIn to get rid of the technical safeguards hampering hiQ from scraping LinkedIn profiles.[78] Does such a deliberate removal of technical safeguards amount to a violation of the GDPR? In such scenario, would LinkedIn be required to report a data breach?[79] Or, to the contrary, would the EU data protection authorities draw an analogy with the recent Google v CNIL judgment,[80] which limited the territorial scope of the GDPR-based right to be forgotten to the EU? On top of that, one could wonder whether the LinkedIn case affects the ***869** European Commission's decision on the adequacy of the level of protection of personal data in the US[81] and consequently, whether this decision should be reviewed.[82] Indeed, the self-certification system established by the Privacy Shield has already raised a number of concerns similar to those which put end to the Safe Harbour.[83] Considering that LinkedIn participates in the said scheme,[84] the problem of data scraping adds yet another question mark to the adequacy of the protection of personal data between the EU and the US.

Conclusions

The Bisnode and hiQ v LinkedIn cases show that the existence of different privacy and data protection standards across the world may lead to an artificial division of the otherwise boundaryless digital world and creates an important hindrance to what both the EU and the US value: the free flow of data. Despite the existence of a regime allowing for international transfers between the EU and the US (Privacy Shield), the existing differences may de facto prevent the flow of data, as the latter is accepted by the EU only when it complies with fundamental rights. Moreover, the tension caused by not only different, but now also conflicting, legal obligations in the relation to personal data may give rise to problems in more and more areas. Web scraping is one example, but in the wake of new technologies such as artificial intelligence or automated decision making, it might soon be followed by others. It remains to be seen how the courts and private actors deal with it.

Zuzanna Gulczynska

Ghent University

Footnotes

1 PhD researcher at the Ghent European Law Institute and Fellow of the Research Foundation—Flanders (Project No.1169920N). I would like to thank Paul Nemitz, Merijn Chamon and the anonymous referees for their valuable comments. Any errors or omissions remain solely mine.

1 "The term Smart Data refers to the challenge of transforming raw data into quality data that can be appropriately exploited to obtain valuable insights … Smart Data was born as the attempt of transforming Big Data to a new form of structured data", after: *I. Cordón et al., "Smartdata: Data preprocessing to achieve smart data in R" (2019), Neurocomputing*, https://doi.org/10.1016/j.neucom.2019.06.006 [Accessed 1 November 2020].

2 *ALM Media, "What Courts have said about the legality of data scraping" (20 July 2017) YAHOO! FIN.*, https://finance.yahoo.com/news/courts-said-legality-data-scraping-090000366.html [Accessed 1 November 2020].

3 *A. Agius, "Legal Perspectives on Scraping Data from the Modern Web" (2017), Law in Society*, https://perma.cc/6J3H-B7C5 [Accessed 1 November 2020] cited in: I. Drivas, "Liability for Data Scraping Prohibitions under the Refusal to Deal Doctrine: an Incremental Step toward More Robust Sherman Act Enforcement" (2019) 86 The University of Chicago Law Review 1903.

4 *M. Hughes, "Bots drove nearly 40% of internet traffic last year — and the naughty ones are getting smarter" (2019) thenextweb*, https://thenextweb.com/security/2019/04/17/bots-drove-nearly-40-of-internet-traffic-last-year-and-the-naughty-ones-are-getting-smarter/ [Accessed 1 November 2020].

5 Drivas, "Liability for Data Scraping Prohibitions under the Refusal to Deal Doctrine" (2019) 86 The University of Chicago Law Review 1903–1904.

6 See, e.g., Facebook, Inc v Power Ventures, Inc 91 U.S.P.Q. 2d 1430 (US District Court for the Northern District of California (N.D. Cal.)); Ryanair Ltd v PR Aviation BV (C-30/14) EU:C:2015:10; [2015] 2 C.M.L.R. 36.
7 hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301 (N.D. Cal., pending).
8 ZSPR.421.3.2018 (15 March 2019), President of the Polish Data Protection Office; English translation available at https://uodo.gov.pl/en/file/314 [Accessed 1 November 2020].
9 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office); English translation: https://uodo.gov.pl/en/file/314[Accessed 1 November 2020].
10 II SA/Wa 1030/19 (11 December 2019), Polish Voivodeship Administrative Court.
11 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017).
12 hiQ Labs, Inc v LinkedIn Corp, No.17-16783 (US 9th Circuit Court of Appeals (9th Cir.), 9 September 2019.
13 Article 267(3) of the Treaty on the Functioning of the European Union [2012] OJ C326 (TFEU).
14 Article 14(1)–(3) of Regulation 2016/679 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 (General Data Protection Regulation) [2016] OJ L119 (GDPR).
15 Artice 14(5)(b) GDPR.
16 Article 34, Ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych, t.j., Dz. U. 2018 poz. 1000 (Polish Data Protection Act); art.34 s.2 GDPR.
17 Article 58(2) GDPR.
18 Article 83 GDPR.
19 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office); English translation: https://uodo.gov.pl/en/file/314[Accessed 1 November 2020].
20 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office), p.6.
21 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office), p.6.
22 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office), p.7.
23 ZSPR.421.3.2018 (15 March 2019), (President of the Polish Data Protection Office), p.10.
24 II SA/Wa 1030/19 (11 December 2019), Polish Voivodeship Administrative Court.
25 A referral by the Supreme Administrative Court is highly probable, since in line with art.267(3) TFEU, national courts adjudicating at last instance have a duty to bring questions over the correct application of EU law before the CJEU. In line with the CILFIT doctrine this obligation may be waived only in three situations: if the question raised before the national court is irrelevant, if it has already been answered by the CJEU or if the correct application of EU law does not leave scope for any reasonable doubt (see: CILFIT SrL v Ministero della Sanita (C-283/81) EU:C:1982:335; [1983] 1 C.M.L.R. 472 at [10], [14] and [16]). In the opinion of the author, none of these circumstances are present in this case.
26 See, e.g., Google Spain (C-131/12) EU:C:2014:317; Planet49 (C-673/17) EU:C:2019:801; Glawischnig-Piesczek v Facebook Ireland Ltd (C-18/18) EU:C:2019:821; [2020] 1 C.M.L.R. 33.
27 Digital Millennium Copyright Act, Pub. L. 105-304, 112 Stat. 2860 (1998).
28 Computer Fraud and Abuse Act 18 USC. § 1030, Pub. L. No. 99-474, 100 Stat. 1213.
29 California Penal Code § 502(c).
30 hiQ's Complaint for Declaratory and Injunctive Relief of 7 June 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, pp.16–19, https://www.courtlistener.com/docket/6071320/1/hiq-labs-inc-v-linkedin-corporation/ [Accessed 1 November 2020].
31 hiQ's Complaint for Declaratory and Injunctive Relief of 7 June 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, pp.16–19.
32 Winter v Nat. Res. Def. Council, Inc 555 US (US Supreme Court, 12 November 2008).
33 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017).
34 hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019).
35 hiQ's Complaint for Declaratory and Injunctive Relief of 7 June 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, pp.12–14.
36 LinkedIn's Opposition to Motion of 26 June 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, p.25.
37 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), p.22.
38 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), pp.7 and 24; see also: hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), p.15.
39 Transcript of the hearing of the Court of Appeals for the 9th Circuit of 27 July 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, p.32, https://www.hiqlabs.com/s/Hearing-Transcripts-and-Exhibits.zip [Accessed 1 November 2020]; see also Exhibit B in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, "LinkedIn User Agreement", point 3.1.c. in conjunction with point 2.5., https://www.courtlistener.com/recap/gov.uscourts.cand.312704.23.1.pdf [Accessed 1 November 2020].
40 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), pp.6–7.

41 hiQ Labs, Inc v LinkedIn Corp, No.17-16783 (9th Cir., 9 September 2019), p.16.
42 Bernhardt v Los Angeles Cty, 339 F.3d 920, 931–32, (9th Cir, 5 August 2003), cited in: hiQ Labs, Inc v LinkedIn Corp, No.17-16783 (9th Cir., 9 September 2019), p.35.
43 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), pp.23–24.
44 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), p.24.
45 hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), p.38.
46 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), p.25.
47 *S. Peers et al. (eds), The EU Charter of Fundamental Rights. A Commentary (Oxford: Hart Publishing, 2014), p.229*.
48 Charter of Fundamental Rights of the European Union [2012] OJ C326 (EU Charter) art.8(2).
49 Recital 7, GDPR.
50 This is also reflected in the American legal terminology which prefers using the term "privacy" over "data protection". See *S. Schiedermair, "Data Protection: is there a Bridge across the Atlantic?" in D. Dörr and R. L. Weaver (eds), The Right to Privacy in the Light of Media Convergence: Perspectives from Three Continents (Berlin: De Gruyter, 2012), p.360*.
51 E.g. Health Insurance Portability and Accountability Act of 1996, Pub.L. 104–191, 110 Stat. 1936; Fair Credit Reporting Act, 15 USC. § 1681, Pub.L. 91-508, 84 Stat. 1114.
52 E.g. Children's Online Privacy Protection Act, 15 USC. §§ 6501–6506, Pub.L. 105–277, 112 Stat. 2681-728; state-level data breach notification laws, such as Cal. Civ. Code §§ 1798.29, 1798.82.
53 Treaty on the European Union [2012] OJ C326 in conjunction with art.8 EU Charter art.6.
54 Decision of 15 March 2019, Ref. ZSPR.421.3.2018 (President of the Polish Data Protection Office), p.4; English translation: https://uodo.gov.pl/en/file/314 [Accessed 1 November 2020].
55 Decision of 15 March 2019, Ref. ZSPR.421.3.2018 (President of the Polish Data Protection Office), p.10; English translation: https://uodo.gov.pl/en/file/314 [Accessed 1 November 2020].
56 II SA/Wa 1030/19 (11 December 2019), (Polish Voivodeship Administrative Court) [original in Polish, English translation by the author].
57 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), p.23 cited in hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), p.14.
58 hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), pp.2–3.
59 hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), p.14.
60 hiQ Labs, Inc v LinkedIn Corp No.17-16783 (9th Cir. 2019), p.16.
61 Article 58 GDPR.
62 Article 79 GDPR.
63 Article 80 GDPR.
64 Recital 11 GDPR.
65 Recital 7 GDPR.
66 Indeed, despite some statutory protection of consumers' privacy in the US, this protection does not cover publicly available personal data. See, FTC, "Data brokers: a call for transparency and accountability" (May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [Accessed 1 November 2020]..
For a broader discussion on the consumers' privacy protection in the US, as compared to the EU horizontal protection, see: P.M. Schwartz and K.N. Peifer, "Transatlantic Data Privacy Law" (2017) 106 Georgetown Law Journal 115.
67 Transcript of the hearing of the Court of Appeals for the 9th Circuit of 27 July 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, https://www.hiqlabs.com/s/Hearing-Transcripts-and-Exhibits.zip [Accessed 1 November 2020].
68 Transcript of the hearing of the Court of Appeals for the 9th Circuit of 27 July 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, p.17.
69 Transcript of the hearing of the Court of Appeals for the 9th Circuit of 27 July 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, p.17.
70 Transcript of the hearing of the Court of Appeals for the 9th Circuit of 27 July 2017 in hiQ Labs, Inc v LinkedIn Corp 3:17-cv-03301, p.17.
71 hiQ Labs, Inc v LinkedIn Corp, No.17-16783 (9th Cir. 2019), p.17.
72 Brief of Amicus Curiae Electronic Privacy Information Center (EPIC) in Support of Neither Party Urging Reversal of 10 October 2017 in hiQ Labs, Inc v LinkedIn Corp, 3:17-cv-03301, p.5, https://epic.org/amicus/cfaa/linkedin/HiQ_Response_Brief.pdf [Accessed 1 November 2020].
73 Brief of Amicus Curiae Electronic Privacy Information Center (EPIC) in Support of Neither Party Urging Reversal of 10 October 2017 in hiQ Labs, Inc v LinkedIn Corp, 3:17-cv-03301, p.9.

74 For a broader analysis of the problem, see: P.M. Schwartz, K.N. Peifer, "Transatlantic Data Privacy Law" (2017) 106 Georgetown Law Journal 115, 119.

75 *J. Brill, Microsoft Corporate Vice President for Global Privacy and Regulatory Affairs and Chief Privacy Officer, "Microsoft's commitment to GDPR, privacy and putting customers in control of their own data" (Microsoft 2018), Microsoft Blog*, https://blogs.microsoft.com/on-the-issues/2018/05/21/microsofts-commitment-to-gdpr-privacy-and-putting-customers-in-control-of-their-own-data/ [Accessed 1 November 2020].

76 Article 3(2) GDPR.

77 Articles 25 and 32 GDPR.

78 hiQ Labs, Inc v LinkedIn Corp 273 F. Supp. 3d 1099 (N.D. Cal. 2017), pp.11 and 25.

79 Article 33 GDPR.

80 Google LLC v Commission Nationale de l'Informatique et des Libertes (CNIL) (C-507/17) EU:C:2019:772; [2020] 1 C.M.L.R. 24.

81 Commission implementing decision 2016/1250 pursuant to Directive 95/46 on the adequacy of the protection provided by the EU-U.S. Privacy Shield [2016] OJ L207.

82 The Commission's decision was invalidated in the judgment of 16 July 2020, Data Protection Commissioner v Facebook Ireland Ltd (C-311/18) EU:C:2020:559, before the publication of this article. Given the findings of the CJEU relating to the unique level of protection required for all international transfers regardless of the legal basis used, the conclusions of this article remain relevant not only for any future adequacy decision but also for transfers based on one of the appropriate safeguards.

83 See, e.g., A.-L. Philouze, "The EU-US Privacy Shield: Has Trust Been Restored" (2017) 3 European Data Protection Law Review 11.

84 Privacy Shield List, https://www.privacyshield.gov/participant?id=a2zt0000000L0UZAA0&status=Active [Accessed 1 November 2020].

E.L. Rev. 2020, 45(6), 857-869