https://doi.org/10.6092/issn.1825-1927/18876 ISSN 1825-1927

Data scraping in criminal law: the possible contribution in preserving the value of data analytics in the modern age

Rosa Maria Vadalà*

Abstract. Criminal investigations increasingly rely on data analytics, including web scraping, which may infringe upon privacy rights. The potential criminal implications of data scraping may intersect with civil disputes related to copyright law violations, breaches of terms of service agreements, and security protocol violations. The challenge for criminal law lies in allowing data scraping while simultaneously preventing abuses that arise from the unacceptable compromise or illegitimate sacrifice of other interests resulting from these technologies. Specifically, this analysis will scrutinize how criminal law tackles the aforementioned challenge, contemplating data scraping as both a tool for criminal investigation and a phenomenon subject to selective criminalization in accordance with the principle of extrema ratio.

Keywords: data analytics, web scraping, copyright offences, data property rights, unauthorized access.

1 Introduction to the "grey area"

We find ourselves amidst a "data tsunami"[1], wherein data repositories encompass a wide spectrum of knowledge, including information susceptible to criminal investigation[2].

Certainly, the cornerstone technologies underpinning the data analytics and criminal analytics framework are data capture and data mining technologies[3]. In this context, the practice of data or web scraping has emerged prominently as an automated means of acquiring online data, with applications ranging from market competition to identifying human trafficking operations[4] to monitoring illicit drug markets[5]. Moreover, web scraping plays a pivotal role in retrieving the open data requisite for training artificial intelligence systems[6], as well as fostering the development of technologies grounded in predictive models for threat detection and forecasting[7].

Specifically, web scraping comprises three distinct phases: website analysis, website crawling, and data organization. The bots, operating as scripts or programs, emulate human web browsing to execute fully automated and repetitive tasks on websites[8]. It is

*University of Verona, Department of Law; ✉ rosamaria.vadala@univr.it

1. Xiaoling , Shu. *Knowledge Discovery in the Social Sciences: A Data Mining Approach.* University of California Press, 2020, p. 3.
2. Oatley, Giles C. "Themes in data mining, big data, and crime analytics". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12.2 (2022.), p. 2.
3. Sicurella, Rosaria, Scalia, Valeria. "Data Mining and Profiling in the Area of Freedom, Security and Justice: State of Play and New Challenges in the Balance between Security and Fundamental Rights Protection". *New Journal of European Criminal Law*, 4(2013), p. 409-460.
4. McAlister, Ruth. "Webscraping as an Investigation Tool to Identify Potential Human Trafficking Operations in Romania". *Ulster University's Research Portal* (2015), p. 2.
5. Maybir, James, Chapman, Brendan. "Web scraping of ecstasy user reports as a novel tool for detecting drug market trends". *Forensic Science International: Digital Investigation*, 37-301172 (2021), p. 2.
6. Monterossi, Micheal W. "Estrazione e (ri)utilizzo di informazioni digitali all'interno della rete internet. Il fenomeno del c.d. web scraping". *Il diritto dell'informazione e dell'informatica*, 2 (2020), p. 332.
7. Tymchyshyn, Andriy, et al. "The use of big data and data mining in the investigation of criminal offences". *Amazonia Investiga*, 11. 56(2022), p. 278-290.
8. Wang, Yuguang, et al. "Review of data scraping and data mining research". *Journal of Physics: Conference Series* 1982, 012161(2021), p. 2.

pertinent to note that bots are "dual-use software"[9], as they can be employed for both constructive and malevolent purposes. The application of web scraping as a tool for criminal investigations, as well as its utilization in journalism[10], academic research[11], and legitimate commercial purposes[12], stands as a testament to its valuable functionalities. Conversely, malicious bots remain a primary source of security vulnerabilities, often targeting websites and contributing to security breaches[13].

Indeed, the legal status of Web Scraping is defined a "grey area"[14]: some web scraping activities, even for legitimate purposes, may potentially violate fundamental rights or lead to criminal liability. In the first sense, web scraping regards specifically personal data and is a part of a general discussion on the limits posed by the rights to privacy and the rights to data protection[15] in cyberspace as a "public" data reserve[16]. In the second sense, the criminal implications are intertwined with complex civil disputes on copyright law and violation of terms of service agreements and security protocols[17].

In both scenarios it is imperative to strike a balance between the commercial interests of private platforms and the privacy rights of consumers and users, while also considering other values such as the promotion of competition, safety, and health[18]. In a broader context, this equilibrium aims to facilitate efficient access and improvement of knowledge extracted from the web[19].

2 Definition of the analysis perspective

The paper aims to outline how criminal law can contribute to achieving this balance while preserving the intrinsic value of knowledge. In the opinion of the author this value is intricately linked to the harnessing of data analytics and criminal analytics technologies. The challenge for criminal law is to simultaneously prevent the abuses stemming from the unacceptable compromise or illegitimate sacrifice of other interests brought about by these technologies.

In this context, the analysis must be guided by two considerations: the absence of 'tyrannical'[20] interests, particularly those with economic significance, and the justification, in specific cases, for 'intrusion' into fundamental rights within the domain of information technology.

In particular, the manner in which criminal law addresses the aforementioned challenge will be examined, considering data scraping both as a tool for criminal investigation and as a phenomenon for selectively criminalizing in accordance with the principle of extrema ratio[21]. In the first sense, concerning publicly available personal information, the admissible privacy interference passes through the clearly defined and proportionate conditions derived from the rights to data protection[22]. For this reason, the third

9. Cfr. Albrecht, Michael. *Die Kriminalisierung von Dual-Use-Software*, 2014, p. 18

10. Diouf, Rabiyatou, et al. "Web Scraping: State-of-the-Art and Areas of Application". *IEEE International Conference on Big Data*, 2019, p. 6041.

11. Luscombe, Alex, Dick, Kevin, Walby, Kevin. "Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences". *Quality & Quantity*, 56 (2022), p. 1024.

12. Khder, Moaiad. "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application". *Int. J. Advance Soft Compu. Appl*, 13.3 (2021), p. 151-154.

13. Rizwan Ur, Rahman, Deepak Singh, Tomar. "A new web forensic framework for bot crime investigation". *Forensic Science International: Digital Investigation*,33-300943 (2020), p. 3.

14. Cfr. Krotov, Vlad, Leigh Redd, John, Leiser, Silvia. "Tutorial: Legality and Ethics of Web Scraping". *Communications of the Association for Information Systems,* 47 (2020), p. 581.

15. This right is established under art. 8 of the EU Charter of Fundamental Rights which sits alongside the right to privacy; in addition, art. 16 of the Treaty on the Functioning of the European Union (TFEU) obliges the EU to lay down data protection rules for the processing of personal data.

16. Edwards, Lilian, Urquhar, Laclhan. "Privacy in Public Spaces: What Expectations of Privacy do we have in Social Media Intelligence?". *International Journal of Law and Information Technology*, 24.3 (2016), p. 310.

17. Ciani Sciolla, Jacopo. *Il pubblico dominio nella società della conoscenza L'interesse generale al libero utilizzo del capitale intellettuale comune*, 2021, p. 200-207.

18. Fan, Mary D. "The Right to Benefit from Big Data as a Public Resource". *New York University Law Review*, 96. 5(2021), p. 1470.

19. Zamora, Amber. "Making Room for Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online". *Journal of Business, Entrepreneurship and the Law*, 12. 1 (2019), p. 227.

20. Recalling the Schmitt's lesson on the tyranny of values, the expression was used by the Italian Constitutional Court in the controversial sentence on ILVA case no. 85/2013.

21. This principle posits that criminal law should be the ultimate resort for social control, thereby minimizing the substantial material and human costs it entails.

22. Kokott, Juliane, Sobotta, Christoph. "The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR". *International Data Privacy Law*, 3.4(2013), p. 226.

paragraph considers specifically the level of the condition posed by these rights to the technology-led policing. In the second sense, while it is accurate to assert that data scraping per se is not a crime, its classification as a criminal act is contingent upon the nature of the data scraped and the manner in which the scraper gains access. Considering this dual perspective, the fourth paragraph further explores transgressions related to the protection of copyright law, with a particular focus on databases and breaches involving technical protection measures. (TPMs)[23].

For other unprotected data, the fifth paragraph examines whether and to what extent there is criminal protection for the interest in data availability against illegitimate interference by third parties. In analysing Italian case law concerning unlawful data appropriation, the formulation of a criminal law anti-intrusion regime against data scraping will commence with the admissibility of data as a subject of property crimes. Subsequently, the consideration of the computer crime of unauthorized access will be explored as the primary tool for safeguarding data availability.

3 The data protection standpoint: investigative and substantial confines

The right to data protection presents a challenging perspective in the context of data scraping of publicly available personal data, primarily due to disparities between the approaches adopted by the European Union and the United States. These disparities are fundamentally rooted in the fact that the United States provides federal data protection laws, but they are largely confined to specific fields[24]. In contrast, the EU's data protection framework represents a comprehensive structure, including the General Data Protection Regulation (GDPR) and the Directive no. 2016/680 pertaining to the safeguarding of personal data processed for law enforcement purposes (Law Enforcement Directive or LED).

These differences give rise to significant consequences in a private context, where federal courts in the United States have typically excluded the relative reasonable expectation of privacy or weighed it against competing interests[25]. Indeed, the most noteworthy disparities pertain to law enforcement activities. In this regard, the Clearview case serves as an illustrative example.

Clearview is a U.S. startup that clandestinely scraped photos from various social networks, including Facebook, Twitter, and LinkedIn, subsequently leveraging individuals' data to construct a facial recognition tool that it marketed to law enforcement agencies and other entities[26]. Within the European Union context, this company received formal notifications demanding the cessation of its "unlawful processing" of facial images, which contravened the provisions of Europe's GDPR[27].

Specifically, the Italian Data Protection Authority contended that the mere public accessibility of these images does not suffice to justify data scraping. It further asserted that data subjects cannot reasonably expect their images to be utilized for facial recognition purposes, especially by a non-EU-based private entity, the existence and activities of which remain largely unknown to most data subjects[28]. It is imperative to underline that the right to data protection serves specifically to address information disparities between data subjects and those responsible for processing their personal data[29]. Consistent with this public value, the right to data protection is safeguarded in Italy through criminal sanctions, also, reformed by Legislative Decree No. 101/2018 and applicable solely in situations characterized by a heightened risk to data subjects with collective significance[30]. Indeed, these crimes are not at the centre of the debate posed by the Clearview case in the Italian system, either. In general, in the European context, this case pertained, instead, precisely to the definition of the limitations of data scraping as a tool of law enforcement.

In particular, the European Data Protection Supervisor (EDPS) clarified that Article 17(2) of the Europol Regulation no. 2016/794 allows Europol to procure personal data from publicly available sources, even when an intermediary is involved, provided that the intermediary strictly operates as a data processor. In the case of Clearview AI, its collection of these images extends beyond

23. About TPMs and RMI, usually indicated together as Digital Right Management (DRM), see Moscon, Valentina. "Misure tecnologiche di protezione (Diritto d'autore)". *Digesto on line* (2013). Baldwin, Peter. *The Copyright Wars. Three Centuries of Trans-Atlantic Battle*. Princeton University Press, 2014, p. 280 ss.

24. Xiao, Geoffrey. "Bad Bot: Regulating the Scraping of Public Personal Information". *Harvard Journal of Law & Technology,* 34.2 (2021), p. 703.

25. In this sense the Opinion by Judge Berzon, US court of appeals for the ninth circuit, HIQ LABS, INC. v. LINKEDIN CORPORATION, 18 April 2022.

26. Dul, Camilla. "Facial Recognition Technology vs Privacy: The Case of Clearview AI". *Queen Mary Law Journal*, 3 (2022), p. 3-4.

27. In this sense, e.g., the Hellenic DPA decision no. 35/2022 and the France DPA Decision no. MED 2021-134.

28. The Italian DPA decision n. 9751362/2022.

29. Lynskey, Orla. "Deconstructing Data Protection; the"added-value" of a right to data protection in the EU legal order". *International and Comparative Law Quarterly*, 63(2014), p. 595.

30. Manes, Vittorio, Mazzacuva, Francesco. "GDPR e nuove disposizioni penali del Codice privacy". *Dir. pen. e proc.*, 2(2019), p. 174.

making them available to Europol, as it uses them to offer its proprietary facial matching algorithm to law enforcement agencies worldwide[31].

The aforementioned Article 17 bears resemblance to Article 4(2) of the LED Directive, which regulates situations involving law enforcement's access to personal data originally generated by private entities for non-law enforcement purposes. It draws a distinction between the initial purpose of collection, regulated by the GDPR, and the purpose of further processing, overseen by the LED Directive. Consequently, if the initial data processing for non-law enforcement purposes does not align with GDPR requirements, the law enforcement agencies may be precluded from utilizing said data[32].

Accordingly, this connection between GDPR and the LED Directive implies that data scraping, whether used directly or indirectly by law enforcement agencies, falls under the ambit of data protection regulations. Nevertheless, a significant issue arises from the absence of a formal connection between the initial (private) and subsequent (law enforcement) objectives of data processing.[33]

As emphasized in recent guidelines by the European Data Protection Board (EDPB) regarding the use of facial recognition technology in law enforcement, the overarching objective of improving efficiency in combating serious criminal offenses does not, on its own, constitute a valid justification for the indiscriminate collection and processing of extensive data quantities[34]. Thus, the admissibility of data scraping also hinges on meeting the general criteria outlined in the LED Directive's necessity and proportionality assessment. Considering the aforementioned points, this legislation serves as a fundamental standard for criminal investigations, albeit not exhaustive.

4 The copyright criminal protection

Concerning the second aspect under analysis, the issue of criminalizing data scraping is closely tied to criminal copyright regulation. The primary characteristic of this regulation is its subsidiarity to civil law. It is impermissible to regard criminal law as an independent tool within the arsenal against copyright infringement. This characteristic is evident in the formulation of criminal offenses and in the fact that exemptions for copyright infringements based on fair use within the American legal framework, as well as the exceptions or limitations within the EU legal framework, also extend to cover criminal offenses. While this interdependence is inherently negative, delineating the legality of data scraping becomes more intricate when considering the criminal copyright regulations applicable to databases and the "anti-circumvention" rules. This complexity arises due to the heightened risk in these two sectors of exceeding the boundaries within which copyright protection can be deemed admissible.

Staring to databases regulation, the American legal regime covers an individual who wilfully and knowingly engages in activities that contravene the exclusive rights of the database owner. Such activities may encompass unauthorized reproduction, distribution, or public display of the database without the owner's consent[35]. However, a web scraper typically employs the search functionalities and systematic organization of data found on publicly accessible websites to create a transformative product, rather than attempting to redirect users or produce a product substantially similar to that of the target website for the purpose of gaining a competitive advantage. In such instances, copyright criminal sanctions may not be applicable. According to American legal doctrine, this inapplicability is contingent upon the limited copyright protection afforded to databases, often referred to as "thin" protection[36].

On the contrary, the EU Database Directive introduced a two-tier system for protecting databases: copyright to safeguard the database's structure and the *sui generis* right to prevent unauthorized appropriation, extraction, or reuse of the contents of non-

31. EDPS Opinion on the possibility to use Clearview AI and similar services at Europol, Case 2020-0372.

32. Gottschalk, Thilo. "The Data-Laundromat? Public-Private-Partnerships and Publicly Available Data in the Area of Law Enforcement". *European Data Protection Law Review*,6.1 (2020), p. 37.

33. Jasserand, Catherine. Law enforcement access to personal data originally collected by private parties: Missing data subjects' safeguards in directive 2016/680?. *Computer law & security review,* 34 (2018), p. 164.

34. EDPS Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement, adopted 26 April 2023.

35. Vasiu, Ioana, Vasiu, Lucian. "Criminal Copyright Infringement: Forms, Extent, and Prosecution in the United States". *University of Bologna Law Review, 4.* 2 (2019), p. 244.

36. Geoffrey, Xiao. "Data Misappropriation: A Trade Secret Cause of Action for Data Scraping and a New Paradigm for Database Protection". *Columbia Science and Technology Law Review*, 24. 1 (2022), p. 129.

original databases resulting from substantial investments in their creation and management. The primary risk associated with the second type of protection could involve the 'monopolization' of information[37].

In spite of this, in implementing this directive, the Italian legislator, through Legislative Decree no. 169/1999, established a criminal sanction for the extraction and reutilization of the entire database or a "substantial part" of it, as well as for the repeated and systematic extraction and reutilization of non-substantial portions thereof, which involves activities contrary to the usual database management or results in unjustified harm to the database creator. These actions result from the integration of the criminal offence, established by art. 171 *bis* (2) of the Italian Author's Right (L.d.a), with the civil concepts of Articles 102 *bis* and 102 *ter* L.d.a, expressly referenced[38].

Furthermore, the pursuit of profit does not effectively distinguish punishable conduct, as it is compatible with both non-monetary benefits and, in itself, violations of the *sui generis* right can be attributed to activities conducted within a competitive business context. In this manner, the Italian criminal *sui generis* right protection primarily serves as a punitive measure for extra-criminal dispositions. Commencing with the definition of what constitutes a database, the application of the criminal provision depends on generic concepts that are filled with meaning through decisions of the Court of Justice of the European Union and national courts, which are not always consistent with each other[39].

The conformity of these offenses to the principle of *extrema ratio* is indeed questionable, taking on the concerning characteristics of a largely unapplied symbolic criminal law[40]. In the author's perspective, the primary critical issue lies in the excessive criminal protection granted to the creator of the database at the expense of significant collective interests.

The central importance of striking this balance was recently emphasized by the CJEU in the CV Online Latvia case, which addressed the classification of scraping publicly available databases of meta tags as "extraction" and "re-utilization". The CJEU's affirmation of this classification signifies a notable departure from prior case law, placing more emphasis on balancing the rights of database creators in recouping their investment and the legitimate interests of competitors and users in accessing the information contained in those databases and the potential to create innovative products based on such information[41]. In particular, in applying this balance, the Court finds an infringement of the *sui generis* right only if the data scraping deprives the creator of income intended to enable them to recoup the costs of their investment.

This balance needs to be present in the consideration of legal consequences of technological protection measures, too. Frequently, data-scraping bots that seek access to a website may encounter various technological protection measures (TPMs) designed to deter or prevent automated access. These measures include login requirements, captcha tests, cookies, scripts, and IP blocking[42]. These techniques have been granted legal protection under the World Intellectual Property Organization (WIPO) treaties. In many signatory states, it is now unlawful for users to bypass TPMs, even when doing so is intended to exercise legitimate exemptions provided by copyright law[43]. American criminal law has followed this trend by penalizing the act of circumventing access controls, in addition to prohibiting the trafficking of circumvention devices, regardless of whether copyright infringement results from such circumvention[44].

The primary criticism raised is the concern that the use of technological measures may contravene the goal of copyright law, that is not "to grant"control" over works of authorship, but rather to accord certain limited rights over some kinds of exploitations"[45]. This risk of criminalizing a form of abuse of economic power to the detriment of other interests and public values also impacts Italian copyright legislation regards TPMs.

While European directives did not directly require criminal sanctions, this approach was adopted by European member states, including the Italian legal system, as it was deemed the most effective and deterrent measure.

37. Grosheide, Willien. "Database Protection. The European Way". Washington University Journal of Law & Policy 8 (2002), p. 45

38. Musso, Luca. "La tutela penale delle banche dati: un bilancio". *DI,* 5 (2018), p. 417.

39. Sganga, Caterina. "Ventisei anni di direttiva Database alla prova della nuova strategia Europea per i dati: evoluzioni giurisprudenziali e percorsi di riforma". *Il diritto dell'informazione e dell'informatica*, 3(2022), p. 684.

40. Haber, Eldar. "The Criminal Copyright Gap". *Stanford Technology Law Review*, 18. 2 (2015), p. 287.

41. CJEU, 3 June 2021, CV-Online Latvia C-762/19, para. 42.

42. Robinson, Pamela, Scassa, Teresa. *The future of open data*. University of Ottawa Press, 2022, p. 89.

43. Jackson, Matt. "Using Technology to Circumvent the Law: The DMCA's Push to Privatize Copyright". *Hastings Comm. & Ent. L.J.*, 23.3(2001) p. 611.

44. Cohen, Loren, O'Rourke, Okedeiji, *Copyright in a global information economy*, 2015, p. 866.

45. Ginsburg, Jane. "Copyright and Control Over New Technologies of Dissemination". 101 COLUM. L. REV., 1613 (2001), p. 1616.

More precisely, the Italian L.d.a does not attribute criminal significance to the actual utilization of technologies for bypassing protective measures. Instead, it penalizes preparatory actions that acquire criminal relevance in two cases. The first is when the action is characterized by the exclusive or predominant intent or commercial use related to circumventing effective technological measures. The second concerns the acts that regard devices primarily designed, produced, or adapted to enable or facilitate such circumvention. Furthermore, the presence of an infringement or a concrete risk of infringing the rights of the author or creator is not a requisite for anti-circumvention offences[46]. Nonetheless, the severity of the penalty is on par with the offences involving such circumstances.

This notably vigorous enforcement strategy is linked to the absence of objective, qualitative, or quantitative legal criteria that would allow the required intent to enable or facilitate the circumvention of effective technological measures to be proven as the prevailing purpose or connotation of the related devices[47].

In addition to the criminal circumvention TPMs, the Italian legislation establishes also the civil and administrative measures for the same conduct. However, the possibility to combine these measures with excessive and undetermined criminal protection generates a risk of overcriminalization which is totally incompatible with the *extrema ratio* principle.

5 The data ownership and criminal law anti-intrusion regime against scraping

In general, web scraping shows how the access to data is a decisive economic factor, independently from the content of the data itself. From a criminal standpoint, this aspect generates a debate on the possibility of considering data as a subject of classical property offences. In response to this issue, legal scholars have considered the challenges associated with identifying the typical elements of property crimes, such as acquisition and the intention to permanently deprive, in cases of "data theft/appropriation"[48].

In particular, the data theft was defined a "non-zero sum theft" because it consists of copying data while the victim retains disposal of his or her property albeit in diminished capacity for the loss of exclusivity[49]. The situation is substantially similar in the case of the misappropriation offence. Notwithstanding the foregoing, in a recent case, the Italian Supreme Court ruled that the act of removal from a company personal computer, entrusted for work purposes, of the computer data stored there, subsequently deleting the same data, and returning the formatted computer, constitutes misappropriation.

The court ruling, which is also relevant to the criminal qualification of data scraping, is the data's suitability for being subject to misappropriation due to its ease of transferability: this ability consists in the fact that a file can be transferred from one system or device to another at significant distances, or can be "stored" in "virtual" environments (corresponding to physical locations where computers store and process computer data)[50]. However, this argument falls short in both the context of the rendered judgment and, more significantly, in the pertinent issue at hand because it does not resolve the incongruity between the inherent multi-shareability of computer data and the specific moment of usurpation that characterizes property offences[51]. In the case of data theft/appropriation, only data copying occurs which does not deprive the owner of their original data[52]. In fact, the Italian court attempts to address this absence of deprivation by concentrating on the fact that the offender deleted the original data. But the act of deletion is a distinct offence, unrelated to the typical deprivative effect of property crimes.

The preceding considerations illustrate how the data acquisition firstly compromises the data availability, which does not align with property rights. In a systematic approach, this conclusion is supported by the European data strategy, which eschews the notion of creating a new type of data ownership right, opting instead to endorse the principles of access and data sharing for the betterment of the public good and technological optimization[53].

46. Flor, Roberto. "Autore (Diritto di) (diritto penale)". *ED Annali*, X (2017) p. 130.

47. Flor, Roberto. "Misure tecnologiche di protezione e tutela penale dei diritti d'autore: l'esperienza applicativa italiana". *Dir. pen. e proc.,* 8 (2011), p. 1010.

48. Clough, Jonathan. "Data theft? Cybercrime and the increasing criminalization of access to data". *Criminal Law Forum*, 22(2011), p. 151-152.

49. Black, Stephen."Cyberdamages". *Santa Clara High Technology Law Journal*, 36. 2 (2020), p. 5.

50. Così Cass. pen., Sez. II, 10.04.2020 n. 11959 in *Diritto penale e processo,* 6/2020, p. 749 ss..

51. Picotti, Lorenzo. "*Reati Informatici". Enc. giur.,* Agg., VIII (2000), p. 2.

52. Kerr, Orin. "Cybercrime's Scope: Interpreting"Access' and 'Authorization' in Computer Misuse Statutes". *New York University Law Review 1596*, (2003), p.1610.

53. Lillà Montagnani, Maria, von Appen, Antonia. "Intellectual property and data ownership in the European strategy for data". *Law, Regulation and Governance in the Information Soci*ety, 2023, p. 333.

In a similar vein, the relationship between data and the act of access must be evaluated in the matter under examination to establish a criminal anti-intrusion framework against data scraping for common content. Specifically, centering the criminal intervention on the moment of data access is justified by the need to prevent the breach of various interests expressed by the contents of data and its subsequent use. Linking the criminal assessment to the existence of harm resulting from access could lead to delayed protection due to the swift pace of the information flow.

Consequently, a pivotal component of this framework is the computer crime of unauthorized access, the precise definition of which assumes paramount importance in distinguishing lawful from unlawful activities within the realm of cyberspace[54]. Other offences within this framework may include crimes, for example the misuse of devices, which penalize preparatory acts aimed at illicit purposes from the outset. The following analysis will cover only the crime of unauthorized access, as it directly safeguards the Internet's social value by defining zones of confidentiality and security, free from the intervention of others.

The American as well as the Italian criminal law justice systems recognize this crime. The Computer Fraud and Abuse Act (CCFAA) penalizes the access to a protected computer when done without authorization. Italian criminal law covers both access and the act of remaining in a computer system without consent when the said system is protected by security measures. (615 *ter* c.p.). In the Italian perspective, security measures are only to be understood as the explicit manifestation of the dominus's will to exclude others, but the infringement of these measures is deemed irrelevant to the perfection of the offence[55].

This requisite of the Italian criminal offence determines that this offence does not cover freely accessible websites. Some legal scholars have proposed eliminating this requirement, to resolve the debate on the punishment of insiders that exceed the limits of the legitimate access operated[56]. Analysing the questions raised concerning the applicability of the American unauthorized access crime is beneficial for evaluating the soundness of this approach in defining web scraping with criminal significance.

In particular, the case law concerning the legality of scraping under the CFAA allows us to assert the absence of authorization when a scraper circumvents code-based or contract-based access restrictions[57]. However, there is controversy surrounding whether the absence of authorization also encompasses cases where the website, upon discovering web scraping activities, takes measures to block it or sends a cease-and-desist letter. In a recent case, this thesis was dismissed, with the assertion that the CFAA is best understood as an anti-intrusion statute and not as a misappropriation statute[58]. This consideration highlights that contractual or technical bans or blocks adopted in reaction to data scraping are often mere specific use restrictions on the information, which are distinct from the preventive access limitations[59].

In light of the above, it is important to recall, in accordance with the explanatory guidelines of the criminalization obligations outlined in the Cybercrime Convention, that the use of specific technical tools, such as bots, may be considered illegal access as defined in Article 2 of this treaty. However, this eventuality does not suffice to categorize the use of these tools as "without right"[60].

To clarify ambiguously worded provisions, the identification of criminally pertinent accesses to third-party computer environments must involve referencing a condition in the crime's description even before addressing any issue regarding the existence and terms of access. In particular, this element can be constituted by the presence of the security measures here discussed. The issue of the legitimacy of the insider's actions can be better resolved by specifying that the access contravenes the conditions regulating authorization methods and contents, rather than by eliminating only the requirement capable of externalizing the illegality of the conduct.

54. Bussolati, Nicola. *Harmonisation of cybercrime law: Past solutions, present tensions, and future challenges*, 2020, p. 61.

55. Flor, Roberto. "Art. 615 ter c.p.: natura e funzioni delle misure di sicurezza, consumazione del reato e bene giuridico protetto". *Dir. pen. e proc.*, 1(2008), p.109.

56. Picotti, Lorenzo, Flor, Roberto, Salvadori, Ivan. "Proposta di riforma dei Reati contro l'inviolabilità del domicilio, la tutela della vita privata e dei segreti, la libertà e la personalità informatica". *www.aipdp.it* (2021), p. 7.

57. Macapinlac, Tess. "The Legality of Web Scraping: A Proposal". *Federal Communications Law Journal,* 71. 3 (2019), p. 411.

58. Addicks, Madison. "Van Buren v. United States: The Supreme Court's Ruling on the Fate of Web Scraping -"Access" to Discovery or Detention?". *Tulane Journal of Technology and Intellectual Property*, 24 (2022), p. 172.

59. Sellars, Andrew. "Twenty Years of Web Scraping and the Computer Fraud and Abuse Act". *Boston University Journal of Science and Technology Law*, 24.2(2018), p. 401.

60. Council of Europe. *Explanatory Report to the Convention on Cybercrime Budapest* (2001), p. 9.

Indeed, the implementation of protective measures demonstrates externally the website owner's right to exclude or restrict third parties' access and can be clearly perceived as such by the scraper[61].These measures, selected in response to various encountered risk factors, serve as guarantees of certainty and the integrity of digital relationships and associated communication channels, irrespective of the nature of the data involved. The protection thus provided actually goes beyond safeguarding the undisturbed enjoyment of one's virtual space, demonstrating a direct functional connection with cybersecurity. This legal good, in its double components of information security and Computer Security, expresses an essential need in the information society[62].

6 Conclusion: possible limits to tyrannical rights

The central issue under examination concerns the "criminal solutions" in the balance of conflicting interests generated by data scraping. For the investigation activity the LED directive doesn't explicitly address the data scraping; in general, it leaves numerous blind spots for Big Data processing practices that are pertinent to criminal investigations and potentially more detrimental to individuals[63]. There is a need for law enforcement not to abuse technology: *dataveillance*[64]cannot be admitted because it constitutes not merely an 'intrusion,' but a definitive privacy compromise.

Regarding the punishment of data scraping, this perspective is grounded in the belief that maintaining knowledge availability in cyberspace relies on unrestricted access, limiting to a minimum the cases in which accessing someone else's information may result in criminal sanctions.

In this regard, a combination of the American and Italian approaches could suggest that an offence of unauthorized access, through the violation of protective measures, different from mere contractual restrictions, could serve as an anti-intrusion framework against data scraping in line with the principle of *extrema ratio*.

As demonstrated by offenses protecting copyright, the residual area of criminal intervention can be dependent on the specific nature of the data in question.

But the analysis conducted on the criminal copyright regulations of the database and Mtp shows a perilous shift from the criminalization of "individual" fraudulent infringements to the prosecution of broader behaviours related to technological self-control rather than the legal control of authors and creators.

In the realm of copyright offenses pertaining to databases, this matter is directly connected to the ambiguity concerning the fundamental elements of the alleged misconduct. Due to this ambiguity, the infringement of the *sui generis* right simultaneously constitutes an offence, necessitating reliance on decisions made by the CJEU to identify the factors of selectivity. The dependence on civil law leads to criminal law playing a part in unwarrantedly redirecting the emphasis of intellectual property law from non-economic aspects, such as the encouragement of science and creativity, to the economic aspect of safeguarding investments[65].

The analogous illegitimate alteration of the scale of value, with a sacrifice of the public interest to knowledge, also characterizes the field of TPMs, which are often employed to restrict the circulation of works for competitive purposes rather than preventing copyright infringement. Consequently, the connected dependent criminal law protection ends up covering this case of abuse, thus risking to disrupt the equilibrium of interests that copyright law aims to maintain[66].

As a result, another factor contributing to this risk is the circumstance that the same acts, which Italian law punishes as crimes, are also subject to administrative penalties, as well as civil remedies. To prevent economic interests from exerting tyrannical control, it is necessary to reserve criminal sanctions for violations that genuinely pose potentially irreparable harm to the rights derived from copyright. As stated by Patry, Mr. Clark's metaphor – "The Answer to the Machine is in the Machine" – represents

61. Seminara, Sergio. "Note sul reato di accesso abusivo a sistemi informatici o telematici da parte di un pubblico agente (art. 615-ter, c. 2, n. 1, c.p.)". *MediaLaws*, 2 (2018), p. 242.

62. Picotti, Lorenzo. "Cybersecurity: quid novi?". *Diritto di Internet*, 1/2020, p.12.

63. Castro Toledo Francisco, Mirò Llinares, Fernando. "Researching Cybercrime in the European Union: Asking the Right Etihcs Questions". *Researching Cybercrimes, Methodologies, Ethics, and Critical Approaches*, 2021, p. 343; De Hert Paul, Sajfert Juraj. "The fundamental right to personal data protection in criminal investigations and proceedings: framing big data policing through the purpose limitation and data minimisation principles of the directive (EU) 2016/680". Brussels Privacy Hub Working Paper, 31 (2021), p. 12-14.

64. Arrigo B. Bruce, Brian Sellers. *The Pre-Crime Society: Crime, Culture and Control in the Ultramodern Age*, 2021, pp. 201-202.

65. Rungrojtanakul, Chana. "Legal Protection of Sui Generis Databases". Golden Gate University School of Law. Paper 15 (2015), p. 168.

66. Wheatley, Christopher. "Overreaching Technological Means for Protection of Copyright: Identifying the Limits of Copyright in Works in Digital Form in the United States and the United Kingdom". *Washington University Global Studies Law Review*, 7.2 (2008), p. 369.

a reformulation of the centuries-old approach of copyright owners, asserting that the law is the solution to the business problem. The problematic aspect of criminal TPMs is that the law grants a delegation to the owners to set, at their discretion and without government control, digital locks with dual civil and criminal relevance [67].

Also in the digital context, especially the criminal protection must to ensure a balance for the benefit of the community and do not serve as a tool of power or, even worse, control. Otherwise, it functions similar to "errant fragments of a «grenade»(..) such fragments can indiscriminately affect other parties or fail to hit the intended target"[68].

7 Bibliography

M. Addicks, *Van Buren v. United States: The Supreme Court's Ruling on the Fate of Web Scraping - "Access" to Discovery or Detention?* in "Tulane Journal of Technology and Intellectual Property", 24, 2022, pp. 161-180.

M. Albrecht, *Die Kriminalisierung von Dual-Use-Software*, Duncker & Humblot, 2014.

B. Arrigo, and B. Sellers (Eds.), *The pre-crime society: Crime, culture and control in the ultramodern age*, Policy Press, 2021.

P. Baldwin, *The Copyright Wars. Three Centuries of Trans-Atlantic Battle,* Princeton University Press, 2014.

S. Black, *Cyberdamages*, in "Santa Clara High Technology Law Journal", 36. 2, 2020, pp. 1-35.

N. Bussolati, *Harmonisation of cybercrime law: Past solutions, present tensions, and future challenges*, 2020.

F. Castro Toledo, F. Mirò Llinares, *Researching Cybercrime in the European Union: Asking the Right Etihcs Questions*, in A. Lavogna and T. J. Holt (Eds.) *Researching Cybercrimes, Methodologies, Ethics, and Critical Approaches*, Palgrave Macmillan, 2021.

J. Ciani Sciolla, *Il pubblico dominio nella società della conoscenza. L'interesse generale al libero utilizzo del capitale intellettuale comune*, Digitalica, Giappichelli, 2021.

J. Clough, *Data theft? Cybercrime and the increasing criminalization of access to data*, in "Criminal Law Forum", 22, 2011, pp. 145–170.

J. E. Cohen, et al., *Copyright in a global information economy*, Aspen Publishing, 2015.

P. De Hert, J. Sajfert, *The fundamental right to personal data protection in criminal investigations and proceedings: framing big data policing through the purpose limitation and data minimisation principles of the directive (EU) 2016/680*, Brussels Privacy Hub Working Paper, 31 (2021).

R. Diouf, et al., *Web Scraping: State-of-the-Art and Areas of Application*, in *IEEE International Conference on Big Data*, 2019, pp. 6040-6042.

C. Dul, *Facial Recognition Technology vs Privacy: The Case of Clearview AI*, in "Queen Mary Law Journal", 3, 2022, pp. 1-24.

L. Edwards, L. Urquhar, *Privacy in Public Spaces: What Expectations of Privacy do we have in Social Media Intelligence?* in "International Journal of Law and Information Technology", 24.3, 2016, pp. 1-29.

M. D. Fan, *The Right to Benefit from Big Data as a Public Resource*, in "New York University Law Review", 96. 5, 2021, pp. 1438-1491.

R. Flor, *Art. 615 ter c.p.: natura e funzioni delle misure di sicurezza, consumazione del reato e bene giuridico protetto,* in "Dir. pen. e proc.", 1, 2008, pp. 106-112.

R. Flor, *Misure tecnologiche di protezione e tutela penale dei diritti d'autore: l'esperienza applicativa italiana*, in "Dir. pen. e proc.", 8, 2011, pp. 1003-1013.

R. Flor, *Autore (Diritto di) (diritto penale)*, in ED Annali X, 2017, pp. 111-135.

67. Patry, William. *How to fix copyright*, Oxford University press, 2011, p. 238.

68. Terracina, Davide. *La tutela penale del diritto d'autore e dei diritti connessi,* 2006, p. 36.

G. Geoffrey, *Data Misappropriation: A Trade Secret Cause of Action for Data Scraping and a New Paradigm for Database Protection*, in "Columbia Science and Technology Law Review", 24.1, 2022, pp. 125-172.

J. Ginsburg, *Copyright and Control Over New Technologies of Dissemination*, in "101 COLUM. L. REV.", 1613, 2001, p. 1613-1647.

W. Grosheide, *Database Protection. The European Way*, in "Washington University Journal of Law & Policy" 8 (2002), p. 39-74.

T. Gottschalk, *The Data-Laundromat? Public-Private-Partnerships and Publicly Available Data in the Area of Law Enforcement*, in "European Data Protection Law Review",6.1, 2020, pp. 21-40.

E. Haber, *The Criminal Copyright Gap*, in "Stanford Technology Law Review", 18. 2, 2015, pp. 247-288.

M. Jackson, *Using Technology to Circumvent the Law: The DMCA's Push to Privatize Copyright*, in "Hastings Comm. & Ent. L.J.", 23.3, 2001, pp. 607-645.

C. Jasserand, *Law enforcement access to personal data originally collected by private parties: Missing data subjects' safeguards in directive 2016/680?* in "Computer law & security review", 34, 2018, pp. 154–165.

O. S. Kerr, *Cybercrime's Scope: Interpreting "Access" and "Authorization" in Computer Misuse Statutes*, in "New York University Law Review" *1596*, 2003, pp.1597-1667.

M. Khder, *Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Applicatio*n, "Int. J. Advance Soft Compu. Appl.", 13.3, 2021, pp. 145-168.

S. Koçer, *An assessment of the implementation of the data protection regulations in criminal proceedings*, in "Law & Justice Review", 26, 2023, pp. 1-14.

V. Krotov, et al. *Tutorial: Legality and Ethics of Web Scraping*, in *Communications of the Association for Information Systems,* 47, 2020, pp. 555-581.

M. Lillà Montagnani, A. von Appen, *Intellectual property and data ownership in the European strategy for data*, in *Law, Regulation and Governance in the Information Society*, Routledge, 2023.

O. Lynskey, *Deconstructing Data Protection; the "added-value" of a right to data protection in the EU legal order*, in "International and Comparative Law Quarterly", 63, 2014, pp. 569-597.

A. Luscombe, et al., *Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences*, in "Quality & Quantity", 56, 2022, pp. 1023-1044.

V. Manes, F. Mazzacuva, *GDPR e nuove disposizioni penali del Codice privacy*, in "Dir. pen. e proc.", 2, 2019, pp. 171-179.

T. Macapinlac, *The Legality of Web Scraping: A Proposal*, in "Federal Communications Law Journal", 71. 3, 2019, pp. 399-422.

J. Maybir, B. Chapman, *Web scraping of ecstasy user reports as a novel tool for detecting drug market trends,* in "Forensic Science International: Digital Investigation", 37, 2021, pp. 1-11.

R. McAlister, *Webscraping as an Investigation Tool to Identify Potential Human Trafficking Operations in Romania*, in Ulster University's Research Portal, 2015, pp.1-3.

M. W. Monterossi, *Estrazione e (ri)utilizzo di informazioni digitali all'interno della rete internet. Il fenomeno del c.d. web scraping*, in "Il diritto dell'informazione e dell'informatica", 2, 2020, pp. 327-369.

V. Moscon, *Misure tecnologiche di protezione (Diritto d'autore)*, in "Digesto on line", 2013, pp. 1-13.

L. Musso, *La tutela penale delle banche dati: un bilancio*, in "DI", 5, 2018, pp. 415-420.

G. C. Oatley, *Themes in data mining, big data, and crime analytics,* in *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(2), 2022, pp. 1-19.

W. Patry, *How to fix copyright*, Oxford University Press, 2011.

L. Picotti, *Internet e diritto penale: il quadro attuale alla luce dell'armonizzazione internazionale*, in "Diritto dell'Internet", 2, 2005, pp. 189-204.

L. Picotti, *Reati Informatici,* in *Enc. giur.,* Agg., VIII, 2000, pp. 1-36.

L. Picotti, *Cybersecurity: quid novi?* in "Diritto di Internet", 1, 2020, pp. 11-14.

L. Picotti, et al., *Proposta di riforma dei Reati contro l'inviolabilità del domicilio, la tutela della vita privata e dei segreti, la libertà e la personalità informatica*, in *www.aipdp.it*, 2021, pp. 1-27.

R. Rizwan Ur, T. Deepak Singh, *A new web forensic framework for bot crime investigation*, in *Forensic Science International: Digital Investigation*,33, 2020, pp. 1-12.

P. Robinson, T. Scassa, *The future of open data*, University of Ottawa Press, 2022.

C. Rungrojtanakul, *Legal Protection of Sui Generis Databases*, Golden Gate University School of Law. Paper 15 (2015).

A. Sellars, *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act*, in "Boston University Journal of Science and Technology Law", 24.2, 2018, pp. 372-415.

S. Seminara, *Note sul reato di accesso abusivo a sistemi informatici o telematici da parte di un pubblico agente (art. 615-ter, c. 2, n. 1, c.p.)*, in "MediaLaws", 2, 2018, pp. 235-250.

C. Sganga, *Ventisei anni di direttiva Database alla prova della nuova strategia Europea per i dati: evoluzioni giurisprudenziali e percorsi di riforma*, in "Il diritto dell'informazione e dell'informatica", 3, 2022, pp. 651-704.

R. Sicurella, V. Scalia, *Data Mining and Profiling in the Area of Freedom, Security and Justice: State of Play and New Challenges in the Balance between Security and Fundamental Rights Protection*, in "New Journal of European Criminal Law", 4, 2013, pp. 409-460.

D. Terracina, *La tutela penale del diritto d'autore e dei diritti connessi,* Giappichelli, 2006.

A. Tymchyshyn, et al., *The use of big data and data mining in the investigation of criminal offences*, in "Amazonia Investiga", 11. 56, 2022, pp. 278-290.

G. Xiao, *Bad Bot: Regulating the Scraping of Public Personal Information*, in "Harvard Journal of Law & Technology", 34.2, 2021, pp. 701-731.

S. Xiaoling, *Knowledge Discovery in the Social Sciences: A Data Mining Approach*, University of California Press, 2020.

Y. Wang, et al., *Review of data scraping and data mining research*, in "Journal of Physics: Conference Series" 1982, 2021, pp. 1-6.

C. T. Wheatley, *Overreaching Technological Means for Protection of Copyright: Identifying the Limits of Copyright in Works in Digital Form in the United States and the United Kingdom*, in "Washington University Global Studies Law Review", 7.2, 2008, pp. 353-372.

I. Vasiu, L. Vasiu, *Criminal Copyright Infringement: Forms, Extent, and Prosecution in the United States*, in "University of Bologna Law Review", *4.* 2, 2019, pp. 229-260.