NEWSLETTER

スクレイピングに関する法的論点及び各国動向

ヨーロッパ & データ保護ニューズレター　　2025 年 8 月 15 日号

執筆者:
石川 智也
n.ishikawa@nishimura.com
水井 大
d.mizui@nishimura.com
山本 希望
no.yamamoto@nishimura.com

1. ウェブスクレイピングの概要

ウェブスクレイピング（web scraping）とは、ウェブクローラーを用いて、ウェブページの HTML データ等を取得して、取得したデータの中から特定のトピックにかかわるデータを抽出、整形し直すことをいう（「スクレイピング」や「データスクレイピング」と呼称することもあるが、同義である。）。例えば、AI の学習用データを収集するにあたっても、ウェブスクレイピングが用いられている。

一方、ウェブスクレイピングをめぐっては、下記の海外動向でみるとおり、昨今ではガイドラインや共同宣言のレベルで、ウェブスクレイピングをされる側での保護措置が求められている。また、General Data Protection Regulation（以下「**GDPR**」という。）[1]との関係から、個人データをウェブスクレイピングにより取得する行為についてデータの管理者又は処理者に対し GDPR が適用されると考えられる。さらに、2024 年 8 月 1 日には、生成 AI を含む包括的な AI の規制である EU AI Act（以下「**AIA**」という。）が発効し、人の生命や基本的人権に対し直接的に脅威をもたらすと考えられる「許容できないリスク」（Unacceptable Risk）が生じる AI は 2025 年 2 月 2 日より禁止されているが、ウェブスクレイピングによって顔認識データベースを作成又は拡張する AI はそのひとつに挙げられている。

本稿は主に AI とウェブスクレイピングをめぐる我が国及び海外における規律・動向を概観する。

2. ウェブスクレイピングと我が国における規律

日本では、ウェブスクレイピングを直接規律する法令等は存在しない。ウェブスクレイピングの対象となった企業等の利用規約でウェブスクレイピングを禁止している場合には、その違反等を理由に民事上の損害賠償、差止め、利用停止や解除をすることが考えられる。また、収集した著作物を含むデータを著作権者の許諾なしに公開、保存、譲渡、販売等をした場合には著作権法等で処理することもあり得る。さらに、収集したデータに（要配慮）個人情報が含まれている場合に、個人情報の保護に関する法律（以下「**個人情報**

[1] 本稿では、特段の指摘の無い限り「GDPR」とは EU GDPR を指す。後掲 Clearview AI Inc v The Information Commissioner 判決では、GDPR 発効後に英国が EU を離脱したが、英国は 2018 年 EU 離脱法を制定し、GDPR を国内法として維持したため（UK GDPR）、UK Data Protection Act 2018（DPA）の解釈に従って EU GDPR が継続的に適用されるとした上で、EU GDPR と UK GDPR の両方について同時に判断したが（判決文 No.70・71）、本稿では、同判決に関する判示について EU GDPR の判断のみを示す。UK GDPR についての判断は、EU GDPR と変わりがない。

保護法」という。）に抵触しないかも、個別の事案に応じ検討が必要となる[2]。

いわゆる 3 年ごと見直しの一環として行われた 2024 年 4 月 3 日付第 279 回個人情報保護委員会「有識者ヒアリング（AI・医療関係）について」における有識者意見では、3 年ごと見直しに関連し得るひとつの論点として、「スクレイピング、特に顔識別 DB 構築の位置付け」が指摘された[3]。また、同年 6 月 27 日に個人情報保護委員会から発表された「個人情報保護法いわゆる 3 年ごと見直しに係る検討の中間整理」[4]の中にはウェブスクレイピングについての言及は存在しないが、当該中間整理に関する意見募集の結果でウェブスクレイピングに関連する意見が複数見受けられた[5]。

日本の個人情報保護法との関係では、ウェブスクレイピングにより要配慮個人情報を取得してしまう場面において、取得を適法化するための本人同意の取得が難しい点が問題となりやすい。現時点では、統計作成等、特定の個人との対応関係が排斥された一般的・汎用的な分析結果の獲得と利用のみを目的とする個人データ等の取扱いについて、同意を要しない制度設計とする方向での個人情報保護法の改正が検討されている。具体的には、本人の同意なく公開されている要配慮個人情報の取得について、要配慮個人情報が統計情報等の作成又は上記規律に基づく本人同意なき個人データ等の第三者提供にのみ利用されることを担保する観点等から、公開されている要配慮個人情報の取得者における一定の事項（取得者の氏名・名称、行おうとする統計作成等の内容又は上記規律に基づく本人同意なき個人データ等の第三者提供を行う目的である旨等）の公表、取得者における目的外利用及び第三者提供（上記規律に基づく本人同意なき個人データ等の第三者提供を行う目的である場合における当該第三者提供を除く。）の禁止を義務付けた上で認めることが検

---

2 個人情報保護法の観点でみれば、要配慮個人情報（個人情報保護法 2 条 3 項）の取得について原則として事前に本人同意を得る必要があるが（同 20 条 2 項各号）、例えば機械学習データを収集する過程でウェブスクレイピングを行う等の場合、無作為的に大量のデータを収集する過程で、同意取得を要しない例外要件にも該当せず、要配慮個人情報を取得してしまうリスクがある。個人情報保護委員会は、収集する情報に要配慮個人情報が含まれないよう必要な取組を行うこと等所定の場合には、要配慮個人情報の「取得」には該当せず、本人同意を要しない旨の説明がなされている（個人情報保護委員会「OpenAI に対する注意喚起の概要」（2023））。

3 生貝直人「AI と個人情報保護：欧州の状況を中心に」（2024）10 頁。

4 個人情報保護委員会「個人情報保護法いわゆる 3 年ごと見直しに係る検討の中間整理」（2024）。

5 個人情報保護委員会「『個人情報保護法　いわゆる 3 年ごと見直しに係る検討の中間整理』に関する意見募集結果」（2024）。本稿に関連する意見を以下に列挙する。

- ✓ 個人情報保護法においても、みだりにデータ化されない自由と基本的人権の保護を求める意見（No.39-2、81）、ディープフェイクをもたらすスクレイピングや生成 AI 学習のためのアクセスについて、個人情報保護法の観点から規制すべき範囲の情報収集及び検討を求める意見（No.52、89、95、99、103、126、136、104）
- ✓ 事業者や個人によるスクレイピングを拒絶するオプトアウトについて、個人情報保護法制の範囲で検討し、他の法域との連携についての提言を求める意見（No.89）
- ✓ 生体データに関する項目においては、個人情報を含み得るスクレイピングデータで作成したデータセットの公開・提供者にも適用されることの検討を求める意見（No.121）
- ✓ オンライン上で公開されている写真の画像をスクレイピングすることで顔特徴データを抽出し、利用者に検索を可能にさせるサービスは、個人本人が気付かぬところで当該個人を追跡や監視しやすくするため、個人の自由や安全が損なわれる恐れが高いものと考えられることから、個人の権利利益にとって高リスクであり禁止されるべき利用法がいかなるものであるかについても検討し、政令やガイドラインで明確化する必要があるとの意見（No.135）
- ✓ 個人情報保護法でスクレイピングや本人が望まない生体データの収集の規制を求める意見（No.154）

討されている[6]。この点については議論もあり[7]、今後の動向が注目される。

3. ウェブスクレイピングをめぐる EU の規律

(1) GDPR

① GDPR の適用（3 条 2 項 b）

個人データ（識別された自然人又は識別可能な自然人（データ主体）に関する情報（GDPR4 条 1 号）を指す。以下同じ。）の処理が EU 域外で行われる場合であっても、処理が EU 域内で行われるデータ主体の行動の監視に関連する場合には、GDPR が適用される（3 条 2 項 b）[8]。

注目すべき事例として、Clearview AI Inc v The Information Commissioner 判決[9]がある。顔認識テクノロジー企業の Clearview AI が、インターネットから無断で収集した画像に基づく違法なデータベースを作成し、英国の個人への公正かつ透明な方法での情報開示が不十分であったことや、生体認証データに要求される高度データ保護基準を提供していなかったこと等を理由として英国個人情報保護監督機関（The Information Commissioner。以下「**ICO**」という。）から約 12 億円の罰金、英国居住者の個人データの取得・使用停止、英国に居住するデータ主体の個人データの削除を命じる執行通知が発出された[10]。

同事案では、Clearview AI が行った(A)画像に関するデータベースの作成（画像のインデックス化を含む。）、開発及び保守、並びに(B)顧客から顧客が識別しようとしている人物の画像（プローブ画像）を受信し、その画像とデータベースにある画像とを照合し、一致する画像に関する特定の情報を顧客に提供する 2 段階のプロセスが問題とされた。

同判決では、GDPR3 条 2 項 b の適用要件について、①問題とされる行為が個人データの処理であること、②個人データが EU のデータ主体であること、③前記①の処理が EU 域内に拠点のない管理者又は処理者によって行われていること、④前記①の処理について、データ主体の行動が EU 域内で行われる限り EU におけるデータ主体の行動の「監視」に「関連」している必要があると判示した（判決文 No.76）。

その上で、まず、問題とされるサービスで英国内の英国人データ主体の行動が「監視」されていたか否かについて、画像のインデックス化の処理自体は、自動化された数学的作業であるため Clearview AI はデータ主体の行動をそれのみでは監視していないが、データベースの検索結果又は当該検索結果をプローブ画像等の情報と併せて検討することにより、Clearview AI の顧客は特定の時点又は期間にわたる人物の行動に関する情報を把握できる場合があることから、このような情報を取得すること又は取得しようとすることはデータ主体の行動の「監視」に相当するとされた（判決文 No.115～129）。そして①～④の要件充足性について

---

6 個人情報保護委員会「個人情報保護法のいわゆる 3 年ごと見直しについて」（2025）。

7 曽我部真裕司会「【座談会】（座談会）個人情報保護法 3 年ごと見直し論議をめぐって」JILIS レポート（2025.05.28）の論点 3。

8 "Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) - Version 2.1 （GDPR の地理的適用範囲（第 3 条）に関するガイドライン）"も参照。

9 Clearview AI Inc v The Information Commissioner, UKFTT 819 (GRC) (2023).

10 ICO "ICO fines facial recognition database company Clearview AI Inc more than £7.5m and orders UK data to be deleted"（2022）.

は、①については争いがなく認められ（判決文 No.130）、②は個人データが EU のデータ主体であり（判決文 No.131）、③プロセス(A)は Clearview AI が単独のデータ管理者であり、プロセス(B)は Clearview AI と顧客が共同データ管理者であると認定した上で、Clearview AI は英国に拠点を置いておらず、その顧客も英国に拠点を置いていないことから要件が充足されるとされた（判決文 No.132 以下）。そして④英国内のデータ主体の行動の監視に「関連」しているといえるかが問題となり（判決文 No.139 以下）、法令等に「関連」の定義はない [11]と適示した後に、個人データの処理と問題となっている行動の監視との間に関係がなければならないと判示した上で、Clearview AI の顧客はプロセス(A)がなければ監視できず、また、プロセス(B)の目的が Clearview AI の顧客による監視が可能となるようにすることであるため、Clearview AI の行為と顧客による監視との間に関連性が存在するとされた。したがって、本判決は、EU GDPR3 条 2 項 b の 4 つの要件は満たされていると判示した [12]。

② 個人データ処理の適法性根拠（6 条 1 項）[13] [14]

GDPR は、個人データの処理は、6 条 1 項に限定列挙されたいずれかの根拠（個人データ処理の適法性根拠）がある限りで適法とする（6 条 1 項）。

具体的には、(a)データ主体が、一つ又は複数の特定の目的のために自己の個人データの処理に同意を与えた場合、(b)データ主体が当事者となっている契約の履行のために処理が必要な場合、又は契約の締結前のデータ主体の求めに応じて手続を履践するために処理が必要な場合、(c)管理者が従うべき EU 又は EU 加盟国の法令による法的義務を遵守するために処理が必要な場合、(d)データ主体又は他の自然人の重大な利益を保護するために処理が必要な場合、(e)公共の利益又は管理者に与えられた EU 又は EU 加盟国の法令に定められた公的権限の行使のために行われる業務の遂行において処理が必要な場合、又は(f)管理者又は第三者によって追求される正当な利益のために処理が必要な場合（ただし、データ主体の、特に子どもがデータ主体である場合の個人データの保護を求めている基本的権利及び自由が、当該利益に優先する場合を除く。）である。

AI モデルや機械学習モデル等をトレーニングするためにウェブスクレイピングを行う場合、多くは(f)正当

---

11 前文 24 項では「処理行為がデータ主体の行動の監視と考えられうるか否かを判断するためには、自然人のプロファイリングを構成する個人データの処理技術が後に使用される可能性を含め、自然人がインターネット上で追跡されているかどうか、特に、データ主体に関連する判断をするため、又は、データ主体の個人的な嗜好、行動及び傾向を分析又は予測するために追跡されているかを確認しなければならない。」とされている（判決文 No.85）。

12 本判決では、以上の 4 要件を満たすとしつつも（判決文 No.139）、GDPR は外国政府の法執行活動の一環として行われるデータ処理には適用されないところ（2 条 2 項 a）、Clearview AI は刑事法執行又は国家安全保障機能の目的で EU 及び英国以外の法執行機関及びその請負業者にのみにサービスを提供していたという争いのない証拠を考慮し（判決文 No.146）、Clearview AI のデータ処理は GDPR の範囲に該当しないと結論付けた（判決文 No.154 及び 157(c)）。

13 2024 年 10 月 9 日、European Data Protection Board（欧州データ保護会議委員会。以下「**EDPB**」という。）は、”Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR（正当な利益に関するガイドライン）“を採択した。同ガイドラインは、後述の ICO による正当利益テストとほぼ類似のテストによる慎重な評価をするよう管理者に求めている。

14 Taner Kuru “Lawfulness of the mass processing of publicly accessible online data to train large language models” （2024)も参照。

利益が認められるか否かが問題となる[15]。正当利益の考え方については、EU データ保護指令 7 条 f の判断基準についての意見書[16]や ICO による正当利益について説明したウェブページ[17]、ICO が作成した正当利益の判断テストのテンプレート[18]が参考になる（GDPR 前文 47～49 項も参照されたい。）。これらによれば、正当利益とは、管理者又は第三者の商業上の利益、個人の利益又はより広範な社会的利益が含まれ、①目的テスト、②必要性テスト、及び③バランステストの 3 つの観点から評価を行うことになる[19]。

① 目的テストにおいては、モデルの特定の正当な目的及び利益等を証明する必要がある。

② 必要性テストは、処理が目的テストで特定された利益を達成するために必要か否か、すなわち代替手段の存否を問うテストである。なお、ICO によれば、現在多くの AI モデルや機械学習モデル等のトレーニングはウェブスクレイピングを通じて取得された大量のデータを使用してのみ可能とされていることから、必要性テストを満たすと考えられる[20]。

③ バランステストとは、データ主体の個人データの保護を求める基本的権利及び自由と、管理者又は第三者による処理によって追求される正当な利益との比較衡量をするテストである。ウェブスクレイピングによるデータ収集行為は「目に見えない処理」であり、また、ウェブスクレイピングをすることによって収集されたデータに基づいて AI モデルや機械学習モデル等をトレーニングする行為は、「新しい技術の使用、又は既存の技術（AI を含む。）の斬新な応用を伴う処理」といえる。なお、ICO のガイダンスによれば、これらの処理は、いずれもデータ保護影響評価（DPIA）の実施を必要とする「高度のリスクをもたらす可能性」[21]があるデータ処理とみなされている[22]。

ウェブスクレイピングそのもの及びウェブスクレイピング後に AI モデルや機械学習モデル等をトレーニングする行為は、上記のうち③バランステストがとりわけ大きく問題となり、DPIA の実施を含め比較衡量の検討を慎重に行う必要があると考えられる。

---

15 前掲注 13）は「GDPR6 条 1 項 f は、他の法的根拠が適用されないとみなされる稀な状況又は予期しない状況に対する『最後の手段』として扱われるべきではない。また、GDPR6 条 1 項 f が他の法的根拠よりも制約が少ないという認識に基づいて自動的に選択されたり、その使用が不当に拡大されたりすべきではない。」としていることに留意されたい。

16 ARTICLE 29 DATA PROTECTION WORKING PARTY "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC"（2014）.

17 ICO "Guide to the General Data Protection Regulation (GDPR), Legitimate interests".

18 ICO "LIA template"（2019）.

19 ICO "Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models"における"Our analysis"と題する章を参照。

20 同上（ICO）参照。

21 European Commission "Guidelines on Data Protection Impact Assessment (DPIA) (wp248rev.01)（データ保護影響評価（Data Protection Impact Assessment：DPIA）、及び処理が EU 規則 2016/679 の目的に照らして「高度のリスクをもたらす可能性」があるかを決定するためのガイドライン）"（2017）8 頁以下を参照。

22 ICO "Examples of processing 'likely to result in high risk'". なお、ICO"Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models"（2024）でも同様。

そのほか、スクレイピングにより特別な種類の個人データを取得してしまう場合には、本人の明示の同意を取得できず（GDPR9 条 2 項(a)参照。）、かつ、本人が明示的に公開したものとも限らないため（同項(e)）[23]、特別な種類の個人データの処理について厳格な要件を定めた 9 条 2 項の要件を充足することが難しいという問題もある。実務的には、極力特別な種類の個人データを取得しないようにする、また、取得してしまった場合には削除するという方向での検討を行うこと等によりリスクを低減することが目指されるが、残された難しい課題である[24]。

(2) AIA

AIA の総則及び禁止される AI に関する条項は、AIA の発効日（2024 年 8 月 1 日）から 6 か月後、すなわち 2025 年 2 月 2 日に適用が開始された。

AIA では、AI システムが、有害な結果を防ぐために人間によって監視される必要があるとし、リスクレベルごとに異なるルールを設ける（前文 26 項）。そして、①許容できないリスク（unacceptable risk）のある AI システム（禁止される AI（Prohibited AI））は市場への投入、運用開始、利用の禁止（5 条 1 項）、②高リスク（high-risk）のある AI システムは事前・事後の厳格な規制、③透明性が必要なリスクのある AI システムについては透明性の義務（transparency obligations）、④他の分類のいずれにも該当しない AI システムについては、提供者・利用者の職員等の AI リテラシーを確保する措置の実施義務をそれぞれ求めている。

その上で、AIA は、インターネット又は CCTV 映像（防犯カメラ映像等）から無作為にスクレイピングした顔画像を用いて顔認識データベースを構築（作成）又は拡張（データの追加又は変更）する AI システムについて、個人の権利・自由が侵害されるおそれがあるとして「禁止される AI」として列挙している（5 条 1 項 e、前文 43 項）。

4. ウェブスクレイピングをめぐる海外動向

本章では、ウェブスクレイピングに関する主な海外動向を紹介する。

(1) 香港の個人情報保護機関（PCPD）による各国との共同宣言

2023 年 8 月 25 日、香港の個人情報保護機関は、アルゼンチン、オーストラリア、カナダ、コロンビア、ジャージー、メキシコ、モロッコ、ニュージーランド、ノルウェー、スイス、イギリスと合計 12 カ国で、ソーシャルメディア・プラットフォームに対する「データスクレイピング及びプライバシー保護に関する共同宣言」（Joint statement on data scraping and the protection of privacy）を発表した[25]。

---

23 EDPB, Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models （2024）, para 17。

24 EDPB は、上記オピニオンにおいて、特別な種類の個人データのウェブスクレイピングの論点を検討スコープから除外している。

25 The Office of the Privacy Commissioner for Personal Data “Data Scraping on Social Media Raises Concerns The PCPD, together with Other Privacy Protection Authorities, Promulgates Global Privacy Protection Expectations and Principles to Social Media Platforms” （2023）.

当該宣言では、SMC（Social Media Companies。以下「**SMC**」という。）及びその他の Web サイトからの大量のデータスクレイピングによって、標的型サイバー攻撃、ID 詐欺、プロファイリング、政治機関又は情報機関による不正な情報収集、不要なダイレクトマーケティング又はスパム等への利用があり得ることなどのプライバシーリスクに関する重大な懸念があるとされる。また、公にアクセスできる個人情報であっても、ほとんどの法域でデータ保護及びプライバシー保護法の対象となること、ソーシャルメディア企業及び一般にアクセス可能な個人データを保有するウェブサイト運営者は、プラットフォーム上の個人情報を、不法なデータスクレイピングから保護するためのデータ保護及びプライバシー法上の義務を有していること、個人情報を収集する大量のデータスクレイピングは多くの法域で報告義務のあるデータ漏えいに該当し得ることを宣言している。

そこで、推奨事項としてであるものの、SMC 及びその他の Web サイトは、例えば、ウェブスクレイピングによるリスクを軽減するため技術的・手続的な多層的なコントロール（multi-layered technical and procedural controls）を備える必要があると述べられている。なお、この共同宣言では、スクレイパーが遵守すべき規律の言及は含まれていない。

① スクレイピングから保護し、スクレイピングを監視し、スクレイピングに対応するための対策を特定・実施するチーム及び（又は）特定の役割を、組織内において指定する。
② ある一つのアカウントが他のアカウントのプロフィールを訪れる回数を 1 時間又は 1 日あたりに制限するレート制限（Rate limiting）を行い、異常な活動が検出された場合にはアクセスを制限する。
③ 新しいアカウントが急速かつ積極的に他のユーザーを探し始めるかを監視し、異常に高い活動が検出された場合、許容できない利用の兆候である可能性がある。
④ ボット活動（bot activity）のパターンを特定してスクレイパーを検出する手段を講じる。例えば、同一の資格情報（same credentials）で複数の場所からプラットフォームにアクセスする際に、疑わしい IP アドレスのグループが検出されることがある。これが短期間内に発生した場合、疑わしいとみなされる。
⑤ CAPTCHA（**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part/コンピュータと人間を区別する自動公開チューリング・テスト）などを用いてボットを検出し、データスクレイピング活動が認められた IP アドレスをブロックする。
⑥ データスクレイピングが疑われる又は確認された場合には、スクレイピング情報の削除を要求し、その削除の確約を取得するため差止命令通知（cease and desist letters）を送付するなど、適切な法的措置を講じる。また、データスクレイピングを禁止する利用規約の執行を行うための他の法的措置を講じる。
⑦ データスクレイピングがデータ漏えいとみなされる法域では、被害を受けた個人及びプライバシー規制当局に通知する。

その後、2024 年 10 月 29 日には、ガーンジー、スペイン、モナコ、イスラエルを加えた 16 か国で、「データスクレイピングとプライバシー保護に関する共同声明の結論」（Concluding joint statement on data scraping and the protection of privacy）を公表した[26]。その結論、声明（Concluding Statement）では、初回の共同声明の発表後における SMC 及びその他の業界関係者との協議に基づく追加の重要なポイン

[26] The Office of the Privacy Commissioner for Personal Data "https://www.pcpd.org.hk/english/news_events/media_statements/press_20241029.html" （2024）.

トとして、以下を挙げている。

- 違法なスクレイピングから効果的に保護するために、組織は複数の保護措置を組み合わせて導入すべきであり、これらの措置は、スクレイピング技術やこれに関連するテクノロジーの進化に対応するために、定期的に見直し更新すべきである。
- 一部の高度なスクレイパーは検出を回避するために人工知能（AI）を使用するが、AI は違法なスクレイピング対策を強化するための解決策の一部にもなり得る。
- 違法なスクレイピングから保護する義務は、大企業のみならず中小企業（Small and Medium Enterprises。以下「**SME**」という。）にも適用される。サービスプロバイダーの支援により、SME がこの義務を履行するための低コストの対策も存在する。
- SMC 及びその他の組織がプラットフォームからの個人データのスクレイピングを契約上許可している場合、その契約条件のみでスクレイピングを適法にできないが、重要な保護措置となる場合がある。
  - 商業的又は社会的に有益な目的を含むどのような目的であっても、個人データのスクレイピングを許可する組織は、合法的な根拠を有するものとし、また、許可するスクレイピングについて透明性を確保し、さらに、法律により必要とされる場合は同意を得ることを徹底しなければならない。
  - 組織は、契約条件及びそれらのモニタリング及び執行を含む適切な対策を実施し、契約上許可されたスクレイピングデータの使用が適用されるデータ保護及びプライバシー法を遵守するようにすべきである。
- 組織が第三者に自らのプラットフォームから公開されている個人データを収集する適法な権限を与える場合、アプリケーション・プログラミング・インターフェース（API）を通じて提供することで、組織がデータをより制御可能とし、不正なスクレイピングの検知の促進及び軽減が可能となる。
- スクレイピングデータセット及び（又は）自社プラットフォームのデータを用いて AI（大規模言語モデルなど）を訓練する SMC 及びその他の組織は、データ保護及びプライバシー法並びに AI 固有の法令が存在する場合は、それらの法令を遵守しなければならない。規制当局が AI モデルの開発と実装に関するガイドラインや原則を提供している場合、組織はそのガイドラインを遵守することが期待される。

(2) オランダ

2024 年 5 月 1 日、オランダデータ保護局（以下「**AP**」という。）は、個人及び組織によるデータスクレイピングに関するガイドライン（以下「**2024AP ガイドライン**」という。）[27]を公開した[28]。

2024AP ガイドラインは、個人データを対象にウェブスクレイピングすることは、直接的にスクレイピングされた者の個人データを保護する権利に対する重大な侵害になると述べ、民間組織又は個人が、スクレイピングによるデータを使用したい場合は、GDPR に基づく原則及び要件に準拠する必要があるとしていた（2024AP ガイドライン 11 章（24 頁））。すなわち、GDPR6 条のいずれかの根拠が成立するかを検討する必要があるところ、通常は「正当な利益」（6 条 1 項 f）のみが根拠になるとした上で、前述の正当利益の 3 要件を定立する（2024AP ガイドライン 5.1 及び 5.2 章（11 頁））。そして、AP は、法的に保護された利益のみが正当利益として認められるとし、純粋に商業的な利益は含まれないという見解に立脚していた

---

27 AP “Richtlijnen scraping door private organisaties en particulieren（民間組織及び個人によるスクレイピングのガイドライン）”（2024）． 当該サイトは、現在リンク切れとなっているため、リンクを付していない。

28 AP “scraping bijna altijd illegaal”（2024）． 当該サイトは、現在リンク切れとなっているため、リンクを付していない。

（2024AP ガイドライン 5.2 章（12 頁）及び 8 章（21 頁））。

以上から、AP は、2024AP ガイドラインを公開した時点では、ICO の見解よりも正当利益として認められる範囲を狭く解していたと思われる。

この点について、GDPR 違反を理由にオランダロイヤルローンテニス協会に罰金を科した AP の決定に関する訴訟手続において、「正当な利益」の概念として商業的な利益が含まれるかが争われ、当該訴訟を審理していたアムステルダム地方裁判所は、2022 年 9 月 22 日付で、欧州連合司法裁判所に対し、EU 機能条約 267 条に基づく先決裁定を求めていた。AP は、当該判決が出るまでこの立場を維持するとしていた（2024AP ガイドライン 5.2 章（12 頁））。

その後、2024 年 10 月 4 日、欧州連合司法裁判所は判決を行い[29]、「正当な利益」の概念は、法律によって定められた利益に限定されないが、主張される正当な利益が合法であることが必要であるとした上で（同判決文 No.40）、「正当な利益」の概念として商業的な利益が含まれるかという問いに対しては正面から回答せず、事例判断的に「管理者の商業的利益を満たすためにスポーツ連盟のメンバーの個人データを対価として開示することから構成される個人データの処理は、その処理が問題の正当な利益のために厳密に必要であり、全ての関連する状況に照らして、それらのメンバーの利益又は基本的な権利と自由がその正当な利益に優先しないという条件でのみ、その規定の意味において、管理者が追求する正当な利益のために必要であるとみなされる可能性があることを意味すると解釈されなければならない」（判決文 No.57）と述べた。

このように、欧州連合司法裁判所は「正当な利益」の概念として商業的な利益が全く含まれないとはしていないため、AP も、これを受けて、この点に関する説明部分については最新ではない旨を、正当な利益について解説しているウェブページに追記していた[30]。

その後、2025 年 4 月 2 日、AP は、個人及び民間組織によるデータスクレイピングに関するガイドライン（以下「**2025AP ガイドライン**」という。）[31]を新たに公開した。2025AP ガイドラインにおいても、2024AP ガイドラインと同様に、GDPR6 条のいずれかの根拠が成立するかを検討する必要があるところ、通常は「正当な利益」（6 条 1 項 f）のみが根拠となるとした上で、前述の正当利益の 3 要件を定立しており、個人データを対象にウェブスクレイピングすることは、直接的にスクレイピングされた者の個人データを保護する権利に対する重大な侵害になると述べ、民間組織又は個人が、スクレイピングによるデータを使用したい場合は、GDPR に基づく原則及び要件に準拠する必要があるとしている（2025AP ガイドライン 5 章及び 11 章（10 頁以下及び 25 頁））。もっとも、2025AP ガイドラインにおいて、AP は、純粋に商業的な利益は含まれないという従前の見解を削除している（2024AP ガイドライン 5.2 章（12 頁）及び 8 章（21 頁）、2025AP ガイドライン 5 章（10 頁以下）対照）。そのため、純粋に商業的な利益は含まれないという、従前の見解を変更したものと考えられる。

(3) イタリア

2024 年 5 月 30 日、イタリアデータ保護局（以下「**GPDP**」という。）は、データ管理者としての立場に

---

29 Judgment - 04/10/2024 - Koninklijke Nederlandse Lawn Tennisbond Case C-621/22 (ECLI:EU:C:2024:857).

30 AP “Grondslagen AVG uitgelegd”（2024）.

31 AP “Handreiking scraping door particulieren en private organisaties（個人及び民間組織によるスクレイピングのガイドライン）”（2025）.

ある公的機関及び民間団体がオンラインで公開する個人データを、生成 AI モデルのトレーニングを目的とするウェブスクレイピングから保護する方法に関するガイドラインを発効した[32]。

このガイドラインは、スクレイパーを対象としておらず、ウェブサイトやオンラインプラットフォームで個人データを公開する公的機関や民間団体を対象とする（ガイドライン 2 頁）。そして、ガイドラインでは、技術的・組織的に保護エリアを設置すること（GDPR5 条 1 項 c が求める「データの最小化（data minimization）」に関連する。）、利用規約でウェブスクレイピングを禁止する条項を設けること、技術的に HTTP リクエスト等を監視すること、CAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）検証を課すこと等により bot の操作を防ぐことが挙げられている（ガイドライン 5 頁）。これらの措置は必ずしも必須ではないが、データ管理者はウェブスクレイピングによる影響を防止又は軽減するために選択的に措置を実施するかを評価する必要があるとされる（ガイドライン 2 頁）。

(4) アイルランド

アイルランドデータ保護委員会（以下「**IDPC**」という。）は、2022 年 11 月 28 日、Facebook のデータ管理者である Meta Platforms Ireland Limited（MPIL）に対し、GDPR25 条 1 項及び 2 項に違反するとして、2 億 6,500 万ユーロの罰金と一連の是正措置を課した[33]。

これは、GDPR の施行日から Facebook ユーザーの個人情報データ流出が判明するまでの間に、同社がリリースした複数の Facebook、Messenger、Instagram についてスクレイピング攻撃対策をしなかったため、GDPR25 条に定める「データ保護バイデザイン及びデータ保護バイデフォルト」という要件を満たしていなかったと IDPC が判断したことに基づくものである。すなわち、許可なくデータを自動収集するスクレイピング等の行為は、Facebook の利用規約違反であるが、その違反に対する措置を講じていなかったことが問題視された[34]。これを受けて同社は、2020 年 10 月に 2 社のウェブスクレイピング企業に対し訴訟を提起したほか（2022 年に和解）[35]、2022 年 7 月にも、2 社のウェブスクレイピング企業及び個人に対し法的措置を講じている[36]。

(5) フランス

フランスのデータ保護当局（以下「**CNIL**」という。）は、2025 年 6 月 19 日（英語版は 7 月 2 日）、スクレイピングに関するコメントを公表した[37]。

---

32 GPDP “Intelligenza artificiale: dal Garante privacy le indicazioni per difendere i dati personali dal web scraping - PUBBLICATO IN GAZZETTA UFFICIALE IL PROVVEDIMENTO”（2024）.

33 IDPC ”Data Protection Commission announces decision in Facebook “Data Scraping” Inquiry” （2022）.

34 Facebook「利用規約」3-2-3 によれば「弊社から事前の許可を得ることなく、自動化手段を用いて弊社製品のデータにアクセスしたり、データを取得したりすること、及びアクセス許可のないデータへのアクセスを試みることは禁止されています。」と規定されている。

35 Meta “Taking Legal Action Against Data Scraping” （2020）.

36 Meta “Taking Action Against Scraping for Hire” （2022）.

37 CNIL “The legal basis of legitimate interests: Focus sheet on measures to implement in case of data collection by web scraping” （2025）.

CNIL は、ウェブスクレイピングを全面的に禁止せず、AI の開発者がウェブスクレイピングを行う際に検討が必要な措置（mandatory measures）と、追加的な保護措置（additional safeguards）のリストを提示している。EDPB がウェブスクレイピングに関するガイダンスを提示するまでの間は [38]、このリストを参照することが GDPR のリスクを低減するために有用であるように思われる。検討が必要な措置については、必要最小限度の原則（GDPR5 条 1 項(c)）の観点から、①事前に正確な収集基準を定義し、不要なデータカテゴリの収集を排除するためのフィルターを適用し、②これらの基準にもかかわらず収集された関連のないデータは、収集直後又はそのようなデータとして特定された時点で直ちに削除されることを確保するべき旨が規定されている。

38 EDPB Work Programme 2024-2025 によれば、Guidelines on generative AI – data scraping が検討されているようである。