https://doi.org/10.6092/issn.1825-1927/18875 ISSN 1825-1927

Web Scraping: A Private Law Perspective

Alessandra Quarta* Michael William Monterossi†

Abstract.

The Article deals with the analysis of the implications of web scraping from a private law standpoint. In the absence of a legal regime regarding online non-personal data, website managers have attempted to protect them from the automatic extraction of data by providing for anti-scraping clauses in the websites' terms of use, by alleging the violation of presumed property rights over the websites and/or their contents or by affirming that scraping could constitute a conduct counter to the principles of professional correctness. The Article provides an analysis of the legitimacy of such strategies and a critical evaluation of the criteria which judicial courts – both in Europe and in the US – tend to elaborate in order to determine the limits within which web scraping is to be deemed a lawful practice.

Keywords:

Web scraping, data, terms of use, property, unfair competition.

1 Introduction

Since the dawn of the Internet, web scraping has represented a fundamental tool for unleashing the potential of a global network of interconnected computers and boosting technological innovation through data. To cite just a few cases of use, the automatic search and extraction of online data has enabled the enhancing of the search engines which underlie the success of the World Wide Web, has favored the emergence of a myriad of aggregator services, allowing users to orient themselves within the vast range of commercial offers circulating in cyberspace and, nowadays, is permitting the collection of huge amounts of data to feed machine learning AI systems for their relentless training sessions.

In spite of its pervasiveness in the digital universe, web scraping has been stigmatized by many business operators, backed up, in some cases, by scholars and public opinion. Besides the problem stemming from the automatic collection of personal data, which within European Union clearly falls under the spectrum of the General Data Protection Regulation[1], the most relevant legal issues have arisen in relation to the extraction of that immense amount of non-personal data that can be found in the firmament of websites that light up cyberspace. From this perspective, the aversion towards scraping is typically fueled by strictly economic interests, linked to the protection of the core commercial activity carried out on the website subjected to data scraping. The complaints mostly derive either from the development – with the data extracted – of new parallel services, which may undermine the website managers' opportunity "to maximize [their] own revenues" or from website misfunctioning caused by uninterrupted automatic server queries.

In the absence of a definite legal regime regulating non-personal online data, website managers have developed a wide variety of factual and/or legal strategies to block or at least dissuade users from employing bots to scrape data. Alongside allegations of IP or quasi-IP rights violations (such as trade secrets and *sui generis* database right), most of the strategies rely on – or have

*University of Turin, Department of Law; ✉ alessandra.quarta@unito.it

†University of Turin, Department of Law; ✉ michaelwilliam.monterossi@unito.it

1. Whenever web scraping involves personal data, the collection and processing must be carried out in accordance with the rules laid down in the European Regulation (EU) 2016/679. In Italy, the scraping of personal data was sanctioned Data Protection Authority even prior to the entering into force of the GDPR See, for example, the provision adopted on the 14th of January 2016 concerning *Elenchi telefonici on line e "ricerca inversa": illegittimi se la fonte non è il d.b.u.*, [doc. *web* n. 6053915].

implications for – private law institutions. In particular, the unlawfulness of web scraping has been grounded on *contract law*, in relation to the violation of specific anti-scraping clauses included in the websites' terms of use; on *property law*, by arguing that the unauthorized use of scraping bots amounts to a violation of website managers' proprietary rights over the website and/or content published therein; and on *unfair competition*, by affirming that the extraction and reuse of data by "freeloading" scrapers could constitute a conduct counter to the principles of professional correctness.

The uncertainty over the legitimacy of this order of allegations has led to a conspicuous number of legal disputes on both sides of the Atlantic. Despite differences in the legal systems, the decisions handed down by the courts provide the jurist with some convergent lines of interpretation, which can give market operators guidelines on the limits within which web scraping can be considered a lawful practice. The article aims to bring to the surface such case-law orientations and to critically evaluate the soundness of their theoretical foundations as well as the impact they generate on the circulation and use of data among market actors.

2 Web Scraping and Terms of Use: A Contract Law Perspective

It is well known that the Internet was developed as a public network to connect U.S. governmental agencies and then, successively, universities and research centres (ARPAnet – Advanced Research Project Agency Network). At the very beginning of its introduction, it was conceived as a pure communication infrastructure, enabling the exchange of information, scientific papers and research findings. The Internet changed its appearance and functions at the beginning of the Nineties when the U.S. government decided to open the network to previously excluded private players.

Companies were interested in entering the network to develop their own businesses also online: the Internet gave them the possibility to find new customers and expand their commercial activities. The Internet hosted mainly those companies that had a real business in the analogue world – as for instance, retailers – and, in fact, they used the network as a global shop window.

This first phase of Internet development is referred to as "web 1.0": this expression describes a static network, in which users did not play an active role, since they could only enjoy the contents – such as texts, pictures or multimedia files – created and uploaded by website managers. These contents were generally organized in web pages and these web pages constituted a website.

The legal qualification of these contents stimulated the introduction of *terms of use*.

To understand this transformation, it is essential to clarify some points.

Firstly, not all the website contents are creative works. In particular, the main problem concerning this statement arises when considering compilations of data, as for instance timetables or price lists. According to the US Copyright Act, a "compilation" is "a work formed by the collection and assembly of pre-existing materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship". Therefore, originality is the essential requirement to benefit from the exclusive rights recognized to the holder of the copyright[2]; otherwise, he or she cannot prohibit the reproduction and the creation of copies.

This legal framework is particularly relevant for commercial entities who use the web: in fact, data were (and continue to be) the essential digital assets used to organize online business. Allowing their copy and reproduction implicates that competing activities can be easily organized. The best example to understand this point is represented by online travel agencies, who generally reproduce prices and schedules of different airline companies by scraping the data published on their websites.

The lack of originality entailed the lack of copyright protection for the website contents and their free reproduction. To avoid this legal effect and maintain the exclusive right to reproduce and make copies, website managers prohibit reproduction by exploiting the binding effects of a contract: to this end, they came up with *terms of use*[3]. In fact, these texts include an essential regulation core, consisting of rules that forbid extraction, copy and reproduction of the data contained on the website.

2. The right to reproduce and make copies of an original work; The right to prepare derivative works based on the original work; The right to distribute copies to the public by sale or another form of transfer, such as rental or lending; The right to publicly perform the work; the right to publicly display the work, and The right to perform sound recordings publicly through digital audio transmission.

3. Lemley, Mark A. "Shrinkwraps in Cyberspace", in *Jurimetrics J*., 35. (1995): 311-323, at 312. In that period, legal scholars mainly focused on the ability of contract law to respond to the problems deriving from the application of tort law and criminal law. Dunne, Robert L. "Deterring Unauthorized Access to Computer: Controlling Behaviour in Cyberspace through a Contract Law Paradigm", in *Jurimetrics J*., 35 1, (1994): 1-15, at 12-13.

However, to be a valid contract, terms of use need the user's acceptance. Two main solutions have been implemented to achieve this goal. Firstly, in the web 1.0 era, website managers introduced a small box which referred both to the terms of use and the privacy policy to be clicked in order to express acceptance. Jurists usually classify such terms of use as "click-wrap contracts". Secondly, in the web 2.0 era, a new solution was implemented: users accept the terms of use when they start using the web site, i.e. when they browse through the pages; a link to the terms of use text is provided at the bottom of the homepage, so that users can read them if they so wish. In this case, terms of use are classified as "browse-wrap contracts"[4].

Thanks to both these models, website managers could build the terms of use as a contract and make the prohibition of the reproduction of contents a binding rule. In this way, users promised not to extract and copy the information presented on the website; the infringement of this term was a breach of contract.

Coming to the EU implementation of these models, we shall notice that terms of use were applied to protect website contents, even if the legal framework was radically different.

With regards to the protection of the database, the EU regulation is more complete than that of the US. In fact, databases are provided with the typical protection of copyright if they can be qualified as original compilations; in alternative, they can be protected with the so-called *sui generis* right if they respond to the description found in art. 7 Directive 96/9/EC[5]. Thus, the holder of copyright or *sui generis* right has exclusive rights and in particular the right to reproduce and make copies of their compilation. According to this double track, it could be sustained that those databases not framed in one of the two legal models, should be considered publicly accessible and their extraction or copy should not be authorized by the creator.

Even if in the EU, protections are more pervasive, the terms of use continue to include provisions to prohibit content reproduction. This solution allows website managers to eliminate the risk that a judicial authority would deny the application of both copyright and *sui generis* right and consider, as a consequence, the extraction and the reproduction of online compilations as legal activities.

The European Union Court of Justice (EUCJ), in deciding the *PR Aviation* case[6], considered this strategy as valid, admitting that exclusive rights can be introduced by contracts when compilations are neither original works nor created with a significant qualitative or quantitative investment. This decision was strongly criticized, since it creates a paradox: in fact, currently, databases that do not match the ones protected by the law are more strongly defended against reproduction than those that benefit from the exclusive rights described in Directive 96/9/EC[7]. Another critique derives from observing that the EUCJ did not valorize the double track system of protection, with the consequence that, currently, the free reproduction of those compilations not protected by legal exclusive rights can always be avoided.

In this scenario, national courts coped with data scraping contrasts by applying contract law and, in particular, by taking into consideration the formation of contracts and the acquisition of the user's acceptance. The most problematic issues concern the application of the *browse-wrap* model, since browsing the website has not been considered as a behavior apt to constitute a contractual legal relationship[8]. Furthermore, national courts have stressed that the relationship between the website manager and the company scraping is not regulated by a contract, since there the latter cannot be deemed to have manifested his or her consent to the terms of use contract.

This interpretation is supported mainly on the grounds of two arguments.

In the other words, the lack of an agreement is inferred, first of all, from the 'devious methods' of data extraction, which by employing a bot 'disguised' as a human user would be aimed precisely at 'evading contractual commitments': in other words, the

4. Lemley, Mark A. "Terms of Use", in *Minnesota L. Rev*., 91 (2006): 459-483.

5. «Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database».

6. Case C-30/2014, Ryanair c. PR Aviation, Judgment of 15 January 2015, in curia.europa.eu.

7. Borghi, Maurizio, and Stavroula Karapapa. "Contractual Restrictions on Lawful Use of Information: Sole-Source Databases Protected by the Back Door?," 37 (8) EIPR, (2015): 505-511; Bottis, Maria. "How Open Data Become Proprietary in the Court of Justice of the European Union", in Sokratis K. Katsikas and Alexander B. Sideridis (eds.), *E-Democracy. Citizen Rights in the World of the New Computing Paradigm*, Springer, Switzerland, 2015. 169-174; Gupta, Idranath, and Vishwas H. Devaiah. "The Database Directive «Contracting Out» Bar: Does It Apply to Unprotected Databases?", J. *Intell. Prop. L. & Pract*. 10.9 (2015): 669-672.

8. Cour d'Appel de Paris, *Ryanair v. Vivacances* - sentence issued on March 23, 2012.

scraper's conduct is manifestly contrary to the terms and conditions, so that it would make no sense to consider that the scraper has entered into an agreement for the sole purpose of violating it[9].

Secondly, the mere fact that a person has engaged in conduct that is not permitted by website managers, who do not use technical means to prevent scraping cannot in itself be equated to a breach of contract: according to the Court in the *Ryanair v. Atrapalo* case, in particular, the anti-scraping clause, where not supported by the provision of adequate technical protection measures, constitutes the mere expression of an intention on the part of the manager of the website, which cannot be binding for third parties[10].

3 Data Scraping as a Violation of Property Rights Over Websites

Besides relying on terms of use, there have been several attempts to base protection against unauthorized automated data mining on an alleged violation of property rights over the website and/or its contents.

In particular, two theories have emerged from a proprietary perspective.

The first theory, developed within US case-law in the first decade of the twenty-first century, sought to link the illicit nature of web scraping to the violation of property over the (tangible) assets that are (or were) at the basis of website functioning. This aim was pursued by exploiting an old action – known as *trespass to chattels* – that can be exercised by the owner of a movable asset (personal property) against anyone who has interfered with its enjoyment[11]. More precisely, US case-law has held that the 'electronic impulses' generated by spiders and other similar automatic programs may cause interference with the enjoyment of the components and resources that allow the website to function (such as, for example, hardware, servers, computer equipment, computer capacity). As a consequence, when the interference has caused damage, the website manager is legitimized to obtain an injunction on the basis of the trespass action[12]. It is worth noting that such a doctrine has also granted protection to the website manager against *danger externalities* associated with the use of bots: in several cases, the action has been granted to a company whose website had been scraped, in view of the danger of (likely) serious harm that could result to the latter should other competitors, different from the defendant in point, undertake similar automated, parasitic mining[13].

The second theory, elaborated within the German doctrinal context, relies on the concept of *virtual property*. Some scholars have hypothesized that the manager has a right of ownership over the virtual space that the website occupies – or, we could say, represents – within the digital sphere: according to the so-called *virtuelles hausrecht*, the manager – like the owner of an immovable good – would enjoy the power to exclude access to his/her website[14]. The debate on such a right has been (to a

9. See, in particular, Tribunale di Milano, sez. spec. in materia di impresa, *Viaggiare c. Ryanair*, 4.6.2013.

10. See, in this regard, Tribunal Supremo - Sala Primera, de lo Civil, *Ryanair v. Atrápalo*, 9.10.2012, STS 572/2012.

11. This is an intentional tort applicable only to movable property, which can be invoked in cases where the interference, while not so serious as to prevent its use (in which case so-called conversion can be invoked), has resulted in the occurrence of harm to the owner (according to the criterion of proximate cause). See Restatement (second) of Torts § 218. Next to trespass to chattels, the American system contemplates a trespass to land, which can be enforced against a person who interferes with another's private property on real estate (real property). For this purpose, it is sufficient to prove unauthorized entry into the subject property, without the need to prove the existence of damage.

12. See in particular the following cases: *eBay, Inc. v. Bidder's Edge, Inc.*, 100 F. Supp. 2d 1058, 1069-72 (N.D. Cal. 2000); *Register.com, Inc. v. Verio, Inc.,* 356 F.3d 393, 396, 404 (2d Cir. 2004); *Ticketmaster Corp. v. Tickets.com, Inc.* (2000), *cit*.; *Ticketmaster Corp. v. Tickets.com, Inc.* (2003), *cit*.; *Craigslist Inc. v. 3Taps Inc*., No. CV 12-03816 CRB, 942 F.Supp.2d 962 (2013). In legal doctrine, see Quilter, Laura. "The Continuing Expansion of Cyberspace Trespass to Chattels", in *Berkeley Tech. L. J*., 17, 2002: 421-443; Lemley, Mark. A. "Place and Cyberspace", in *Cal. L. Rev*. 91 (2003): 521-542; Bellia, Paul L. "Defending Cyberproperty", in *N.Y.U. L. Rev*. 79 (2004): 2164-2273; Hirschey, Jeffrey K. "Symbiotic Relationships: Pragmatic Acceptance of Data Scraping", in *Berkeley Tech. L.J.* 29 (2014): 897-928, at 902.

13. *Register.com, Inc., v. Verio, Inc.,* 126 F. Supp.2d 238 (S.D.N.Y. 2000). This consideration would also seem to have guided the decision in eBay, Inc. v. Bidder's Edge, Inc. the first in which the action for trespass to chattels was used to counter the use of bots. Here, in fact, the automated extraction activity had involved very little use of eBay's server resources and, therefore, no harm had been found to eBay's subject property. What warranted the granting of the action, however, was the finding of the risk of damage to computer systems and, therefore, to eBay's business associated with the possibility of other competitors performing the extraction of data found on its site, for the development of which the plaintiff had incurred a very high cost. See, on this front, the statements made by the court in *Ticketmaster Corp. V. Tickets.com Inc*. (2000), *cit*. with respect to the *eBay* case. In such cases, however, the plaintiff has to demonstrate that the risk of damage.

14. This right was recognized for the first time in the decision handed down by the Bonn Regional Tribunal at the end of the 1900s (LG Bonn, 16.11.1999, 10 O 457/99) in order to give the website "owner" of *chat room* the faculty to block access to certain users. Subsequently, the right was recognized also for the manager of an online discussion forum (LG München I, 25.10.2006 - 30 O 11973/05); of an online community (LG Ulm, 13.1.2015 - 2 O 8/15) and for the owner of an online shop (see OLG Hamm, 23.10.2007 - 4 U 99/07; LG Hamburg, 28.8.2008 - 315 0 326/08). For a thourogh analysis of the theory of *virtuelles hausrecht* see Ladeur, Karl-Heinz, "Ausschluss von Teilnehmern aus Diskussionsforen im Internet: Absicherung von Kommunikationsfreiheit

large extent) triggered by the conflicts deriving from data scraping and has been invoked in court cases by website operators to prevent extraction, usually by competitors, deemed (at least by plaintiffs) to be an obstacle to the exercise of their own freedom of enterprise[15].

The theory has been criticized by both legal scholars and judges. It has been pointed out that the construction of the "virtual right of domicile" cannot be justified from a theoretical perspective due to the necessary 'relativity' of the interest to exclude others from one's own digital space. Unlike the landlord who has an interest in excluding third parties from his property, the essence of the website lies precisely in the fact that the space corresponding to it is "visited" by users and, therefore, that third parties can generally access it[16].

4 From the Property of Websites to the "De Facto" Control of Online Data: The Role of Technological Access Barriers in the Online Data Legal Regime

Even though the theories regarding the property of websites have not been widely accepted, according to a recent line of case law the lawfulness of web scraping depends on whether the website manager has employed – and the data scraper circumvented – code-based authentication mechanisms to limit access to the website or to a portion of the same.

In the US, the relief has emerged in the *hiQ Labs v. LinkedIn* case[17], in connection with the assessment of the Federal Computer Fraud and Abuse Act (CFAA), a law that punishes (criminally and civilly) anyone who intentionally accesses a protected computer without authorization in order to obtain the information contained therein. According to the US courts, the conduct of scraping can be deemed as an unauthorized access to a "computer" only as long as the person carrying out scraping activities has circumvented an 'authentication gate', set up by the website manager in order to protect its website[18]. Otherwise – i.e. if the website is freely accessible – access to the *data* published therein will not be "without authorization": in fact, the failure to provide technical protection measures in order to regulate access to the website conveys the social norm that the space corresponding to it is open and accessible to all Internet users, in the same way as a 'modern public square'[19].

Although grounded on different legal bases, similar conclusions have been reached by the German Federal Supreme Court, in the *Flugvermittlung im Internet* decision, concerning the *Ryanair v. CheapTickets.de* case[20]. The Supreme Court stated that when entrepreneurs decide to make their offer accessible to the public, they must accept the fact that – in the light of the general interest in the functioning of the Internet – the *information* they publish may be subject to automated procedures set up by standard search services and be made available to users in a form appropriate to their search needs. To put it bluntly: website managers can always take advantage of the open and shared structure of the Internet, but, in that case, they will also have to bear the negative stakes arising from the option for a non-restricted website access model.

---

durch „netzwerkgerechtes" Privatrecht", in *MMR* (2001): 787 ss; Feldmann, Thorsten and Joerg Heidrich, "Rechtsfragen des Ausschlusses von Usern aus Internetforen", in *CR* (2006): 406-412; Maume, Philipp. "Bestehen und Grenzen des virtuellen Hausrechts", in *MMR* (2007): 620-625; Klickermann, Paul H., "Virtuelle Welten ohne Rechtsansprüche", in *MMR* (2007): 766-769; Piras, Gabriella, *Virtuelles Hausrecht?*, Mohr Siebeck, Tübingen, 2016.

15. LG Hamburg, 28.8.2008, *cit*.

16. OLG Frankfurt a.M., (5.3.2009), *cit*.

17. *HiQ Labs, Inc. v. Linkedln Corp.,* 31 F.4th 1180 (9th Cir. 2022).

18. The principle has been recalled in the *Memorandum opinion and order* by the United District Court for the District of Delaware in the recent *Ryanair DAC v. Booking Holdings Inc. et al* case, where Ryanair has alleged, among others, the violation of CFAA due to the access of Booking (or its affiliates) in the "myRyanair" online page in order to book flights on behalf of their clients. For other decisions in line with the statement see, e.g., *Greenburg v. Wray*, No. 22-cv-122, 2022 WL 2176499, at 2 (D. Ariz. June 16, 2022); *Meta Platforms, Inc. v. BrandTotal Ltd.*, No. 20-CV-07182, 2022 WL 1990225, at 24 (N.D. Cal. June 6, 2022).

19. It should be noted that there is no agreement on the technological protection measures which should be considered suitable to mark the aforementioned distinction. The uncertainties regard in particular the technical measure known as CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) and the one implemented by blocking the users' IP address. In the *hiQ Labs* case the Court, agreeing with an argument put forward by legal doctrine (see Kerr, Orin S., "Norms of Computer Trespass", in *Columbia Law Rev*., 116 (2016): 1143 ff), ruled that these technical measures, unlike the password gate, are not suitable to prevent access to certain users, but merely to slow down their ability to obtain information from the relevant servers. It follows, according to this reading, that the conduct of "unauthorized access" cannot be said to have taken place, even when the scraper, in order to extract the data, had to overcome the aforementioned technical protection measures.

20. See, in particular, OLG Hamburg, *Ryanair v. CheapTickets.de*, 24.10.2012, 5 U 38/10, which was reformed by the Bundesgerichtshof, BGH, in the decision *Flugvermittlung im Internet*, of 30.4.2014, I ZR 224/12.

Up to now, such a distinction has been elaborated only in view of averting that web scraping activity on freely accessible websites would be sanctioned as illicit. From a theoretical viewpoint, this would confirm that website managers cannot claim a proprietary stance either on an open website or on the data available therein (unless, of course, eligible for IP protection). This does not mean that, in such cases, website managers are not entitled to obtain relief if damaged by data scraping activity. However, they will be protected against unauthorized access by means of a mere *liability rule*, according to which users are allowed to freely access and visit the website (also through bots), provided that they compensate the manager for any damage deriving from the misfunctioning of the website. This is the case, for example, of the damage that could be caused to the economic activity of the website manager when the use of scrapers leads to an overloading of the network, with correlative loss of customers.

Still, the second hypothesis – the one in which the manager actually made use of technological measures to restrict access to his or her site – has significant theoretical implications which should not be underestimated. Now, by drawing the logical consequences from the argument, it should follow that the adoption of technological access barriers allows the website manager to legitimately avoid being subjected to web scraping activity. Although this seems to be true, it is not clear whether the consequent unlawfulness of web scraping would depend on the unauthorized access to the website or to the data available on it. To put it the other way around: it is not clear whether technological barriers have the effect of rendering the website a "private space" under the control of website managers, so that they can decide who can access it, and under which conditions; or, on the contrary, if the use of authentication rules directly affect the legal status of the data, which – undergoing a sort of enclosure – ceased to be public.

The difference is not merely theoretical: indeed, according to the former hypothesis, only the unauthorized access to the website could be sanctioned, whereas, in light of the latter one, the same use of the data extracted would amount to an illicit conduct, independently of whether the person using the data is the one who originally accessed the website without authorization. Although the first option appears to be more consistent with the premises underlying the argument, courts tend to link the effects of the use of technical measures directly to the legal regime of data. This has been made explicit, although in an *obiter dictum*, in the *hiQ Labs v. LinkedIn* case. The Court suggested that had the website in question been "protected by [a] username and password authentication system" rather than being "available to anyone with a web browser", the user's accessing of that *data* might have been without authorization. Similarly, in the *Flugvermittlung im Internet* decision, the German Supreme Court seems to rely on the free accessibility of the website as an argument in support of the public nature of the information contained therein. In this way, however, the Court is indirectly suggesting that whenever website managers decide to limit – by technological barriers – access to websites, data contained therein fall under their own «de facto» control.

Despite being in line with the traditional mode of "appropriation" of non-personal data in the data-driven economy, where the power to control and profit from data is relegated to those actors who can exclude third parties by relying on technological means, this interpretation seems to run into the same theoretical paradox which emerged from a contractual viewpoint: that of legitimizing the 'proprietarization' of data which show neither originality nor special investment efforts as a consequence of a mere expression of website managers private self-determination.

5 Web Scraping and Unfair Competition: Insights for a New Perspective

The distinction between restricted and non-restricted websites has also been identified – at least implicitly – as the decisive factor in order to evaluate the adherence of web scraping activity to the principles underlying fair competition. As a matter of fact, in many of the cases of disputes, the unlawfulness of web scraping has been claimed on the basis of the violation of the principles of professional correctness among entrepreneurs. As noted above, this is particularly evident in the constellation of *Ryanair* cases, since the alleged damage referred to the financial losses – in terms of decrease in sales of packages or complementary services offered to supplement journeys or of loss of advertising revenues – suffered as a consequence of the diverted online traffic caused by the aggregation flight services websites.

On this front, judicial courts have shown the propensity to link the unfair nature of the conduct to the (hypothetical) scraper's circumvention of the website's protection devices[21]. The reasoning of the Courts is generally the following: if owners of freely accessible websites intend, as part of their commercial strategy, to avoid the diversion of Internet traffic from their website, then they must make this intention manifest by adopting technical protection measures.

This approach seems to be questionable from several perspectives.

21. See, in particular, the aforementioned decisions by the Tribunale di Milano, *Viaggiare c. Ryanair*, and Tribunal Supremo - Sala Primera, de lo Civil, *Ryanair v. Atrápalo*.

From a strictly interpretative angle, the eventual circumvention of authentication gates does not appear per se sufficient to determine an infringement of unfair competition rules. Hypothetically, a party could extract data, by circumventing the technological barriers, without even reusing them. Therefore, it seems that the assessment of unlawfulness, under this profile, cannot rely on the mere extraction of data without authorization, but should revolve around the impact that the activity of reuse of the data has in the sphere of the website manager's business activity: it is the reuse of data, and not their mere extraction, that may lead to a behavior qualifiable as deceptive, parasitic and so on and, thus, be harmful to the website manager from a competitive point of view.

By shifting the perspective in this direction, the criteria to assess the unfair nature of web scraping would transcend the mere private interests of the parties involved, so as to take into account also the concomitant interests of the general public in the services or products enabled by the use of the scraped data. In order to identify the limits within which the use of such technique is permissible, courts would have to assess whether, and to what extent, the activity based on the reuse of extracted data produces benefits for the public, facilitating, for example, the matching of supply and demand, promoting price transparency, or even ensuring innovative services and products capable of increasing efficiency and economic competition. Therefore, only when the reuse of data reveals itself to be detrimental (also) to the market, website managers would be entitled to obtain a compensatory remedy (if they are able to demonstrate that they have suffered actual damage) or injunctive relief (when needed to protect their own business activity).

Furthermore, from a policy perspective, the very distinction between restricted and unrestricted websites may be counterproductive, especially in light of the current industrial policy which – at least within the European Union – fosters a wider circulation of (non-personal) data so as to maximize their value in the data-driven economy[22]. On one hand, such an approach incentivizes the "enclosure" of online data, by stimulating the managers to impose on their visitors simple authentication processes as a way to legitimately exclude undesired extraction of data by third parties. On the other hand, this perspective can result in the locking of potential data value since it would reserve to the website manager the power to decide whether to make such data available on request of other economic actors, even when they are not in competition with them. Should this line of reasoning be carried forward, it would follow, for example, that the practice of gathering from the Internet non-personal and non IP-eligible data in order to train (generative) AI systems is illegal, whenever the websites from which they are extracted employed forms of technological barriers.

In light of this, it seems more opportune to envision rules and/or criteria which can stimulate the sharing of online data between the parties, independently of whether the websites are freely accessible or not, provided that the data obtained are not used to unfairly implement products or services in direct competition with the website manager. In this perspective, the recent EU proposal for a *Data Act* concerning the circulation of data generated from IoT[23] may provide technical solutions to be built upon in order to regulate access even to those online data which do not fall under a specific legal regime. Indeed, the Data Act, despite recognizing – at least implicitly – the "de facto" power of control of IoT data by manufacturers, provides for a series of rules which will force them to share raw, non-personal data with other economic actors (the IoT device users, in first instance, and then those with whom the users decide to share the data generated). At the same time, the Regulation prohibits those who have accessed the data generated by the IoT device produced by the manufacturer to develop a product that competes with the one from which the data originate. By relying on a similar balancing mechanism, the access to and extraction of online non-personal data could be deemed legitimate, despite authentication gates, under the condition that the data are used to produce innovative products or services which do not undermine, in an unfair way, the original business activity of the website manager. Given the importance of web scraping activity in the digital sphere, the determination of its legitimacy should be the result of careful theoretical and policy considerations and not based on specific, individual economic interests.

6 Bibliography

P. L. Bellia, *Defending Cyberproperty*, in "N.Y.U. L. Rev.", 79 (2004): 2164-2273.

M. Borghi, and S. Karapapa, *Contractual Restrictions on Lawful Use of Information: Sole-Source Databases Protected by the Back Door?* in "EIPR", 37 (8), (2015): 505-511.

22. See the European Commission Communication on *A European strategy for data*, Brussels, 19.2.2020, COM(2020) 66 final.

23. See *Proposal for a Regulation on harmonised rules on fair access to and use of data (Data Act)*, Brussels, 23.2.2022, 68 final, 2022/0047 (COD).

M. Bottis, *How Open Data Become Proprietary in the Court of Justice of the European Union*, in S. K. Katsikas and A. B. Sideridis (eds.), *E-Democracy. Citizen Rights in the World of the New Computing Paradigm*, Springer, Switzerland, 2015. 169-174.

R. L. Dunne, *Deterring Unauthorized Access to Computer: Controlling Behaviour in Cyberspace through a Contract Law Paradigm*", in "Jurimetrics J.", 35 1, (1994): 1-15.

T. Feldmann, and J. Heidrich, *Rechtsfragen des Ausschlusses von Usern aus Internetforen*, in "CR" (2006): 406-412.

I. Gupta, and V. H. Devaiah. *The Database Directive «Contracting Out» Bar: Does It Apply to Unprotected Databases?* In "J. Intell. Prop. L. & Pract.", 10.9 (2015): 669-672.

J. K. Hirschey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, in "Berkeley Tech. L.J.", 29 (2014): 897-928.

P. H. Klickermann, *Virtuelle Welten ohne Rechtsansprüche*, in "MMR", (2007): 766-769.

K-H. Ladeur, *Ausschluss von Teilnehmern aus Diskussionsforen im Internet: Absicherung von Kommunikationsfreiheit durch „netzwerkgerechtes*" *Privatrecht*, in "MMR", (2001): 787 ss.

M. A. Lemley, *Place and Cyberspace*, in "Cal. L. Rev.", 91 (2003): 521-542.

M. Lemley, *Shrinkwraps in Cyberspace*, in "Jurimetrics J.", 35. (1995): 311-323.

M. A. Lemley, *Terms of Use*, in "Minnesota L. Rev.", 91 (2006): 459-483.

P. Maume, *Bestehen und Grenzen des virtuellen Hausrechts*, in "MMR", (2007): 620-625.

G. Piras, *Virtuelles Hausrecht?* Mohr Siebeck, Tübingen, 2016.

L. Quilter, *The Continuing Expansion of Cyberspace Trespass to Chattels*, in "Berkeley Tech. L. J.", 17, 2002: 421-443.