ALLEN & OVERY

PRIVACY LAWS & BUSINESS

Data Scraping

Karishma Brahmbhatt, Dr Philipp Raether and David Smith

12 October 2021

Agenda

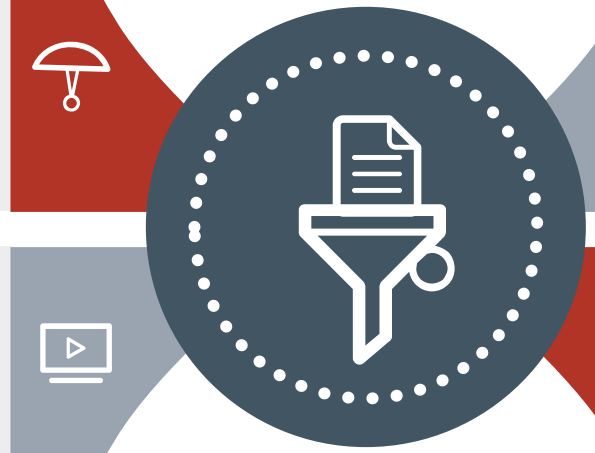
1. Data scraping – the basics
2. Data protection considerations
3. Data ethics
4. Data scraping – cases in point
5. Other legal issues
6. Top tips

Data scraping – the basics

What are we talking about?

Same same but different…
- Web scraping
- Screen scraping
- Web data harvesting

1



What

Method of retrieving large volumes of publicly available data from various online sources. Typically involves use of technology to 'scrape' the data.

2

How

Data is extracted, aggregated and combined into a more digestible format for the user

3

Why

Various purposes related to the generation of business value or efficiencies…

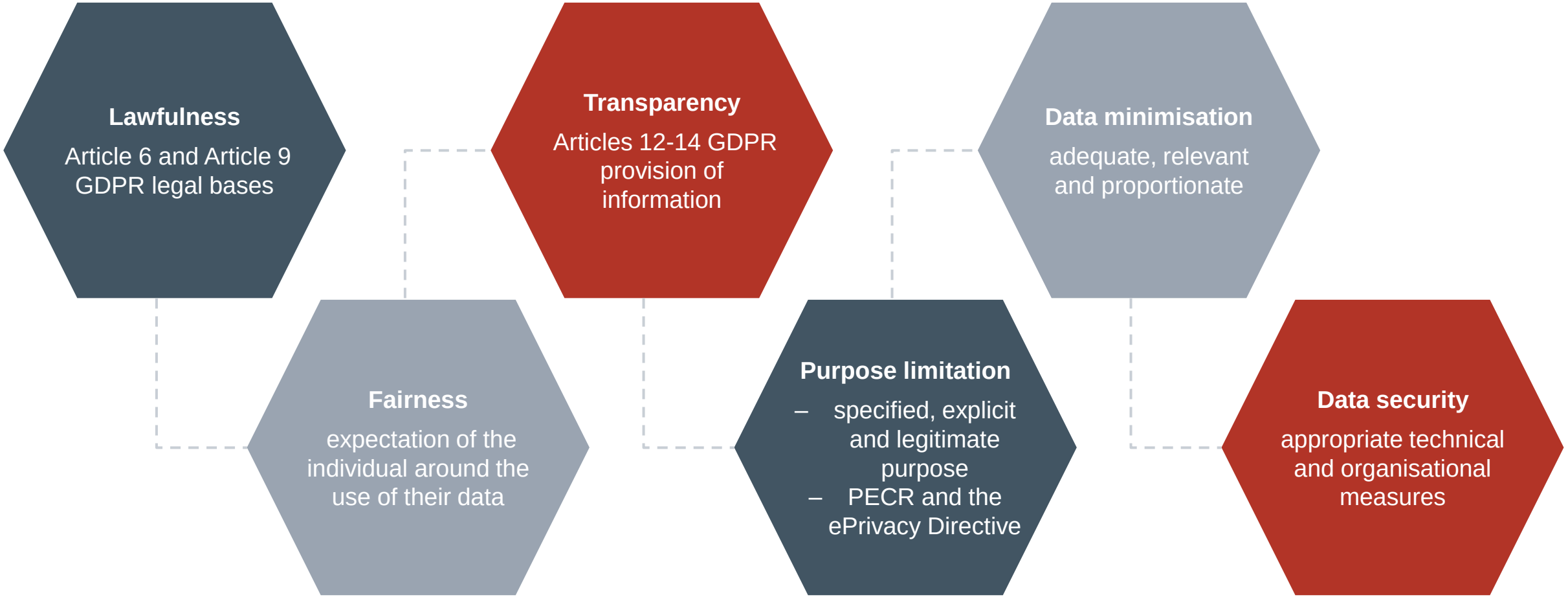
4

Examples of use cases

- Direct marketing
- Lead/sale generation
- Recruitment
- Customer risk assessments
- Identifying trends
- Industry insights
- Competitor monitoring
- Reputation monitoring

Data protection considerations

Key data protection principles



Regulatory guidance on data scraping

The CNIL's guidance on data scraping

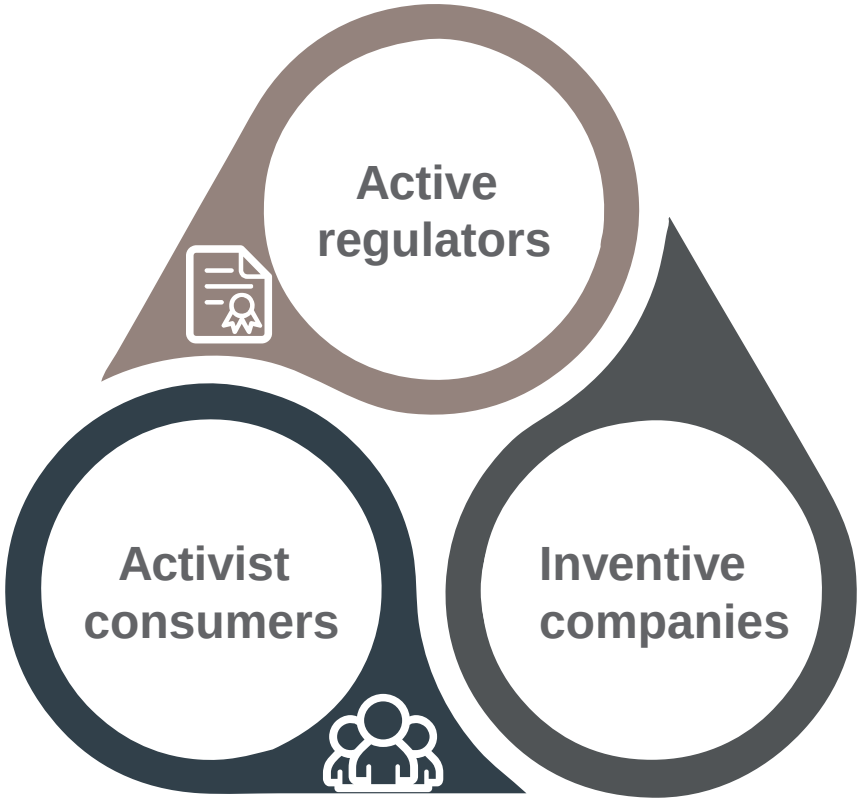
Personal data that is publicly accessible cannot be freely re-used by companies without the individual's consent for direct marketing purposes

The individual's consent must be freely given, specific, informed and unambiguous

The individual's right to object or withdraw consent must be respected

Data ethics

Data ethics – what is it and why should you care?



Just because you can, doesn't mean you should ...

Key points to note

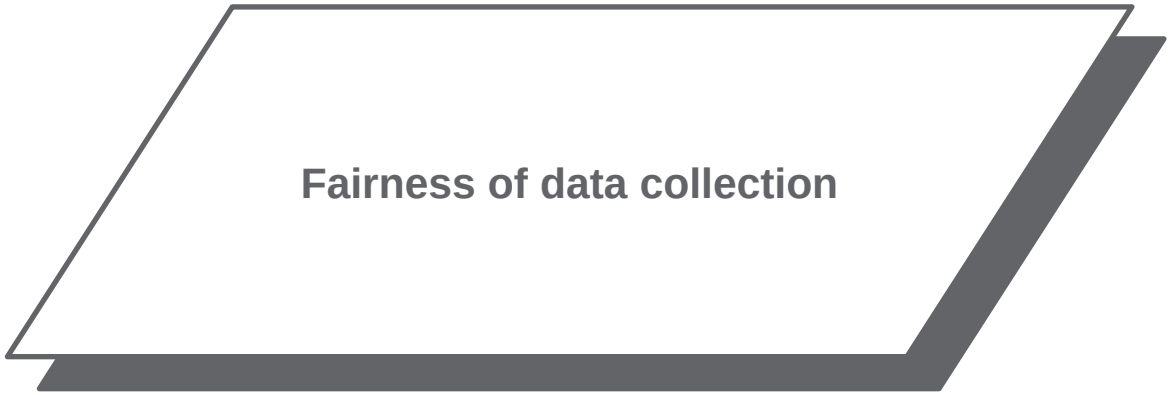
Use of legislation to penalise organisations that fail to process data in an ethical manner

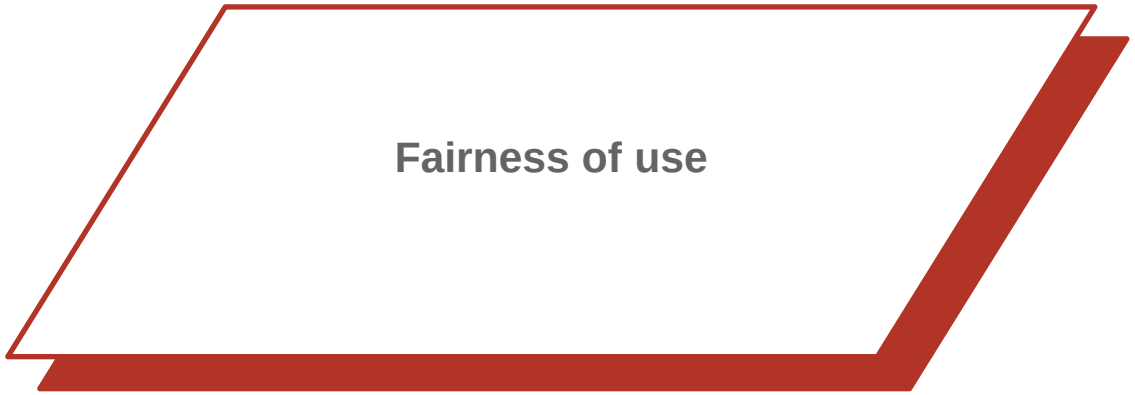
What is considered to be an acceptable use of data is fluid and changing

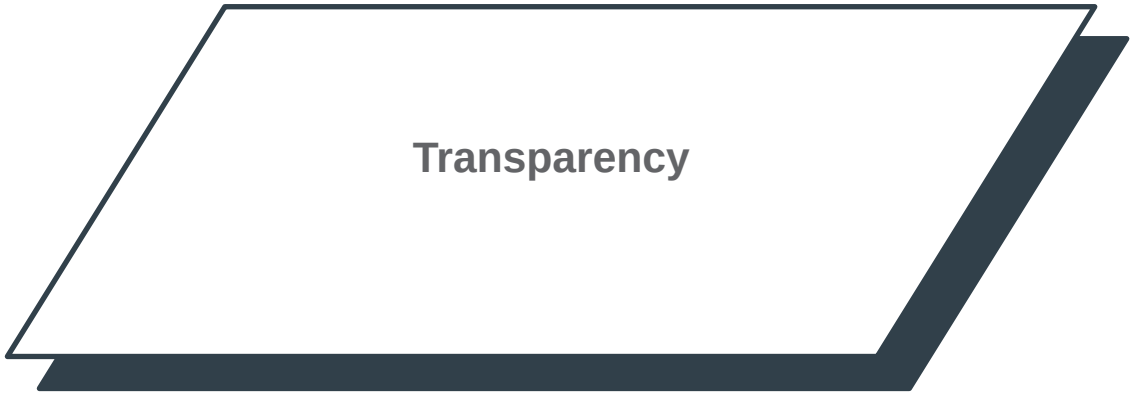
Embedding data ethics is one of the most complex risk management challenges

Data ethics principles are broadly aligned with privacy laws

Key considerations include…







Data scraping – cases in point

Data scraping – A case in point (US)

Who	LinkedIn
When	September 2019
What	– *HiQ v Linkedin* was held in the US Ninth Circuit Court. – LinkedIn sought an injunction to prevent HiQ, a data scraping company, from scraping data from its website. – LinkedIn argued that HiQ did not have authorisation to scrape its users' data as it had put measures in place to prevent this. – The Court refused the injunction and held there was no violation of the CFAA because the data was publicly accessible and as such HiQ did not access the data 'without authorisation'. – The Supreme Court has asked the Ninth Circuit Court to re-consider its decision based on the ruling in the case of *Van Buren v United States*. – The parties have agreed a trial date of 5 December 2022.

Data scraping – A case in point (Global)

Who	LinkedIn
When	Spring 2021
What	– Personal data of 700 million LinkedIn users was scraped from its website and posted for sale on a hacking forum. – The personal data included full names, phone numbers, email addresses, geolocation records and personal and professional experiences and backgrounds. – LinkedIn issued a statement saying that the data posted for sale comprised "*an aggregation of data from a number of websites and companies. It does include publicly viewable member profile data that appears to have been scraped from LinkedIn. This was not a LinkedIn data breach…*". – A similar incident occurred in April 2021 where personal data of 500 million users was leaked.

Data scraping – A case in point (Ireland)

Who	Facebook
When	April 2021
What	– Personal data of over 530 million Facebook users was published on a hacking forum. – The personal data included, for example, phone numbers, full names, locations, bio information, birthdates and in some cases email addresses. – Facebook said the data was scraped prior to September 2019 by malicious actors but that it had rectified the issue. – The Irish DPC made enquiries into the leak – A class action against Facebook is being commenced by Digital Rights Ireland.

Data scraping – A case in point (France)

Who	NESTOR
When	December 2020
What	– NESTOR SAS scraped data from professionals on LinkedIn to compile a mailing list for its meal preparation and delivery services. – CNIL found a breach of Articles 12 and 13 GDPR as the individuals were not informed about the collection of their data for this purpose. – CNIL also found that NESTOR had failed to seek consent of the individuals to direct marketing. – CNIL fined NESTOR €20,000 for violation of the GDPR and ePrivacy Directive.

Data scraping – a case in point (Spain)

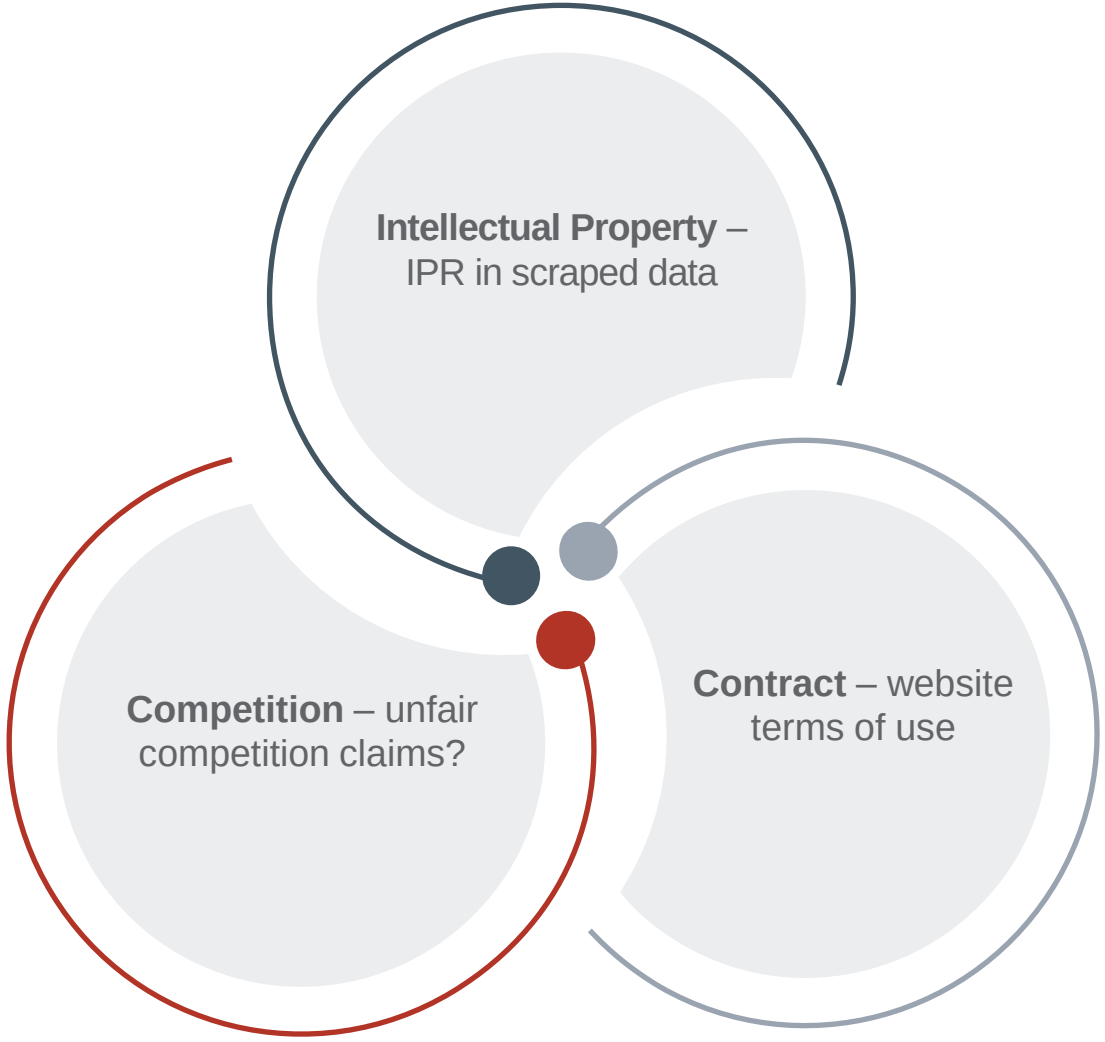
Who	EQUIFAX
When	April 2021
What	– Equifax, a multinational consumer credit reporting agency scraped data from public sources for using in credit reports information from tax authorities and other government sources. – The personal data included information about individuals' outstanding debts. – 96 people complained about their data being used in Equifax's reports and around four million people could be affected. – Spain's data protection authority fined Equifax for violation of the Spanish GDPR and imposed a fine that amounted to EUR 1m. – Spain's data protection authority ordered Equifax to stop collecting data this way and to delete the data already collected.

Data scraping – a case in point (UK)

Who	Ticketmaster
When	November 2020
What	– AI company Inbenta Technologies Inc. was contracted to incorporate a chat bot onto Ticketmaster's website. – The chat bot was used on various pages of Ticketmaster's website, including the payment page. – In February 2018 malicious code infiltrated the chat bot and several banks reported fraudulent transactions. – The malicious code scraped personal data provided by the user on the payment page, which included payment card names, numbers, expiry dates and CVV numbers. – Ticketmaster was fined £1.25 million by the ICO in 2020 as it had failed to put appropriate security measures in place to prevent a cyber-attack on the chat bot. – Ticketmaster has appealed the fine on the basis that it has not breached its security obligations under the GDPR. The appeal has been stayed until the outcome of the High Court proceedings commenced by 795 Ticketmaster customers; and a separate Part 20 action commenced by Ticketmaster against Inbenta.

Other legal issues

Is it lawful?


Intellectual Property – IPR in scraped data
Competition – unfair competition claims?
Contract – website terms of use

Unfair competition, intellectual property… or unauthorised access?

Who	Facebook
When	May 2017
What	– Power Ventures was a social media aggregator using data scraping to allow users to amalgamate their profiles from various sites. It continued to operate after refusing to sign Facebook's developer terms of use and despite an IP address block. – December 2008: Facebook brings an action alleging copyright and trademark infringement, violation of the CAN-SPAM Act, and unauthorised access under the Computer Fraud and Abuse Act (CFAA) and the California Penal Code § 502. – February 2012: District Court decides the **§ 502**, **CFAA**, and **CAN-SPAM Act** claims in Facebook's favour and dismisses the other claims. – December 2016: Ninth Circuit reverses CAN-SPAM finding but affirms other violations. – May 2017: District Court determination of remedies.

Contract as the final line of defence?

Who	**Ryanair**
When	**January 2015**

What	– Preliminary ruling of the CJEU in Case C-30/14, *Ryanair v PR Aviation*, where the Netherlands Supreme Court had asked whether the operation of the Database Directive 96/9 extended to online databases protected neither by copyright nor by the sui generis database right.
	  – SGDR: statutory industrial property right in the EU (+ UK) aiming to protect the contents of a database against reuse by third parties who did not invest time and effort into its creation, distinct from copyright protection of database structure as an intellectual creation.
	  – Third parties may utilise insubstantial parts of the database, subject to conditions.
	  – The CJEU determined that the lawful user rights do not apply to databases which are covered by neither copyright nor the sui generis database right.
	– In national proceedings, Ryanair successfully argued that the use of automated screen scraping by the price-comparison site PR Aviation was in breach of its website terms and conditions.
	  – Contract law can apply where IP law would not.

Top Tips

Top tips

If you are scraping data

- Verify the nature and origin of the data being scraped.
- Minimise data collection.
- Issue a privacy notice to the individuals whose data is being scraped.
- Carry out a Data Protection Impact Assessment if necessary.

If you are using a data scraping provider

- Conduct due diligence.
- Implement a data processing agreement.

If your data is liable to being scraped

- Ensure appropriate monitoring and safeguards.
- Consider other methods of protection.
- Implement technical measures to prevent data scraping.

Speakers

Karishma Brahmbhatt
Allen & Overy LLP
Senior Associate - London
Tel +44 20 3088 2158
karishma.brahmbhatt@allenovery.com

Dr Philipp Raether
Allianz SE
Group Chief Privacy Officer
philipp.raether@allianz.com

David Smith
Allen & Overy LLP
Special adviser - London
david.a.smith@allenovery.com

Questions?