***PROVABLE SCRAPING COMPLIANCE*:**
desain tanggung jawab korporasi atas *data scraping* melalui *enforced self-regulation* menuju *accountable data* di Indonesia

Oleh: **Muhammad Rafli Mawla Rakhman** [1]

[1] *Universitas Brawijaya Jalan Veteran No.10-11, Ketawanggede, Kecamatan Lowokwaru, Kota Malang, Jawa Timur 65145*

email: raflimawla@student.ub.ac.id

Informasi Artikel Received:	08-08-2026	**Accepted:**	25-10-2026

ABSTRAK

Aktivitas *data scraping* telah menjadi bagian penting dalam ekonomi digital, namun juga menghadirkan persoalan hukum terkait batas legalitas ekstraksi, pemanfaatan ulang data, dan pertanggungjawaban korporasi terhadap risiko yang timbul. Penelitian ini bertujuan menganalisis kesenjangan normatif dalam tata kelola *web scraping* di Indonesia serta merumuskan desain *Provable Scraping Compliance* sebagai model akuntabilitas korporasi berbasis *enforced self-regulation.* Penelitian ini menggunakan metode penelitian hukum normatif dengan pendekatan perundang-undangan, konseptual, dan komparatif melalui analisis terhadap rezim pelindungan data Indonesia, prinsip *accountability* dalam General *Data Protection Regulation* (GDPR), serta konsep tata kelola data berbasis risiko. Hasil penelitian menunjukkan bahwa regulasi Indonesia telah memberikan dasar perlindungan terhadap pemrosesan data pribadi, tetapi belum menyediakan mekanisme pembuktian kepatuhan yang secara spesifik mampu mengakomodasi karakteristik ekstraksi data otomatis. Penelitian ini merumuskan empat pilar *Provable Scraping Compliance,* yaitu *Pre-Scraping Legality and Impact Assessment, Scraping Evidence Registry, Auditable Extraction Trail,* dan *Rebuttable Presumption of Compliance* sebagai instrumen untuk menghubungkan kewajiban hukum dengan bukti kepatuhan yang dapat diverifikasi. Model ini diperlukan karena kompleksitas *web scraping* tidak hanya terletak pada legalitas akses, tetapi juga pada kemampuan korporasi mempertanggungjawabkan siklus hidup data secara transparan. Konsep yang dikembangkan memperkuat *accountable data governance* di

Indonesia tanpa melakukan transplantasi penuh terhadap model regulasi asing.

Kata kunci : *data scraping; personal data protection; corporate accountability; enforced self-regulation; accountable data governance*

ABSTRACT

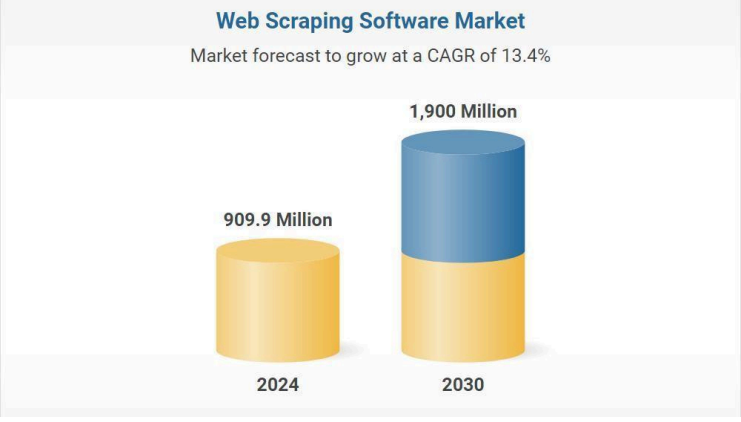
Data scraping has become an essential component of the digital economy while simultaneously creating legal challenges concerning the legality of extraction, secondary data utilization, and corporate accountability for resulting risks. This study aims to examine the normative gap in Indonesia's governance of web scraping and formulate the Provable Scraping Compliance framework as a corporate accountability model based on enforced self-regulation. This research employs normative legal research using statutory, conceptual, and comparative approaches by examining Indonesia's personal data protection regime, the accountability principle under the UE GDPR, and risk-based data governance concepts. The findings demonstrate that Indonesian regulations have established fundamental obligations regarding personal data processing but have not provided a specific mechanism to demonstrate compliance in automated data extraction activities. Proposes four pillars of Provable Scraping Compliance: Pre-Scraping Legality and Impact Assessment, Scraping Evidence Registry, Auditable Extraction Trail, and Rebuttable Presumption of Compliance as mechanisms that connect legal obligations with verifiable compliance evidence. This framework is necessary because the complexity of web scraping lies not only in determining lawful access but also in ensuring transparent accountability throughout the data lifecycle. The proposed model contributes to accountable data governance in Indonesia by adapting global accountability principles without adopting foreign regulatory frameworks.

Keywords: *data scraping; personal data protection; corporate accountability; enforced self-regulation; accountable data governance*

PENDAHULUAN

Transformasi ekonomi digital telah menempatkan data bukan lagi sekadar representasi informasi, melainkan sebagai aset strategis yang menentukan nilai ekonomi (*data is the new oil*), inovasi teknologi, dan proses pengambilan keputusan. Dalam ekosistem digital yang semakin bergantung pada pengumpulan dan pengolahan informasi, aktivitas ekstraksi data secara otomatis *(automated data*

extraction) melalui mekanisme *web scraping* menjadi salah satu instrumen penting dalam memperoleh, mengagregasi, dan menganalisis data dalam skala besar (Khder, 2021). Praktik tersebut memberikan kontribusi signifikan terhadap berbagai aktivitas digital, mulai dari pengindeksan mesin pencari, analisis pasar, penelitian akademik, hingga pengembangan kecerdasan buatan. Nilai ekonominya tercermin dari pasar perangkat lunak *web scraping* global yang diperkirakan mencapai USD909,9 juta pada 2024 dan diproyeksikan menembus USD1,9 miliar pada 2030, dengan pertumbuhan tahunan sebesar 13,4% (Market Glass, 2026).



Gambar 1: Nilai Ekonomi Proyeksi Pasar *Web Scraping*

Sumber: *Research and Market, 2026*

Skala aktivitas ekstraksi otomatis tersebut menunjukkan bahwa pengumpulan data digital tidak lagi berlangsung melalui interaksi manusia secara konvensional. Tampak nyata melalui lalu lintas internet yang berasal dari aktivitas otomatis terus mengalami peningkatan, lebih dari 53% pada 2025 mencapai lebih dari separuh keseluruhan aktivitas internet dengan sebagian besar aktivitas tersebut berasal dari bot yang melakukan pengambilan informasi secara otomatis (Imperva, 2026). Kondisi tersebut menunjukkan adanya perubahan karakter ruang digital, yaitu ketika proses pengumpulan informasi tidak hanya dilakukan oleh pengguna, tetapi juga oleh sistem otomatis yang mampu mengambil, menyalin, dan memproses data dalam jumlah besar. Pada titik ini, *web scraping* tidak lagi dapat dipahami semata-mata sebagai persoalan teknis, melainkan sebagai praktik tata kelola data yang menentukan bagaimana informasi diperoleh, digunakan kembali, dan dikonversi menjadi nilai ekonomi.

Meskipun demikian, keberadaan *web scraping* menghadirkan ambivalensi dalam perspektif hukum. Di satu sisi, aktivitas tersebut menjadi fondasi bagi berbagai inovasi digital karena memungkinkan efisiensi pengumpulan informasi dan pengembangan teknologi berbasis data. Di sisi lain, mekanisme yang sama dapat menciptakan risiko

ketika data diperoleh tanpa transparansi yang memadai, digunakan melampaui tujuan awal, atau melibatkan informasi yang berkaitan dengan individu tertentu (Paguay-Chimarro et al., 2025). Kendati persoalan utama dalam praktik *scraping* tidak terletak pada keberadaan teknologi tersebut, melainkan pada bagaimana menentukan batas antara ekstraksi data yang sah, proporsional, dan dapat dipertanggungjawabkan dengan praktik pengumpulan data yang berpotensi mereduksi hak subjek data.

Perdebatan mengenai batas legalitas *web scraping* salah satunya terlihat dalam perkara hiQ Labs Inc. v. LinkedIn Corp. Perkara ini bermula ketika hiQ Labs, sebuah perusahaan analisis tenaga kerja, melakukan pengumpulan data publik dari profil pengguna LinkedIn untuk menghasilkan layanan *people analytics* dan prediksi risiko kehilangan pekerjaan. Pengadilan Banding Amerika Serikat untuk *Ninth Circuit* dalam putusannya pada tahun 2022 menegaskan bahwa akses terhadap informasi yang tersedia secara publik tidak secara otomatis dikategorikan sebagai akses tanpa hak berdasarkan CFAA. Meskipun demikian, perkara tersebut tidak memberikan legitimasi mutlak terhadap seluruh aktivitas *scraping* karena persoalan lain seperti pelanggaran kontrak, penggunaan data di luar tujuan awal, serta perlindungan data pribadi tetap menjadi isu hukum yang terbuka. Perkara ini menunjukkan bahwa sifat publik suatu data tidak serta-merta menghilangkan kebutuhan terhadap mekanisme akuntabilitas dalam proses ekstraksi dan pemanfaatannya kembali (Franklen, 2017).

Urgensi pengaturan *web scraping* di Indonesia tidak lagi bersifat hipotetis, sebab teknik tersebut telah digunakan dalam penyelenggaraan kebijakan publik berskala nasional. Badan Pusat Statistik (BPS) mulai mengintegrasikan data harga pasar daring ke dalam penghitungan Indeks Harga Konsumen pada 2024 melalui program berbasis *Python* yang mengekstraksi harga dan ketersediaan produk dari platform perdagangan elektronik (Dinda Decembria, 2023). BPS menyatakan bahwa proses tersebut diselenggarakan dengan memperhatikan ketentuan penggunaan platform, kendati pada saat yang sama mengidentifikasi kepatuhan hukum, keterbatasan akses platform, perubahan struktur situs, dan etika pemanfaatan data sebagai tantangan utama. Tak hanya itu, dimensi risiko dari ekstraksi data berskala masif tercermin dalam skandal Facebook–Cambridge Analytica. Berdasarkan informasi yang dirilis pemerintah pada 2018, sekitar 1.096.666 pengguna Facebook Indonesia diperkirakan terdampak pengumpulan data oleh aplikasi pihak ketiga, menempatkan Indonesia sebagai negara dengan jumlah pengguna yang berpotensi terpapar terbesar ketiga setelah Amerika Serikat dan Filipina (Indonesia, 2018). Data tersebut tidak diperoleh melalui peretasan konvensional, melainkan melalui aplikasi dan antarmuka pemrograman Facebook yang memungkinkan pengumpulan informasi pengguna beserta jejaring pertemanannya untuk kepentingan pemrofilan.

Kedua fenomena tersebut menyingkap ambivalensi ekstraksi data dalam ekosistem digital Indonesia. BPS memperlihatkan bahwa pengumpulan otomatis dapat melayani kepentingan publik ketika sumber, tujuan, sampel, dan metode penggunaannya terdokumentasi. Sebaliknya, perkara Facebook–Cambridge Analytica menunjukkan bahwa akses teknis yang tersedia dapat menghasilkan ekstraksi data dalam skala masif tanpa pengetahuan sebagian besar subjek data, sementara mekanisme pemulihannya justru terhenti pada hambatan prosedural sebelum substansi pertanggungjawaban korporasi diperiksa.

Indonesia kemudian membangun rezim perlindungan data pribadi melalui Undang-Undang Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi (UU PDP). Kehadiran regulasi tersebut menandai pergeseran paradigma dari pengaturan yang sebelumnya tersebar dalam berbagai regulasi sektoral menuju suatu kerangka hukum yang lebih komprehensif. UU PDP mengatur prinsip pemrosesan data pribadi, dasar pemrosesan, hak subjek data pribadi, kewajiban pengendali dan prosesor data pribadi, serta kewajiban menjaga keamanan pemrosesan data. Salah satu prinsip penting dalam rezim tersebut adalah prinsip akuntabilitas *(accountability principle)*, yaitu kewajiban pengendali data untuk tidak hanya mematuhi ketentuan perlindungan data, tetapi juga mampu menunjukkan pertanggungjawaban atas pemenuhan kewajiban tersebut (Syailendra et al., 2024).

Namun, perkembangan teknologi ekstraksi data otomatis memperlihatkan adanya celah antara prinsip normatif dan mekanisme pembuktian kepatuhan dalam praktik. UU PDP telah memberikan kewajiban umum mengenai pemrosesan data pribadi, tetapi belum secara spesifik menjelaskan bagaimana korporasi yang melakukan *web scraping* harus menunjukkan bahwa aktivitas ekstraksi yang dilakukan memiliki dasar hukum, tujuan yang proporsional, mekanisme pembatasan, serta dokumentasi yang dapat diverifikasi (Supeno et al., 2025). Kondisi tersebut melahirkan apa yang dalam penelitian ini disebut sebagai *provability gap*, yaitu kesenjangan antara kewajiban untuk bertindak sesuai hukum dan kemampuan untuk membuktikan bahwa tindakan tersebut benar-benar telah dilakukan secara patuh. Kendati demikian melahirkan suatu *quo vadis*, yaitu permasalahan hukum tidak lagi hanya berkaitan dengan penentuan apakah suatu aktivitas *scraping* melanggar hukum, tetapi juga mengenai bagaimana korporasi dapat menunjukkan bahwa aktivitas tersebut dilakukan melalui tata kelola yang dapat diaudit.

Penelitian terdahulu telah banyak membahas *web scraping* dari perspektif hak cipta, akses tanpa hak, perlindungan data pribadi, maupun etika penggunaan data digital. Selain itu, kajian mengenai prinsip akuntabilitas dalam rezim perlindungan data global, khususnya melalui *General Data Protection Regulation* (GDPR), telah

menunjukkan bahwa pengendali data memiliki kewajiban untuk membuktikan kepatuhannya melalui langkah teknis dan organisatoris yang sesuai. Meskipun demikian, kajian yang secara khusus merumuskan bagaimana prinsip akuntabilitas tersebut diterjemahkan menjadi suatu mekanisme pembuktian kepatuhan bagi korporasi yang melakukan ekstraksi data otomatis masih terbatas. Penelitian ini menawarkan konsep ***Provable Scraping Compliance*** sebagai model *enforced self-regulation* yang menempatkan pembuktian kepatuhan sebagai elemen utama tata kelola *web scraping*. Pertanyaan penelitian yang diajukan ialah bagaimana *status quo* hukum Indonesia mengonstruksikan kewajiban korporasi dalam aktivitas *web scraping* dan sejauh mana pengaturan tersebut menyediakan instrumen untuk membuktikan kepatuhan; serta bagaimana *Provable Scraping Compliance* berbasis *enforced self-regulation* dapat dirumuskan untuk menutup *provability gap* tanpa menegasikan pemanfaatan data yang sah. Penelitian ini ditujukan untuk menilai koherensi pengaturan yang berlaku sekaligus merumuskan arsitektur pembuktian kepatuhan yang menghubungkan penilaian pra-ekstraksi, pencatatan teknis, verifikasi independen, dan konsekuensi hukum berbasis tingkat kepatuhan.

Penelitian ini memiliki keterkaitan dengan *Sustainable Development Goals* (SDGs) 16: Peace, Justice and Strong Institutions, khususnya Target 16.6 mengenai pembangunan institusi yang efektif, akuntabel, dan transparan. Dalam konteks tata kelola data digital, penguatan akuntabilitas korporasi dalam aktivitas *web scraping* merupakan bagian dari upaya membangun institusi digital yang mampu menjamin keseimbangan antara inovasi teknologi dan perlindungan hak individu. Selain itu, penelitian ini turut mendukung SDGs 9: Industry, Innovation and Infrastructure melalui pengembangan kerangka regulasi yang memungkinkan pemanfaatan teknologi berbasis data tetap berjalan secara inovatif dengan standar tata kelola yang bertanggung jawab.

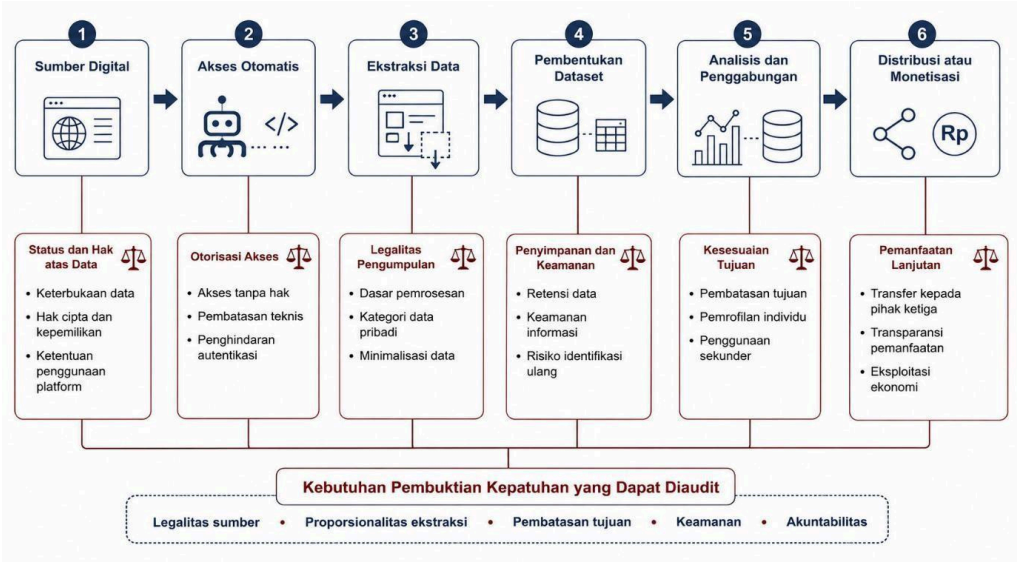
METODE

Penelitian ini merupakan penelitian hukum yuridis-normatif yang ditujukan untuk menelaah sekaligus merekonstruksi tata kelola *web scraping* dalam rezim pelindungan data pribadi melalui pengkajian terhadap norma, asas, dan doktrin hukum yang relevan. Penelitian menggunakan pendekatan perundang-undangan *(statute approach)* untuk menganalisis koherensi pengaturan dalam hukum Indonesia, pendekatan konseptual *(conceptual approach)* untuk mengonstruksikan prinsip akuntabilitas dan *enforced self-regulation* sebagai fondasi *Provable Scraping Compliance*, serta pendekatan perbandingan (*comparative approach*) terhadap Uni Eropa melalui *General Data Protection Regulation* dan Tiongkok melalui *Personal Information Protection Law*, yang dipilih secara purposif berdasarkan perbedaan

orientasi regulatif keduanya dalam tata kelola data. Bahan hukum primer mencakup peraturan perundang-undangan nasional, instrumen hukum dari yurisdiksi pembanding, dan putusan pengadilan yang relevan; bahan hukum sekunder meliputi buku, artikel ilmiah, laporan penelitian, dan pendapat ahli; sedangkan bahan hukum tersier mencakup kamus hukum, dan sumber penunjang lain yang memberikan penjelasan terhadap istilah teknis. Seluruh bahan hukum dikumpulkan melalui studi kepustakaan, kemudian dianalisis secara preskriptif menggunakan penafsiran gramatikal, sistematis, konseptual, dan komparatif untuk merumuskan desain pertanggungjawaban korporasi yang sesuai dengan sistem hukum Indonesia.

HASIL DAN PEMBAHASAN

Secara operasional, ekstraksi data berlangsung melalui suatu mata rantai yang lebih kompleks daripada tindakan mengakses laman publik. Perangkat otomatis terlebih dahulu mengirimkan permintaan kepada sistem sumber, membaca struktur laman atau antarmuka pemrograman, memilah elemen informasi tertentu, lalu menyimpannya ke dalam himpunan data yang dapat digabungkan dengan sumber lain. Himpunan tersebut selanjutnya dapat dianalisis, dipertukarkan, digunakan untuk melatih sistem algoritmik, atau dimonetisasi sebagai bagian dari kegiatan usaha (Wibowo & Karyati, 2021). Setiap tahapan melahirkan titik singgung hukum yang berbeda. Permintaan akses berhadapan dengan otorisasi dan pembatasan sistem; pemilahan data bersentuhan dengan kategori serta sensitivitas informasi; penyimpanan dan agregasi memunculkan persoalan dasar pemrosesan, minimalisasi, retensi, dan keamanan; sedangkan distribusi atau monetisasi menuntut penilaian baru terhadap kesesuaian tujuan dan hak pihak yang terdampak.



Gambar 2. Alur Ekstraksi Data Otomatis dan Titik Risiko Hukumnya

***Sumber**: Kreasi Penulis, 2026*

Karakter tersebut menggugurkan pandangan simplistis bahwa legalitas *web scraping* cukup ditentukan oleh sifat publik atau tertutupnya suatu informasi. Keterbukaan teknis hanya menunjukkan bahwa data dapat dilihat tanpa hambatan tertentu; ia tidak dengan sendirinya menghadirkan persetujuan untuk melakukan pengambilan massal, penggabungan lintas sumber, ataupun pemanfaatan ulang bagi kepentingan yang tidak pernah diketahui subjek data. Informasi yang tampak tidak sensitif ketika berdiri sendiri bahkan dapat berubah sifat setelah dihimpun dalam jumlah besar dan dikorelasikan dengan data lain. Nama, lokasi pekerjaan, riwayat aktivitas, atau preferensi yang terpencar dapat membentuk profil komprehensif mengenai seseorang ketika diproses secara agregatif. Pada titik inilah *scraping* bergeser dari sekadar reproduksi informasi menuju tindakan pemrosesan yang memiliki daya intervensi terhadap privasi, otonomi informasional, dan kepentingan ekonomi.

Ambivalensi tersebut diperuncing oleh asimetri pengetahuan antara korporasi dan pihak yang datanya diekstraksi. Korporasi menguasai arsitektur teknis, tujuan bisnis, volume pengambilan, masa penyimpanan, serta pola pemanfaatan dataset, sedangkan subjek data dan regulator acap kali hanya mengetahui keberadaan pemrosesan setelah timbul sengketa atau kerugian. Model penegakan yang semata bertumpu pada pembuktian pelanggaran setelah akibat muncul menjadi tidak memadai (*es post*), sebab bukti yang menentukan legalitas justru berada dalam sistem internal pelaku ekstraksi. Persoalan tersebut melahirkan kesenjangan antara kewajiban substantif dan kapasitas pembuktiannya. Suatu korporasi dapat

menyatakan telah mematuhi prinsip legalitas, pembatasan tujuan, proporsionalitas, dan keamanan, tetapi pernyataan tersebut tidak memiliki bobot yang memadai apabila tidak ditopang oleh jejak keputusan serta catatan teknis yang dapat diaudit. Kebutuhan hukum karenanya tidak berhenti pada penentuan data apa yang boleh atau tidak boleh diekstraksi. Selanjutnya lebih mendesak ialah membangun standar yang memungkinkan legalitas seluruh siklus *scraping* diuji sejak tahap perencanaan, pelaksanaan, penyimpanan, sampai penggunaan lanjut. Dari kerangka inilah *status quo* pengaturan Indonesia perlu ditelaah: apakah rezim yang ada telah menyediakan instrumen untuk menilai sekaligus membuktikan kepatuhan, atau baru menyusun larangan dan kewajiban dalam rumusan yang bersifat umum.

A. Fragmentasi Pengaturan dan Kesenjangan Pembuktian Kepatuhan *Web Scraping*

Status hukum *web scraping* dalam sistem hukum Indonesia tidak ditentukan oleh ada atau tidaknya istilah tersebut di dalam peraturan perundang-undangan. Kualifikasi hukumnya bergantung pada objek yang diekstraksi, cara sistem memperoleh akses, dasar pemrosesan yang digunakan, dan tujuan pemanfaatan data setelah pengumpulan. Ketika informasi yang dihimpun berkaitan dengan orang yang teridentifikasi atau dapat diidentifikasi, kegiatan *scraping* memasuki rezim Pasal 16 ayat (1) UU PDP yang memasukkan perolehan dan pengumpulan, pengolahan dan penganalisisan, penyimpanan, penampilan, pengumuman, transfer, penyebarluasan, hingga penghapusan sebagai rangkaian pemrosesan data pribadi. Korporasi yang menentukan tujuan dan mengendalikan pemrosesan tersebut berkedudukan sebagai Pengendali Data Pribadi, sekalipun pengoperasian perangkat *scraper* diserahkan kepada penyedia jasa teknologi.

Masuknya *web scraping* ke dalam cakupan pemrosesan tidak menjadikan seluruh pengambilan data otomatis terlarang. Pasal 20 ayat (2) UU PDP mengakui enam dasar pemrosesan, yakni persetujuan eksplisit, pemenuhan perjanjian, kewajiban hukum, pelindungan kepentingan vital, pelaksanaan tugas untuk kepentingan umum atau pelayanan publik, serta pemenuhan kepentingan sah lainnya. Dalam praktik komersial, korporasi yang mengambil data dari sumber terbuka cenderung tidak memperoleh persetujuan langsung sehingga dapat berupaya mendasarkan pemrosesan pada “kepentingan yang sah lainnya” sebagaimana Pasal 20 ayat (2) huruf f. Dasar ini tetap mensyaratkan penilaian terhadap tujuan, kebutuhan, keseimbangan kepentingan pengendali, dan hak subjek data; keterbukaan teknis suatu informasi tidak serta-merta menggantikan pengujian tersebut. UU PDP, bagaimanapun, belum mengharuskan hasil uji kepentingan sah dituangkan ke dalam dokumen dengan format, parameter, dan masa simpan tertentu. Akibatnya, klaim mengenai kebutuhan bisnis dan proporsionalitas ekstraksi dapat

berhenti sebagai penilaian internal yang baru dipersoalkan setelah muncul keberatan atau kerugian.

Persoalan serupa terlihat pada kewajiban transparansi. Pasal 21 ayat (1) UU PDP mengharuskan pengendali menyampaikan legalitas, tujuan, jenis dan relevansi data, masa retensi, rincian informasi yang dikumpulkan, serta jangka waktu pemrosesan apabila pemrosesan didasarkan pada persetujuan. Pasal 24 selanjutnya membebankan kewajiban pembuktian persetujuan kepada pengendali. Formulasi tersebut cukup tegas untuk pengumpulan langsung berbasis persetujuan, tetapi tidak menyediakan ketentuan khusus yang setara bagi data pribadi yang diperoleh secara tidak langsung dari laman, platform, atau basis data milik pihak lain. Kekosongan ini relevan bagi *web scraping* karena subjek data acap kali tidak berinteraksi dengan korporasi yang mengambil datanya dan bahkan tidak mengetahui keberadaan pemrosesan tersebut (Wibowo & Karyati, 2021). Prinsip transparansi dalam Pasal 16 ayat (2) tetap berlaku, tetapi belum diterjemahkan ke dalam kewajiban pemberitahuan tidak langsung yang memuat identitas sumber, waktu ekstraksi, kategori data, dan penggunaan lanjutan.

UU PDP sebenarnya telah membangun embrio tata kelola berbasis bukti (*evidence-based policy*). Pasal 31 mewajibkan pengendali merekam seluruh kegiatan pemrosesan, sedangkan Pasal 32 memberi subjek data akses terhadap data pribadi beserta rekam jejak pemrosesannya. Pasal 34 mewajibkan penilaian dampak apabila kegiatan memiliki potensi risiko tinggi, antara lain pemrosesan dalam skala besar, pemantauan sistematis, pencocokan atau penggabungan kumpulan data, penggunaan teknologi baru, pemrosesan data spesifik, dan pengambilan keputusan otomatis yang menimbulkan akibat hukum atau dampak signifikan. Karakteristik tersebut dapat melekat secara kumulatif pada aktivitas *scraping* yang digunakan untuk pemrofilan, analisis perilaku, atau pelatihan sistem algoritmik (Cahyono, 2021). Pasal 35 sampai dengan Pasal 39 kemudian mengharuskan langkah teknis-operasional, penentuan tingkat keamanan, pengawasan terhadap pihak yang terlibat, pengamanan sistem, serta pencegahan akses tidak sah. Puncaknya, Pasal 47 menegaskan bahwa pengendali wajib bertanggung jawab atas pemrosesan dan menunjukkan pertanggungjawaban dalam pelaksanaan prinsip pelindungan data pribadi.

Konstruksi tersebut menolak anggapan bahwa UU PDP sama sekali tidak memiliki prinsip akuntabilitas. Persoalannya justru berada pada tingkat granularitas bukti. Pasal 31 tidak menentukan muatan minimum rekaman pemrosesan, sedangkan Pasal 34 ayat (3) menyerahkan pengaturan lebih lanjut mengenai penilaian

dampak kepada Peraturan Pemerintah. Untuk aktivitas *web scraping*, rekaman yang hanya mencantumkan tujuan umum, kategori data, dan masa retensi belum cukup menunjukkan bagaimana data diperoleh. Verifikasi memerlukan informasi yang lebih spesifik, seperti alamat sumber, waktu ekstraksi, identitas perangkat atau *crawler*, elemen data yang diambil, volume dan frekuensi permintaan, dasar pemrosesan untuk setiap kategori data, pembatasan teknis yang ditemui, perubahan pada himpunan data, penggunaan sekunder, transfer kepada pihak lain, hingga pelaksanaan penghapusan (Shahrullah et al., 2025). Ketiadaan standar ini menyebabkan kewajiban "merekam" dan "menunjukkan pertanggungjawaban" belum otomatis menghasilkan artefak bukti yang seragam dan dapat diperiksa lintas korporasi.

Di luar rezim pelindungan data, metode akses yang digunakan oleh *scraper* bersinggungan dengan hukum pidana sistem elektronik. Rujukan terhadap Pasal 30 UU ITE perlu diperbarui karena sejak berlakunya KUHP nasional pada 2 Januari 2026, ketentuan akses tanpa hak tersebut dialihkan ke Pasal 332 KUHP. Pasal 622 KUHP mengganti pengacuan Pasal 30 dan Pasal 46 UU ITE dengan Pasal 332, dan rumusan Pasal 332 kemudian disesuaikan melalui Undang-Undang Nomor 1 Tahun 2026. Pasal 332 ayat (1) mengancam akses tanpa hak atau melawan hukum terhadap komputer atau sistem elektronik milik orang lain; ayat (2) mengatur akses yang ditujukan untuk memperoleh informasi atau dokumen elektronik; sedangkan ayat (3) memperberat perbuatan yang dilakukan dengan melanggar, menerobos, melampaui, atau menjebol sistem pengamanan.

Pengaturan tersebut tidak dapat ditafsirkan bahwa setiap *scraping* merupakan akses tanpa hak. Pengambilan data dari halaman yang dapat diakses umum, pengambilan setelah autentikasi, dan pengambilan yang menerobos sistem pengamanan memiliki kualitas hukum yang berbeda. Pelanggaran *terms of service* atau pengabaian terhadap *robots.txt* juga tidak serta-merta identik dengan tindak pidana akses ilegal tanpa pemeriksaan terhadap sifat pembatasan, pengetahuan pelaku, otorisasi yang diberikan, dan cara akses yang digunakan. Batas pidananya menjadi lebih konkret ketika sistem melakukan penghindaran autentikasi, pergantian identitas untuk mengatasi pemblokiran, pemecahan batas permintaan, atau penerobosan pengamanan. Sehingga ragmentasi tampak di sini, UU PDP menilai legalitas pemrosesan dan dampaknya terhadap subjek data, sementara Pasal 332 KUHP berpusat pada keabsahan metode akses. Aktivitas yang tidak memenuhi unsur penerobosan masih dapat melanggar UU PDP apabila data pribadi dikumpulkan tanpa dasar pemrosesan yang sah atau digunakan melampaui tujuan yang dapat dipertanggungjawabkan.

UU PDP juga menyediakan jalur pidana yang berorientasi pada objek data, melalui Pasal 65 yang melarang perolehan atau pengumpulan data pribadi secara melawan hukum dengan maksud menguntungkan diri sendiri atau orang lain yang dapat merugikan subjek data. Pasal 67 mengatur ancaman pidananya, sedangkan Pasal 70 memungkinkan pertanggungjawaban terhadap korporasi, pengurus, pemberi perintah, pemegang kendali, pemilik manfaat, dan pihak yang bertindak untuk kepentingan korporasi. Norma ini dapat menjangkau *scraping* yang memang memenuhi unsur perolehan melawan hukum, motif keuntungan, dan potensi kerugian, karenanya tidak menyediakan pedoman bagi korporasi untuk menilai serta mendokumentasikan kepatuhan sebelum ekstraksi dimulai. Wataknya tetap represif dan berbasis pelanggaran, bukan preventif (*ex ante*) dan berbasis pembuktian.

Keterbatasan tersebut diperuncing oleh keadaan kelembagaan. Pasal 57 UU PDP memberikan dasar bagi sanksi administratif, termasuk denda paling tinggi 2% (dua) persen persen dari pendapatan atau penerimaan tahunan terhadap variabel pelanggaran. Pasal 58 sampai dengan Pasal 60 mengamanatkan lembaga yang bertanggung jawab kepada Presiden serta memberinya kewenangan pengawasan, pemeriksaan dokumen, pemanggilan pihak, pemeriksaan sistem elektronik, dan penjatuhan sanksi. Pasal 61 kembali menyerahkan tata cara pelaksanaan kewenangan itu kepada Peraturan Pemerintah. Laporan kinerja resmi Kementerian Komunikasi dan Digital yang diterbitkan pada 2026 masih mencatat bahwa Rancangan Peraturan Pemerintah pelaksanaan UU PDP dan Rancangan Peraturan Presiden mengenai Badan Pelindungan Data Pribadi berada dalam proses finalisasi (Susana Rita Kumalasanti, 2026). Artinya, kapasitas untuk menguji catatan pemrosesan dan melakukan audit atas aktivitas ekstraksi belum ditopang (*vacuum of power*) oleh arsitektur kelembagaan yang sepenuhnya operasional.

Perbandingan dengan Uni Eropa memperlihatkan bahwa kekuatan *General Data Protection Regulation* (GDPR) tidak terletak pada penyebutan *web scraping* secara eksplisit, melainkan pada mata rantai akuntabilitas yang lebih terperinci. Pasal 5 ayat (2) GDPR meletakkan prinsip *accountability* dengan membebankan kewajiban mematuhi sekaligus membuktikan pemenuhan prinsip pemrosesan. Pasal 6 menetapkan dasar legalitas, sedangkan Pasal 14 secara khusus mengatur informasi yang harus diberikan ketika data tidak diperoleh langsung dari subjek data, meliputi identitas pengendali, tujuan, dasar hukum, kategori data, penerima, dan sumber perolehan data. Ketentuan ini relevan bagi *scraping* karena hubungan langsung antara pengumpul dan subjek data umumnya tidak pernah terbentuk.

Dimensi pembuktiannya diperkuat oleh Pasal 24 GDPR yang mewajibkan langkah teknis dan organisatoris agar pengendali mampu menunjukkan kepatuhan, Pasal 25 mengenai *data protection by design and by default*, serta Pasal 30 mengenai catatan kegiatan pemrosesan yang sekurang-kurangnya memuat tujuan, kategori subjek dan data, penerima, transfer lintas negara, masa penghapusan, dan langkah keamanan. Pasal 35 secara tegas mewajibkan *Data Protection Impact Assessment* (DPIA) sebelum pemrosesan berisiko tinggi berlangsung, disusul konsultasi pendahuluan dengan otoritas berdasarkan Pasal 36 apabila risiko residunya tetap tinggi (Purtova, 2018). Mekanisme sertifikasi dalam Pasal 42, kewenangan investigatif otoritas pengawas berdasarkan Pasal 58, dan sanksi Pasal 83 melengkapi hubungan antara dokumentasi internal dan verifikasi eksternal. Walaupun lebih operasional daripada UU PDP, GDPR tetap belum menentukan catatan teknis yang khusus bagi *scraping*, seperti identitas *crawler*, sumber setiap elemen data, frekuensi permintaan, pembatasan akses yang dihadapi, atau riwayat pembentukan dataset.

PIPL Tiongkok menawarkan konstruksi yang lebih eksplisit terhadap data yang tersedia secara publik. Pasal 13 angka 6 memperbolehkan pemrosesan informasi pribadi yang diumumkan sendiri oleh individu atau telah diumumkan secara sah, tetapi hanya dalam lingkup yang wajar. Batas tersebut dipertegas oleh Pasal 27, yaitu informasi publik dapat diproses dalam lingkup yang wajar kecuali individu menyatakan penolakan, sementara pemrosesan yang berdampak besar terhadap hak dan kepentingan individu tetap memerlukan persetujuan. Pasal 17 juga membebankan kewajiban pemberitahuan sebelum pemrosesan, mencakup identitas pemroses, tujuan, metode, kategori data, masa penyimpanan, serta cara pelaksanaan hak. Keterbukaan data karenanya diposisikan sebagai dasar terbatas, bukan legitimasi tanpa syarat untuk melakukan ekstraksi dan pemanfaatan ulang.

Kelebihan PIPL yang paling relevan bagi penelitian ini berada pada infrastruktur auditnya. Pasal 54 mewajibkan audit kepatuhan secara berkala. Pasal 55 mengharuskan *Personal Information Protection Impact Assessment* (PIPLA) dilakukan sebelum pemrosesan data sensitif, pengambilan keputusan otomatis, pemberian data kepada pihak lain, publikasi, transfer lintas batas, dan kegiatan lain yang berdampak besar; hasil pemrosesan juga harus dicatat (Aho & Duffield, 2020). Berdasarkan Pasal 56, penilaian tersebut harus menguji legalitas, kepatutan dan kebutuhan pemrosesan, dampak terhadap hak individu, risiko keamanan, serta kesepadanan langkah pelindungan, sementara laporan dan catatannya wajib disimpan sekurang-kurangnya tiga tahun. Pasal 63 memberi otoritas akses terhadap kontrak, catatan, perangkat, dan objek yang berkaitan

dengan pemrosesan, sedangkan Pasal 64 memperkenankan regulator mewajibkan audit oleh lembaga profesional. Pasal 66 menyediakan denda hingga RMB50 juta atau lima persen omzet tahun sebelumnya untuk pelanggaran berat, dan Pasal 69 membebankan tanggung jawab kepada pemroses apabila ia tidak mampu membuktikan ketiadaan kesalahan (Dorwart et al., 2026).

Perbedaan ketiga yurisdiksi dapat dipadatkan sebagai berikut:

Tabel 1. Matriks Perbandingan Peraturan Tata Kelola Ekstraksi Data

Dimensi	Indonesia	GDPR	PIPL
Data yang diperoleh tidak langsung	Prinsip transparansi berlaku, tetapi Pasal 21 berfokus pada pemrosesan berbasis persetujuan	Pasal 14 secara khusus mengatur pemberitahuan ketika data tidak diperoleh dari subjek	Pasal 17 mewajibkan pemberitahuan sebelum pemrosesan
Data yang telah tersedia secara publik	Tidak diatur secara khusus sebagai kategori tersendiri	Dinilai melalui prinsip Pasal 5 dan dasar pemrosesan Pasal 6	Pasal 13 angka 6 dan Pasal 27 membatasi pemrosesan pada lingkup yang wajar
Rekaman pemrosesan	Pasal 31 mewajibkan perekaman seluruh kegiatan tanpa merinci elemen minimum	Pasal 30 merinci komponen catatan pemrosesan	Pasal 55–56 mewajibkan pencatatan dan penyimpanan hasil penilaian selama tiga tahun
Dimensi	**Indonesia**	**GDPR**	**PIPL**
Penilaian dampak	Pasal 34 mewajibkan untuk pemrosesan berisiko tinggi; teknisnya didelegasikan kepada PP	Pasal 35 secara eksplisit mengharuskan penilaian sebelum pemrosesan	Pasal 55 secara eksplisit mengharuskan penilaian sebelum kategori pemrosesan tertentu
Audit	Kewenangan pemeriksaan terdapat dalam Pasal 60, tetapi mekanisme kelembagaannya belum sepenuhnya operasional	Audit ditopang DPA, sertifikasi, dan kewenangan investigatif	Pasal 54 mewajibkan audit berkala dan Pasal 64 memungkinkan audit profesional atas perintah regulator
Standar bukti khusus *scraping*	Tidak tersedia	Tidak tersedia secara khusus	Tidak tersedia secara khusus

Sumber: *Kreasi Penulis, 2026*

Komparasi tersebut memperlihatkan bahwa *status quo* Indonesia bukan merupakan kekosongan hukum absolut. UU PDP telah mengatur dasar pemrosesan, pencatatan, penilaian dampak, keamanan, dan pertanggungjawaban; KUHP mengatur akses tanpa hak terhadap sistem elektronik; sementara rezim kontraktual dan kekayaan intelektual dapat berlaku menurut karakter sumber yang diekstraksi. Kelemahannya terletak pada tiga keterputusan: norma pemrosesan belum diterjemahkan menjadi artefak

pembuktian khusus *scraping*; legalitas data dan legalitas metode akses masih diperiksa melalui rezim yang berlainan; serta dokumentasi internal belum terhubung secara mapan dengan audit dan konsekuensi regulatif. *Provability gap* lahir dari konfigurasi tersebut, yakni keadaan ketika korporasi telah dibebani kewajiban untuk bertindak patuh, tetapi hukum belum menentukan bukti teknis dan yuridis apa yang harus dihasilkan agar kepatuhan itu dapat diuji. Temuan inilah yang menjadi landasan bagi perumusan *Provable Scraping Compliance* pada subbab berikutnya.

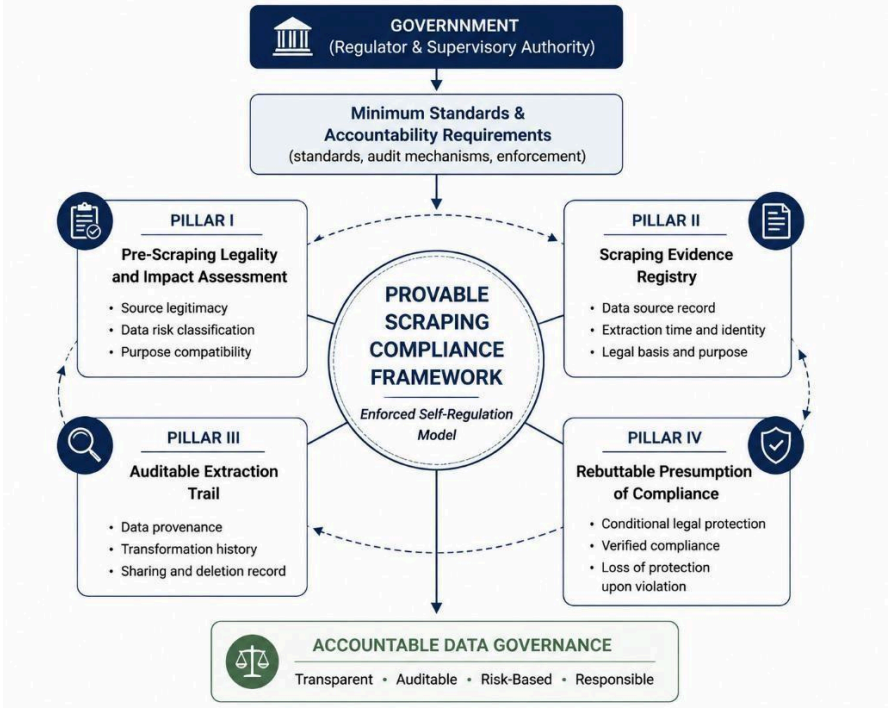
B. *Provable Scraping Compliance* sebagai Desain *Enforced Self-Regulation* dalam Pertanggungjawaban Korporasi

Ketiadaan standar pembuktian yang spesifik dalam aktivitas *web scraping* menunjukkan bahwa persoalan utama dalam tata kelola ekstraksi data bukan semata-mata terletak pada keberadaan larangan hukum, melainkan pada kemampuan untuk memastikan bahwa setiap keputusan pemrosesan dapat dipertanggungjawabkan secara objektif. Prinsip akuntabilitas dalam pelindungan data pada dasarnya telah menggeser orientasi hukum dari sekadar menentukan kewajiban substantif menuju kewajiban untuk menunjukkan kepatuhan (*demonstrable accountability*). Namun, tanpa instrumen yang mampu merekam alasan, proses, dan konsekuensi dari aktivitas ekstraksi, akuntabilitas berpotensi berhenti sebagai norma deklaratif yang sulit diuji (Protection, n.d.).

Berdasarkan kondisi tersebut, penelitian ini merumuskan konsep *Provable Scraping Compliance* sebagai desain *enforced self-regulation* yang menempatkan korporasi sebagai pihak yang bertanggung jawab untuk membangun mekanisme kepatuhan internal, tetapi tetap berada dalam kerangka standar minimum, pengawasan, dan evaluasi eksternal. Model ini tidak berangkat dari asumsi bahwa seluruh aktivitas *scraping* harus dibatasi melalui larangan absolut, sebab karakter teknologi digital tidak memungkinkan pendekatan regulasi yang sepenuhnya preskriptif. Sebaliknya, pendekatan ini membangun suatu arsitektur yang memungkinkan korporasi membuktikan bahwa proses ekstraksi data telah melalui evaluasi legalitas, pengendalian risiko, serta dokumentasi yang dapat diverifikasi.

Mengutip dalam perspektif *reflexive regulation*, negara tidak menentukan seluruh konfigurasi teknis yang digunakan oleh korporasi, tetapi membentuk mekanisme agar korporasi mampu mengendalikan risiko melalui prosedur internal yang dapat diperiksa (Refslund, 2022). Prinsip tersebut kemudian diterjemahkan melalui empat instrumen utama, yaitu penilaian sebelum ekstraksi dilakukan (*Pre-Scraping Legality and Impact Assessment*), pencatatan aktivitas melalui *Scraping Evidence Registry*, pelacakan perjalanan data melalui *Auditable*

Extraction Trail, serta pemberian perlindungan hukum bersyarat melalui *Rebuttable Presumption of Compliance.* Keempat instrumen tersebut membentuk suatu siklus kepatuhan yang menghubungkan tahap perencanaan, pelaksanaan, penggunaan, hingga evaluasi aktivitas *web scraping.*



Gambar 3. Desain Tata Kelola Provable Scraping Compliance Framework

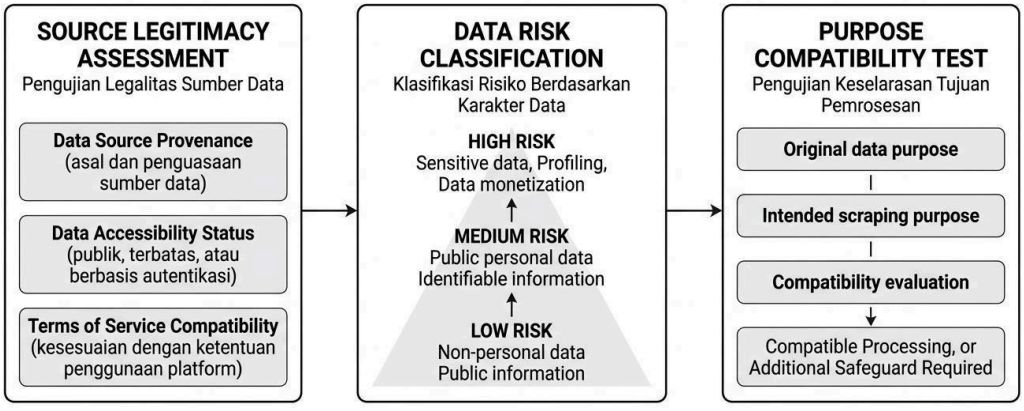
***Sumber**: Kreasi Penulis, 2026*

PILAR I. Pre-Scraping Legality and Impact Assessment

Pilar pertama berangkat dari premis bahwa legalitas suatu aktivitas *web scraping* tidak dapat sepenuhnya diuji setelah proses ekstraksi selesai dilakukan. Ketika data telah dikumpulkan, digabungkan, dan diproses lebih lanjut, sebagian risiko hukum telah terwujud dan pilihan teknis yang menentukan sifat aktivitas tersebut telah terbentuk. Oleh karena itu, pengujian kepatuhan harus dimulai sebelum suatu korporasi melakukan ekstraksi melalui mekanisme penilaian awal yang menguji sumber data, karakter informasi, tujuan penggunaan, serta potensi dampaknya terhadap subjek data.

Pendekatan ini merupakan bentuk konkret dari prinsip *privacy by design* dalam Pasal 25 GDPR dan prinsip pelindungan data berdasarkan risiko dalam berbagai rezim modern. Fokus utamanya bukan menilai apakah suatu teknologi dapat digunakan, tetapi apakah teknologi tersebut digunakan melalui proses pengambilan keputusan yang mempertimbangkan hak individu, kebutuhan bisnis,

dan risiko hukum secara proporsional. Di Indonesia, mekanisme tersebut memperkuat penerapan Pasal 34 UU PDP mengenai kewajiban melakukan penilaian dampak terhadap kegiatan pemrosesan yang berisiko tinggi, khususnya ketika aktivitas *scraping* melibatkan pemantauan sistematis, penggabungan data, teknologi baru, atau pemrofilan.



Gambar 4. Alur Pilar I Sumber: Kreasi Penulis, 2026

Penilaian *pra-scraping* terdiri atas tiga komponen utama, diantaranya:

a. Source Legitimacy Assessment: Pengujian Legalitas Sumber Data

Tahap pertama diarahkan untuk memastikan bahwa korporasi memiliki dasar yang memadai untuk memperoleh data dari sumber tertentu. Keterbukaan suatu informasi dalam ruang digital tidak secara otomatis menciptakan hak tanpa batas untuk melakukan ekstraksi. Oleh karena itu, korporasi harus melakukan evaluasi terhadap beberapa aspek:

Pertama, asal dan penguasaan sumber data *(data source provenance*). Korporasi perlu mengetahui apakah data berasal dari laman publik, layanan berbasis akun, API resmi, penyedia pihak ketiga, atau sumber lain yang memiliki batasan akses tertentu.

Kedua, status keterbukaan informasi. Data yang dapat dilihat oleh publik memiliki tingkat aksesibilitas berbeda dengan data yang tersedia setelah autentikasi atau pembatasan teknis. Perbedaan tersebut menjadi relevan karena metode memperoleh akses dapat menentukan apakah suatu tindakan berada dalam ruang penggunaan yang wajar atau justru memasuki wilayah akses tanpa hak.

Ketiga, kesesuaian dengan ketentuan penggunaan platform (*terms of service compatibility*). Meskipun pelanggaran kontraktual tidak selalu identik dengan pelanggaran hukum pidana, keberadaan pembatasan penggunaan tetap menjadi

faktor yang harus dipertimbangkan dalam menilai itikad baik dan kewajaran suatu ekstraksi data.

b. Data Risk Classification: Klasifikasi Risiko Berdasarkan Karakter Data

Setelah sumber data dinilai, langkah berikutnya adalah menentukan tingkat risiko berdasarkan karakter informasi yang akan diekstraksi. Model ini tidak menggunakan pendekatan biner berupa kategori "boleh" dan "dilarang", melainkan menggunakan pendekatan berbasis risiko (*risk-based approach*) yang mempertimbangkan konteks penggunaan. Klasifikasi tersebut dapat dibagi menjadi tiga tingkatan.

- Risiko Rendah *(Low Risk)*

Kategori ini mencakup informasi nonpersonal yang tidak memiliki hubungan langsung maupun tidak langsung dengan individu tertentu, seperti data harga umum, informasi geografis terbuka, atau statistik agregat. Aktivitas pada kategori ini tetap membutuhkan pencatatan, tetapi tidak memerlukan tingkat pengawasan yang sama dengan data pribadi.

- Risiko Menengah *(Medium Risk)*

Kategori ini mencakup data pribadi yang tersedia secara publik, seperti identitas profesional, informasi profil terbuka, atau informasi yang dapat dikaitkan kembali dengan individu tertentu. Meskipun tersedia secara publik, pemrosesan tetap harus mempertimbangkan prinsip pembatasan tujuan, transparansi, dan kemungkinan dampak terhadap individu.

- Risiko Tinggi *(High Risk)*

Kategori ini mencakup data pribadi spesifik, penggabungan beberapa dataset untuk membangun profil individu, penggunaan untuk keputusan otomatis, pelatihan model kecerdasan buatan, dan monetisasi dataset personal. Kategori ini memerlukan pengujian tambahan karena kemampuan analisis dan agregasi dapat menghasilkan konsekuensi baru yang tidak terlihat pada saat data dikumpulkan (Kaminski & Malgieri, 2021).

c. Purpose Compatibility Test: Pengujian Keselarasan Tujuan Pemrosesan

Komponen terakhir dalam pilar pertama adalah pengujian hubungan antara tujuan awal keberadaan data dan tujuan baru yang hendak dilakukan melalui *scraping*. Prinsip ini berangkat dari gagasan bahwa data tidak kehilangan konteks hanya karena tersedia secara digital. Informasi yang dipublikasikan untuk satu tujuan tertentu tidak selalu dapat dialihkan kepada tujuan ekonomi lain tanpa evaluasi tambahan (Iman, 2024). Misalnya, informasi profesional yang tersedia untuk kepentingan jejaring karier tidak secara otomatis memberikan legitimasi

untuk digunakan membangun sistem pemeringkatan kandidat kerja tanpa transparansi dan perlindungan tambahan. Demikian pula, data harga yang dikumpulkan untuk analisis pasar tidak memiliki karakter risiko yang sama dengan pengumpulan profil perilaku pengguna untuk kepentingan iklan tertarget.

PILAR II. Scraping Evidence Registry

Apabila Pilar I berfungsi sebagai mekanisme pengendalian sebelum aktivitas *web scraping* dilakukan, maka Pilar II berfokus pada pembentukan infrastruktur pembuktian selama proses ekstraksi berlangsung. Dalam kerangka *Provable Scraping Compliance*, kepatuhan tidak lagi dipahami sebagai keadaan internal yang cukup dinyatakan melalui kebijakan perusahaan (*internal compliance declaration*), melainkan sebagai kondisi hukum yang harus dapat ditunjukkan melalui rekam jejak yang objektif (*verifiable compliance evidence*)(Zulfikar et al., 2020).

Kebutuhan terhadap mekanisme tersebut berangkat dari karakteristik aktivitas *web scraping* yang berbeda dengan bentuk pemrosesan data konvensional. Pada pemrosesan langsung, hubungan antara subjek data dan pengendali relatif dapat diidentifikasi karena terdapat interaksi, pemberitahuan, atau mekanisme persetujuan sebelumnya. Sebaliknya, dalam aktivitas ekstraksi otomatis, data dapat diperoleh melalui sistem yang bekerja tanpa interaksi langsung dengan individu yang datanya dikumpulkan. Kondisi tersebut menciptakan jarak antara tindakan teknis dan subjek yang terdampak, sehingga pembuktian mengenai sumber, metode, serta tujuan penggunaan data menjadi semakin kompleks.

Atas dasar tersebut, *Scraping Evidence Registry* dirancang sebagai mekanisme pencatatan terstruktur yang merekam seluruh informasi material mengenai aktivitas ekstraksi data. Berbeda dengan dokumentasi administratif biasa, registri ini berfungsi sebagai *compliance ledger* yang menghubungkan keputusan hukum sebelum *scraping* dilakukan dengan tindakan teknis yang benar-benar terjadi (Wu et al., 2019). Dengan demikian, registri tidak hanya menjawab pertanyaan “apakah perusahaan memiliki kebijakan pelindungan data?”, tetapi juga menjawab pertanyaan yang lebih spesifik: “apakah perusahaan dapat menunjukkan bagaimana, kapan, dari mana, dan untuk tujuan apa suatu data diperoleh?”

Selanjutnya penulis memperkenalkan konsep *Scraping Evidence Registry* yang merupakan suatu sistem pencatatan internal yang wajib dibangun oleh korporasi untuk merekam seluruh aktivitas ekstraksi data otomatis yang dilakukan secara langsung maupun melalui pihak ketiga. Konsep ini berangkat dari gagasan *data*

provenance, yaitu prinsip bahwa setiap elemen data harus memiliki riwayat asal-usul yang dapat ditelusuri sejak diperoleh hingga digunakan, dipindahkan, atau dihapus.

Dalam tata kelola data modern, keberadaan data tidak dapat dipandang sebagai objek yang bersifat statis. Sebuah dataset mengalami berbagai transformasi sepanjang siklus hidupnya, mulai dari pengumpulan awal, pembersihan (*cleaning*), penggabungan (*aggregation*), analisis, hingga distribusi kepada pihak lain (Fernandes et al., 2023). Tanpa mekanisme pencatatan yang mampu mengikuti perjalanan tersebut, korporasi akan mengalami kesulitan untuk membuktikan apakah penggunaan akhir suatu dataset masih memiliki hubungan yang sah dengan tujuan awal pengumpulannya. Agar berfungsi sebagai instrumen pembuktian, registri tidak dapat hanya berisi informasi umum mengenai keberadaan proyek *scraping*. Setiap aktivitas harus memiliki elemen pencatatan minimum yang memungkinkan proses verifikasi dilakukan.

Tabel 2. Elemen Minimum dalam *Scraping Evidence Registry*

Elemen Registri Kepatuhan	Informasi yang Wajib Dicatat
Identitas dan Sumber Data *(Data Source Identification)*	● Sumber atau alamat digital tempat data diperoleh (*data source URL/location*). ● Identitas pemilik atau pengelola sumber data apabila dapat diidentifikasi. ● Waktu pertama kali akses atau ekstraksi dilakukan. ● Jenis sumber data, meliputi website 14ystem, API resmi, platform pihak ketiga, atau sumber digital lainnya. ● Status akses terhadap sumber data, termasuk apakah data tersedia secara terbuka, memerlukan autentikasi, atau memiliki pembatasan teknis tertentu.
Identitas Sistem dan Metode Ekstraksi (*Crawler Identity and Extraction Method*)	● Identitas 14ystem, perangkat lunak, atau *crawler* yang digunakan dalam proses ekstraksi. ● Versi perangkat lunak atau konfigurasi 14ystem yang digunakan saat aktivitas berlangsung. ● Waktu pelaksanaan dan durasi operasi ekstraksi. ● Metode akses yang diterapkan dalam memperoleh data. • Batas permintaan (*request limitation*) untuk memastikan intensitas akses tetap proporsional. ● Perubahan konfigurasi 14ystem atau parameter teknis selama proses ekstraksi berlangsung.

Sumber: *Kreasi Penulis, 2026*

Elemen ini diperlukan karena risiko hukum sering kali tidak muncul dari keberadaan ekstraksi itu sendiri, melainkan dari metode yang digunakan. Penggunaan sistem otomatis yang menghormati batas teknis suatu sumber

memiliki karakter berbeda dengan sistem yang secara aktif menghindari mekanisme pengamanan atau pembatasan akses.

Selanjutnya, setiap aktivitas ekstraksi harus memiliki keterkaitan langsung dengan dasar hukum pemrosesan dan tujuan yang telah ditentukan sebelumnya. Maka dari itu registri harus memuat dasar pemrosesan berdasarkan UU PDP, tujuan spesifik ekstraksi, kategori data yang dikumpulkan, alasan kebutuhan pengambilan data, dan pihak yang akan menggunakan dataset. Fungsinya untuk mencegah terjadinya *purpose drift,* yaitu kondisi ketika data yang awalnya diperoleh untuk suatu tujuan tertentu kemudian digunakan untuk kepentingan lain yang tidak pernah dinilai sebelumnya (Te Molder et al., 2023). Sebagai contoh, pengambilan data untuk analisis tren pasar memiliki karakter berbeda dengan penggunaan dataset yang sama untuk membangun profil individu atau sistem keputusan otomatis. Meskipun sumber datanya sama, risiko terhadap subjek data dapat berubah akibat perubahan tujuan penggunaan.

PILAR III. Auditable Extraction Trail

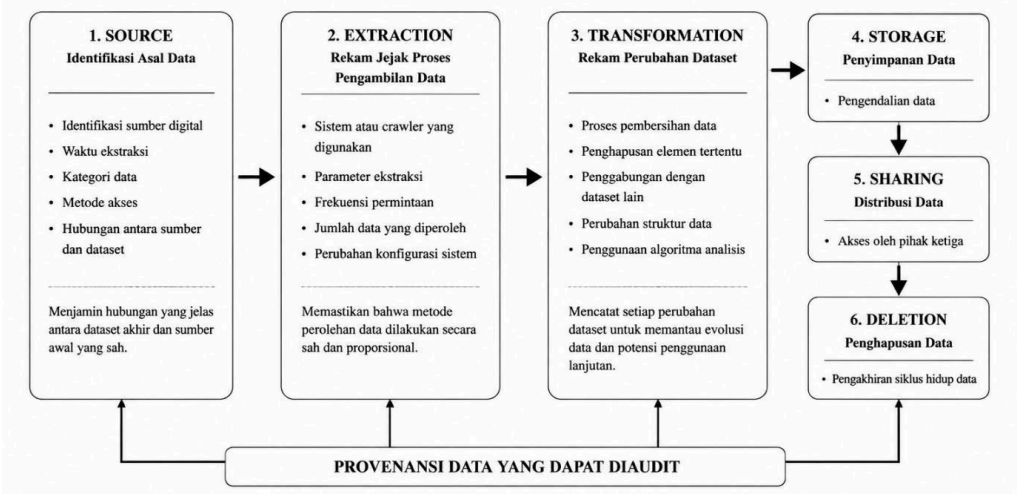
Aktivitas *web scraping* tidak berakhir ketika data berhasil diekstraksi dari sumber digital. Setelah diperoleh, dataset umumnya mengalami berbagai bentuk transformasi, mulai dari penyaringan (*filtering*), pembersihan (*cleaning*), penggabungan dengan sumber lain (*data integration*), analisis, hingga distribusi kepada pihak internal maupun eksternal. Perubahan tersebut menciptakan persoalan hukum baru karena karakter dan risiko suatu dataset tidak selalu tetap sama dengan kondisi ketika pertama kali diperoleh.

Dalam konteks tersebut, keberadaan *Scraping Evidence Registry* sebagaimana dijelaskan pada Pilar II belum sepenuhnya menjawab kebutuhan pembuktian kepatuhan. Registri mampu menunjukkan bahwa suatu aktivitas ekstraksi pernah dilakukan, tetapi belum mampu menjelaskan bagaimana data tersebut mengalami perubahan sepanjang perjalanan pemrosesannya. Oleh karena itu, *Provable Scraping Compliance* memerlukan instrumen lanjutan berupa *Auditable Extraction Trail*, yaitu mekanisme yang memastikan setiap tahapan perjalanan data dapat ditelusuri, diverifikasi, dan dievaluasi secara independen.

Karakter utama dari data digital adalah kemampuannya untuk mengalami reproduksi, modifikasi, dan distribusi dalam jumlah besar tanpa menghilangkan bentuk awalnya. Berbeda dengan objek fisik yang perubahan kondisinya relatif mudah diamati, data dapat berpindah antar sistem, dikombinasikan dengan informasi lain, dan menghasilkan nilai baru tanpa perubahan yang terlihat secara langsung. Situasi tersebut menciptakan tantangan hukum karena suatu dataset

yang awalnya diperoleh melalui mekanisme yang sah dapat berubah menjadi aktivitas berisiko ketika digabungkan dengan dataset lain, digunakan untuk tujuan baru, menghasilkan profil individu, diberikan kepada pihak ketiga, digunakan untuk kepentingan komersial yang berbeda (Supriyadi, 2023). Oleh karena itu, pembuktian kepatuhan tidak cukup berhenti pada tahap awal pengumpulan. Korporasi harus mampu menunjukkan asal-usul data (*data origin*), perubahan yang terjadi (*data transformation*), serta penggunaan akhir (*data utilization*) (Galyani et al., 2026). Dalam model *Auditable Extraction Trail*, setiap dataset diperlakukan sebagai objek yang memiliki riwayat perjalanan (*data lifecycle record*). Riwayat tersebut memungkinkan regulator maupun auditor memahami apakah perubahan terhadap dataset masih berada dalam batas tujuan pemrosesan yang telah ditentukan atau telah mengalami penyimpangan tujuan (*purpose deviation*).

Mekanisme *Auditable Extraction Trail* dibangun melalui enam tahapan utama yang menggambarkan perjalanan data sejak diperoleh hingga dihapus.



Gambar 5. Kerangka *Auditable Extraction Trail* dalam Siklus Hidup Data

***Sumber**: Kreasi Penulis, 2026*

Source: Identifikasi Asal Data

Tahap pertama merekam titik awal keberadaan data. Informasi yang perlu tersedia meliputi:

- sumber digital tempat data diperoleh;
- waktu ekstraksi;
- kategori data;
- metode akses;

- hubungan antara sumber dan dataset yang terbentuk.

Tahap ini memastikan adanya hubungan yang jelas antara dataset akhir dan sumber awalnya. Tanpa identifikasi sumber, korporasi tidak dapat membuktikan apakah suatu informasi berasal dari sumber yang sah atau diperoleh melalui mekanisme yang melampaui batas akses yang diperbolehkan.

2. *Extraction*: Rekam Jejak Proses Pengambilan Data

Tahap kedua berkaitan dengan bagaimana data diperoleh. Catatan yang perlu tersedia mencakup:

- sistem atau crawler yang digunakan;
- parameter ekstraksi;
- frekuensi permintaan;
- jumlah data yang diperoleh;
- perubahan konfigurasi sistem.

Tahap ini memiliki relevansi khusus karena legalitas *web scrapin*g tidak hanya bergantung pada objek yang diambil, tetapi juga metode yang digunakan untuk memperolehnya. Dua aktivitas yang menghasilkan dataset serupa dapat memiliki konsekuensi hukum berbeda apabila salah satunya dilakukan melalui akses terbuka, sedangkan lainnya dilakukan dengan melewati mekanisme perlindungan sistem.

3. *Transformation*: Rekam Perubahan Dataset

Setelah ekstraksi, data biasanya mengalami proses transformasi agar dapat digunakan untuk kebutuhan tertentu. Tahap ini harus mencatat:

- proses pembersihan data;
- penghapusan elemen tertentu;
- penggabungan dengan dataset lain;
- perubahan struktur data;
- penggunaan algoritma analisis.

Sebagai contoh, beberapa informasi yang secara individual tampak tidak sensitif dapat menghasilkan profil yang signifikan apabila dikombinasikan dengan sumber lain. Fenomena tersebut menunjukkan bahwa risiko data bersifat kontekstual dan tidak hanya ditentukan oleh kategori awal informasi.

Selanjutnya, penulis mencoba mengonsepkan *data storage* di dalam tahap penyimpanan untuk memastikan bahwa dataset yang telah diperoleh tetap berada dalam mekanisme pengamanan yang sesuai.

Tabel 3. Elemen Pengendalian Penyimpanan dan Distribusi dalam *Auditable Extraction Trail*

Elemen Pengendalian Penyimpanan Data (*Storage Control*)	**Elemen Pengendalian Distribusi Data (*Data Sharing Control*)**
Lokasi penyimpanan data (*data storage location*), termasuk identifikasi sistem, server, atau infrastruktur digital tempat dataset disimpan.	**Pihak penerima data** (*data recipient*), termasuk identitas unit internal, afiliasi perusahaan, atau pihak ketiga yang memperoleh akses terhadap dataset.
Pihak yang memiliki akses terhadap data (*authorized data access*), termasuk identifikasi individu atau unit yang memiliki kewenangan untuk melihat, mengelola, atau memproses dataset.	**Dasar pemberian akses** (*access authorization basis*), termasuk alasan hukum, kebutuhan operasional, kontrak, atau mekanisme persetujuan yang menjadi dasar distribusi data.
Tingkat keamanan data (*security level*), meliputi klasifikasi perlindungan, standar pengamanan teknis, serta tingkat risiko apabila terjadi akses tidak sah.	**Tujuan penggunaan data** (*purpose of utilization*), yaitu tujuan spesifik penggunaan dataset oleh pihak penerima setelah memperoleh akses.
Periode penyimpanan data (*retention period*), mencakup jangka waktu penyimpanan dan dasar penentuan masa retensi sebelum dilakukan penghapusan.	**Batasan penggunaan data** (*usage limitation*), termasuk pembatasan penggandaan, pengalihan, pemanfaatan ulang, atau penggunaan di luar tujuan awal.
Mekanisme pengendalian akses (*access control mechanism*), mencakup autentikasi,	**Tanggung jawab masing-masing pihak** (*allocation of responsibility*), termasuk kewajiban
Elemen Pengendalian Penyimpanan Data (*Storage Control*)	**Elemen Pengendalian Distribusi Data (*Data Sharing Control*)**
pembatasan hak akses, pencatatan aktivitas pengguna, serta mekanisme pengawasan terhadap penggunaan data.	perlindungan data, pengamanan, kepatuhan terhadap tujuan penggunaan, dan pertanggungjawaban apabila terjadi penyalahgunaan.

Sumber: *Kreasi Penulis, 2026*

Nilai utama dari *Auditable Extraction Trail* terletak pada kemampuannya mengubah aktivitas pemrosesan data yang sebelumnya bersifat tidak terlihat (*opaque processing*) menjadi proses yang dapat diperiksa (*auditable process*). Dalam praktiknya, mekanisme ini memungkinkan auditor atau regulator melakukan rekonstruksi terhadap suatu aktivitas scraping melalui pertanyaan seperti dari mana data berasal?, bagaimana data diperoleh?, apa perubahan yang dilakukan terhadap dataset?, siapa yang menggunakan data tersebut?, dan apakah penggunaan akhirnya masih sesuai dengan tujuan awal?**.** Pertanyaan tersebut menunjukkan bahwa audit bukan sekadar memeriksa keberadaan dokumen administratif, tetapi melakukan rekonstruksi perjalanan data secara menyeluruh.

PILAR IV. Rebuttable Presumption of Compliance

Tiga pilar sebelumnya membangun fondasi preventif dan evidensial dalam aktivitas *web scraping*. Namun, suatu model kepatuhan tidak akan memiliki daya

dorong yang optimal apabila hukum hanya menempatkan dokumentasi tersebut sebagai kewajiban administratif tanpa memberikan konsekuensi hukum terhadap korporasi yang secara nyata telah membangun sistem kepatuhan. Atas dasar tersebut, pilar keempat memperkenalkan konsep *Rebuttable Presumption of Compliance*, yaitu suatu mekanisme perlindungan hukum bersyarat yang memberikan pengakuan awal bahwa korporasi telah menjalankan kewajiban kehati-hatian (*due diligence*) apabila mampu menunjukkan bukti kepatuhan berdasarkan standar *Provable Scraping Compliance.* Perlindungan ini bukan merupakan kekebalan hukum (*legal immunity*) dan tidak menghapus kemungkinan pemeriksaan maupun pertanggungjawaban apabila ditemukan pelanggaran. Sebaliknya, mekanisme ini berfungsi sebagai instrumen insentif agar korporasi memiliki motivasi untuk membangun tata kelola internal yang transparan dan dapat diverifikasi. Konsep ini berangkat dari pemikiran *enforced self-regulation*, bahwa regulasi terhadap korporasi tidak harus selalu berbentuk larangan langsung atau intervensi represif, tetapi dapat dilakukan melalui kombinasi antara kewajiban internal, pengawasan eksternal, serta konsekuensi hukum yang proporsional (Braithwaite, 1982). Dalam konteks *web scraping*, korporasi yang mampu menunjukkan bahwa aktivitas ekstraksi telah melalui penilaian risiko, memiliki dokumentasi yang lengkap, dan menyediakan jejak audit yang dapat diperiksa, berada dalam posisi hukum yang berbeda dengan korporasi yang melakukan ekstraksi tanpa prosedur pengendalian apa pun. Perbedaan tersebut penting karena karakter aktivitas digital sering kali menciptakan kesulitan pembuktian. Tanpa mekanisme diferensiasi, korporasi yang telah berinvestasi dalam sistem kepatuhan akan memiliki posisi yang sama dengan pelaku yang mengabaikan seluruh kewajiban kehati-hatian. Situasi tersebut justru dapat menciptakan *compliance disincentive*, yaitu kondisi ketika perusahaan tidak memiliki alasan ekonomi maupun hukum untuk membangun tata kelola yang lebih baik (Bisack & Clay, 2021).

Rebuttable Presumption of Compliance harus dipahami sebagai praduga awal, bukan keputusan final mengenai legalitas suatu aktivitas *scraping*. Ketika korporasi mampu menunjukkan seluruh elemen kepatuhan, terdapat alasan hukum untuk menganggap bahwa perusahaan telah bertindak dengan itikad baik dan menerapkan langkah pengendalian yang memadai (Prayoko, 2015). Namun, praduga tersebut tetap dapat dibantah apabila terdapat bukti yang menunjukkan bahwa kepatuhan hanya bersifat formal atau manipulatif. Dalam hal ini, keberadaan dokumen registri, laporan audit, atau kebijakan internal tidak otomatis menghapus tanggung jawab apabila substansi praktiknya menunjukkan adanya penyimpangan.

Pendekatan ini diperlukan untuk mencegah terjadinya fenomena yang dikritik Michael Power sebagai *audit ritualism*, yaitu kondisi ketika organisasi menghasilkan berbagai dokumen kepatuhan semata-mata untuk memenuhi tuntutan administratif, tetapi tidak benar-benar mengendalikan risiko yang muncul (Hood, 1998). Oleh karena itu, nilai pembuktian tidak hanya ditentukan oleh keberadaan dokumen, tetapi oleh konsistensi antara dokumen, aktivitas teknis, dan dampak aktual terhadap subjek data.

Korporasi yang hanya memenuhi kewajiban dasar, seperti pencatatan sumber data dan tujuan pemrosesan, memperoleh tingkat pengakuan kepatuhan yang terbatas. Sementara itu, korporasi yang menerapkan penilaian dampak, memiliki audit trail yang lengkap, melakukan evaluasi berkala, serta mampu menunjukkan pengendalian terhadap pihak ketiga dapat memperoleh tingkat perlindungan yang lebih kuat. Pendekatan bertingkat tersebut mencerminkan prinsip proporsionalitas. Tidak seluruh aktivitas *web scraping* memiliki tingkat risiko yang sama. Ekstraksi data nonpersonal dalam jumlah terbatas tentu tidak dapat disamakan dengan pengumpulan jutaan data pribadi untuk kepentingan pemrofilan, pemasaran tertarget, atau pengembangan sistem kecerdasan buatan. Dengan demikian, perlindungan hukum tidak diberikan berdasarkan status korporasi semata, tetapi berdasarkan kualitas proses kepatuhan yang dapat dibuktikan.

Sebagai mekanisme yang dapat dibantah, perlindungan tersebut harus memiliki batas yang jelas. Praduga kepatuhan kehilangan dasar apabila korporasi terbukti melakukan tindakan yang secara fundamental bertentangan dengan prinsip akuntabilitas. Beberapa keadaan yang dapat menggugurkan perlindungan tersebut antara lain:

- memperoleh data melalui penerobosan sistem pengamanan atau akses tanpa hak;
- memberikan informasi yang tidak benar dalam dokumen kepatuhan;
- mengubah tujuan penggunaan data tanpa melakukan evaluasi ulang;
- menghilangkan atau memanipulasi rekam jejak ekstraksi;menyerahkan data kepada pihak lain tanpa pengendalian yang memadai; dan
- tetap menggunakan data setelah mengetahui adanya risiko terhadap subjek data.

Keadaan tersebut menunjukkan bahwa korporasi tidak hanya gagal memenuhi kewajiban administratif, tetapi juga gagal menjalankan prinsip accountability yang menjadi dasar utama model ini.

Dalam perspektif pertanggungjawaban korporasi, *Rebuttable Presumption of Compliance* menciptakan keseimbangan antara dua kepentingan yang sering berhadapan dalam ekonomi digital. Di satu sisi, korporasi membutuhkan kepastian hukum agar inovasi berbasis data tetap dapat berkembang. Di sisi lain, subjek data membutuhkan jaminan bahwa aktivitas ekstraksi tidak dilakukan secara eksploitatif tanpa mekanisme pengendalian. Model ini tidak menggeser beban tanggung jawab secara mutlak kepada subjek data untuk membuktikan adanya pelanggaran, tetapi juga tidak menempatkan korporasi dalam posisi rentan terhadap setiap klaim tanpa mempertimbangkan upaya pencegahan yang telah dilakukan. Korporasi yang mampu menunjukkan kepatuhan memperoleh posisi hukum yang lebih terlindungi karena telah memenuhi standar kehati-hatian yang dapat diverifikasi. Dengan demikian, *Rebuttable Presumption of Compliance* menjadi elemen akhir yang melengkapi konstruksi *Provable Scraping Compliance*. Jika tiga pilar sebelumnya memastikan bahwa kepatuhan dibangun dan dicatat, maka pilar keempat memastikan bahwa kepatuhan tersebut memiliki konsekuensi hukum yang nyata. Melalui mekanisme ini, akuntabilitas tidak lagi dipahami sebagai kewajiban untuk sekadar menyatakan kepatuhan, tetapi sebagai kemampuan untuk mempertahankan klaim kepatuhan melalui bukti yang dapat diuji.

KESIMPULAN

Penelitian ini menunjukkan bahwa kerangka hukum pelindungan data pribadi di Indonesia masih menghadapi keterbatasan dalam merespons kompleksitas aktivitas *web scraping*, khususnya karena belum tersedianya mekanisme yang menghubungkan kewajiban akuntabilitas korporasi dengan instrumen pembuktian kepatuhan yang dapat diverifikasi. Persoalan utama dalam tata kelola ekstraksi data bukan semata-mata mengenai batas antara aktivitas scraping yang sah dan melawan hukum, melainkan mengenai kemampuan korporasi untuk menunjukkan bahwa proses pengumpulan, penggunaan, dan pengelolaan data telah dilakukan secara bertanggung jawab. Sebagai bentuk rekonstruksi terhadap pendekatan kepatuhan yang bersifat deklaratif, penelitian ini menawarkan konsep *Provable Scraping Compliance* berbasis *enforced self-regulation* yang menempatkan pembuktian kepatuhan sebagai elemen utama dalam pertanggungjawaban korporasi. Model ini dibangun melalui keterhubungan antara penilaian legalitas sebelum ekstraksi (*Pre-Scraping Legality and Impact Assessment*), pencatatan aktivitas *melalui Scraping Evidence Registry*, keterlacakan perjalanan data melalui *Auditable Extraction Trail*, serta perlindungan hukum bersyarat melalui *Rebuttable Presumption of Compliance*. Konsep tersebut tidak dimaksudkan untuk membatasi inovasi berbasis data melalui

larangan yang rigid, melainkan membangun keseimbangan antara kepentingan ekonomi digital dan perlindungan hak subjek data melalui mekanisme tata kelola yang transparan, berbasis risiko, dan dapat diaudit. Implementasi lebih lanjut dari model ini memerlukan pengembangan pedoman teknis oleh otoritas pelindungan data di Indonesia yang menetapkan standar minimum dokumentasi, auditabilitas, dan evaluasi risiko bagi korporasi yang melakukan aktivitas ekstraksi data otomatis.

REFERENSI

Aho, B., & Duffield, R. (2020). Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China. *Economy and Society*, *49*(2), 187–212.

Bisack, K. D., & Clay, P. M. (2021). Behavioral responses to competing incentives and disincentives: Compliance with marine mammal protection. *Marine Policy*, *132*, 104674.

Braithwaite, J. (1982). Enforced self-regulation: A new strategy for corporate crime control. *Michigan Law Review*, *80*(7), 1466–1507.

Cahyono, N. (2021). Ekstraksi Informasi Terstruktur Profil Pengguna Website Iklan Baris. *Jurnal Buana Informatika*, *12*(1), 39–48.

Dinda Decembria. (2023). *BPS Ubah Indikator Inflasi 2024, Belanja Online Masuk Hitungan*. https://www.bloombergtechnoz.com/detail-news/23490/bps-ubah-indikator-inflasi-2024-belanja-online-masuk-hitungan/2

Dorwart, H., Zanfir-Fortuna, G., & Clarisse Girot. (2026). *China's New Comprehensive Data Protection Law: Context, Stated Objectives, Key Provisions*. https://fpf.org/blog/chinas-new-comprehensive-data-protection-law-context-stated-objectives-key-provisions/

Fernandes, A. A. A., Koehler, M., Konstantinou, N., Pankin, P., Paton, N. W., & Sakellariou, R. (2023). Data preparation: A technological perspective and review. *SN Computer Science*, *4*(4), 425.

Franklen, A. (2017). *Data scraper's case v. LinkedIn pits free speech against CFAA, DMCA*. https://www.reuters.com/article/world/data-scrapers-case-v-linkedin-pits-free-speech-against-cfaa-dmca-idUSKBN19B2WD/

Galyani, S. S., Leviza, J., & Rizky, F. K. (2026). Analisis Hukum Nasional Indonesia dan Hukum Internasional terhadap Hilirisasi Batubara Menjadi Dimetil Eter (DME) dalam Perspektif Transisi Energi Berkelanjutan, Perlindungan Lingkungan, dan Perdagangan Internasional. *Jurnal Media Akademik (JMA)*, *4*(6).

Hood, C. (1998). Michael Power, The Audit Society: Rituals of Verification, Oxford, Oxford University Press 1997£ 19.99, ISBN 0-19-828947-2. *Journal of Public Policy*, *18*(1), 103–104.

Iman, N. (2024). The fight for our personal data: analyzing the economics of data and privacy on digital platforms. *International Journal of Law and Management*, *66*(6), 774–791.

Imperva. (2026). *Bad Bot Report 2026: The Internet Is No Longer Human and It's Changing How Business Works*. Imperva.Com. https://www.imperva.com/blog/bad-bot-report-2026-bots-agentic-age/

Indonesia, S. K. R. (2018). *Soal Kebocoran Data Pengguna, Pemerintah Sebut Jawaban Facebook Belum Lengkap*. https://setkab.go.id/soal-kebocoran-data-pengguna-pemerintah-sebut-jawaban-facebook-belum-lengkap/

Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *International Data Privacy Law*, *11*(2), 125–144.

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, *13*(3), 144.

Market Glass, I. (2026). *Web Scraping Software - Global Strategic Business Report.* https://www.researchandmarkets.com/reports/6107159/web-scraping-software-global-strategic?

Paguay-Chimarro, C., Cevallos-Salas, D., Rodríguez-Hoyos, A., & Estrada-Jiménez, J. (2025). Transparency unleashed: Privacy risks in the age of e-government. *Informatics*, *12*(2), 39.

Prayoko, R. (2015). Doktrin Business Judgment Rule: Aplikasinya Dalam Hukum Perusahaan Modern. *Yogyakarta: Graha Ilmu*.

Protection, P. D. (n.d.). *The Regulations of the Supervisory Officer Personal Data Protection-Based Accountability Principle*.

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, *10*(1), 40–81.

Refslund, B. (2022). How Self-regulating can it be? Explaining Limitations in Firms' Compliance with Reflexive Regulation. *International Review of Public Policy*, *4*(4: 1), 55–113.

Shahrullah, R. S., Park, J., & Irwansyah, I. (2025). Examining personal data protection law of Indonesia and South Korea: The privacy rights fulfilment. *Hasanuddin Law Review*, *10*(1), 1–20.

Supeno, S., Rosmidah, R., & Iqbal, S. M. U. (2025). Personal data protection in review of legal theories and principles. *Journal of Law and Legal Reform*, *6*(3), 1349–1376.

Supriyadi, D. (2023). The Regulation of Personal and Non-Personal Data in the Context of Big Data. *Journal of Human Rights, Culture and Legal System*, *3*(1), 33–69.

Susana Rita Kumalasanti. (2026). *Sidang di MK, Pemerintah Bantah Ada Kekosongan Pengawasan Perlindungan Data Pribadi.* https://www.kompas.id/artikel/sidang-di-mk-pemerintah-bantah-ada-kekosongan-pengawasan-perlindungan-data-pribadi

Syailendra, M. R., Lie, G., & Sudiro, A. (2024). Personal data protection law in Indonesia: Challenges and opportunities. *Indon. L. Rev.*, *14*, 175.

Te Molder, R., Fedorova, M., Dubelaar, M., & Lestrade, S. (2023). The principle of purpose limitation in data-driven policing: A guiding light or an empty shell? *New Journal of European Criminal Law*, *14*(4), 512–533.

Wibowo, T. M., & Karyati, C. M. (2021). Implementasi Data Mining Menggunakan Algoritma Apriori pada Transaksi Penjualan Produk (Studi kasus Toko Kanaya Rooftop Garden). *Explore*, *11*(2), 127.

Wu, Z., Williams, A. B., & Perouli, D. (2019). Dependable public ledger for policy compliance, a blockchain based approach. *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 1891–1900.

Zulfikar, R., Lukviarman, N., Suhardjanto, D., Ismail, T., Dwi Astuti, K., & Meutia, M. (2020). Corporate governance compliance in banking industry: The role of the board. *Journal of Open Innovation: Technology, Market, and Complexity*, *6*(4), 137.